PROCEEDINGS OF THE 12TH INTERNATIONAL CONFERENCE ON APPLIED INNOVATIONS IN IT

VOLUME 12

Issue 1

Proceedings of the 12ᵗʰ International Conference on Applied Innovations in IT

Volume 12 | Issue 1

Koethen, Germany
7 March 2024

This journal contains publications of the International Conference on Applied Innovations in IT (ICAIIT), which took place in Koethen March 7th 2024. The conference is devoted to problems of applied research in the fields of automation and communication technologies. The research results can be of interest for researchers and development engineers who deal with theoretical base and the application of the knowledge in the respective areas.

Content

Communication Technologies and Hardware

Proceedings of the 12th International Conference on Applied Innovations in IT (ICAIIT)

Turing Machine Development for High-Secure Data Link Encoding in the Internet of Things Channel

Victor Tikhonov[1], Abdullah Taher[2], Serhii Tikhonov[3], Kateryna Shulakova[1,4],
Vlad Hluschenko[1] and Andrii Chaika[1]

[1]*Department of Computer Engineering and Information Systems, State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, 65023 Odesa, Ukraine*
[2]*Department of Electronic and Communication Engineering, College of Engineering, Al-Qadisiyah University, 58002 Qadisiyah, Iraq*
[3]*Department of Computer Systems, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine*
[4]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*
victor.tykhonov@suitt.edu.ua, abdallahqays@gmail.com, od.sergii.tihonov@gmail.com, katejojo29@gmail.com, vlad.gluschenko.97@gmail.com, chayka.and@gmail.com

Keywords: Internet of Things, Information Security, Data Link Encryption, Data Framing, Formal Grammar, Turing Machine, Packet-Based Streaming.

Abstract: The work considers the issues of data stream encoding in the IoT systems and networks in the context of information security provision, with particular focus on the data link layer in wireless and wired telecommunication channels. The problem state of the art shows a great progress in the sphere of IT cybersecurity based on Advanced Encryption Standard. However, new tasks and related threats emerge, such as unmanned mobile devices of mass disposable with remote control, for which known approaches are not reliable and efficient enough. In this respect, the common algorithms of data link frame encoding are studied, and original formalization of typical frames proposed for static and dynamic representation. An original method introduced for high-secure data encoding with variable frame structure. A formal grammar of an abstract Turing machine (TM) is developed for data link encoding, which is based on ternary line-signaling and three-bit command system. Constructed typical frame-patterns with the use of TM syntax. General principles are formulated for the high-secure frame encoding with variable structure for packet-based streaming on the data link layer with the use of TM algorithm. The results of the work intend to improve the mobile objects cyber-threats protection, as well as to remote vehicle control and other IoT real-time applications.

1 INTRODUCTION

The term "Internet of Things" (IoT) was first used in 1999 for physical objects with sensors connected to a packet network. Currently, IoT means the transition to a qualitatively new stage of the Internet evolution, which services extend to many devices, sensors and actuators [1].

IoT sensor networks are close to the mass user of Internet services, and this will dominate the Internet access networks. The further development of the Internet in the segments of concentration and distribution of traffic is formulated in the ITU concept of Next Generation Networks (NGN), which integrates a variety of data types and network services based on IP [2].

The core of the Internet is a transport system of optical lines based on high-speed coherent optical communications. Global satellite mobile communications networks such as Starlink are also being actively implemented. Each of these aspects has its own characteristics and challenges [3].

This paper considers information encoding at Data Link Layer (DLL) of IoT computer systems interfaces, for which a big challenge is personal data protection. Vulnerability links of IoT systems are radio channels, controller's software, traffic hubs. IoT cyber protection is complicated by a large amount of data and documents of various types, and each of them is a potential source of threats. Files "exe" and "com" may contain virus program bookmarks, which in themselves are not virus carriers, but refer to Web resources that contain them [4].

The variety of threats and countermeasures, as well as the growing risks of cyber-attacks in the modern world, require further theoretical and practical researches on the cyber protection methods of the IoT telecommunication channels, considering the characteristics of network interfaces at different levels of computer system architecture.

The Internet of Things cyber security architecture implies, that wireless connectivity is integrated into IoT devices and sensors, enabling them to transmit data to the network. The wired Ethernet along with the wireless WiFi links provide a standardized interface for these devices, allowing them to communicate with gateways or directly with cloud platforms [5].

Up to days, a ground-breaking progress achieved in network security technologies based on Advanced Encryption Standard (AES). However, new problems have arisen, such as unmanned mobile objects, for which the known approaches are not sufficient enough. Therefore, new researches needed in the IoT-cybersecurity.

The transmission of frames in a telecommunication channel can be taken as a sequence of commands and data for some abstract automaton that operates according to a certain algorithm and converts the incoming stream code into the structural-parametric code of individual frames.

The common principles of device operation, that transform information according to a certain algorithm, are studied by the automata theory. An important section of automata theory is the TM as a mathematical model that reduces the logical structure of an arbitrary processor to its basics (i.e. formalized rules for building grammatical structures on a set of alphabet symbols) [6].

In connection with the above, an actual scientific-applied task arises in the IoT field, that is, development the methods and algorithms for secure data transmission in the local networks and communication channels. This work intends to develop a formal grammar for the high-secure data encoding in the IoT data link channels on the base of an abstract TM algorithm.

Section 2 analyses the IoT cybersecurity state of the art. Section 3 formulates the objectives of the work. Section 4 presents formalization of the data link framework. Section 5 develops a TM grammar and general principles of high-secure packet-data encoding based on the ternary line-signaling and three-bit TM-command system. Section 6 summarizes results of the work.

2 THE IOT CYBER SECURITY STATE OF THE ART

Methods for transmitting, processing and protecting data in telecommunication systems and networks depend on the selected mathematical model for representing digital information. Let $A=\{a_n\}$, $n\in[1, N]$ be a set of discrete alphabet symbols used for interaction between network participants; $p(n, k)$, $(n,k)\in[1, N]$ - a priori probability of erroneously perceiving the symbol a_n as if it were a_k; $Q=\{p(n, k)\}$ - quantum uncertainty matrix.

The representation of digital information using a quantum uncertainty matrix Q corresponds to the quantum data model (QDM) in promising digital IT technologies that has been actively developing in recent years [7].

In the case of $p(n, k)\to 0$ when $n \neq k$, the matrix Q can be approximated by a diagonal matrix $Q\approx diag(P)$, where $P=\{p(n)\}$ is a quantum credibility vector for symbols, $p(n)$ is a priori probability of correct perception of the symbol a_n during the interaction of the parties. This case of representation the digital information using a quantum credibility vector P corresponds to a reduced quantum data model (RDM).

Finally, in the case of $p(n, k)\to 0$ for $n\neq k$ and $p(n,k)\to 1$ for $n=k$, the RDM-model turns into a classical data model (CDM). The difference between these three models (QDM, RDM and CDM) is rather conditional and not clearly defined. Commonly known IT-technologies primarily use the classical CDM-model of digital information representation.

We will consider here the cybersecurity issues with respect to the classical CDM-model of digital information representation, in which data reception-transmission errors are possible, but not critical.

Cybersecurity is a broadly used term, whose definitions are highly variable and often subjective. It includes organization, resources, processes, and structures used to protect cyberspace systems from occurrences that misalign property rights [8]. The cybersecurity also refers to improving the integrity of information management systems or infrastructure and addressing present and emerging challenges [9].

In this work, we will understand the cybersecurity for IoT networking as "a set of technologies, processes and practices to protect and defend networks, devices, software and data from attack, damage or unauthorized access" (Laboratory for Open Systems and Networks, Jožef Stefan Institute, Ljubljana [10].

The cybersecurity is becoming complex because of the exponential growth of interconnected devices, systems and networks. This is exacerbated by advances in the digital economy and infrastructure, leading to a significant growth of cyberattacks with serious consequences. In addition, researchers report the continued evolution of nation-state-affiliated and criminal adversaries, as well as the increasing sophistication of cyberattacks, which are finding new and invasive ways to target even the savviest of targets. This evolution is driving an increase in the number, scale and impact of cyberattacks, and necessitating the implementation of intelligence-driven cybersecurity to provide a dynamic defense against evolving cyberattacks and to manage big data. Advisory organizations, such as the National Institute of Standards and Technologies (NIST), are also encouraging the use of more proactive and adaptive approaches by shifting towards real-time assessments, continuous monitoring and data-driven analysis to identify, protect against, detect, respond to, and catalogue cyberattacks to prevent future security incidents [10].

One of the Internet protocol architects D. Clark pointed out cybersecurity as a serious problem of the global network. The Internet lacks of built-in security, and its shortcomings have resulted in a decreased ability to accommodate new technologies. We might just be at the point where the utility of the Internet stalls – and perhaps turns downward. It's time to rethink the Internet's basic architecture and start over with a fresh design. This is not about building a technology innovation but about architecture – pulling the pieces together in a different way. Improving the Internet is not so much about delivering the latest cool application, it's about survival. L. Peterson, a computer scientist at Princeton University, thinks that we've been on a track of incrementally making improvements to the Internet and fixing problems that we see. We see vulnerability, we try to patch it. That approach is one that has worked for 30 years. But there is reason to be concerned. Without a long-term plan, if you are just patching the next problem you see, you end up with an increasingly complex and brittle system. It makes new services difficult to employ. It makes it much harder to manage because of the added complexity of all these point solutions that have been added. T. Leighton (member of the President's Information Technology Advisory Committee) believes herewith, that there are more and more holes, and more resources are going to plugging the holes, and there are less resources being devoted to fundamentally changing the game, to changing the Internet [11].

During the global Network development, authoritative institutions proposed various models of Internet architecture. Each of these models reflects a specific vision of the organization on the current problems of the state and prospects of the Network, and from this point of view, defines multi-level interfaces for the open systems interconnection. Each layer and type of Internet model has its own cybersecurity vulnerabilities [12].

Commonly known are the following Internet models: 7-layer OSI/ISO reference model (1978-80), 4-layer TCP/IP model by IETF (1978-80), 3-layer NGN/ITU-T model (2004), 4-layer NGN/ITU-T model (2012), 7-layer IoT model by IoT World Forum (2014), 4-layer NIST-SP model (2016), 3-layer Industrial IoT model by IIC (2022).

Among the seven listed above Internet architecture models, three latter have been produced by new emerged institutions, and they reflect transition of traditional Internet to the Internet of Things. Similar to early Internet model's evolution by shifting the number of layers in the "7-4-3" pattern, the latter IoT-models reproduce this pattern but not in all. The last two models are of particular interest.

The NIST-SP model issued in 2016 by the Information Technology Laboratory (ITL) at the US National Institute of Standards and Technology (NIST) as a Special Publication 800-183 with particular focus on distributed computer systems security. The document offers "an underlying and foundational science to IoT based on a belief that IoT involves sensing, computing, communication, and actuation" [13].

Unlike previous Internet models, the NIST-model includes the human factor as a critical element of the IoT system security; it formulates five system primitives of the generalized Network of Things (NoT): Sensors, Aggregators, Communications, external Utilities (e-Utilities), Decision Triggers.

The first four ones refer to technical entities of the model, whereas Decision Trigger refers to human factor, which acts similarly to aggregator a special case of it. The model does not specify what is or is not a 'thing. In physical space, things consider humans, vehicles, residences, computer, switches, routers, smart devices, road networks, office buildings, etc. In virtual space, they consider software, social media threads, files, data streams, virtual machines, virtual networks, etc. An important primitive of the model is communication channel is a medium by which data is transmitted (wireless, wired, verbal, etc.). Since data is the "blood" of a NoT, communications are "veins" and "arteries" [13].

The latest known model of the Industrial Internet of Things (I-IoT), proposed by the I-IoE consortium in 2022, is quite consistent with the NIST model, and in a certain sense develops it. It contains three hierarchical categories: End Devices (sensors and actuators), I-IoT platform and Users (devices, individuals); two primitives of the NIST-model (Aggregators and Communications) merged into the "I-IoT platform".

Both IoT latest models consider communication channels as a critical element for the overall system security chain. There is a wide variety of telecommunication channels. In this paper we will focus on packet-based digital communication channels that operate in non-stationary or extreme conditions. This primarily concerns wireless channels for communication with mobile devices and objects, e.g. in WiFi networks.

Information in a digital packet network is supplied via a communication channel in the form of sequences of symbols (letters of the alphabet and syntactic signs), similar to grammatical text in school. Letters are combined into separate words according to spelling rules. Words are combined into sentences according to syntactic rules. A sequence of sentences forms a message (for example, a section of a document). An ordered set of messages forms a document file [14].

In a packet telecommunications channel, the role of grammatical sentences is played by frames, e.g. wired Ethernet or wireless WiFi frames, that operate on the physical and data link OSI-layers. The OSI network layer deals with IP-packets which carry distinct messages. The four upper OSI-layers are responsible for the file-documents delivering.

We will highlight three main components of the data protection process in the communication channel, regardless of the level of packet data transmission during remote interaction between the parties. These are:
a) method of plane text encrypting;
b) method of authentication and key provision;
c) protocol for packet generation.

Next, we will focus the IoT cybersecurity state of the art by tracking the WiFi wireless technology evolution, which is the most popular today in the Internet of Things applications. Over the past three decades, WiFi technology has evolved from a simple WEP-40 interface (Wired Equivalent Privacy, 1997, still widely used today), up to the modern WPA3 interface (WiFi Protected Access 3, certified since 2018), which is considered now the most advanced, and is only beginning to be deployed.

There are four main WiFi encrypted security protocols currently valid and available on the telecom market: WEP (IEEE 802.11 standard released in 1997 and ratified by WiFi Alliance in 1999); WPA (2003, IEEE 802.11i − an amendment to the original 802.11 standard), WPA2 (2004) and WPA3 (has been in use since 2004, WiFi Alliance certifying began since 2018). The features and characteristics of these interfaces are discussed below:

A) WEP interface:

1) *Method of plane text encrypting.* WEP is based on RC4 stream cipher (Rivest Cipher 4, designed in 1987 by R. Rivest from the American computer and network security company RSA) for data encryption with 8÷2048 bit key. Actually, WEP supports the RC4-encryption with 64/128-bit keys (40/104-bit static secret key-part + 24-bit initialization vector as a variative open key-part). Due to its speed and simplicity, RC4 is now widely used in Secure Socket Layer (SSL) and Transport Security Layer (TSL) of the IEEE 802.11 wireless LAN. However, because of its vulnerabilities, RC4 is no longer recommended for critical applications [15].

RC4 has 4 variants as follows.
- SPRITZ - building the cryptographic hash functions and deterministic random bit generator.
- RC4A - proposed to be faster and stronger than the average RC4 cipher, though it was found to have not truly random numbers in cipher.
- VMPC - Variably Modified Permutation Composition − found to have not truly random numbers, like RC4A.
- RC4A+ - an advanced version of RC4A that is longer and more complex than RC4 and RC4A, but is stronger as a result of its complexity as well.

RC4 operates on a stream of data byte by byte with 64/128/256-bit keys. Yet, a flaw was found in RC4 where the 128-bit encryption key could be cracked in seconds. Another RC4-vulnerability was discovered in 2013 while it was being used as a workaround for a cipher block chaining issue (2011), while RC4 did not use itself the block-chain operational mode. It was found a way around RC4, with only a slight increase in processing power necessary in the previous RC4 attack. Now RC4 is considered a legacy method.

2) *Method of authentication and key provision.* The WEP-encryption for any local WiFi connection implies predefined static secret 40/104-bit key along with the 24-bit dynamic open key (initialization vector IV), that is (pseudo) randomly generated for each frame. The WEP secret key can be provisioned automatically by the WiFi access point for a WiFi LAN host by arbitrary chose among the set of default keys when host authorization. Up to four default keys can be pre-set in the WiFi access point; default key is identified by their number 0, 1, 2, or 3 in the WEP-frame extended header. The WEP secret key can also be assigned as the WiFi LAN access password of 5/13 ASCII characters or 10/26 HEX characters (40/104 bits).

3) *WEP protocol for packet generation.* WEP was introduced as part of the original 802.11 standard released be IEEE in 1997, and ratified by WiFi Alliance in 1999 [16]. It specifies the set of medium access control (MAC) and physical layer (PHY) protocols for implementing wireless local area network (WLAN) communication. The WLAN MAC frame structure consist of a MAC header, frame body, and a frame check sequence (FCS). The 802.11 MAC frames can be either protocol version 0 or version 1. The algorithm of WEP-frame design is described in [17]. It is applicable for any frame or a packet that consist of a header and a payload:

Input: Header, Payload;
ICV = CRC(Header, Payload);
Data = (Payload, ICV);
IV = Initiation_Vector generator;
{**Key_Nr, Secret_key**} = Key_looup generator;
WEP_seed = (IV, Secret_key) // RC4 key;
Payload_Script = RC4(Data, WEP_seed);
WEP_Frame = (Header, IV, Key_Nr, Payload, ICV).
  Output: WEP_Frame.

B) **WPA-interface.** The WPA protocol (WiFi Protected Access) is similar to WEP, but it is more secure due to increased *256-bit encryption* key. Besides, it uses an enhanced RC4-based TKIP cipher (TKIP) that includes three new features: a) 64-bit *message integrity check* (MIC) instead of CRC to re-initialize the sequence number each time when a new temporal key is used; b) *packet sequencing control* to protect against replay attacks (packets received out of order are rejected by AP; c) *per-packet key mixing* function that combines the secret key with the initialization vector before passing it to the cipher.

WEP, in comparison, merely concatenated the initialization vector to the root key, and passed this value to the RC4 routine. Compare to WEP authentication (Open system/Shared key), WPA-personal provides a *pre-shared key* (PSK) for Small Office Home Office (SOHO) wireless LAN.

C) **WPA2-interface.** WPA2 is a security standard for wireless networks based on the AES – Advanced Encryption Standard technology (NIST, 2001). It is used with the IEEE 802.11a/b/g/n/ac WiFi standards as the successor to WPA and has superseded WEP encryption [18]. Compare to WPA with the TKIP cipher, the WPA2 standard implemented an advanced cipher CCMP (Counter Mode with Cipher Block Chaining Message Authentication Code Protocol) that uses the AES 256-bit encryption algorithm; latter is one of the most robust encryption methods that is commercially available today.

While it is theoretically true that AES 256-bit encryption is harder to crack than AES 128-bit encryption, AES 128-bit encryption seems not to been cracked yet. Starting in August 2023, AES256-bit in cipher block chaining mode (AES256-CBC) will be the default encryption mode across all applications using Microsoft Purview Information. AES-256 encryption is virtually uncrackable using any brute-force method. It would take millions of years to break it using the current computing technology and capabilities [19].

D) **WPA3-interface.** The WPA3 standard was announced in 2018 by WiFi Alliance as a replacement to WPA2; it is now considered a cutting-edge security protocol to the market, and is a mandatory certification for WiFi CERTIFIED™ devices. WPA3 brings better protections to individual users by providing more robust password-based authentication, even when users choose passwords that fall short of typical complexity recommendations. This capability is enabled through Simultaneous Authentication of Equals (SAE, IEEE 802.11s standard issued in 2011 and superseded in 2012 when became part of the IEEE 802.11 standard of 2012). The technology is resistant to offline dictionary attacks where an adversary attempts to determine a network password by trying possible passwords without further network interaction.

The new standard uses 128-bit encryption in WPA3-Personal mode (192-bit in WPA3-Enterprise). The WPA3 standard also replaces the pre-shared key (PSK) with SAE as defined in IEEE 802.11-2016 resulting in a more secure initial key-exchange in personal mode. The WiFi Alliance also claims that WPA3 will mitigate security issues posed by weak passwords and simplify the setting up devices with no display interface. WPA3-Enterprise also offers Extensible Authentication Protocol for transport layer security (EAP-TLS) using Elliptic Curve Diffie-Hellman (ECDH) exchange and Elliptic Curve Digital Signature Algorithm (ECDSA) with 384-bit elliptic curve.

At the same time, the widespread adoption of WPA3 is a gradual process with potential limitations and challenges. Among them are weak backward compatibility and limited device support, vulnerabilities in implementation (errors or weaknesses in the deployment and configuration of WPA3 can undermine its security benefits). Regular firmware updates and adherence to security best practices are crucial to mitigate these potential weaknesses. However, when WPA3 gains wider adoption and attention from potential attackers, new vulnerabilities may emerge. The continuous discovery of security flaws and the need for timely patches and updates are ongoing challenges in maintaining the security of any wireless protocol, including WPA3. The weaknesses mentioned above do not undermine the overall benefits that WPA3 brings to wireless security [20].

The IoT-cybersecurity state of the art may conclude the following.
1) To date, principal results have been achieved in the field of security of networks and telecommunication channels. The top AES-256 encryption method is used to protect digital data, which is virtually uncrackable. The wireless channels of networks and systems commonly use a set of WiFi interfaces of different complexity and resistance to attacks like WEP, WPA, WPA2, while the most advanced WPA3 interface is now being actively implemented. At the same time, new threats and challenges emerge. One of them is provoked by an open packet structure for any WiFi standard; other things being equal, this factor facilitates to hack information at the data link layer.
2) Despite great achievements in the field of cybersecurity, we are now experiencing a major crisis in the global security system.

3) New operational and tactical tasks have arisen, for the solution of which conventional IT technologies turned out to be insufficiently effective. One of these unsolved problems is the creation of a simple and reliable interface for interaction with unmanned mobile objects of disposable use.

Based on the analysis of the surveyed publications, the objectives of this work are formulated in the context of protecting information on the data link layer in distributed systems of the IoT with the use of deterministic digital data model.

3 OBJECTIVES OF THE WORK

This work aims to develop a formal grammar for high-secure data link encoding with variable framework in the Internet of Things channel.
To achieve this, the following objectives are set:
1) Formalization the static and dynamic data link framework.
2) Construction a TM formal grammar for data structuring.
3) Developing a TM for high-secure data link encoding with variable framework.

4 FORMALIZATION THE DATA LINK FRAMEWORK

The wired and wireless LAN-technologies are widely used on the Data Link Layer (DLL) in today's IoT. As it was spoken above, the algorithms of DLL encryption and framing are tolerant to concrete frame structure, whenever the static frame consist of a header and a payload body. The WiFi wireless frame structure differs from the 802.3 wired standard, and it is more complicated. Therefore, we will further discuss the principal issues of DLL framing that refer to 802.3 wired LAN. The specifications DIXv2.0 (aka Ethernet-2) and IEEE-802.3 (aka Ethernet-3) historically remain two independent basic standards, that are not directly compatible. The later specification IEEE-802.3x unites both these two standards on the principle of backward compatibility. It defines the structure and parameters of the Data Link frame, invariant to the method of its transmission (structural-parametric code), as well as its mapping into a sequence of digital symbols for transmission over the communication channel – the stream code of the frame.

The Data Link streaming code is formed due to the frame structure and parameters, in accordance with related cable type and line-signaling. At the physical layer binary linear codes are popular (with two-digit physical signals, aka MII-codes) or three-digit (MLT-3, 100 BASE-TX). A simple binary code of type MII is the interference-resistant NRZ-PAM2 with a small frequency band; it is popular in the SerDes communications, which convert parallel code into serial one and vice versa. Its drawback is a DC component and weak synchronization in monotonous symbol sequences. The 10Mbitps interfaces implement a binary self-synchronized inverse Manchester code (MC-code of the IEEE 802.3 convention), in which any bit is at the same time a clock signal; the original MC-code is referred to as Thomas convention encoding. The interfaces 100BASE-T4/T1/1000BASE-T1 use the 8B6T block code in the 3-character PAM3 alphabet. For instance, the 100Base-T4 specification (IEEE 802.3u standard) uses the 4-twisted pair cable of category 3; two of the four pairs are oriented in opposite directions, and the other two are switched in the direction of transmission. At any moment, three of the four pairs are transmitting data, and one is listening to the line for collisions. For every 2 cycles, three twisted pairs transmit a block of 6 ternary symbols ($3^6 = 729$ numbers), which encode 256 octet-numbers according to redundant table scheme. The 1000BASE-T interface implements the PAM5 code, in which 2 positive and 2 negative amplitude levels encode 2 bits of information in a cycle (4-character alphabet), and zero-signal is not used. The 2.5GBASE-T interface uses the PAM16 code.

The DLL stream contains the structure-parametric components, embedded into a digital symbols sequence due to a certain formal grammar, considering the properties of the communication line and the counteracting parties' protocols. It can be shaped in a permanent commands sequence for an abstract digital processor. Let introduce the static Frame Parameter Code (**FPC**) as a system of vectors:

$$FPC = \{[AD], [ET], [D]\}, \qquad (1)$$

where [AD] = (DMAC, SMAC) − vector of physical destination and source addresses (DMAC, SMAC) with 2×6=12 bytes in the network adaptor RAM; [ET] − vector of the DLL-frame type of 2, 5, 6 or 10 bytes in RAM, depending on the first two bytes, and besides, on the fifth byte of this vector; [D] = (DL, DAP) − vector of the payload data of the total size DL bytes in RAM, allocated by the Data Access Point (DAP).

The static **FPC** code is a temporary data record, which is necessary to generate a dynamic Frame Streaming Code (**FSC**) for the data transmission over the physical communication line between any two adjacent nodes within a LAN. The **FPC** code plus CRC-checksum (4 bytes) in the stream code of the 802.3x frame, must be 64÷1518 bytes length, (RFC 894, 1984). Hence,

|**FPC**|+4≤1518, or (2÷6+|ET|+DL)+4≤1518.

From this follows the data field length bound DL≤1502−|ET|. When |ET|=2, we have 0≤DL≤1500; at |ET|=10 we get (0≤DL≤1492) bytes.

Consider the three following **FPC** formats depending on the size of the ET field.

1) |ET|=2 bytes; this corresponds to the DIXv2.0 (Ethernet-2) frame, and the values ET=1536÷65534 encode the frame type.

2) |ET|=5 or 6 bytes; this corresponds to an IEEE-802.3 (Ethernet-3) frame, where the ET field is expanded into a D data field by 3 or 4 bytes of the LLC header (IEEE 802.2 specification). The first two bytes in the ET field determine the number of bytes of useful data DL=0÷1497. In the streaming code, the data field DL of less than (64−18) =46 bytes, has to be padded up to the minimal size of 46 bytes. The two least significant bits "xx" in the control byte of the extended ET field of the IEEE-802.3 frame define one of the three following modes of the LLC protocol:

- **U** (Unnumbered frame, "xx"="11") − datagram mode of transmission of frames as independent packets of the channel level, LLC=3 bytes, |ET|=2+3=5 bytes;
- **I** (Information frame, "xx"="00" or "10") − transmission of frames by multiplex streams over logical connections with LLN numbers (Logical Link Numbers) in the sixth byte of the extended ET field; LLC=4 bytes, |ET|=2+4=6 bytes;
- **S** (Supervisor frame, "xx"="01") − frames of control messages contained in the sixth byte of the extended ET field; LLC=4 bytes, |ET|=6 bytes.

3) |ET|=10 bytes; this indicates the SNAP frame of the IEEE-802.3x standard (Sub-Network Access Protocol) with DL=0÷(1500−8) =1492 payload data size in the first 2 bytes of extended ET-header. Though, the DL-size in the stream code of such a frame must be at least (46−8) =38 bytes, otherwise it is padded to 38 bytes.

Unlike the IEEE-802.3 frame, the first 3 bytes of the LLC header in the extended ET-field of the SNAP standard frame contain a dummy LLC with a fixed value $(AA`AA`03)_{16}$ or $(AB`AB`03)_{16}$. The first 3 of the next 5 bytes of the SNAP header contain a 24-bit OUI (IEEE Organizationally Unique Identifier) that identifies the manufacturer and vendor of the network equipment. The last 2 bytes contain the 16-bit frame type number (similar to the ET field in the DIXv2.0 frame header). The expansion of the LLC header has been due to the gradual increase in frame types, despite the limited 6-bit type code in the 3rd byte of the previous LLC version. Since the hexadecimal values $(AA`AA)_{16}$ and $(AB`AB)_{16}$ for the first two bytes of the LLC were previously idle (reserved), they were used as "commands" for the extension of the LLC header.

The static DLL frame structure **FPC** (Frame Parameter Code) introduced above reflects the necessary and sufficient set of parameters for the node's interaction within a local network. An **FPC** code of a frame is formed at the OSI data link layer for temporal storage in the network adapter RAM for their further transfer over the physical layer link, or vice versa, to be passed to the network layer driver.

The **FPC** code cannot be directly transmitted into the channel under various technical circumstances. Instead, it must be injected into a digital symbol sequence of the Frame Streaming Code (**FSC**). The latter includes additional fields: an inter-frame guard interval (IFG), synchronization preamble (PRB), Start Frame Delimiter byte (SFD), and the Cyclic Redundancy Checksum (CRC, 4 bytes) from the vectors AD, ET and D of the **FPC**-code.

The **FSC**-code is limited in minimal size in order to distinguish the frame from the channel garbage when frames collision occurs. For that reason, the net-size data vector D of the **FPC** code is mapped into extendable gross-size payload data field PL of the **FSC**, while the set of parameters (AD, ET, PL, CRC) is kept for 64 bytes length or more (RFC 894). Thus, the Frame Streaming Code **FSC** (without the IFG guard interval) must be not less than 72 bytes.

Let define a *dynamic Data Link frame structure* **FSC** (*Frame Streaming Code*) as vector-system with a minimal size of 72 bytes:

$$\textbf{FSC} = \{[PRB]_7, [SFD]_1, [\textbf{FPC}]_{60}, [CRC]_4\}, \quad (2)$$

where PRB is preamble (7 bytes), SFD is the Start Frame Delimiter (1 byte), CRC is the checksum (4 bytes), and **FPC** is the spoken above static frame structure (Frame Parameter Code) of minimal 60 bytes length: $\textbf{FPC} = \{[AD]_{12}, ([ET], [D])_{48}\}$.

5 TURING MACHINE FOR DATA LINK ENCODING WITH VARIABLE FRAMEWORK

The TM is commonly understood as an abstract digital processor, that is simple, intuitive, generic and formalizes computation performed by a human mind.

Most concrete digital processors perform the so called "program code", or just a "program" (consequence of commands and data allocated at ROM/RAM memory with the 8-bit structure of bytes); here, one byte is the elementary syntax unit (symbol) in program text. Instead, an abstract TM-processor can handle the program texts with symbols of 1 bit and more.

Generic TM operates with the ternary alphabet: an abstract "space" and two "letters" (0 and 1). Each ternary symbol carries the $\log_2(3) \approx 1.585$-bit information. While constructing the TM formal grammar for the DLL-interface, we will use the two formalized above objects: **FPC** (1) and **FSC** (2).

The first formal object (**FPC**) is the TM-basis of the high-secure protocols generation for data transmission within an arbitrary local area network (LAN). The DLL-interaction presumes the common 6-byte MAC-addressing, with no care of the DLL-standards limitations in the streaming code (e.g. the minimal frame length, fixed **FSC** shape, mandatory guard interval IFG, CRC format etc.).

The second formal object (**FSC**) is the TM-prototype for simulation the conventional data transmission and receiving via the spoken above 100BASE-T4 standard with ternary line-signalling PAM3 and 8B6T block encoding. The modern physical channels implement ternary symbol encoding also by the other modulation methods (e.g. 3-PSK in coherent communications).

5.1 The TM-Formal Grammar Definition

Consider the TM grammar (**TMG**) as (3):

$$\textbf{TMG} = \{\textbf{A}, \textbf{Syn}, \textbf{Sem}\}, \quad (3)$$

where $\textbf{A}\{-, 0, 1\}$ is the TM ternary alphabet of abstract line-signalling symbols: syntax sign "space" $(-)$, and two "letters" (0, 1); **Syn** is the formal TM-syntax; **Sem** is the unformal TM-semantic.

Next, a TM-code **TMC**{**S**} is defined on the set **S** of ternary alphabet symbols $s \in (\textbf{A})$, along with the TM-grammar syntax:
1) An TM-word is a non-empty sequence of "letters" 0 or 1 separated by one or more "spaces".

2) An TM-command is a TM-word with no more than three "letters"; others words are TM-data.

3) The TM-command basic set is shown in Figure 1. The overall number of TM-commands equals 14, including two 1-letter words (0 and 1), four 2-letter words (00,01,10,11) and eight 3-letter words (000÷111). The first 13 command of the basic set are tags; the command 14 is reserved for extended commands (13 additional commands plus next ones).

4) The common punctuation tags follow the syntax rules of ordinary spoken languages. Two or more "spaces" are equivalent a single one; parenthesis (), brackets [] and braces {} must be paired; repetition of tags prohibited, etc.

5) Nesting of tags {{...}} or [[...]] prohibited.

6) Tags {...} define an opened/closed sequence of multiple framing objects FPC and/or FSC.

7) The **TMC** codes are consequently transmitted in sessions {...}:

$$\{...\} \{TMC_1, TMC_2 ... TMC_N\} \{...\}-, \ N \geq 0.$$

8) The **TMC** may include the components:

$$\textbf{TMC} = [AD(...); DAT(...); CRC(...)],$$

where AD is the address vector (DMAC, SMAC); DAT is data vector (Type, Payload); CRC is checksum of AD and DAT. All the **TMC** components could be dummy, as well as the **TMC** itself.

№	Code	Syntax	Command
1	0	.	Dot
2	1	,	Comma
3	00	;	Semicolon
4	01	:	Colon
5	10	(	Left Parenthesis
6	11	)	Right Parenthesis
7	000	AD	Address Vector
8	001	DAT	Data
9	010	CRC	Checksum
10	011	[	Left Bracket
11	100	]	Right Bracket
12	101	{	Left Brace
13	110	}	Right Brace
14	111	$	Extension

Figure 1: The basic set of TM commands.

In contrast to common grammars with clearly defined rules, the TM-semantic is beyond the syntax formalism. Instead, it is rather to be introduced by related patterns collection. In practice, the semantic of a formal grammar can be embodied in appropriate algorithms and related computer programs made by human mind, or due to machine learning techniques, based on the sufficient AI training sequence of human

mind design. Further on, we bring several TMC patterns of data transmission in a digital channel:

A) Pattern 1. IoT Real-Time data Streaming.
The telemetry real-time data stream (RTS) is characteristic in a sequence of short messages (one or several bytes each one) and fast label-switching. The following pattern fits this.

$$\left[\begin{array}{l} \textbf{TMC} = ... \{RTS.RTS.RTS\} ... \\ RTS = [AD \ 0001 \ DAT \ xxxxxxxx \ CRC]. \end{array} \right.$$

The telemetry stream is designed as an open sequence of sessions {...}; each session transmits three RTS segments (i.e. frame parameter codes). Each **FPC** includes the 4-bit label (0001) of the predefined logical connection, 8-bit data field, and 1-bit CRC (parity bit "p"). In case of data field "10101011", the **FPC** TM ternary code is

$$-011-000-0001-001-10101011-010-1-100-$$

Here, 37 ternary symbols of the **FPC** line-code occupy $37 \times \log_2(3) \approx 58.6$-bit, while carrying 8-bit payload with information efficiency $8/58.6 \approx 0.137$, compare to DIXv2.0 framing with 10 times worse efficiency of $8/672 \approx 0.012$:
12(IFG)+7(PRB)+1(FSD)+6(DMAC)+6(SMAC) + +2(ET)+46(PLD)+4(FCS) = 84bytes = 672 bit.

B) Pattern 2. DIXv2.0 Frame Simulation.

$$\left[\begin{array}{l} \textbf{TMC} = ... \{\textbf{FSC.FSC.FSC}\} ... \\ \textbf{FSC} = DMAC \ SMAC \ ET \ Data \ CRC. \end{array} \right.$$

A session { } is presented by 3 consequent frames, each of 5 semantic parameters in regular format. The Data field has 0.5÷1500 byte without padding. The extra **FSC**-fields (IFG, PRB, SFD) are not needed. According to above TM syntax, the data-words are of 4 and more bits.

C) Pattern 3. Frame Parameter Transmission.

TMC = {AD (DMAC, SMAC) DAT(Data)}.

A session { } is presented by 2 pseudo-frames: the first one carries DMAC, SMAC, and the second − 0.5÷1500 byte of Data field without padding. This pattern can be used in case of low signal-to-noise rate (SNR) in the channel.

5.2 Principles of TM-Based High-Secure Encoding

The high-secure data link encoding with TM-encryption technique includes the following.
1) The set of TM-codes allows variety of TM-based formal grammars for encoding the

DLL-streaming data by permutation the rows in the TM-codes table. So, a pre-shared permutation number can be used as an extra secret key for DLL-stream encryption, in addition to conventional encryption keys.

2) Specific unmanned disposable mobile objects with remote control may allow to provide a simple and low-cost but unique encryption scheme at the data link layer using the one-time secret keys generated and pre-shared before their launching.

3) Based on the DLL-frameworks, a family of untypical frames can be outworked and secure-indexed to provide each object-mission by a unique subset of frameworks and related secret keys. This makes the channel harder to hack.

6 CONCLUSIONS

The IoT-cybersecurity state of the art indicates significant progress in network security both in wired and wireless communications. Nevertheless, new tasks and security challenges emerge, for which known approaches are not sufficient enough.

The work introduces an original TM formal grammar for high-secure data link encoding to increase the wireless and wired channel protection in special IoT-applications.

General principles proposed for data link encoding by the use of TM-encryption with variable framework to enable an unpredictable dynamic switching of the streaming frames structure and make the communication channel harder to hack.

The concrete TM-based security mechanisms are the objects of further researches. The results of the work intend to apply in unmanned disposable mobile objects with remote control and special IoT-systems.

ACKNOWLEDGMENTS

REFERENCES

[1] D. Hanes, et al., "IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things," Cisco Press, USA, 2017, pp. 543.

[2] International Telecommunication Union, "ITU-T's Definition of NGN," 2004.

[3] S. Wali, "The Internet: Opportunities and Challenges," 2022.

[4] M. Aqeel, et al., "A Review of Security and Privacy Concerns in the Internet of Things (IoT)," 2022.

[5] IEEE Computer Society, "The Internet of Things (IoT) to Cloud: Computing and Communication Gateways," 2023.

[6] J.E. Hopcroft, et al., "Introduction to Automata Theory, Languages, and Computation," 3rd ed., Pearson Education, 2008, pp. 554, [Online]. Available: https://openlibrary.org/books/OL24357442M/Introduction_to_automata_theory_languages_and_computation.

[7] F.J. Duarte and T.S. Taylor, "Quantum entanglement in matrix notation," in "Quantum Entanglement Engineering and Applications," 2021, ch. 7, pp. 1-9.

[8] D. Craigen, et al., "Defining Cybersecurity," 2014, [Online]. Available: https://www.researchgate.com/publication/267631801_Defining_Cybersecurity.

[9] F. Schiliro, "Towards a Contemporary Definition of Cybersecurity," 2022, [Online]. Available: https://arxiv.org/ftp/arxiv/papers/2302/2302.02274.pdf.

[10] R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions," Elsevier, Information Fusion, vol. 97, 2023.

[11] F.D. Talbot, "The Internet is broken," MIT Technology Review, 2006, [Online]. Available: https://www.technologyreview.com/2006/02/15/229667/the-internet-is-broken/.

[12] A.C. Franco da Silva and P. Hirmer, "Models for Internet of Things Environments. A Survey," Information, vol. 11, no. 10, p. 487, 2020.

[13] J. Voas, "NIST Special Publication 800-183. Networks of Things," Computer Security Division, Information Technology Laboratory, 2016.

[14] C. Vleugels, "Human communication," [Online]. Available: https://osuokc.edu/sites/default/files/documents/arts/SPCH-1113-Speech-%26-Communication.pdf.

[15] V. Kota, "RC4 Cipher," 2021, [Online]. Available: https://www.linkedin.com/pulse/rc4-cipher-venkata-siva-naga-sai-kota?trk=public_profile_article view.

[16] D. Ghimiray, "Wi-Fi Security: WEP vs WPA or WPA2," [Online]. Available: https://www.avast.com/c-wep-vs-wpa-or-wpa2.

[17] G. Singh, "WEP Encryption and Its Vulnerability in Detail," 2023, [Online]. Available: https://tbhaxor.com/wep-encryption-in-detail.

[18] "How WPA2 differs to WEP and WPA," [Online]. Available: https://www.nfon.com/en/get-started/cloud-telephony/lexicon/knowledge-base-detail/wpa2.

[19] IDERA Inc., "What is AES 256-bit Encryption?", 2024, [Online]. Available: https://www.idera.com/aes-256-bit-encryption.

[20] "What is WPA3 vs. WPA2?" [Online]. Available: https://www.portnox.com/cybersecurity-101/wpa3.

Method of Grouping Complementary Microservices Using Fuzzy Lattice Theory

Oleksandra Dmytrenko[1] and Mariia Skulysh[1,2]

[1]*Institute of Telecommunication Systems, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiskyi Avenue 37, 03056 Kyiv, Ukraine*
[2]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*
olexandra.dmytrenko@gmail.com, mskulysh@gmail.com

Keywords: Microservices, Cloud Energy Efficiency, Fault Tolerance, Shared Instance Group, Cluster.

Abstract: This paper contains ideas on how to optimize the costs of running a microservice system. Currently, there is much done to provide high fault tolerance of a microservice and a system as a whole. Cloud providers come up with new ways to guarantee the high speed of newly launched instances. This leads to a ubiquitous run of redundant servers with possible cold or hot standby mode. This is often crucial because the ability to use some applications quickly and on time can be important to many users, potentially saving lives. At the same time, it's important to prioritize ecological preservation and minimize overuse of the Earth's resources. In the context of cloud, and specifically, server computing, that would involve using resources in a way that extends their lifespan, minimizing the creation of slowly decomposing waste, and avoiding excessive energy consumption. Cloud providers, such as Amazon, Google, and Azure, discard millions of underused hardware units due to the necessity of ensuring service guarantees to their customers. In the article, method to optimize the usage of servers by organizing microservices in complementary sets are described. As a result, server resources will be used most efficiently. The method of grouping the microservices can be likened to the principles of lattice theory. The ideas in the article could be useful for the systems like Kubernetes scheduler in the stage of picking the right set of instances to run a new microservice, or to cloud providers. As a result, less energy and hardware resources will be used to provide the same quality of fault tolerance.

1 INTRODUCTION

These days, many businesses try to make their program products scalable in order to handle high loads efficiently. That's why designing a program as a group of microservices that can be replicated across multiple instances is becoming increasingly popular. To support this kind of structure, a microservice manager platform is needed. Docker, Kubernetes, and different cloud platforms offer assistance in managing load and scalability by deploying new microservice instances that could help overcome the limitations of a single instance.

In most cases, application loads are predictable. The impact of sudden load surges has a significant influence on system stability. The main idea is to distribute microservices among server groups first taking into account prioritizes balancing the overall load of the group, then followed by predictability and ensuring that a certain bearable maximum that the server group can provide is not exceeded.

2 GROUPING COMPLEMENTARY MICROSERVICES ACCORDING TO PRINCIPLES OF FUZZY LATTICE THEORY

2.1 Fuzzy Lattice Theory in Relation to Microservices

In this paper, the complementarity of microservices and demonstrate how their coordinated utilization can significantly reduce wasteful resource usage will be discussed. Let's consider a definition of complementary objects from the lattice theory. A lattice b is complemented to lattice a is a bounded lattice that satisfies

$$a \vee b = 1 \text{ and } a \wedge b = 0 \qquad (1)$$

In linguistic terms, the initial formula can be interpreted to signify the absence of overlapping resource utilization, wherein concurrent usage of a single resource by both microservices does not occur. Simultaneously, both microservices collectively

leverage all available resources during operation. A complementary lattice need not necessarily be unique. A bounded lattice means that the minimal element is 0 and the maximal element is 1 [1]. The lattice theory contains definitions of a relatively complemented lattice and orthocomplementation, which could also be useful in showing the parallelism with the idea of organizing microservices into groups.

Let us expand this idea to microservices running on a single resource group, or simplifying, on one server. It means that the microservices are complementary when, if running together, they fully utilize the server's capacity, and if they are off, the server remains idle.

To generalize, taking in consideration that it might be difficult to find an ideal complementary type of microservice, let's assume that b can be recursively replaced with a set of two other microservices, c and d, that follow the same rule as in (1) by sharing the total load in proportions that will never exceed 1 and that won't fully interact with a:

$$a \vee (c \vee d) = 1, c \wedge d = 0, a \wedge d = 0, a \wedge c = 0 \quad (2)$$

This kind of recursion can occur multiple times, meaning there is no limit to how many microservices a resource group can contain. The main principle is to achieve the efficient usage of the group.

2.1.1 Adding Fuzziness

When dealing with resource load, incorporating fuzziness can be more appropriate. To that end, we modify the definition of the microservices to allow partial resource load. It can be proportional or not, but the total load at any moment won't exceed the full possible load.

Let us define the following terms:
- C: a certain processor characteristic, such as processor load, channel throughput, user number, or memory usage;
- T_j: a certain time interval, e.g. an hour or "morning", during which C is measured. The cycles are repeated every 24 hours. Certain differences can be also made for holidays and weekends;
- j: a time unit. We consider that time is divided into reasonable units. For example, what interests us are the hourly measurements during the day or fuzzy definitions of "morning", "day", "evening" and "night" time.
- i: an index of a microprocessor that runs from 1 to some positive integer N.
- $\mu T(C_i)$: the degree of a certain characteristic C of the microprocessor i during time T.

If microservices 1 and 2 are the two complementary microservices, then $\mu_T(C_1)$ and $\mu_T(C_2)$ are the degrees of presence of a given characteristics C during a selected time T. Together they should form a load of a server resource approaching to a full possible load. Based on given definitions above, we can rewrite a formula of fuzzy complementation presented in [2] in the following way:

$$\mu_{Tj}(C_2) = 1 - \mu_{Tj}(C_1) \quad (3)$$

Every microservice has its usual, average, minimal, and maximal loads. When referring to (2) it means that disregarding the exact load of one microprocessor, the load of the second one should be adopted. The complementation should be based using the definition: maximal intersecting load at any same time should not exceed full load. In the words of formulas, that is:

$$\max(\mu_{Tj}(C_1), \mu_{Tj}(C_2)) < 1 \quad (4)$$

Till now, we have spoken about having 2 complementary microservices. With the same success, we can extrapolate the formulas to more elements. To bound N - the maximal number of the microservice instances that could run on one server group, let us take into consideration a number of threads K on the group processor(s). Then $i \in [1, K]$. Making generalization, the formulas (3) and (4) would be the following:

$$\mu_{Tj}(C_N) = 1 - \mu_{Tj}(C_1) - \mu_{Tj}(C_2) - ... - \mu_{Tj}(C_{N-1}) \quad (5)$$

$$\max(\mu_{Tj}(C_1), \mu_{Tj}(C_2),..., \mu_{Tj}(C_{N-1}), \mu_{Tj}(C_N)) < 1 \quad (6)$$

2.1.2 Illustration of an Idea

To illustrate the idea of creating balanced resource groups, consider a simple example. Suppose there are 2 companies with private hostings: a ticket company, which experiences a high load during the daytime, and an online casino, whose main activity takes place at night. To optimize the number of hardware resources spent for hosting these 2 enterprises, a common hosting could be considered. The same servers would be reused by an application that has to handle more load at a particular moment. The total common load remains the same, but fewer resources for running both programs are used. The transition in resource usage by a specific application would be relatively gradual.

Potential spikes in load for each application during non-standard times could occur due to sales and special offers, as well as unpredictable sudden events, such as heavy weather conditions and

disasters. If the applications provide services in the same region, evidently, in the second case people would be more likely to prioritize buying tickets and saving their lives over playing in casinos, both during the day and at night.

On the other hand, the first case is more difficult, necessitating either a localized or a comprehensive resolution. A potential local resolution involves negotiating between companies regarding staggered timing for special offers. This possibility is unlikely, as it would require the companies to be managed by friends or the same person. This approach also entails considering the risk of one or both companies having their income undermined. A general resolution implies providing additional resources that are rarely used, but which are crucial for safe and stable working conditions. Such resources would also be useful when components of the hardware degrade. Greater hardware durability can be anticipated when operating below maximum capacity. Additional resources would also serve to provide longer life expectancy for the servers.

Increased resource allocation and simultaneous higher loads for both applications could raise concerns about profitability of shared hardware. This is where the concept of dividing an application into microservices would act as a resource optimizer. Every microservice has different tasks and different general loads in the same period compared to other microservices of the same application. In online stores, people spend significantly more time on product selection than on ordering and payment procedures. As a result, much fewer login activities are to be expected than filtering goods activities. Hence, combining low-load microservices from one application with high-load microservices from another application is essential for achieving a balanced group.

As a result, when having high activities in both applications at the same moment, a balanced group of microservices will not require a substantial amount of resources. As a consequence, a general margin of safety for a shared space among several microservice applications will be reduced compared to using private spaces. At the same time, general fault tolerance will be higher [3].

2.1.3 Comparison of the Example with the Fuzzy Lattice Theory

In the context of lattice theory, a complementary microservice that achieves the same maximum state of processor usage should be identified, ensuring that only one microservice is active at any given moment.

Let us denote by a a microservice of the ticket company that works in the daytime, and by b, a microservice of the casino that works at night. Let us also denote by 1 a capacity of the server/cluster/resource group that runs both microservices and at the same time is a potential maximal safe capacity of one of the microservices. Likewise, let 0 denote a state when everything is idle or which should not happen.

Following (1) and its fuzzy counterpart (3), we can say that we want to achieve state 1 when either a or b is working. The state of partial load is also acceptable: if it's 80% of a, then it's no more than 20% of b, and similarly for (4) and (6). States where one is 60% loaded and the other one is 10% loaded are also acceptable, as they comply with the formulas for fuzzy sets. In that case, a server could run something else in addition during this time.

The maximum resource capacity is designed only to support partial load. In case when both microservices need more resources, then replication should be activated and a state of 0 must be prevented.

Taking into consideration that there may be several complementary paths, it is worth mentioning that a microservice that could fit as a dual pair to the first one is not unique. Not only ticket companies work at night. A bank microservice could similarly fit the casino one in this regard.

2.2 Grouping Strategies

The example provided is reasonable, however, having a standard set of rules or considerations for grouping microservices would be beneficial. A group by itself implies that the physical servers are located as close as possible and are connected in a local group via wires, or this is a single supercomputer with plenty of resources. Below, we list criteria that may influence decision-making.

2.2.1 Security or Multi-Tenancy

Microservices can be grouped by security reasons or user access rights. This is the most secure way of grouping. However, the challenges posed by multi-tenancy can be solved in several ways, depending on specific needs and risks.
1) Strong division of common space among several running applications can be performed by virtualization.
2) Containerization can be used in addition to an already running operating system.

3) One can launch applications from different users in terms of one operating system, where every user has its predefined space.

4) In a public cloud environment, it is possible to segregate the same physical server into secure logical spaces for every user or tenant, thereby ensuring a high level of security.

5) Single tenancy, which requires separate microservices for separate user groups or users, can also be provided in terms of one group.

At the same time, extremely stringent security requirements are relatively uncommon. To provide an adequate level of security, it is not necessary to launch microservices that require the same user access rights in the same group. Unless it is required, it would be more effective to base the division on alternative guidelines, from a resource efficiency perspective.

In Kubernetes, multi-tenancy often involves many teams sharing the same cluster. It also has a so-called SaaS tenancy, where multiple clusters with different applications are provided to a single user team [4].

2.2.2 Shared Resources

Most applications are not standalone but have separate resources such as databases. In a microservice architecture, it is common to have a shared schema registry, DTO classes, domains, libraries, and protocols that are stored separately, exclusively for use by two or more microservices. Grouping can also be based on using the same shared resource.

If several microservices communicate with each other often, it is beneficial to have a system design with a focus on maximal failure prevention, particularly during data transmission. Data inconsistencies caused by partially sent and processed data can be challenging to roll back. Addressing such inconsistencies may necessitate actions from each microservice that processed the information. This is described in the SAGA pattern. Alternatively, creating specialized clean-up procedures could be an option. These procedures would know all potential failure scenarios and would be capable of making manual rollbacks, fetching information from the backup files, or recalculating updated values and writing previous consistent values.

Rollback procedures in a microservice architecture are risky. There is no guarantee that some part of the application won't read temporary incorrect values. Both strategies are time-consuming and include multiple operations, rather than completing everything in a single step. The human factor is also present because when the code is updated, a developer might forget to accordingly modify the rollback procedure. The whole process is complicated by communication problems. If the communication issue happens on the forward path, it may also happen in the reverse direction. Therefore, eliminating the problem is more beneficial than solving it.

In a microservice architecture, the following problems are possible: latencies, outages in one part of the system, failure to establish the connection because of the connecting hardware or its settings updates, limitations of queues and their operational peculiarities, and other physical or logical issues. It is always better to eliminate as many potential failures as possible and increase fault tolerance, as long as it does not affect the quality of work and does not increase the downsides of a software product dramatically. Grouping by shared resources will reduce many possible issues related to connectivity problems. In case of outages, a whole group will fail, which will keep the information in the database consistent with a higher probability. At the same time, a downside of such grouping is not being able to respect the load of microservices and to provide effective usage of the resources.

2.2.3 Channel Throughput

There are cases when the most busy resource is the channel resource, when a microservice needs to handle massive and/or very frequent data chunks. This may not produce a high processor load but requires many threads to be able to pick up the incoming information. If the activity for such a microservice increases, scaling becomes necessary.

Supporting such a microservice requires a significant amount of resources. The best option would be to combine it with other microservices that would either accept rare requests and possibly have high processor load, or function similarly, but during different hours. Needless to say, such microservices are risky, and if possible, it is advisable to avoid heavy traffic between application components by redesigning the system and the purpose of microservices. There are techniques to prevent it. Introducing an additional stage of data preprocessing or storing intermediary results in the database could be beneficial. Data could undergo primary filtering and be routed in several directions.

2.2.4 Time and State of Processor Load

There are microservices designed specifically for data processing, such as machine learning routines, which require many computational resources [5]. Such tasks may be time-consuming and require high load

on the processor. It might be possible to split high processor-load microservices into several ones to reduce the amount of operations they have to perform. Consequently, this will reduce the time of processing a single request and accordingly increase the throughput of a single instance of such a microservice. In many cases, though, researchers and engineers prefer having all logic in one place and reusing its components in various places, rather than dispersing it across multiple microservices, thus duplicating the code or increasing the traffic between the application components [6].

This kind of microservice would normally be the heart of the application and would receive fewer requests compared to the other components. This stems from preceding procedures of filtering erroneous data and data aggregating activities, which reduce the number of initial requests. To create a resource optimizing group, such microservices could be paired with others that process their tasks quickly and do not experience heavy traffic concurrently with the active phase of the first set. Complementary microservices should not tie up threads for extended periods, allowing them to quickly become available for new tasks. In other words, we should combine the microservices when one needs 2 seconds to process one request, while its counterpart processes 10 requests in 1 second.

2.2.5 Geographical and Active Time Reasoning

The speed of user response decreases as the distance to the processing server increases. Hence, globally distributed applications would prefer to run separate instances of their applications on regional servers or cloud centres to provide the best user experience [7]. For example, Amazon servers are currently located in 32 geographical regions [8]. Consider, on the other hand, a company that is present on one continent, e.g. North America, and also has a presence on another continent where working hours do not significantly overlap, e.g. Europe. Depending on the goals of the company and the most likely sources of its profit, there might be no compelling reasons to establish costly hosting on several continents. People using the services in unpopular locations would agree to wait for a response from a distant server and would stay loyal to the service provider they use.

As an example, consider cell phone providers working in a certain county. When their customers travel, the calling services are provided by local providers instead, but the mobile application should work regardless of the user's current location. When on a tourist trip to another country and waiting for information updates in a mobile app, a user may be more ready to receive a delayed response than usual. This response time will not influence his choice of a cell phone provider when he comes back. Awareness of the existence of companies and applications that operate in this manner gives us reasons to plan for the more optimal utilization of shared space.

In terms of grouping, this implies that we could combine microservices to provide services not only at different times for the same region but also for different geographical regions, whose time zones can be considered opposed due to their lack of overlap in active usage times. In the above example, it could be convenient to make one logical and possibly physical group of servers as a set of instances hosting the same microservices, with each one providing services to the opposite geographical regions with low user intensity.

Table 1 and Figure 1 presents an outline of all reasonable ways of combining microservices, allowing them to form a complementary group and be collocated within a single server group.

2.3 Existing Solutions

To understand the existing research in this area, we performed a basic review of the most popular designs of microservice resource usage optimization.

2.3.1 What are the Kubernetes Scheduling Solutions?

In Kubernetes documentation, it is mentioned that there are many factors to determine the server for running a new instance and that is the most feasible. The major ones include individual and collective resource requirements, hardware/software/policy constraints, affinity and anti-affinity specifications, data locality, and inter-workload interference [9]. These approaches are based on physical abilities and the best possible speed of intercommunication. That doesn't include the optimal solution from the economic or ecological points of view, which are of high importance these days.

2.3.2 How Amazon Cloud Manages the Load Balancing?

Amazon Cloud has its Elastic Load Balancing (ELB) which includes three elements. The first one is the Application Load Balancer. It supports host-based and path-based routing, meaning that the traffic is routed based on its content or its headers, the domain name. This idea is not unlike geographical grouping,

Table 1: The grouping possibilities to optimize resources usage.

Group Name	Primary (problematic) microservice	Complementary microservice	When to use	Problems
Security	A microservice with certain security standards.	Other microservices with the same security standards.	To manage multi-tenant access to a shared resource.	Might not respect the load of microprocessors and the group won't optimally use resources.
Shared resource	A microservice that uses a certain external resource that is launched on the same subcluster.	Other microservices that use the same shared resource.	When there is a need to have maximal speed of communication between external resources and microservice.	Might not respect the balance of microservice group load, and might not provide effective usage of the resources.
Heavy channel throughput	Massive and/or very frequent chunks of data should pass through a microservice.	Microservices that would either accept rare requests and possibly have high processor load, or work in a similar way, but in the other daytime.	When many threads are needed to pick up the incoming information.	Heavy traffic microservices are risky because of possible lack of space in the queues, more often failing to send the information and delays in processing.
Heavy processor load	Tasks that are implemented during a long time and/or heavily load the processor.	Complementary microservices should not hold threads for a long time and so let them be ready for new tasks quickly.	When not many threads are needed, but high processor capacity is called for.	Time of returning response can be a problem in real-time applications.
Geographically separated	Applications handling a large number of concurrent connections resulting in high load by channel throughput and/or processor load.	Microservice that has a high load of a similar type but which working hours are not intersecting.	When the time of heavy load is stable and limited to a certain part of the day,	Problems may arise if the complementary microservices have sudden unexpected picks of activity in the intersecting time.

Group 1 - a security group of microservices that share the same server resource (supercomputer, cloud group)

Microservice 1   Microservice 1   Microservice 2   Microservice 2

Microservice 2

Microservice 3

Microservice 3

LAN

Cloude

Microservice 3

Microservice 3

Group 2 - a group for microservices that share the same DB, libraries of files, and/or communicated with each other

Shared server that runs all the microservices (e.g. supercomputer, local, cloud server group)

Groups 3, 4, 5 - High Channel Throughput, processor load or geographically separated microservices

Figure 1: Possible strategies of organizing servers into groups mentioned in the Table 1.

but in fact is opposite in spirit. Groupings are done by close locations rather than the opposite ones.

The second one is the Network Load Balancer. Its purpose is to prevent connection draining before a target is considered unhealthy and evenly distribute the traffic in a cross-zone mode which makes the optimal resource utilization. The flow-based distribution algorithm is used to make the network load balancing particularly suitable for applications that benefit from predictable and consistent connection handling, making it well-suited for a variety of use cases, including those with stateful and connection-oriented requirements

The third element is the combination of the two mentioned ones, called a Classic Load Balancer [10], [11].

The methods of load balancing are very well-thought-out and are meticulously designed and organized to serve users in the most effective manner possible. The techniques are also formulated to ensure equitable sharing of resources, thereby extending the longevity of the last serve [12]. The resources are turned off as soon as they are not needed, which prevents wasteful energy consumption. Nevertheless, in Cloud management theory, there is not information on how to reduce the number of needed servers in order to optimize the cloud activity. The methods described in Table 1 could help to achieve this.

Google Cloud [11] offers similar functionality.

3 CONCLUSIONS

Running multiple microservices and scaling them to ensure fault tolerance for the entire application, as well as its ability to effectively handle any number of users, is a critically important task. When choosing the right node, cluster, or instance group to run a certain instance of the microservice at cloud and Kubernetes, one of the many criteria to be taken into account is the ecological component. This means fullutilization of resources of one server, reducing their number to an absolute minimum. Such an approach also helps minimize electricity bills, which can be huge for constantly operating machine loads.

The article proposes forming instance groups of complementary microservices using the rules of a complemented lattice in fuzzy logic. The most efficient can be the groupings based on
- Difference in the time zones.
- Difference in the style of work of the applications.

- Balance between high or long processor load and amount of fast requests that the application handles.

At present, such criteria are not included in the list of factors that Kubernetes and popular cloud services use to decide on which server to run an instance.

REFERENCES

[1] G. A. Gratzer, "General lattice theory." Pure and Applied Mathematics: A Series of Monographs and Textbooks, no. 75, Academic Press, New York, 1978.

[2] N. Ajmal and K. V. Thomas, "Fuzzy lattices," Information Sciences, vol. 79, no. 3-4, pp. 271-291, Jul. 1994, doi: 10.1016/0020-0255(94)90124-4.

[3] O. Dmytrenko and M. Skulysh, "Fault Tolerance Redundancy Methods for IoT Devices," Infocommunication Comput. Technol., vol. 2, no. 04, University "Ukraine," pp. 59-65, Dec. 2022.

[4] Kubernetes Team, "Multi-tenancy in Kubernetes," Kubernetes, [Online]. Available: https://kubernetes.io/docs/concepts/security/multi-tenancy/, [Accessed: 23 Dec 2023].

[5] "A review of in-memory computing for machine learning: architectures, options," International Journal of Web Information Systems, Dec. 2023, doi: 10.1108/IJWIS-08-2023-0131.

[6] F. Wilhelmi, D. Salami, G. Fontanesi, L. Galati Giordano, and M. Kasslin, "AI/ML-based Load Prediction in IEEE 802.11 Enterprise Networks," 2023.

[7] A. Meir, "Does Location Matter In Cloud Computing?," Ridge Cloud, [Online]. Available: https://www.ridge.co/blog/location-in-cloud-computing/, [Accessed: 23 Dec 2023].

[8] Amazon Team, "AWS Global Infrastructure". [Online]. Available: https://aws.amazon.com/about-aws/global-infrastructure/?nc1=h_ls, [Accessed: 22 Dec 2023].

[9] Kubernetes Team, "Kubernetes Scheduler." [Online]. Available: https://kubernetes.io/docs/concepts/scheduling-eviction/kube-scheduler/, [Accessed: 14 Dec 2023].

[10] Y. Sharma, "Key Strategies for Implementing AWS Network Load Balancer." Sep. 27, 2023. [Online]. Available: https://dev.to/aws-builders/key-strategies-for-implementing-aws-network-load-balancer-35fc.

[11] S. Al-Raheym, S. C. Açan, and Ö. T. Pusatli, "Investigation Of Amazon And Google For Fault Tolerance Strategies In Cloud Computing Services," AJIT-E Online Acad. J. Inf. Technol., vol. 7, no. 23, pp. 7-22, Nov. 2016, doi: 10.5824/1309-1581.2016.4.001.x.

[12] L. Globa, M. Skulysh, and A. Zastavenko, "The method of resources allocation for processing requests in online charging system," in The Experience of Designing and Application of CAD Systems in Microelectronics, 2015, pp. 211-213, doi: 10.1109/CADSM.2015.7230838.

Using a Genetic Algorithm for Telemedicine Network Optimal Topology Synthesis

Roman Tsarov[1], Lesya Nikityk[1], Iryna Tymchenko[1], Vladyslav Kumysh[2], Kateryna Shulakova[1,3], Serhii Siden[1] and Liliia Bodnar[4]

[1]*State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, 65023 Odesa, Ukraine*
[2]*VitalSource Technologies LLC, Fayetteville Str. 227, Suite 400, Raleigh, NC 27601, United States*
[3]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*
[4]*South Ukrainian National Pedagogical University, Staroportofrankyvska Str. 26, 65020 Odesa, Ukraine*
rcarev@gmail.com, lesyanikityuk579@gmail.com, mentalitett@gmail.com, vlad.kumish@gmail.com,
katejojo29@gmail.com, ssiden@suitt.edu.ua, bodnarl79@pdpu.edu.ua

Keywords: Telemedicine, Telemedicine Network, Topology Synthesis, Topology Optimization, Genetic Algorithm.

Abstract: A method based on a genetic algorithm is proposed for synthesizing the optimal topological structure of telemedicine network, ensuring that the distribution of users (with a known location) by telemedicine stations (the number and location of which are also known) is optimal in terms of signal delay time during transmission and the cost of network deployment. The method uses: random generating of a base population, a tournament selection of chromosomes among two pairs for crossover, and a homogeneous crossover operator. The results of benchmarking the proposed method are presented. The experiment reveals that the resulting solution is indeed close to optimal, i.e. due to the use of a genetic algorithm, the method avoids falling into the trap of a local extremum. While the current study focused on a specific telemedicine network, future research could explore the scalability of this genetic algorithm approach for larger-scale networks and consider additional factors such as energy efficiency and fault tolerance.

1 INTRODUCTION

Ensuring a healthy lifestyle and accessibility of medical care is an important component of the sustainable development of humanity [1]. The concept of telemedicine was developed to ensure and to increase the accessibility of medical services. In essence, telemedicine refers to the delivery of medical services using information and communication technologies, particularly when distance poses a significant barrier. [3]. The experience of recent years has shown that telemedicine is often the only solution for providing assistance to people who do not have the physical opportunity to visit a medical facility.

The provision of telemedicine services is implemented through a telemedicine network (TMN). TMN can be classified as distributed information and communication network. In most cases, TMN is built on a hierarchical principle and includes telemedicine centers and telemedicine stations at the network core level, and in turn telemedicine service consumers and

their access terminals to information networks establish the access level (Figure 1) [3].

In practice, stationary telemedicine stations are often located on the territory of existing medical institutions or in places equipped with ready-made telecommunications and engineering infrastructure. As a result, their location and number are known, mainly determined by the number of suitable locations and the amount of funding. Each telemedicine station can serve a limited number of patients, which is determined by the class of the telemedicine station [2, 3].
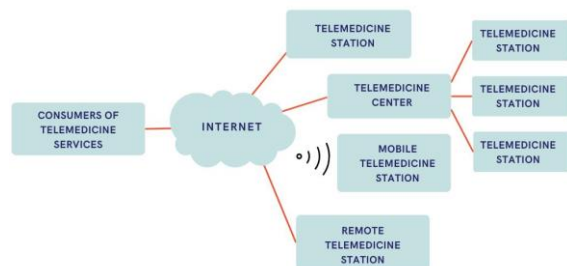
Figure 1: Basic layout of a telemedicine network.

As a rule, employees of telemedicine services can connect to only one telemedicine station (the one with which they have signed a contract). The location of user terminal is determined by the places of deployment (residence, study, work) or public places provided with access to information networks.

Thus, in practice, synthesizing the optimal topological structure when designing a new TMN or optimizing the topology of an existing TMN often implies the optimal distribution of users (with a known location) by telemedicine stations (the number and location of which are also known) and is an actual task.

2 STATE OF THE ART AND PAPER GOAL

When conducting a feasibility study to choose a network topology, characteristics such as throughput, reliability, quality of service, and the cost of all its elements are usually taken into account. It's worth mentioning that parameters like signal delay time, which characterize the quality of service, also directly affect the resulting data transfer rate. Thus, as criteria for optimizing the topological structure of the TMN in the scope of this work we propose to use the minimum signal delay time and the minimum cost of network deployment.

In terms of optimization theory, the problem can be formulated as follows. Let there be a certain set of nodes of two types: telemedicine stations and terminal equipment for telemedicine service consumers. Each node is specified by certain coordinates in space and a number. Terminal equipment corresponds to the maximum amount of traffic that it can generate per unit of time. In turn, telemedicine stations are specified by the maximum amount of traffic that they can process without

additional delay. Also, each telemedicine station has a certain capacity, i.e. it can serve a limited number of patients.

It is necessary to synthesize a TMN topology that has a minimum total weight of all connections between nodes and a minimum total average delay when a signal passes through the TMN. In this case, the following restrictions are imposed on the topology: each patient can be connected to only one telemedicine station and there should be no unconnected patients in the network.

As the weight characteristics of the TMN communication lines we will use the Euclidean distances between the corresponding nodes.

The formulated problem is an NP-complete integer nonlinear problem, similar to the well-known traveling salesman problem. The exact solution of such problems can be found by brute-force search, however, for solving large-scale problems, finding a solution requires a significant time resource. To solve similar problems, heuristic methods are usually used, which allow to get a solution as close as possible to the exact one, but at the same time with the minimization of time costs. Known heuristic methods for solving nonlinear optimization problems include [4]: Penalty Function Method, Projected Gradient Method, Interior Point Method, Branch and Bound Method. The principle of operation of these methods consists in a pseudo-random search with constant analysis and evaluation of the accumulated data on the already reviewed part of the solution space. In Figure 2 the traveling salesman problem solution space with at five nodes is shown in different projections.

As can be seen from Figure 2, the set of solutions contains a significant number of local extremums, which could be the potential solutions to the problem.

All the above-mentioned methods have a common drawback - they can fall into the trap of a "local" extremum, and thus provide a non-optimal solution.
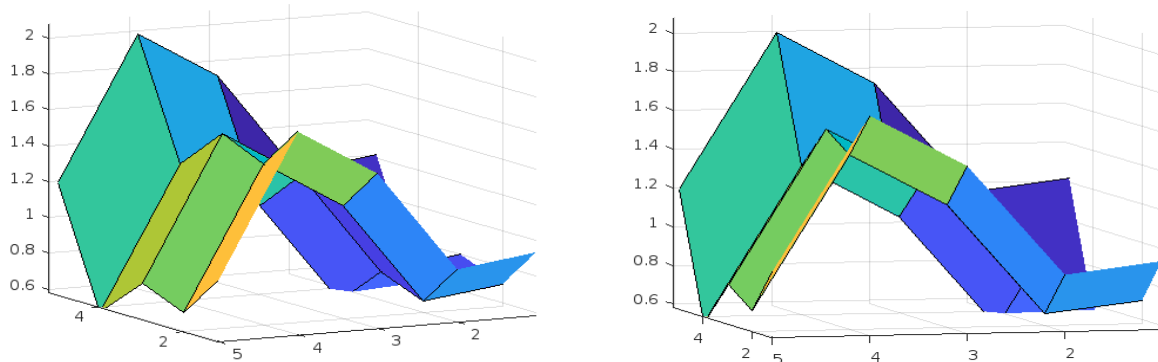
Figure 2: Projections of traveling salesman problem solution space.

It is known that Genetic algorithms do not have this drawback. In such algorithms, a mutation operation is used, which randomly changes the current parameters of the problem, and thus "bypasses" local extremums. Also, the advantages of genetic algorithms include: the ability to work with a large number of variables, the absence of the need to calculate derivatives, the possibility of parallelization [5-7].

Genetic algorithms are actively used to solve various optimization problems in the field of information communications. Thus, work [8] investigates the issue of optimizing the routing process, for the solution of which authors have developed a genetic algorithm, which, unlike the Dijkstra and Depth-First Search algorithms, allows you to find a set of shortest routes that have the same cost. The work [9] solves the problem of creating a resource scheduler to maximize the bandwidth of network channels. The problem of optimizing the structure of a wireless network using genetic algorithms in order to achieve an optimal balance between network performance and resource consumption is studied in [10]. The paper [11] analyzes the possibility of using genetic algorithms to solve the problem of energy efficiency of a wireless network. In [12] the problem of load balancing is considered during data transmission in a telemedicine network, which is built on the basis of a modified adaptive genetic algorithm.

From the above it follows that the development of method empowered by genetic algorithm for synthesizing the optimal topological structure of TMN, ensuring that the distribution of users (with a known location) by telemedicine stations (the number and location of which are also known) being optimal in terms of signal delay time and the cost of network deployment is an actual scientific problem. Solving this scientific problem is the goal of this work.

3 PROPOSED METHOD

Genetic algorithms use combinations of such genetic operations as selection, crossing and mutation [6-8]. In terms of genetic algorithms, the solution of a formalized problem is called a chromosome, i.e. coded TMN topology.

The matrices of coordinates in space and the maximum volume of traffic for two types of nodes (telemedicine stations and terminal equipment) are used as the input data. The matrix of telemedicine stations also includes the maximum number of patients that these nodes can serve.

Also, the input data include an adjacency matrix $A = (a_{ij})$, where element $a_{ij} =1$, if there is a connection between the i-th telemedicine station and the j-th terminal device, and 0 otherwise. So, for example, the configuration of a network consisting of 4 stations and 12 terminal devices may look like Figure 3.

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Figure 3: Discrete network representation.

The proposed method can be described as follows:

Step 1. Building a matrix of connection weights. The structure of the weight matrix is similar to the structure of the adjacency matrix, but instead of 1 it contains the value of the distance between the corresponding nodes.

Step 2. Initializing a genetic algorithm with the base population and number of generations. In this case, the base population means a set of initial topologies in which each terminal device has a connection with one and only one telemedicine station.

Step 2.1. Setting the number of generations G (a number of iterations of the genetic algorithm).

Step 2.2. Setting the population size S (a number of initial topologies).

Step 2.3. Generating the base population. In this work, it is proposed to use a random method for generating the base topology.

Step 3. Calculating the fitness function for each topology in the population. The problem to be solved is multicriteria, therefore it is proposed to use one of the classical methods for solving such problems - the weighted function method [4].

As the fitness function F we use the sum of three functions F1, F2 and F3 with different weights:

$$FF = w_1F_1 + w_2F_2 + F_3,$$

here w_1 , w_2 are weighting coefficients such that $w_1 + w_2 = 1$; function F_1 calculates the sum of the weights of all TMN communication lines; F_2 calculates the average total delay when passing through the TMN; F_3 is a penalty function that significantly increases the value of the fitness function if in topology the number of terminal devices connected to any telemedicine station exceeds its maximum capacity.

Parents:

$$A_1 = \begin{array}{|cccccccccccc|} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array}$$

$$A_2 = \begin{array}{|cccccccccccc|} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}$$

Children:

$$A'_1 = \begin{array}{|cccccccccccc|} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array}$$

$$A'_2 = \begin{array}{|cccccccccccc|} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}$$
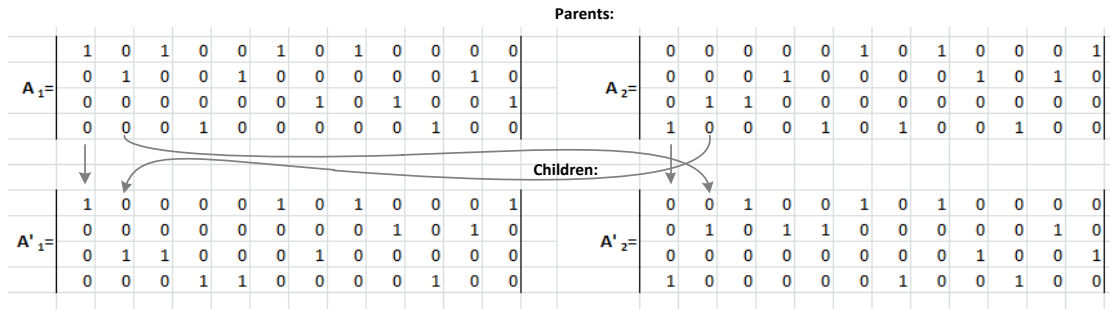
Figure 4 – Crossover operation.

Original chromosome:

$$A = \begin{array}{|cccccccccccc|} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & ① & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}$$

Mutation:

$$''A = \begin{array}{|cccccccccccc|} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & ① & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}$$
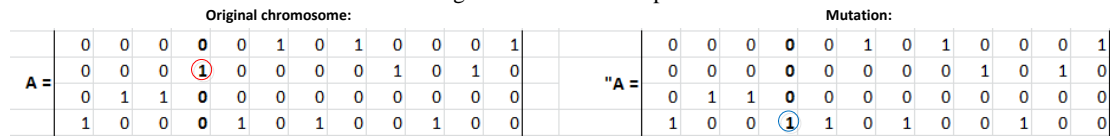
Figure 5 – Mutation operator.

Let's define the components of the fitness function:

$$F_1 = \sum_{\tau_i \in \tau} m(\tau_i),$$

here τ is the set of communication lines of the TMN graph; $m(\tau_i)$ – weight of the corresponding TMN communication line,

$$F_2 = \frac{\sum_{\theta_i \in \theta} D(\theta_i)}{N},$$

here Θ is the set of paths between any two TMN graph vertices; $D(\Theta_i)$ – delay time along the path Θ_i; $N = |\Theta|$ – total number of paths.

Step 4. Selecting chromosomes from the population for crossover. When solving this problem, we will use tournament selection of chromosomes, carried out among the two pairs [6]. Each of these two pairs is selected from the population with a probability that follows a uniform distribution. Then, from each pair, the individual with the best fitness function value is selected.

Step 5. Crossover on the winning chromosomes. The winners of each of the two tournaments held in the previous step get combined with each other. To solve this problem, it is proposed to use the homogeneous crossover operator. An element (in this case, a column of matrix) of each of the pair of descendants is selected with uniform probability from the first or second parent. If the element for the first descendant was taken from the second parent, then the same element of the second descendant should be taken from the first parent [6-8]. An example of using this operator is shown in Figure 4.

Step 6. Mutation of descendants. The mutation operator is applied to the descendants resulting from crossover in the previous step. The main purpose of using mutation in genetic algorithms is to avoid premature convergence of the population to a solution that is not good enough [6]. When solving this problem, the best results were obtained with probability $P = 1/S$. With some probability, a terminal device is selected and connected to another telemedicine station. The probability of choosing one or another telemedicine station is subject to a uniform distribution. An example of a mutation of this kind is shown in Figure 5.

Then steps 3-6 of the method are repeated until the number of generations is equal to G, as specified at the Step 1.

4 EXPERIMENTAL RESULTS

For the experiment, we the optimal topological structure of TMN from 5 telemedicine stations and 40 telemedicine service consumers was synthesized. Genetic algorithm parameters: population size – 50, number of generations – 250. Penalty function

$$F_3 = 1000*n,$$

here n is the number of terminal devices, connected to the telemedicine station in excess of its maximum capacity.

Weighting coefficients $w_1 = w_2 = 0,5$. The maximum capacity of telemedicine stations is TS1=7, TS2=10, TS3=8, TS4=15, TS5=9. Genetic algorithm was emulated in the MatLab modeling environment.

Figure 6 shows one of the basic topologies, and Figure 7 - the optimal topology of the TMN, resulting from the proposed method. Table 1 presents the dynamics of changes in the population fitness function.
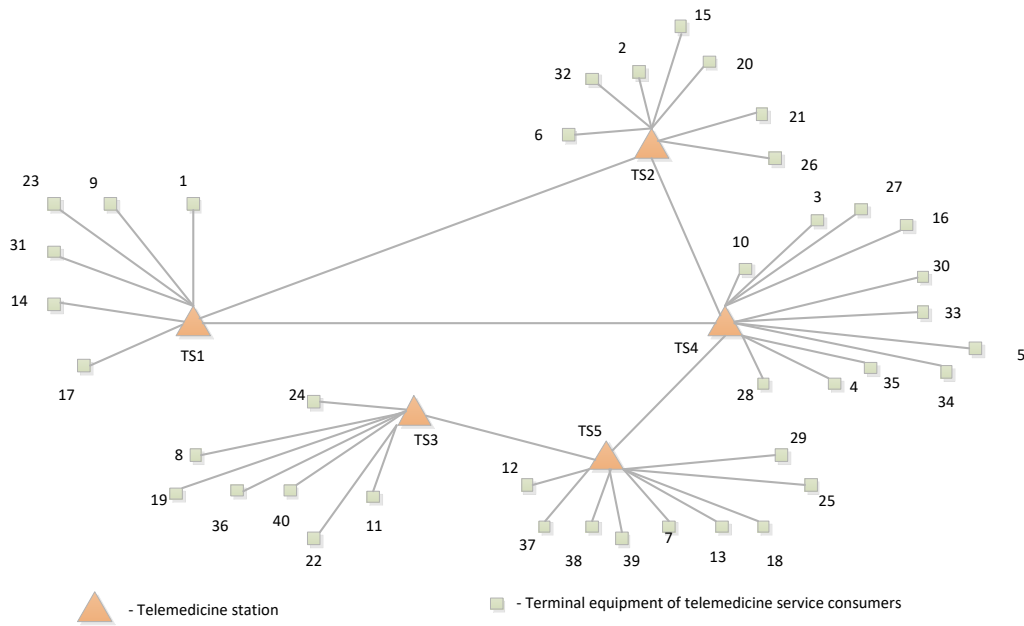
Figure 6: One of the basic topologies of TMN.
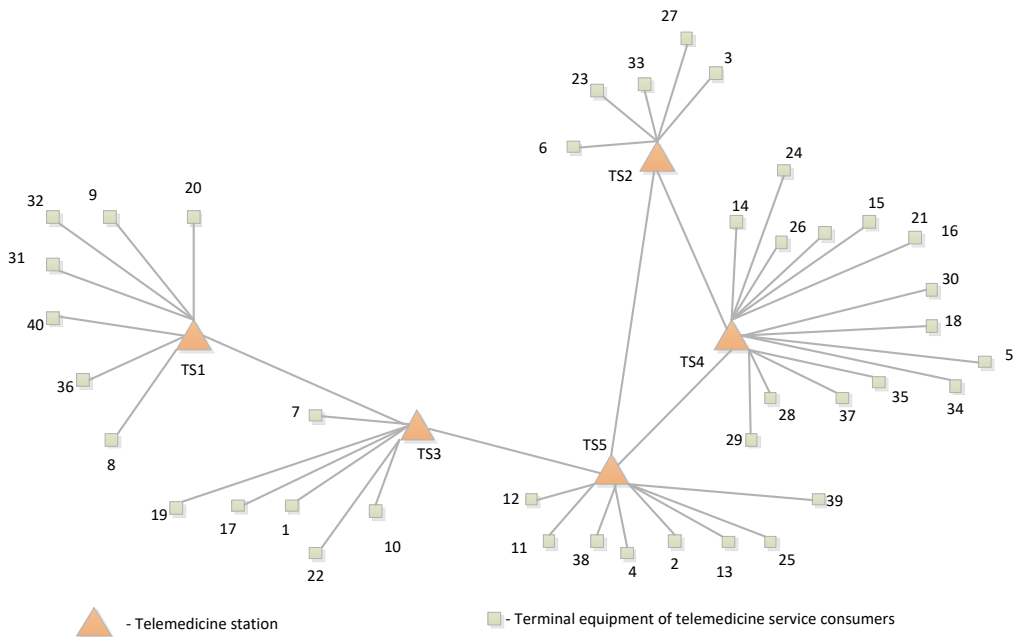
Figure 7: Optimal topology of TMN.

Table 1: Dynamics of the fitness function change.

Generation	The dynamics of changes in the average value of fitness function, %	Dynamics of changes in the best value of fitness function, %
1-10	27,00	9,2
11-20	24,30	6,8
21-30	11,15	2,14
31-40	3,4	0,95
41-50	1,17	0,02

5 CONCLUSIONS

The proposed method, based on a genetic algorithm, can be used to synthesize the optimal topological structure of a TMN. This ensures that the distribution of users (with known geographical coordinates) by telemedicine stations (whose number and location are also known) is optimized for both signal transmission delay time and the cost of network deployment. The experiment revealed that the resulting solution achieves near-optimal performance , i.e. the use of a genetic algorithm significantly reduces the risk of falling into a local extremum. In this method, the accuracy of the obtained results depends on the number of generations specified during the initialization stage of the genetic algorithm. The computational complexity of the algorithm can be further reduced by employing a more efficient method for finding the base population.

Further research will be aimed at developing a generalized method that allows synthesizing the optimal TMN topology when the required number of telemedicine stations and their location are unknown in advance. It is also important to modify the resulting method to optimize the TMN topology with reference to a larger number of parameters.

ACKNOWLEDGMENTS

REFERENCES

[1] United Nations, "Goals in the field of sustainable development," [Online]. Available: https://www.un.org/sustainabledevelopment/ru/health.

[2] S.M. Talha, S. Siden, R. Tsaryov, and L. Nikityuk, "Assessment of the Possibility of Using 5G to Build Telemedicine Networks in Various Environment," 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Dortmund, Germany, 2023, pp. 1125-1129, doi: 10.1109/IDAACS58523.2023.10348929.

[3] R. Tsarov, I. Tymchenko, V. Kumysh, K. Shulakova, and L. Bodnar, "Extended Classification Model of Telemedicine Station," in Proceedings of International Conference on Applied Innovation in IT, vol. 11, no. 1, pp. 37-42, doi: 10.25673/101908.

[4] L.P. Feldman, "Numerical methods in computer science: a study guide," BHV Publishing Group, Kyiv, 2009, 479 p.

[5] J. Holland, "Genetic algorithms, computer programs that evolve in ways that even their creators do not fully understand," Scientific American, pp. 66-72, 1975.

[6] M. Srinivas and L. M. Patnaik, "Genetic algorithms: A survey," Computer, vol. 27, no. 6, pp. 17-26, 1994.

[7] K.-F. Man, K.-S. Tang, and S. Kwong, "Genetic algorithms: concepts and applications," IEEE Transactions on Industrial Electronics, vol. 43, no. 5, pp. 519-534, 1996.

[8] C.W. Ahn and R.S. Ramakrishna, "A genetic algorithm for shortest path routing problem and the sizing of populations," Evolutionary Computation, IEEE Transactions on, vol. 6, no. 6, pp. 566-579, 2002.

[9] D. Gozupek and F. Alagoz, "Genetic algorithm-based scheduling in cognitive radio networks under interference temperature constraints," International Journal of Communication Systems, vol. 24, no. 2, pp. 239-257, 2011.

[10] H.T. Cheng and W. Zhuang, "Novel packet-level resource allocation with effective QoS provisioning for wireless mesh networks," Wireless Communications, IEEE Transactions on, vol. 8, no. 2, pp. 694-700, 2009.

[11] A. Roy, N. Banerjee, and S. K. Das, "An efficient multi-objective qosrouting algorithm for wireless multicasting," in Vehicular Technology Conference, 2002, VTC Spring 2002, IEEE 55th, vol. 3, pp. 1160-1164, IEEE, 2002.

[12] D. Lin and F. Labeau, "Accelerated genetic algorithm for bandwidth allocation in view of emi for wireless healthcare," in Wireless Communications and Networking Conference (WCNC), 2012 IEEE, pp. 3312-3317, 2012.

Design of a Model of an Information and Communication System for Solving the Problem of Preventing Car Collisions

Vitalina Babenko[1,2], Oleg Pursky[3], Bohdan Karpishen[1] and Mykhaylo Fedorchuk[4]

[1]*Kharkiv National Automobile and Highway University, Yaroslav Mudryi Str. 25, 61002 Kharkiv, Ukraine*
[2]*Daugavpils University, Vienības Str. 13, LV - 5401 Daugavpils, Latvia*
[3]*State University of Trade and Economics, Kioto Str. 19, 02156 Kyiv, Ukraine*
[4]*Mykolayiv National Agrarian University, General Karpenko Str. 73, 54038 Mykolaiv, Ukraine*
vitalinababenko@karazin.ua, o.pursky@knute.edu.ua, karpishen.bogdan@gmail.com, mfedorchyk01@gmail.com

Keywords: Vehicle-to-Vehicle Communication, Advanced Driver Assistant Systems, Dedicated Short-Range Communication, Model of Prevention Car Collisions System, Basic Safety Massage.

Abstract: Vehicle-to-vehicle (V2V) communication is a special type of vehicular ad-hoc network communication that has attracted much attention from researchers, industries, and the government due to its important application to improve driving safety for the next generation of vehicles. However, if the warning comes too late, the vehicle behind may not have enough time to stop smoothly or slow down. Especially when it comes to car accidents involving multiple vehicles. The collision warning system is important for avoiding rear-end collisions. When the vehicle in front slows down or the risk of a rear-end collision increases, the system sends a warning. Fast-moving vehicles and unreliable wireless communication links can lead to communication difficulties between vehicles, affecting the performance of multi-vehicle systems. The article describes the general perspective of the development of automobile networks. This material examines the possibilities of communication modeling and the possibility of looking at development problems, as well as researching and solving issues of network operation, the behavior of driver assistance applications and the interaction of many cars. A model of system operation based on the "receiver-transmitter" principle for BSM transmission using DSRC communication was created using MATLAB/Simulink software. The resulting model allows for various analyzes of the system for the future improvement of information and communication systems in cars.

1 INTRODUCTION

The rapid development of transportation systems has brought many conveniences to our daily lives, allowing the safe and reliable transportation of both people and goods within the country and abroad. It is estimated that more than one billion cars are owned by people around the world. It is estimated that this number will double within a decade or two. However, a number of issues related to this growth cause concern. From a safety perspective, more than 42,915 people were killed in US highway crashes in 2021 [1].

Vehicle-to-vehicle (V2V) communication between multiple connected vehicles improves the safety and efficiency of our transportation systems.

This is achieved through the use of traffic management systems that rely on on-board sensors and V2V communication. Communication mainly provides real-time status information

(e.g., acceleration, speed, location) of the front vehicle or vehicles [2].

Assessments regarding the importance of this technology:

- Improving road safety. V2V communication allows cars to exchange information about their location, speed and other parameters in real-time. This helps to avoid accidents, reduce the number of collisions and improve the response to danger. For example, the system can warn drivers about possible risks, such as an emergency situation on the road, obstacles or dangerous overtaking.
- Reducing congestion and improving traffic flow. V2V technology helps drivers choose the optimal route and speed to avoid traffic jams. This can improve traffic flow and reduce travel time.

- Accident reduction. Thanks to V2V communication, cars can share information about dangerous road conditions, such as slippery asphalt, limited visibility or bad weather. This helps drivers adapt their driving style to specific conditions and reduces the risk of accidents [3].

Demand for advanced driver assistance systems (ADAS) [4] – those that help perform monitoring, warning, braking, and steering tasks - will grow over the next decade, driven largely by regulatory and consumer interest in safety programs that protect drivers and to reduce the number of accidents.

Currently, ADAS systems are seen as an ever-evolving industry that continues to improve and evolve. Although there are already many different ADAS systems on the market, they are constantly being improved and supplemented with new features. At the system development stage, various challenges and problems arise, such as interaction between systems, data security, compliance with regulatory requirements, integration with other systems and car components, intelligent data management, and others.

The development of ADAS systems is a complex and multifaceted process that requires a detailed study of various aspects, from technical capabilities to socio-economic and legal aspects.

In this study, we focus on services that create or correlate with one-time messaging, especially using a V2V communication platform.

2 LITERATURE REVIEW

Based on the literature, we present various applications of V2V communication, classified according to two broad purposes: security purposes and non-security purposes. The main goals in this category are to minimize safety problems by providing directions or other information to the driver to prevent or predict traffic accidents, such as pre- or post-crash situations, blind spot prediction, intersection assistance, etc.

Vehicle-to-vehicle messaging aims to minimize potential crashes and enhance safe in-car driving with driver assistance features for both autonomous and non-autonomous vehicles. In addition, V2V communication will enhance safety support in five-level autonomy in autonomous vehicles, where the combination between artificial intelligence, vehicle technology, the Internet of Things and communication capabilities will accelerate the mass adoption of autonomous vehicles in the [5].

Nowadays, with V2V communication through technologies such as IEEE 802.11p-based DSRC and new 5G solutions, various IFTs become possible as a vehicle can communicate with vehicles outside its immediate environment [6].

If we take the characteristics of the transmission channel itself, there is a possibility of improving communication by interfering with the structure of the 802.11p protocol, which covers the field that has the value of the size of the information packet. Work [7] focused on developing a powerfully adaptive packet size structure that depends on the value of the signal-to-noise ratio. There are neural network controls that are trained by potential packet size values derived from an equation obtained by practical testing of the multiple relationship between packet error rates and packet size values [8]. Adjusting the size of the information packet results in a reduction of the packet transmission error rate.

The paper [9] proposes a predictive distributed model predictive control (DMPC) method for multi-vehicle system control in switched communication topologies. An open-loop optimization problem is formulated, in which penalties and constraints for neighbor deviation and self-deflection are included to ensure stability. The DMPC algorithm is designed for systems with multiple vehicles subject to switching topologies. As a result, a system control controller with several vehicles in switching communication topologies was created.

For example, the goal of developing person-based adaptive signal control (PB-ACA) was to investigate optimal signal plans on an isolated connection. In this study, as a decentralized coordinated control, the coordinated person-based signal control (C-PBC) algorithm allows the local controller at each intersection in the road network region to independently control the PB-ACA based on the data of the connected vehicle within the wireless range to optimize personnel-based signal plans. For each intersection, the communication range is defined as a circle with a radius of 250 m from the center of the intersection, and the planned intersection can only receive data within this communication range [10].

With the rapid development of wireless communication technologies, the behavior of the driver in front of the vehicle can be transmitted to the following vehicle to improve the performance of the system.

This paper proposes a forward collision warning (FCW) system that detects the driving intention of the preceding vehicle and transmits the information to the following vehicle using V2V communication technologies. The proposed driving intention

recognition method provides better performance of the FCW system and gives the following vehicle additional time for smooth braking [11].

In all these cases, modeling is an important stage in the development and testing of systems. The paper [12] describes a mathematical model of V2V, overtaking and minimum overtaking distance analysis using fuzzy logic. Experiments were designed using PreScan/MATLAB. The model shows the effectiveness of the created algorithm.

In this work, the collision warning system was created using the information and communication data transmission model based on wireless communication technology using MATLAB/ Simulink [13].

3 MATERIALS AND METHODS

Wi-Fi technology, known as Dedicated Short-Range Communication (DSRC) between each vehicle, and GPS technology, which offers detailed positioning by exchanging data with similarly equipped vehicles. DSRC is a special-purpose communication tool designed for a vehicle to provide short-range communication with a neighboring vehicle or with the environment to achieve a shared driving situation. DSRC uses 75 MHz spectrum for automotive communications and uses IEEE 802.11p-based radio technology with a bandwidth of 3 to 27 Mbps [14]. Several components are required to provide V2V communication:

1) DSRC is a special radio unit that works as a data receiver and transmitter;
2) GPS receiver [15], responsible for determining the position of the car in space and time; this data will be the input data for DSRC;
3) OBU (On-Board Unit), which collects vehicle status data such as speed, steering angle, acceleration, brake status, etc. It also installs an application and a screen for displaying interface information.

We model communication (V2V) and build a model based on the "receiver-transmitter" principle (Figure 1). The model is an integral part of the project and has a pre-prepared "scene" (model of the road surface with obstacles for the transmitter), "scenario" (route of movement) and "actors" (transport with defined characteristics). The model uses DSRC basic radio transmission safety message (BSM) [16] message.

Communication relies on channel characteristics to determine the probability of successful message reception.

The Transmitter V2V subsystem generates a Basic Security Message (BSM) for each target vehicle using the received information for that actor. The transmitter reads the "actor" information and passes it through the Inertial Navigation System (INS) and the Global Navigation Satellite System (GNSS) to apply noise to the "actor" information. The subsystem also transforms the spatial location information of target vehicles from Cartesian coordinates to geographic coordinates using "scene" information. The subsystem then generates BSMs for all target vehicles. The SendMessage block inside the transmitter converts the signal into a Simulink message and delivers it to the object queue. Queues are organized as first-in-first-out (FIFO) queues.

The V2V receiver subsystem implements the behavior of the vehicle object receiver. The receiver receives the precomputed channel characteristics as the mask parameter and the transmitted BSM, scene, and car information as input. When a transmitter delivers a message to an object queue, it triggers the V2V receiver subsystem. For each target vehicle, the receiver calculates the distance from that target vehicle to its vehicle and then finds the corresponding bandwidth using the precomputed channel characteristics.

When the bandwidth exceeds the generated random number, the receiver receives the BSM and stores it on the output bus BSMOut. A FIFO queue models a message receiving interface that works based on message availability.

The received message is further forwarded to the message processing unit, which converts the BSM input data into physical values and generates object detection reports for the input data to track multiple objects [17].

4 THE RESULTS

The results of the model show the dependence between the distance from the transmitter to the receiver and the signal-to-noise ratio (SNR) [18] for different transmission bands (Figure 2). In this example, a difference of 50 and 150 m is compared. We can also see the relationship between distance and bandwidth for the specified range (Figure 3).
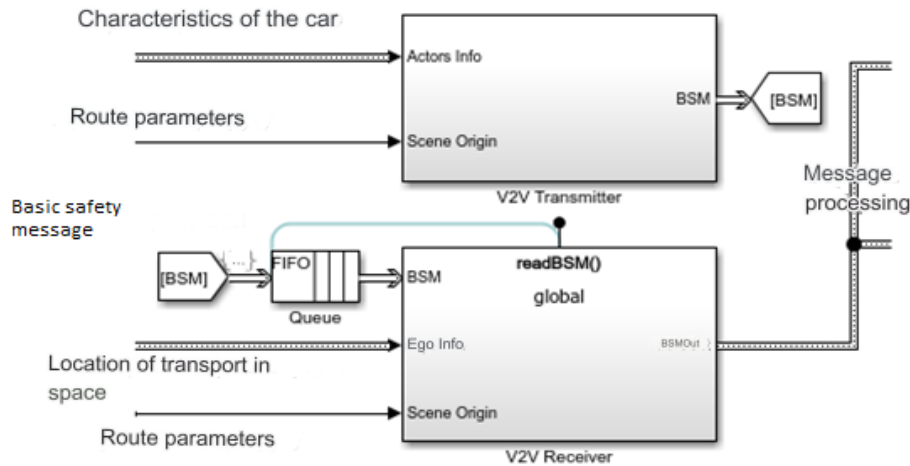
Figure 1: Vehicle to Vehicle Communication model.
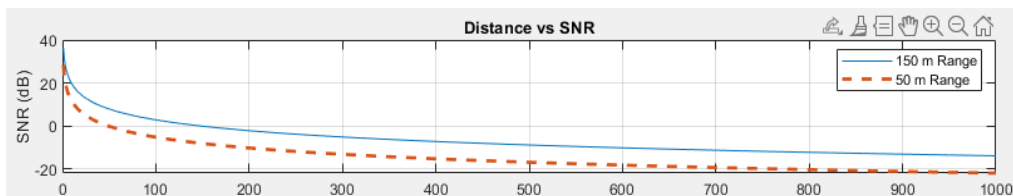
Figure 2: Dependence of distance and signal-to-noise ratio.
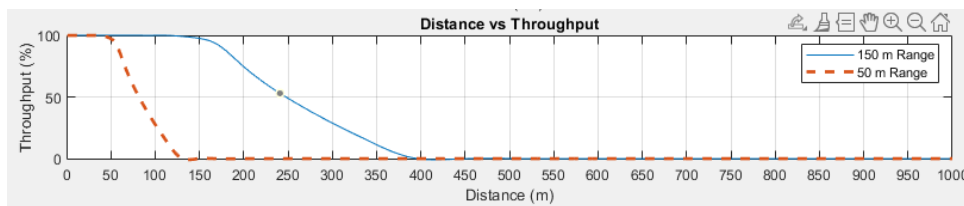
Figure 3: Dependence versus bandwidth relationship.

Throughput means the expected probability of detecting a packet. When the range is 150 m, the graph shows that the probability of packet detection is almost 100% up to 150 m, and then it gradually decreases until it reaches 0% at about 400 m, the probability of detecting a package decreases faster and at a distance of 150 m is close to 0%. During operation, the model visualizes important data and issues the following information (Figure 4):

- The ratio of transmitted and received messages, which reflects the number of transmitted and received messages at each time step.
- V2V Communication Data - Displays information about the transmission and

reception of BSM data and the signal-to-noise ratio for each message received.

BSM Message Received – Displays latitude, longitude, speed, course, longitude, and latitude for each target whose BSM message is received.

After receiving the BSM to generate a possible collision warning, the paths for the target and neighboring vehicles are estimated using their current position, speed, and heading angles. The estimated trajectory of each vehicle is a straight line connecting the vehicle's initial position and its predicted position 20 seconds later. To assess the collision risk, the analyzer checks whether the estimated path of the neighboring vehicle intersects with the estimated path

Message Transmitted vs Received

V2V Communicaion Data

	Target ID	Message Transmitted by Target	Message Received by Ego	Target-Ego Distance (m)	Received Message SNR (dB)
1	2	☑	☑	75.9258	4.8700
2	3	☑	☑	96.1348	3.1584
3	4	☑	☑	19.6898	14.6607
4	5	☑	☑	93.6371	3.3494
5	6	☑	☑	107.6260	2.3396
6	7	☑	☑	111.2922	2.0966
7	13	☑	☑	50.6118	7.8118

Received BSM Message

	Target ID	Latitude (deg)	Longitude (deg)	Speed (m/s)	Heading (deg)	Width (m)	Length (m)
1	2	42.300049	-83.698656	2.2	271.7	1.8	4.7
2	3	42.300532	-83.697632	20.1	92.2	1.8	4.7
3	4	42.300899	-83.698708	7.0	246.3	1.8	4.7
4	5	42.300614	-83.697642	19.9	91.1	1.8	4.7
5	6	42.300194	-83.697678	30.0	178.4	1.8	4.7
6	7	42.299850	-83.698116	19.9	179.8	1.8	4.7
7	13	42.300273	-83.698813	10.0	179.3	1.8	4.7

Figure 4: Visual display of received and transmitted messages from actors.

Table 1: Warning level according to time conditions.

Actor Arrival Time Condition	Time Gap Condition	Warning Level
ActArrivalTime < minArrivalTime	timeGap < minTimeGap	High
ActArrivalTime < minArrivalTime	timeGap >= minTimeGap	Moderate
ActArrivalTime >= minArrivalTime	timeGap < minTimeGap	Low
ActArrivalTime >= minArrivalTime	timeGap >= minTimeGap	Low

of the target. In the case of an intersection, the collision warning analyzer calculates the arrival time of the "actor" at the intersection point and determines the absolute difference between the arrival time of both vehicles at the intersection point.

Next, the arrival time of the "actor" and the value of the time interval are compared with the corresponding predetermined threshold values. Based on the results of the comparison (Table 1), the corresponding warning level is set [19]. Medium and high warning levels require driver awareness and response.

5 CONCLUSIONS

In this work, a communication system was proposed in the problem of warning about a possible collision. Communication parameters rely on channel characteristics to determine the probability of successful message reception. V2V technology was used to implement wireless transmission and reception of information between cars. The result of the work is receiving information from connected vehicles in the range of the system and the reaction of the information system to the processed data in the form of external signals [20-24]. At the same

time, the research in this paper also accelerates the use of V2V technology in the field of intelligent vehicles and improves the ability of intelligent vehicle to perceive the environment.

As for improving the performance of driver assistance systems, some solutions already exist. In the case of solving the task of ensuring message transmission for the frontal collision warning system, DSRC technologies were used. The results of the FCW experiments demonstrated that the system provided earlier warning than the previous result [24, 25]. This proposed system not only provided early warnings to prevent rear collisions, but also contributed to more effective braking [11].

Thanks to recent breakthroughs in wireless control networks, the quality of wireless communication can be controlled in a predictable manner [19, 26, 27, 28], which opens the door for the collaborative design of a wireless car network [6].

Thanks to the joint control of the movement of several vehicles connected by wireless communication, some or all of the following advantages of the transport system are possible:

1) The capacity of the road can be increased by reducing the gaps between vehicles;
2) Energy consumption and pollutant emissions can be reduced by reducing unnecessary speed changes and aerodynamic drag of following vehicles;
3) Driving safety is potentially improved as detection and response times are reduced compared to manually operated vehicles;
4) Consumer comfort can be improved because system behavior is more responsive to changes in traffic, and shorter following intervals can deter the inclusion of other vehicles [2].

The model shows the operation of the system based on the "receiver-transmitter" principle using DSRC communication using MATLAB/ Simulink software [29]. The obtained model allows conducting various system analyzes for the future improvement of information and communication systems in cars.

REFERENCES

[1] USDOT. "Progress Report on the National Roadway Safety Strategy", 2023, [Online], Available: https://rosap.ntl.bts.gov/view/dot/66387.

[2] Z. Wang, Y. Bian, SE Shladover, G. Wu, SE Li, and M.J. Barth, "A Survey on Cooperative Longitudinal Motion Control of Multiple Connected and Automated Vehicles," in IEEE Intelligent Transportation Systems Magazine, vol. 12, no. 1, pp. 4-24, Spring 2020, doi:10.1109/MITS.2019.2953562.

[3] H. Xie, Y. Wang, X. Su, S. Wang, and L. Wang, "Safe Driving Model Based on V2V Vehicle Communication," in IEEE Open Journal of Intelligent Transportation Systems, vol. 3, pp. 449-457, 2022, doi: 10.1109/OJITS.2021.3135664.

[4] "SAE: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems", SAE On-road Automated Vehicles Standards Committee and others, 2021, [Online], Available: https://www.sae.org/standards/content/j3016_202104.

[5] K. Bintoro, "A Study of V2V Communication on VANET: Characteristics, Challenges and Research Trends". JISA (Jurnal Informatika dan Sains).no. 4. pp. 46-58, 2021, doi:10.31326/jisa.v4i1.895.

[6] S.E. Li, et al., "Dynamical Modeling and Distributed Control of Connected and Automated Vehicles: Challenges and Opportunities," in IEEE Intelligent Transportation Systems Magazine, vol. 9, no. 3, pp. 46-58, 2017, doi: 10.1109/MITS.2017.2709781.

[7] S. Alyassri, M. Ilyas, A. Marhoon, and O. Bayat, "Reduction of Packet Error Rate in V2V Communication Based on Machine Learning," 2021 International Conference on Communication & Information Technology (ICICT), Basrah, Iraq, pp. 110-115, 2021, doi: 10.1109/ICICT52195.2021.9568487.

[8] J. Ploeg, E. Semsar-Kazerooni, G. Lijster, N. van de Wouw, and H. Nijmeijer, "Graceful degradation of cooperative adaptive cruise control," IEEE Trans. Intel. Transp. Syst., vol. 16, no. 1, pp. 488-497, 2015, doi: 10.1109/TITS.2014.2349498.

[9] K. Li, Y. Bian, S. Li, B. Xu, and J. Wang, "Distributed model predictive control of multi-vehicle systems with switching communication topologies," Transportation Research Part C: Emerging Technologies, vol. 118, 2020, doi: 10.1016/j.trc.2020.102717.

[10] Z. Wu, B. Waterson, and B. Anvari, "Developing and evaluating a coordinated person-based signal control paradigm in a corridor network," Transportation Planning and Technology, vol. 45:6, pp. 498-523, 2022, doi: 10.1080/03081060.2022.2134128.

[11] W. Yang, B. Wan, and X. Qu, "A Forward Collision Warning System Using Driving Intention Recognition of the Front Vehicle and V2V Communication," in IEEE Access, vol. 8, pp. 11268-11278, (2020), doi: 10.1109/ACCESS.2020.2963854.

[12] C. Mo, Y. Li, and Z. Ling, "Simulation and Analysis on Overtaking Safety Assistance System Based on Vehicle-to-Vehicle Communication", Automotive Innovation. vol. 1, 2018, doi:10.1007/s42154-018-0017-9.

[13] The MathWorks, Inc., Automated driving toolbox. Automated Driving Toolbox Documentation, 2023, [Online], Available: https://ch.mathworks.com/help/driving/index.html?s_tid=CRUX_lftnav.

[14] J.B. Kenney, "Dedicated Short-Range Communications (DSRC) Standards in the United States," in Proceedings of the IEEE, vol. 99, no. 7, pp. 1162-1182, 2011, doi: 10.1109/JPROC.2011.2132790.

[15] L. Cheng, B. Henty, D. D. Stancil, F. Bai, and P. Mudalige, "A fully mobile, GPS enabled, vehicle-to-vehicle measurement platform for characterization of the 5.9 GHz DSRC channel," 2007 IEEE Antennas and Propagation Society International Symposium, Honolulu, HI, 2007, pp. 2005-2008, doi: 10.1109/APS.2007.4395917.

[16] SAE J2735, V2X Communications Message Set Dictionary; SAE International DSRC Committee, Warrendale, PA, USA: 2022, [Online], Available: https://www.sae.org/standards/content/j2735set_20 2211/.

[17] B.S. Karpishen, "Modeling work information and communication systems on cars", Computer technologies and mechatronics. Collection scientific works by materials of the V international scientific and methodical conferences, Kharkiv, Khnadu, pp. 17-20, 2023.

[18] H. Farahneh, F. Hussian, and X. Fernando, "De-Noising Scheme for VLC-Based V2V Systems"; A Machine Learning Approach, Procedia Computer Science, vol. 171, pp. 2167-2176, 2020, [Online], Available: https://doi.org/10.1016/j.procs.2020.04.234.

[19] Z. Hongwei, C. Xin, L. Xiaohui, and J. Xi, "Adaptive Instantiation of the Protocol Interference Model in Wireless Networked Sensing and Control". ACM Transactions on Sensor Networks (TOSN), vol. 10:2, pp 1-48, 2014, doi: 10.10.1145/2530286.

[20] L.M. Karpenko, S.R. Koev, N.B. Kashchena, V. Grushko, and A. Khorosheniuk, "Formation of the stakeholders account model in conditions of sustainable development", Academy of Accounting and Financial Studies Journal, vol. 23, no. 2, pp. 1-7, 2019, [Online], Available: https://repo.btu.kharkov.ua/bitstream/123456789/92 97/3/Kashchen%d0%b0_Formation_of_the_Stakeh olders_article_2019.pdf.

[21] N., Savytska, V., Babenko, H., Chmil, O., Priadko, and I., Bubenets, "Digitalization of Business Development Marketing Tools in the B2C Market", Journal of Information Technology Management", vol. 15, no. 1, pp. 124-134, 2023, doi: 10.22059/jitm.2023.90740.

[22] N. Bezrukova, L. Huk, H. Chmil, L. Verbivska, O. Komchatnykh, and Y. Kozlovskyi, "Digitalization as a Trend of Modern Development of the World Economy', WSEAS Transactions on Environment and Development, vol. 18, pp. 120-129, 2022, doi:10.37394/232015.2022.18.13.

[23] D. Solokha, N. Trushkina, L. Potemkin, and R. Mirkurbanova, "Use of multi-agent simulation modeling for predicting the sales of wholesale trade companies", Journal of Management Information and Decision Sciences, vol. 22, no. 4, pp. 483-488, 2019, [Online], Available: https://www.abacademies.org/articles/use-of-multi-agent-simulation-modeling-for-predicting-the-sales-of-wholesale-trade-companies.pdf.

[24] V. Babenko, M. Pasmor, Ju. Pankova, and M. Sidorov, "The place and perspectives of Ukraine in international integration space", Problems and Perspectives in Management, vol. 15, no. 1, pp. 80-92, 2017, doi: 10.21511/ppm.15(1).2017.08.

[25] V. Babenko, N. Chebanova, N. Ryzhikova, S. Rudenko, and N. Birchenko, "Research into the process of multi-level management of enterprise production activities with taking risks into consideration", Eastern-European Journal of Enterprise Technologies, vol. 1, no 3 (91), pp. 4-12, 2018, [Online], Available: https://doi.org/10.15587/1729-4061.2018.123461.

[26] K.A. Pylypenko, I.V. Babiy, N.V. Volkova, L.K. Feofanov, and N.B. Kashchena, "Structuring economic security of the organization", Journal of Security and Sustainability, vol. 9, no. 1, pp. 7–38, 2019, [Online], Available: https://doi.org/10.9770/jssi.2019.9.1(3).

[27] V.A. Babenko, "Formation of economic-mathematical model for process dynamics of innovative technologies management at agroindustrial enterprises," Actual Problems of Economics, vol. 1, no. 1, pp. 182-186, 2013.

[28] Y.V. Batygin, S.F. Golovashchenko, and A.V. Gnatov, "Pulsed electromagnetic attraction of sheet metals - Fundamentals and perspective applications", Journal of Materials Processing Technology, vol. 213, no. 3, pp. 444-452, 2013, [Online], Available: https://doi.org/10.1016/j.jmatprotec.2012.10.003.

[29] A. Galkin, L. Obolentseva, I. Balandina, E. Kush, V. Karpenko, and P. Bajdor, "Last-mile delivery for consumer driven logistics", Transportation Research Procedia, vol. 39, pp. 74-83, 2019, [Online], Available: https://doi.org/10.1016/j.trpro.2019.06.009.

Practical Experience in DevOps Implementation

Liliia Bodnar[1], Mykola Bodnar[2], Kateryna Shulakova[3,4], Oksana Vasylenko[4], Roman Tsarov[3], and Eduard Siemens[4]

[1]*South Ukrainian National Pedagogical University, Staroportofrankyvska Str. 26, 65020 Odesa, Ukraine*
[2] *LLC B&B Solutions, Dukivska Str. 5, 65000 Odesa, Ukraine*
[3]*State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, 65023 Odesa, Ukraine*
[4]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*
bodnarl79@pdpu.edu.ua, katejojo29@gmail.com, oksana.vasylenko@hs-anhalt.de, rcarev@gmail.com,
eduard.siemens@hs-anhalt.de

Keywords: DevOps, CI/CD, Commercial Experience, Educational Resources, IT Professionals.

Abstract: This scientific article serves as a comprehensive exploration of the amalgamated technical expertise derived from hands-on involvement in commercial projects. Its primary objective is to harmonize the training of upcoming DevOps engineers with the dynamic demands of the real-world market. The focus of our investigation lies in the intricate stages of continuous integration (CI) processes, meticulously shaped by the practical experiences of seasoned DevOps engineers. Delving into these stages, we unveil invaluable insights that extend their applicability to benefit both IT instructors and active DevOps practitioners. The elucidation of these stages is firmly grounded in the pragmatic utilization of a diverse array of DevOps practices. By drawing from the multifaceted experiences encountered in real-world scenarios, our article presents a nuanced understanding of the challenges and triumphs that permeate the field of DevOps. This nuanced approach not only enhances the theoretical knowledge imparted to future engineers but also provides a reservoir of practical wisdom that can be readily applied in professional settings. That is, our undertaking encapsulates the core objective of bridging the disjunction between theoretical knowledge and practical application within the domain of DevOps.

1 INTRODUCTION

In recent years, DevOps has emerged as one of the most popular methodologies that bridges the gap between software development and operations. The DevOps methodology involves the collaboration of development teams and operational specialists to create software with the goal of swiftly deploying IT solutions to the market. Research [1] has determined that the primary objective of DevOps is efficient implementation of new features and reduction of time frames from development to production. These findings underscore the impact of this aspect on improving software development outcomes. However, DevOps is more than just a methodology; it is a culture of collaboration and automation that contributes to reduced burnout and enhanced productivity [2]. Research involving over 36,000 professionals worldwide regarding DevOps practices [3] supports successful software delivery and operational efficiency. Automation, in particular, helps ensure code quality and reduces the number of

defects [4]. Additionally, a systematic literature review [5, 6] emphasizes the importance of integrating security into the DevOps process to ensure secure development and effective product delivery. This highlights how modern enterprises in the digital transformation era face new challenges, compelling IT teams to adapt to emerging technologies and embrace customer-centric approaches. Consequently, traditional software development models like the Waterfall Model and Spiral Model have given way to agile practices such as Kanban and Scrum [7]. DevOps has become the most relevant approach in software development, leading to an increased demand for individual IT solutions. Rapid software deployment is no longer just a competitive advantage; it's a survival requirement in today's market. The demand for DevOps specialists has steadily grown over the past few years, and this trend is expected to continue. Alongside demand, the responsibilities of DevOps engineers have expanded. Today, they are not only responsible for process automation and infrastructure configuration but also collaborate with

developers, testers, and system administrators. DevOps engineers are highly sought after in the job market, yet existing educational programs don't always provide them with the necessary up-to-date knowledge and skills.

In this study, we explore key technical competencies and practical experience that make DevOps engineers marketable. Our research aims to provide recommendations for preparing DevOps professionals who meet modern requirements. By analyzing current trends and drawing from the practical experience of DevOps engineers, we identify the essential components that contribute to a successful CI process. Understanding these stages and their nuances empowers aspiring DevOps specialists to navigate the complexities of CI effectively.

2 ANALYZE THE STATE OF THE PROBLEM

As a result of our analysis, we've identified key components that make DevOps engineers highly sought after in the job market (Figure 1), considering both recent trends [8] and practical insights [9-11]:

- Deep understanding of Software Development Life Cycle (SDLC). A DevOps professional must have a clear understanding of all SDLC stages, from planning and design to testing and deployment. This understanding enables effective collaboration with development, QA, and operations teams to ensure a smooth and efficient software development and release process.
- Proficiency in Infrastructure as Code (IaC) principles. DevOps engineers should be able to describe and manage infrastructure using code, leveraging tools like Ansible, Puppet, Chef, or Terraform. IaC allows for automated infrastructure creation and configuration, making it more efficient, reliable, and scalable.
- Thorough configuration management knowledge. DevOps experts need to grasp configuration management principles and utilize tools such as CMDB, Puppet, Chef, or Ansible for centralized configuration management of infrastructure and software. This ensures consistent and coordinated configuration across all systems, reducing the risk of errors and improving reliability.
- Confident mastery of Continuous Integration and Continuous Delivery/Deployment (CI/CD). DevOps professionals should be

adept at setting up CI/CD processes, which automate software build, testing, and release. CI/CD significantly reduces software release time and enhances its quality.
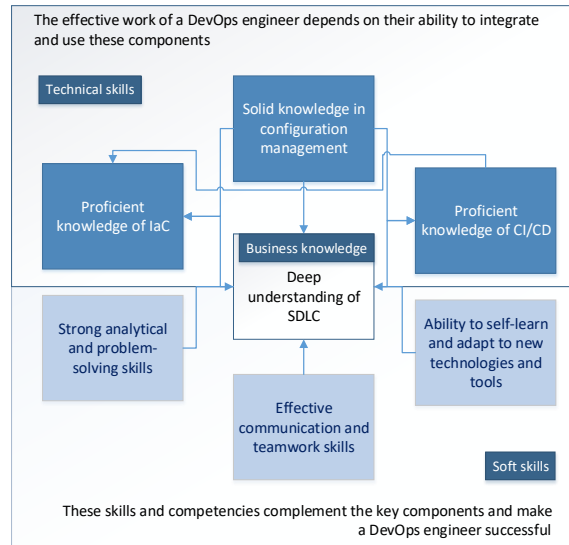
Figure 1: Components of a successful DevOps engineer.

In addition to these key components, a successful DevOps engineer should also possess [12]:

- Strong analytical and problem-solving skills.
- Effective communication and teamwork abilities.
- Adaptability to learn and embrace new technologies and tools.

Mastering these knowledge areas, skills, and competencies makes a DevOps professional a valuable asset for any organization striving to successfully implement DevOps practices.

Currently, information about DevOps specialists is scattered across the internet, and there is no single guide that describes best practices and offers solutions for effectively combining them to achieve optimal results. In light of this, we have decided to create recommendations that allow us to consolidate our experience and catalogs, showcasing a systematic approach to problem-solving. From Figure 1, it is evident that describing all components is a rather labor-intensive process. Therefore, within the scope of this article, we have developed an approach for one of the CI/CD (Continuous Integration/Continuous Delivery/Continuous Deployment) components: continuous deployment. CI/CD represents the practice of automatically delivering and deploying applications to the production environment after successfully passing all CI stages (Figure 2). One of

its distinguishing features is that continuous delivery deploys code continuously using manual triggers, while continuous deployment automates deployment without human intervention. While continuous delivery and deployment processes do not require special approaches, continuous integration involves developing a more detailed strategy.
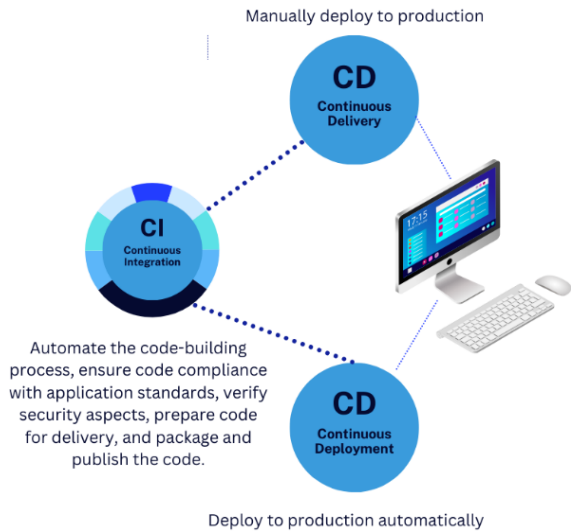
Figure 2: CI/CD basic workflow.

The core idea of CI is to automate the code build process, verify code compliance with application standards, assess security factors, prepare code for delivery, package it, and publish it. Imagine assembling a car. Each developer creates individual components, such as the engine or doors. CI acts as the conveyor belt that brings these components together. It automatically "transfers" them from one stage to another as soon as they are ready. This approach allows for early error detection and correction during development, ensuring swift and reliable delivery of new features. By reducing the time between developing new functionality and putting it in users' hands, CI also minimizes the risk of deployment issues.

Analyzing works [13-14], we can identify key aspects of CI:

- Automated Builds. Developers commit their code changes frequently, triggering automated builds. These builds compile the code, run code quality tests, and generate deployable artifacts.
- Unit Testing. CI systems execute unit tests to validate individual components of the codebase. Early detection of defects helps maintain code quality.

- Quality Assurance Testing. CI ensures that code changes integrate seamlessly with existing code.
- Artifact Generation. The CI process produces deployable artifacts (e.g., binaries, containers) ready for deployment.
- Feedback Loop. Developers receive immediate feedback about their code changes, allowing them to address issues promptly.

3 DEVELOPMENT OF A SYSTEMATIC METHOD OF SOLUTION

Despite the fact that the CI workflow shares commonalities with Agile methodologies, it has its specific elements. While there isn't a universal algorithm for CI processes, we can identify a general structure for the workflow. Figure 3 illustrates the six stages involved:

1) Source code build stage. This initial step involves compiling and building the source code.
2) Self-test stage. Automated tests are executed to validate the functionality of the code. These tests cover unit testing, code quality testing, and other quality checks. It ensures that the code adheres to coding standards and conventions.
3) Artifact preparation stage. Here, the code is packaged into deployable artifacts. These artifacts can be binaries, containers, or other forms suitable for deployment.
4) Quality Assurance (QA) stage. QA processes, including manual testing and user acceptance testing, ensure that the artifacts meet the desired quality standards.
5) Security checking stage. Security scans and vulnerability assessments are performed to identify and address any security risks.
6) Artifact publishing stage. Finally, the validated and secure artifacts are published to the appropriate environment (e.g., staging or production).

CI is a cornerstone of successful DevOps practices, enabling faster development cycles, improved collaboration, and higher-quality software delivery.

Let's analyze each stage of CI and formulate a systematic approach for each stage.

Figure 3: The general structure of the workflow CI.

3.1 Source Code Build Stage

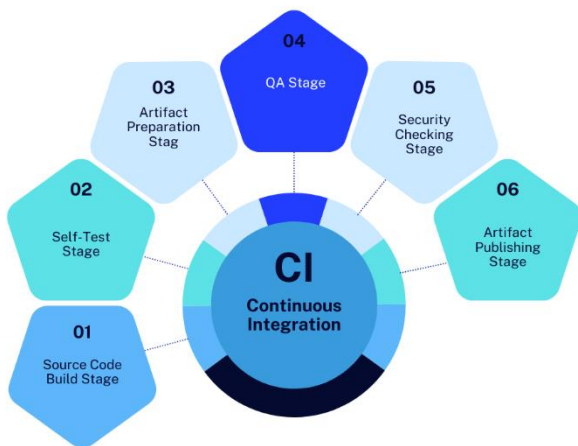In software development pipelines, the build stage is crucial for transforming source code into executable artifacts. While developers often handle this stage, a proficient DevOps Engineer should also understand the process thoroughly. After all, the DevOps Engineer will be responsible for configuring and maintaining this critical step in the pipeline.

The primary challenge lies in the fact that each programming language has its own specific build tools. For instance:

- .NET: Uses the dotnet builder.
- Java: Relies on the Maven build system.
- React (JavaScript): Requires its own set of build tools.

However, there are broader solutions that can simplify the process by taking a more universal approach. Examples include Gradle and NPM. These tools cater to specific domains: NPM is excellent for building web user interface (UI) components based on JavaScript frameworks like Angular and React, while Gradle serves as a versatile solution for Java-related applications (including Kotlin, Groovy, and Android). Unfortunately, neither Gradle nor NPM is suitable for building .NET applications.

So, how does a DevOps Engineer navigate this challenge? Let's explore some strategies:

A) Developer Responsibility Approach:
 1) Some DevOps teams consider the build stage purely a developer's responsibility. They allow developers to configure the build process locally, assuming it will work seamlessly in a multiuser environment.
 2) However, this approach has limitations. Developers often tailor their local setups to

specific use cases, which may not translate well to a shared pipeline. This can lead to instability and friction between developers and DevOps engineers.

B) Universal Builders Approach:
 1) A more pragmatic solution involves using universal builders. DevOps Engineers create general build configurations based on these tools.
 2) Developers then describe their specific build processes within these universal solutions. For example:
   - The DevOps Engineer ensures the correct execution of the builder (e.g., Gradle or NPM).
   - Developers focus on defining the build steps and dependencies within the chosen tool.
 3) This compromise streamlines the DevOps part, making it more predictable, easier to manage, and stable.
 4) It also clarifies the division of responsibilities: DevOps handles the orchestration, while developers focus on the specifics of their application builds.

In summary, a collaborative approach that leverages universal builders strikes a balance between stability, flexibility, and clear responsibilities in the build process. DevOps Engineers play a crucial role in ensuring smooth execution, while developers contribute their domain-specific knowledge to the configuration.

3.2 Self-Test Stage

In software development pipelines, the build stage plays a critical role. It serves as the last checkpoint where developers can review their code before it proceeds to subsequent stages. These subsequent stages are beyond their control, as they fall outside the developers' direct responsibility.

The build stage is an ideal place to execute unit tests, lint checks, and other validations. These tests are typically short and efficient, minimizing the time spent in this phase. However, the main challenge lies in the human factor. Writing comprehensive unit tests requires disciplined processes and a strong commitment, which developers may not always prioritize. Common reasons include a lack of clear code standards and insufficient requirements for adhering to those standards.

While this leniency might be acceptable in startup environments, it becomes unacceptable for stable projects. In mature projects, where the codebase is

well-established, any deviation from coding standards can lead to unpredictable outcomes.

Another challenge in this stage pertains to reporting and handling failures. CI pipelines often run in the background, lacking the UI where developers can directly view test results. Additionally, when multiple pipeline runs occur concurrently, the previous results may get overwritten by the current run. To address this, we rely on middleware tools such as SonarQube or SonarCloud. These tools integrate seamlessly with CI platforms and serve as repositories for test reports.

However, merely collecting reports isn't sufficient. We also need mechanisms to halt the pipeline when tests fail. This functionality is typically implemented within the chosen middleware.

Here are two trade-offs to consider for this step:

1) Base Code Formation. If the base code is still evolving, and the development team lacks strict code standards, you might choose to omit this stage initially. However, it's essential to add it as the team matures and establishes clearer guidelines.
2) Middleware Integration. For projects with well-defined standards, integrate middleware tools like SonarQube or SonarCloud into your CI process. These tools not only store reports but also provide stop triggers to prevent pipeline progression upon test failures.

In summary, striking a balance between flexibility and stability ensures that the DevOps role remains effective in managing the build process. Developers contribute their expertise within the universal framework, while DevOps Engineers orchestrate the overall pipeline execution.

3.3 Artifact Preparation Stage

The stage is optional, but in line with the modern trend of creating platforms based on microservices, the importance of stages is growing. These stages serve as containers for microservices, where each microservice can reside. The use of Docker containers for this purpose is particularly advantageous. By preparing a standardized environment for each service within a container, you can ensure that the application will function consistently from scratch.

Additionally, there are significant benefits for DevOps engineers. Working with Docker containers, rather than independent applications, allows for the unification and automation of the delivery and runtime processes. However, a key challenge arises during the CI/CD pipeline: how to package an application into a Docker image. When using a virtual machine (VM) as a build server, installing the Docker daemon and running the docker build command with a prepared Dockerfile is straightforward. But what if your CI/CD pipeline relies on Docker containers as build servers, as is the case with GitHub Actions, GitLab CI/CD, or Jenkins with Kubernetes plugins? In such scenarios, you must address several issues, including providing the Docker daemon inside a Docker container with access to the Linux socket.

To tackle this problem, the Google team developed an application called 'Kaniko.' Kaniko can create a Docker image based on a Dockerfile without requiring the Docker daemon. Instead, it treats the Docker image as a simple archive with a specific internal structure. Kaniko parses the Dockerfile based on the defined run points, creates the archive package (Docker image), and pushes it to any Docker registry. Kaniko strikes an ideal compromise for this stage of the process.

3.4 QA Stage

The stage is ideal for running short-lived quality assurance tests. Specifically, we focus on tests like smoke tests, rather than larger ones such as performance or integration tests.

However, the main challenge with using a stage is that some teams treat the CI tool as the primary environment for running all tests conducted by QA engineers. This approach is fundamentally flawed because CI tools serve multiple purposes. While one of their functions is to execute scripts triggered by events, this functionality is primarily suited for running automatic test scripts.

Unfortunately, CI tools cannot fully provide the range of QA functionality required. For instance, they may lack proper mechanisms for collecting and viewing test reports or adjusting the runtime environment to consistently run the same set of tests automatically and manually.

To strike a balance, consider splitting your test suite into two groups:

1) Short-lived tests. These can be seamlessly integrated into CI pipelines.
2) Independent tests. These should be executed using dedicated QA tools like TestKube or similar alternatives.

When integrating QA tools with CI/CD pipelines, choose wisely. For instance:

- CI integration makes sense for controlling and automatically running short-lived, CI-compatible tests.

- CD integration is more suitable for other types of tests, as it accommodates specific triggers that may only be available during CD pipelines.

Selecting the right integration ensures correct automation and flexibility.

3.5 Security Checking Stage

The stage is optional, but in recent times, its importance has been growing. First and foremost, security considerations need to be defined. The field of security testing is vast, with multiple levels, each having its own penetration goals and scenarios. While it's not feasible to run the entire spectrum of security tests during the continuous integration (CI) pipeline due to various constraints (such as context or the inability to accurately assess application status), certain types of tests can indeed be executed within this pipeline.

Specifically, let's focus on vulnerability tests for applications. The goal of these tests is to identify security vulnerabilities in the code before it reaches the production stage. Scanning code for vulnerabilities is a challenging task that demands extensive knowledge not only about potential weaknesses in your application but also in the dependencies it relies upon.

The primary challenge with this stage lies in the lack of a universally understood approach. Engineers often seek the simplest solution—one that works out of the box without requiring deep dives into complex configurations. Unfortunately, such one-size-fits-all solutions don't exist.

A reasonable compromise is to explore the various CI plugins available in the market. By delving into these options, you can select the most suitable solution for your project, covering at least 50% of your security needs.

In different projects with varying requirements, we've encountered solutions like BlackDuck, which is intricate to understand and may not always provide stable results. On the other hand, X-Ray scanning, integrated into JFrog Artifactory, offers a simpler experience. JFrog takes care of X-Ray's functionality, although it might be less powerful in terms of where scanning results are visible.

Given the abundance of solutions in the global market, it's essential to find the one that best fits your project's specific context.

3.6 Artifact Publishing Stage

The stage is the finishing stage. The appearance and functionality of this stage are determined by the programming language used. Each language has its own build process and produces specific types of artifacts. An artifact is a compiled and packaged file ready for deployment within the application process. For instance, in Java, artifacts could be JAR files, application packages, WAR files, or even Docker image files.

However, a critical challenge arises: how to manage these artifacts effectively. Each type of artifact requires storage in a specific registry. While some artifacts (like JAR, WAR, and application files) can share a common repository (such as Maven or Gradle), others (like .NET artifacts) necessitate separate repositories (e.g., NuGet for .NET or npm for JavaScript).

To address this, consider using universal storage solutions like Sonatype Nexus Repository or JFrog Artifactory. These platforms offer a wide range of repository types, allowing you to create storage tailored to your project's needs. By integrating just one solution with your CI/CD tools, you simplify artifact management.

Beyond this primary compromise, there are additional tips to reduce repository complexity. For example:

- Microservices and Docker. If you're building a microservices platform based on Docker containers, you don't need to store individual application artifacts. Instead, focus on storing Docker images containing the application artifacts. This approach streamlines repositories within a single Docker registry.
- Helm Chart Repositories. Helm can utilize a Docker registry as a Helm chart repository. While this approach has limitations, it's worth considering if needed.

In summary, aim for a repository setup that aligns with your project's requirements while minimizing unnecessary complexity. However, native repositories offer features that can be highly beneficial for the CI/CD process. If you have the opportunity to use them, it's advisable to do so.

4 CONCLUSIONS

In our research, we delved into the key components that make DevOps engineers highly sought-after in the job market. Despite widespread adoption, there remains a scarcity of comprehensive resources that consolidate practical knowledge for aspiring DevOps professionals. Our study aims to bridge this gap by

offering a holistic view of DevOps practices based on real-world commercial experience.

Our primary focus was on CI. We developed recommendations outlining six approaches for implementing CI. Additionally, we provided examples and approaches for applying these practices in real-world projects.

By considering these critical factors, IT companies can strengthen their position in the international market by hiring skilled DevOps engineers, thereby creating a dynamic environment that encourages growth, innovation, and sustainable global collaboration.

Our research contributes to the existing body of knowledge by providing a comprehensive overview of DevOps practices and their associated benefits.

These findings can be leveraged by:

- Universities: To develop and implement effective DevOps education programs.
- IT companies: To recruit and train skilled DevOps engineers.
- DevOps engineers: To enhance their skills and knowledge.

Future research directions include:

- Investigating other key components of DevOps, such as infrastructure as code (IaC) and containerization.
- Developing case studies on the successful implementation of DevOps practices in different industries.

We believe that this research will help to advance the field of DevOps and make a contribution to the IT industry.

ACKNOWLEDGMENTS

REFERENCES

[1] L.E. Lwakatare, P. Kuvaja, and M. Oivo, "Relationship of DevOps to Agile, Lean and Continuous Deployment," in P. Abrahamsson et al. (eds), Product-Focused Software Process Improvement. PROFES 2016. Lecture Notes in Computer Science, vol 10027, pp 399-415, Springer, Cham, 2016, [Online]. Available: https://doi.org/10.1007/978-3-319-49094-6_27.

[2] M. Sánchez-Gordón and R. Colomo-Palacios, "Characterizing DevOps Culture: A Systematic Literature Review," in I. Stamelos et al. (eds), Software Process Improvement and Capability Determination. SPICE 2018. Communications in Computer and Information Science, vol. 918, pp. 3-15, Springer, Cham, 2018, [Online]. Available: https://doi.org/10.1007/978-3-030-00623-5_1.

[3] "2023 State of DevOps Report," [Online]. Available: https://cloud.google.com/devops/state-of-devops, [Accessed: 28 Dec 2023].

[4] D. Belcher, "Three Trends That Will Transform DevOps in 2023," 2023, [Online]. Available: https://devops.com/three-trends-that-will-transform-devops-in-2023, [Accessed: 28 Dec 2023].

[5] M. Gasparaite, K. Naudziunaite, and S. Ragaisis, "Systematic Literature Review of DevOps Models," in M. Shepperd et al. (eds), Quality of Information and Communications Technology. QUATIC 2020. Communications in Computer and Information Science, vol. 1266, pp. 184-198, Springer, Cham, 2020, [Online]. Available: https://doi.org/10.1007/978-3-030-58793-2_15.

[6] T. Leppanen, A. Honkaranta, and A. Costin, "Trends for the DevOps Security. A Systematic Literature Review," in B. Shishkov (eds), Business Modeling and Software Design. BMSD 2022. Lecture Notes in Business Information Processing, vol. 453, pp. 200–217, Springer, Cham, 2022, [Online]. Available: https://doi.org/10.1007/978-3-031-11510-3_12.

[7] M. Rütz, "DevOps: A Systematic Literature Review," Twenty-Seventh European Conference on Information Systems (ECIS2019), Stockholm-Uppsala, Sweden, 2019, pp. 1-16, [Online]. Available: https://www.researchgate.net/publication/335243102_DEVOPS_A_SYSTEMATIC_LITERATURE_REVIEW.

[8] M. Modi, "The Future of DevOps: 2024 and Beyond," 2023, [Online]. Available: https://www.knowledgehut.com/blog/devops/future-of-devops, [Accessed: 29 Dec 2023].

[9] "DevOps practice in Ukraine, 2024," [Online]. Available: https://bandbsolution.com.ua/en, [Accessed: 11 Jan 2024].

[10] "DevOps practice in Germany, 2024," [Online]. Available: https://www.fme.de/dienstleistungen/technology-services, [Accessed: 10 Jan 2024].

[11] "DevOps practice in USA, 2024," [Online]. Available: https://www.fme-us.com. [Accessed: 10 Jan 2024].

[12] G.B. Ghantous and A. Gill, "DevOps: Concepts, Practices, Tools, Benefits and Challenges," PACIS 2017 Proceedings, pp. 1-13, 2017, [Online]. Available: http://aisel.aisnet.org/pacis2017/96.

[13] L. Gryzun, V. Pikalova, and L. Bodnar, "Issues of Selecting an Instrument of Continuous Integration for Software Automated Testing," International Scientific and Practical Forum «Digital Reality» 2023 [«FDR2023»]: Cybersecurity and information technologies in the hybrid wars conditions, [Accessed: 2 Dec 2023].

[14] "DevOps Tools - GitLab CI/CD," [Online]. Available: https://digitize01.com/devops-tools-gitlab-ci-cd, [Accessed: 29 Dec 2023].

The Use of ICT for the Development of Foreign Language Communicative Competence in Hospitality Industry Students

Natalia Plakhotniuk, Oksana Chernysh and Oleg Makarevych

Zhytomyr Polytechnic State University, Chudnivska Str. 103, 10005 Zhytomyr, Ukraine

nat.pl2002@gmail.com, chernyshoxana@gmail.com, makarevoleg@ukr.net

Keywords: Information and Communication Technologies, Foreign Language Communicative Competence, Future Hospitality Specialists, Educational Blog.

Abstract: The article considers the didactic potential of the use of educational information technologies in the process of learning a foreign language for specific purposes by hospitality students. Based on the analysis of scientific works, the study proves the importance of introducing information and communication technologies in foreign language training of future specialists. With the help of a questionnaire and an interview clarifies the state of information and communication technologies in forming foreign language communicative competence of hospitality students. The paper specifies the peculiarities of foreign language communicative competence of hospitality students. Moreover, it highlights its components. The study presents the experience of introduction of information and communication technologies in foreign language training of future hospitality specialists. It proves that video hosting on YouTube, online presentations, mind maps, and educational blogs effectively influences the formation of students' foreign language communicative competence. Thus it increases motivation to master foreign language material and develops analytical, creative, speaking, writing, listening, and reading skills. The article describes the didactic possibilities of the authors' educational blog to organize independent work of students and partnership learning. Hence, it develops self-realization skills and improves foreign language communicative competence. Furthermore, the authors consider the didactic potential of mind maps in forming foreign language communicative competence and actualizing their mental activity. Finally, the paper experimentally tests the effectiveness of introducing information and communication technologies in learning a foreign language for specific purposes in hospitality students.

1   INTRODUCTION

Rapid and constant progress and improvement of the higher education system aim to train modern competitive professionals in various economic, social, and cultural fields to create a set of innovative educational and methodological training for students. The development of information and communication technologies (ICT) and their introduction into higher education institutions' educational process significantly contribute to acquiring knowledge and improving skills. It puts skills at a qualitatively new level, ensures creative personal growth, and highly qualified training for future professionals.

"Strategy for Higher Education in Ukraine for 2021-2031" highlights the significance of the informatization of education and modernization of information and resource provision. In addition, it outlines the importance of the digitalization of educational processes, continuous improvement of education on an innovative basis, and improving information training students. The priority of education lies in the use of ICT. ICT contributes to a new level of training for future professionals and ensures the availability and effectiveness of education, preparing the younger generation to function in the information society [1].

In recent years, there has been a tendency to increase the number of hours for foreign language training for students in most educational institutions, particularly for learning a foreign language in a professional field. A proper level of foreign language communication competence (FLCC) as a component of professional competence allows the future specialist to become a full-fledged competitive entity in the labour market and master modern technologies necessary for personal and

professional development. The effective development of the FLCC requires innovative methods and means of its formation in the modern education system. They allow to increase the motivational component of the educational process, promote the development of analytical skills and critical thinking, skills of independent work, and processing of a large amount of information in a short period. At the same time, according to our observations, in the frames of learning a foreign language, students use ICT mainly to create presentations on the topic, electronic translators in the Google system, and video conferencing during distance learning.

Thus, the importance and urgency of integrating ICT for the development of foreign language communicative competence among students as future professionals cannot be overstated. In today's interconnected global landscape, proficiency in foreign languages is a valuable skill that enhances professional opportunities and fosters intercultural understanding. Here's why leveraging ICT is crucial in this endeavour due to the following factors:

1) Access to Authentic Resources: ICT provides students with access to a vast array of authentic resources in the target language, such as articles, videos, podcasts, and social media platforms. Engaging with these resources exposes students to real-world language usage, enriching their vocabulary, comprehension skills, and cultural awareness.
2) Global Communication: ICT facilitates communication and collaboration with native speakers and language learners worldwide through video conferencing, social networking sites, and online language exchange platforms. Engaging in authentic conversations in the target language fosters linguistic confidence, fluency, and intercultural competence, preparing students to effectively communicate in diverse professional contexts.
3) Multimedia Integration: Incorporating multimedia elements such as audio recordings, videos, and interactive simulations into language instruction enhances engagement and comprehension. ICT enables the creation of multimedia-rich learning materials that cater to different learning preferences and stimulate multiple senses, leading to deeper understanding and retention of language concepts.
4) Adaptation to Technological Advancements: In today's rapidly evolving digital landscape, proficiency in ICT is essential for professional

success across industries. Integrating ICT into language learning equips students with valuable digital literacy skills, including information literacy, critical thinking, and problem-solving, which are integral to navigating and thriving in the digital age.

The article aims to explore didactic possibilities and potential of ICT tools to develop future hospitality professionals' foreign language communicative competence.

2 RESEARCH BACKGROUND

Many outstanding scientists have already considered the advantages of ICT for educational purposes, namely V. Bykov, L. Fiorella, R. Gurevich, M. Zhaldak, M. Kademiya, N. Morse, R. Mayer, Y. Ramsky, V. Trainev, J. Anderson, K. Ratheesvari, D. Sharma and many others. According to K. Ratheesvari, ICT significantly changes and modernizes the educational process, providing teachers and students with more personal and professional growth opportunities, developing critical thinking, and expanding information potential and communication skills [2]. Implementing ICT in the educational framework of institutions fosters the enhancement of cognitive abilities such as critical thinking, memory retention, analytical aptitude, communication proficiency, creativity, research acumen, self-expression, and self-motivation [3].

ICT tools play a significant role in improving the educational process by providing various opportunities such as storing vast amounts of information, facilitating access to diverse data sources, delivering relevant information to students, and offering versatility in presenting information [3].

Significant research proves the benefits of ICT in the process of professional training of future teachers of various specialties, including philologists (S. Galetsky, T. Gorokhova, M. Zlativ, K. Klimova, G. Korytska, I. Kostikova, N. Ostapenko, T. Yatsenko, etc.), mathematicians (M. Zhaldak, Y. Trius), junior classes (D. Mazokha, L. Petukhov, I. Shaposhnikov). ICT tools increase motivational and cognitive activities through interactive content, provide individualization and differentiation of learning, and optimize time spent searching and analyzing information. Moreover, they develop research skills, ensure free access to global information systems and expand control over acquiring knowledge, skills, and abilities [4].

Scientists admit that ICT tools is an effective method of achieving educational goals which widens the teaching opportunities. They help to make the process of learning quicker and more interesting, provide absolutely new methods of teaching and studying [5].

The issue of introducing ICT tools in the process of learning foreign languages and forming the FLCC of students of non-language faculties is the subject of scientific research by S. Amelina, I. Gotsynets, S. Johnson, S. Kostrytska, O. Kraevska, V. Krasnopolsky, L. Rosencrance, I. Semeryak, O. Tinkalyuk, and many others.

Scientists are considering using the latest digital technologies to develop FLCC future professionals in various specialties: engineers, economists, lawyers, marketers, customs officers, etc. The scientific work of I. Skrill is of particular interest. It deals with forming professionally oriented English lexical competence of hospitality students using video materials. The dissertation substantiated the expediency of using YouTube, web forums, and video chats to effectively assimilate and reproduce foreign lexical material. According to the scientist, authentic videos provide student-centric, communicative, and interactive approaches to developing foreign language fluency. Thus, it promotes self-realization, self-assessment, and self-correction of learning outcomes taking into account future professionals' didactic interests and needs [6].

Despite the significant achievements in the study of innovative digital methods in the professional training of non-language students, the problem of using ICT in the foreign language practice of hospitality students requires more consideration. Moreover, observations of students learning a foreign language for professional purposes show that only a small proportion of teachers use modern ICT tools in their professional activities. Therefore, it is appropriate to find and use ICT for the practical study of a foreign language by future specialists in non-language specialties, particularly students majoring in hospitality.

Systematic ICT introduction in the process of professional training of future specialists creates opportunities to ensure the implementation of high-level educational activities. Moreover, it combines traditional and innovative education. Thus it allows participants to realize their abilities and learning needs following the Law of Ukraine "On Higher Education" [7].

Many outstanding scientists emphasize the relevance and effectiveness of ICT tools in students' foreign language training. For example, according to

V. Krasnopolsky, educational information technologies modernize the implementation of didactic principles and learning strategies and significantly improve the content and means of organizing future professional activities of students of different specialties [8].

When multimedia materials are thoughtfully chosen to align with learners' requirements, processing methods, and strategies, they yield particular results in shaping intricate skills and capabilities, fostering the development of the competencies they aim to attain. [9].

ICT learning is the purposeful unity of software methods, tools, and resources that provide joint activities for teachers and future professionals. It is the communication, creation, storage, and transformation of information to solve educational problems [3, 10, 11, 12].

The main means of ICT in the educational activities of the higher educational institutions are multimedia (video conferencing, video devices for information design, Power Point presentations, YouTube video hosting), software (computer testing systems, virtual dialogue, computer simulators, electronic textbooks, virtual laboratories, cloud-based educational environment) and telecommunications (educational Internet sites and projects, virtual educational platforms Moodle, Whiteboard) [14].

The didactic potential of ICT in the context of ICC formation contributes to sustainable motivation to learn a foreign language through new forms of language development, individualization and creativity in learning language units and grammatical structures, and dynamic development of dialogic speech by simulating typical situations in future professional activities using multimedia technologies, effective control and reflection processes.

Scientific studies prove that the nature, structure, and ways of ICC formation of non-language faculties students are an integrative system of interconnected components, characterized by the ability to freely understand and produce speech to solve professionally important tasks [15, 16, 17, 18].

FLCC includes professional vocabulary knowledge and knowledge of professional subjects, a sufficient level of use of a common foreign language, its culture, and its features. Regarding the FLCC of hospitality students, foreign language knowledge and skills are relevant given Ukraine's entry into the European space and expanding the scope of foreign contacts and updated labour market requirements. Hospitality specialists should establish

contacts with foreign partners, communicate with foreign clients, and analyze and transform professionally relevant information from foreign sources. Thus, we view foreign language communicative competence as an integrative quality of a personality and system-value education [13]. It aims at obtaining and processing professionally relevant information in a foreign language. It also includes creative application in hotel and restaurant services and skills in freely conducting international activities to address professionally important issues.

Scientific works review [15, 17, 19] allowed to highlight the following components of FLCC hospitality students: motivational and value (motives, aspirations, goals, understanding of the importance of learning and using a foreign language in professional activities); cognitive-activity (lexical, phonetic, orthographic, linguistic knowledge and skills of their creative use in professional discourse); reflexive-evaluative (skills of self-assessment and further design of foreign language communicative activities).

3 METHODOLOGY

The aim of the study requires the use of the following methods:

- theoretical: analysis, synthesis, comparison, and systematization of scientific views on the problem of the nature and structure of the FLCC of non-language faculties students, content, and components of ICT for educational purposes in educational institutions;
- empirical: observation of the work of freelance teachers, the survey of teachers and students, pilot study to determine the frequency of ICT use in foreign language training of hospitality students, interviews, generalization of pedagogical experience in the professional training of future specialists to assess their level of ICC formation;
- Smirnov method of statistical data processing to determine the effectiveness of ICT in the formation of FLCC hospitality students.

The investigation presents the survey results with hospitality students of Zhytomyr Polytechnic State University. It includes 83 hospitality students of the 3d and 4th courses who answered the following questions:

1) What ICT tools do foreign language teachers use in the classroom?

2) What ICT tools do foreign language teachers use when conducting on-line classes?
3) What ICT tools do you use to learn a foreign language?
4) How do ICT help to increase motivation to learn a foreign language, better master the lexical and grammatical material of the topic and develop communication skills?
5) What ICT tools do you consider the most effective in learning a foreign language?
6) What ICT tools have you experienced as the most successful ones in developing communication skills?
7) What ICT would you like to use in foreign language classes?
8) What prevents you from using the ICT in foreign language learning?

4 RESULTS AND DISCUSSION

The study of ICT use in the formation of FLCC of future hospitality specialists undergoes three stages:

- orientation and search (determining the needs of students in theoretical knowledge and practical application of ICT tools to master a foreign language professionally);
- cognitive activity (formation of lexical-grammatical, stylistic, linguistic knowledge and their creative use in solving professional problems in a foreign language with the help of ICT);
- generalizing-reflexive (analysis, systematization and generalization of the results of experimental work, determination of further goals of formation of FLCC of future specialists).

Thus, during the first stage of the study, students took part in a questionnaire, group, and individual survey on the use of ICT in foreign language teaching for professional purposes both in practice and in independent work, and their interest in using innovative ICT to improve language practice. The results show that teachers often use ICT in foreign language classes, but mainly during distance learning (videoconferencing, uploading tasks on the University educational portal, using Viber and Telegram to send tasks to students and get feedback). During the classes on campus, not many teachers use ICT to teach foreign languages. For example, only 28% of foreign language teachers use a multimedia projector and 26% of them provide YouTube videos to design lexical and grammatical

tasks and 32% create computer module tests (Figure 1). Instead, most teachers use e-versions of the textbooks and audio recordings downloaded to their phones, tablets, and laptops.
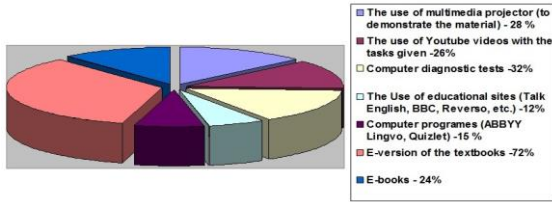

Figure 1: The use of ICT by foreign language teachers.

When preparing foreign language assignments, 88% of respondents always use on-line dictionaries, mainly Google Translate and Cambridge Dictionary, to translate words and phrases, select synonyms and successful stylistic units. 64 % search for information on Google and create multimedia presentations. However, a small percentage of students (6%) use on-line apps in learning a foreign language (Figure 2). Among the reasons for not paying a decent attention to on-line apps for solving educational tasks in a foreign language, the respondents mentioned their ignorance in using them, and they didn't understand their role in the development of their foreign language communicative competence.
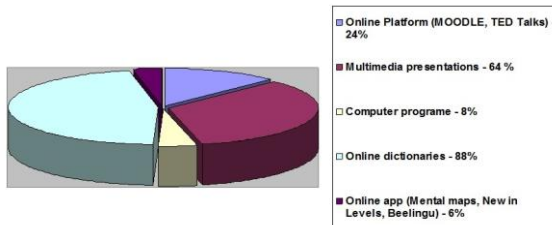

Figure 2: The use of ICT tools by students in foreign language classes preparation.

Most future hospitality specialists think that ICT increases motivation to learn a foreign language, master topic lexical and grammatical material, and develop communication skills. However some students do not consider the importance of ICT tools in developing their FLCC, believing that traditional tools (textbooks, dictionaries, communication with native speakers) have a better effect on language acquisition. The only problem is that it is not easy to come across a native speaker to practice the necessary skills. It is quite convenient to find such a person on-line, searching for some blogs, educational and entertainment sites.

The respondents name the most effective electronic means of forming FLCC (Figure 3). They are: videos (with designed tasks); blogs providing an opportunity to communicate with the teacher and other participants in the learning process; an educational portal with relevant materials and assessment tasks posted; cloud services; applications, computer tests. It should be mentioned that all the apps and other tools are numerous but are aimed at learning and mastering general and business English. When analyzing the ICT tools for developing communicative competence of hospitality industry students we haven't found much. That's why we decided to create some educational tools for them.
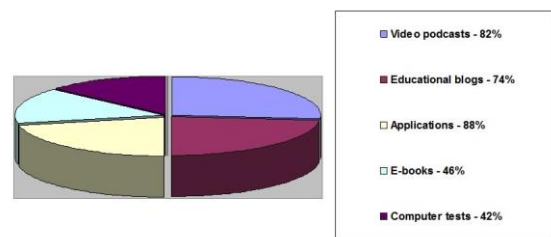
Figure 3: The most effective ICT tools in forming FLCC of hospitality students (according to their opinion).

The cognitive-activity stage of the experiment is aimed to form and develop lexico-grammatical, stylistic, and linguistic knowledge and skills of hospitality students with the help of ICT. One of the effective means of developing FLCC for future professionals is YouTube video hosting, which provides the ability to select, download and watch videos on the selected topic, create educational products, and communicate with teachers and other students [6, 16].

The YouTube channel has many authentic podcasts for teaching students of different levels, created by native speakers in many training courses. However, most videos outline general topics. Therefore, in preparation for professional foreign language classes, we have selected videos from the YouTube channel and developed tasks for mastering and creatively reproducing the necessary topic vocabulary, training dialogical and monological speech and writing skills. Authentic videos help to understand different situations of a professional nature, reproduce, model the necessary verbal behaviour, and learn communication skills in a foreign language environment for students with different levels of training. In addition, visual support helps to understand the communicative goals of the individual, the means to achieve them, and models of nonverbal behaviour [21].

One of the post-watching tasks was to complete the project and present it in the form of a presentation about the hotel's service, its premises, and its facilities. Students used alternative PowerPoint free versions of Google Docs and Haiku Deck to make their presentations. In particular, Google Docs makes it easy to create presentations using text, images, video, and audio.

The students can share their presentations with others to collaborate on. The teacher and other students can also add their comments by monitoring and evaluating the work on the project. They can set tasks directly in the files. The participants of the team are able to paste graphics from Google Sheets and sent their comments directly from Gmail. Besides, students can search for the necessary content and pictures on the Google Disc and on the Web not leaving Google Slides. Performing professionally-oriented tasks in a foreign language with the help of presentations in Google Docs improves lexical and grammatical skills of both oral and written speech.

Mental maps (knowledge maps) are an effective method of data synthesis and analysis and the construction of logical connections. For educational purposes, they are used to present information, conduct research, present opinions, develop projects, plan essays, improve writing skills, etc. In addition, mental maps are a powerful graphic technology that helps discover mental activity's potential and visual representation of hierarchical information that contains the central idea and associative connections [22].

Knowledge maps use various platforms for presenting ideas, building diagrams, and charts: mindmeister, lucidchart, zapmeta, mindmap, bubbl.us, coggle, gitmind, canvas, and others. Each of these software has advantages and disadvantages. For example, the web application mindmeister helps capture, structure, distribute ideas, and add context to each topic with links, tabs, and media files.

Mental maps can serve as a visual ground for memorizing new vocabulary or grammar phenomena. They give the possibility to return to the information, fixed in the picture, with the aim to specify and edit it, add new information and make corrections. They also provide the opportunity to quickly generalize ideas and rules, stimulate imagination and memory, to plan further steps in completing the task.

However, it is impossible to save mental maps in free versions of the application. The alternative is to take screenshots. Hospitality students created mental

maps for projects, essays, and resumes. Figure 4 shows a mental map for writing resume.

An effective didactic tool for forming FLCC is the blog[1] "English for Hospitality Students" (Figure 5), created by the authors of the article for teaching students in rooms and studying individually. The blog has several sections: to develop reading, vocabulary, writing, and listening skills. There are also links to additional online resources and electronic dictionaries. Feedback between students and the teacher is held through questions. The developed tests help to assess students' progress.
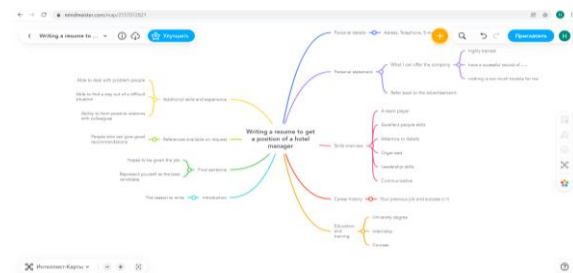
Figure 4: A mental map by a 3rd-year hospitality student.

Using the blog the students improve FLCC being surrounded not only by the English environment but also by an interesting and inspiring content, which they can use at any time available. There are authentic texts and developed tasks aimed at mastering speaking skills and vocabulary knowledge. Video podcasts are of great help for enriching the vocabulary and listening skills. The content of the blog is created to help students improve their knowledge not only of the English language but also their majors.
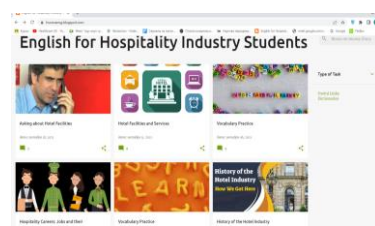
Figure 5: Home page of the blog[1].

The third stage of the research deals with analyzing the formation of FLCC components of hospitality students by ICT to compare the educational activities of future professionals in control (CG) and experimental (EG) groups. To define the development of the components of foreign language communicative competence of hospitality students by means of information communication technologies we processed and analyzed the answers

---

[1] https://horecaeng.blogspot.com

of the students to the following questionnaire (at the beginning and the end of the experiment):

A) *Section 1. Motivational and Value Component:*

1) Are you motivated to learn a foreign language for your career in hospitality? Rank from 1 to 5 (1- not motivated at all, 5- highly inspired).

2) How important do you think foreign language proficiency is in the hospitality industry? Rank from 1 to 5 (1-not important at all, 5-extremely important)

3) How do you perceive the role of foreign language proficiency in achieving your professional aspirations in the hospitality sector?

4) On a scale of 1 to 5, how important do you believe it is to be proficient in a foreign language for success in the hospitality industry?

5) Have you set specific goals regarding your foreign language learning? If yes, please describe them.

B) *Section 2. Cognitive-Activity Component:*

1) How often do you use online resources (websites, apps, etc.) to practice vocabulary related to hospitality in a foreign language?

2) Can you give an example of how you creatively use linguistic knowledge in professional discourse in a foreign language?

3) How confident are you in your pronunciation of foreign language words related to hospitality? (Rate on a scale of 1 to 5)

4) How comfortable are you with using information communication technologies to enhance your foreign language learning experience? (Rate on a scale of 1 to 5)

5) What specific ICT tools or platforms do you find most useful in improving your foreign language skills for hospitality?

C) *Section 3. Reflexive-Evaluative Component:*

1) How do you assess your current level of foreign language communicative competence?

2) Describe a recent instance where you evaluated your own foreign language communicative performance.

3) What strategies do you employ to enhance your foreign language communicative skills after self-assessment?

4) How do you plan your foreign language communicative activities for further improvement?

5) Are there any challenges you face in using ICT for foreign language learning? If yes, please describe.

We used the method of O. Smirnov to calculate the results of how many students of the two groups are at each level of the formation of FLCC components. Accordingly, at the beginning of the experiment in CG and EG, students were at approximately the same level of foreign language proficiency according to a diagnostic test and an oral interview. There were 42 students in EG and 41 in KG. In KG, the teachers used the traditional method, and in EG, they used the ICT as mentioned above tools. We have used a test, creative tasks and a questionnaire to evaluate the levels of FLCC formation of hospitality students at the beginning and the end of the experiment (Table 1).

Table 1: The results of the experiment in EG and CG.

Levels of ICC formation		ICT Components					
		Motivational and value		Cognitive-activity		Reflexive-evaluative	
		EG	CG	EG	CG	EG	CG
		%	%	%	%	%	%
Low	Before the experiment	28	30	24	26	36	36
	After the experiment	4	18	10	16	14	30
Average	Before the experiment	44	46	48	46	42	44
	After the experiment	30	48	22	48	34	40
Sufficient	Before the experiment	22	18	20	22	18	14
	After the experiment	38	22	50	28	38	20
High	Before the experiment	6	8	8	6	4	6
	After the experiment	28	12	18	8	14	10

As a result, ICT increased the sufficient FLCC motivational and value component level from 22% to 38% and from 6% to 28%. Accordingly, in the CG, this indicator increased from 18% to 22% (sufficient level) and from 8% to 12% (high level). EG's low motivational and value component level decreased from 28% to 4%. Indicators of the formation level of the cognitive-activity component of FLCC in EG also had significant positive dynamics. Sufficient level rose from 20% to 50%, high – from 8% to 18%. The low level of the reflexive-evaluative component of FLCC in EG decreased from 36% to 14%, and the sufficient one increased from 18% to 38% (Figure 6).
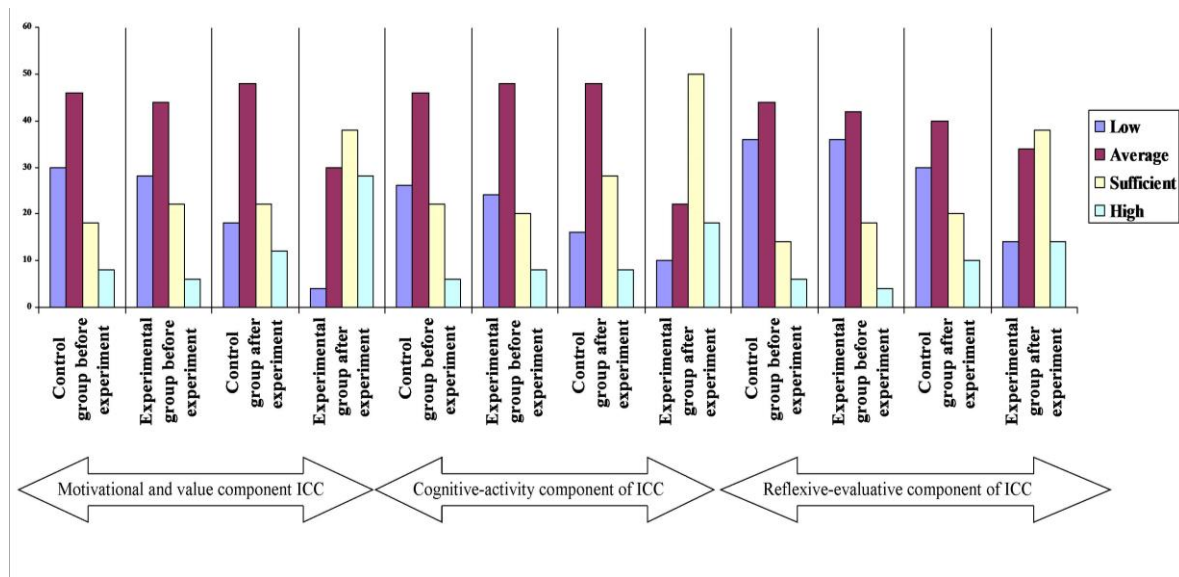
Figure 6: The comparative analysis of FLCC in EG and CG at the beginning and the end of the experiment.

The results prove the effective use of ICT for the development of FLCC hospitality students. Therefore, the formation of all FLCC components is at a sufficient and high level.

5 CONCLUSIONS

Thus, the review of the scientific work makes it possible to clarify the concept of FLCC hospitality students. It is a complex unity that contributes to obtaining and analyzing professionally relevant information in a foreign language and using it to solve professional problems. The paper outlines the peculiarities of ICT use in foreign language teaching to future hospitality specialists. The analysis required conducting questionnaires, observations and surveys. Thus it proves the didactic potential of ICT. However, teachers use ICT insufficiently. Consequently, they do not contribute to the full extent to the effective formation of students' FLCC.

The study proves that integrating ICT tools in learning a foreign language for professional purposes increases the motivational and value aspect in the formation of FLCC. After the implementation of ICT technologies (mainly professionally oriented Youtube podcasts, blogs, the creation of mind maps) into the language learning process more students started to set specific goals in their language learning: expand general vocabulary to facilitate effective communication with international guests and colleagues; gain cultural insights into the countries or regions where the target language is

spoken; to access industry-related resources, such as online articles, podcasts, webinars, and forums to stay updated on hospitality trends and best practices; learn and regularly review industry-specific vocabulary related to different departments within the hospitality sector. At the beginning of the experiment the students couldn't properly define the goals for language learning. Mainly the respondents didn't set the goals at all or they just wanted to improve the level of speaking skills.

The importance of a foreign language learning proficiency increased as well, as more students started to rank it 4 and 5 on the scale, instead of 2, 3 at the start of the learning process. They also admitted that they started to be more persistent in the language learning and get more interested in the process itself after the implementation of the ICT tools. They contributed greatly to their desire to learn a foreign language.

As for the cognitive-activity component of the FLCC the number of students using ICT for mastering the language increased. The rates on the scale in the answers to the question about using information communication technologies to enhance a foreign language learning experience were mainly 4 instead of 2, 3 at the beginning of the experiment.

In addition, the use of ICT in a foreign language learning of hospitality students promotes the formation and development of language skills and stimulates thinking, creativity, self-realization, and self-improvement. Furthermore, using YouTube video hosting resources, mental maps, educational blogs, etc. helps students realize their educational

potential and provides access to up-to-date professional information, communication, and cooperation of participants in the learning process. In particular, working on authentic videos contributes to mastering lexical and grammatical material and developing communication and listening skills.

Watching videos contributes to a better understanding of professional sphere realities, the behavior of communicators in different situations, and the mastery of models for solving professional problems. With the help of mental maps, it is possible to realize the creative potential, train critical thinking skills, and update and systematize vocabulary. Educational blogs are an effective means of forming FLCC, contributing to developing communication skills, self-esteem, and self-reflection.

After watching authentic podcasts and creating mental maps more hospitality students started to consider their achievements in the language learning, especially in speaking and pronunciation. Though the respondents still experience difficulties in planning their communicative activities for further improvement in a language learning.

We see the prospect of further research in the study of pedagogical conditions of ICT usage in teaching foreign languages to future specialists, including postgraduate professional training, incorporating the applicable criteria that follow.

REFERENCES

[1] Ministry of Education and Science of Ukraine, "Strategies for the development of higher education in Ukraine for 2021-2031", [Online], Available: https://mon.gov.ua/storage/app/media/rizne/2020/09/25/rozvitku-vishchoi-osviti-v-ukraini-02-10-2020.pdf.

[2] K. Ratheeswari, "Information Communication Technology in Education", Journal of Applied and Advanced Research, vol. 3, no. 1, pp. 45-47, 2018, [Online], Available: https://doi.org/10.21839/jaar.2018.v3iS1.169.

[3] O. Nestorenko, M. Wierzbik-Stronska, Innovation and Information Technologies in the Social and Economic Development of Society: Monograph, Katowice, Poland: Wydawnictwo Wyższej Szkoły Technicznej w Katowicach, pp. 180-182, 2018, [Online], Available: http://www.wydawnictwo.wst.pl/uploads/files/b25fdb8a9297745c3598096184f155d7.pdf.

[4] I. Kotiash, I. Shevchuk, M. Borysonok, I. Matviienko, M. Popov, V. Terekhov, O. Kuchai, "Possibilities of Using Multimedia Technologies in Education", International Journal of Computer Science and Network Security, vol. 22, no. 6,

pp. 727-732, 2022, [Online], Available: https://doi.org/10.22937/IJCSNS.2022.22.6.91.

[5] G.R. Korytska, "School language education in the development of electronic learning environment", in E learning in the theory and practice of teaching social sciences and humanities: a collective monograph, for general. ed. G.R. Korytska, T.M. Putiy, Ivano-Frankivsk, Ukraine: Symphoniya forte, 2017, pp. 40-54.

[6] I.V. Skril, "Formation of professionally oriented English lexical competence in speaking in future professionals in the hotel and restaurant business with the use of authentic videos", PhD thesis, Ternopil National University ped. Univ. V. Hnatyuk, Ternopil, p. 52-53, p. 692019.

[7] The Verkhovna Rada of Ukraine, Law of Ukraine "On Higher Education", [Online], Available: http://zakon0.rada.gov.ua/laws/show/ukr/1556 - 181.

[8] V.E. Krasnopolsky, "Foreign language training of students of non-philological specialties based on the creation and use of web-technologies", PhD thesis, East Ukrainian National Univ. V. Dalia, Kyiv, 2019.

[9] R. Mayer and L. Fiorella, "The Cambridge handbook of multimedia learning (3rd ed): a Textbook". GB: Cambridge University Press, 2021, 628 p.

[10] B. Tillinghast, "Developing an open educational resource and exploring OER-Enabled Pedagogy in higher education", IAFOR Journal of Education: Technology in Education. vol. 8, no. 9, 2020, [Online], Available: https://doi.org/10.22492/ije.8.2.09.

[11] E. Ofori and B. Lockee, "Next generation mobile learning: Leveraging message design considerations for learning and accessibility", IAFOR Journal of Education: Technology in Education, vol. 9, no. 4, 2021, [Online], Available: https://doi.org/10.22492/ije.9.4.07.

[12] G.G. Shvachych, V.V. Tolstoy, L.M. Petrechuk, Yu. S. Ivashchenko, O.A. Gulyaeva, and O.V. Sobolenko, "Modern information and communication technologies: a Textbook", Dnipro, Ukraine: NMetAU, p. 50, 2017.

[13] S.M. Haletsky, "Formation of communicative competence of future teachers of foreign languages by means of information and communication technologies", PhD thesis, Zhytomyr State Ivan Franko University, Zhytomyr, 2020.

[14] L. Morska, J. Skibska, V. Sulym, and V. Masztalir, "Didactic Potential of the Integrated Approach to Teaching Future Programmers Professional Communicative Competence in a Foreign Language", Information technologies and learning tools, vol. 64, no. 2, pp. 1-12, 2018, [Online], Available: https://doi.org/10.33407/itlt.v64i2.2174.

[15] D. Demchenko, O. Shynkarov, and L. Zaichenko, "Experimental work of forming professional foreign language competence of law students", 9 (SPE1), e941, 2021, [Online], Available: http://dx.doi.org/10.20511/pyr2021.v9nSPE1.941.

[16] S. Yekimov, T. Nepokupna, A. Shukanova, A. Ryabev, and M. Tonkoshkur, "Using social media to increase the communicative competence of future specialists in the field of ecological tourism", E3S Web of Conferences. 273(5): 09013, 2021, [Online], Available: https://doi.org/10.1051/e3sconf/202127309013.

[17] Ye. Provorova, T. Harasym, A. Zubryk, O. Ladyka, and V. Shepitchak, "Formation of foreign language professionally oriented competence of future foreign language teachers with the use of electronic educational resources", Journal for Educators, Teachers and Trainers, vol. 11 (1), pp. 1-12, 2020, doi: 10.47750/jett.2020.11.01.001.

[18] S.V. Symonenko, "Formation of communicative competence of specialists in software engineering in higher education institutions", PhD thesis, Melitopol State. ped. Univ. B. Khmelnytsky, Zaporizhia, 2019.

[19] N. Iukhno, "The essence and structure of the formation of foreign language communicative competence of specialists", The Scientific Notes of the Pedagogical Department, vol. 51, pp.113-119, 2022, doi: 10.26565/2074-8167-2022-51-13.

[20] A.O. Turgeneva, A.S. Popova, I.H. Sizonenko, K.V. Petrovska, and O.I. Hurenko, "YouTube video hosting as a means of information and cultural socialization of primary school children", Information technologies and learning tools, vol. 81, no. 1, p. 60-81, 2021.

[21] M.M. Greb and N.V. Grona, "Information and communication technologies as a means of forming linguodidactic competence of future primary school teachers", Information technologies and learning tools, vol. 82, no. 2, p. 109-125, 2021, [Online], Available: https://doi.org/10.33407/itlt.v82i2.3288.

[22] Z. Baksh, "A study of mind mapping in education and teaching through the technology", International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, no. 11, 2016, [Online], Available: https://doi.org/10.15680/ijircce.2016.0411041.

Integration of Digital Technologies in Modeling the Educational Environment of a Bachelor in the Conditions of Martial Law

Olena Kosovets, Mariana Kovtoniuk, Olena Soia and Denis Koval

Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Ostrozki Str. 32, 21100 Vinnytsia, Ukraine
kosovets.op@vspu.edu.ua, kovtonyukmm@vspu.edu.ua, soia.om@vspu.edu.ua, d.koval197@ukr.net

Keywords: Virtual Math Education Space, WordPress Content Management System, WordPress Plugins.

Abstract: The article reveals the features of the integration of digital technologies in the modeling of the educational environment of a higher education student, which can significantly improve the quality of education and preparation of students for real life, help in providing a qualified workforce and increasing the innovative potential of the state. The purpose of the article is to model and develop a virtual educational environment that will help to effectively form general and professional competencies of students, as well as develop cooperation and communication skills, in particular, in the conditions of the state of war. We used research methods: theoretical; empirical and mathematical-statistical; methods of developing a virtual mathematical educational environment. The potential of modeling a virtual mathematical educational environment on the example of a website has been studied, and it has been determined which tools and technologies can be used to create it. The possibility of obtaining free access to educational materials and a testing platform for consolidating acquired knowledge has been developed, which is quite relevant in the conditions of the state of war. The content management system (CMS) WordPress was chosen as the platform for creating the website. The article analyzes the results of a survey of applicants regarding the effectiveness of using a virtual educational environment in the process of studying mathematical disciplines and the degree of their familiarity with the use of digital technologies in modeling such an environment. The practical significance of the research results lies in the modeling and development of a virtual mathematical educational environment for a student of higher education, which, according to the authors of the article, will help students better understand complex mathematical concepts and principles, as well as develop cooperation and communication skills.

1 INTRODUCTION

The need for innovations and their effective transfer to the economy of Ukraine is extremely urgent today. This is primarily related to the need to transform the raw material-oriented low-tech economy of Ukraine into an economy focused on innovation and high technologies.

In these conditions, innovations can be an effective means of restoring and modernizing production and transitioning the Ukrainian economy to a higher technological level. First of all, this can be ensured by the transformation of Industry 4.0 technologies (Internet of Things, artificial intelligence, big data analytics, robotics, cloud computing, virtual and augmented reality, etc.), which fundamentally change the technical processes of production of goods and services, marketing and the institutional structure of society [1].

To evaluate the innovative capacity of Ukraine, we will use international comprehensive ratings and evaluations.

Table 1 shows the ranking of Ukraine according to the Global Innovation Index (GII) for the last three years (2020-2022). The Global Innovation Index ranks world economies according to their innovation potential. It consists of approximately 80 indicators grouped into innovative contributions and results. The GII aims to capture the multidimensional aspects of innovation. The table below shows the ratings of Ukraine for the last three years, noting that data availability and changes in the structure of the GII model affect the annual comparison of GII ratings. The statistical confidence interval for the ranking of Ukraine in GII 2022 is between 48 and 59 places.

According to the data provided in Table 1, in 2022, Ukraine shows better indicators of innovative results than innovative contributions. In 2022,

Ukraine ranked 75th in terms of innovative power, which is higher than in 2021, but lower than in 2020. Ukraine ranks 48th in the volume of innovative results. This position is lower for both 2021 and 2020. Ukraine ranks 4th among 36 countries with incomes below the average and ranks 34th among 39 European economies [2].

Table 1: Ranking of Ukraine according to the Global Innovation Index (GII) for 2020-2022 [2].

Reporting period	Ranking according to the GII	Subindex "Innovation Input"	Subindex "Innovation Output"
2020	45	71	37
2021	49	76	37
2022	57	75	48

In 2022, Ukraine ranked 75th in terms of innovative power, which is higher than in 2021, but lower than in 2020. Ukraine ranks 48th in the volume of innovative results. This position is lower for both 2021 and 2020. Ukraine ranks 4th among 36 countries with incomes below the average and ranks 34th among 39 European economies [2].

According to the GII indicator, innovation capacity is based on knowledge and technology (human capital and research, as well as the results of knowledge and technology), while indicators of infrastructure development, institutions, and market complexity have a negative impact on innovation capacity (Table 2).

Table 2: Rating of Ukraine by components of the Global Innovation Index (GII) for 2020-2022 [2].

Reporting period	GII	Institution	Human capital and research	Infrastructure	Market complexity	Business complexity	Results of knowledge and technologies	Creative results
2020	45	93	39	94	99	54	25	44
2021	49	91	44	94	88	53	33	48
2022	57	97	49	82	102	48	36	63

Note that according to the rating, one of the best developed drivers in Ukraine is human capital and research, as well as the results of knowledge and technology.

Considering the consequences of the aggressive war of the Russian Federation against Ukraine and the general decline of the world economy as a whole, which will negatively affect the innovative potential of Ukraine, it is necessary to concentrate efforts on preserving and strengthening the strengths - the development of secondary and higher education and digital technologies.

The purpose of the article is to model and develop a virtual educational environment that will help to effectively form general and professional competencies of students, as well as develop cooperation and communication skills, in particular, in the conditions of the state of war.

2 ANALYSIS OF CURRENT RESEARCH

The evolution of educational technologies led the scientific community of mathematics education researchers to the conclusion that the use of digital technologies performs two main functions: (a) as a support for the organization of the teacher's work and (b) as support for new ways of working and presenting mathematics [3]. In today's environment, the use of technology for mathematics teachers has begun to have a third function: as a reinforcement of communication, interaction in communities, communication and sharing of materials. In [4] added a fourth function, commercial-industrial, which consists in supporting more independent work of students and is focused on practicing and evaluating previously taught mathematical knowledge and skills in various online formats.

Digital and non-digital tools in teaching mathematics were investigated by Monaghan et al. [5]. Digital technologies for assessment are discussed in detail in the journal "Mathematics Education in the Digital Era", in papers [6-8].

The authors [9] investigated the innovativeness of vocational education and described the architecture of digital technologies in the educational environment of the teacher [10] as the transfer of innovations to the economic space of the state.

Lipeikiene in his scientific work [11] defines the main requirements for a virtual educational environment: access control; creation of professional training material of the course and mastering the skills of managing it; ensure the use of all types of multimedia content; links to digital libraries and other important information on the Internet; automatic glossaries, indexes, search systems; communication tools; connection to the software necessary for training (if available); a personal space for students to share and store materials; tools for automatic assessment and self-assessment; monitoring of students' success, systematization of their grades,

statistical information about the educational process for the teacher.

The authors [12] investigated the issue of foresight modeling of the synergistic educational space of obtaining higher education and described the architecture of digital technologies in the educational environment of the teacher [13] as the transfer of innovations to the economic space of the state.

3 DEVELOPMENT OF VIRTUAL MATHEMATICS EDUCATIONAL SPACE IN LMS WORDPRESS

The use of modern digital technologies in education has led to the emergence of a virtual educational environment, within which continuous self-realization, self-development of an active and creative personality is possible under the conditions of the organization and functioning of educational systems with realized openness, integrability and adaptability through educational websites.

In the conditions of a pandemic and wartime, the use of the latest technologies and the global Internet network in the educational process is the main need of today. Web technologies are developing rapidly, and developed websites contain more and more modern features, becoming convenient for users both in professional use and in everyday life. The ability to design and develop dynamic, functional and convenient educational websites is an important component of the teacher's professionalism. After all, the success of his students depends on how he will be able to implement an educational web project in the digital educational Internet space.

Some teachers are inclined to think that only a specialist programmer can create a website. But this is not the case, as there are a large number of simple digital tools with ready-made templates for managing content and presenting educational material in the form of an author's website, with the help of which a teacher can independently, without outside help, create his own website.

In the article, we consider the development of a virtual mathematical educational space using the popular open source content management system WordPress [14]. This content management system allows you to develop a powerful educational website, where registered users get access to various mathematics courses and a platform for testing the acquired knowledge.

Installing the MasterStudy LMS plugin [15] for WordPress enables the teacher to create and customize a personalized learning process by building a separate educational trajectory for the individual user profiles that registered students have access to. It offers an easy-to-use interface that simplifies the management of the site's resources with educational materials. Administrators can create an unlimited number of courses and lessons, organize them by category and set prerequisites for course enrolment.

MasterStudy LMS offers a built-in grading system that allows for easy grading and feedback.

The virtual mathematical educational space, created by the WordPress content management system with the MasterStudy LMS plugin, involves the registration of users who can access various mathematical courses and a testing platform to consolidate knowledge. Website developers add educational videos and lecture texts, create tests, combine them into different courses, receive feedback and course evaluations from registered users.

The main page of the educational site contains brief information about the purpose of the website, the "Join for free" button to view the list of courses and topics, user login and registration, information about popular courses, information about the teacher and the courses published by him.

Useful for site developers is a page editor with a visual interface, Elementor, which allows you to create and edit WordPress pages without using programming code [16]. Among its main advantages is a set of ready-made templates that can be used to create different types of pages; a large number of widgets; adding animation effects to elements; creating and configuring forms for collecting data or applications; providing the ability to view the page in real time while editing it, which allows you to visually evaluate the results, etc.

The optional WP Mail SMTP plugin [17] improves email inflow by reconfiguring WordPress to require a proper SMTP provider for sending emails. WP Mail SMTP Pro plugin features email logging, advanced email reports and statistics, connection backup, email alerts, smart conditional routing, and more.

To add video materials to the courses, we created a YouTube channel on which we created several playlists and uploaded the corresponding video lectures (Figure 1). This YouTube channel will act as a server for the site's video materials, thus removing the extra load from the site's server by transferring the

processing and playback of video materials to the YouTube service.
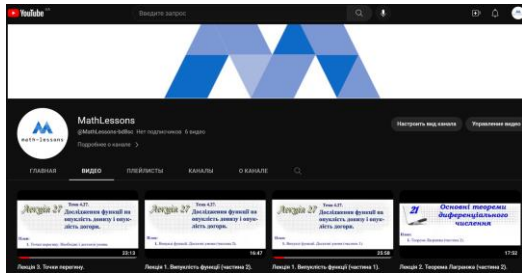
Figure 1: Created YouTube channel for hosting video materials.

In the page settings section on the mathematical educational website, we specify the type of lesson "Video" and add a link to the video material posted on the YouTube channel (Figure 2).
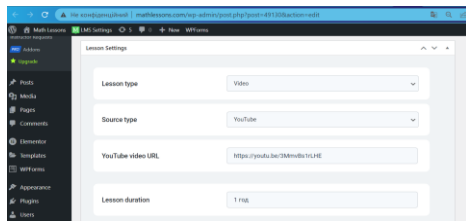
Figure 2: Adding video material from a YouTube channel to a lesson.

You can create, edit and delete courses as an administrator from the site control panel (Figure 3).
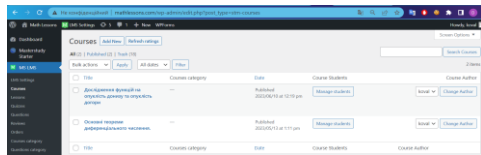
Figure 3: Courses control panel.

An unregistered student can only familiarize himself with the list of educational materials placed in the chosen course. To gain access to educational materials, the student must register on the site using the "Login / Registration" command on the main page. Next, he fills in the fields name, e-mail address of the mailbox, password and confirms the registration. At the specified address, a letter with account activation will automatically arrive in the mailbox by going to the site with a unique vocation.

After activating the account, the management page will open, where the student can change the name and surname, the password of the account and get information about the courses and tests for which he is registered. The site administrator will receive

information about the new registered student in the website control panel menu and has the ability to manage this student (edit or remove name, photo, password, etc.) (Figure 4).
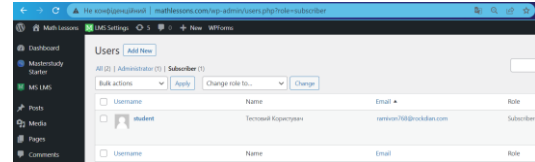
Figure 4: New user information in the website control panel.

After the student has registered on the site, he starts studying, for example, on the course "Basic Theorems of Differential Calculus" (Figure 5).
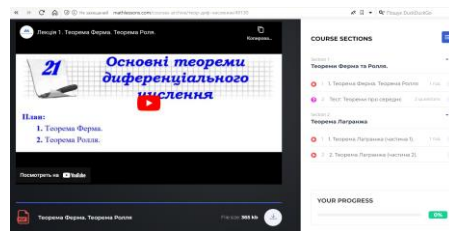
Figure 5: Course materials "Basic Theorems of Differential Calculus".

The course progress is tracked in the window "Your progress".

The site administrator will receive a question in the "Comments" menu of the site control panel (Figure 6) and will send an answer to the user (Figure 7).
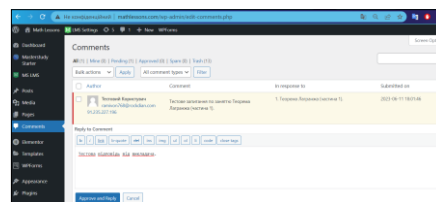
Figure 6: User question in the site control panel.
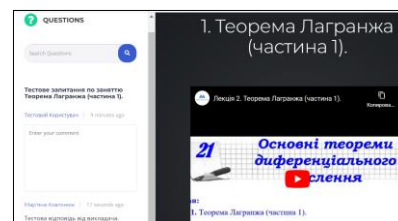
Figure 7: Provided response from the teacher to the user.

After completing the course, the student has the opportunity to leave his evaluation and comments on

this course. The administrator will receive an assessment for the developed course (Figure 8).
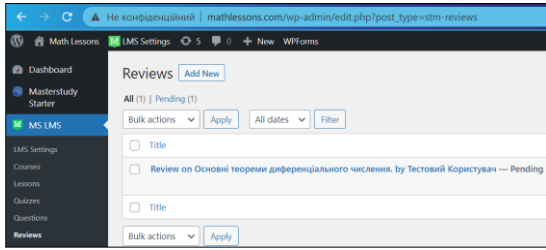
Figure 8: Received rating from the user for the course.

4 DISCUSSION

To test the relevance of developing a virtual mathematical educational space for bachelor's degree and master's degree students of the Faculty of Mathematics, Physics and Computer Sciences of the Mykhailo Kotsiubynskyi Vinnytsia State Pedagogical University, we will use the correlation coefficients of Chuprov K and Pearson C to assess the relationship in a situation where a quality feature consists of more than two groups [18]. The Chuprov coefficient K (1) is used in case of an unequal number of rows and columns of the conjugation table ($k_1 \neq k_2$):

$$K = \sqrt{\frac{\varphi^2}{\sqrt{(k_1-1)(k_2-1)}}}, \qquad (1)$$

where k_1 i k_2 are the number of groups of the first and second characteristics (parameters X i Y).

Pearson correlation coefficient C (2) is used when the number of rows and the number of columns in the conjugation table are the same ($k_1 = k_2$):

$$C = \sqrt{\frac{\varphi^2}{1+\varphi^2}}, \quad \varphi^2 = \sum_{y=1}^{k_1}\left(\frac{\sum_{x=1}^{k_2}\left(\frac{n_{xy}^2}{n_x}\right)}{n_y}\right) - 1. \quad (2)$$

The values of the Chuprov K and Pearson C coefficients vary from 0 to 1.

Enter the emp irical data in Table 3-4 and perform the calculations.

Let's write the expression φ^2 in more detail (3), based on the conditions $k_1 = 3$ i $k_2 = 5$:

$$\varphi^2 = \sum_{y=1}^{k_1}\left(\frac{\sum_{x=1}^{k_2}\left(\frac{n_{xy}^2}{n_x}\right)}{n_y}\right) - 1 == \frac{\sum_{x=1}^{k_2}\left(\frac{n_{xy}^2}{n_x}\right)}{n_1} +$$

$$+ \frac{\sum_{x=1}^{k_2}\left(\frac{n_{xy}^2}{n_x}\right)}{n_2} + \frac{\sum_{x=1}^{k_2}\left(\frac{n_{xy}^2}{n_x}\right)}{n_3} - 1. \qquad (3)$$

Let's find the numerical value of the Chuprov K correlation coefficient.

Let's calculate individual components:

$$A_1 = \frac{\sum_{x=1}^{k_2}\left(\frac{n_{xy}^2}{n_x}\right)}{n_1} = \frac{\frac{11^2}{21} + \frac{8^2}{35} + \frac{8^2}{21} + \frac{2^2}{3} + \frac{1^2}{2}}{30} \approx 0{,}42;$$

$$A_2 = \frac{\sum_{x=1}^{k_2}\left(\frac{n_{xy}^2}{n_x}\right)}{n_1} = \frac{\frac{6^2}{21} + \frac{17^2}{35} + \frac{10^2}{21} + \frac{0^2}{3} + \frac{1^2}{2}}{34} \approx 0{,}45;$$

$$A_3 = \frac{\sum_{x=1}^{k_2}\left(\frac{n_{xy}^2}{n_x}\right)}{n_1} = \frac{\frac{4^2}{21} + \frac{10^2}{35} + \frac{3^2}{21} + \frac{1^2}{3} + \frac{0^2}{2}}{18} \approx 0{,}24.$$

Let's define a parameter φ^2:

$$\varphi^2 = 0{,}42 + 0{,}45 + 0{,}24 - 1 \approx 0{,}11.$$

Table 3. Distribution of students according to the results of pedagogical research[1].

A contingent of students (parameter Y)	1. In your opinion, can the virtual mathematical educational space affect your motivation to study mathematics? (parameter X)					Total
	Yes[1]	Yes[2]	Neutral[3]	No[4]	No[5]	
1-2 courses, bachelor's degree	11	8	8	2	1	30
3-4 courses, bachelor's degree	6	17	10	0	1	34
1-2 courses, master's degree	4	10	3	1	0	18
Total	21	35	21	3	2	82

---

[1] Yes, it has a positive effect on the motivation to study mathematics.

[2] Yes, it can boost motivation as it allows you to use interactive tools and track your progress.

[3] Neutral, the motivation to study mathematics does not depend on the form of the educational space.

[4] No, it doesn't affect motivation because the math remains challenging regardless of the format.

[5] No, it can even reduce motivation due to the lack of personal contact with the teacher and the community in general.

We will obtain the numerical value of the Chuprov K correlation coefficient:

$$K = \sqrt{\frac{\varphi^2}{\sqrt{(k_1 - 1)(k_2 - 1)}}} = \sqrt{\frac{0{,}11}{\sqrt{(3 - 1)(5 - 1)}}} \approx 0{,}19.$$

The value of Chuprov's coefficient $K \approx 0{,}19$ indicates an insignificant mutual relationship between the parameters X, Y. From a practical point of view, this is a confirmation of the hypothesis that the influence of the virtual mathematical educational space on the motivation to study mathematics actually depends little on the distribution of students by courses of study. Junior year students began blended learning at school during the COVID-19 pandemic. And at the university, all students are forced to adapt to wartime conditions. Therefore, the use of a virtual mathematical educational tool is relevant for all students. 25.6% of respondents believe that the virtual educational space has a positive effect on the motivation to study mathematics. 42.7% of students are inclined to think that the virtual educational space can stimulate motivation, as it provides an opportunity to use interactive tools and track their progress. 25.6% of respondents have a neutral attitude, their motivation to study mathematics does not depend on the form of educational space. 3.7% of students think that the virtual educational space does not affect motivation, because mathematics remains difficult regardless of the format. 2.4% of respondents answered that the virtual educational space can even reduce motivation due to the lack of personal contact with the teacher and the community in general.

We will make similar calculations with the survey results presented in the Table 4.

Let's calculate individual components:
- $A_1 \approx 0{,}40$; $A_2 \approx 0{,}46$; $A_3 \approx 0{,}23$.
- Parameter $\varphi^2 \approx 0{,}09$.
- Numerical value of Chuprov's correlation coefficient $K \approx 0{,}18$.

Table 4. Distribution of students according to the results of pedagogical research[2].

A contingent of students (parameter Y)	2. Do you think that the virtual mathematical educational space makes mathematics more accessible to students? (parameter X)					Total
	Yes[1]	Yes[2]	Neutral[3]	No[4]	No[5]	
1-2 courses, bachelor's degree	17	8	4	0	1	30
3-4 courses, bachelor's degree	9	14	10	0	1	34
1-2 courses, master's degree	8	7	3	0	0	18
Total	34	29	17	0	2	82

The value of the Chuprov coefficient $K \approx 0{,}18$ indicates a slight mutual relationship between the parameters X, Y. In fact, this is a confirmation of the hypothesis that the influence of the virtual mathematical educational space on the accessibility of mathematics for students depends little on the distribution of students by courses of study. Therefore, the use of a virtual mathematical educational space is a sought-after resource for all students. 41.5% of respondents agree that thanks to convenient learning and interactive tools, mathematics has become more accessible to them. 30.5% of students tend to think "Yes", especially for those who have limited access to traditional resources. 25.6% of respondents have a neutral attitude, because for them the availability of a virtual mathematical educational space depends on the individual circumstances of the student. None of the students think that the virtual mathematics educational space does not make mathematics more accessible to all students. 2.4% of respondents answered that the virtual mathematical educational space may be less effective for understanding mathematical concepts.

---

[1] Yes, thanks to convenient learning and interactive tools.

[2] Yes, especially for those with limited access to traditional resources.

[3] Neutral, because the availability of the virtual mathematical educational space depends on the individual circumstances of the student.

[4] No, does not make math more accessible to all students.

[5] No, may be less effective for understanding mathematical concepts.

5 CONCLUSIONS

In martial law, when access to traditional education may be limited, digital technologies provide the opportunity to continue learning and develop students' mathematical skills remotely.

A virtual mathematics website based on WordPress provides students with an effective tool for learning mathematics in an interactive and adaptive environment that ensures the implementation of effective student-centered learning. Modeling the educational process with the help of digital technologies allows students to get access to relevant educational content, interact with teachers and classmates, as well as perform dynamic tasks and tests online.

As we can see, the changing trends in the development of society require from students the ability to adapt to new conditions of the organization of the educational process, and from teachers - the desire to change, to get used to living and working in conditions of uncertainty [19]. In the era of informatization (digitalization) of society and education, in particular, "a modern teacher must possess innovative practices for implementing adaptive, mixed, distance, cloud, and mobile learning" [20]. This prompts teachers to pedagogical design and development of digital technologies for the transformation of traditional approaches to learning and increasing work productivity, revealing one's own individuality.

REFERENCES

[1] T. Vasylieva and Y. Petrushenko, "Digital technologies in education: modern experience, problems and prospects," In: monograph. Sumy: Sumy State University, 2022, 150 p.

[2] Economy reports & analysis. Global Innovation Index. 2022, [Online], Available: https://www.globalinnovationindex.org/analysis-economy.

[3] N. Sinclair and O. Robutti, "Teaching practices in digital environments," In S. Lerman (Ed.), Encyclopedia of mathematics education, 2nd ed, Dordrecht: Springer, 2020. pp. 845-849.

[4] A. Clark-Wilson, O. Robutti, and M. Thomas, "Teaching with digital technology," ZDM Mathematics Education, vol. 52, 2020, pp. 1223-1242, [Online], Available: https://doi.org/10.1007/s11858-020-01196-0.

[5] J. Monaghan, L. Trouche, and J. Borwein, "Tools and mathematics, Dordrecht: Springer, 2016, 483 p.

[6] G. Aldon and J. Trgalova (Eds.), "Technology in Mathematics Teaching," Selected Papers of the 13th ICTMT Conference, vol. 13, 2017, [Online], Available: https://hal.science/hal-01632970/file/book-ictmt13-en.pdf.

[7] P. Black and D. Wiliam, "Developing the theory of formative assessment," Educational assessment, Evaluation and Accountability, vol. 21, no. 1, 2009, pp. 5-31.

[8] G. Aldon and M. Panero, "Can digital technology change the way mathematics skills are assessed?" ZDM Mathematics Education, vol. 52, 2020, pp. 1333-1348, doi: 10.1007/s11858-020-01172-8.

[9] L. Antoniuk, O. Soia, O. Kosovets, A. Klimishyna, M. Kovtoniuk, and O. Tryfonova, "Ensuring the Innovativeness of Vocational Education: Neuropedagogical Aspect". BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 2022, no. 13(4), pp. 242-263, [Online], Available: https://doi.org/10.18662/brain/13.4/386.

[10] M. Kovtoniuk, O. Kosovets, O. Soia, and I. Leonova, "The architecture of digital technologies in the educational environment of the teacher as a transfer of innovations to the economic space of the state," Modern information technologies and innovative teaching methods in the training of specialists: methodology, theory, experience, problems, vol. 68, 2023, pp.93-106, [Online], Available: https://doi.org/10.31652/2412-1142-2023-68-93-106.

[11] J. Lipeikiene, "Virtual Learning Environments as a Supplement to Traditional Teaching," Informatics in Education, vol. 2, no. 1, 2003, pp. 53-64.

[12] P. Mulesa and O. Semenikhina, "Pedagogical conditionsfor preparing future mathematicsand computer science teachersfor the usage of virtual clarity tools in professional activities," Phisical and Mathematical Education, vol. 38, no. 2, 2023, pp. 37-42, [Online], Available: https://doi.org/10.31110/2413-1571-2023-038-2-006.

[13] M. Mintii, N. Sharmanova, A. Mankuta, O. Palchevska, and S. Semerikov, "Selection of pedagogical conditions for training STEM teachers to use augmented reality technologies in their work," Journal of Physics: Conference Series, 2023, doi: 10.1088/1742-6596/2611/1/012022.

[14] WordPress, [Online], Available: https://wordpress.com.

[15] MasterStudy LMS WordPress Plugin – for Online Courses and Education, [Online], Available: https://wordpress.org/plugins/masterstudy-lms-learning-management-system/.

[16] Elementor Website Builder – More than Just a Page Builder, [Online], Available: https://en-ca.wordpress.org/plugins/elementor/.

[17] WP Mail SMTP by WPForms – The Most Popular SMTP and Email Log Plugin, [Online], Available: https://wordpress.org/plugins/wp-mail-smtp.

[18] V. Rudenko, "Mathematical statistics," Study guide, Kyiv: Centre for Educational Literature, 2012, 304 p.

[19] R. Hurevych, V. Kobysia, A. Kobysia, S. Kizim, L. Kutsak, and N. Opushko, "Use of digital services and tools in professional training of future teachers," Modern information technologies and innovative teaching methods in the training of specialists: methodology, theory, experience, problems, vol. 64. 2022, pp. 5-22, [Online], Available: https://doi.org/10.31652/2412-1142-2022-64-5-22.

Scratch PoseNet Exergame Prototyping for Learning Process Support in Physical Education

Oleksandr Blazhko, Viktoriia Podhorna, Anastasiia Kokotieieva and Oleksii Ivanov

Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine

{blazhko, podgorna.v.v, kokotieieva.a.s, o.v.ivanov}@op.edu.ua

Keywords: Exergame, PoseNet, Scratch.

Abstract: This paper suggests the Exergaming approach of computer game control for a human motion in sport computer games in order to increase the level of physical activity of students in online classes, which takes into account the initial knowledge of the specialised courses of the IT curriculums. The approach is based on open access repository of computer programs in the Scratch-based block language (47 games for 28 summer sports, 16 games for 10 winter sports). Scratch programs for human motion recognition are based on the PoseNet neural network, which allows changing the principle of controlling game characters through the keyboard and mouse. In October 2023, ninety two first-year students majoring in Computer Science participated in a five-week online Exergaming as second part of the Physical Education course, among them 70 students (76%) successfully completed all stages. 57% of students felt motivated to continue using the Exergame approach and demonstrated the desire to achieve decent results, that indicates skill "Positive Thinking" in the discipline "Physical Education", seven Hard Skills and four Soft Skills of Computing Curricula 2020 were indicated to compensate for the lack of classroom learning.

1 INTRODUCTION

The report "Computing Curricula 2020 (CC2020): Paradigms for Global Computing Education. Association for Computing Machinery" defines competences in the Knowledge-Skill-Disposition framework, where element "Skill" includes not only Computing Knowledge (Hard Skills), but also Foundational and Professional Knowledge (Soft Skills), for example: Analytical Thinking, Collaboration and Teamwork, Mathematics, Problem Solving and Trouble Shooting, Quality Assurance/Control, Self-Starter/Learner, Time Management and others [1]. The basis for quickly acquiring and maintaining these skills is physical activity with benefits for individual health and well-being, improved mental health, sleep and cognitive function. Therefore, an academic discipline "Physical Education" can be a useful tool to promote life/soft skills and to add skill "Positive Thinking" into previous Soft Skills [2]. Daily exercise and participation in sports competitions build competencies "Self-Starter/Learner" and "Collaboration and Teamwork" for students.

Unfortunately, in many universities due to COVID restrictions in 2020-2021 and in the Ukrainian universities due to war activities since 2022, a lot of students have not access to sports facilities for Physical Education classes. Therefore, the "Physical education and sports" department of Odesa Polytechnic National University decided to adapt the curriculum of this course for students of IT specialties only through the use of modern software technologies in cooperation with the "Information Systems" department. Today, there are tools for human-computer interaction through human motion recognition that allow you to control events in entertaining or educational computer games using Exergaming technologies ("exercise" + "game") [3].

For many years, non-contact motion control sensors are used to recognize human movements, for example, the MS Xbox 360/One game console, which contains an MS Kinect sensor with an infrared camera [4] to support 3D computer sports games Kinect Sport: Tennis, Table Tennis, Golf, Skiing, Baseball, Soccer, Basketball, Boxing, Track and Field (Sprint, Javelin, Discus, Long Jump and Hurdles). However, the use of Exergaming technologies in the curriculum for first-semester

students has limitations when conducting online classes:

1) lack of access to infrared cameras;
2) lack of access to commercial computer games;
3) lack of knowledge to create computer games.

The first limitation can be partially removed by using webcam image processing programs based on neural network, for example, PoseNet [5]. The second limitation can be partially removed by using the Scratch programming environment [6], a tool for teaching computational thinking skills in schools. Scratch allows students quickly create prototypes of 2D-computer programs and provides to modify computer games created by students around the world and posted in open file repositories [7]. Unfortunately, there are no examples of sports games with motion control in the Scratch repository. However, work [8] presents special programs for displaying human movement by using the MS Kinect v.1 cameras in Scratch v.1.4. These programs use different scenarios of games [9]: "Alien attack" controls the movements of body to the left/right and raising the arms above the head, "Hungry shark" controls the movements of the head to up/down and joining the hands together, "Hungry ant" controls the movements of the right hand. The work [10] presents additional games for the rehabilitation of stroke patients: "Whack-a-mole", "Harvest carrots", "Picking apples", "Bowling" and "Boxing". These games can be easily adapted for simple Webcams instead of MS Kinect cameras by using free software libraries for connecting neural networks with the Scratch environment, for example, PoseNet [11].

The analysis of the previous works determined the need to modify the Physical Education curriculum for students in IT specialties in learning process of first semester. These modifications should partially compensate for the lack of classes with students in the gym and take into account the list of discipline competencies of the current and future semesters.

This work presents the Scratch-Based Exergame Educational Approach (Exergame approach) for computer game prototyping on the topics of Olympic sports. This Exergame approach should take into account the simple technical requirements for students during online classes and low requirements for knowledge of programming languages:

- access to simple webcams and computers;
- knowledge of the basics of programming in the Scratch environment;
- access to the Scratch software library for connection to the PoseNet neural network;
- access to the open Scratch repository of Olympic sports computer games.

The purpose of this work is to establish a connection between Hard/Soft Skills of CC2020 and skills of Exergaming approach during online classes (staying students at home) of the discipline "Physical Education" for students in IT specialties.

2 EXERGAMING IN PHYSICAL EDUCATION

2.1 Education Concept Based on Exergaming Approach

The proposed Exergaming approach is based on (Figure 1):

1) the Open Scratch repository of 2D-computer game programs;
2) the open Scratch Web-server library for connection to the PoseNet neural network;
3) Scratch program code templates for human motion recognition based on Webcam and the PoseNet neural network.
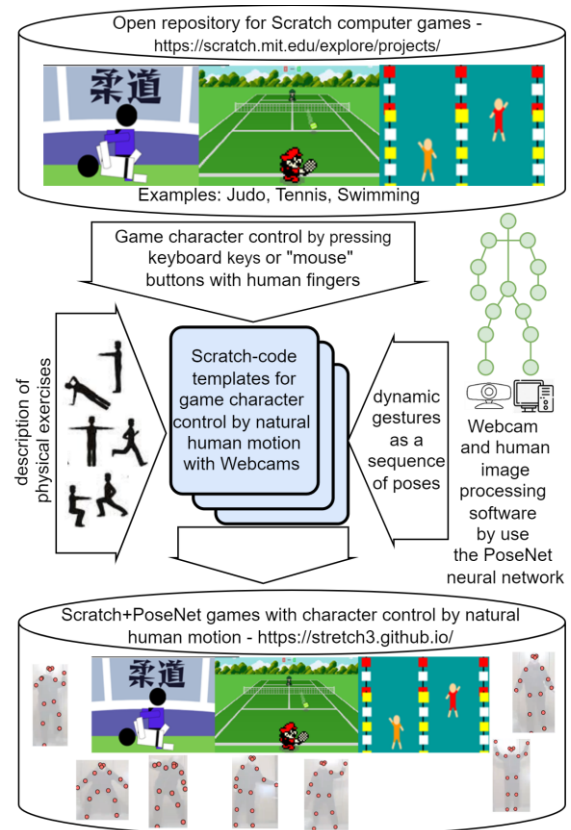

Figure 1: Student learning concept based on the Exergame approach.

The proposed approach gives recommendations for choosing a method for controlling game characters by using Webcams and descriptions of the characteristics of human dynamic gestures as a sequence of transitions between poses (static gestures).

The learning process includes the following stages:

1) students choose an example of a physical exercise and an example of a computer game in the Scratch open repository;
2) students analyze physical exercises and choose Scratch program code templates based on recommendations;
3) students use templates and make changes to the Scratch program code of computer games to control the game characters with gestures.

To determine the list of computer games that students will explore and modify during online classes in the Scratch-repository, we searched and selected 63 computer games: 47 games for 28 summer Olympic sports; 16 games for 10 winter Olympic sports. The Scratch version 3 uses an open source code and can be hosted on any Web server with additional software library PoseNet2Scratch to recognize human motion based on the PoseNet neural network [11]. The PoseNet allows present the human body in the form of 17 reference points (Figure 2).
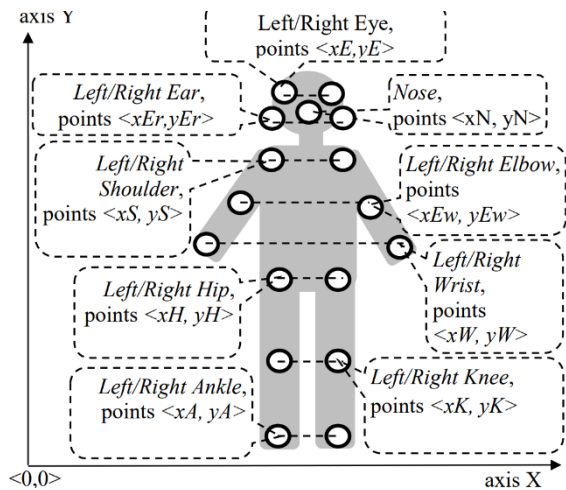
Figure 2: Reference points in PoseNet neural network.

Each description of a movement can be considered as a sequence of poses – positions taken by the human body, head and limbs in relation to each other. These poses can be considered as gestures: actions or movements of a human body. Exergaming approach uses the following terms:

- static gestures (sg) – unchanging position of the human body with fixed relationships of its various parts to each other;
- dynamic gesture (dg) – sequential transition in time between two static gestures.

For static gestures their coordinate description is proposed, which takes into account the following:

- for each pair of static gestures, reference points with significant difference between static gestures are searched, and a logical expression comparing the values of the coordinates of these reference points is described;
- these gestures must use a minimum number of comparisons between reference points, but guaranteeing a significant difference in their description between different static gestures.

Table 1: Examples of coordinate description of gestures.

A kind of sport	Screen-shot of 1th static gesture	Screen-shot of 2nd static gesture	Coordinate description of static gestures
Judo	sg1 – protection	sg2 – attack	sg1: $(xK < xH)$ and $(xW < xA)$ and $((ySL - yHL) = (xHR - xHL))$ sg2: $(yWL = yN)$ and $(yWR < yEwR)$ and $(xN = xH)$
Tennis	sg1 – racket ready to hit	sg2 – hit with a racket	sg1: $(yWL > yEwR)$ and $(xWL > xEwL)$ sg2: $(yWL > yEwR)$ and $(xEwL = xEwL)$
Swimming	sg1 – swing by right hand	sg2 – swing by left hand	sg1: $(yWR > yN)$ and $(yWL < yEwR)$ sg2: $(yWL > yN)$ and $(yWR < yEwL)$

Table 1 presents three computer games examples of Web camera screenshots in the form of reference points of a human body with static gestures in the

Scratch program with a connected PoseNet neural network and a coordinate description of gestures.

2.2 Description of Scratch PoseNet Program Library Templates

To control anchor points, the PoseNet2Scratch software library adds a section of green blocks to Scratch. Figure 3 shows fragments of the sprite placement rendering Scratch-blocks associated with the anchor points of a right wrist. In addition, to program the recognition conditions of the specified gestures, it is recommended to create variables, for example, *nose_y, left_wrist_y, right_wrist_y*, the values of which are set in the software modules for visualization of reference points.

Figure 3: Fragments of Scratch-blocks for visualization of anchor points for a right wrist.

It is proposed to use an algorithmic template (example in C-language style) to recognize dynamic gestures by using system states of static gestures:

```
old_static_gesture := 0;
new_static_gesture := 0;
while (true) {
if static_gesture_1_definition() {
 // change system state
 old_static_gesture=new_static_gesture;
 new_static_gesture = 1;
}
  …
if static_gesture_n_definition() {
 // change system state
 old_static_gesture=new_static_gesture;
 new_static_gesture = n;
    }
}
if dynamic_gesture_1_definition(
 new_static_gesture,old_static_gesture)
 dynamic_gesture = 1;
…
if dynamic_gesture_n_definition(
 new_static_gesture,old_static_gesture)
 dynamic_gesture = n;
}
```

The training task includes programming algorithms for recognizing two dynamic gestures.

Students translate this algorithmic template into blocks of the Scratch language. For example, to recognize two static gestures "Both hands below nose" and "Both hands above nose" program code can include the following steps:
1) checking the condition of recognition of the first static gesture, which is defined by the logical expression "*sg1: (yN > yWL) AND (yN > yWR)*";
2) if the static gesture returns true, then set variable *new_static_gesture = 1* (state *sg1*);
3) checking the condition of recognition of the second static gesture, which is defined by the logical expression "*sg2: (yWL > yN) AND (yWR > yN)*";
4) if the static gesture returns true, then set variable *new_static_gesture = 2* (state *sg2*).

Dynamic gesture recognition blocks take into account the transition between a previously recognized static gesture and a new recognized static gesture. For the above example, two dynamic gestures may be recognized:
- *dg1* – "hands raised above the nose" through previously programmed recognition processes of two static gestures by using order $sg_1 \rightarrow sg_2$;
- *dg2* – "hands lowered below the nose" through previously programmed recognition processes of two static gestures by using order $sg_2 \rightarrow sg_1$.

To establish the fact of recognition of dynamic gestures, the change *dynamic_gesture* is used, which can later be used in the program code of the analysis of the conditions of controlling the game character to replace the classic version of control in the game through a computer keyboard or a computer mouse.

The exergaming approach offers students only to make changes to an existing computer game. However, to understand the basics of game design, students consider classic model "Mechanics-Dynamics-Aesthetics" [12]. Particular attention is paid to the Dynamics-component, which takes into account the actions of the players, depending on their psychophysiological abilities, for example, the speed of the brain's reaction, the speed of the fingers movement when pressing the keys on a keyboard. Tuning the Dynamics-component should balance the game. The computer games are already balanced. But after replacing the key control with gesture control, the computer game becomes unbalanced. Therefore, students should find Scratch-blocks for balancing the game and conduct experiments to adjust the new characteristics of game dynamics.

During the analysis of the steps of using the Exergame approach, connections between Hard/Soft skills of CC2020 and skills of the Exergame approach were established. Besides, two important steps determined connections with Hard Skills in the Physical Education Curricula (Figure 4).
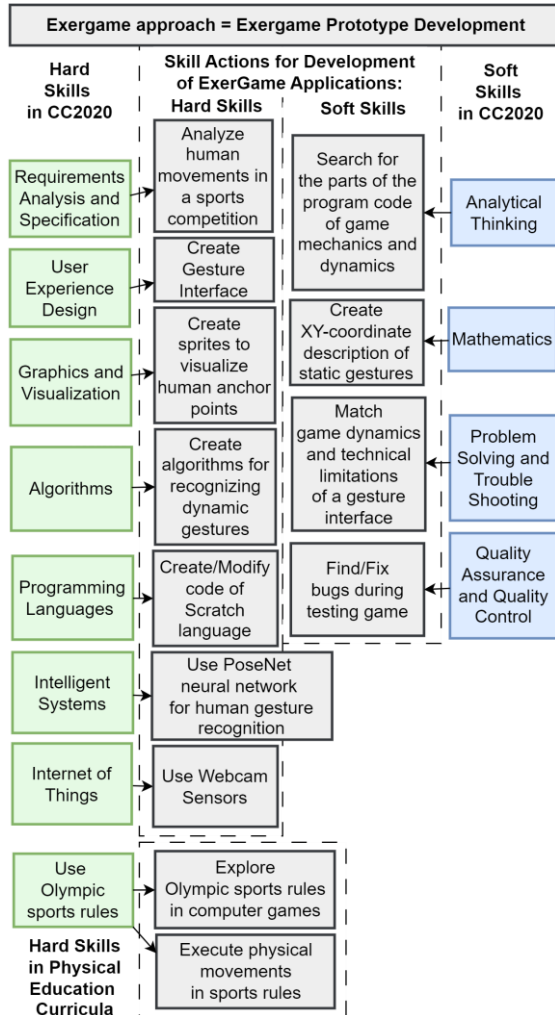

Figure 4: Examples of connections between Hard/Soft Skills of CC2020 and skills of the Exergame approach.

3 CONCLUSIONS

In October 2023, ninety two students participated in five-week online classes of the "Physical Education" discipline. Seventy students (76%) successfully completed all steps. The analysis of the results confirmed that students starting to study "Physical Education" in the first semester and lacking sufficient skills in professional programming, were already able to create computer games with 2D graphics and could assess/test their motor skills. An effect of implemented the Exergame approach consists of:

- 57% of students felt motivated to continue using the Exergame approach and demonstrated the desire to achieve decent results, that indicates skill "Positive Thinking" in "Physical Education";
- seven Hard Skills and four Soft Skills of CC2020 were indicated to compensate for the lack of classroom learning.

REFERENCES

[1] "CC2020 Task Force. Computing Curricula 2020: Paradigms for Global Computing Education". Association for Computing Machinery, New York, USA, 2020.

[2] S. Aliberti and G. Manzo, "Physical and sport education for life and soft skills development". Formazione & Insegnamento, vol. 21(1S), pp. 79-85.

[3] J.M. Krause and S.E. Jenny, "Physical educators' exergaming integration experiences, attitudes, and self-efficacy beliefs". The Physical Educator, vol. 80, 2023, pp. 91-107.

[4] M. Kourakli, I. Altanis, S. Retalis, M. Boloudakis, D. Zbainos, and K. Antonopoulou, "Towards the improvement of the cognitive, motoric and academic skills of students with special educational needs using Kinect learning games", International Journal of Child-Computer Interaction, vol. 11, 2017, pp. 28-39.

[5] J.-L. Chung, L.-Y. Ong, and M.-C. Leow, "Comparative Analysis of Skeleton-Based Human Pose Estimation", Future Internet", vol. 14(12), 2022, doi: 10.3390/fi14120380.

[6] S. Grover, "Teaching and Assessing for Transfer from Block-to-Text Programming in Middle School Computer Science", Research in Mathematics Education, Springer, Cham, 2021.

[7] "Scratch project Explore", [Online], Available: https://scratch.mit.edu/explore/projects/games/.

[8] S. Howell, "Install Scratch and Kinect2Scratch", In J. St. Jean (Ed.), Kinect Hacks: Tips and Tools for Motion and Pattern Detection, Sebastopol, CA: O'Reilly, 2012, pp. 184-206.

[9] S. Howell, "Kinect2Scratch (Version 1.5) [Computer Software]", [Online], Available: https://stephenhowell.github.io/kinect2scratch/.

[10] J.-W. Hung, et al., "Comparison of Kinect2Scratch game-based training and therapist-based training for the improvement of upper extremity functions of patients with chronic stroke: A randomized controlled single-blinded trial", European Journal of Physical and Rehabilitation Medicine, vol. 55, 2019, pp.542-550.

[11] J. Ishihara, "PoseNet2Scratch", [Online], Available: https://github.com/champierre/posenet2scratch.

[12] L.C. Duarte and A.L. Battaiola, "Distinctive features and game design", Entertainment Computing, vol. 21, 2017, pp. 83-93.

Data Analysis and Processing

Machine Learning-Based Forecasting of Bitcoin Price Movements

Darko Angelovski, Bojana Velichkovska, Goran Jakimovski, Danijela Efnusheva and Marija Kalendar

Faculty of Electrical Engineering and Infomation Technologies, SS. Cyril and Methodius University,
Rugjer Boshkovikj Str. 18, 1000 Skopje, N. Macedonia
darkoangelovski666@gmail.com, {bojanav, goranj, danijela, marijaka}@feit.ukim.edu.mk

Abstract: In the volatile realm of cryptocurrency markets, this research explores the intricate dance of Bitcoin price dynamics through the lens of machine learning. Employing a multifaceted approach, we harness the power of Long Short-Term Memory (LSTM) networks, Gradient Boosting, LightGBM (LGBM) Regressor, and Random Forest algorithms to unravel the complexities of price movements. We perform a comprehensive analysis, and observe patterns and dependencies within historical data at hour-long intervals in the last 30 and 45 days, by using a holdout technique with 80% of the data used for training and 20% used for testing. We evaluate the models using four standard regression metrics. The training data incorporates a diverse range of features capturing hourly trends, day-of-the-week variations, and the correlation between opening and closing prices. Our study delves into the ability for forecasting Bitcoin price movements using ensemble algorithms and LSTM. The results show best performance for the LSTM models, especially when trained on longer training intervals. Namely, our LSTM model obtains R2 of 0.98 when trained on 30 days and 0.99 when trained on 45 days. In comparison, the ensemble methods show volatility and lower predictive ability.

1 INTRODUCTION

In the dynamic realm of cryptocurrency markets, the ability to anticipate price movements holds immense significance for investors and traders. In 2008, Nakamoto [1] has introduced an electronic peer-to-peer cash system Bitcoin to the world. Over the last decade, the cryptocurrency market has grown tremendously, whereby individual cryptocurrency prices have exhibited large volatility [2]. Being able to understand and predict said volatile changes is an ongoing challenge, which if successful can significantly influence the cryptocurrency market.

Cryptocurrency relies on the Blockchain [3]. By implementing an access management mechanism, Blockchain systems provide ways of ensuring the privacy and protection of user data. The currency is based on a decentralized peer-to-peer network that creates currencies and management of transactions without central authority. All Bitcoin transactions are posted in blocks to an open directory, which is called Blockchain. As the flagship cryptocurrency, Bitcoin's volatile nature poses a captivating challenge, necessitating sophisticated tools for accurate predictions.

This research focuses on predictive modeling, by employing advanced machine learning (ML) algorithms: Long Short-Term Memory (LSTM) [4], Gradient Boosting [5], Light Gradient Boosting Machine (LGBM) [6], and Random Forest [7].
Our investigation encompasses the challenges faced in preprocessing historical Bitcoin data, the representation of predictive features, and the comparative evaluation of several algorithms in predicting future price changes of Bitcoin. As we navigate the landscape of Bitcoin price prediction, our aim is to contribute valuable insights to the broader discourse on the application of ML in cryptocurrency markets.

2 RELATED WORK

Researchers have already tackled this problem. Majority use deep learning methodologies, however some venture into the application of standard ML approaches. In spite of the approach, both long-term and short-term strategies are tested. In [8], the authors predict Bitcoin prices at one-minute intervals for data collected from 2012 to 2018. The

authors of [9] focus on the performance of deep learning models for three popular cryptocurrencies. In [10], the authors use ensemble learning methods and neural networks to analyze the closing price of several cryptocurrencies for a seven-day dataset collected in 2019.

In [11] the authors observe 20-day history of price, volume, and market capitalization in order to preform one-day trading decisions, whereas the research in [12] LSTM is used for data analysis in a span of three years. The results indicate promise in the realm of cryptocurrency trading. Another daily cryptocurrency approach is presented in [13], where the authors predict binary relative daily market movements of the 100 largest cryptocurrencies and their results show statistically viable predictions. In [14], the time analysis performed results with best model performance that carries a mean average error of 227. The authors of [15] assess various neural network approaches for predicting daily Bitcoin prices, and the findings reveal that a gated recurrent unit implementation with recurrent dropout stands out as the top performer on their dataset. In the research detailed in paper [16], advanced artificial intelligence frameworks are harnessed to understand Bitcoin, Ethereum, and Ripple's business dynamics. The study reveals that the Artificial Neural Network capitalizes on longer-term historical data, whereas the LSTM specializes in capturing swift dynamics. In the context of paper [17], the focus is on extracting and comparing the accuracy of Bitcoin predictions using diverse ML.

3 METHODOLOGY

This section outlines the methodology employed for the research, i.e., the data used and its preprocessing, the algorithms applied, and the metrics used for model evaluation.

3.1 Dataset

The bitcoin historical prices were collected using the python-binance API [18]. Two separate analyses were performed. The first considered collecting data in a period of 30 days from 25th of December 2023 at 01:00h to 24th of January 2024 at 01:00h and the second in a period of 45 days from 10th of December 2023 at 01:00h to 24th of January 2024 at 01:00h. In both cases, we consider the following features: open, high, low and close. Additionally, the API returns a datetime information for when each of these features was observed. For the purpose of the study, we do

not consider the entire datetime. Instead, we only extract and use the hour of the day and the day of the week. Moreover, for data brevity purposes and light model building, we do not consider the additional data received as response from the API, namely information on volume, number of trades, taker base volume, etc. are excluded. As most features (open, high, low, close) were retrieved as object types, the processing included adequate conversion to a numeric value. The open feature shows the starting price for every hour, the high represents the highest price in each hour, low is the lowest price for every hour, whereas close the closing price or final price for every hour. The additionally added features are self-explanatory. The target variable we are focusing on is the price movement for each hour, i.e., the difference between the starting and ending prices of bitcoin in each hour. The six features which are the input variables for our models and the output variable (price movement) are given in Table 1.

Table 1: Features for the model. The last row of the table is the target feature, which the model should predict.

Feature	Description	Unit
open	start price at every hour	USDT
high	highest price at every hour	USDT
low	lowest price at every hour	USDT
close	end price at every hour	USDT
hour	hour in the day (extracted from datetime)	Integer value [0-23]
week day	day in the week (extracted from datetime)	Integer value
price movement	difference between ending and starting price at every hour	USDT

As the data covers a significant range of values, we normalize the data before proceeding with training. Moreover, the data is divided in a holdout technique with 80% of the data being used for training and the remaining 20% used for testing purposes, for both the 30-day and the 45-day interval. The reason behind observing historical data for every hour in a period of 30 and 45 days specifically is the question whether successful predictions can be made by capturing shorter-term trends and patterns.

3.2 Algorithms

This study focuses on comparing the performance of ensemble methods to that of LSTM neural networks. From ensemble methods, we are analyzing Random Forest (RF), Gradient Boosting (GB) and Light

Gradient Boosting Machine (LGBM). As ensemble models combine predictions of multiple individual models to form their final assessment, they enhance generalization and can improve overall performance. The idea behind ensemble modeling is that by aggregating the predictions of multiple diverse models, the ensemble can often outperform any individual model in terms of accuracy and robustness, which is why we aimed to compare their performance to a LSTM network.

RF, as an ensemble algorithm creates diverse trees by using bootstrapped samples and random subsets of features, with the final prediction is obtained by averaging. This accounts for reducing overfitting, handling noisy data, and for feature importance. GB builds a strong predictive model also by sequentially improving weak learners, often decision trees. It corrects errors made by the previous models, and as such excels in handling complex relationships in data. LightGBM is a gradient boosting framework that efficiently handles large datasets. It uses a tree-based learning algorithm and employs a novel technique called Gradient-based One-Side Sampling to reduce the amount of data used for building trees.

Compared to ensemble algorithms, LSTM networks focus on understanding long-term dependencies in sequential data, and are particularly effective for time-series data, hence the application. The LSTM contain memory cells which can store and retrieve information over extended periods, particularly relevant information.

During the training phase for every ensemble model, we are going to use 100 decision trees, maximum depth of 15 layers per decision tree and additionally, GB and LGBM have a learning rate of 0.001. On the other hand, the LSTM neural network is trained on 100 epochs, batch size of 32, with a callback for early stopping which monitors the root mean squared error with patience of 50 epochs. The LSTM uses Adam optimizer, as an iterative optimization algorithm which can best minimize the loss function. The LSTM model consists of two LSTM layers with 256 units, two dropout layers, one after each LSTM layer, with a dropout rate of 20% and one dense layer with 1 output unit. Moreover, the first LSTM layer uses the hyperbolic tangent function (tanh) as an activation function.

3.3 Metrics

In order to evaluate the performance of the model, we calculated the following metrics: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination R^2. MSE quantifies the average squared difference between the actual and predicted values, and the formula is given in Figure 1. The RMSE shows the average magnitude of the errors between predicted and actual values, and is calculated as the square root of the MSE. The formula is given in Figure 2. MAE quantifies the accuracy of predictions by measuring the average error magnitude between predicted and actual values. Unlike MSE and RMSE, MAE uses the absolute values of the errors. The formula for MAE is given in Figure 3. R2 gives a sense of how well a model captures the relationship between independent and dependent variables. Unlike MSE, RMSE and MAE, R2 is a score between 0 and 1 where a value of 1 means that the model perfectly predicts target variables based on the provided features. In some cases, the score can be negative, meaning the model is not capturing any meaningful relationships between the independent and the dependent variables. The formula for R2 is given in Figure 4 where SSR (sum of squares) defines the sum of squared differences between the predicted value and the actual values and SST (total sum of squares) represents the total sum of squares of the differences between each data point and the mean of the dependent variable (target) in the dataset.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(actual_i - predicted_i)^2$$

Figure 1: Formula for MSE.

$$RMSE = \sqrt{MSE}$$

Figure 2: Formula for RMSE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|actual_i - predicted_i|$$

Figure 3: Formula for MAE.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n}(actual_i - predicted_i)^2}{\sum_{i=1}^{n}(actual_i - mean_i)^2}$$

Figure 4: Formula for R-squared.

4 RESULTS

The obtained results observe performance of four algorithms in their ability to predict Bitcoin price movements using the features given in Table 1, i.e., three ensemble RF, GB, and LGBM, and one deep learning approach, i.e., LSTM. We compare two different time intervals, i.e. we observe price movements in the last 30 and last 45 days. The results are given in Table 2. The results in this table show the predictive performance of the models, with values showing the differences between predicted and actual value.

As can be observed, there is a significant error for all ML models when training on 30 days. The worst performance can be observed by GB, with RF giving the best predictions. As RF offers best performance from all ensemble algorithms, only these predictions are shown in the figures. The performance of the RF model compared to the original price movements is given in Figure 5, with the actual movements denoted with green, and the predicted movements in purple. As the figure shows, RF can almost always predict the direction of the

price movements. However, the model struggles with following the amplitude of the change, which is also demonstrated by the MAE of 96.91 seen in Table 2. This can be distinctly and firstly noted around the 10th hour, where the prediction detects a drop in the data, even if on smaller scale compared to the original. Also, the results for RF around the 45th hour in, predicted price movements are in totally different direction compared to the actual price movements.

Table 2 Results from training the algorithms with 30-day and 45-day intervals. Three metrics are observed.

Algorithm	MSE	RMSE	MAE	R^2
30 days				
RF	56945.22	238.63	96.91	0.11
GB	63313.60	251.62	149.01	0.01
LGBM	62864.45	250.73	150.95	0.01
LSTM	641.77	25.33	15.63	0.98
45 days				
RF	18857.53	137.32	76.01	0.66
GB	54632.86	233.74	149.96	0.02
LGBM	52927.02	230.06	143.10	0.05
LSTM	257.29	16.04	11.90	0.99
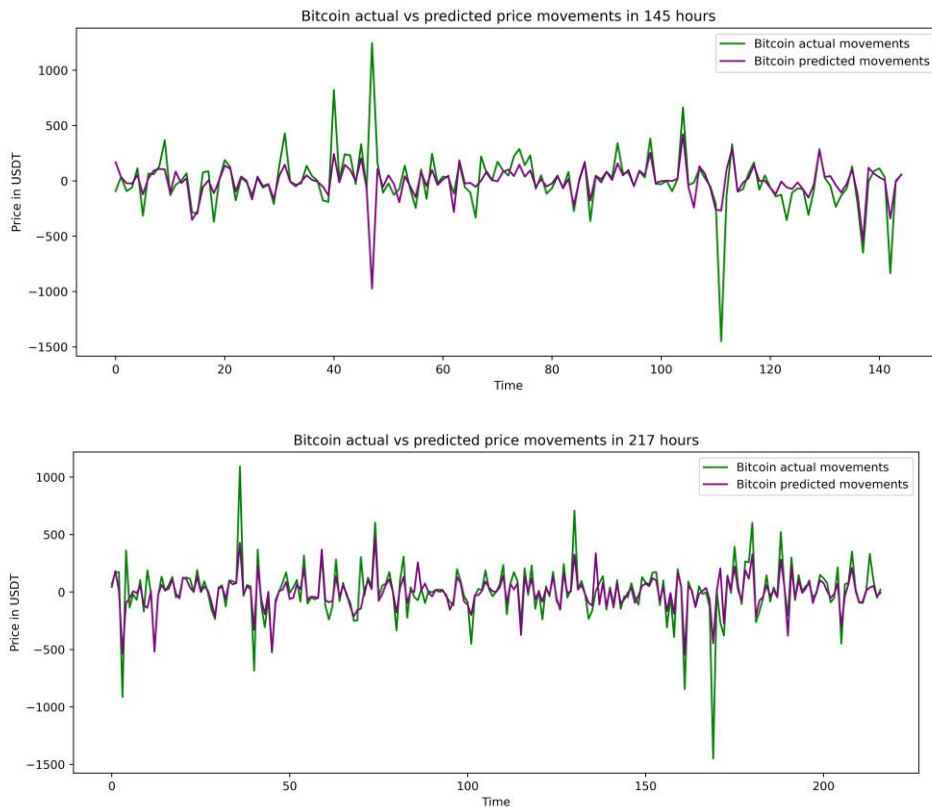
Figure 5: Actual and predicted price movements for RF trained on a 30-day period (above) and 45-day period (under).
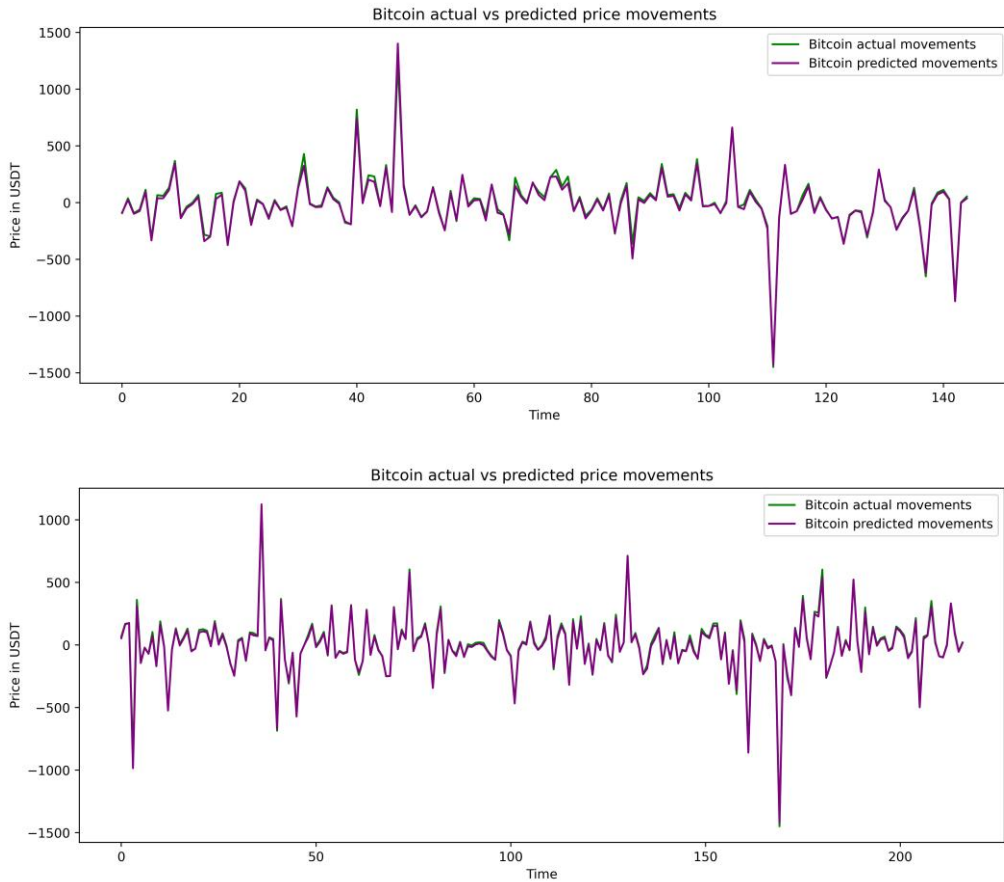
Figure 6: Actual and predicted price movements for LSTM trained on a 30-day period (above) and 45-day period (under).

On the other hand, improvement in performance can be noted in all models when trained on 45-day interval, as Table 2 clearly shows. This is visible in both Figure 5 and Figure 6, i.e., the improvement can be observed all throughout the timeline. It can be seen that predicted price movements are following the line of the actual price movements. The results in Table 2 show highest improvement in performance for the LSTM approach. Namely, from a MAE of 15.63 for the 30-day interval, the model reaches a MAE of 11.9. Moreover, the MAE of 11.9 when trained and tested on data from 45 days is the lowest for all models trained and tested in the 45-day interval. Thus, not only does the LSTM provide highest improvement, it also gives the best results overall.

The actual and predicted price movements for the LSTM for the 30-day and 45-day intervals are given in Figure 6. The figure further illustrates the significant improvement in predictions between the two approaches, showing how LSTM benefits from the added 15 days in understanding the performance better. The error which occurred around the 45th hour when applying the RF does not occur here, as LSTM has a better ability to memorize previous trends in data compared to ensemble algorithms.

The approach provides an accurate prediction model when tested on data from the selected time period. This is one of the model's limitations, i.e., the model has only been evaluated for testing data in a close period to the data used in training. However, the time frame within which these predictions remain valid can only be determined by further exposing the model to real time data and re-observing its performance, which is intended as a future step in this research. Moreover, as Bitcoin prices have volatile market movements future research can focus on understanding the time interval of model validity before retraining is a necessity.

5 CONCLUSIONS

The study observed how different models perform when predicting cryptocurrency price movements with only six features used as input at an hour-long interval. Four different models were trained, three ensemble algorithms and one LSTM approach, and changes in model performance was observed over the course of both 30 and 45 days of data for the training and testing of the models. The results show better performance for the 45-day interval, and particular improvement can be noted with the RF and LSTM algorithms, and it was also concluded that the LSTM algorithm has significantly better performance compared to the ensemble algorithm.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf.

[2] Coinmarketcap, [Online]. Available: https://coinmarketcap.com/.

[3] P. Morgen, "Reinforcing the links of Blockchain," IEEE Spectrum Magazine special edition "Blockchain World," 2017.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-80, 1997.

[5] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," Frontiers in Neurorobotics, vol. 7, 2013.

[6] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[7] G. Louppe, "Understanding Random Forests: From Theory to Practice," 2015. arXiv:1407.7502.

[8] T. Phaladisailoed and T. Numnonda, "Machine learning models comparison for bitcoin price prediction," in 2018 10th International Conference on Information Technology and Electrical Engineering, pp. 506-511, 2018.

[9] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach," Fractal and Fractional, vol. 7, no. 2, pp. 203, 2023.

[10] R. Chowdhury, M. A. Rahman, M. S. Rahman, and M. R. C. Mahdy, "An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning," Physica A: Statistical Mechanics and its Applications, vol. 551, pp. 124569, 2020.

[11] C. Betancourt and W. H. Chen, "Reinforcement learning with self-attention networks for cryptocurrency trading," Appl Sci, vol. 11, pp. 7377, 2021.

[12] S. McNally, J. Roche, and S. Caton, "Predicting the price of bitcoin using machine learning," in Proceedings of 2018 Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, pp. 339-343, 2018.

[13] P. Jaquart, S. Kopke, and C. Weinhardt, "Machine learning for cryptocurrency market prediction and trading," The Journal of Finance and Data Science, vol. 8, pp. 331-352, 2022.

[14] M. Iqbal, M. S. Iqbal, et al., "Time-Series Prediction of Cryptocurrency Market using Machine Learning Techniques," Endorsed Transactions on Creative Technologies, vol. 8, no. 28, 2021.

[15] A. Dutta, S. Kumar, and M. Basu, "A gated recurrent unit approach to bitcoin price prediction," J Risk Financ Manag, vol. 13, 2020.

[16] W. Yiying, "Cryptocurrency Price Analysis With Artificial Intelligence," in 2019 5th Int. Conf. Inf. Manag., pp. 97-101, 2019.

[17] K. Rathan, S. V. Sai, and T. S. Manikanta, "Crypto-Currency price prediction using Decision Tree and Regression techniques," in 2019 3rd Int. Conf. Trends Electron. Informatics, pp. 190-194, 2019.

[18] Python-Binance API, [Online]. Available: https://python-binance.readthedocs.io/en/latest.

Comparative Analysis of Holt-Winters Algorithms on the Oracle Machine Learning Platform

Oleksandr Redych[1] and Ruslan Boichuk[2]

[1]State Tax University, Universytetska Str. 31, 08205 Irpin, Ukraine
[2]Main Department of the State Tax Service in Ivano-Frankivsk region, Nezalezhnosti Str. 20,
76000 Ivano-Frankivsk, Ukraine
redich2@gmail.com, if.dps@ukr.net

Keywords: Machine Learning, Oracle Cloud, APEX, Holt-Winters, Personal Income Tax.

Abstract: The Oracle Cloud Machine Learning toolkit was used in the work to study time series forecasting methods. The choice of tools was dictated by conducting research at the State Tax University (Irpin) for further use of the results in the work of regional and central divisions of the state tax service, whose information systems are developed on the Oracle platform. Holt-Winters exponential smoothing algorithms, currently represented by the functions of the application programming interface for machine learning models of the PL/SQL package DBMS_DATA_MINING, were investigated. The research used personal income tax data for 2015-2023, obtained from the tax office in the Ivano-Frankivsk region. Higher accuracy (MAE=3,57%) was shown by the algorithm of the Holt-Winters multiplicative exponential smoothing model with a fading multiplicative trend and multiplicative seasonality. Oracle Machine Learning for SQL implements exponential smoothing using a state of the art state space method that incorporates a single source of error (SSOE) assumption which provides theoretical and performance advantages. Access to research results is organized using the APEX web application creation tool. The considered toolkit will help in making decisions when assessing the base of the tax potential of the region and planning tax revenues. The results of the Holt-Winters exponential smoothing model algorithms research are presented and the losses of personal income tax in Ivano-Frankivsk region in 2022 are estimated.

1 INTRODUCTION

The information infrastructure of the Ministry of Finance of Ukraine, the State Tax Service, the National Bank and other central bodies of Ukraine is being developed on software products of the Oracle Corporation [1-3]. The core of the information systems of these institutions is the Oracle DBMS, which has proven itself to be a secure and reliable system. An important component of the operational activity of financial institutions is data analysis, detection of hidden patterns, forecasting, classification. This paper shows the results of the Holt-Winters exponential smoothing algorithm research on the Oracle Machine Learning platform for forecasting Personal Income Tax (PIT) revenues in the Ivano-Frankivsk region (Ukraine). The relevance of personal income tax research is determined by its high social and economic significance. Before the start of the war, during 2019-2021, the specific weight of personal income

tax in the structure of tax revenues and revenues of the Consolidated Budget of Ukraine was more than 20%; in the structure of tax revenues of local budgets – more than 60% [4]. During the war in 2022-2023, the share of personal income tax in the total revenues of the Consolidated Budget decreased to 19,5% and 15,6%, respectively, which is largely explained by the mobilization of taxpayers and the absence of taxation of the financial support of conscripts.

Forecasting PIT revenues enables more accurate budgeting and expenditure planning for the government. This allows for the adaptation of budget priorities, rationalization of expenses, and ensuring financial stability. Forecasting tax revenue is a crucial tool for tax policy planning. Based on it, decisions can be made regarding the adjustment of tax rates, defining priorities in tax collection, and improving administrative procedures [5].

2 ORACLE MACHINE LEARNING PLATFORM

Oracle Cloud Machine Learning (OML) supports algorithms for data exploration, training and modeling with machine learning using SQL, R, Python, REST. OML includes more than 30 high-performance algorithms for working with databases that create models for use in applications [6]. It allows data analysts and other data professionals to quickly build models by automating key elements of the machine learning lifecycle. The technological advantage of using the OML platform is Oracle Corporation's focus on expanding cloud services, in particular, providing machine learning directly on the Oracle database. OML offers machine learning features such as classification, regression, clustering, feature extraction, anomaly detection, association (market basket analysis), time series, BigData, and more. Each machine learning function has several variants of calculation algorithms. OML uses built-in Oracle database features for maximum scalability and performance [7].

Using open source packages of R and Python, users can extend the set of methods and algorithms, as shown in the examples on the OML cloud environment (in the OML4Py and OML4R notebooks). In Oracle DB21c, Oracle Data Mining has been renamed to Oracle Machine Learning for SQL (OML4SQL), but the name of the PL/SQL package DBMS_DATA_MINING has not changed. The DBMS_DATA_MINING package exposes APIs that are leveraged by the Oracle Machine Learning for SQL. Users who wish to create machine learning models in their own schema require the CREATE MINING MODEL system privilege.

2.1 Holt-Winters Models in Time Series Forecasting

In its simplest form, exponential smoothing is a one-parameter moving average method that models the exponential decrease in the influence of past levels on future values. Winters [8] developed Holt's [9] exponential smoothing model with a trend and added seasonality to it. The advantage of this method lies in its ability to make forecasts over a long period. However, in order to make a forecast, for example, for 1 year, data for at least 2 complete years are required, preferably for 3-5 complete years.

Let's consider the Holt-Winters Triple Exponential Smoothing model, which takes into account three smoothing parameters: level, trend, and seasonality . This model uses additive trend and multiplicative seasonality [10]:

1) An additive trend implies that the trend changes over time with a constant absolute change. For example, if the trend value is 10, then at each time step it will increase by 10. This is suitable for data where trend changes occur with constant absolute values.
2) Multiplicative seasonality means that seasonal fluctuations change as a percentage of the baseline. For example, if the baseline data has a value of 100 and the seasonal fluctuation is 10%, then during the season the data will fluctuate between 90 and 110. This approach is more suitable when relative changes in seasonality are more significant than absolute values.

The Holt-Winters method introduces two types of smoothing: exponential smoothing of the level and exponential smoothing of the trend. These smoothing parameters are denoted as α (alpha) for the level, β (beta) for the trend, and γ (gamma) for seasonality. The following formulas are used in this Holt Winters model:

Level:

$$L(t) = \alpha * (Y(t) / S(t\text{-}L)) + (1 - \alpha) * (L(t\text{-}1) + T(t\text{-}1)),$$

Trend:

$$T(t) = \beta * (L(t) - L(t\text{-}1)) + (1 - \beta) * T(t\text{-}1),$$

Seasonality:

$$S(t) = \gamma * (Y(t) / L(t)) + (1 - \gamma) * S(t\text{-}L),$$

Forecast:

$$F(t\text{+}m) = (L(t) + m * T(t)) * S(t\text{+}m)$$

where: $L(t)$ is the smoothed value for the current period; $Y(t)$ is the actual value for the current period; $S(t)$ is the seasonality coefficient for the current period; $T(t)$ is the trend value for the current period; $F(t\text{+}m)$ forecast for m periods ahead; α, β, γ smoothing coefficients for: series level, trend and seasonality, respectively:

- Level: This represents the average value of the time series. The model assumes that the series fluctuates around a certain level. The level indicates the baseline level of the time series without the influence of trend and seasonality.

Table 1: Algorithms of the Holt-Winters exponential smoothing model.

#	Holt-Winters model algorithm	Value of the model type parameter
1	Holt-Winters triple exponential smoothing model, additive trend, multiplicative seasonality. Offering a robust model for data with both linear trend and proportional seasonal variation.	EXSM_WINTERS
2	Holt-Winters multiplicative exponential smoothing model with damped multiplicative trend, multiplicative seasonality. Effectively moderating the exponential increase or decrease of both trend and seasonal components over time.	EXSM_WINTERS_MUL _TREND_DMP
3	Holt-Winters multiplicative exponential smoothing model with damped trend, additive trend, multiplicative seasonality. Moderating the linear trend over time while still capturing proportional seasonal changes.	EXSM_WINTERS_DAM PED

- Trend: This denotes the direction and speed of changes in the time series. It indicates whether the series is increasing, decreasing, or remaining constant over time. The trend can be linear or nonlinear.
- Seasonality: This refers to regular oscillations or patterns in the time series that repeat at fixed intervals. For example, annual seasonal effects or monthly fluctuations.

A total of 14 exponential smoothing algorithms are supported. OML implements exponential smoothing using a state-of-the-art state-space method that incorporates the Single Source of Error (SSOE) assumption, providing theoretical and operational advantages. An appealing feature of SSOE is that the estimates of the state variables converge in probability to their true values, thereby leading to a formal inferential structure for the ad-hoc exponential smoothing methods for forecasting [11].

The maximum accuracy of the forecast in OML is achieved by successive selection of smoothing coefficients of series, trend and seasonality α, β, γ in the range from 0 to 1. This type of model considers various combinations of additive and multiplicative trend, seasonality and error, with and without trend attenuation. With various extensions, exponential smoothing covers a wider class of models than the Box-Jenkins auto-regression integrated moving average (ARIMA) approach.

2.2 Setting Parameters of the Holt-Winters Model for Forecasting

In this work, OML exponential smoothing algorithms for SQL (*"OML4SQL Time Series ESM"*) are investigated on PIT data for 2015-2023(Jan-May) and for its forecasting in 2023-2024. The "OML4SQL Time Series ESM" algorithm was selected because it was the sole time series forecasting algorithm available on the OML platform at the time of the study. Examples of OML algorithms of models on training datasets are provided in the form of notebooks (OML Notebooks), which can be taken as a basic version and used. This paper highlights the results of the three most accurate exponential smoothing algorithms studied (Table 1).

The studied time series is the personal income tax data from January 2015 to May 2023 as of the 1st of each month (PIT1523 time series table with 101 values). Date - column of dates of receipts and column PIT - actual receipts of personal income tax. The analysis of the series showed the presence of a linear trend and seasonality after 12 months, so in the Holt-Winters forecasting model, the value of the seasonality parameter can be set to 12 and the forecast window can also be specified as 12 months.

The main stage is setting the parameters of the selected model. The parameter ALGO_NAME defines the main algorithm of the model, which we have exponential smoothing

```
- ALGO_EXPONENTIAL_SMOOTHING.
```

The EXSM_MODEL parameter defines the algorithm type itself. In OML, the Triple Exponential Holt-Winters Smoothing Model algorithm is called "EXSM_WINTERS". So, we set the parameters of the EXSM_WINTERS model (Figure 1):

- Sets the interval of the data set or the interval size unit, such as day, week, month, and so on. This setting applies only to a time column with a date and time type. For example, we need to provide for monthly PIT, so we set this parameter to `EXSM_INTERVAL _MONTH`. The values of the parameters are taken from the OML documentation [12]:

```
v_setlst('EXSM_INTERVAL')
:= 'EXSM_INTERVAL_MONTH';
```

```
'EXSM_WINTERS' Create a Holt-Winters Model using monthly parameters          FINISHED
                                                                             ▷ ⅀ 🕮 ⚙ ▾
%script

BEGIN DBMS_DATA_MINING.DROP_MODEL('PIT1523');      -- name of the model
EXCEPTION WHEN OTHERS THEN NULL; END;
/
DECLARE
    v_setlst DBMS_DATA_MINING.SETTING_LIST;
BEGIN

    v_setlst('ALGO_NAME')          := 'ALGO_EXPONENTIAL_SMOOTHING'; -- name of the algorithm
    v_setlst('EXSM_INTERVAL')      := 'EXSM_INTERVAL_MONTH';        -- accumulation int'l = month
    v_setlst('EXSM_PREDICTION_STEP') := '12';                       -- prediction step = 12 months
    v_setlst('EXSM_MODEL')         := 'EXSM_WINTERS';               -- ESM model Holt-Winters
    v_setlst('EXSM_SEASONALITY')   := '12';                         -- seasonal cycle = 12 months

    DBMS_DATA_MINING.CREATE_MODEL2(
        MODEL_NAME        => 'PIT1523',              -- name of the model
        MINING_FUNCTION   => 'TIME_SERIES',          -- Oracle Data Miining  function = TIME_SERIES
        DATA_QUERY        => 'select * from PIT1523', -- SELECT from table PIT1523
        SET_LIST          => v_setlst,               -- reference table
        CASE_ID_COLUMN_NAME => 'Date',               -- Date colamn from  PIT1523
        TARGET_COLUMN_NAME =>'PIT');                 -- PIT data from a time series table PIT1523
END;
```

Figure 1: The Holt-Winters algorithm EXSM_WINTERS with the set parameters in the OML notebook.

- Setting the forecast window parameter. If we want to display each value representing a quarter, then a value of 4 gives four values (quarters) of forecasting for the future. For forecasting monthly PIT, this parameter is set to "12":

```
v_setlst('EXSM_PREDICTION_STEP'):=
'12';
```

Setting the algorithm of the exponential smoothing model to be used.

```
v_setlst('EXSM_MODEL'):='EXSM_WINTE
RS';
```

- Set the seasonality parameter. The parameter defines a positive integer value as the duration of the seasonal cycle. The value it takes must be greater than 1. For example, 4 means that each group of four values forms a seasonal cycle (a quarter for monthly data), which makes sense if you use 4 quarters to represent a year. In our case, when forecasting PIT, we set this parameter to 12.

```
v_setlst('EXSM_SEASONALITY'):=
'12';
```

- Setting the parameter of the method for processing missing values. In time series, the special value EXSM_MISS_AUTO indicates that if the series contains missing values, it should be treated as an invalid time series.

```
v_setlst('EXSM_SETMISSING'):=
'EXSM_MISS_AUTO';
```

- Also sets the values of the main parameters that point to the target value column and date column, etc:

```
CASE_ID_COLUMN_NAME => 'Date' and
TARGET_COLUMN_NAME => 'PIT';
```

By default, the confidence level is set to 0.95 and the EXSM_NMSE parameter is set to 3, which determines the window used to calculate the average mean square error (AMSE) metric.

2.3 Evaluation of Results

The tables of views listed in Table 2 contain information about the input data, model quality assessment, and global information. The names of the views include the model name "PIT1523".

Table 2: Tables views of the Holt-Winters model results.

#	Views with the model Details	Description
1	DM$VGPIT1523	Global Name-Value Pairs
2	DM$VPPIT1523	Exponential Smoothing Forecast
3	DM$VSPIT1523	Computed Settings
4	DM$VWPIT1523	Model Build Alerts
5	DM$VRPIT1523	Time Series Regression Build
6	DM$VGPIT1523	Global Name-Value Pairs

Table 3: Estimating the parameters of ESM algorithms.

#	Description	Name	Algorithm (see Table 1)		
			1	2	3
1	Smoothing constant, α	ALPHA	3860,9018	3849,8462	3855,8791
2	Trend smoothing constant, β	BETA	0,0001	0,0001	0,015
3	Seasonal smoothing constant, γ	GAMMA	0,0001	0,0001	0,0001
4	Akaike information criterion	AIC	3867,3779	3857,2197	3863,2526
5	Bayesian information criterion	BIC	3902,7437	3894,3033	3900,3361
6	Model mean absolute error	MAE	0,0455	0,0412	0,0433
7	Model standard error	STD	0,0565	0,0531	0,0551

These views are generated in the user schema and can be used to display modeling results in an APEX-based application. More details about Exponential Smoothing model parameters here [12].

Table 3 shows the estimated parameters of the model from the views DM$VGPIT1523 for all three algorithms when forecasting data for the period from June 2023 to June 2024. Some explanations for the estimates:

- *Akaike's Information Criterion* (AIC) is a measure of the trade-off between model accuracy and complexity. A lower AIC value indicates a better model.
- *Bayesian Information Criterion* (BIC). BIC is an alternative information superiority criterion that penalizes more complex models. A lower BIC value indicates a better model.

As can be seen from Table 3, by most estimates, the best algorithm for our time series is the Holt-Winters Multiple Exponential Smoothing Model with Damped Multiple Trend, Multiple Seasonality (EXSM_WINTERS_MUL_ TREND_DMP).

Given sufficient accuracy, this algorithm was used to estimate the decrease in PIT in 2022 due to the armed aggression of the Russian Federation, since the model was trained on pre-war time series and the predicted values could correspond to the peaceful state of the economy. The forecasting results for 2021-2023 and the absolute errors are shown in Table 4.

As can be seen from the table, the absolute error of the simulated PIT revenues for 2021 is 0,65%, and the forecast error for 2022 is 11,26%, which is explained by the fact that the model could not predict the war and its impact and the model error, in fact, determines the amount of PIT losses. Probable PIT losses in 2022 may reach an average of 674 ± 1% million UAH. An estimate of the simulated PIT values for 5 months of 2023 gives a 6.4% absolute forecast error. It should be noted that the mean absolute error (MAE) shows that the mean absolute difference between the observed values and the predicted values for all predictions is < 4%, which is a good result.

Figure 5 shows a graph of actual and simulated PIT data for 2015-2024 with confidence intervals (confidence level 0,95), obtained in the OML notebook (Holt-Winters model algorithm #2, Table 1). You can see significant differences between the forecast values and the actual values at the beginning of the pandemic (2020) and during the armed aggression of the Russian Federation. However, the following values are already taken into account by the model and the quality of the forecast improves.

As already mentioned, the simulation results can be used when creating analytical applications using Oracle APEX (Application Express). Application Express is a supported feature of the Oracle Database and is included, at no additional cost, with every Oracle Database, both on-premises and in the cloud. There are no additional licensing costs based on the number of developers, applications or end-users. Application Express is also included with every Oracle Database Cloud Service, from the low-priced Oracle Database Exadata Express Cloud

Table 4: Estimating actual and forecast revenues PIT.

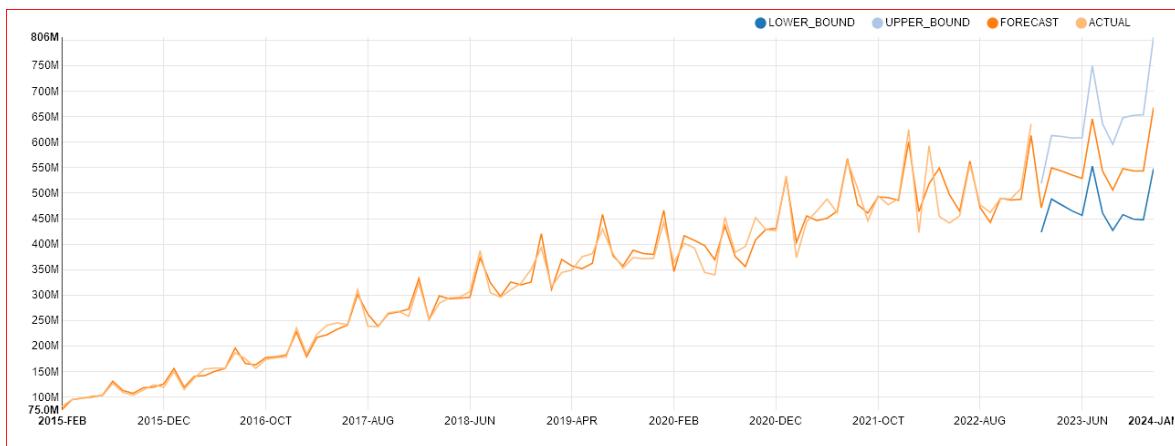| Year	Actual revenues PIT, UAH	Forecast of revenues PIT, UAH	Absolute error,per year, UAH	Relative error of revenue PIT per year, %	MAE
2021	5 836 364 394,25	5 798 285 924,96	-38078469,29	0,65%	0,0367
2022	5 986 078 790,44	6 659 949 560,32	673870769,88	-11,26%	0,0357
2023 5 months	2 470 910 417,81	2 628 752 677,00	157842259,19	6,40%	0,0397

Figure 5: Chart of actual and forecast values of PIT in Ivano-Frankivsk region 2015-2024 (confidence level of 0.95).

Service all the way up to the Oracle Database Exadata Cloud Service [13].

Features of creating an Apex schema user to access OML:

1) Create the Workspace and the user schema on APEX.
2) When creating a Workspace, a user schema with the same name is automatically created in Cloud DB (ATP or ADW).
3) Register as an Oracle Cloud Database Administrator and go to the user administration tab.
4) Select the created user, activate the access to OML button in the edit menu.

For example, in the view DM$VPPIT1523, the model forms columns of actual, forecast and calculated values for the lines of the confidence level specified in the parameters (lower_bound and upper_bound). All calculated values of the model coefficients and detailed information can also be displayed on the APEX pages.

3 CONCLUSIONS AND FURTHER RESEARCH

The following results were obtained in the study:

1) Among the 14 Holt-Winters algorithms implemented on the OML platform and used in the forecasting of PIT receipts (data from the Ivano-Frankivsk region for 2015-2023), the best forecasting accuracy was shown by the algorithm of Holt-Winters multiplicative exponential smoothing model with damped multiplicative trend and multiplicative seasonality (MAE = 3,57%).

2) The effectiveness of the algorithms can be explained by the use of an extended state space method for exponential smoothing, including the assumption of a single source of error (SSOE).

3) The results of modeling tax revenues in 2022 indicate that estimated losses of PIT in Ivano-Frankivsk region. due to the aggression of the Russian Federation, they may amount to UAH $654 \pm 1\%$ million UAH.

4) APEX, which is a supported and free feature of the Oracle database, allows fast, low-code creation of applications using simulation results. This enables convenient access to the OML toolkit and simulation results for a wide range of employees of corporations, enterprises, regulatory agencies, in particular, divisions of the State Tax Service, whose information systems are being developed on the Oracle platform.

Further research will be focused on exploring the algorithms of Oracle Machine Learning for Python (OML4Py), which enable the utilization of specialized libraries for data processing, machine learning, and graphical analysis available through the Oracle Autonomous Database service. OML4Py is a module that allows users to manipulate data in Oracle database tables using Python syntax. OML4Py functions and methods transparently transform selected Python function sets into SQL commands for execution directly on the Oracle database.

Subsequent investigations are planned to consider the integration of OML, REST, and Oracle Big Data services for a wide range of data collection and utilization tasks, validation of the obtained

results on tax revenue data from other regions, utilization of the discussed toolkit in studies of tax policy, and the development of methodological recommendations for the implementation of machine learning technologies in the operational activities of tax authorities [14, 15].

REFERENCES

[1] O. Puhachenko and T. Fomina, "Information support for expert studies and tax audits of the procedure for calculating and paying taxes and fees," Economics and society, no. 52, 2023, [Online]. Available: https://doi.org/10.32782/2524-0072/2023-52-72.

[2] T.V. Donchenko and A.V. Oliynyk, "Stages of formation of the tax system of Ukraine in the context of the development of the digital economy," Collection of scientific works of the State Fiscal Service University of Ukraine, no. 1, pp. 45-62, 2019, [Online]. Available: https://core.ac.uk/download/pdf/287302574.pdf.

[3] S.P. Kanzyuba, R.M. Matviychuk, Y.M. Sydorovych, and P.M. Musienko, "Electronic document flow. Reengineering of administrative processes in public authorities," Electronic governance and electronic democracy. Education manual, vol. 9, 64 p., 2017, [Online]. Available: https://books.google.de/books?id=paW4DwAAQBAJ&printsec=frontcover&hl=uk&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.

[4] O. Totska and I. Dmytruk, "Individual income tax in Ukraine: national and regional dimension," FCS, vol. 1, no. 8, pp. 30-39, Mar. 2023, [Online]. Available: https://doi.org/10.26565/2786-4995-2023-1-04.

[5] E.I. Maslennikov and A.O. Gusev, "Income tax of individuals in Ukraine: current state and ways of reform," Black Sea Economic Studies, no. 57, pp. 47-50, 2020, [Online]. Available: https://doi.org/10.32843/bses.57-8.

[6] Oracle, "Oracle Machine Learning for SQL Use Cases," 2023, [Online]. Available: https://docs.oracle.com/en/database/oracle/machine-learning/oml4sql/21/mlsql/machine-learning-functions.html#GUID-E5D374FB-F7D5-49ED-BBD4-272B8B8CADC2.

[7] A. Png and H. Helskyaho, "Oracle Machine Learning in Autonomous Database. In: Extending Oracle Application Express with Oracle Cloud Features," Apress, Berkeley, CA, pp.139-191, 2022, [Online]. Available: https://doi.org/10.1007/978-1-4842-8170-3_5.

[8] P.R. Winters, "Forecasting Sales by Exponentially Weighted Moving Averages," Management Science, vol. 6, no. 3, pp.324-342, 1960.

[9] C.E. Holt, "Forecasting Seasonals and Trends by Exponentially Weighted Averages," ONR Memorandum, no. 52, Carnegie Institute of Technology, Pittsburgh, USA, 1957. Available from the Engineering Library, University of Texas at Austin, 2003. Published in International Journal of Forecasting, vol. 20, pp. 1-13, 2004, with additional commentaries.

[10] C. Chatfield, "The Holt-Winters Forecasting Procedure," Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 27, no. 3, pp. 264-279, 1978, [Online]. Available: https://doi.org/10.2307/2347162.

[11] J. K. Ord, et al., "Time Series Forecasting: The Case for the Single Source of Error State Space Approach," Working Paper, Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia, April 2005, [Online]. Available: https://robjhyndman.com/papers/SSOE.pdf.

[12] Oracle, "DBMS_DATA_MINING - Algorithm Settings: Exponential Smoothing," 2023, [Online]. Available: https://docs.oracle.com/en/database/oracle/oracle-database/21/arpls/DBMS_DATA_MINING.html#GUID-A95A0A38-8A5A-4470-B49F-80D81C588BFC.

[13] Oracle, "Oracle® APEX," 2023, [Online]. Available: https://docs.oracle.com/en/database/oracle/apex/23.1/htmrn/index.html#Oracle%C2%AE-APEX.

[14] L. Canali, Z. Baranowski, and P. Kothuri, "Integration of Oracle and Hadoop: Hybrid Databases Affordable at Scale," Journal of Physics: Conference Series, vol. 898(2), 2017, doi: 10.1088/1742-6596/898/4/042055.

[15] A. Png and H. Helskyaho, "Oracle Machine Learning REST Services. In: Extending Oracle Application Express with Oracle Cloud Features," Apress, Berkeley, CA, pp.193-232, 2022, [Online]. Available: https://doi.org/10.1007/978-1-4842-8170-3_6.

Computational Breakthroughs in Aquatic Taxonomy: The Role of Deep Learning and DNA Barcoding

Nadiia Kasianchuk [1,2], Sofiia Harkava[3], Sofiia Onishchenko[4], Olesia Solodka[5],
Daria Shyshko[6,7], Eduard Siemens[8], Halina Falfushynska[8,9] and Taras Ustyianovych[10]

[1]Faculty of Biology, Adam Mickiewicz University in Poznan, Uniwersytetu Poznańskiego Str. 6, 61712 Poznań, Poland
2Faculty of Pharmacy, Bogomolets National University, Taras Shevchenko Str. 13, 01601 Kyiv, Ukraine
[3] Dnipro regional branch of the Ukrainian Junior Academy of Sciences, Gagarin Avenue 26, 49000 Dnipro, Ukraine
[4] Kharkiv regional branch of the Ukrainian Junior Academy of Sciences, Skrypnyka Str. 14, 61000 Kharkiv, Ukraine
[5] Rivne regional branch of the Ukrainian Junior Academy of Sciences, Symon Petliura Str. 17, 33028 Rivne, Ukraine
[6] Faculty of Biology and Ecology, Oles Honchar Dnipro National University, Gagarin Avenue 72,49000 Dnipro, Ukraine
[7] Klinik für Psychiatrie und Psychotherapie, Charite - Universitätsmedizin Berlin, Hindenburgdamm Str. 30,
12203 Berlin, Germany
[8] Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany
[9] University of Rostock, Albert Einstein Str. 3, 19059 Rostock, Germany
[10] Department of Artificial Intelligence Systems, Lviv Polytechnic National University,
Kniazia Romana Str. 5, 79005 Lviv, Ukraine
nadkas2@st.amu.edu.pl, sofiiaharkava@gmail.com, sonyastudy982@gmail.com, olesiasolodka@gmail.com,
daria.shyshko@charite.de, eduard.siemens@hs-anhalt.de, halina.falfushynska@uni-rostock.de,
taras.o.ustyianovych@lpnu.ua

Keywords: Aquatic Ecosystems; Deep Learning; Taxonomic Identification; Metagenomic Sequences; Environment Modeling; Machine Learning.

Abstract: Aquatic ecosystems are crucial in maintaining environmental equilibrium and sustaining human well-being. However, the traditional manual methods used in hydrobiological research have limitations in providing a far-reaching understanding of these intricate ecosystems. Data science, machine learning, and deep learning techniques offer a variety of opportunities to overcome these limitations and unlock new insights into aquatic environments. This study highlights the impact of computational tools in areas such as taxonomic identification, metagenomic sequence analysis, and water quality prediction. Deep learning techniques have demonstrated superior accuracy in classifying organisms, including those previously unidentified by conventional methods. In metagenomic sequence analysis, machine learning aids in effectively assembling DNA sequences, aligning them with known databases, and addressing challenges related to sequence repeats, errors, and missing data. Furthermore, predictive models have been developed to provide insights into water quality parameters, such as eutrophication events and heavy metal concentrations. These advancements lead to informed conservation measures and a deep understanding of the intricate relationships within aquatic ecosystems. However, challenges persist, including data quality issues, model interpretability, and the need for robust training datasets. Thus, data integration strategies designed specifically for environmental and genomic studies are necessary. Data fusion and imputation can help address data scarcity and provide a comprehensive view of hydrobiological processes. As the study of aquatic ecosystems continues to evolve, the synergy between computational methods and traditional hydrobiological techniques holds immense potential. By leveraging the power of data science and cutting-edge technologies, researchers can gain a deep understanding of aquatic environments, monitor changes in biodiversity, and develop informed strategies for sustainable management amidst global environmental shifts.

1 INTRODUCTION

Hydrobiological research stands as a cornerstone in the comprehensive understanding of aquatic ecosystems, which are critical to both environmental equilibrium and human sustenance [1]. Historically, hydrobiological inquiries were primarily anchored in manual sampling techniques and observational methodologies, offering insights that were often limited by the scope and resolution of available tools.

With the advent of modern technologies, there has been an exponential surge in the volume and granularity of data obtainable from aquatic environments, necessitating the integration of more sophisticated analytical tools [2].

The taxonomic identification of organisms within aquatic ecosystems is of paramount importance in hydrobiological research [3]. It serves as a fundamental pillar in assessing the health and biodiversity of these ecosystems, which in turn impacts their ecological functions and resilience to environmental changes. Accurate taxonomic identification allows scientists to monitor the presence of indicator species, assess shifts in community composition, and detect potential invasive species that can disrupt the balance of aquatic ecosystems [4]. However, the field of taxonomic identification faces common challenges, including the vast diversity of aquatic organisms, taxonomic ambiguities, and the need for rapid and cost-effective identification methods [5]. Additionally, environmental stressors such as climate change and pollution further underscore the importance of robust taxonomic identification to monitor and mitigate their effects on aquatic ecosystems [6].

Furthermore, global environmental challenges, including climate change, pose significant threats to aquatic ecosystems. Climate-related shifts in temperature, precipitation, and water chemistry impact species distributions, increase the prevalence of cyanobacterial blooms with toxic effects, disrupt hydrological patterns, and lead to ocean acidification and glacial retreat [7, 8]. These global issues, while vast in their implications, are intricately linked to the microcosmic interactions occurring within water bodies.

The integration of data science, machine learning, and deep learning into hydrobiological research offers an unprecedented opportunity to decipher these complex interactions, predict future trends, and formulate strategies for sustainable management and conservation [9]. The sheer volume, complexity, and multidimensionality of data sourced from aquatic ecosystems have transcended the analytical capabilities of conventional methods. Data science, machine learning, and deep learning offer sophisticated frameworks that can seamlessly process, analyze, and interpret this deluge of information, enabling researchers to glean deeper insights and uncover subtle patterns that were previously elusive. These computational tools not only enhance the precision and accuracy of analyses but also empower researchers to model intricate ecological interactions, forecast environmental shifts, and optimize interventions for aquatic conservation [10]. In a realm as dynamic and intricate as hydrobiology, the fusion of computational prowess with biological inquiry heralds a new era of informed decision-making and robust ecosystem management.

2 ADDRESSING COMMON CHALLENGES IN METAGENOMIC ANALYZES

In recent years, the biotechnological landscape has undergone a huge transformation. Traditionally, DNA sequencing was predominantly executed on either cultivated cells or genetic material derived from a specific organism with known taxonomic classification. That is to say, homology-based approaches for sequence analysis have emerged as a popular solution for taxonomic classification [11]. They are grounded in the principle that sequences sharing a common ancestry will exhibit similarities. These approaches compare a query sequence against a database of known sequences, seeking matches or alignments that indicate a common typology. These methods are characterized by the very high strength and precision, especially when the genome is already cataloged in the database.

However, a significant challenge arises from the vast number of sequences that remain unclassified. Estimates suggest that at most, only 9% of ocean species have been described [12]. Furthermore, the effectiveness of taxonomic classification varies depending on factors like the sample origin, desired taxonomic level, and database specifics. Such limitation became even more evident with the rise of DNA sequence assembly aims at reconstructing the original structure of the DNA in question, by aligning [13] and merging fragments of a DNA sequence.

Sequence assembly, the process of reconstructing full DNA sequences from fragmented reads, relies heavily on advanced computational algorithms and specialized software [14]. While these tools offer precision, technical challenges, such as missing data and genomic intricacies like sequence repeats and heterozygosity, can impede accurate reconstruction. Such oversights can compromise the assembly's integrity [15]. As the demand for computational efficiency and optimal resource utilization grows [16], the integration of data science techniques becomes increasingly crucial, promising improved accuracy and streamlined assembly workflows (Figure 1).
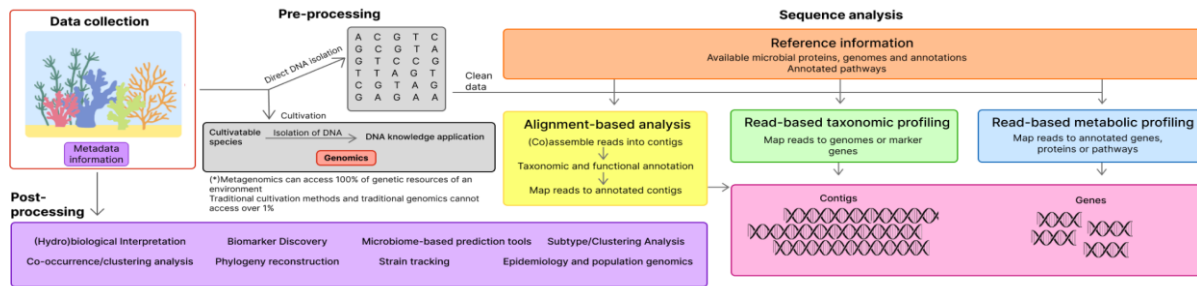
Figure 1: Schematic workflow for metagenomic sequences analysis. The workflow depicts the journey from raw sequence data acquisition, through preprocessing and quality control, to assembly and annotation.

The development of graph algorithms has brought significant advancements to the field, with three primary categories traditionally employed: Overlap/Layout/Consensus (OLC), de Bruijn Graph, and greedy graph methods [13]. In classic computer science terminology, a graph is an abstraction comprising nodes interconnected by edges. Within the overlap graph approach, sequencing reads are depicted as nodes, while their overlaps are represented by edges. This graph can also be equipped with additional attributes to differentiate between the 5′ and 3′ ends of reads, forward and reverse complement sequences, read lengths, overlap lengths, and overlap types (either suffix-to-prefix containment [17]). Researchers at Leiden University have explored the use of overlap graphs for assembling genome sequences from Ciliates found in water bodies. While their findings underscored the potential of this approach for genome assembly, they also suggested refinements to enhance its efficiency and accuracy. One such refinement was the introduction of a 'partial' model, characterized by specific forbidden induced subgraphs. Notably, this model does not have a counterpart for the simple double string rule in graph rules [18]. Furthermore, the team introduced a method to directly construct the reduction graph from its overlap graph, emphasizing the capability to recover structural information seemingly lost in the overlap graph [19] . While these results are promising, further investigations are needed to determine the consistency of this method's effectiveness across various overlap graphs and organism genera.

De Bruijn approach for sequence assembly introduced great advancements in the field and underlied the development of a number of modern sequence assembly approaches [20]. In this approach, the nodes represent all possible fixed-length strings and the edges represent suffix-to-prefix perfect overlaps. One of the important forms of de Bruijn graphs is the K-mer graph (Figure 2a). Its edges represent all the fixed-length overlaps between subsequences that were consecutive in the larger sequence. According to one approach, each K-mer starting at a base corresponds to an edge, with nodes representing overlaps of K-1 bases [21]. In contrast, each K-mer starting at a base is depicted as a node, while edges signify overlaps of K-1 bases [22]. In the realm of WGS assembly, K-mer graphs can depict multiple sequences, with each read represented as a distinct path. When reads perfectly overlap, they share a common path, allowing the implicit detection of these overlaps without the necessity for pair-wise sequence alignment calculations (Figure 2b). While overlap graphs merge paths at longer repeats within a read, K-mer graphs do so at perfect repeats with a length of K or more, given that K is shorter than the read length. This makes K-mer graphs particularly susceptible to sequencing errors and repeats. A singular sequencing error can generate up to K incorrect nodes in the K-mer graph. Consequently, these erroneous nodes might align with other nodes, leading to unintended path convergences [13].

A number of approaches incorporate de Bru graphs to the sequence assembly with positive outcomes. For instance, Velvet stands out as a widely-used de novo assembler, tailored specifically for short-read sequencing data from next-generation sequencing platforms [22]. Thanks to its mechanism, Velvet is capable of managing exceptionally short reads and read pairs, facilitating the construction of meaningful genomes. Its versatility is evident from its widespread application across diverse genomic investigations, often juxtaposed with other assembly tools to gauge its competence [23]. On the other hand, Velvet's single-threaded design restricts its operation to a single processor, potentially curtailing its scalability with expansive datasets that are common in environmental and water body research [24]. That is to say, the application relies on an in-memory representation of the de Bruijn graph, which can be

memory-intensive for larger genomes [25]. Other common de Bruijn graph-based assemblers include, ALLPATHS, HaVec, ABySS, MEGAHIT, SOAPdenovo , and YAGA.

The challenge of handling high-dimensional data is recurrent across various methods. Consequently, several strategies have been developed to address this concern. Among these, DBGPS, which employs the de Bruijn graph combined with a greedy path search, stands out. This approach not only effectively manages high-dimensional data, demonstrated by its successful handling of 6.8 MB of data, but also adeptly addresses issues related to DNA breaks, rearrangements, and indels [26].

Next important step of the analysis of sequenced data is the alignment. Alignment involves matching the sequenced data to a reference genome or other sequences, a process fraught with challenges. Common difficulties [27] include handling mismatches due to genetic variations, dealing with gaps or insertions, and navigating through repetitive regions that can confound traditional alignment algorithms [28]. Metagenomic samples add layers of complexity to the analysis, amplifying the inherent challenges. Due to the vast diversity of microbial species in a single sample, distinguishing between closely related organisms can be problematic [29]. The presence of rare species means reference genomes might be absent, making traditional

alignment methods inadequate [30]. Moreover, horizontal gene transfer events, common in microbial communities, can create chimeric sequences, complicating alignment [31].

There were several approaches that tried to solve the issues of conventional alignment methods. One of those is the greedy x-drop algorithm that performs sequence alignment by making locally optimal choices at each stage with the hope of finding a global optimum. While it is generally quicker than other conventional alignment algorithms, its primary disadvantage is that it doesn't always guarantee an optimal global alignment. However, due to its efficiency, it remains a popular choice in scenarios where approximate alignments are acceptable or when dealing with shorter DNA sequences [32]. Several other attempts existed to incorporate machine learning algorithms to optimise the alignment process.

Recent research underscores the promise of reinforcement learning algorithms in tackling key challenges associated with sequence alignment, demonstrating encouraging results. Specifically, these algorithms have demonstrated superior performance compared to conventional matrix-based tools, such as ClustalW and MAFFT, especially when applied to multiple sequence alignments of several benchmark datasets [33].

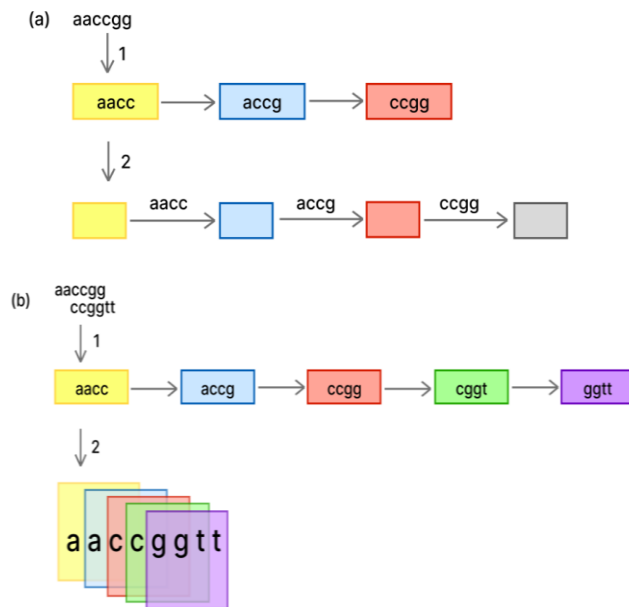Figure 2: De Bruijn approach: a) a single read is mapped using two K-mer graph models to highlight the node-edge relationship for overlapping K-mers. Ideal for depicting simple paths due to minimal repeats. b) Pair-wise Overlap via K-mer Graph demonstrates error-free overlap of two reads and their unified representation in a K=4 K-mer graph, facilitating straightforward consensus sequence reconstruction through a simple path.

However, a prevalent constraint among these methods is their tendency to align sequences of a particular length, a limitation stemming from fixed input and network dimensions. To surmount this obstacle, Song and Cho (2021) integrated the DQNalign approach with the x-drop algorithm [34]. This amalgamation enhanced alignment performance, reduced complexity, and minimized computational time. Notably, when employed in the alignment of the E. coli genome, this method's accuracy paralleled that of conventional techniques. Another notable reinforcement learning-based solution is EdgeAlign, strategically tailored for the effective alignment of DNA sequences on edge devices. This innovative approach showcases the integration of a highly compact yet robust deep Q-network (DQN) agent, akin to the previously described method, ensuring a uniform hardware resource footprint regardless of sequence lengths [35].

In summary, recent advancements in genomics have ushered in a new era of metagenomics and DNA sequence analyses, challenging traditional methodologies. The integration of data science approaches, exemplified by use of machine learning methods, promises to enhance sequence alignment efficiency and accuracy. Altogether, it will speed-up advances in the hydrobiology due to high processing throughput of these systems. The automation of time-consuming and routine activities will contribute to high-quality data-driven approaches to tackle urgent climate change issues and its impact on water ecosystems and resources.

3 COMPUTATIONAL ADVANCES IN THE AQUATIC TAXONOMY

The taxonomic classification of organisms within water bodies plays a crucial role in ecological research and environmental monitoring [36]. Over time, this classification has progressed significantly, from traditional homology-based methods to the adoption of advanced deep learning techniques. Understanding the composition and dynamics of aquatic ecosystems is essential for assessing water quality, tracking changes in biodiversity, and studying the impact of environmental factors, such as climate change and pollution, on aquatic habitats [37].

From a technical standpoint, most classification tools hinge on the similar methodologies as described above, namely local alignments, k-mers, Burrow–Wheeler transformations, minimizers, or hybrid methodologies [38]. For instance, the discriminative k-mers method has been applied to classify metagenomic sequences, demonstrating remarkable accuracy, especially for short metagenomic reads. Notably, this algorithm exhibited impressive speed, capable of processing up to 32 million metagenomic short reads per minute [38]. Its performance aligns with other k-mer-based tools known for their accuracy, speed, and minimal memory requirements [39]. However, it's important to note that these methods primarily focus on the genus and/or species levels, with limited evaluation at higher taxonomic ranks. Overall, k-mer-based approaches excel in terms of speed, but their recall and precision may exhibit variability. On the contrary, local alignments, while highly precise, can be computationally intensive and may yield restricted recall rates.

To address these challenges, deep neural network (DNN) approaches have emerged as a promising solution. These deep learning techniques, rather than merely relying on database similarities, model intricate relationships between DNA sequences and their taxonomic classes. An example of such an algorithm is a recently developed model called BERTax, that is based on the state-of-the-art natural language processing architecture BERT (bidirectional encoder representations from transformers) updated with additional layers. BERTax is able to classify the sequences in question on 3 taxonomic levels (superkingdom, phylum and genus). A notable advantage of such methods compared to the conventional approaches is that it does not focus on local similarity, but on the overall image and, therefore, is not subjected to the common restrictions of comparable tools. As a result of such novelty, the tool was able to deal with novel organisms, not existing in the initial databases, which still remained a challenge for similar tools [40].

DNA barcoding, another widely-adopted genomic approach for taxonomic classification, owes its popularity to several key attributes [41]. This method utilizes a standardized region of DNA, enabling efficient species identification, insights into molecular lineage, and applications in conservation biology [42]. Traditionally, DNA barcoding relied on similarity-, character-, and tree-based methodologies. However, as computational capabilities have advanced, the integration of machine and deep learning techniques has emerged as a transformative approach. The evaluation of various algorithms using both empirical and synthetic datasets has demonstrated the remarkable efficacy of the k-nearest neighbors algorithm in addressing these tasks, outperforming other techniques such as Naive Bayes, Random Tree, and SVM [43]. While these algorithms

exhibit high performance on synthetic datasets, real-world data introduces additional challenges, including the high dimensionality of DNA barcode sequences, limited interspecific sequence variation, and numerical constraints due to the diversity of species. Consequently, a recent study has employed a sophisticated deep learning model to tackle these complexities in classifying fish from different families. This novel approach combines an Elastic Net-Stacked Autoencoder (EN-SAE) with Kernel Density Estimation (KDE), effectively mitigating the aforementioned challenges and enhancing classification accuracy [44].

Computer vision techniques have also been effectively employed in the taxonomic detection and classification of aquatic organisms. A noteworthy study utilised the Faster Region-based Convolutional Neural Network (R-CNN) extended with a supplementary classification branch. Remarkably, the model posted mean average precision scores of 74.64% at the genus level and 81.17% at the class level source, however the dataset's uneven distribution, with varying instance percentages of specific genera and biological classes, influences the model's efficacy. This limitation underscores the impending need for well-balanced, high-quality datasets for the algorithm training, as this step directly influences the model's performance [45]. Another successful application of taxonomy in aquatic habitats is exemplified by the work of Memmolo et al. (2020), where they conducted algorithm training using diatom test slides. Leveraging a substantial dataset comprising 8,731,800 elements, an average of 174.636 augmented phase-contrast images were generated from a single hologram record. This extensive dataset contributed to training a model that achieved an impressive classification accuracy of 98% [46]. These findings underscore the practicality and cost-effectiveness of utilizing species test slides as a valuable approach for training classification models in taxonomic studies.

In conclusion, the adoption of advanced computational methods, including deep learning and DNA barcoding, holds significant promise for enhancing aquatic taxonomic classification. These approaches offer notable advantages, such as improved accuracy and scalability, enabling more precise species identification and ecological analysis. However, challenges related to data quality, model interpretability, and the need for robust training datasets remain. Despite these obstacles, the potential benefits for biodiversity assessment, environmental monitoring, and conservation efforts are substantial. Further research and refinement of these methodologies are essential to unlock their full potential in advancing our understanding of aquatic ecosystems.

4 CONCLUSIONS AND FUTURE

The intersection of machine learning and hydrobiological research has opened doors to a plethora of possibilities in understanding aquatic ecosystems. These computational tools offer a sophisticated framework for handling the increasing volume and complexity of aquatic data, making previously elusive patterns more discernible. The taxonomic identification of organisms, a pivotal aspect of hydrobiological research, has seen remarkable advancements with the integration of deep learning techniques, improving precision and enabling detection of novel organisms. Furthermore, in the domain of metagenomic sequences analysis, machine learning has addressed challenges related to sequence assembly, alignment, and taxonomy, paving the way for more efficient and accurate methods.

However, despite the significant strides made, challenges persist. The quality of data, model interpretability, and the need for robust training datasets are among the hurdles faced. Therefore, there is a need for developing robust data integration and augmentation strategies, custom-tailored specifically for environmental and genomic studies. Most of the studies utilize common practices, such as image rotation and cropping, synthetic data generation, and manual modification. To overcome the data scarcity, usage of data fusion techniques can become helpful, however such approaches are not precisely studied in terms of hydrobiological research. Additionally, utilization of data imputation might fill gaps in the datasets and provide a precise view of the explored hydrobiological processes. Furthermore, while computational models offer enhanced accuracy and efficiency, their real-world application necessitates a comprehensive understanding of local aquatic nuances and characteristics.

However, the main challenges and opportunities for future studies in this area include high complexity and non-linearity of the data, noise, lack of covariates, and compositionality. This particular problem can be solved with a large model well-versed in the specifics of the domain area, similarly the way large language models are trained.

In essence, the fusion of computational methodologies with traditional hydrobiological techniques holds immense potential. As research continues and technologies evolve, the synergy of

these domains will undoubtedly lead to more informed strategies for the preservation and understanding of aquatic ecosystems. The emerging developments of data science in hydrobiology are mainly in their early stages and are becoming a powerful assistance tool for regular research activities. Their applications are vital for water quality monitoring, environment and climate change processes modeling. We expect the number of data science models and tools to grow in the incoming years with improvements in predictive performance and data quality. It is worth preparing solid data resources to make such studies possible and consolidate the available knowledge into a high-quality model.

The leveraged data and machine learning technologies are offering a significant boost to the development of specific areas of hydrobiology and aquatic environment modeling. The reviewed studies explore usage models and algorithms with a varying level of complexity depending on the input requirements. We consider that despite the exponential growth and usage of deep learning technologies, classical linear- and tree-based machine learning algorithms provide enough efficiency, flexibility, and accuracy to assist with modeling of hydrobiological properties and characteristics.

We observe a rising number of deep learning applications for unstructured data classification. It is worth mentioning that such an approach is very robust when dealing with image, sensor or text data formats. Nevertheless, the usage of deep learning networks requires large amounts of data in order to be efficient, which is the primary issue of the reviewed studies. Data collection for these research areas is still a major challenge due to a number of factors: sensor design and technologies, their deployment and high costs associated with them, a need for appropriate equipment and computational resources. From a hydrobiological perspective the following difficulties arise during data assemblage: vast diversity of unclassified microbes; heterogeneity, repeats, and duplications; quick genetic changes complicate the clarity of taxonomic categorization; external impact (environment or reagents contamination).

ACKNOWLEDGMENTS

REFERENCES

[1] D. B. Oerther, L. Gautham, and N. Folbre, "Environmental engineering as care for human welfare and planetary health," Journal of Environmental Engineering, vol. 148, no. 6, Jun. 2022, doi: 10.1061/(asce)ee.1943-7870.0002013.

[2] D. Y. Kwon, J. Kim, S. Park, and S. Hong, "Advancements of remote data acquisition and processing in unmanned vehicle technologies for water quality monitoring: An extensive review," Chemosphere, vol. 343, p. 140198, Dec. 2023, doi: 10.1016/j.chemosphere.2023.140198.

[3] K. I. Suh, J. M. Hwang, Y. J. Bae, and J. H. Kang, "Comprehensive DNA barcodes for species identification and discovery of cryptic diversity in mayfly larvae from South Korea: Implications for freshwater ecosystem biomonitoring," Entomological Research, vol. 49, no. 1, pp. 46-54, Jan. 2019, doi: 10.1111/1748-5967.12334.

[4] V. Gomez-Alvarez, H. Liu, J. G. Pressman, and D. G. Wahman, "Metagenomic Profile of Microbial Communities in a Drinking Water Storage Tank Sediment after Sequential Exposure to Monochloramine, Free Chlorine, and Monochloramine," ACS ES&T Water, vol. 1, no. 5, pp. 1283-1294, Mar. 2021, doi: 10.1021/acsestwater.1c00016.

[5] C. O. Coleman and A. Radulovici, "Challenges for the future of taxonomy: talents, databases and knowledge growth," Megataxa, vol. 1, no. 1, Jan. 2020, doi: 10.11646/megataxa.1.1.5.

[6] A. C. Staudt, et al., "The added complications of climate change: understanding and managing biodiversity and ecosystems," Frontiers in Ecology and the Environment, vol. 11, no. 9, pp. 494-501, Nov. 2013, doi: 10.1890/120275.

[7] H. Falfushynska, N. Kasianchuk, E. Siemens, E. Henao, and P. Rzymski, "A review of common cyanotoxins and their effects on fish," Toxics, vol. 11, no. 2, p. 118, Jan. 2023, doi: 10.3390/toxics11020118.

[8] R. C. Allen, B. E. Rittmann, and R. Curtiss, "Axenic Biofilm Formation and Aggregation by Synechocystis sp. Strain PCC 6803 Are Induced by Changes in Nutrient Concentration and Require Cell Surface Structures," Applied and Environmental Microbiology, vol. 85, no. 7, Apr. 2019, doi: 10.1128/aem.02192-18.

[9] K. Malde, N. O. Handegard, L. Eikvil, and A.-B. Salberg, "Machine intelligence and the data-driven future of marine science," Ices Journal of Marine Science, vol. 77, no. 4, pp. 1274-1285, Apr. 2019, doi: 10.1093/icesjms/fsz057.

[10] R. H. Medina, et al., "Machine learning and deep learning applications in microbiome research," ISME Communications, vol. 2, no. 1, Oct. 2022, doi: 10.1038/s43705-022-00182-9.

[11] R. Harr, P. Hagblom, and P. Gustafsson, "Two-dimensional graphic analysis of DNA sequence homologies," Nucleic Acids Research, vol. 10, no. 1, pp. 365-374, Jan. 1982, doi: 10.1093/nar/10.1.365.

[12] C. Mora, D. P. Tittensor, S. M. Adl, A. G. B. Simpson, and B. Worm, "How many species are there on Earth and in the ocean?," PLOS Biology, vol. 9,

no. 8, p. e1001127, Aug. 2011, doi: 10.1371/journal.pbio.1001127.

[13] J. R. Miller, S. Koren, and G. Sutton, "Assembly algorithms for next-generation sequencing data," Genomics, vol. 95, no. 6, pp. 315-327, Jun. 2010, doi: 10.1016/j.ygeno.2010.03.001.

[14] A. M. Phillippy, "New advances in sequence assembly," Genome Res., vol. 27, no. 5, pp. xi-xiii, May 2017, doi: 10.1101/gr.223057.117.

[15] K. L. Korunes and K. Samuk, "pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data," Mol. Ecol. Resour., vol. 21, no. 4, pp. 1359-1368, May 2021, doi: 10.1111/1755-0998.13326.

[16] J. Mbatchou, et al., "Computationally efficient whole-genome regression for quantitative and binary traits," Nat. Genet., vol. 53, no. 7, pp. 1097-1103, Jul. 2021, doi: 10.1038/s41588-021-00870-7.

[17] E. Csuhaj-Varjú, I. Petre, and G. Vaszil, "Self-assembly of strings and languages," Theor. Comput. Sci., vol. 374, no. 1, pp. 74-81, Apr. 2007, doi: 10.1016/j.tcs.2006.12.004.

[18] R. Brijder and H. J. Hoogeboom, "Combining overlap and containment for gene assembly in ciliates," Theor. Comput. Sci., vol. 411, no. 6, pp. 897-905, Feb. 2010, doi: 10.1016/j.tcs.2009.07.047.

[19] R. Brijder, H. J. Hoogeboom, and G. Rozenberg, "REDUCTION GRAPHS FROM OVERLAP GRAPHS FOR GENE ASSEMBLY IN CILIATES," Internat. J. Found. Comput. Sci., vol. 20, no. 02, pp. 271-291, Apr. 2009, doi: 10.1142/S0129054109006553.

[20] B. Ekim, B. Berger, and R. Chikhi, "Minimizer-space de Bruijn graphs: Whole-genome assembly of long reads in minutes on a personal computer," vol. 12, no. 10, pp. 958-968.e6, Oct. 2021, doi: 10.1016/j.cels.2021.08.009.

[21] R. M. Idury and M. S. Waterman, "A new algorithm for DNA sequence assembly," Journal of Computational Biology, vol. 2, no. 2, pp. 291-306, Jan. 1995, doi: 10.1089/cmb.1995.2.291.

[22] D. R. Zerbino and E. Birney, "Velvet: algorithms for de novo short read assembly using de Bruijn graphs," Genome Res., vol. 18, no. 5, pp. 821-829, May 2008, doi: 10.1101/gr.074492.107.

[23] Y. Endo, F. Toyama, C. Chiba, H. Mori, and K. Shoji, "Memory Efficient de novo Assembly Algorithm using Disk Streaming of K-mers," scitepress.org, Jan. 2016, doi: 10.5220/0005798302660271.

[24] E. Costa and G. Silva, "The velvet assembler using OpenACC directives," EPiC Series in Computing, May 2023, doi: 10.29007/pzbt.

[25] R. Chikhi and G. Rizk, "Space-efficient and exact de Bruijn graph representation based on a Bloom filter," Algorithms Mol. Biol., vol. 8, no. 1, p. 22, Sep. 2013, doi: 10.1186/1748-7188-8-22.

[26] L. Song, et al., "Robust data storage in DNA by de Bruijn graph-based de novo strand assembly," Nature Communications, vol. 13, no. 1, Sep. 2022, doi: 10.1038/s41467-022-33046-w.

[27] J. Thompson and O. Poch, "New challenges and strategies for multiple sequence alignment in the Proteomics Era," in Humana Press eBooks, 2005, pp. 475-492. doi: 10.1385/1-59259-890-0:475.

[28] T. N. Petersen, et al., "MGmapper: Reference based mapping and taxonomy annotation of metagenomics sequence reads," PLOS ONE, vol. 12, no. 5, p. e0176469, May 2017, doi: 10.1371/journal.pone.0176469.

[29] T. Wolf, P. Kämmer, S. Brunke, and J. Linde, "Two's company: studying interspecies relationships with dual RNA-seq," Current Opinion in Microbiology, vol. 42, pp. 7-12, Apr. 2018, doi: 10.1016/j.mib.2017.09.001.

[30] C. Anyansi, T. J. Straub, A. L. Manson, A. M. Earl, and T. Abeel, "Computational methods for Strain-Level microbial detection in colony and metagenome sequencing data," Frontiers in Microbiology, vol. 11, Aug. 2020, doi: 10.3389/fmicb.2020.01925.

[31] I. L. Brito, "Examining horizontal gene transfer in microbial communities," Nature Reviews Microbiology, vol. 19, no. 7, pp. 442-453, Apr. 2021, doi: 10.1038/s41579-021-00534-7.

[32] R. A. DeVore, G. Petrova, and P. Wojtaszczyk, "Greedy algorithms for reduced bases in banach spaces," Constructive Approximation, vol. 37, no. 3, pp. 455-466, Feb. 2013, doi: 10.1007/s00365-013-9186-2.

[33] R. Jafari, M. M. Javidi, and M. K. Rafsanjani, "Using deep reinforcement learning approach for solving the multiple sequence alignment problem," SN Applied Sciences, vol. 1, no. 6, May 2019, doi: 10.1007/s42452-019-0611-4.

[34] Y.-J. Song and D.-H. Cho, "Local alignment of DNA sequence based on deep reinforcement learning," IEEE Open Journal of Engineering in Medicine and Biology, vol. 2, pp. 170-178, Jan. 2021, doi: 10.1109/ojemb.2021.3076156.

[35] A. Lall and S. Tallur, "Deep reinforcement learning-based pairwise DNA sequence alignment method compatible with embedded edge devices," Scientific Reports, vol. 13, no. 1, Feb. 2023, doi: 10.1038/s41598-023-29277-6.

[36] M. Muthulakshmi, "A Novel Feature Extraction from Genome Sequences For Taxonomic Classification Of Living Organisms," Turkish Journal of Computer and Mathematics Education, Apr. 2021, doi: 10.17762/turcomat.v12i2.1364.

[37] F. J. Wrona, T. D. Prowse, J. Reist, and W. F. Vincent, "Climate change effects on aquatic biota, ecosystem structure and function," ResearchGate, Dec. 2006, doi: 10.1579/0044-7447(2006)35.

[38] V.-K. Bui and C. Wei, "CDKAM: a taxonomic classification tool using discriminative k-mers and approximate matching strategies," BMC Bioinformatics, vol. 21, no. 1, Oct. 2020, doi: 10.1186/s12859-020-03777-y.

[39] R. Ounit, S. Wanamaker, T. J. Close, and S. Lonardi, "CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers," BMC Genomics, vol. 16, no. 1, Mar. 2015, doi: 10.1186/s12864-015-1419-2.

[40] F. Mock, F. Kretschmer, A. Kriese, S. Böcker, and M. Marz, "Taxonomic classification of DNA sequences beyond sequence similarity using deep neural networks," Proceedings of the National Academy of Sciences of the United States of America, vol. 119, no. 35, Aug. 2022, doi: 10.1073/pnas.2122636119.

[41] B. H. Mendoza-Ramírez, L. Páiz-Medina, T. Salvatierra-Suárez, N. Del Socorro Hernández, and J. A. Huete-Pérez, "A survey of aquatic macroinvertebrates in a river from the dry corridor of Nicaragua using biological indices and DNA barcoding," Ecology and Evolution, vol. 12, no. 11, Nov. 2022, doi: 10.1002/ece3.9487.

[42] H.-T. Vu and L. Le, "Bioinformatics Analysis on DNA Barcode Sequences for Species identification: A review," Annual Research & Review in Biology, pp. 1-12, Dec. 2019, doi: 10.9734/arrb/2019/v34i130142.

[43] M. Emu and S. Sakib, "Species Identification using DNA Barcode Sequences through Supervised Learning Methods," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Feb. 2019, doi: 10.1109/ecace.2019.8679166.

[44] L. Jin, J. Yu, X. Yuan, and X. Du, "Fish Classification Using DNA Barcode Sequences through Deep Learning Method," Symmetry, vol. 13, no. 9, p. 1599, Aug. 2021, doi: 10.3390/sym13091599.

[45] P. Qian, et al., "Multi-Target Deep Learning for Algal Detection and Classification," In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2020, doi: 10.1109/embc44109.2020.9176204.

[46] P. Memmolo, et al., "Learning Diatoms Classification from a Dry Test Slide by Holographic Microscopy," Sensors, vol. 20, no. 21, p. 6353, Nov. 2020, doi: 10.3390/s20216353.

Transient Phenomena in Information Technology for Branching Processes with an Infinite Set of Types

Sergii Degtyar[1], Oleh Kopiika [2] and Yurii Shusharin[1]

[1]*Kyiv National Economic University named after Vadym Hetman, Beresteiskyi Avenue 54/1, 03057 Kyiv, Ukraine*
[2]*Institute of Telecommunications and Global Information Space of NASU, Chokolivskiy Blvd. 25, 03186 Kyiv, Ukraine*
svdegtyar@gmail.com, okopiyka@gmail.com, u.shusharin@gmail.com

Abstract: Branching processes as a mathematical concept has applications in various fields, including information technology. In information technology, branching processes can be used to model and analyze various scenarios, such as the propagation of data or information in a network, the growth of computer viruses, the spread of software bugs, and more. Branching processes are particularly useful for understanding the dynamics of systems where events can lead to multiple new events in a probabilistic manner. Overall, branching processes provide a valuable mathematical framework for modeling and analyzing various aspects of information technology, helping to make informed decisions and optimize IT systems and networks. We have studied transient phenomena for branching processes with an infinite number of types close to critical. The analytical apparatus for this study is Markov renewal theorems. Branched processes were used to evaluate the performance of IT systems and predict their behavior under different conditions. This is important for capacity planning and resource allocation.

1 INTRODUCTION

Branching processes are one of the most interesting sections of probability theory. The theory of branching processes has now become a much branched field of probability theory and a powerful research tool in various areas of mathematics, such as the theory of algorithms, queuing theory, random mapping theory, the theory of leakage, as well as in many branches of other sciences, which include, in particular, physics, chemistry, biology and information technology.

The theory of branching random processes is an important method in the theory of Markov queuing models. Similar processes arise, in particular, in the description of queuing systems (for example, in multiprocessor or distributed computer processing of data). This apparatus is especially important in the theory of reliability, where it is used to describe failures in complex equipment (failure of one element gives rise to failures of other elements).

Here are some ways branching processes can be applied in IT:

1) Network Propagation: In the context of computer networks and communication systems, branching processes can model how data packets or information propagate through the network. This is useful for understanding network congestion, data transfer rates, and the spread of information in social networks [1-8].

2) Virus and Malware Propagation: Branching processes can be used to model the spread of computer viruses, malware, and other malicious software within a network. This can help in analyzing the potential impact of a malware outbreak and developing strategies to mitigate it [9-10].

3) Software Development and Bugs: In software development, branching processes can be applied to model the occurrence and propagation of software bugs and issues. This can aid in understanding the factors that contribute to software quality and reliability [11-13].

4) Fault Tolerance and Redundancy: In designing fault-tolerant systems, branching processes can be used to model the reliability of components within a system. This helps in making informed

decisions about redundancy and backup strategies [14-18].

5) Queuing Systems: In IT systems, queuing models often involve branching processes to analyze the flow of requests or tasks through a system, such as in call centers, web servers, and database management [19-21].

6) Social Media Engagement: In the context of social media platforms and online communities, branching processes can be used to model the spread of content (e.g., viral videos, news articles) and user engagement patterns, helping to optimize content distribution strategies.

7) Performance Analysis: Branching processes can be used to assess the performance of IT systems and predict their behavior under different conditions. This is valuable for capacity planning and resource allocation [22-27].

2 TRANSIENT PHENOMENA OF MARKOV RENEWAL THEORY

The basic statements of the classical renewal theory can be extended to the so-called Markov renewal equation

$$f(x,t) = g(x,t) +$$

$$+ \int_E \int_0^t G(x, dy \times du) f(y, t-u), t \geq 0, x \in E.$$

where E is a given phase space, $G(x, dy \times du)$ is so-called semi-Markov kernel, $g(x,t)$ is a given function of $x \in E$, and $t \geq 0$, and $f(x,t)$ is the function to be found. Its solution is the convolution

$$f(x,t) = U * g(x,t) =$$

$$= \int_E \int_0^t U(x, dy \times du) g(y, t-u), t \geq 0, x \in E.$$

where $U(x, dy \times du)$ is the potential of the semi-homogeneous kernel $G(x, dy \times du)$.

Generally, the renewal theory has wide range of applications in mathematical practice. Markov renewal theorems are an analytical tool for studying the limiting behavior of Markov and related processes, including semi-Markov and regenerative processes.

For example we can consider Crump-Mode-Jagers branching process [31] with an arbitrary set of types (say E). Denote by $M_t(x,A)$ the conditional mean number of particles at moment $t \geq 0$ whose types belong to set $A \subset E$ under the condition that there was one new-born particle of type $x \in E$ at the initial moment $t_0 = 0$.

It is well-known that $M_t(x,A)$ as function of $x \in E$ and $t \geq 0$ satisfies the

$$M_t(x,A) = L_t(x,A) +$$

$$+ \int_E \int_0^t K(x, dy \times du) M_{t-u}(y,A),$$

where, roughly speaking, $K(x, dy \times du)$ is equal to the conditional mean number (under the same condition) of particles of type y produced by the initial particle during an infinitely small interval $[u, u+du)$, and $L_t(x,A)$ is equal to the conditional probability (under the above condition) that the type of the still existing initial particle belongs to the set A.

W. Feller introduced the very important notion of direct Riemann integrability.

Namely, a Borel function $g(t), t \geq 0$, is called directly Riemann-integrable if

$$\sum_{k=0}^{\infty} \sup_{k \leq t \leq k+1} |g(t)| < \infty, \quad (1)$$

and

$$\delta \sum_{k=0}^{\infty} \left\{ \sup_{k\delta \leq t \leq k\delta+\delta} g(t) - \inf_{k\delta \leq t \leq k\delta+\delta} g(t) \right\} \underset{\delta \to 0}{\longrightarrow} 0. \quad (2)$$

Under conditions (1) and (2) the function $g(t)$ is absolutely integrable on $[0, \infty)$ and

$$\int_0^{\infty} g(t)dt = \lim_{\delta \to 0} \left[\delta \sum_{k=1}^{\infty} g(t_k) \right],$$

where $t_k \in [k\delta, k\delta + \delta]$, in contrast to the usual definition of the improper Riemann integral as a limit of integrals over finite intervals.

That is why such a function $g(t)$ is called directly Riemann-integrable.

In the theory of branching processes in the case of a finite phase space, the Markov renewal equations are often considered

$$f_i(t) = g_i(t) + \sum_{j=1}^{d} \int_0^t f_j(t-u) \, dM_{ij}(u), \quad (3)$$

where the functions $M_{ij}(u)$ do not decrease, have bounded variation, and $M_{ij}(0) = 0$.

The asymptotic properties of the solution to this equation are essentially determined by the maximum eigenvalue λ (the Perron root) of matrix

$$M(\infty) = \left\| M_{ij}(\infty) \right\|_{i,j=1}^{d}.$$

Equations (3) is called supercritical if $\lambda > 1$, critical, if $\lambda = 1$, and subcritical, if $\lambda < 1$. Critical equations with indecomposable matrix $M(\infty)$ simply reduce to the equations of Markov renewal. If the Perron root of an indecomposable matrix $M(\infty)$ is equal to one, then it has left and right invariant vectors $v = (v_1, v_2, \ldots, v_d)$ and $u = (u_1, u_2, \ldots, u_d)$ respectively with positive coordinates, i.e.

$$\sum_j M_{ij}(\infty) u_j = u_i, \quad \sum_i v_i M_{ij}(\infty) = v_j.$$

In the case when the criticality parameter λ of (3) is close to one, but possibly not equal to it, then the problem of transient phenomena arises. One of the exact formulations of this problem is as follows. Let $M^{(n)}(t) = \left\|M_{ij}^{(n)}(t)\right\|_{i,j=1}^d, n = 1,2,\ldots$, is a sequence $d \times d$ matrices, whose elements are non-decreasing functions of bounded variation with $M_{ij}^{(n)}(0) = 0$. Let also for all $i, j = 1,2,\ldots,d$ the sequence $M_{ij}^{(n)}(t), n = 1,2,\ldots$, converges to $M_{ij}(t)$ at the points of continuity of the limit function, and the matrix $M(\infty) = \left\|M_{ij}(\infty)\right\|_{i,j=1}^d$ is indecomposable and its Perron root is equal to one. This assumption allows, without loss of generality, to assume that the matrices $M^{(n)}(\infty)$ are also not decomposable, and that their Perron roots λ_n converge as $n \to \infty$ to one, and the right and left eigenvectors $u^{(n)}$ and $v^{(n)}$ converge respectively to the right-left invariant vectors $u = (u_1, u_2, \ldots, u_d)$ and $v = (v_1, v_2, \ldots, v_d)$ of the matrix $M(\infty)$.

Let $f_i^{(n)}(t)$ at each $n = 1,2,\ldots$ be a solution of the

$$f_i^{(n)}(t) = g_i^{(n)}(t) + \sum_{j=1}^d \int_0^t f_j^{(n)}(t-u)\, dM_{ij}^{(n)}(u),$$

functions sequence $g_i^{(n)}(t)$, $n = 1,2,\ldots$, is uniformly directly Riemann integrable on $[0, \infty)$. Then if

$$\sup_{n \geq 1} \int_t^\infty s\, dM_{ij}^{(n)}(s) \xrightarrow[t \to \infty]{} 0,$$

and the matrix $M(t)$ is nonlattice, then

$$f_i^{(n)}(t) - \frac{u_i}{m} e^{c/m} \sum_j v_j \int_0^\infty g_j^{(n)}(u)\, du \to 0,$$

as $t \to \infty, n \to \infty, t(\lambda_n - 1) \to c$, where

$$m = \sum_{ij} v_i \int_0^\infty t\, dM_{ij}(t) u_j, \quad \sum_j v_j u_j = 1.$$

In the infinite-dimensional case, let $f_\varepsilon(x,t)$ at every $\varepsilon > 0$ be a solution of the Markov renewal equation

$$f_\varepsilon(x,t) = g_\varepsilon(x,t) +$$

$$+ \int_E \int_0^t G_\varepsilon(x, dy \times du) f_\varepsilon(y, t-u), t \geq 0, x \in E.$$

Transient phenomena for the solution of the Markov renewal equation

$$f_\varepsilon(x,t) = U_\varepsilon * g_\varepsilon(x,t) =$$

$$= \int_E \int_0^t U_\varepsilon(x, dy \times du) g_\varepsilon(y, t-u), t \geq 0, x \in E,$$

as $t \to \infty, \varepsilon \to 0, t(1 - \lambda_\varepsilon)/m \to c$, where λ_ε is the Perron root of the basis $G_\varepsilon(x, dy)$ of the kernel $G_\varepsilon(x, dy \times dt)$,

$$m = \int_E \int_E \int_0^\infty l(dx) G(x, dy \times dt) h(y) t,$$

$$\int_E l(dx) h(x) = 1,$$

h and l are eigenfunction and eigenmeasure respectively of the basis of the limit semi-homogeneous kernel $G(x, dy \times dt)$, were investigate [28].

We denote $\gamma_\varepsilon = (1 - \lambda_\varepsilon)/m$. In [28] it was proved, if
1) the basis $G(x, dy)$ of the kernel $G(x, dy \times dt)$, that is $G(x, dy) = G(x, dy \times [0, \infty))$, is conservative [32], and its Perron root equals 1,
2) there exists a Borel function $g(t)$ such that

$$\int_0^\infty \left| g(t) - \int_E l(dx) g_\varepsilon(x,t) \right| dt \xrightarrow[\varepsilon \to 0]{} 0,$$

then

$$\lim_{\substack{\varepsilon \to 0 \\ t \to \infty \\ \gamma_\varepsilon t \to c}} U_\varepsilon * g_\varepsilon(x,t) = e^{-c} \frac{h(x)}{m} \int_0^\infty g(s)\, ds,$$

uniformly in $x \in E$.

The another polar case is the degeneracy of the basis of the limit kernel. The asymptotics of the solution of the Markov renewal equation when the

basis $G_\varepsilon(x, dy) = G_\varepsilon(x, dy \times [0, \infty))$ of the kernel $G_\varepsilon(x, dy \times dt)$ close to the singular kernel $I(x, dy)$ on a given measurable phase space $(E, \mathfrak{B})$ was studied in [29]. The main result of that study was formulated in the form of a theorems.

The formal definition of branching processes with an arbitrary number of types of particles, the transformations of which may depend on their age, is rather cumbersome. By contrast, a descriptive description is simple and short. Any such branching process is associated with the following evolving population, consisting of particles of several types: each of the particles existing at a given time, regardless of its origin and the presence of other particles, after the expiration of its existence, turns into a certain (possibly empty) set of newborn particles. The progeny of a particle depends only on its type and the age at which the transformation took place.

Critical processes have the most interesting asymptotic properties. However, the question of the criticality of the actually observed branching process is not simple and it is not always possible to give an unambiguous answer to it. From this point of view, it is very important to study the asymptotic of branching processes, when over time the criticality parameter (in this case, the Perron root) tends to one. The resulting phenomena are called transient.

This article studies transient phenomena for branching processes with an infinite number of types close to critical. The analytical apparatus for this study is Markov renewal theorems. The asymptotic properties of the solution of the Markov renewal equation were studied in [32].

3 A BRANCHING PROCESS WITH AN INFINITE SET OF TYPES

First, we describe a model for the evolution of a population with an infinite number of types.

Let be E - abstract set, which we will call the set of types.

Suppose that on the set E is distinguished a σ-algebra of its subsets $\mathfrak{B}$ which contains all one-point sets and, moreover, is generated by a countable number of its elements.

A population is considered, which consists of a certain number of particles, for each of which a certain type is assigned, that is, an element of the set E. The law of population evolution is as follows. A newborn particle of type x, regardless of the

presence of other particles and the previous history of the development of the population, lives a random time $\tau(x) > 0$, at the end of which it turns into some (possibly empty) set of newborn particles of different types.

Let us denote $(\Omega, M, \mathbb{P})$ - the basic probability space; $\zeta_t(x)$ - is the type of particle at the end of t units of its lifetime, $\zeta_0(x) = x, 0 \le t < \tau(x)$; $\eta(x, S)$ - is the number of immediate descendants with types from the set $S \in \mathfrak{B}$ of one particle that had type x, $\xi_t(x, S, v)$ - is the number of particles formed during time t from one newborn particle of the type x whose age is at least v and whose types belong to the set $S \in \mathfrak{B}$.

Suppose that all introduced characteristics depend measurably on the set of variables $x \in E, \ t \ge 0, \omega \in \Omega$.

Let us fix the number $v > 0$, bounded non-negative $\mathfrak{B}$ - measurable function φ and put

$$f(x, t) = A(x, t) = \mathbb{E} \int_E \xi_t(x, dy, v)\varphi(y),$$
$$g(x, t) = 0 \ at \ 0 \le t < v,$$

$$g(x, t) = \mathbb{E}\left[\varphi(\zeta_t(x))I_{\{t < \tau(x)\}}\right] \ at \ t \ge v,$$

$$G(x, dy \times du) = \mathbb{E}\left[\eta(x, dy)I_{\{\tau(x) \in du\}}\right].$$

Here and in what follows, the symbol $\mathbb{P}$ is denoted the main probability measure and the symbol $\mathbb{E}$ denoted corresponding mathematical expectation.

By slightly modifying the thinking from [30], we can show that the function $A(x, t)$ satisfies the Markov renewal type equation

$$A(x, t) = g(x, t) + \int_E \int_0^t G(x, dy \times du) A(y, t - u).$$

According to the total probability formula, we have

$$A(x, t) = \mathbb{E}\left[\varphi(\zeta_t(x))I_{\{t < \tau(x)\}}\right] +$$

$$+ \int_0^t \mathbb{E}\left[\int_E \xi_t(x, dz, v)\varphi(z)I_{\{\tau(x) \in du\}}\right].$$

Further at $u < t$ we have

$$\mathbb{E}\left[\int_E \xi_t(x, dz, v)\varphi(z)I_{\{\tau(x) \in du\}}\right] =$$

$$= \mathbb{E}\left[\int_E \xi_{\tau(x)+t-u}(x,dz,v)\varphi(z)I_{\{\tau(x)\in du\}}\right] =$$

$$= \int_E \mathbb{E}\left[\eta(x,dy)I_{\{\tau(x)\in du\}}\right] \mathbb{E}\int_E \xi_{t-u}(y,dz,v)\varphi(z)$$

$$= \int_E G(x,dy\times du)A(y,t-u).$$

That's why

$$A(x,t) = \int_E \int_0^t U(x,dy\times du)\,g(y,t-u), \quad (4)$$

where

$$U(x,dy\times du) = \sum_{k=1}^{\infty} G^{k*}(x,dy\times du)$$

is recovery kernel that matches the kernel

$$G(x,dy\times du) = \mathbb{E}\left[\eta(x,dy)I_{\{\tau(x)\in du\}}\right].$$

The representation (4) allows one to give (by the same methods as in [30] a complete description limiting behavior of the mean $A(x,t)$ as $t\to\infty$ under the assumption that the kernel $G(x,dy) = G(x,dy\times [0,\infty))$ is critical and conservative (in [30] such kernels called critical and recurrent).

4 TRANSIENT PHENOMENA FOR BRANCHING PROCESSES WITH AN INFINITE SET OF TYPES CLOSE TO CRITICAL

We will assume that all introduced random variables depend on the small parameter $\varepsilon > 0$. Accordingly, we denote

$$\zeta_t(x) = \zeta_t^\varepsilon(x), \ \eta(x,S) = \eta^\varepsilon(x,S),$$

$$\xi_t(x,S,v) = \xi_t^\varepsilon(x,S,v),$$

$$\tau(x) = \tau^\varepsilon(x), \qquad x\in E, \qquad t\geq 0.$$

We fix a number $v > 0$, a bounded non-negative $\mathfrak{B}$ - measurable function φ and put

$$f_\varepsilon(x,t) = A_\varepsilon(x,t) = \mathbb{E}\int_E \xi_t^\varepsilon(x,dy,v)\varphi(y),$$

$$g_\varepsilon(x,t) = 0 \text{ at } 0\leq t < v,$$

$$g_\varepsilon(x,t) = \mathbb{E}\left[\varphi\big(\zeta_t^\varepsilon(x)\big)I_{\{t<\tau^\varepsilon(x)\}}\right] \text{ at } t\geq v.$$

The function $A_\varepsilon(x,t)$ satisfies the Markov renewal type equation

$$A_\varepsilon(x,t) = g_\varepsilon(x,t) +$$

$$+ \int_E \int_0^t G_\varepsilon(x,dy\times du)\,A_\varepsilon(y,t-u), \quad (5)$$

where

$$G_\varepsilon(x,dy\times du) = \mathbb{E}\left[\eta^\varepsilon(x,dy)I_{\{\tau^\varepsilon(x)\in du\}}\right],$$
$$x\in E, \qquad t\geq 0.$$

Thus, it can be argued that the solution of the equation (5) has the form

$$A_\varepsilon(x,t) = \int_E \int_0^t U_\varepsilon(x,dy\times du)\,g_\varepsilon(y,t-u),$$

where $U_\varepsilon(x,dy\times du)$ - the potential of the kernel $G_\varepsilon(x,dy\times du)$.

We will be interested in the asymptotic behavior $A_\varepsilon(x,t)$ at large values $t\geq 0$ and small values $\varepsilon > 0$.

We impose a number of the following conditions. Let the random process $\zeta_t^\varepsilon(x)$ converges to the random process $\zeta_t^0(x)$ with life time $\tau^0(x)$ particles with type x in that sense

$$\lim_{\varepsilon\to 0} \mathbb{P}\big\{ \zeta_{t_1}^\varepsilon(x)\in A_1,\dots,\zeta_{t_n}^\varepsilon(x)\in A_n, t_1,t_2,\dots,t_n <$$
$$\tau^\varepsilon(x)\big\} = \mathbb{P}\big\{ \zeta_{t_1}^0(x)\in A_1,\dots,\zeta_{t_n}^0(x)\in$$
$$A_n, t_1,t_2,\dots,t_n < \tau^0(x)\big\} \quad (6)$$

at points of continuity $t_1,\dots,t_n$ limiting probability distribution, $n = 1,2,\dots$, $A_i \in \mathfrak{B}$, $i = 1,2,\dots,n$, uniformly on $x\in E$.

We denote

$$G_\varepsilon(x,A) = \mathbb{E}\left[\eta^\varepsilon(x,A)\right]$$

and suppose that

$$\sup_{x\in E}\sup_{A\in\mathfrak{B}} |G_\varepsilon(x,A) - G(x,A)| \xrightarrow[\varepsilon\to 0]{} 0, \quad (7)$$

where

$$G(x,dy) = \mathbb{E}\left[\eta^0(x,dy)\right].$$

We will assume that the kernel $G(x,dy)$ is conservative and its Perron root is equal to one (the kernel is critical). This guarantees the existence of a non-trivial σ - finite measure l and a positive $\mathfrak{B}$ - measurable l - almost everywhere finite function h such that

$$\int_E l(dx)G(x,A) = l(A), \ A\in\mathfrak{B},$$

$$\int_E G(x, dy)h(y) = h(x), \quad x \in E.$$

Let be

$$0 < \inf_{x \in E} h(x) < \sup_{x \in E} h(x) < \infty. \qquad (8)$$

From (6), (7) the existence of a set of measures l_ε on $(E, \mathfrak{B})$ and functions $h_\varepsilon(x)$ is follows, such that

$$\int_E l_\varepsilon(dx)G_\varepsilon(x, A) = \lambda_\varepsilon l_\varepsilon(A), \quad A \in \mathfrak{B},$$

$$\int_E l_\varepsilon(dx)h_\varepsilon(x) = 1,$$

$$\sup_{A \in \mathfrak{B}} |l_\varepsilon(A) - l(A)| \underset{\varepsilon \to 0}{\longrightarrow} 0, \quad h_\varepsilon \underset{\varepsilon \to 0}{\longrightarrow} h,$$

where λ_ε - is the Perron root of the kernel $G_\varepsilon(x, dy)$, $\lambda_\varepsilon \underset{\varepsilon \to 0}{\longrightarrow} 1$.

Suppose the condition

$$0 < m = \int_E \int_E \int_0^\infty l(dx)G(x, dy \times dt)h(y)t < \infty. (9)$$

Applying the Markov renewal theorem [28], we obtain the following statement.

Theorem. Let in (6) - (9) the kernel $G(x, dy)$ is critical and conservative, the kernel

$G(x, dy \times du) = \mathbb{E}\left[\eta^0(x, dy)I_{\{\tau^0(x) \in du\}}\right]$ is non-lattice and a random process $\varphi(\zeta_t^\varepsilon(x))$ at $t < \tau^\varepsilon(x)$, stochastically continuous uniformly on $\varepsilon > 0$, $x \in E$, that is

$$\sup_{\varepsilon > 0} \sup_{x \in E} \mathbb{P}\{|\varphi(\zeta_t^\varepsilon(x)) - \varphi(\zeta_u^\varepsilon(x))| > \delta, t, u < \tau^\varepsilon(x)\} \underset{t-u \to 0}{\longrightarrow} 0$$

for all δ>0, then, if

$$\sup_{\varepsilon > 0} \int_E \int_E \int_T^\infty l(dx)G_\varepsilon(x, dy \times dt)h(y)t \underset{T \to \infty}{\longrightarrow} 0.$$

$$\sup_{\varepsilon > 0} \sup_{x \in E} \sup_{t \geq 0} A_\varepsilon(x, t) < \infty, \int_E l(dx)\,\mathbb{E}\,\tau^0(x) < \infty,$$

$$\lim_{\varepsilon \to 0} \int_E l(dx) \int_v^\infty \mathbb{E}\left[\varphi(\zeta_t^\varepsilon(x))I_{\{t < \tau^\varepsilon(x)\}}\right]dt =$$

$$= \int_E l(dx) \int_v^\infty \mathbb{E}\left[\varphi(\zeta_t^0(x))I_{\{t < \tau^0(x)\}}\right]dt = l_v(\varphi),$$

then

$$\lim_{\substack{\varepsilon \to 0 \\ t \to \infty \\ \gamma_\varepsilon t \to c}} A_\varepsilon(x, t) = e^{-c}\frac{h(x)}{m}l_v(\varphi),$$

uniformly on $x \in E$.

5 CONCLUSIONS

In branching processes, random variables are typically used to represent the number of offspring generated by each event, and the process can be analyzed using probability theory and stochastic processes. This allows IT professionals and researchers to make probabilistic predictions and decisions regarding system behavior and performance.

Overall, branching processes provide a valuable mathematical framework for modeling and analyzing various aspects of information technology, helping to make informed decisions and optimize IT systems and networks.

We have studied transient phenomena for branching processes with an infinite number of types close to critical. The analytical apparatus for this study is Markov renewal theorems.

Branched processes were used to evaluate the performance of service-oriented information technology to solve problems of sustainable environmental management and uniform information platform for the national automated ecological information and analytical system.

This is important for capacity planning and resource allocation.

REFERENCES

[1] B. Kamal, El F. Abdeslam, and E.E. Abdelbaki, "Software-defined networking (SDN): A survey", Security and Communication Networks, 2016, 9 (18), pp. 5803-5833, doi:10.1002/sec.1737.

[2] M. Ahmadreza, "Software-defined load-balanced data center: design, implementation and performance analysis", 2020, Cluster Computing, 24 (2), pp. 591-610, doi: 10.1007/s10586-020-03134-x. ISSN 1386-7857. S2CID 220490312.

[3] M. Ahmadreza, "Software-defined Internet of Multimedia Things: Energy-efficient and Load-balanced Resource Management, " IEEE Internet of Things Journal, 9 (3), pp. 2432-2442, doi: 10.1109/JIOT.2021.3095237. ISSN 2327-4662. S2CID 237801052.

[4] "Software-defined networking is not OpenFlow, companies proclaim," 2024, searchsdn.techtarget.com.

[5] "InCNTRE's OpenFlow SDN testing lab works toward certified SDN product," 2024, https://www.techtarget.com/news.

[6] M. Kimura, K. Saito, R. Nakano, and H. Motoda, "Extracting influential nodes on a social network for information diffusion. Data Min. Knowl. Discov. 2010, 20, pp. 70-97.

[7] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks, " In: International AAAI Conference on Weblogs and Social Media, 2009.

[8] F. Hartog, A. Raschella, F. Bouhafs, P. Kempker, B. Boltjes, and M. Seyedebrahimi, "A Pathway to solving the Wi-Fi Tragedy of the Commons in apartment blocks, " In 2017 27th International Telecommunication Networks and Applications Conference (ITNAC), 2017, pp. 1-6.

[9] J. Stuart, "Engineering Information Security: The Application of Systems Engineering Concepts to Achieve Information Assurance, ", John Wiley & Sons, 2015. ISBN 9781119104711.

[10] T. Fox-Brewster, "Netflix Is Dumping Anti-Virus, Presages Death Of An Industry," Forbes. Archived from the original on September 6, 2015.

[11] C. Claire, N. Holtschulte, E.K. Smith, Y. Brun, P. Devanbu, S. Forrest, and W. Weimer, "The ManyBugs and IntroClass Benchmarks for Automated Repair of C Programs". IEEE Transactions on Software Engineering. 2015, 41 (12), pp. 1236-1256, doi:10.1109/TSE.2015.2454513. ISSN 0098-5589.

[12] W. Chwan-Hwa, and D. Irwin, "Introduction to Computer Networks and Cybersecurity. CRC Press. 2016, p. 500. ISBN 978-1-4665-7214-0.

[13] S, Behzad, B. Morteza, and A. Masayoshi, "Efficient feature extraction model for validation performance improvement of duplicate bug report detection in software bug triage systems". Information and Software Technology, October 1, 2020, 126: 106344, doi:10.1016/j.infsof.2020.106344. S2CID 219733047.

[14] M. Rigger, D. Pekarek, and H. Mössenböck, "Context-Aware Failure-Oblivious Computing as a Means of Preventing Buffer Overflows", Network and System Security, Lecture Notes in Computer Science, Cham: Springer International Publishing, 2018, vol. 11058, pp. 376-390, arXiv: 1806.09026, doi: 10.1007/978-3-030-02744-528, ISBN 978-3-030-02743-8.

[15] A. Herzberg, and H. Shulman, "Oblivious and Fair Server-Aided Two-Party Computation". 2012 Seventh International Conference on Availability, Reliability and Security. IEEE. 2012. pp. 75-84, doi: 10.1109/ares.2012.28. ISBN 978-1-4673-2244-7. S2CID 6579295.

[16] B. Al-Shaikh and S.G. Stacey, "Essentials of Equipment in Anaesthesia, Critical Care, and Peri-Operative Medicine, p. 247, 2017.

[17] E. Dubrova, "Fault-Tolerant Design," Springer, 2013, ISBN 978-1-4614-2112-2.

[18] L. Globa, S. Dovgiy, O. Kopiika, and O. Kozlov, "Approach to building uniform information platform for the national automated ecological information and analytical system," In: CEUR Workshop Proceedings, 2021, 3021, pp. 53-65.

[19] V. Sundarapandian,"7.Queueing Theory," Probability, Statistics and Queueing Theory. PHI Learning. ISBN 978-81-203-3844-9.

[20] L.W. Dowdy, A.F. Almeida, and D.A. Menasce, "Performance by Design: Computer Capacity Planning by Example," Archived from the original on 2016-05-06, [Accessed: 8 Dec 2023].

[21] K. Schlechter, "Hershey Medical Center to open redesigned emergency room," The Patriot-News. Archived from the original on June 29, 2016.

[22] "How to find the performance bottleneck in C# desktop application?" Stack Overflow, 2012.

[23] K. Kirk, "Performance Profiling with a Focus," (2017), Develop for Performance.

[24] "What is code profiling? Learn the 3 Types of Code Profilers," Stackify Developer Tips, Tricks and Resources. Disqus, 2016.

[25] E. Lawrence, "Getting Started with Profile Guided Optimization," testslashplain, 2016.

[26] K. Krauss, "Matching Wildcards: An Improved Algorithm for Big Data," Develop for Performance, 2018.

[27] "List of .Net Profilers: 3 Different Types and Why You Need All of Them," Stackify Developer Tips, Tricks and Resources, Disqus, 2016.

[28] S. Degtyar, "Markov renewal limit theorems, Theory of Probability and Mathematical Statistics," 76, pp. 33-40, 2008.

[29] S.V. Degtyar, Y.V. Shusharin, and Y.O. Zhuk, "Motion of Symmetric Gyroscope under Action of Gravity with Discrete Random Change of Parameters of System," International Applied Mechanics, November 2022, vol. 58, pp. 634-644.

[30] L. Zonenashvili, "Some remarks on the asymptotics of the mean number of particles in branching processes," Bulletin of the Georgian National Academy of Sciences, January 1986.

[31] P. Jagers, "Branching processes with biological applications," New York, Wiley, 1975, 20, pp. 848-851, 1975.

[32] V.M. Shurenkov, "Ergodic Theorems and Related Problems," "Nauka", Moscow, 1989; English transl., VSP International Science Publishers, Utrecht, 1998. MR1690361 (2000i:60002).

Solution of the Inverse Incorrectly Posed Problem by the Library Method

Valentin Davydov, Vladimir Brunetkin and Ganna Lysiuk

Department of Computer Technologies of Automation, Odesa Polytechnic National University, Shevchenko Avenue 1,
65044 Odesa, Ukraine
{davydov, brunetkin.v.o, lysjuk }@op.edu.ua

Abstract: The presented study considers a range of problems to be interpreted as inverse problems. The complexity of solving incorrectly posed inverse problems due to their unique features and the lack of a universal method for their solution is highlighted. In this regard, the importance of considering the application of different methods on the example of specific problems is emphasised. The research analyses a method of solving the problem of determining the composition of gaseous hydrocarbon fuel in the process of its combustion in real time. It is emphasised that the way of measuring technological parameters allows to consider this problem as a complex problem of interpretation. The method of "library" (selection) as the most universal one is chosen. For its implementation the method of library formation in the form of a three-dimensional array is proposed, where the initial data for each solution of the direct problem are presented in the form of a single number, using the positional principle of recording decimal numbers. In the process of forming the working array, the method of comparing the oxidant excess ratio and the ratio of oxidant and fuel volume flow rates was used. This made it possible to use the results of the solution of the direct problem to determine the composition of fuel by the measured temperature of combustion products within the framework of the inverse problem. A method of searching for a solution among the elements of the working array based on the results of technological measurements of the combustion product temperature and the ratio of oxidant and fuel volume flow rates was developed.

1 INTRODUCTION

The solution of technical problems is based on a list of initial data. Their quantitative characteristics can be set and kept constant or determined by direct measurements during operation. But it is not always possible. For example, quantitative characteristics of the composition of substances in the process of work may change, and there is no possibility of their determination. Or in principle they are measurable, but cannot be provided with the necessary promptness. Also, the elements that determine the configuration of an object may be known, but their quantitative parameters are not known and are not directly measurable, etc. In all cases, an approach related to the solution of inverse problems can be used. In the process of its realisation it is necessary to determine, for example, the composition of initial substances or characteristics of elements included in the considered object by measured parameters of manifestation of some processes. In this case, the qualitative composition of substances (a set of chemical elements) or the list of elements and the order of their combination are known. It is necessary to determine their quantitative characteristics.

The generality of the approach to research by solving the inverse problem is accompanied by a variety of external conditions and, as a consequence, of possible solution methods. These methods depend on the problem to be solved. The type of method can be selected or developed in relation to a specific problem. For this reason, it is relevant to consider the application of each method on the example of solving a specific problem relevant in modern society. In this case, the method of determining the variable composition of the mixture of gaseous fuel in the process of its combustion is considered. Unlike certified fuel, the variable composition is due to the use of alternative sources of gaseous fuel, including those that do not generate greenhouse effect (gas generators, pyrolysis gases of woodworking waste, agricultural waste, coke, blast furnace gases, etc.).

The optimal combustion process of gaseous fuel of variable composition can in principle be organised on the basis of direct measurement methods.

Gas analysers. Mixtures of gases can be analysed for their qualitative and quantitative composition using chemical, physico-chemical and physical methods. The advantages of physico-chemical and physical methods over chemical methods are the rapidity of execution and the possibility of automating the analysis. Today, the most commonly used methods are gas chromatography [1]. Their main disadvantages are the sequential (not simultaneous) measurement of different gas components, long analysis time, and difficulties encountered in the detection of some components.

With the advent of small, reliable laser sources and highly sensitive multichannel photodetectors, the development of gas analysers based on the use of spontaneous Raman spectroscopy began. The development of such an instrument has been considered in the applied plan [2, 3], although it is still at the stage of laboratory development.

Extreme regulators. Despite the fact that they do not allow determining the composition of the combusted gas, they can theoretically be used to organise the combustion process with maintaining the temperature of combustion products at an optimal (maximum) level. All types of extreme regulators have a common disadvantage. When approaching the extremum point, the gradient of the measured parameter decreases (down to zero). This applies both to the case of determining the maximum temperature in the combustion chamber (furnace) [4] and to the case of determining the maximum temperature of the heat transfer medium [5]. The same property of the function is also manifested when determining the minimum (zero) concentration of carbon monoxide (CO) at the minimum concentration of oxygen in flue gases [6]. Taking into account that measuring instruments have a threshold of sensitivity, as well as there are noises in the measuring channels, at the point of extremum there is a maximum measurement error, which often excludes the possibility of using extreme regulators.

Determination of the reduced formula of the mixture. If the enthalpy and gross formula of a mixture of gases are known, some calculations of combustion process parameters are performed on the basis of chemical kinetics methods for a single substance [7,8]. In [9], a method for determining such a gross formula is proposed. It is based on the measurement of the temperature of combustion products (CP) and the flow rates of an unknown gaseous fuel and air. The formulated problem is an inverse one. On the basis of measurements of technological parameters, the composition of the combusted fuel is determined. In [9] it is shown that the solution of this problem exists and is unique. But the attempt to find the solution showed its instability. Thus, while this approach is promising, it leads to the necessity of solving the inverse, incorrectly formulated problem.

There is no single method for solving the whole variety of inverse incorrectly posed problems. Each of these problems requires an individual approach and contains an element of heuristics. As a result, there is a need to develop a new solution method capable of coping with the complexity and variability of such problems.

2 AIM AND OBJECTIVES OF THE STUDY

The aim of the study is to solve the problem of determining the composition of flared gas using the constraint method.
In order to achieve the objective, the following objectives were set:

- justify the choice of a method for solving the inverse incorrectly posed problem;
- develop a method of filling the library of solutions to a direct problem as a working three-dimensional array;
- to develop a method for finding a solution to a complex interpretation problem on the basis of using data from a working three-dimensional array.

3 METHOD FOR SOLVING THE PROBLEM

3.1 Justification of the Chosen Solution Method

To solve the inverse problem, you must first master the methods for solving the direct problem.

Let us consider the process on the example of solving the problem of determining the composition of flared gas.

During the solution of the direct problem, one of the initial parameters is the oxidant excess coefficient, denoted by α. However, a different parameter is used in the calculations:

$$\chi = \alpha \cdot \chi_o, \qquad (1)$$

where χ is the molar ratio of the components; χ_0 is the molar stoichiometric ratio of the components.

In the process of analysing the combustible and oxidising agent represented as ideal gases, the value of χ can be calculated on the basis of equality of molar volumes of any gases

$$\chi = \dot{V}_{ox}/\dot{V}_f. \qquad (2)$$

Here $\dot{V}_{ox}$ is the oxidiser (air) volume flow rate; $\dot{V}_f$ is the fuel volume flow rate. In other words, the value of χ (1) can be calculated on the basis of technological measurements of flow rates, while the chemical composition of the combustible fuel and oxidiser is not fixed. The model for solving the problem of determining the composition of combustible fuel is built on this property, using technological measurements of CP temperature and volumetric flow rates of combustible and oxidant (inverse problem).

Suppose that the fuel composition is known in advance in the form of the gross formula $C_{b_C} H_{b_H} O_{b_O} N_{b_N}$ with a list of chemical elements.

We will assume that during a certain period of time the fuel composition remains constant, ensuring the constancy of the values of the required parameters b_C, b_H, b_O, b_N, I_t in the process of their determination. The ratio between fuel and oxidiser is set with measurement of their volume flow rates ($\dot{V}_{ox}^1$ and $\dot{V}_f^1$). The temperature of the CP T^1 is measured during the combustion process. This data is used to form part of the model as a set of equations (in this case, 13 equations) that allow the determination of 12 partial pressures P_j^1 and one of the quantities sought, e.g., b_C. Then the volume flow ratios ($\dot{V}_{ox}^i$ and $\dot{V}_f^i$) are changed several times, and for each case the temperature of the PS is measured - T^i. These data serve as a basis for the formation of additional sets of equations necessary to determine the remaining required parameters - b_H, b_O, b_N, I_t. At the same time, the corresponding number of sets of partial pressures P_j^i of the CP is also determined. As a result, a system of equations is formed, consisting of five sets of 13 equations each (65 equations in total), designed to determine the five required quantities and the corresponding groups of partial pressures. Each group corresponds to different ratios of fuel and oxidiser volume flow rates.

In [10], the most universal approach to solving such a direct problem is outlined.

## 3.2 Method of Solving a Direct Problem

At the stage of preparation for determination of the gross formula of the fuel it is necessary to fill in the library of solutions for the direct problem with variation of six parameters. These parameters include the possible number of atoms of the corresponding elements in the gross formula of the fuel (b_C, b_H, b_O, b_N), possible values of the oxidant excess factor α and enthalpy of the fuel I_t.

The number of intervals can be chosen arbitrarily, and for concreteness we will take it equal to 100. This will allow tracking the change of each of the values b_C, b_H, b_O, b_N, in the interval 0.05...4.95 with a step of 0.05, the value α - in the interval 0.025...2.5 with a step of 0.025. It will require $(100)^6 = 10^{12}$ solutions of the direct problem with fixation of all results. At this stage, it is possible to reduce the dimensionality of the problem.

Thus the stoichiometric combustion process of ethane (and any other substance) can be organised both on the basis of its complete formula and at a normalised number of atoms.

$$2X2H6+7O2=4XO2+6H2O$$

or $\qquad (3)$

$$4X1H3+7O2=4XO2+6H2O.$$
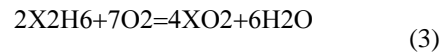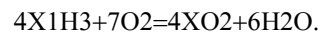
Here the normalisation is carried out by the number of carbon atoms. In the gross formula, the number of atoms is not necessarily an integer, and normalisation can be performed with respect to any chemical element. In this problem, carbon normalisation is applied, where the number of carbon atoms is fixed and equal to 1 and the gross formula is defined as follows:

$$C_1 H_{(b_H/b_C)} O_{(b_O/b_C)} N_{(b_N/b_C)} \text{ or } C_1 H_{d_H} O_{d_O} N_{d_N} \qquad (4)$$
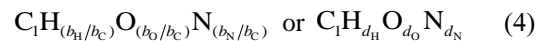
Thus, the number of solutions to the direct problem is reduced to $(100)^4 = 10^8$ realisations as a result.

At the next stage, the proposed method is used to convert the obtained results into a form convenient for further use in finding a solution based on the conducted process measurements of the CP temperature and volume flow rates of the fuel and oxidiser:

1) By varying dH, dO, dN and from (4), for each realisation of the solution of the direct problem, a certain fuel composition is set. For each case, the molar stoichiometric ratio of the components $\chi 0$ is determined. The change in the value of the oxidiser excess ratio α is converted into a change

in the ratio of the measured process parameters $\dot{V}_{ox}/\dot{V}_f$.

2) At all possible combinations of initial data $\dot{V}_{ox}/\dot{V}_f$ (α), dH, dO, dN and It, arising at variation of their values, an attempt is made to solve a direct problem on definition of temperature of CP. In some cases, the ratio of fuel composition and enthalpy is incompatible, and such variants are excluded from consideration.

3) For the remaining (successful) realisations of the solution to the direct problem, the set of input data dH, dO, dN and It is represented as a single number using the positional principle of decimal number representation.

First, the data is written in normalised form:

$$\bar{d}_H = \frac{d_H - d_H^l}{d_H^r - d_H^l} \cdot 100. \qquad (5)$$

Here d_H - number of hydrogen atoms in the considered realisation of the direct problem; d_H^l, d_H^r - left and right boundaries of the considered interval of change in the value of the number of hydrogen atoms during the solution of the direct problem. The obtained values are combined into one number "E" by setting to its corresponding digits (Figure 1). Although the order of formation of this number can be any, let us assume that the value I_t . is placed in the first two digits. As a result, complexes are formed in which each pair of values of the results of the solution of the direct problem (T and $\dot{V}_{ox}/\dot{V}_f$) corresponds to one number "E" with packed values of the initial data corresponding to them.

4) The calculation results (T and $\dot{V}_{ox}/\dot{V}_f$) are structured. Their max [Tr, ($\dot{V}_{ox}/\dot{V}_f$)r] and min [Tl, ($\dot{V}_{ox}/\dot{V}_f$)l] values are determined. The ranges of variation of these values are divided into a certain number of intervals. The number of intervals and the value of the step of change of each value are correlated with the technological accuracy of measurement (T and $\dot{V}_{ox}/\dot{V}_f$) in the process of fuel combustion.

5) The results of the solution (T and $\dot{V}_{ox}/\dot{V}_f$) of the direct problem and their corresponding initial data, summarised into a single number "E", are used to form a three-dimensional array (Figure 2).
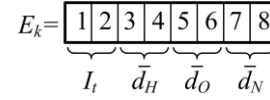

Figure 1: Diagram of positional representation in the form of a single number of normalized values of the initial data.
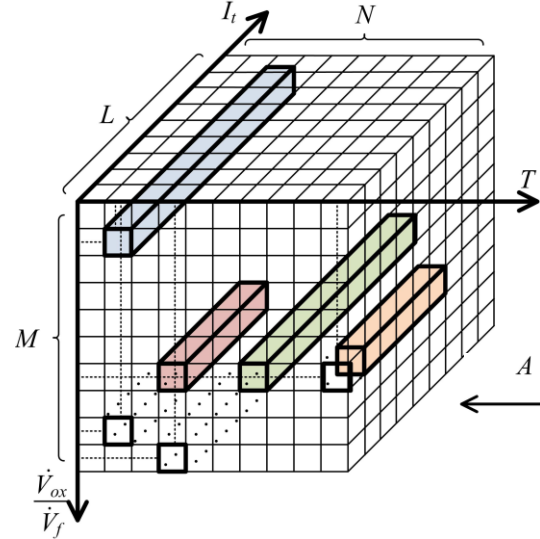

Figure 2: Diagram of forming a structured three-dimensional working array based on the initial data and the results of solving the direct problem.

6) Each cell corresponds to a vector of dimension L (Figure 2), the elements of which are numbers "E". The value of the number is determined by the value in the first (highest) digit. In this case, the values in the first digits are related to the enthalpy value. Therefore, the vector axis is labelled as enthalpy axis It .

Parameters L, M, N, are determined on the basis of the required accuracy of consideration of the corresponding values. Let's assume, for example, that N=M=100. In this case, 104 cells are formed on the plane T -$\dot{V}_{ox}/\dot{V}_f$. At the number of realisations of the solution of the direct problem 108 each cell can correspond to the results of ~108/104 =104 solutions. The same number of numbers "E_k" associated with these solutions are placed in the elements of the corresponding vectors.

7) In the elements of the vector are placed the values "E_k", which are in their respective boundaries:

$$E_k \in [(I_t^l)_k, (I_t^r)_k]. \qquad (6)$$

Since not all combinations of fuel composition ($\bar{d}_H$, $\bar{d}_O$, $\bar{d}_N$) can be realised with the

considered list of enthalpy values I_t, some elements of the vectors are unfilled. The elements of the vectors are ordered in ascending order. Examples of filling of some vectors are shown in Figure 2 and marked in colour.

3.3 Method for Finding a Solution to the Interpretation Problem

A method for determining a fuel composition using a generated working three-dimensional array includes the following steps:

1) The CP temperature T_i is measured for different values of the oxidiser-fuel volume ratio $(\dot{V}_{ox}/\dot{V}_f)$i . The number of measurements is determined by the number of quantities to be determined. So in case of fuel of the type (4) it is necessary to determine 4 quantities $(\bar{d}_H, \bar{d}_O, \bar{d}_N, I_t)$ and, accordingly, to perform 4 sets of measurements.

2) Based on the measurements, 4 cells (highlighted by dashed lines in Figure 2) and associated vectors (highlighted in colour) are defined on the T - $\dot{V}_{ox}/\dot{V}_f$ plane of the working three-dimensional array. Each vector in E_k records all possible combinations of defined quantities $(\bar{d}_H, \bar{d}_O, \bar{d}_N, I_t)$ that provide the temperature T_i at $(\dot{V}_{ox}/\dot{V}_f)$i .

3) The projections of the selected vectors on the plane $\dot{V}_{ox}/\dot{V}_f$ - It (view A, Figure 2), more precisely on the axis It (Figure 3) are determined.

4) Since the composition and enthalpy of the determined fuel remain constant at any ratios of changes in volume flow rates, the corresponding values of the numbers E_k in all vectors under consideration must be the same. Therefore, the area D of intersection of the set of elements of each vector is determined (Figure 3). The elements of vectors exceeding the area are removed. Further actions are aimed at reducing the width of the area to 1 element, which will determine the desired solution.

5) The selected numbers (area D, Figure 3) are rearranged in stack order. The values of the next value to be determined are transferred to the upper digits The value from the upper digits (Figure 1) is transferred to the end of the number.

The rearranged numbers are sorted, resulting in a scheme similar to the one shown in Figure 3.

6) The steps from step 4 are repeated until all values from the list of values to be determined ($I_t, \bar{d}_H, \bar{d}_O, \bar{d}_N$) have been used. With each pass the number of remaining numbers E_k decreases. In the end there remains one value E_k in which the determined parameters ($I_t, \bar{d}_H, \bar{d}_O, \bar{d}_N$) correspond to all measured Ti at the given $\dot{V}_{ox}/\dot{V}_f$. This is the solution.
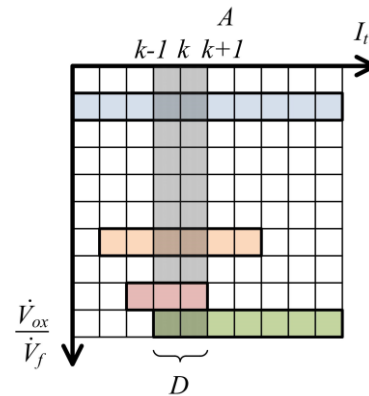

Figure 3: Example of the projection of selected vectors.

4 EXAMPLES OF SOLVING PROBLEMS OF FUEL COMPOSITION DETERMINATION

4.1 Simulation of Process Measurements

Initial data in the form of process measurements were simulated by the results of solving the direct problem for the corresponding substances. Ethanol was chosen as the fuel. The choice of ethanol as a combustible material is justified by the availability of reference data [9] on its combustion process in oxygen (α, T, composition of the CP). This ensures the reliability control of the results in solving the direct problem. Two variants of initial data setting are considered:

The following input data variation ranges were adopted: $\bar{d}_N$ from 0 to 4.95 in steps of 0.05; $\bar{d}_O$, from 0 to 4.95 in steps of 0.05; α from 0.025 to 2.5 in steps of 0.025; I_t from -13500 to 8800 in steps of 100.

Thus, $2.24 \cdot 10^8$ solutions of the direct problem were performed. This took several hours with a

personal computer with the following parameters Intel Core i5-4590 3.3 GHz, 32 GBt RAM, HDD 2 TBt. For each combination of input data, the CP temperatures were calculated and the oxidiser excess coefficients α were recalculated with respect to the oxidiser and propellant volume flow rates $\dot{V}_{ox}/\dot{V}_f$ Part of the combinations of input data turned out to be incompatible, which was revealed during the solution process. After removing these incompatible combinations, a working data array was generated. When using the text format, it required ~2.7 GBt of memory to accommodate it. After conversion to binary form, the volume decreased to 1.2 GBt.

5 DISCUSSION OF THE RESULTS

In the process of forming a working three-dimensional array for the test problem, 10^8 variants of solutions were calculated, which took several hours. At the same time, the gross formula of the fuel included only three chemical elements: C, H and O. The addition of each next element led to an avalanche-like increase in the number of solutions to the direct problem. Thus, it is necessary to consider the limits of applicability of the proposed solution method.

In most cases, the gases whose composition is determined are organic compounds, and their composition can be determined using six chemical elements. In addition to the three discussed above, nitrogen (N), sulphur (S), phosphorus (P) should be considered. All initial data determining the possible composition of a certain gas are represented as one number of normalised values of the initial data (Figure 1). Therefore, increasing the number of chemical elements taken into account (from three to six) will only double the length of these numbers. The required memory required to store the data will increase to 2.4 GBt if they are represented in binary form.

Accounting for each successive chemical element in the gross fuel formula increases the number of solutions by a factor of 100, which also increases the complexity of the model and the duration of individual solutions. When considering this problem, it should be taken into account:

- phosphorus (P) is rarely found in combustible gases and can be excluded from the calculations;
- the amount of nitrogen (N) and sulphur (S) in the fuel gross formula is significantly less than the amount of other elements taken into account.

The range of their possible variation can be reduced while maintaining the same discretisation step.

The result is an increase in the number of solutions and thus the time spent, but not as catastrophic as originally imagined. The impact of hardware, in particular computing speed, also plays a role. Modern personal computers, even household ones , have high computing power compared to the hardware used in the test task. In addition, the solution of a direct problem may be distributed among several computers. The problem can be solved by using the time of a supercomputer in a network.

It is important to note that the solution of the whole set of direct tasks and formation of the working array is performed once at the preliminary stage without time limits. The work can continue for a week or a month. While the solution of the inverse problem, using the formed array, is performed on an ordinary personal computer and takes only a few seconds.

6 CONCLUSIONS

The paper has accomplished the following:
1) In this paper, a general type of problem, characterised as an inverse incorrectly posed problem, was identified and a narrower class of complex interpretation problems was distinguished. In order to solve this class of problems efficiently, a method based on the use of a library and the matching principle was chosen.
2) A method of structured and compactly placing the results of direct problem solving has been developed to populate the library. In its framework:
   - a method for converting the oxidant excess coefficient α as the initial data for the direct problem with respect to the volume flow rates of oxidant and fuel is proposed. The obtained value is used as a measurable technological parameter in solving the inverse problem of fuel composition determination;
   - a method of transformation of other initial data of the direct problem (number of atoms of chemical elements in the gross formula and enthalpy of fuel) into one number, which is formed on the basis of the positional principle of recording decimal numbers, is proposed.

3) A method for solving the complex problem of data interpretation based on analysing the generated 3D array was developed. This method involves the application of constraints and sequential selection of areas where the data match. By step-by-step reduction of the amount of magnitude analysis, it was possible to achieve a single result in a limited number of steps that corresponds to the solution sought. An important difference between this method and the previously used method of direct enumeration of library data is its ability to effectively reduce the amount of data analysed, ensuring that the solution sought is found accurately and quickly.

REFERENCES

[1] J.P. Donval, V. Guyader, and E. Boissy, "A simple method for the preparation and injection of gas mixtures into a gas chromatograph using a two-component device," Journal of Chromatography, vol. 1631, 2020, [Online]. Available: https://doi.org/10.1016/j.chroma.2020.461579.

[2] M.A. Buldakov, B.V. Korolev, I.I. Matrosov, D.V. Petrov, and A.A. Tikhomirov, "Raman gas analyzer for determining the composition of natural gas," Journal of Applied Spectroscopy, vol. 80 (1), pp 124-128, March 2013, doi: 10.1007/s10812-013-9731-6.

[3] S. Schorsch, J. Kiefer, S. Steuer, T. Seeger, A. Leipertz, and S. Gonschorek, "Development of an Analyzer System for Real-time Fuel Gas Characterization in Gas Turbine Power Plants," Chemie Ingenieur Technik, vol. 83(3), pp. 247-253, March 2011, [Online]. Available: https://dx.doi.org/10.1002/cite.201000095.

[4] Z. Dou and L. Sun, "Design of Temperature Controller for Heating Furnace in Oil Field," International Conference on Applied Physics and Industrial Engineering. Physics Procedia, vol. 24, pp. 2083-2088, March 2012, [Online]. Available: https://dx.doi.org/10.1016/j.phpro.2012.02.305.

[5] V.M. Larionov, Yu.V. Vankov, and E.R. Sayfullin, "Sposob avtomaticheskoj optimizacii processaszhiganiya topliva peremennogo sostava [Method of fuel with variable composition combustion process automatic optimization]. Patent RUS, no. 2647940, 2018, [Online]. Available: https://yandex.ru/patents/doc/RU2647940C1_20180321.

[6] J. Pitel', J. Mizáková, and A. Hošovský, "Biomass Combustion Control and Stabilization Using Low-Cost Sensors," Advances in Mechanical Engineering, vol. 5, 2013, [Online]. Available: http://dx.doi.org/10.1155/2013/685157.

[7] D.L. Ferreira Bárbara, J.M. Paulo, J.P. Braga, R.C.O. Sebastião, and F.J.P. Pujatti, "Methane combustion kinetic rate constants determination: an ill-posed inverse problem analysis" Química. Nova vol. 36 no. 2 São Paulo 2013, [Online]. Available: http://dx.doi.org/10.1590/S0100-40422013000200011.

[8] M. Orbegoso, E.F. da Silva, L. Fernando, and A. Novgorodcev "On the predictability of chemical kinetics for the description of the combustion of simple fuels," Journal of the Brazilian Society of Mechanical Sciences and Engineering, vol. 33, pp. 492-505, 2011, doi: 10.1590/S1678-58782011000400013.

[9] A.I. Brunetkin and M.V. Maksimov, "The method for determination of a combustible gase composition during its combustion," Science Bulletin of the National City University, no. 5, pp. 83-90, 2015, [Online]. Available: http://nbuv.gov.ua/UJRN/Nvngu_2015_5_16.

[10] V.E. Alemasov and V. Glushko, "Thermodynamic and thermophysical properties of combustion products" = Termodinamicheskie i teplofizicheskie svoistva produktov sgoraniya : translated from Russian [by R. Kondor, and Ch. Nisenbaum], Jerusalem : Israel Program for Scientific Translations, v.1, [Online]. Available: https://searchworks.stanford.edu/view/892711.

Algorithms for Optimizing Vocabulary Acquisition in Language Learning

Tetiana Verhun[1], Maryna Davydovych[1], Kateryna Parnus[1], Marina Kosheleva[1], Iryna Biliak[1] and Olga Khorosh[1]

[1]*Zhytomyr Polytechnic State University, 103 Chudnivska Str., 10005 Zhytomyr, Ukraine*
kim_vtm@ztu.edu.ua, kim_dms@ztu.edu.ua,katemarchuk38@gmail.com, maryshelle68@gmail.com,
ktpl_biv@ztu.edu.ua, olyakhorosh@gmail.com

Keywords: Dictionaries, Foreign Language Classes, Vocabulary, Algorithm, Word.

Abstract: This study addresses the critical role of dictionaries in language learning, emphasizing the need for effective navigation strategies to enhance vocabulary acquisition. As globalization intensifies, the ability to learn new languages efficiently becomes increasingly important, making the act of selecting and using dictionaries more significant. Traditional dictionary use, while beneficial, often lacks the precision and adaptability required by modern learners. This research introduces navigational algorithms designed to optimize dictionary use, making vocabulary learning more efficient and tailored to individual needs. Employing a mixed-methods approach, the study analyzes data from students at the State University "Zhytomyr Polytechnic" to explore current usage patterns and the challenges faced in vocabulary acquisition. The study's methodology included quantitative analysis of survey responses and qualitative evaluation of the algorithms' impact on learning outcomes. The development and application of navigational algorithms aim to facilitate learners' independence, improving their ability to understand and retain new words effectively. The findings reveal that students utilizing these algorithms demonstrate notable improvements in vocabulary retention and contextual application, suggesting a deeper linguistic comprehension. These results not only highlight the practical benefits of incorporating navigational algorithms into language learning but also underscore the theoretical implications for applied lexicography and educational technology. By bridging the gap between conventional dictionary usage and the dynamic requirements of contemporary language learners, this research contributes to the fields of language education and lexicography. It underscores the potential of combining traditional linguistic resources with modern technology to meet the needs of today's learners. This study advocates for a paradigm shift towards the development of more interactive and user-friendly dictionaries, leveraging algorithms to support language acquisition in an increasingly digital and globalized learning environment.

1 INTRODUCTION

In an era marked by increasing globalization, proficiency in multiple languages has emerged as a critical element of effective communication and cultural understanding. The role of dictionaries as indispensable tools in the realm of language learning has never been more significant. As learners embark on the journey of acquiring a new language, the initial step of learning new words and phrases necessitates not just any dictionary, but the right one tailored to their specific needs. The ability to select the most appropriate dictionary is crucial, as it significantly influences the learner's ability to form a solid foundation in their new language, thereby facilitating a smoother learning experience and success in subsequent stages of language mastery.

The centrality of dictionaries in academic settings, particularly within university education, cannot be overstated, particularly in university settings where they enhance comprehension and serve as catalysts for improved communication in foreign languages. With their capacity to aid in vocabulary expansion and language skill development, dictionaries—especially monolingual versions—are highly effective. Yet, the vast array of available dictionaries poses a challenge, requiring learners to navigate through this diversity to find the most suitable type.

While bilingual dictionaries may offer convenience, they often fall short of providing the depth of information available in student-oriented dictionaries, which include multiple definitions, tense variations, plural forms, pronunciation guides, parts of speech, exemplar sentences, and idiomatic expressions. Such comprehensive resources can significantly enrich the language learning experience, leading to more effective acquisition strategies and enhanced learning outcomes, as supported by existing research [3].

The urgency of this research stems from the critical need to streamline the process of dictionary selection and use in an increasingly multilingual world. As language learning evolves, so too must the tools and strategies employed by learners. This study aims to address this gap by developing a structured approach that leverages navigational algorithms to guide learners in selecting and utilizing dictionaries more effectively. By integrating these algorithms, we propose a solution that not only simplifies the search process within dictionaries but also tailors it to the unique needs of each learner, thereby enhancing vocabulary acquisition and overall language proficiency.

This research seeks to:
1) Examine the current challenges faced by language learners in selecting and using dictionaries efficiently.
2) Develop and assess the effectiveness of navigational algorithms designed to optimize dictionary use for language learning.
3) Explore the practical implications of these algorithms for educators and learners, and their potential to transform language education.

By fulfilling these aims, this study aspires to contribute to the fields of applied linguistics and language education, offering innovative solutions that align with the needs of today's global communicators.

2 THE EFFECTIVE USE OF VARIOUS DICTIONARIES WHEN LEARNING A FOREIGN LANGUAGE

In the spectrum of language acquisition, thesauri (thesaurus) and collocations dictionaries are two types of dictionaries that should be used regardless of the level of foreign language proficiency. Learning a new language is not just about learning a meaning or a definition of a word. While the conventional approach to language learning often focuses on the mere extraction of word meanings, the significance of delving into associated phrases, expressions, idioms, and sentences cannot be overstated.

At the initial stages of language acquisition, the judicious use of thesauri and collocations dictionaries proves instrumental in broadening the contextual understanding of words. These resources not only provide alternative expressions but also illuminate the intricate web of linguistic associations, fostering a more holistic comprehension of language usage.

Recent meta-analyses have shed light on the multifaceted nature of dictionary use in language learning. Studies have consistently demonstrated the positive impact of dictionary use on vocabulary acquisition, though the magnitude of this effect varies widely across different learning contexts and dictionary types. For instance, the meta-analysis conducted highlights the significant yet varied effectiveness of dictionary use in assisting second language (L2) learners to expand their lexical repertoire. The analysis distinguishes between treatment-related variables, methodological variables, and learner-related variables, providing a comprehensive understanding of the factors that influence the effectiveness of dictionaries in language learning [4].

The advancement of technology has notably affected learners' choice of language learning tools, shifting preferences towards electronic dictionaries due to their convenience. However, the debate on whether electronic dictionaries outperform their paper counterparts in scaffolding L2 vocabulary learning remains unresolved. Studies suggest that the medium of the dictionary (electronic vs. paper) may have different implications for immediate learning and delayed retention of vocabulary knowledge, indicating a need for further exploration in this area [1].

Moreover, the type of dictionary (monolingual vs. bilingual) and the target lexical unit (single words vs. multi-word units) are identified as crucial factors affecting vocabulary acquisition. While monolingual dictionaries are often cited for their precision and depth, bilingual dictionaries are preferred by a majority of learners for their ease of use [8]. The effectiveness of dictionaries in learning multi-word units, an area that has been somewhat neglected in lexicography and language learning research, also warrants further investigation.

The accessibility of dictionaries to students has evolved in tandem with technological advancements, providing a spectrum of options for learners to

enhance their vocabulary. Currently, students can access such avenues for free:

- printed dictionaries, they can be purchased online or in a bookstore. But with the widespread use of online dictionaries, printed dictionaries seem a bit old-fashioned.
- online dictionaries, access to such dictionaries can be obtained by subscription, for example, the Oxford English Dictionary. You can find everything you need in such dictionaries, because they provide comprehensive information about the vocabulary [5].

Using a dictionary to learn vocabulary is a complex process that requires mental effort and in-depth study of various strategies. Students should undergo through training in how to use a dictionary transforming the seemingly intricate task of word retrieval into a seamless process. Therefore, acquiring skills and algorithms for effective dictionary use has become a focal point in recent decades, giving rise to a new field of knowledge, which is defined as applied lexicography.

Applied lexicography encompasses an array of subjects, including the study of vocabulary and its application. Within this, a variety of studies have emerged, such as nuances of dictionary use strategies, purposes behind dictionary usage, user attitudes toward dictionaries and the learning of dictionary use. In recent decades, there has been a growing interest among many scholars (Schofield 1982; Hartmann 2001; Wingate, 2004; Bishop 2001) in the study of various strategies used by dictionary users in the process of searching for the necessary information. In addition, researchers have investigated and conceptualized that students use dictionaries for two functions - decoding and encoding [2]. One prominent model, developed by Schofield, refers to a series of algorithms guiding students in the process of dictionary use. In fact, Scholfield suggests that the strategy of finding and understanding the information needed by the acquirer can be divided into the following steps:

1) Identify Unclear Words and Phrases. (The initial step involves recognizing words or phrases that pose comprehension challenges).
2) Remove Inflections for Search. (If the unknown word is inflected, stripping away the inflection allows for a renewed search using the base word form).
3) Alphabetical Order Search. (Conduct a search for the unknown word in alphabetical order within the dictionary).

4) Advanced Strategies for Unidentified Words. If unable to identify a basic form, apply the following strategies:
   - Explore basic elements if the word resembles a fixed expression, idiom, or phrase.
   - Identify the stem if the word appears to have a suffix.
   - Address incorrect declined forms or spelling variants by examining nearby words and using dictionary applications.
5) Shorten Multiple Meanings or Homographs. (If a word has multiple meanings or is a homograph, condense and differentiate the possibilities).
6) Choose Appropriate Definition and Integrate. (Select the most relevant definition and incorporate it into the context where the unknown word was encountered).
7) Infer Meaning if None Match. (If no found meanings align, infer a potential meaning. If multiple meanings are applicable, seek additional context clues in the source text for clarification) [10].

These steps encompass a variety of strategies, with each step requiring distinct approaches. For example, in the initial step, readers employ complex strategies to identify new words or phrases. In instances where understanding an idiomatic phrase proves challenging, even when the individual words are understood, the reader may need to explore multiple dictionary entries to comprehend the phrase's intended meaning. The iterative nature of these steps reinforces a persistent and methodical approach until the desired understanding is achieved.

However, another scholar, Wingate, suggested that the steps described by Scholfield do not offer a comprehensive algorithm, since these steps are only for those involved in reading. He brings forth a valuable consideration regarding the nuanced nature of dictionary use across different language skills. While Scholfield's steps primarily cater to the needs of readers, Wingate contends that a comprehensive algorithm must encompass additional steps when the purpose shifts to writing. This acknowledgment underscores the multifaceted role dictionaries play in language acquisition. Wingate asserts that the steps proposed by Scholfield, while effective for readers, may fall short of providing a comprehensive algorithm for individuals engaged in writing activities. Writing involves distinct linguistic demands, necessitating a more elaborate set of steps within the dictionary utilization process [13].

Bishop's compilation of recommendations on the use of bilingual dictionaries for learning a foreign language represents a valuable guide tailored to both written exams and personal language study. This comprehensive set of guidelines aims to enhance the effectiveness of dictionary utilization in diverse language learning contexts. Here's a concise representation of Bishop's contribution, the algorithm consists of 10 steps:

1) Reviewing Dictionary Contents. A preliminary step involves exploring the diverse types of information contained within the dictionary, laying the foundation for effective use.
2) Identifying Part of Speech. Learners are advised to discern the part of speech of the desired word, accompanied by an understanding of symbols denoting verbs, nouns, and other grammatical nuances.
3) Synonym Exploration. Utilizing the dictionary to look up synonyms emerges as a strategy to expand and enrich one's vocabulary repertoire.
4) Specialized Vocabulary Accumulation. For academic pursuits, particularly in paper writing, the dictionary becomes a tool for accumulating specialized vocabulary, enhancing linguistic precision.
5) Grammar Information Utilization. Emphasis is placed on understanding and effectively applying the grammar information provided in the dictionary to construct grammatically accurate sentences.
6) Phonetic Symbol Proficiency. A crucial skill involves learning to read and interpret the phonetic symbols listed in the dictionary, enhancing accurate pronunciation.
7) Understanding Contextual Nuances. Mastery extends to discerning formal/informal and written/spoken contextual nuances, ensuring appropriate language use.
8) Cross-Referencing in Bilingual Dictionaries. Learners are encouraged to cross-reference bilingual dictionaries, checking translations in both directions to ensure accurate understanding.
9) Post-Writing Proofreading. A pivotal step involves proofreading written work after completion, refining language expression for clarity and coherence.
10) Dictionary as Aid, Not Substitute. The algorithm concludes with a reminder that while dictionaries are indispensable aids, they cannot replace the necessity for a confident mastery of one's vocabulary.

Bishop's algorithm provides learners with a structured and holistic approach to leverage dictionaries effectively, promoting nuanced language acquisition and proficiency.

Obviously, the divergence between Bishop's recommendations and Scholfield's algorithm highlights the varied nuances of dictionary use in language learning, particularly when considering different skills and types of dictionaries. Bishop's emphasis on vocabulary skills geared towards writing and the specific application to bilingual dictionaries underscores the need for tailored guidance.

Wingate's strategy provides a nuanced approach, categorizing strategies into vocabulary-oriented, language-oriented, and meaning-focused. This classification acknowledges the multifaceted nature of language learning activities and the diverse requirements of reading [13]. However, Wingate acknowledges the limitations of her research, calling for further exploration into other language activities, such as writing.

In light of these findings, it is clear that dictionary use in language learning is a complex process influenced by a variety of factors. The integration of technological advancements and the consideration of learners' preferences and needs are essential in developing effective dictionary use strategies. This complexity underscores the importance of tailored guidance in dictionary selection and use, highlighting the role of applied lexicography in enhancing language learning outcomes.

3 THE IMPORTANCE OF RECORDING AND KEEPING A VOCABULARY JOURNAL

The acknowledgment of the diversity in language activities and the varied purposes of learning necessitates a flexible approach in developing algorithms. Students can choose and adapt methods that align with their specific learning objectives and preferences. The proposal of maintaining an individual vocabulary journal, formatted as a table, aligns with the personalized and adaptable nature of language learning. Such a method allows students to systematically record and review new words, reinforcing the learning process across different language activities.

Table 1: An example of an individual vocabulary journal [3].

Word	Part of Speech	Definition	Example Sentence	Source / Context	Personal Notes / Collocates
acquire	verb	to gain or obtain something	The students aim to acquire new language skills through immersive experiences.	Reading - Article about language learning	Remember to use 'acquire' in the context of learning.
					to acquire competence, expertise, knowledge, skill and understanding
wanderlust	noun	a strong desire to travel to different places	Her endless wanderlust led her to discover hidden gems around the globe.	Reading - short abstract about traveling	Consider using it in contexts that highlight the passion and adventurous spirit associated with the desire to wander
					Satisfy wanderlust, Embrace wanderlust, Wanderlust-filled journey
adaptable	adjective	Capable of adjusting or changing easily to different conditions.	Our family is incredibly adaptable, seamlessly navigating through life's challenges and changes.	Vocabulary Work on the topic Family	Consider using it to highlight the flexibility and resilience within the family unit.
					Adaptable nature, Family adaptability, Cultivate adaptability

The journal, presented in a structured table format, encapsulates essential components aimed at fostering a comprehensive understanding of each encountered word (Table 1). Each entry includes:

1) Word: The target vocabulary word.
2) Part of Speech: The grammatical category of the word (e.g., noun, verb, adjective).
3) Definition: The meaning of the word.
4) Example Sentence: A sentence demonstrating the word's usage in context.
5) Source/Context: Where the word was encountered, providing additional context.

Personal Notes / Collocates: Any personal comments or reminders for the learner regarding the word's usage.

When delving into a new language, it's common to face initial confusion. To comprehend how to use specific words, begin by creating at least one example sentence. While the illustrations provided in this article target English learners, this methodology seamlessly extends to the acquisition of other foreign languages. The essence of learning a new language fundamentally commences with acquiring unfamiliar words [3].

Incorporating algorithms and technology into foreign language classes can significantly enhance the learning experience for students. Here are several algorithms and techniques that can be used with dictionaries in foreign language classes [11]:

1) Flashcard Algorithms:
   - Spaced Repetition: Use algorithms like Leitner system or SuperMemo to schedule flashcards for review at optimal intervals. This helps students memorize vocabulary effectively.
   - Algorithmic Sorting: Sort flashcards algorithmically based on difficulty level or the frequency of mistakes, ensuring students focus on challenging words more frequently.
2) Pronunciation Assistance:
   - Phonetic Algorithms: Implement algorithms to match and compare phonetic representations of words, helping students with correct pronunciation.
   - Speech Recognition: Utilize speech recognition algorithms to assess students' pronunciation accuracy and provide instant feedback.

3) Language Translation:
- Machine Translation: Integrate machine translation algorithms to provide instant translations of words, phrases, or sentences. Google Translate API, for instance, can be used for this purpose.
- Contextual Translation: Algorithms can analyze the context of a sentence to provide more accurate translations, teaching students the nuances of the language.

4) Grammar and Syntax:
- Parsing Algorithms: Algorithms that analyze the structure of sentences can help students understand the grammar and syntax rules of the language.
- Error Detection Algorithms: Implement algorithms that detect common grammatical errors in students' writing or speaking exercises, offering targeted feedback.

5) Vocabulary Expansion:
- Word Association Algorithms: Provide related words or phrases based on user input, encouraging students to learn words in context.
- Frequency Analysis: Algorithms can analyze texts and suggest commonly used words, helping students prioritize learning high-frequency vocabulary.

6) Cultural Context:
- Geolocation-based Context: Utilize algorithms that consider the geographical location of students to provide language learning materials relevant to their region or dialect.
- Cultural Relevance Algorithms: Recommend learning materials and topics based on students' interests and cultural background, making learning more engaging.

7) Interactive Learning:
- Chatbots: Implement chatbots powered by natural language processing (NLP) algorithms, enabling students to have interactive conversations in the foreign language [12].
- Gamification Algorithms: Introduce game-like elements with algorithms that adapt the game difficulty based on students' language proficiency, motivating them to progress [9].

8) Assessment and Feedback:
- Automated Assessments: Algorithms can automatically grade quizzes, essays, or speaking exercises, providing instant feedback to students and saving teachers' time.
- Sentiment Analysis: Analyze students' written or spoken responses with sentiment analysis algorithms to gauge their emotional tone and tailor feedback accordingly.

9) Collaborative Learning:
- Collaborative Filtering: Recommend study groups or conversation partners based on students' learning preferences and proficiency levels, encouraging collaborative learning.
- Peer Evaluation Algorithms: Implement algorithms for peer-to-peer evaluation, where students can assess each other's language skills, promoting active engagement.

10) Feedback Loop and Personalization:
- Learning Analytics: Use data analytics algorithms to track students' progress, identify learning patterns, and personalize the learning experience.
- Adaptive Learning: Employ algorithms that adapt learning materials and exercises based on individual students' strengths and weaknesses, ensuring personalized learning paths.

The traditional practice of maintaining a vocabulary journal, while seemingly straightforward, holds untapped potential for innovation in language acquisition strategies. The acknowledgment of the diversity in language activities and the varied purposes of learning necessitates not just a flexible approach but also a technologically enhanced one. Integrating algorithms and advanced data analytics into the process of keeping a vocabulary journal can revolutionize how students interact with new vocabulary, making this practice a cornerstone of personalized and adaptive language learning.

The proposal of enriching vocabulary journals with technology aligns with the personalized and adaptable nature of modern language learning. By employing algorithms that analyze recorded vocabulary in terms of usage frequency, context variability, and learning progress, students can receive tailored recommendations for further study, practice, and review. This method allows for a dynamic interaction between the learner and the journal, far beyond simple record-keeping. Each entry can become a data point in a larger linguistic analysis, fostering a comprehensive understanding of language usage, patterns, and progress over time.

Key enhancements to the vocabulary journal could include:
1) Automated Contextual Analysis: Leveraging natural language processing (NLP) to provide contextual insights and usage examples from a vast database of texts, making each vocabulary entry a launchpad for deeper learning.
2) Adaptive Learning Recommendations: Algorithms could analyze a student's interaction with the journal to suggest personalized review schedules, similar words to expand their vocabulary, or areas requiring additional focus.
3) Progress Tracking and Analytics: Incorporating learning analytics to visualize progress, identify trends, and adjust learning strategies accordingly.

Incorporating such technology into foreign language classes can significantly enhance the learning experience for students, aligning with several cutting-edge teaching methodologies and algorithms [14]. For instance, spaced repetition algorithms can be integrated into the journaling software to optimize review times for each word based on the forgetting curve, ensuring that students review vocabulary just as they are about to forget it, thereby maximizing retention.

By transforming the vocabulary journal from a static record into an interactive, algorithm-driven learning tool, educators can foster a more engaging, effective, and scientifically grounded approach to language acquisition. This innovative use of technology not only enhances the methodological novelty of the vocabulary journal but also aligns with contemporary research in applied linguistics and educational technology, marking a significant leap forward in personalized language learning.

4 RESEARCH METHODS AND TECHNIQUES

To delve deeper into the landscape of dictionary usage among English learners, we conducted comprehensive research among students at the State University "Zhytomyr Polytechnic." This empirical study aims to identify a student's need for a dictionary, the method of working with the dictionary, the frequency of using the dictionary and which dictionary is in constant use by a student. The insights gained from this research contribute valuable perspectives to the ongoing discourse on language acquisition and dictionary utilization.

1) Usage of dictionaries. Examining the survey responses on the use of dictionaries among 300 students of the 1st-3rd years, we received a complex vision of their language learning practices. A significant 66% of students affirmatively report using dictionaries, showcasing a prevalent reliance on these tools to support their language acquisition journey.

Conversely, 27% of students indicate that they do not use dictionaries, signaling a noteworthy minority who may adopt alternative strategies or resources for language learning.

A smaller subset, comprising 4%, occasionally turns to dictionaries, highlighting a small yet existing use of these resources.

Furthermore, 2% of students express a seldom use of dictionaries, suggesting infrequent reliance on these tools in their language learning routine.

An additional 2% reports using dictionaries only when no one else can provide the translation, indicating a situational reliance on dictionaries for specific language challenges.

The majority of students actively embrace dictionaries as integral tools in their language learning journey. However, a notable minority opts for alternative approaches or limits their dictionary usage to specific situations. These findings provide educators and language learning platforms with valuable insights to tailor resources and support systems that align with the diverse preferences and habits of students in their language learning endeavors.

2) Purpose. Analyzing the diverse purposes for which students use dictionaries sheds light on their specific language learning needs (Figure 1). A significant 67% primarily use dictionaries to look for translations, underscoring the foundational role of translation in language acquisition.

A smaller yet substantial portion, 9%, seeks both translation and pronunciation, indicating an awareness of the importance of mastering both the written and spoken aspects of language.

Another 8% focuses on definitions, emphasizing a desire for in-depth understanding and clarity in the meaning of words.

A subset of 5% utilizes dictionaries to find synonyms or antonyms, showcasing a commitment to expanding their vocabulary range.

Similarly, 5% of students turn to dictionaries for examples of usage, highlighting a contextual approach to language learning.
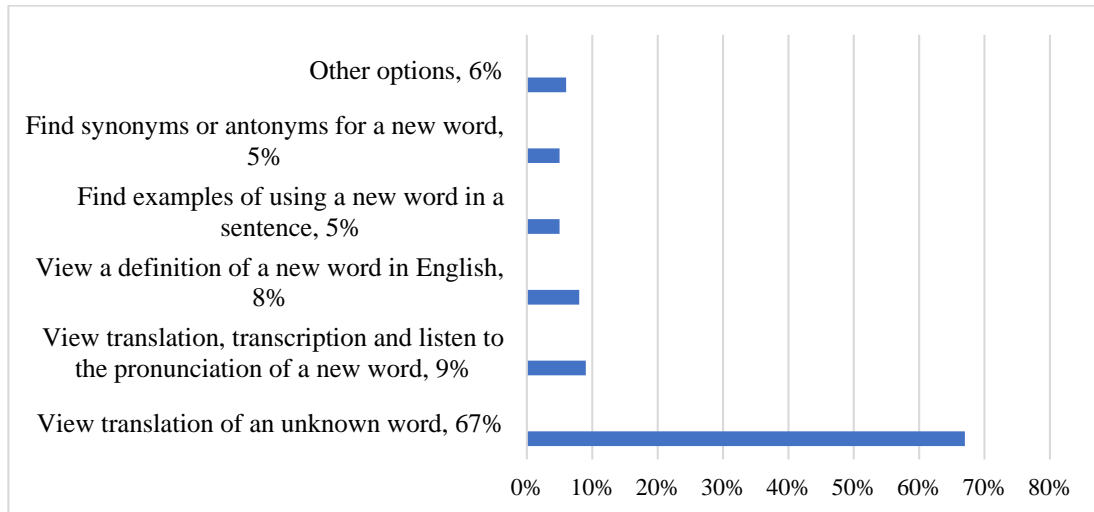
Figure 1: For what purposes students turn to dictionaries (sourse: own research).

A minimal 2% adapts their dictionary use depending on the specific need, showcasing a strategic and flexible approach to language challenges.

An additional 2% acknowledges the value of all mentioned components, suggesting a comprehensive utilization of dictionaries.

A mere 1% expresses a preference for almost everything, indicating a holistic reliance on dictionaries for various language learning aspects.

Contrastingly, 1% almost never uses dictionaries, signaling a unique learning approach that minimizes reliance on these tools. In summary, the varied purposes for dictionary usage among students reflect the diverse facets of language learning. Educators and language learning platforms can leverage these insights to tailor resources that align with the specific needs and preferences of students, fostering a more targeted and effective language learning experience.

3) Algorithm for using dictionaries. Analysis of the survey results reveals that a significant majority of students, 56%, do not have their own algorithm for using dictionaries. This majority suggests a potential gap in the development of individualized language learning methodologies among the student population. Conversely, 44% of respondents affirmatively report having their own algorithm, showcasing a diverse range of approaches in optimizing dictionary utilization. These results underscore the importance of understanding and addressing the varied needs and preferences of students when it comes to integrating dictionaries into their language learning journey.

4) Prevalence of the dictionary types. Analyzing the data on the types of dictionaries students use provides valuable insights into their preferred tools for language learning. The majority, comprising 57%, leans towards the convenience of Google Translator, indicating a prevalent reliance on online translation services. Furthermore, 20% of students opt for the indirect approach of using Google Search to identify suitable dictionaries, showcasing a resourceful strategy in dictionary selection.

A notable 23% of students demonstrate a diversified approach, distributing their preference across renowned dictionaries such as Longman, Oxford, Cambridge, Collins, as well as digital platforms like Multitran, Reverso, and Deepl. This diversity suggests a nuanced user base with varying preferences, possibly influenced by the specific features and strengths offered by each dictionary.

The prevalence of translation-centric preferences underscores the practical and immediate language learning needs of the student population. This insight is crucial for educators and language learning platforms to tailor resources that align with the specific requirements of students, ensuring a more targeted and effective learning experience.

Analyzing the data on students' dictionary preferences in terms of online and printed resources reveals a clear trend in favor of digital tools. The overwhelming preference for online dictionaries among 72% of students underscores the digital transformation in language learning.

This substantial portion reflects the convenience, accessibility, and real-time nature of online resources, aligning with the fast-paced and tech-driven learning environment.

Furthermore, 26% of students demonstrate a flexible approach, choosing either online or printed dictionaries based on the situation. This indicates an adaptable learning strategy, where students may leverage the advantages of online dictionaries for quick reference but turn to printed dictionaries in specific contexts.

A mere 2% of students exclusively rely on printed dictionaries, signaling a minimal preference for traditional, tangible resources in the digital age. This group may value the tactile experience or specific features offered by printed dictionaries that cater to their learning preferences.These insights are crucial for educators and educational platforms to tailor language learning resources that align with the diverse preferences and needs of students in the modern learning landscape.

5) Usage of individual vocabulary journals. Examining the responses on the presence of individual vocabulary journals among students provides valuable conclusions into their study habits. A majority of 53% indicate that they do not maintain an individual vocabulary journal. This suggests that over half of the student population may not engage in the practice of systematically recording and reviewing new words.

Conversely, 38% of students affirmatively report having their individual vocabulary journal. This dedicated group recognizes the value of maintaining a personal record to enhance their vocabulary acquisition and retention. A smaller subset of 9% admits to either never having, sometimes having, or rarely having an individual vocabulary journal. This minority may not consistently incorporate this practice into their language learning routine.

In essence, while a substantial portion of students do not currently maintain individual vocabulary journals, there exists a notable segment that actively recognizes and employs this effective language learning strategy. This information is pivotal for educators and language learning platforms to encourage and support the integration of vocabulary journals into the study routines of students, fostering a more comprehensive and personalized approach to language acquisition.

6) Aspects students are looking for. Examining students' preferences for aspects beyond translation in language learning reveals diverse interests and priorities. A significant 39% express an interest in transcription or pronunciation, showcasing a recognition of the importance of mastering the auditory and phonetic dimensions of language. Another substantial group, constituting 28%, values example sentences, indicating a desire for contextual understanding and application of newly acquired words.

A smaller subset of 12% expresses interest in definitions, emphasizing a focus on comprehending the nuanced meanings and applications of words.

Surprisingly, only 15% show interest in translations exclusively, suggesting a lesser emphasis on native language equivalents.

The granular aspects of language, such as part of speech and collocations, attract the interest of 1% each. These individuals demonstrate a keen interest in delving into the structural and contextual intricacies of words.

A minority of 4% expresses a holistic interest in everything, underlining a comprehensive approach to language learning that encompasses various facets.

In conclusion, the survey findings illuminate the multifaceted nature of students' interests in language learning. Educators and language learning platforms can leverage this information to tailor resources that align with the diverse preferences and learning styles of students, fostering a more engaging and effective language learning experience.

5 CONCLUSIONS

The journey of vocabulary acquisition is foundational to achieving fluency in a foreign language, encompassing not merely the understanding of word meanings but also their adept utilization in varied contexts. This journey, traditionally underemphasized in favor of communicative competence, has seen a paradigm shift with the recognition of dictionaries and translation tools as indispensable assets in language education. Our investigation into the practices and perceptions of 300 students regarding dictionary use has unveiled a landscape rich in diversity, reflecting a broad spectrum of strategies that students employ in their quest for language mastery.

Notably, the ascendancy of online dictionaries and platforms like Google Translator signifies a pivotal move towards digital resources, underscoring the need for educational strategies to evolve in tandem with technological advancements. This shift not only highlights the predominance of translation in language learning but also signals the potential for incorporating other critical aspects of language use, such as pronunciation, definitions, and contextual applications, into learning strategies.

The underutilization yet apparent benefits of maintaining an individual vocabulary journal, especially when enhanced with algorithms for personalized learning and linguistic analysis, point towards an untapped resource in language education. Such innovations can transform traditional methods into dynamic, interactive tools that cater to the individual needs of learners, thereby fostering a more engaging and effective learning experience.

Moreover, the feedback loop created by integrating advanced technologies and adaptive learning algorithms into language learning resources represents a significant advancement. It allows for a more nuanced and responsive approach to language education, one that accommodates the diverse needs, preferences, and practices of students. By embracing these technological innovations, educators can offer more personalized, efficient, and impactful language learning experiences.

In essence, the future of language education lies in leveraging technology to create adaptive, personalized learning environments. This approach not only enhances vocabulary acquisition but also ensures a more holistic and profound mastery of languages. As we move forward, the key to success in language learning will increasingly reside in our ability to harness the potential of digital tools and resources, making the learning journey not only more effective but also more aligned with the digital age.

REFERENCES

[1] A. Dziemianko, "Paper or electronic. The role of dictionary form in language learning," ReCALL, 22(3), pp. 338-357, 2010.

[2] A. Lounas, "Dictionary Use strategies as an effective way to learn vocabulary of English as a Foreign Language," Revue El-Bahith, pp. 5-17, 2017.

[3] A. Smith, B. Jones, and C. Johnson, "The Impact of Dictionary Types on Foreign Language Acquisition," Journal of Language Education Research, 25(3), pp. 123-145.

[4] B. Laufer and M. Hill, "What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention?" Language Learning and Technology, 3(2), pp. 58-76, 2000.

[5] D. Mutlu, "How to Use Dictionaries Effectively When Learning a Foreign Language," 2021, [Online]. Available: https://www.redefinerswl.org/post/how-to-use-dictionaries-effectively-when-learning-a-foreign-language.

[6] Enyota Learning, "Natural Language Processing (nlp) in eLearning: Improving Learner Interaction," 2023, [Online]. Available: https://enyotalearning.com/blog/natural-language-processing-nlp-in-elearning-improving-learner-interaction/.

[7] G. Bishop, "Using Quality and Accuracy Ratings to Quantify the Value Added of a Dictionary Skills Training Course," Language Learning Journal 24, pp. 62-69, 2001.

[8] K.J.Tsai, "Effects of Dictionary Use on Second Language Vocabulary Acquisition: A Meta-Analysis," International Journal of Lexicography, 32(4), pp. 418-444, 2019.

[9] M. Duffin, "What Is Natural Language Processing?," A Comprehensive Guide, 2023, [Online]. Available: https://rareconnections.io/what-is-natural-language-processing/.

[10] P. Scholfield, "Using the Dictionary for Comprehension," TESOL Quarterly 16.2, pp. 185-194, 1982.

[11] S. Hajmi, "The importance of dictionaries in learning a foreign language," 2022, [Online]. Available: https://hozir.org/the-importance-of-dictionaries-in-learning-a-foreign-language.html.

[12] S. Medewar, "The Role Of Natural Language Processing In eLearning," 2023, [Online]. Available: https://elearningindustry.com/the-role-of-natural-language-processing-in-elearning.

[13] U. Wingate, "Dictionary use – the need to teach strategies," Language Learning Journal, 29 (1), pp 5-11, 2004.

Bert Embedding and Scoring for Scientific Automatic Essay Grading

Abeer Abdulkarem and Anastasia Krivtsun

Department of Information Technologies and Management, Platov South-Russian State Polytechnic University (NPI),
Prosveshchenie Str. 132, 346428 Novocherkassk, Russia
abeerabdulsalam15@gmail.com, anastasia.srstu@gmail.com

Keywords: Automatic Essay Grading, Word Embedding Techniques, BERT Techniques, Neural Network.

Abstract: The educational landscape is experiencing a surging demand for Automated Essay Grading (AEG), prompting the need for innovative solutions. This paper introduces a cutting-edge methodology that harnesses the power of Bidirectional Encoder Representation from Transformers (BERT) to embed and score essays in the scientific AEG domain. Tackling challenges such as Out-of-Vocabulary (OOV), BERT's contextual embedding proves instrumental. The study meticulously evaluates a hybrid architecture on a prototype incorporating non-English essay answers, establishing a benchmark against state-of-the-art studies. Beyond the expeditious grading of essays, particularly in scientific realms, this paper makes a substantial contribution to the ever-evolving field of educational technology. The AEG task revolves around the automation of essay response grading, where input data encompasses essay answers, and output data comprises assigned scores. The adopted mathematical model seamlessly integrates BERT for contextual embedding and subsequent scoring. The evaluation uncovers compelling results, underscoring the effectiveness of the proposed BERT-based model. The model's architecture, characterized by bidirectional layers and a dense output, encompasses a notable 2,243,401 parameters. Significantly, the Kappa Score achieved by the model impressively stands at 0.9725, highlighting its superiority over existing methodologies.

1   INTRODUCTION

In the dynamic landscape of educational assessment, the advent of Automatic Essay Grading (AEG) stands as a transformative force, significantly impacting the evaluation processes within esteemed institutions such as the Educational Testing Service (ETS) [1]. The imperatives of efficiently and accurately appraising a substantial volume of student assignments have driven the adoption of machine grading systems. Noteworthy, standardized examinations, including SAT, TOEFL, and GRE, have seamlessly integrated machine grading methodologies, and industry giants such as Pearson.org and ETS.org have pioneered the development of proprietary AEG systems, streamlining the assessment of a myriad of student essays with unprecedented efficiency [2].

Traditionally, AEG methodologies have leaned on manually crafted attributes, a paradigm increasingly complemented by the integration of Deep Learning (DL) techniques, exemplified by the utilization of Recurrent Neural Networks (RNNs).

However, these approaches encounter a common limitation an inherent reliance on finite datasets for model training, hindering their capacity to comprehensively discern the nuanced contextual intricacies prevalent in well-articulated essays. Furthermore, conventional word embedding models based on lookup tables confront the formidable challenge of capturing grammatical correctness, an indispensable factor in ensuring precise essay scoring [3].

The earliest research efforts on AEG depicted were concentrating on determining structural features such as number of paragraphs, sentences, words, spelling and grammatical mistakes [4]. Apparently, the assessment of an essay answer using only the previously mentioned structural criteria would seem insufficient. It is necessary to examine the morphological and semantic aspects of the answer in order to give accurate scoring. Therefore, afterward researches have started to include lexical analysis, some sort of semantic analysis using external knowledge sources such as dictionaries, lexicons or ontology, or morphological. However, the emergence

of word embedding techniques which are based on neural network architectures has contributed toward a revolutionary progress in terms of the semantic analysis. Most text analysis/mining tasks such as document classification, topic modeling, question-answering and even AEG task have begun to take the advantage of word embedding due to its capability of determining the vector of a particular word in multiple dimensions where its lexical, semantic and syntactic aspects can be captured. One of the most commonly used datasets in essay grading is the Automatic Student Assessment Prize (ASAP).

Word embedding is the task of processing a series of tokens/terms through a neural network architecture in which the output of such a network is the prediction of consequent term. During the neural network training, a vector embedding starts to be emerged in the hidden layer of the network. Such a vector would consist of multiple values that indicate the position of the word, its lexical and semantic perspective. Figure 1 shows the traditional word embedding architecture known as Word2Vec. Word2Vec is a technique in natural language processing (NLP) for obtaining vector representations of words [5].
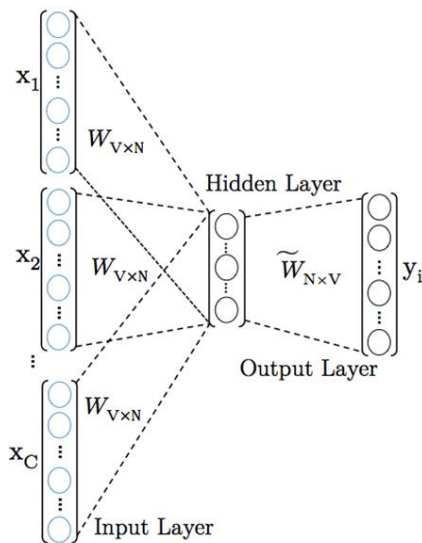

Figure 1: Word2Vec architecture.

A remarkable issue behind using Word2Vec is that it needs a large amount of text to train on in order to acquire better vector representation. For this purpose, have used a pretrained architectures of Word2Vec and GloVe for an AEG task. GloVe is also a very popular unsupervised algorithm for word embeddings that is also based on distributional

hypothesis – "words that occur in similar contexts likely have similar meanings". Those architectures had been trained on vast amount of text and saved for future usage. After turning the template and students answers into vector representation, the authors have applied a regression analysis. The same ASAP dataset has been used and results of accuracy was 56% [6].

In the same regard, it have also utilized a pretrained word embedding model of GloVe in order to conduct AEG task. A mapping process has been performed to turn the answers' text into the GloVe representation. Then, a regression analysis has been conducted on ASAP dataset and the acquired accuracy was 77%. [7].

It also have been proposed an enhanced word embedding representation known as Bag of Super Word Embedding. This representation aims to provide embedding vectors based on significant terms. In other words, it focuses on specific important terms then, it gives relevant embedding to each term based on its relevancy to the pre-specified terms. Using a regression analysis based on ASAP dataset, the proposed method showed an accuracy of 78% [8].

ELMo is another popular word embedding framework. It was developed at Allen Institute for AI. It addresses the fact that the meanings of some words depend on the context. ELMo does not produce a fixed vector representation for a word. Instead, ELMo considers the entire sentence before generating embedding for each word in a sentence. This contextualized embedding framework produces vector representations of a word that depends on the context in which the word is actually used. ELMo uses a deep bi-directional language model giving the model a better understanding of not only the next words, but also the preceding ones [9].

Lastly, it have been proposed an AEG method based on Word2Vec representation and an advance prediction approach known as Gated Recurrent Unit (GRU). After replacing the answers' text into its Word2Vec representations, the proposed GRU has been applied to predict the score. Using ASAP dataset, accuracy result was 86% [10].

The state of the art in AEG task is mainly focusing on utilizing word embedding techniques. The majority of literature concentrated on traditional Word2Vec or pretrained Word2Vec and GloVe. However, these word embedding techniques suffer from a remarkable issue known as Out-of-Vocabulary (OOV).

This problem occurs when a word embedding model is trained on set of text tokens and later encounter a word that has no embedding vector on its model. In other words, such an unseen word would have no presence in the training text. In that case, researchers attempt to avoid this problem by normalizing a fixed vector (usually full of zeros) to substitute the absence of vector embedding for OOV words. In addition, the traditional word embedding techniques deal with text in a word-level rather than sentences or paragraphs. Since the AEG task is mainly depending on sentences or paragraphs answers thus, it would pose another problem.

On the other hand, most of the researches in AEG task have utilized datasets that are related to second language tests such as the ASAP. These datasets contain essay answers related to general topics. Yet, not all exams would have general topic essay answers. Some subjects especially related to science require a domain specific answer to a particular question. Customizing the AEG task to a particular scientific domain of interest would facilitate mastering its questions and answers.

2 RESEARCH METHODOLOGY

The methodology adopted for the Automated Essay Grading (AEG) task encompasses a two-fold approach aimed at addressing critical challenges in the existing landscape. This section provides a comprehensive overview of the proposed methodology's key components and phases,

A crucial facet in the landscape of Automated Essay Grading (AEG) research is the selection of datasets, a factor that significantly influences the efficacy and generalizability of proposed methodologies. Notably, the majority of AEG studies have traditionally gravitated towards datasets associated with second language tests, exemplified by widely used benchmarks like ASAP.

While these datasets offer valuable insights and facilitate a comprehensive understanding of AEG challenges, they predominantly comprise essay answers related to general topics. However, the diverse array of academic disciplines implies that not all examinations yield responses aligned with generalized themes. In particular, scientific subjects demand a domain-specific approach, where questions

are tailored to assess knowledge and comprehension within a particular scientific domain.

Customizing the AEG task to a specific scientific domain holds the promise of enhancing the relevance and accuracy of grading. This involves formulating questions within the chosen scientific domain, constructing model answers, and subsequently testing students.

The manual assessment of student responses by teachers provides labeled scores, which undergo statistical analysis to derive average scores. The culmination of this process yields a meticulously curated dataset comprising questions, model answers, and labeled student responses.

In this subsection, we scrutinize the prevalent practice of relying on generic datasets and emphasize the need to tailor AEG research to specific scientific domains. The ensuing sections detail the proposed methodology's dual phases, addressing the OOV challenge and sentence-level embedding problems using Bidirectional Encoder Representation from Transformers (BERT) on benchmark datasets like ASAP and a newly developed scientific dataset.

BERT was released with two versions: BERT-base and BERT-large, and each has a cased and uncased iteration (there is also a Chinese BERT for Chinese and a Multilingual BERT that was originally trained on 102 different languages). BERT-base has twelve layers (Transformer blocks), each with twelve self-attention heads and 768 hidden neurons. BERT-base consists of approximately 110 million parameters: approximately 24 million from the embeddings, 85 million from the transformers, and one million from the pooler (BERT-large, comparatively, has roughly 340 million parameters). BERT-large is over three times as large, but it has not been shown to outperform BERT-base by an equally significant margin. BERT-base's architecture can be seen in Figure 2, where twelve encoders are stacked sequentially. Each encoder is a transformer with its own attention heads. It is also important to note that only the encoder portion of the transformer (shown) is included in BERT's architecture, as BERT is not a generative model and does not implement a decoder [11].

The application of Bidirectional Encoder Representation from Transformers (BERT) in Automated Essay Grading (AEG) represents a pivotal aspect of this study, leveraging cutting-edge techniques for enhanced performance. BERT's model

architecture, inspired by Vaswani et al., stands as a beacon of innovation, purposefully designed to navigate the intricacies of scientific text analysis [12].
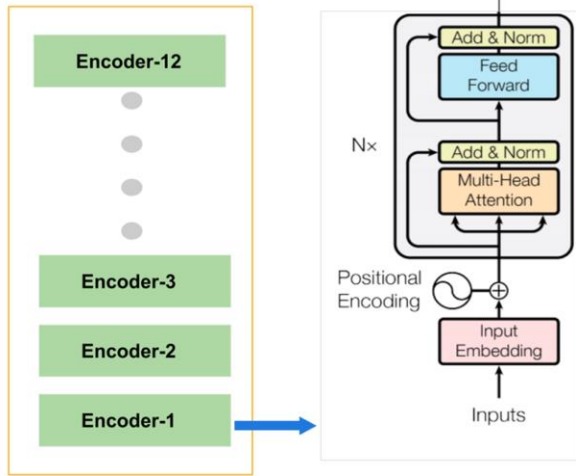

Figure 2: BERT architecture.

The methodology of this study consists of two main phases as shown in Figure 3.

The first phase represents the application of the proposed BERT embedding that is intended to solve both OOV and sentence-level embedding problem for the AEG task. For this purpose, the benchmark dataset of ASAP that has been widely examined by the literature, will be used in this phase. After applying the BERT embedding and scoring, an evaluation task will take a place in order to assess the automatic scoring produced by the proposed BERT method and comparing its results against the state-of-the-art. Once, the proposed method demonstrated superior results in terms of accuracy compared to the literature, the second phase will take a place.

The second phase represents the development of new scientific dataset for the AEG task. To do so, there are multiple procedures will be conducted. First, a set questions in specific domain of science will be initiated. Then, a model answer for each question will be formulated. Consequentially, the questions will be given to students in order to be tested. The tested answers produced by the students will be assessed by different teachers in order to give a manual score for each essay answer. The scores given by the teachers will be undergoing statistical analysis in order to take the average score. Lastly, a dataset of questions, model answers, and tested and labeled/scored answers.

Finally, the proposed BERT embedding and scoring will be applied on the dataset. The results of BERT scoring will be compared against teachers' scorings.
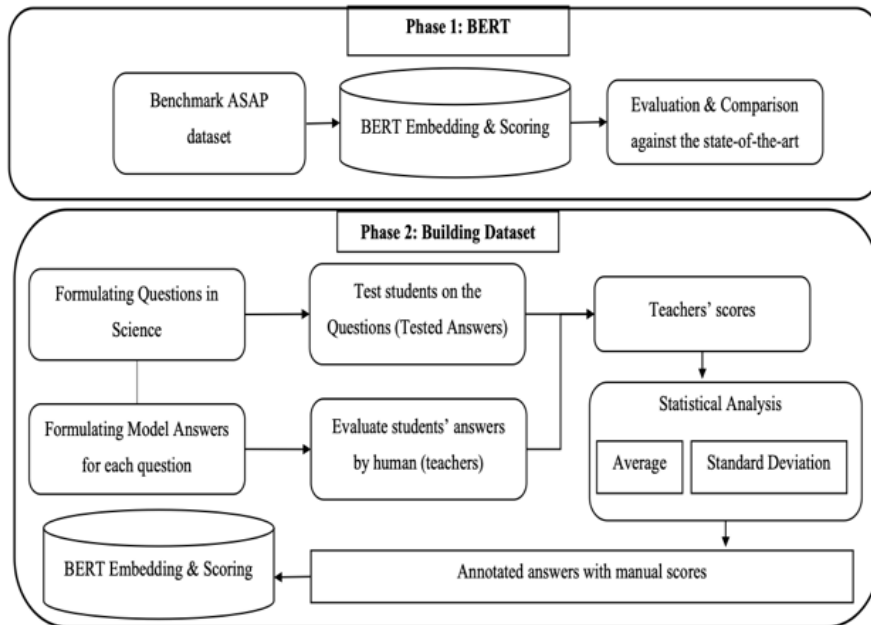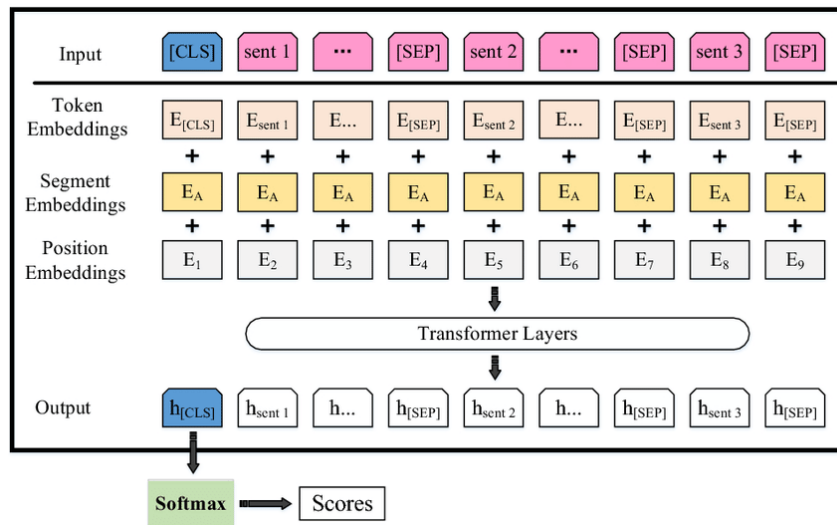

Figure 3: Methodology.

Figure 4: BERT's Input Representation Strategy.

To fully harness the capabilities of Bidirectional Encoder Representation from Transformers (BERT) in Automated Essay Grading (AEG), it's imperative to delve into the intricacies of BERT's architecture. BERT represents a paradigm shift in natural language processing, offering a sophisticated framework that aligns seamlessly with the nuances of scientific text analysis. BERT stands tall with its multi-layer bidirectional Transformer encoder. Departing from conventional models, this architecture introduces Transformers with bidirectional self-attention, a key element in capturing contextual dependencies with unparalleled sophistication. The versatility of BERT shines through in its adept handling of both single sentences and sentence pairs, providing a nuanced representation ideal for the complexities of scientific essays [13].

Notably, BERT employs Transformers with bidirectional self-attention, a departure from conventional models. This departure is significant, as it enables BERT to capture contextual dependencies with unparalleled sophistication.

BERT's prowess in handling a spectrum of downstream tasks is underpinned by its versatile input representation. It seamlessly accommodates both single sentences and sentence pairs, thereby accommodating the complex nature of scientific essays. Leveraging WordPiece embeddings with a 30,000-token vocabulary, BERT employs a special classification token ([CLS]) at the sequence's outset. The final hidden state corresponding to this token serves as the aggregate sequence representation for classification tasks [14].

A notable feature is the unambiguous representation of sentence pairs. Employing a special token ([SEP]) to separate sentences and adding a learned embedding to distinguish sentence A from sentence B BERT orchestrates a comprehensive input representation strategy, exemplified in Figure 4.

Pre-training, the inaugural phase in BERT's journey, unfolds over extensive unlabeled data using two unsupervised tasks. The first task involves Masked Language Modeling (MLM). Unlike traditional language models constrained by left-to-right or right-to-left conditioning, BERT employs a masked token approach. A percentage of input tokens are randomly masked, and the model predicts these masked tokens, paving the way for deep bidirectional understanding [11].

The second task, Next Sentence Prediction (NSP), addresses the intricacies of sentence relationships. Generating a binary next sentence prediction task, BERT establishes a foundational understanding of the contextual interplay between sentences. A noteworthy aspect is the synergy between these pre-training tasks, showcasing BERT's commitment to bidirectionality while mitigating mismatches between pre-training and fine-tuning [15].

BERT consists of two main architectures; language modeling and fine-tuning. The first architecture aims at process an answer sentences where each sentence is represented by its tokens along with two tags of 'CLS' and 'SEP' which refer

to the beginning and ending of a sentence respectively as shown in Figure 5. The output of this architecture is the language modeling where BERT would have the ability to understand the answer text.

The second architecture of BERT contains the fine-tuning where the processed answer in the previous architecture will be processed as an input to such an architecture. Fine-tuning, a streamlined process enabled by BERT's self-attention mechanism, unfolds seamlessly across a spectrum of downstream tasks. The model's inherent bidirectionality proves pivotal, allowing it to encode and comprehend text pairs efficiently.

In the scientific automatic essay grading context, this adaptability is invaluable. Fine-tuning involves plugging in task-specific inputs and outputs, thereby configuring BERT for tasks ranging from question answering to text classification. The bidirectional cross attention facilitated by BERT's self-attention mechanism elegantly unifies the encoding of text pairs, offering a versatile solution for the demands of scientific essay grading.
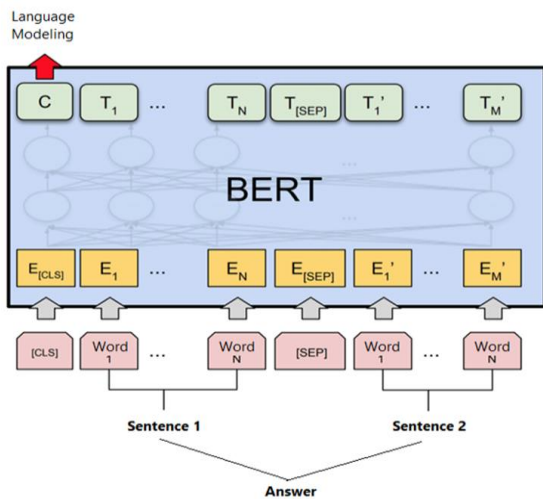

Figure 5: BERT language modeling.

The output of this architecture is the answer score prediction as shown in Figure 6. Afterwards., the evaluation metrics that intended to examine the performance of the proposed architecture will be applied. The common accuracy metric used for AEG task is the Quadratic Weighted Kappa (QWK) [11].

Such a metric is intended to calculate the agreement between human score and automatic score. It can be computed as follows:

$$QWK = \frac{p_o - p_e}{1 - p_e}, \qquad (1)$$

where p_o is the observed agreement and p_e is the agreement by chance between human and automatic rater.
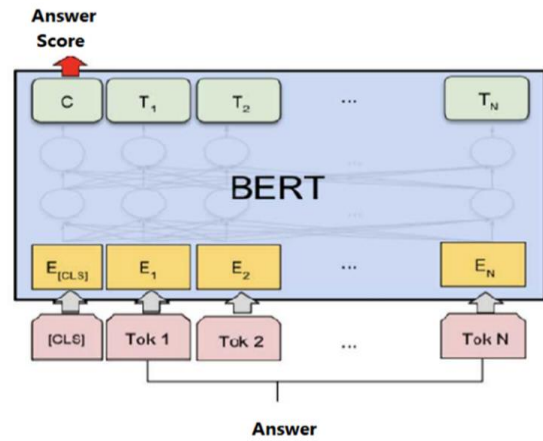

Figure 6: BERT fine-tuning.

Since this study is aiming to conduct the experiments on ASAP dataset which is a benchmark that contain scored answers by multiple teachers. Therefore, the evaluation would take a straightforward style where the score generated by the proposed BERT architecture will be compared directly with the original or actual score within the dataset.

After evaluating the scores produced by the proposed hybrid architecture, it is necessary to consider the state-of-the-art studies in the comparison. This is to point out the actual novelty or improvement depicted by the proposed hybrid architecture. Hence, all the state-of-the-art studies that have considered BERT embedding, bidirectional LSTM or CNN will be considered.

This evaluation paradigm will take a place on the proposed prototype where a collection of non English essay answers will be acquired. Consequentially, the proposed BERT architecture will be applied on such essays. Lastly, three teachers will be asked to score these answers and comparison between their scores and the proposed hybrid architecture's scores will be accommodated.

The model's architecture, characterized by bidirectional layers and a dense output, encompasses a notable 2,243,401 parameters. Significantly, the

Kappa Score achieved by the model impressively stands at 0.9725, highlighting its superiority over existing methodologies.

3 CONCLUSIONS

One of the significant processes within the education is the assessment of student's performance. Oral and MCQ examinations are the easiest parts to be automated where a web/mobile portal can be developed and facilitate the communications and exam conducting. However, the assessment of essay answers is still representing the most challenging task to be turned into a fully-automatic task.

The automation of essay answer grading requires a wide range of textual analysis including the lexical, morphological, semantic and syntactic aspects to train the computer to give a grade for a particular essay answer. Improving the AEG task would be a significant contribution to the educational process especially in the online manner where plenty of time spent by the instructor/teacher to give the grade will be saved.

In this research a new embedding technique based on Bidirectional Encoder Representation from Transformers (BERT) were proposed to overcome the OOV and sentence-level embedding problems in AEG task. The proposed method, incorporating BERT's contextual embedding, outperforms traditional approaches. Notably, the model architecture, featuring bidirectional layers and a dense output, comprises 2,243,401 parameters. The Kappa Score of 0.9725 attests to the model's exceptional performance.

The automation of essay grading, leveraging BERT's contextual embedding, stands as a breakthrough. The model, trained to analyze lexical, morphological, semantic, and syntactic aspects, significantly contributes to the educational process.

Secondly, this research study also curates a new question answering dataset in the science domain with the assistance of manual/human labelling scores as far as annotation is concerned. Improving the AEG task would be a significant contribution to the educational process especially in the online manner where plenty of time spent by the instructor/teacher to give the grade will be saved.

REFERENCES

[1] F. Li, X. Xi, Z. Cui, D. Li, and W. Zeng, "Automatic Essay Scoring Method Based on Multi-Scale Features," Applied Sciences (Switzerland), vol. 13, no. 11, Jun. 2023, doi: 10.3390/app13116775.

[2] L. Blecher, G. Cucurull, T. Scialom, and R. Stojnic, "Nougat: Neural Optical Understanding for Academic Documents," Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.13418.

[3] F. Nadeem, H. Nguyen, Y. Liu, and M. Ostendorf, "Automated Essay Scoring with Discourse-Aware Neural Models," 2019, [Online]. Available: www.smashwords.com.

[4] M.V. Koroteev, "BERT: A Review of Applications in Natural Language Processing and Understanding."

[5] Y. Farag, H. Yannakoudakis, and T. Briscoe, "Neural Automated Essay Scoring and Coherence Modeling for Adversarially Crafted Input," Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.06898.

[6] Y.-J. Jong, Y.-J. Kim, and O.-C. Ri, "Improving Performance of Automated Essay Scoring by using back-translation essays and adjusted scores." [Online]. Available: https://github.com/j-y-j-109/asap-back-translation.

[7] H. Zhang and D. Litman, "Co-Attention Based Neural Network for Source-Dependent Essay Scoring," Aug. 2019, doi: 10.18653/v1/W18-0549.

[8] C. T. Lim, C. H. Bong, W. S. Wong, and N. K. Lee, "A comprehensive review of automated essay scoring (Aes) research and development," Pertanika Journal of Science and Technology, vol. 29, no. 3. Universiti Putra Malaysia Press, pp. 1875-1899, 2021, doi: 10.47836/pjst.29.3.27.

[9] J. Liu, Y. Xu, and Y. Zhu, "Automated Essay Scoring based on Two-Stage Learning," Jan. 2019, [Online]. Available: http://arxiv.org/abs/1901.07744.

[10] D. Ramesh and S. K. Sanampudi, "An automated essay scoring systems: a systematic literature review," Artif Intell Rev, vol. 55, no. 3, pp. 2495–2527, Mar. 2022, doi: 10.1007/s10462-021-10068-2.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018, arXiv e-prints, page arXiv:1810.04805.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, et. al., "Attention Is All You Need," 2017, arXiv:1706.03762.

[13] K. Taghipour and H. T. Ng, "A Neural Approach to Automated Essay Scoring," [Online]. Available: https://www.kaggle.com/c/asap-aes.

[14] B.F. Dhini, A.S. Girsang, U.U. Sufandi, and H. Kurniawati, "Automatic essay scoring for discussion forum in online learning based on semantic and keyword similarities," Asian Association of Open Universities Journal, Dec. 2023, doi: 10.1108/AAOUJ-02-2023-0027.

[15] E. Mayfield and A.W. Black, "Should You Fine-Tune BERT for Automated Essay Scoring?," 2020, [Online]. Available: https://course.fast.ai.

Adaptive Clustering for Distribution Parameter Estimation in Technical Diagnostics

Galina Shcherbakova, Svetlana Antoshchuk, Daria Koshutina and Kiril Sakhno

National University Odesa Polytechnic, Shevchenko Avenue 1, 65044 Odesa, Ukraine
{ galina.sherbakova, asg, d.v.koshutina}@op.edu.ua

Keywords: ASTD, Automated Systems for Technical Diagnostics, Electronic Components, Reliability Parameters, Adaptive Clustering, Wavelet Transform, Noise Immunity, Rejection Systems, Exponential Distribution, DN Distribution.

Abstract: A novel approach has been introduced to estimate the parameters of exponential and DN distributions during the rejection testing of electronic devices, accompanied by a detailed procedure for its implementation. This innovative method enhances noise immunity and minimizes the error associated with the rejection process through the application of a clustering technique involving wavelet transform. The effectiveness of the method has been verified using resistors, employing criteria such as noise level and stability. The substantial improvement in noise immunity and the reduction in rejection procedure errors are achieved by incorporating an adaptive clustering method coupled with wavelet transform. Notably, in clustering with a signal-to-noise ratio by amplitude of 1.17, the relative error in determining the minimum of the test function was reduced to 8.32%. These promising outcomes substantiate the recommendation of the developed method for the automated selection of resistors, particularly those designated for long-term operational equipment with critical applications. The presented method thus contributes significantly to enhancing the reliability and accuracy of electronic device testing and selection processes.

1 INTRODUCTION

Automated Systems for Technical Diagnostics (ASTD) in the production of electronic components (EC) are utilized to address various tasks, including the assessment of reliability parameters. Given the current state of electronics characterized by increased multifunctionality and miniaturization of EC, information in such systems is presented as multidimensional arrays of correlated parameters. The application of ASTD for evaluating reliability parameters of complex EC is indispensable.

During EC production, it is often necessary to select more reliable components, requiring the choice of a reliability assessment model and estimation of its parameters. In cases where the impact of sudden failures needs evaluation, the assessment often involves a single parameter – the failure intensity of the exponential distribution λ. The accuracy of reliability parameter estimation is enhanced by employing two-parameter distributions, such as the two-parameter DN-distribution [1-4].

The duration of accelerated tests is significantly shorter than the actual durability of EC. Due to the high reliability of EC, these tests are conducted until failure occurs in a small portion of the sample, resulting in substantial error. The source of this error is the high variance of the distribution DN parameter estimates, further exacerbated by the noisy reliability data (objective and subjective errors in collection and recording).

In such studies, for instance, during accelerated tests of complex EC for durability, it was revealed that the failure intensity has a bimodal nature (the first peak characterizes failures of anomalous complex EC, and the second – failures of the main part of the sample); meanwhile, the mean time to failure for these groups may differ by four orders of magnitude [5]. In such situations, when screening potentially unreliable EC, it is necessary to enhance robustness and reduce the error in separating the batch of EC into two groups.

The state of EC during screening is described by the values of their parameters, which vary from object to object and over time during testing. When there are many correlated parameters, automated approaches

are applied for EC separation, implemented using one of the pattern recognition methods – statistical classification. Such classification consists of two stages: clustering and classification. Clustering methods are used to group EC into clusters with common properties and varying degrees of monotonic dependence of controlled parameters on operating time. Assessing the rate of change of cluster parameters is possible using an adaptive approach, where initial cluster center parameters are determined from the analysis at a previous time.

EC is divided into clusters based on the compactness of parameters through quality functional optimization. The optimization method is chosen considering the properties of this quality functional, which may not be explicitly known and may possess a multi-extremal, noisy surface due to screening being performed on small samples. Clustering methods based on gradient search do not provide sufficient robustness under these conditions. It is proposed to apply adaptive clustering using wavelet transformation (WT), allowing for low-error clustering in the presence of high data noise and small sample sizes [6].

2 PROBLEM STATEMENT

The purpose of this work is to develop a method for determining the parameters of the exponential and *DN* distributions for reliability estimation of a batch of electronic components (ECs) to improve noise immunity and reduce the error of automated rejection in ASTD using adaptive clustering with the use of wavelet transform (WP) and the procedure for implementing this method.

3 CHALLENGES IN SCREENING WITHIN ASTD

The creation and use of information technologies and automated systems for technical diagnostics in the production of electronic products are largely determined by both the capabilities of obtaining and processing the information necessary for diagnostics, as well as the availability of methodological foundations that allow making decisions based on this information. ASTD in electronics are used to solve various problems, including the estimation of reliability parameters. Since the current state of electronics is characterized by an increase in the multifunctionality and miniaturization of electronic

components (ECs), the information in such systems is presented in a complex form, in the form of multidimensional arrays of correlated parameters. The use of ASTD for complex ECs has no alternative. In the course of their operation, ASTD use procedures for the formation and processing of information that allow making a decision about the degree of compliance of the object's parameters with the requirements of normative technical documentation and, if necessary, to form the corresponding control actions during production. It is for this reason that reducing the error of the decisions made in ASTD, especially in the case of selecting reliable ECs, especially for applications where the cost of an incorrect decision is high, is a rather important task.

The scope of application of such systems for the selection of reliable ECs is limited by a complex of contradictions. On the one hand, the growth of the quality characteristics of information formation and display systems of ASTD is due to the high cost of their hardware (testing systems, precision mechanics, and optics). On the other hand, the rapid improvement of the element base and assembly technologies is the reason for the growth of the number of organizations focused on small-scale, pilot production, and prototype production, for which the use of such ASTD is limited for economic reasons. The reduction of the product life cycle in electronics requires reducing the time for control and diagnostic operations when selecting reliable ECs, which necessitates conducting such operations on small data samples, increasing their noise; the noise of data in ASTD also grows due to interference in communication channels, uneven temperature changes of the object or sensors, etc., and the decision-making procedures in ASTD based on gradient iterative algorithms of successive approximation are characterized by low noise immunity. In connection with the growth of the complexity of electronic equipment to ensure the safety systems of various industries, the requirements for its reliability are growing. On the other hand, the diagnostic decision-making procedures developed on the basis of noise-resistant iterative algorithms of successive approximation are characterized by high error. In addition, there are no methodological foundations that allow, on the basis of existing methods for choosing reliability distributions and estimating their parameters, to increase the degree of automation of the selection of reliable ECs.

Most often, in the case when it is necessary to estimate the reliability in the absence of running-in failures or when the aging phenomena are assumed to be insignificant, the reliability of ECs and devices

based on them is estimated using the exponential distribution. This approach simplifies the processing of experimental results, which is essential in cases where, for technical, economic reasons, or due to lack of time, it is impossible to conduct tests for complete samples of large volumes. For this distribution, the probability of failure-free operation over the given interval *(t, t + τ)* does not depend on the time of previous operation *t*, but only on the interval *τ*. That is, if it is known that the EC is in working order, then its future state does not depend on the past. This model takes into account, mainly, sudden failures of a random nature, while failures that occur as a result of irreversible physicochemical changes in the parameters of the EC do not obey the exponential law. Therefore, the use of the exponential failure model in the experimental evaluation and prediction of the reliability of ECs makes this estimate approximate and leads to significant errors. The exponential distribution has one parameter – the failure intensity λ, which does not change over time. The assumption of the constancy of λ over time increases the error of reliability forecasts compared to two-parameter distributions. Therefore, it is recommended to use the exponential distribution to conduct a relative reliability assessment at the stage of conceptual design, when it is necessary to evaluate the reliability of various options, and on the basis of their analysis, choose the element base.[3]

To identify and remove ECs with actual and potential failures from a finished batch before delivery to the customer, rejection tests are carried out at the final control. To assess the compliance of the selected ECs with the customer's requirements, their reliability parameters are usually evaluated in the process of accelerated censored tests. In the course of such tests, during the analysis of degradation processes leading to EC failures, it was revealed that they have a random nature, and the change of these processes has both monotonic and non-monotonic character. For example, complex ECs, such as integrated circuits (ICs), are simultaneously exposed to the influence of many factors. These factors, uncorrelated and correlated with each other, form the overall process of IC degradation. The parameters of degradation processes, exceeding which certain values can cause a failure of an IC component, have different physical nature: accumulation of dislocations, plastic and elastic deformations, fatigue mechanical destruction, electrochemical corrosion, generation and movement of charges on the surface of a semiconductor crystal, etc. [1, 3].

A comparative analysis of probabilistic-physical distributions that take into account the physical processes leading to failure is presented in [3]. Based on this analysis, the calculated reliability estimates of such complex ECs as ICs, based on the results of accelerated tests, have an error of no more than 10%, if the two-parameter diffusion DN distribution is used as a failure model, corresponding to a non-monotonic Markov process [1, 4]. In the case when it is not possible to establish the prevailing degradation processes leading to failures, it is also recommended to use this distribution for objects consisting of ECs [4].

4 ESTIMATION OF DISTRIBUTION PARAMETERS

The task of estimating distribution parameters in the reliability assessment of electronic components (EC) is as follows. Suppose there are N objects (EC), each characterized by a set of k parameters. Vector $x_j(t) = \left(x_j^1(t), x_j^2(t), \dots, x_j^k(t)\right)$ characterizes the state of the *j*-th object at time *t*. This means that the mutual arrangement of the set of points $x_1(t), \dots, x_N(t)$ in the k-dimensional parameter space X reflects the real classification (grouping by parameters over time) of the studied objects. To identify this structure, the cluster analysis with WT [6] is used.

Using this method, at the time of rejection t_1, the separation of n points in the X space into *2* classes (clusters) is performed. The concept of the cluster center $a_i(t)$, $i = 1, \dots, r$ is introduced.

At time t_2, each point $x_j(t_2)$ is classified into one of the classes obtained at the first step. Then, the values of the cluster centers $a_i(t_2)$, $i = 1, \dots, r$ are recalculated and the distances between the points $x_j(t_2)$ and the new centers $i = 1, \dots, r; j = 1, \dots, n$ are calculated. This procedure is performed for all m time points.

At the next stage, the failure intensity λ is estimated for each of the two groups [4]. The equivalent device hours (EDH) are calculated for each cluster of ECs. The failure intensity is defined as $\lambda = \frac{r}{EDH}$. Here, r is the number of ECs that failed during the test time.

Then, the parameters of the DN-distribution are estimated for each of the two groups [4, 8]:

$$F(t) = DN(t; \mu, \nu) = \Phi\left(\frac{t-\mu}{\nu\sqrt{\mu t}}\right) + \\ + \exp(2\nu^{-2})\,\Phi\left(-\frac{t+\mu}{\nu\sqrt{\mu t}}\right)$$

where μ is the scale parameter, which coincides with the expected value of the random variable t; v is the shape parameter, which is equal to the coefficient of variation of the distribution of the variable t.

The proposed method was tested on a sample of resistors [7]. The noise level and the expected value of the resistance change in the groups were used as predictive parameters. The data from the first control (after 24 hours of operation in the loaded mode) were divided into two clusters using adaptive clustering: the first cluster included groups from 1 to 8, and the second cluster included group 9, based on the noise level [8, 9].

The calculated failure intensities were $\lambda_1 = 0{,}35\times10^{-5}$ hour^{-1} in the absence of rejection, $\lambda_2 = 0{,}19\times10^{-5}$ hour^{-1} for resistors of the first cluster, and $\lambda_3 = 0{,}19\times10^{-5}$ hour^{-1} for resistors of the second cluster.

Furthermore, the coefficients of variation of the diagnostic parameter, the expected value of the resistance change, were calculated for the first, second cluster, and the entire sample of resistors without rejection, using the formula.

$$v_j = \frac{\sqrt{n}}{\sum_{i=1}^n \Delta x_{ji}} \sqrt{\sum_{i=1}^n \left(\Delta x_{ji} - \frac{1}{n}\sum_{i=1}^n \Delta x_{ji}\right)^2},$$

where $\Delta x_{ji} = x_{ji+1} - x_{ji}$, index j corresponds to the cluster number $j=1,2$ or $j=3$ for the entire sample of resistors; index i corresponds to the time point (operating time) $t_i = (i = 1, 2, \dots, n)$.

The intervals between measurements $\Delta t = t_{i+1} - t_i$ were taken to be non-uniform (the expected value of the resistance change was determined after 24, 168, 1000, 5000, and 10000 hours after the start of measurements) [8]. Further, according to the methodology [4], taking into account that the values of v_j were $v_1=1.38$, $v_2=1.43$, and $v_3=1.55$, the scale parameters are estimated by solving the equation.

$$F_j(t_i, \mu_{ji}, v_i) = \widetilde{F}_j(t_i), \tag{1}$$

where $\widetilde{F}_j(t_i) = \frac{i}{N}$, $(i = 1, \dots, r)$; i - failure number in the corresponding cluster $(i = 1, \dots, r)$; t_i – is operating time of the i-th failed resistor, since the exact time of failure is not available [8], the right boundary of the time interval during which it failed was considered as the moment of failure in the calculations.

The average estimate of the parameter $\overline{\mu}_j$ for each of the three groups of resistors was determined as

$$\overline{\mu}_j = \frac{1}{r}\sum_{i=1}^r \overline{\mu}_{ji},$$

where $\overline{\mu}_{ji}$ - results of solving (1) at $(i = 1, \dots, r)$.

As a result, the calculated values of the average operating time to failure using the DN distribution were 4,8 years in the absence of rejection, 7,5 years for the resistors of the first cluster, and about 2 years for the resistors of the second cluster.

5 CONCLUSIONS

A method has been developed for determining the parameters of the exponential and DN distributions for the rejection testing of electronic components (ECs) and a procedure for implementing this method. The method was tested on the example of rejecting resistors based on noise level and stability. The improvement in noise immunity and the reduction of the error of the rejection procedure is achieved by applying the method of adaptive clustering with the use of the VP (the relative error in determining the minimum of the test function during clustering with the signal-to-noise ratio by amplitude of 1.17 was 8.32% [6, 10]). This result allows us to recommend the developed method for the automated selection of ECs intended for long-term operating equipment, especially of responsible purpose.

REFERENCES

[1] E. Wyrwas, L. Condra, and A. Hava, "Accurate quantitative physics-of-failure approach to integrated circuit reliability", 13 p., 2011, Las Vegas, NV: IPC APEX Expo.

[2] A.K.S. Jardine and J.A. Buzacott, "Equipment reliability and maintenance. European Journal of Operational Research", vol 19(3), 285-296 p., 1985, doi: 10.1016/0377-2217(85)90124-9.

[3] V.P. Strelnikov and P.V. Strelnikov, "Defining the nomenclature of the spare parts sets and calculating the number of single sets of spare parts. Mathematical machines and systems", pp. 1-10, 2022, doi: 10.1007/s10238-022-00904-7.

[4] V.P. Strelnikov and P.V. Strelnikov, "Basic provisions, requirements and initial data for the calculation of spare parts sets", Mathematical machines and systems, vol. 3, pp. 133-137, 2021, doi: 10.34121/1028-9763-2021-3-133-137.

[5] H.V.G.R. Hoskote, N.Prabhu, and M.R. Champaka, "Wavelet transform based fault location estimator for STATCOM compensated lines", International Journal of Electrical Engineering and Technology, vol. 11(4), pp. 309-317, 2020.

[6] S.G. Antoshchuk, S.B. Kondratyev, G.Y. Shcherbakova, and M.A. Hodovychenko, "Depth map generation for mobile navigation systems based on objects localization in images", BCIT, vol. 5, no. 1, pp. 11-18, March 2021.

[7] M. Gupta, M.A. Khan, R. Butola, and R.M. Singari, "Advances in applications of non-destructive testing (NDT): A review", Advances in Materials and Processing Technologies, vol. 8(2), pp. 2286-2307, 2022, doi: 10.1080/2374068X.2021.1909332.

[8] G. Shcherbakova, V. Krylov, W. Qianqi, B. Rusyn, A. Sachenko, et. al., "Optimization methods on the wavelet transformation base for technical diagnostic information systems", In 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), vol.11 pp. 1-6, 2021, July 26-29, IEEE, doi: 10.1109/IDAACS52987.2021.9623431.

[9] A.E. Rubanenko, O.O. Rubanenko, I.A. Hunko, and V.V. Gasych, "Determination of residual resource of measuring current transformers using fuzzy simulation", Herald of Khmelnytskyi National University, vol. 295, pp. 214-219, 2021, doi: 10.31891/2307-5732-2021-295-2-214-219.

[10] K. Lipianina-Honcharenko, C. Wolff, A. Sachenko, I. Kit, and D. Zahorodnia, "Intelligent Method for Classifying the Level of Anthropogenic Disasters," Big Data and Cognitive Computing, MDPI, vol. 7, no. 3, pp. 157, 2023.

Ontological Model of Technological Process for the Production of Complex Shape Details

Larysa Globa[1], Rina Novogrudska[1,2], Oleg Kopiika[3] and Mariia Skulysh[1,2]

[1]*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteisky Avenue 37, 03056 Kyiv, Ukraine*

[2]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*

[3]*Institute of Telecommunications and Global Information Space of NASU, Chokolivskiy Blvd. 25, 03186 Kyiv, Ukraine*

lgloba@its.kpi.ua, rinan@ukr.net, kopiika@nas.gov.ua, mskulysh@gmail.com

Abstract: The paper presents the research devoted to the description of technological process using ontological modelling. For the correct management of the modern technological process it is very important to take into account the components of knowledge management and representation that are used and generated during the execution of such production processes. Ontological models are proposed to be used for this purpose. Such type of models is one of the most frequently used for presenting knowledge nowadays. It allows not only to structure and systematize knowledge, but also to describe certain processes of interaction between data and information for the extraction of new knowledge. The paper gives the research on how to systematize, structure and link information on technological process objects by means of ontology aiming to support knowledge management and representation of such process Specific ontological system is proposed in the paper, that allows to set description of complex technological process.

1 INTRODUCTION

Nowadays the term "knowledge" is considered from different points of view [1]. One of them positions knowledge as a condition for access to information. It is argued that the representation of knowledge should be organized in such a way that it is easy to access and retrieve it.

Authors [2] argue that knowledge representation requires different strategies and different types of knowledge management tools and technologies. For example, if knowledge is viewed as an object, then knowledge management initiatives should emphasize the importance of creating knowledge stocks in organizations. In this case, a knowledge management system such as knowledge repositories should have the ability to capture knowledge stocks. Similarly, if knowledge is viewed as a process, then knowledge management initiatives must operate with knowledge flows in the processes of knowledge creation, knowledge sharing, and knowledge dissemination.

Today, there are various models of knowledge representation [3, 4], each of which has its own advantages and disadvantages. Ontological modelling is one of the most frequently used methods of presenting knowledge at the moment [5, 6, 7], as it allows not only to structure and systematize knowledge, but also to describe certain processes of interaction between data for the extraction of new knowledge.

Various SAP systems and technologies for describing business processes and workflow are currently used to automate production and technological processes. However, for the correct management of modern work processes, it is also very important to take into account the components of knowledge management and representation, that are used and generated during the execution of such production processes.

The aim of the paper is to propose specific ontological system for the description of the technological process. The paper is structured as following: Section 1 sets the introduction; Section 2 presents the review of approaches and models used for knowledge representation, Section 3 describes ontological model of technological process, that is proposed by authors, Section 5 depicts the research conclusion.

2 APPROACHES TO KNOWLEDGE REPRESENTATION REVIEW

Knowledge representation includes various schemes, namely logical schemes, procedural schemes, network schemes, and structured schemes. The network diagram is constantly growing over time. On the other hand, a logical and procedural scheme works with a fixed set of symbols and instructions that are limited as the information to be encoded increases. Structured schemes use a complex structure for a node in a graph, that limits its widespread use. Network diagrams have simple nodes in a graph that stores data and allows a huge amount of data to be built into the system in the form of nodes.

Tolman introduced the concept of cognitive maps [8]. Cognitive maps provide a mental representation of spatial information as knowledge [9]. However, the cognitive map does not have its own cognitive processing. Kosko presented the combination of fuzzy logic and cognitive map as fuzzy cognitive maps (FCM). FCM uses fuzzy logic to calculate relations strength. In 1976, John F. Sowa developed concept graphs (CG) to represent semantic network-based logic using a graph [10, 11]. A CG is a finite connected bipartite graph, where a node either represents a concept or a conceptual relation. Arcs are allowed only between a concept and a conceptual relation, but not between two concepts or two conceptual relations. Sowa later proposed a semantic network as another area of knowledge representation. It is a graphical structure of interconnected nodes, where nodes are connected by an arc to represent knowledge [12].

Artificial intelligence applications for knowledge representation and manipulations use computer architectures that support semantic network processing [13]. The semantic web is based on the understanding that relations between concepts define knowledge.

The classic classification of knowledge representation models includes the following types of models :
- production models,
- logical models,
  frame models,
- network models.

The choice of the type of knowledge representation model depends on the requirements for the designed system. Thus, from the considered models, it is advisable to use production models in the case of developing a system aimed at the implementation of expert conclusions with a rigid "cause-effect" structure; frame models are used to implement a scenario approach in order to create systems oriented to forecasting and modeling future events; logical models are used to present information in the form of formal mathematical structures, which are expedient to use when it is necessary to optimize or automate certain processes during the design of information systems.

Another type of knowledge representation models are ontologies. An ontology is a conceptual model of subject domain that uses one or more taxonomies and includes descriptions of concepts, relations, properties of concepts, and constraints on property values. Ontologies are one of the modern means of describing the semantics of subject domain (for the description of information resources including), that use the main results of previously known knowledge models. At the same time, the ontological approach is quite transparent regarding its implementation within simple logical formalisms that have the property of permissibility. This approach is adopted as one of the basic methodologies in the implementation of ontological systems.

Ontologies are used to describe knowledge about some subject domain. Ontology describes the concepts of a subject domain, as well as the relations that exist between these concepts.

In the era of Big Data, the management and integration of large data sets provide enormous opportunities for the formation of new knowledge. Ontologies, which represent knowledge about the subject domain in the form of concepts, their classes and relationships between them, are a universal tool for overcoming barriers in the integration of data and knowledge from disparate sources of information, thereby facilitating the search for new knowledge.

Ontologies can help organize and analyze large volumes of data that are too large for a single system to manage. Ontological models can be applied to represent different knowledge. It is ontologies that will allow not only to structure or classify knowledge, but also to integrate information through the formalization of standard terminology. Ontological models are well integrated into intelligent decision-making support systems that is why information, as a result of its presentation with the help of ontology, can be further presented and processed by different SAP.

3 ONTOLOGICAL MODEL OF TECHNOLOGICAL PROCESS

The given study is devoted to the development of an ontology that characterizes the production technological process.

In the study, the technological process is defined as a set of technological operations that are performed systematically and consistently in time and space on homogeneous or similar objects, as a result of which the aggregate state, location, or properties of the object of research change. Its peculiarity is that it has a finished character for the production purpose, as well as a set of its elements and sub-processes is known a priori.

Such ontology is a system of interrelated components, each of which is a complete ontology in itself. In total, a set of such components allows not only to describe all the elements and stages of the technical process, but also the corresponding equipment, software and technical support used in the technical process, as well as to introduce and describe some performance indicators of the described technological process. The system of ontologies for the description of the technological process is presented in Figure 1.

The ontology of technological processes is the main ontology of the system. It describes the characteristics and elements of the technological process, as well as the connections between them. Three lower-level ontologies are included in the system:

- ontology of production environment,
- ontology of software,
- ontology of subject domain.

The ontology of the production environment describes a variety of technological means, such as equipment, facilities, tools and materials, that are used in the technological process; the ontology of the software environment describes the software and technical means used in the technical process; the ontology of the subject domain is based on the direct description of the concepts of the subject domain and sets the structures for their "connection" with other ontologies of the system, it contains all the information related to the processes and objects that take part in the technical process, depending on the subject domain for which this technological process is applied.

In the process of each introduced ontologies development, it is necessary to describe its elements. The ontology includes the following elements [14]:

- classes,
- attributes,
- relations,
- types of attribute values,
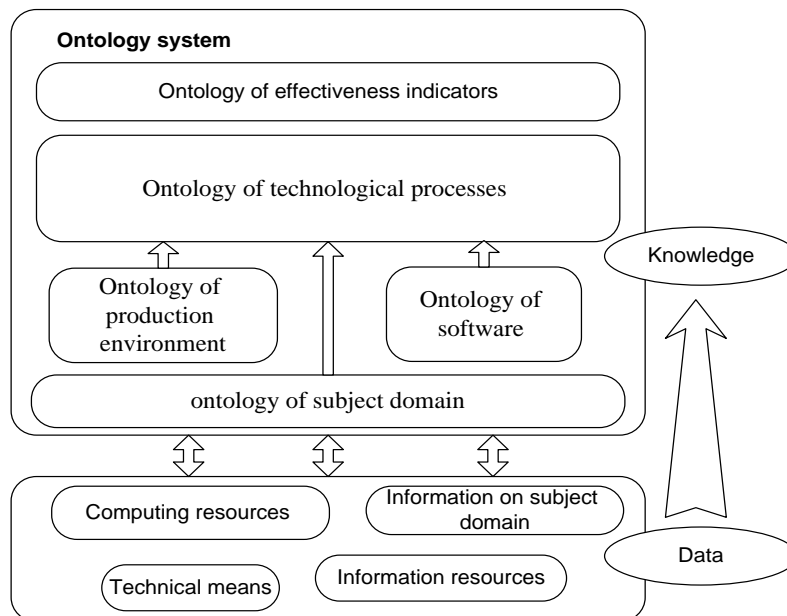- restrictions on attribute values,
- instances of classes.

Figure 2: System of ontologies for the description of the technological process.

It is this set of elements that allows us to define ontology as a complex model. Thus, ontology is a hierarchy of concepts (or classes) connected by relations. The use of associative type relations allows defining the ontological model not only as a hierarchical structure, but also as a structure that takes into account the substantial meaning of the connection between real objects. The various properties of each concept are described based on the attributes of the concepts and the constraints imposed on the domain of their values.

The technological process ontology contains a set of classes that are described by the corresponding attributes. Classes are an element of an ontological model that describes the concepts of some subject or problem domain. Attributes are elements of the ontological model that describe the properties of classes and relations. Attribute value types define standard types for class attribute values. The restriction on the values of attributes of classes and relations is not used for all attributes, but only for those whose values must lie in some area. They cannot be less than/greater than a given value, or are determined by a certain rule.

For example, let's describe the classes and their attributes for the technological process ontology (Figure 2):

- Class "Technical process": name, environment, structure, function, internal properties, external properties, technological route.
- Class "Transitions": type (complex/simple), number, operations that combine.
- Class "Operations": name, type, type of surface for which it is applied.
- Class "Technical object (detail)": name, identification data (item number, drawing number), classification data (batch number, classification number), technological data (material, quality), design data (shape, dimensions, tolerances) , data for processing.
- Class "Construction drawing": name, number, set of structural constituent elements (surfaces) of the technological object, drawing (drawing) characteristics.
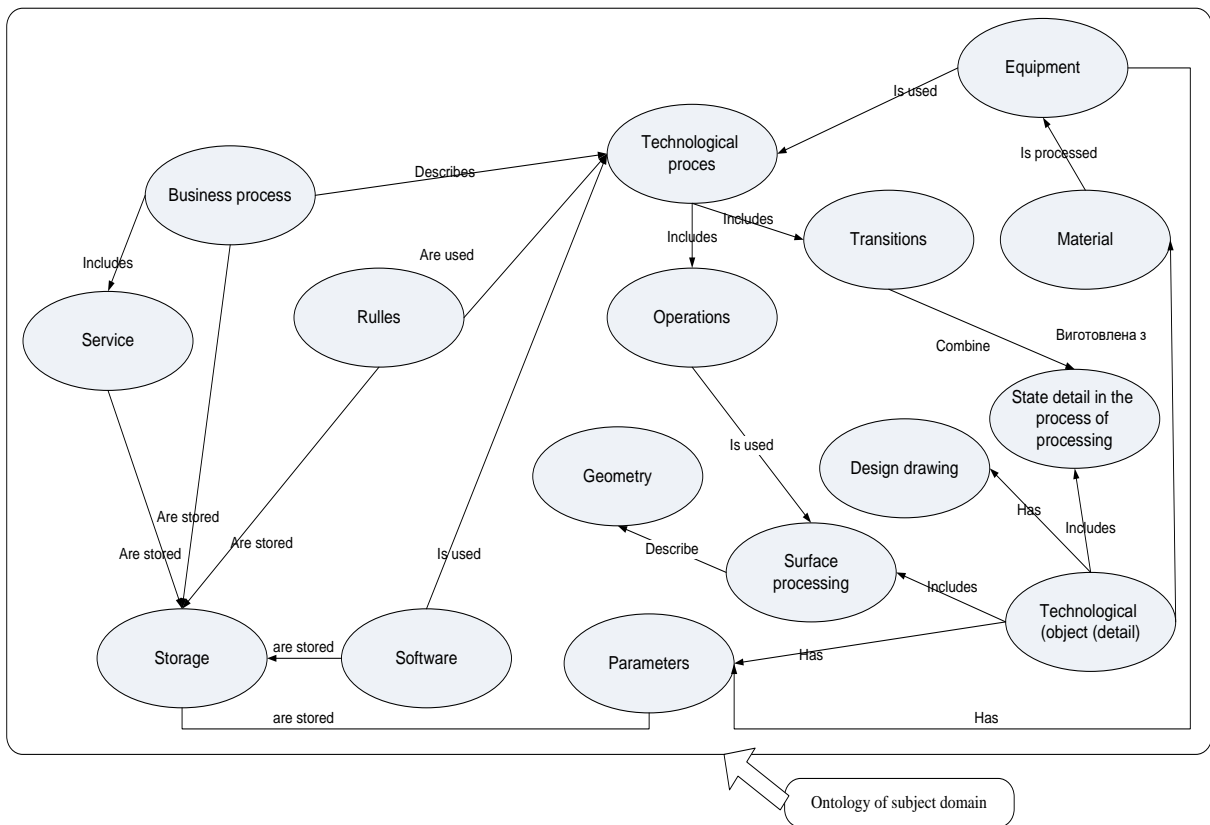

Figure 2: Fragment of technological process ontology.

- Class "Surface treatment": name (type), type (characteristics), dimensions, design parameters of the surface (roughness, size tolerance), operations.
- Class "Geometry": type of surface, dimensions, drawing (drawing) characteristics. Class "State of detail in the process of processing": general information (code and name of the detail, general information about the product, volume of the optimal batch of details, data about the workpiece), processing surface, operations, surface code, description of the workpiece on the surface, parameters that determine the presence of identical surfaces, surface roughness and parameters that impose, restrictions on the cutting process, parameters that describe the accuracy of detail manufacturing, a data set that defines the profile, surfaces and its own geometric constraints, a data set that represents the location of the surface in the coordinate system of detail.

An ontology contains a set of relations that are defined on classes and reflect either the relationship of classes to each other or the relationship of classes to data or attributes. The following types of relations exist: associative relations, part-whole relations, inheritance relations, and class-data relations [16].

Thus, such a description of ontology elements allows managing all elements of technological process, as well as setting the semantic connectivity through them.

4 CONCLUSIONS

In this study, an original approach to the design and storage of data regarding objects of design and technological preparation of production (parts, equipment, materials, design procedures, etc.) is proposed in the form of an ontological model of the technology of manufacturing complex details.

The proposed approach makes it possible to obtain a functionally complete list of elementary, algorithmically correct functional tasks intended to develop regulatory documentation for design and technological preparation of production; methods of making technical decisions using artificial intelligence; principles of flexible adaptation to a specific production environment and requirements of the modern market.

REFERENCES

[1] D. Randall, S. Howard and S. Peter, "What Is a Knowledge Representation?," in AI Magazine, vol. 14, 2002, pp. 17-33.

[2] M. Alavi and D. Leidner, "Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," in MIS Quarterly, vol. 25, no. 6, 2009, pp. 95-116.

[3] V. Kavitska, V. Liubchenko, and A. Lysyuk, "A knowledge representation model for knowledge management systems," in Odes'kyi Politechnichnyi Universytet. Pratsi, 2013, pp. 167-172, doi: 10.15276/opu.3.42.2013.33.

[4] P. Tanwar, T. Prasad and K. Dutt, "A Tour Towards the Various Knowledge Representation Techniques for Cognitive Hybrid Sentence Modeling and Analyzer," in International Journal of Informatics and Communication Technology (IJ-ICT), vol. 7, 2018, pp. 124.

[5] L. Globa., N. Gvozdetska, and R. Novogrudska, "Ontological model for data processing organization in information and communication networks," in System Research and Information Technologies, 2021, no. 1, pp. 47-60, doi: 10.20535/SRIT.2308-8893.2021.1.04.

[6] K. Tkachenko, "Using Ontological Modeling by Intellectualization of Learning Processes. Digital Platform," in Information Technologies in Sociocultural Sphere, vol. 5, 2022, pp. 261-269, doi: 10.31866/2617-796X.5.2.2022.270130.

[7] A. Guraliuk, M. Rostoka, A. Koshel, Y. Skvorchevska, and Luchaninova, Olga, "Ontological Modeling of Electronic Educational Resources," 2022, doi: 10.1007/978-3-030-93907-6_71.

[8] E. Tolman, "Cognitive maps in rats and men," in Psychological Review, vol. 55, no. 4, 1948, pp. 189–208, doi:10.1037/h0061626.

[9] J. Whittington, D. McCaffary, J. Bakermans, and T. Behrens, "How to build a cognitive map," in Nature Neuroscience, vol. 25, 2022, pp. 1-16, doi: 10.1038/s41593-022-01153-y.

[10] J. Sowa, "Conceptual Structures: Information Processing in Mind and Machine," in The Systems Programming Series, Addison-Wesley, MA, 1984, ISBN 978-0201144727.

[11] V. Macedo, L. Thurler, E. Dias, and M. Cavalcanti, "Knowledge graph: A strategy for knowledge management?," in Seven Editoria, 2023, doi: 10.56238/sevened2023.006-041.

[12] J. Sowa, Principles of Semantic Networks, SanMateo, 1992.

[13] P. Pirnay-Dummer, D. Ifenthaler, and N. Seel, "Semantic Networks," in Encyclopedia of the Sciences of Learning, Springer, Boston, MA, 2022, doi: 10.1007/978-1-4419-1428-6_1933.

[14] L. Globa, R. Novogrudskaya, B. Zadoienko, and O. Y. Stryzhak, "Ontological Model for Scientific Institutions Information Representation," in 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 2020, pp. 255-258, doi: 10.1109/PICST51311.2020.9467984.

Digital Transformation of Energy Infrastructure in the Conditions of Global Changes: Bibliometric Analysis

Viktoriia Khaustova, Mykola Kyzym, Nataliia Trushkina and Mykyta Khaustov

Research Center for Industrial Problems of Development of the National Academy of Sciences of Ukraine,
Inzhenernyi Lane 1a, 61165 Kharkiv, Ukraine
{v.khaust, m.kyzym, khaustov.mkt}@gmail.com, nata_tru@ukr.net

Keywords: Energy System, Energy Infrastructure, Energy Network, Digital Transformation, Digitization, Information Technologies, Ecosystem.

Abstract: At the current stage of transformational transformations, the problems of energy infrastructure transformation are extremely important. And these issues are especially relevant in the conditions of rapid development of the digital economy. According to the calculations of The Boston Consulting Group experts, the volume of the digital economy by 2035 will amount to 16 trillion dollars. USA. At the same time, it is worth noting that the digital transformation observed in the energy sector as a critically important branch of the national economy has led to the emergence of new challenges and risks of information and cybersecurity, which should be given special attention in order to increase competitiveness and achieve sustainable development. In view of this, the purpose of this study is to determine the trends and key areas of research on the transformation of energy infrastructure in the era of digitalization based on bibliometric analysis using the VOSviewer software. The article specifies the meaning of the concepts "energy infrastructure", "transformation", "digital transformation". A bibliometric analysis of the relationship between the terms "energy infrastructure", "energy system", "energy network", "digital transformation", "digitalization", "information technologies" was performed. The author's approach to the interpretation of the concept of "digital transformation of the energy infrastructure in the conditions of global changes" is proposed. It has been established that the use of information systems and digital technologies in recent years has been recognized as a strategic direction for the transformation of energy infrastructure in many European countries. A comparative analysis of existing international methods for evaluating the development of energy systems, taking into account the information component, was performed. It has been proven that the priority direction of research in the future should be the substantiation of the theoretical and methodological provisions of the formation of the digital energy ecosystem.

1   INTRODUCTION

Currently, the development of the energy system in the global world takes place in the conditions of constant natural disasters, climate changes, terrorist acts, and cyber threats. Therefore, the strategic directions of ensuring national security in the international security system of the countries of the world are the protection and security of energy infrastructure as an important element of critical infrastructure [1]; application of digital technologies to achieve sustainable development of energy infrastructure, taking into account adaptation to climate change; implementation of the concept of public-private partnership of protection strategies in ensuring the stability of the energy infrastructure in crisis situations, etc.

It is worth noting that the transformation of the energy infrastructure should be considered from the point of view of digitalization [2]. According to estimates by Forbes experts, 67% of the leaders of companies from the Global 2000 list chose digital transformation as a priority goal of their corporate strategy in 2018.

According to a study by analysts of the International Data Corporation, the total global spending on digital technologies will grow by 16.8% annually and reach 2.1 trillion dollars in 2019. According to forecasts of The Boston Consulting Group, the volume of the digital economy by 2035 will amount to 16 trillion dollars. Research by

Huawei and Oxford Economics showed that intelligent network interaction will trigger the growth of the digital economy, which will reach 23 trillion dollars by 2025. This increase will be 78.3% compared to 2017 (12.9 trillion dollars). By 2025, it is planned to increase the share of the digital economy by 7.2 percentage points, or from 17.1 to 24.3% of global GDP.

Therefore, the digital transformation observed in many sectors of the national economy (including the energy sector) has led to the emergence of new challenges and risks of information and cybersecurity, which should be given special attention in order to increase competitiveness and achieve sustainable development.

2 LITERATURE REVIEW

The process of changes and transformations is defined by the concept of "transformation", which is quite complex and multifaceted. It is believed that the concept of "transformation" became widely used in the social sciences in the second half of the 20th century, to characterize the latest processes associated with radical structural changes in national economies.

As the analysis of the literature [3-8] shows, the concept of "transformation" is defined as:
1) the process of adapting elements of economic systems at different levels to the regularities of the functioning and development of the national economy;
2) movement from form to form through the negation of the old form and the formation of a new one;
3) changing the structure of any object;
4) the process of transformation and formation of the system;
5) the general form of development of economic systems associated with evolutionary and revolutionary changes;
6) qualitative transformations of the economic system;
7) quantitative and qualitative changes in system components of various scales.

Therefore, transformation means the process of changing the form, appearance, nature or character of a society or individual structure. This is the transformation of essential components of society, all aspects of spheres of economic activity and social life as a whole.

It is proposed to interpret the transformation as a qualitative transformation and formation of the energy system, which enables the transition to a fundamentally new level of its functioning and development, which is carried out consistently and continuously at all stages.

Based on the generalization of special literature, it was established that currently there are many definitions of the concept of "digital transformation". For the most part, scientists and practitioners [9-13] understand digital transformation as:
1) maximum use of the potential of digital technologies in all aspects of business;
2) large-scale business transformation, which concerns the entire set of functions of the enterprise; the path to the implementation of digital technologies and business models to increase productivity in quantitative terms;
3) use of technologies to radically increase productivity or availability of resources for enterprises;
4) a change in business thinking in the new conditions of the digital economy, the driver of which is the modern consumer and the changing culture of communications;
5) the process of transition of the organization to new ways of thinking and working based on the use of social, mobile and other digital technologies;
6) the process of a radical change in the form of the economic system, as a result of the search, development, implementation and use of digital technological innovations to increase the efficiency of the performance of their functions by all structural divisions;
7) conscious, management-initiated process of radical improvement of business processes both in the company's internal and external environment based on research and development, as well as further implementation and use of digital technologies;
8) radical changes in the complex of business processes, starting from product development and ending with customer service;
9) implementation of modern digital technologies in the organization of business processes at enterprises.

Therefore, digital transformation involves the integration of digital technologies and solutions in all areas of business. This is a cultural and technological shift that requires organizations to make fundamental changes in their work methods and customer experience management. Under this term, it is

proposed to consider a fundamental rethinking of customer experience, business models and operations. This is the search for new ways of creating value, generating revenue and improving the efficiency of business processes.

At the same time, it should be noted that energy infrastructure belongs to strategic objects of critical infrastructure. This is due to the fact that interruptions in the functioning of the energy sector negatively affect other critical infrastructure facilities. Traditionally, the energy infrastructure is considered as a complex system consisting of many components designed for the production, distribution and supply of energy [14].

In this work, energy infrastructure is understood as a set of objects, enterprises and energy networks that are priority and strategically important for the development of a competitive national economy and ensuring an adequate level of energy security, the neutralization or malfunction of which can harm the national interests of the state, lead to the emergence possible threats and risks, significant damages and losses (economic, financial, investment, transport, logistics, environmental, informational, etc.).

The digital transformation of the energy infrastructure is considered as the strategic implementation of digital technologies, which allows the elements of the energy system to function and develop in the digital age. This concept involves the use of digital tools and platforms to transform traditional business processes, improve interaction with customers, introduce innovative technologies and form a digital energy ecosystem. The main components of the digital transformation of the energy infrastructure include:

1) digital technologies (cloud computing, artificial intelligence, big data analytics and the Internet of Things (IoT), which allow energy infrastructure objects to collect, analyse and use huge amounts of data to make informed decisions, automating processes and providing a personalized experience);
2) organizational changes (restructuring of processes to increase their flexibility, introduction of new methodologies (DevOps, Agile), formation of a digital culture);
3) customer orientation and integration of digital channels (the use of websites, mobile applications and social network platforms will improve the quality of customer service, provide personalized content, and ensure smooth interaction at various contact points).

3 RESULTS

Let's have a deeper look at the genesis of scientists' views on the issue of digital transformation of the energy infrastructure. For this, a bibliometric analysis of scientific publications on the development of energy infrastructure in the conditions of digitalization, which are indexed in the Scopus database using the VOSviewer toolkit, was carried out. 19,955 documents were found by the title of articles, abstracts and keywords.

As the analysis showed, the first publication on the selected topic appeared in the international scientometrics database Scopus in 1974 [15]. Until the 1990s, an insignificant level of publishing activity was observed. After that, the works of scientists, including C. Hull [16], appeared in scientometrics databases. A. Farrell et al. [17]; V. Gungor et al. [18]; D. Mengnuo [19] and others, in which attention is focused on the development of the concept of energy infrastructure protection in crisis phenomena; application of smart technologies and intelligent systems in the energy sector.

Among the most cited publications, the following articles deserve special attention:

1) O. Younis, S. Fahmy [20], published in 2004, which was cited in a journal indexed by the Scopus scientometrics database 4232 times. In this paper, the authors proposed a new approach to distributed clustering for long-lived special sensor energy networks.
2) X. Fang et al. [21], published in 2012, which was cited, 2295 times. The article examines three main systems, namely, a smart infrastructure system, a smart management system, and a smart protection system.
3) A.-H. Mohsenian-Rad et al. [22], published in 2010, which was cited, 2258 times. The paper presents an autonomous and distributed user demand-side energy management system that takes advantage of the two-way digital communication infrastructure envisioned in the future smart grid. Based on game theory, an energy planning game is formed, where players are users and their strategies are daily schedules of their household appliances and loads.
4) V. C. Güngör et al. [23], published in 2011, which was cited in a journal indexed by the Scopus scientometrics database 2058 times. The authors consider important issues of smart grid technologies, primarily from the point of view of problems and opportunities of information and communication technologies.

Among the key publications that publish works on the selected topic, the following can be mentioned: Energies (358 documents); IEEE Access (310); Applied Energy (191); Energy Policy (133); Sustainability Switzerland (124); Energy (122); Renewable and Sustainable Energy Reviews (117).

There are 36 documents of scientist A. Tzanakaki in the Scopus database; 30 documents – D. Simeonidou; 27 documents each – N. Javaid, H. Mouftah; 25 documents - scientist P. Siano; 23 documents each – S. Rinaldi, N. Shah; 22 documents each – M. Anastasopoulos, A. Flammini; on 21 documents - H. Gabbar, J. Guerrero, M. Longo, B. Mohammadi-Ivatloo, A. Orgerie, J. Rodrigues, C. Verikoukis, C. Wietfeld.

The key organizations involved in solving the problems of digital transformation of energy infrastructure are Politecnico di Milano (155 documents); Imperial College London (142); Delft University of Technology (139); Tsinghua University (129); CNRS Center National de la Recherche Scientifique (123); Universitat Politécnica de Catalunya, ETH Zürich (117); Beijing University of Posts and Telecommunications, Aalborg University, Massachusetts Institute of Technology (113 documents each).

The results of the analysis show that most of the works on the studied issue are published by scientists from the USA (3898 documents), India (2521), China (1975), Great Britain (1574), Italy (1,378), and Germany (1351).

According to the types of documents, scientific works can be ranked as follows: materials of conferences (9430 documents or 47.3% of the total number); scientific articles (8034 or 40.3%); chapters of books or monographic publications (1031 or 5.2%); review articles (670 or 3.4% of the total).

For the most part, scientific works on the problems of digital transformation of energy infrastructure are published in the following fields of knowledge: engineering (10,726 documents); computer science (10423); energy (5073); ecology (1966); social sciences (1730); business and management (651 documents).

The main sponsors that finance scientific publications on the problems of energy infrastructure transformation in the conditions of the digital economy include the following: National Natural Science Foundation of China (672 documents); National Science Foundation (546); Horizon 2020 Framework Program (429); European Commission (341); Engineering and Physical Sciences Research Council (298); U.S. Department of Energy (245); European Regional Development Fund (173 documents).

So, the analysis of publishing activity confirmed that starting from the beginning of the 90s of the 20th century, there is a growing scientific interest in the study of transformational changes in the energy infrastructure using information technologies and systems. At the same time, the interdisciplinary nature of research is followed, and the geography of scientists and researchers studying this topic is diverse (but with a noticeable predominance of scientists and institutions from the USA, India, China, Great Britain, Italy, Germany, Canada, Spain, France).

Further processing and analysis of bibliographic data was carried out using the VOSviewer software, which is a software tool for constructing and visualizing maps of bibliometric networks [24]. VOSviewer software was used to construct network maps of relationships between keywords based on bibliographic records from Scopus databases. The visual results of the obtained map of the bibliometric network are shown in Figure 1.

The map of the bibliometric network displays the frequency of use of terms by the size of the circle and the intensity of communication, and allows you to track variants of combinations of terms both within clusters and between them. The colour of the circle indicates that the keyword belongs to a certain cluster. The larger the diameter of the circle, the more often this term appears in scientific publications. Links on the map show the frequency of repetition of keywords in publications, while the smaller the distance between keywords, the stronger the connection between them [24].
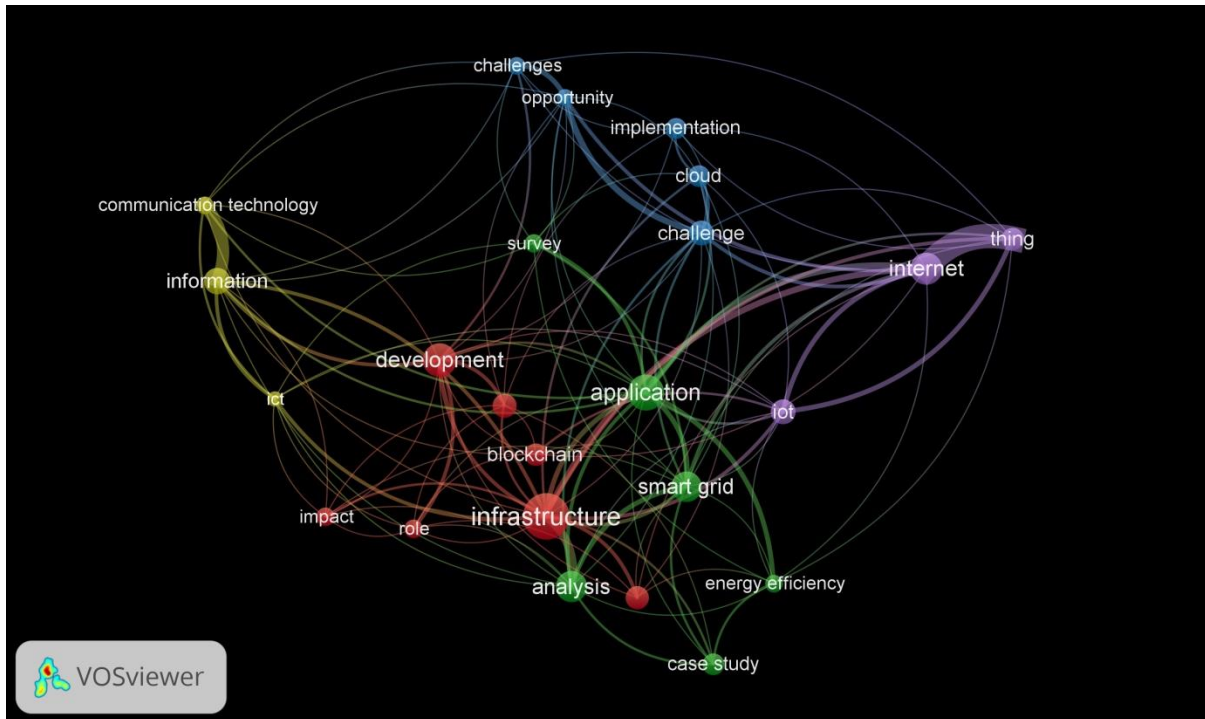
Figure 1: Network visualization of citations of articles on issues of digital transformation of energy infrastructure, implemented using the VOSviewer tool[1].

According to Fig. keywords can be grouped into 5 clusters using the VOSviewer program. A generalized description of clusters of key phrases in scientific research on the digital transformation of energy infrastructure is presented in Table 1.

Table 1: Characterization of clusters of key phrases in scientific studies of transformational changes in energy infrastructure in the context of digitalization[2].

Cluster	The most used term	Number of keywords	Related keywords
1 (red)	infrastruc-ture	7	development, digital transformation, blockchain, impact, digitalization, role
2 (green)	applica-tion	6	analysis, energy efficiency, smart grid
3 (blue)	implemen-tation	5	challenge, cloud, opportunity,
4 (yellow)	Informa-tion	3	communication technology, ICT
5 (purple)	intelligent systems	3	Internet, Internet Things

Thus, based on the results of the study, the following conclusions can be drawn:

1) The number of publications indexed in Scopus, whose titles, abstracts and keywords contain the term "energy infrastructure", is growing at an accelerated pace every year. Research on the transformational changes of the energy infrastructure in the conditions of the rapid development of the digital economy is gaining more and more popularity, starting from the 90s of the 20th century. The main reasons for the growing popularity of these scientific studies are: the search for new ideas, the transformation of the energy management paradigm, the introduction of the concept of sustainable development, the development of digital technologies and intelligent systems.

2) The term "energy infrastructure" has an interdisciplinary nature, it is used in studies of various branches of science, namely: it is found in publications on engineering, computer science, ecology, social science, management, business, decision science, etc.

3) Visualization of the network map of keywords based on bibliographic data made it possible to single out 5 clusters that characterize the key areas of research: infrastructure, energy system,

---

[1]Built on the basis of data from the Scopus scientometrics database using the VOSviewer program.

[2]Created by the authors using the VOSviewer program.

digital transformation, information technologies in energy, intelligent systems.

The Network Readiness Index (NRI) was used to determine the current problems of the development of the energy system and the implementation of information and communication technologies in the countries of the world; ICT Development Index (IDI).

The Network Readiness Index (NRI, Portulans Institute, USA) (Table 2) was developed to assess the network (technical) readiness of the countries of the world to use the possibilities of information and communication technologies and systems. From 2002 to 2016, this index was published by the World Economic Forum as part of the Global Information Technology Report. Today, on the basis of NRI, 134 countries of the world are evaluated according to 60 indicators, which are systematized according to 4 main groups: technologies, personnel, management, influencing factors. NRI covers issues from artificial intelligence technologies and the Internet of Things to the role of the digital economy in achieving the Sustainable Development Goals.

The ICT Development Index (IDI) is a combined indicator that characterizes the achievements of the countries of the world in terms of the development of information and communication technologies. The index was developed according to the methodology of the International Telecommunication Union, a specialized division of the United Nations, which defines global standards in the field of ICT. The IDI is calculated on the basis of summing up 11 indicators to an integral criterion, which allows comparing the achievements of the countries of the world in the development of the ICT sector. The index can be used as a benchmarking tool at the global, regional and national levels.

Table 2: Network Readiness Index on the example of the Surveyed European Countries (source [25]).

Country	Year		Rank changes
	2020	2022	
Germany	9	9	unchanged
Czech Republic	28	25	↑+3
Portugal	31	29	↑+2
Estonia	23	22	↑+1
Poland	33	34	↓-1
Croatia	43	45	↓-2
Latvia	37	39	↓-2
Hungary	39	41	↓-2
Romania	49	52	↓-3
Lithuania	29	33	↓-4

As the analysis shows, the leaders in the field of information and communication technologies in 2007 were Sweden, South Korea, Denmark, the Netherlands, and Iceland. In 2017, Iceland, South Korea, Switzerland, Denmark, and Great Britain were the leading countries in the field of ICT.

During 2007-2017, the positions of some European countries worsened. Thus, the rank of Ukraine in the rating of the development of information and communication technologies decreased by 28 points; Portugal and Hungary – by 13; Romania – by 12; Poland – by 10; Lithuania – by 8; Czech – by 3 points. At the same time, Estonia's position improved by 9 points, and Croatia's by 7 (Table 3).

Table 3: The ICT Development Index of the Surveyed European Countries (source [26, 27]).

Country	Year				
	2007	2010	2014	2015	2017
Croatia	43	31	37	42	36
Czech Republic	40	37	41	34	43
Estonia	26	33	21	20	17
Hungary	35	34	46	48	48
Latvia	36	40	33	37	35
Lithuania	33	35	40	40	41
Poland	39	38	44	44	49
Portugal	31	27	43	43	44
Romania	46	48	58	59	58

For the purpose of rating the countries of the world according to the "Energy Efficiency" indicator, the American Council for an Energy-Efficient Economy (ACEEE) has developed a world rating of global energy efficiency. The organization studied the energy-saving models of 23 countries that consume the largest amount of electricity (75% of all consumption in the world) and ranked them in order. In 2016, Germany ranked 1st, Italy and Japan – 2nd, France – 4th.

In addition, the Energy Information Administration (EIA) compiled a list of countries on electricity consumption. Average electricity consumption per capita reflects the level of socio-economic development. The higher this indicator, the more "comfortable" a person lives, the more electrical equipment can be used in production.

In view of the above, it can be noted that digital transformation of the energy infrastructure is the process of introducing modern technologies into the business processes of enterprises in the energy sector to increase productivity, improve service and customer service. These are radical changes in

management methods, corporate culture and ethics, external and internal interaction of stakeholders and energy market participants.

Therefore, under the digital transformation of energy infrastructure, it is proposed to consider the process of transition of the energy system as an important element of critical infrastructure to a qualitatively new state, taking into account such a modern challenge as digitalization.

4 CONCLUSIONS

Based on the above, we can come to the following conclusion. Currently, the problems of energy infrastructure transformation in the context of the implementation of the digital strategy of the European Union are gaining special relevance. At the same time, the EU's single digital market is the most attractive for the development of digital energy. In the future, this may become a "development driver" of the national economies of European countries.

As a result of the study, it was established that many methodologies are used to determine the current problems and trends in the development of energy infrastructure in the countries of the world. But it is worth emphasizing that all methodologies for evaluating the development of energy infrastructure are imperfect and do not fully correspond to modern economic conditions. As the analysis showed, there is no integral index that would take into account the impact of information and communication technologies on the development of the energy system. In this regard, there is currently a need to develop a proper methodical approach to assessing the development of energy infrastructure in the countries of the world, taking into account the global challenges of the modern digital environment.

For this, it is necessary to make changes and additions to the Energy Strategies of the European countries, in which to provide for the use of information technologies, intelligent and smart systems for the transformation of the energy infrastructure. This, in turn, can become one of the tools for guaranteeing energy and information security as important components of the national security of states.

Therefore, the digital transformation of the energy infrastructure is a key component of the overall strategy for the transformation of the energy system. Correctly selected digital technologies in combination with the competencies of employees, processes and operations will allow energy infrastructure facilities to quickly adapt to crisis

situations, use promising opportunities to modernize work processes, meet new and constantly changing customer needs, stimulate growth and implement innovative and management solutions in energy industry.

For effective digital transformation of the energy infrastructure on a practical level, it is advisable to pay attention to such key aspects as:
- having a clear vision and strategy that meets business goals;
- involvement of interested parties (stakeholders) and ensuring the interest of the entire organization;
- constant monitoring and assessment of the implementation of digital transformation initiatives;
- an adaptive and iterative approach that allows you to navigate the changing digital landscape.

Prospects for further research consist in substantiating the theoretical and methodological provisions of the formation of the digital energy ecosystem.

REFERENCES

[1] M.O. Kyzym, V.Ye. Khaustova, and N.V. Trushkina, "The Essence of the Concept of "Critical Infrastructure" from the Standpoint of National Security of Ukraine," Business Inform, no. 12, pp. 58-78, 2022, [Online]. Available: https://doi.org/10.32983/2222-4459-2022-12-58-78.

[2] A. Kwilinski, N. Trushkina, I. Birca, and Yu. Shkrygun, "Organizational and Economic Mechanism of the Customer Relationship Management under the Era of Digital Transformations," E3S Web of Conf., vol. 456, 05002, 2023, [Online]. Available: https://doi.org/10.1051/e3sconf/ 202345605002.

[3] J. A. Schumpeter, "Das wissenschaftliche Lebenswerk Eugen von Böhm-Bawerks," Zeitschrift für Volkswirtschaft, Sozialpolitik und Verwaltung, no. 23, pp. 454-528, 1914, [Online]. Available: https://anno.onb.ac.at/cgi-content/anno-plus?aid=zvs&datum=1914&page= 464&size=45.

[4] F. Hayek, "The Road to Serfdom". 1st ed. London: George Routledge & Sons Ltd, 1944, [Online]. Available: https://ctheory.sitehost.iu.edu/img/Hayek_The_Road_to_Serfdom.pdf.

[5] A. Toffler, "The Third Wave". 1st ed. New York: William Morrow and Company, Inc., 1980. https://ia801200.us.archive.org/9/items/TheThirdWave-Toffler/The-Third-Wave_-_Toffler.pdf.

[6] P. Drucker, "The age of social transformation," Atlantic Monthly, vol. 274, pp. 53-80, 1994, [Online]. Available: https://www.theatlantic.com/past/docs/politics/ecbig/ soctrans.htm.

[7] D. Bell, "The Cultural Contradictions Of Capitalism." 20th Anniversary ed. New York: Basic Books, 1996, [Online]. Available: https://voidnetwork.gr/wp-content/uploads/2016/08/ The-Cultural-Contradictions-of-Capitalism-by-Daniel-Bell.-Book.pdf.

[8] J. Stiglitz, "Whither Reform? – Ten Years of the Transition." In: Annual Bank Conference on Development Economics (Washington, April 28-30, 1999). Washington DC: World Bank, 1999, [Online]. Available: https://www.gsid.nagoya-u.ac.jp/sotsubo/stiglitz.pdf.

[9] S. Park, et al., "Smart Fire Safety Management System (SFSMS) Connected with Energy Management for Sustainable Service in Smart Building Infrastructures," Buildings, vol. 13, no. 12, 3018, 2023, [Online]. Available: https://doi.org/10.3390/buildings 13123018.

[10] A. Temel, M. Ayaz, and V. Aygül, "Digital transformation for improved energy and product efficiency in tire production: a case study," Eng. Res. Express, vol. 5, no. 4, 045048, 2023, [Online]. Available: https://doi.org/10.1088/2631-8695/ad0877.

[11] A. Bartusiak, et al., "First step into automation of security assessment of critical infrastructures," Sustain. Energy Grids Netw., vol. 36, 101139, 2023, [Online]. Available: https://doi.org/10.1016/j.segan.2023.101139.

[12] L. Li, et al., "Exploring the mechanism of digital transformation empowering green innovation in construction enterprises," Dev. Built Environ., vol. 15, 100199, 2023, [Online]. Available: https://doi.org/10.1016/ j.dibe.2023.100199.

[13] A. Chwiłkowska-Kubala, et al., "The impact of resources on digital transformation in energy sector companies. The role of readiness for digital transformation," Technol. Soc., vol. 74, 102315, 2023, [Online]. Available: https://doi.org/10.1016/j.techsoc.2023.102315.

[14] V. Khaustova, I. Hubarieva, D. Kostenko, T. Salashenko, and D. Mykhailenko, "Rationale for the Creation and Characteristics of the National High-Tech Production of Motor Biofuel." In: Zaporozhets, A. (eds.), Systems, Decision and Control in Energy V. Studies in Systems, Decision and Control (pp. 569-583), vol 481. Springer, Cham, 2023, [Online]. Available: https://doi.org/10.1007/978-3-031-35088-7_31.

[15] G. Zaltman, "A note on an international invisible college for information exchange," J. Am. Soc. Inf. Sci., vol. 25, no. 2, pp. 113-117, March-April 1974, [Online]. Available: https://doi.org/10.1002/asi.4630250206.

[16] C. W. Hull, "Engineering in a global economy. A United States perspective," Technology in Society, vol. 12, no. 2, pp. 107-120, 1990, [Online]. Available: https://doi.org/10.1016/0160-791X(90)90003-U.

[17] A. Farrell, H.Zerriffi, and H. Dowlatabadi, "Energy infrastructure and security," Annu Rev Environ Resour ., vol. 29, pp 421-469, 2004, [Online]. Available: https://doi.org/10.1146/annurev.energy.29.062403.10 2238.

[18] V. C. Gungor, et al., "Smart grid and smart homes: Key players and pilot projects," IEEE Ind. Electron. Mag., vol. 6(4), pp. 18-34. 2012, [Online]. Available: https://doi.org/10.1109/MIE.2012.2207489.

[19] D. Mengnuo, "Research on the securitization of electric vehicle charging network," CICED, 8592335, pp. 296-301, 2018, [Online]. Available: https://doi.org/10.1109/ CICED.2018.8592335.

[20] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," IEEE Trans Mob Comput, vol. 3, no. 4, pp. 366-379, October 2004, [Online]. Available: https://doi.org/10.1109/TMC.2004.41.

[21] X. Fang, et al., "Smart grid - The new and improved power grid: A survey," IEEE Commun. Surv. Tutor., vol. 14, no. 4, pp. 944-980, 2012, [Online]. Available: https://doi.org/10.1109/SURV.2011.101911.00087.

[22] A.-H. Mohsenian-Rad, et al., "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," IEEE Trans Smart Grid, vol. 1, no. 3, pp. 320-331, December 2010, [Online]. Available: https://doi.org/10.1109/ TSG.2010.2089069.

[23] V.C. Güngör, et al., "Smart grid technologies: Communication technologies and standards," IEEE Trans Industr Inform, vol. 7, no. 4, pp. 529-539, November 2011, [Online]. Available: https://doi.org/10.1109/ TII.2011.2166794.

[24] VOSviewer – Visualizing scientific landscapes, [Online]. Available: https://www.vosviewer.com.

[25] S. Dutta and B. Lanvin, "The Network Readiness Index 2022," Portulans Institute, Washington DC, USA, 2022, [Online]. Available: https://download.networkreadinessindex.org/ reports/nri_2022.pdf.

[26] Measuring the Information Society Report 2017 (International Telecommunication Union, Geneva Switzerland, 2017, [Online]. Available: http://handle.itu.int/11.1002/ pub/80f52533-en.

[27] The ICT Development Index. ITU. 2017, [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Pages/IDI/ Background.aspx.

Research on Energy Efficiency of Wi-Fi IoT Systems on Renesas DA16200 Platform

Serhii Kondratiev, Daria Koshutina, Olha Liubomska and Illia Baskov

National University Odesa Polytechnic, Shevchenko Avenue 1, 65044 Odesa, Ukraine
{kondratiev, d.v.koshutina, baskov.i.o}@op.edu.ua, olyalyubomska@gmail.com

Keywords: Internet of Things, Wi-Fi, Energy Efficiency, Controller, Renesas, DA16200, ESP8266, ESP32, Comparative Analysis, Power Consumption, Crystal-on-Chip Microcontrollers.

Abstract: This research focuses on a comprehensive analysis of the energy efficiency of the Renesas DA16200 microcontroller. The investigation adopts a comparative approach, directly contrasting the power consumption of the DA16200 with the widely used ESP8266 controller under identical operating conditions. The primary metric employed to assess energy efficiency is average battery life. Additionally, a detailed examination of current consumption is conducted across various operational modes, encompassing active states like data exchange, reception, and transmission, as well as low-power sleep mode. This analysis extends beyond simply measuring peak current draw. Transient current profiles are captured, providing time-resolved insights into how current consumption fluctuates throughout different operational phases. This granular data enables a deeper understanding of the microcontrollers' energy utilization patterns. Furthermore, the research explores and evaluates techniques for minimizing energy consumption specifically in the ESP8266. These findings are then juxtaposed against the inherent energy-saving features of the DA16200 microcontroller. To facilitate a precise and verifiable comparison, a custom test bench accommodating both the DA16200 and ESP8266 is designed and implemented. This controlled environment ensures consistency in operating conditions and minimizes external variables that could influence the results. The culmination of this research is the presentation of a comprehensive analysis, detailing the comparative energy consumption profiles of the studied microcontrollers. This data forms the foundation for objectively evaluating their suitability for various low-power the Internet of Things applications.

1 INTRODUCTION

The Internet of Things (IoT) is a network of physical objects, embedded with sensors and connected to the internet. These objects can collect and exchange data, enabling them to interact with each other and with people. The IoT enables these objects to be sensed and controlled remotely across existing network infrastructure, creating opportunities for more direct integration of the physical world into computer-based systems, and resulting in improved efficiency, accuracy, and economic benefits [1]. Controllers play a vital role in IoT systems, as they are used for data collection, processing, and device control.

Power consumption is a critical issue for IoT systems. IoT devices typically have limited power sources, and their power consumption must be as low as possible to ensure their autonomy.

There are some ways to reduce the power consumption of IoT systems. One way is to use more energy-efficient components, such as low-power controllers. Another way is to optimize the software of IoT systems to improve energy efficiency.

Reducing the power consumption of IoT systems has several benefits. It can improve the autonomy of IoT devices, reduce operating costs, and extend their lifespan. Additionally, reducing the power consumption of IoT systems can help protect the environment.

In the ubiquitous era of the IoT, millions of interconnected devices are revolutionizing everything from healthcare to environmental monitoring. However, a hidden challenge lies beneath the surface of this transformative technology - power consumption. As these devices often operate on limited battery life or remote power sources, their energy efficiency is paramount. Consider a wearable health tracker constantly monitoring vital signs – even small reductions in power consumption can translate to significantly longer battery life, empowering more seamless and uninterrupted monitoring.

This article delves into the critical issue of energy efficiency in IoT devices, focusing specifically on low-power microcontroller solutions. We compare three popular controllers - the Renesas SoC DA16200, the Espressif ESP32, and the widely used ESP8266. Our primary objective is to evaluate their performance and power consumption under various operating conditions, ultimately seeking to identify optimal solutions for specific IoT applications. Through this analysis, we explore key factors like active and sleep mode power draw, processing capabilities, and software optimization techniques, aiming to provide valuable insights for developers and researchers in the field of energy-efficient IoT design.

2 PROBLEM STATEMENT

The Wi-Fi protocol is a high-speed standard designed for efficient transmission of large volumes of media information and internet data through radio frequency exchanges at 2.4 GHz and 5 GHz frequencies. Due to the high data transmission speeds and the volumes of transmitted information, the radio block of a Wi-Fi controller cannot ensure low power consumption at the wattage level [2]. In transmission mode, the average current of a Wi-Fi controller varies from 120 mA to 270 mA [3], depending on the radiation level (+13…+20 dB), making the use of AA/AAA/tiny batteries impractical for several years, which is crucial for IoT devices.

The high power consumption of Wi-Fi controllers is also evident in the receiving mode, where the average power consumption of the receiver fluctuates between 50 mA and 90 mA, which does not meet the power requirements for IoT batteries and does not allow powering IoT devices with batteries for a year or longer. As a result, the Wi-Fi protocol is applied in cases of IoT with stationary power. In contrast, low-power consumption protocols such as Bluetooth Low Energy (BLE), LoRa, 6LoWPAN, and others are used for battery-powered IoT applications, resulting in lower transmission speeds.

However, connecting IoT devices using other protocols like BLE, LoRa, and 6LoWPAN requires specialized gateways/routers, presenting certain challenges due to the diversity of devices, standards, and configuration complexities. Nevertheless, almost every household has a standard Wi-Fi router, providing easy internet access without the need for additional purchases.

In consideration of the aforementioned, concerted efforts are being directed towards a substantial reduction in the power consumption of Wi-Fi IoT controllers by orders of magnitude, to satisfy battery power requisites spanning multiple years. Initially successful in environments characterized by low IoT operational demands, these initiatives encompassed sporadic information transmissions over Wi-Fi, ranging from daily to monthly intervals. However, the evolution of IoT into the realm of actuators has necessitated a more frequent exchange of data, precipitating a thousandfold reduction in controller response time, bringing it down to 1 second or less.

As an illustrative example, a door lock control system is expected to achieve a response time of no more than a few seconds. This necessitates IoT transmission not on a scale of 1-2 times per day, but rather at intervals of 1 time every 1-3 seconds. Power consumption can only be curtailed by reducing the operational time of the Wi-Fi radio transmitter/receiver to a few milliseconds, coupled with the subsequent transition of the IoT into a deep sleep state during the interludes between transmissions.

Furthermore, in light of the stringent cost constraints associated with IoT microcontrollers, this article restricts its consideration to Espressif SoC microcontrollers, such as ESP8266 and ESP32, for comparative analysis, as other Wi-Fi microcontrollers no longer fall within the confines of this pricing category.

1) *Transmitter activation optimization.* The transmitter is activated infrequently to acknowledge requests from the router. An optimally chosen extended pause before responding helps avoid losing communication with the controller, considering variable times dependent on router settings (e.g., Wireless Inactivity Timeout, Connection Timeout, Client Timeout).

2) *Periodic receiver activation window.* The receiver is activated not for every router polling message (DTIM) but periodically, for example, every second message or less frequently.

3) *Infrequent activation of IoT receiver.* Receiver activation occurs less frequently to align with Wi-Fi exchange standards. For instance, the maximum interval between router polling messages (DTIM) is set to 1000 milliseconds.

4) *Minimization of transmitted data size.* The length of transmitted data from the controller is reduced to the minimum value, thereby reducing transmitter operation time.

5) *Minimization of router polling message (DTIM) transaction time.* The minimum transaction time

6) *Low-level microcontroller programming.* Providing access to the lowest level of microcontroller programming for maximum acceleration of the above requirements.

7) *Setting minimum transmission level and optimizing controller placement.* Configuring the transmitter to the minimum transmission level and placing the controller as close as possible to the unobstructed router for radio wave communication.

8) *Optimizing receiver/transmitter on/off times.* Technologically reducing the time it takes to turn on/off the receiver and transmitter to minimal values.

9) *Minimizing receiver current consumption.* Technologically ensuring minimal receiver current consumption to reduce overall energy consumption.

10) *Use of supercapacitors for powering.* Applying supercapacitors for powering the controller with slow charging (from 5 mA to 50 mA) from weak batteries for pulse-powering the transmitter at a level of 300 mA.

Wi-Fi controllers for IoT devices exhibit high power consumption, limiting their applicability in battery-powered devices [4]. In transmit mode, the average current of Wi-Fi controllers ranges from 120 mA to 270 mA [5], rendering the use of batteries impractical for extended periods. For standard IoT controllers like ESP8266 [6] and ESP32 [7], the hardware implementation of real-time methods is absent. Attempts to optimize performance through low-level register code are constrained due to limited access to the internal architecture of these controllers, provided in a closed form [8]. This imposes restrictions on significant improvements in response time and operation of these devices in IoT mode.

In this context, the DA16200 controller stands out as it is proclaimed to be a device originally designed to support sub-second response time in the IoT domain. It is capable of maintaining a constant connection with the router and achieving a power consumption level that enables it to be powered by batteries for a minimum of one year and even longer.

The objective of this article is to compare the energy efficiency of three Wi-Fi controllers (ESP8266, ESP32, and DA16200) for IoT devices.

3 POWER CONSUMPTION CHARACTERIZATION OF DA16200 CONTROLLER

3.1 Analysis of the DA16200 Controller's Pulse Current Consumption

For the current consumption investigation, an experiment was conducted to measure the current when breaking the P2 contacts. To achieve this, a shunt with a resistance of 0.5 ohms was connected to the P2 contacts, and a 1:1 probe of the Hantek DSO5102P oscilloscope with a bandwidth of 100 MHz was connected to the shunt contacts, as shown in Figures 1 and 2. The measured voltage drop across the shunt at a current of 250 mA was 0.125 V, which had no significant impact on the controller's operation. To reduce interference during measurement, the DA16200 Module Evaluation Kit (EVK) was powered from a power bank via USB.

3.2 Calculation of Average Current Consumption

Another crucial indicator of energy efficiency is the calculation of the controller's average current consumption. Average current consumption (I_{avg}) is the amount of energy consumed by a system in a unit of time. It is calculated as the ratio of the total amount of energy consumed by the system over a given time period to that time period [9].

There are several methods for calculating the average current consumption, such as the geometric and current integration methods. In this study, the geometric method was chosen and calculated using (1). This method is based on the average current (I_{avg}) being equal to the area under the consumption pulse (A) divided by the pulse period (T):

$$I_{avg} = \frac{A}{T}. \qquad (1)$$

To calculate the area under the consumption pulse, data from the oscilloscope measurement results, as shown in Figures 1 and 2, were used. The area of the consumption pulses is equal to the sum of the areas of all pulse shapes, as indicated in (2):

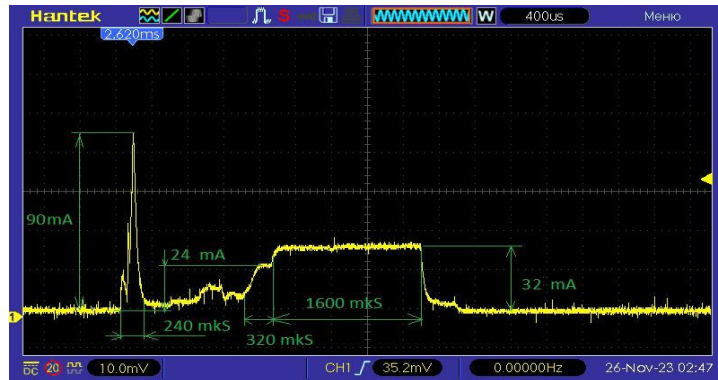$$A = A_r + A_{r.t.} + A_{i.t.}. \qquad (2)$$

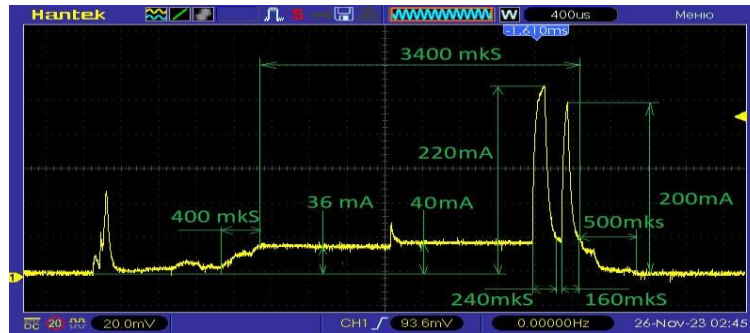Figure 1: Pulse current consumption of the DA16200 in reception mode.

Figure 2: Pulse current consumption of the DA16200 in transmission mode.

The area of rectangular consumption pulses (*Ar*) was calculated using the method of multiplying the sides, that is, multiplying time (*t*) by the current amplitude (I$_m$):

$$A_r = t * I_m.$$

The area of rectangular triangles (A$_{r.t}$), representing the fall and rise of the pulse, is calculated as half the product of the area $A_{r.t} = \frac{t * Im}{2}$, and the area of isosceles triangles (A$_{i.t.}$), is calculated as one-third of the product $A_{i.t.} = \frac{t * Im}{3}$. The measurement time *(t)* was chosen to be 30 seconds.

As a result of oscilloscope measurements, we obtained the following picture: 30 reception pulses (pulse shapes) and 1 transmission pulse during the measurement time. During this period, when there is neither transmission nor reception, the controller is in Sleep 1 mode, consuming a current of 0.2 μA.

The area of one session of consumption pulses during reception is equal to:

$$A_t = 1600 * 32 + \frac{320 * 24}{2} + \frac{240 * 90}{3} = 62.240 \ (mA \cdot \mu s).$$

The area for 30 sessions is equal to:

$$A_t = 62.240 * 30 = 1\ 867\ 200 \ (mA \cdot \mu s).$$

The average receive current is:

$$I_{avg\_t} = \frac{1\ 867\ 200}{30\ 000\ 000} = 62 \ (\mu A).$$

The area of consumption pulses during transmission is equal to:

$$A_r = (1400 * 36) + (110 * 40) + \frac{400 * 36}{2} +$$
$$+ \frac{240 * 220}{2} + \frac{160 * 200}{2} + \frac{500 * 40}{3} = 150\ 667 \ (mA \cdot \mu s).$$

The average current during transmission is:

$$I_{avg\_r} = \frac{150\ 667}{30\ 000\ 000} = 5 \ (\mu A).$$

To obtain the overall average current, it is necessary to add the Sleep 1 current to the average reception current and the average transmission current.

$$I_{avg} = 62 + 5 + 0.2 = 67 \ (\mu A).$$

It turned out unexpectedly that the receiver contributes the main share (92%) to the current consumption - 62 μA out of 67 μA, indicating the potential for further current reduction while maintaining IoT Wi-Fi responsiveness once per second.

The investigation of the DA16200 controller's current consumption revealed that the receiver contributes the majority, consuming 62 μA out of the total 67 μA. This accounts for 92%. Therefore, there is significant potential to reduce power consumption by optimizing the receiver's operation.

3.3 Battery Operating Time of DA16200

The measurement of the average current consumption of the DA16200 controller showed that it is 67 μA. This figure is exceptionally low for devices of this class. However, to ensure prolonged operation of an IoT device, it is essential to consider that the operating time also depends on the supply voltage and battery capacity. The supply voltage of the DA16200 controller ranges from 2.1 to 3.6 V [10]. This voltage range allows the use of various types of batteries to power the controller, including two AA or AAA batteries, one Li-ion battery, or a single lithium battery.

Table 1: The capacities of different batteries.

Battery Type	CR2032	AAA	AA	18650
Capacity, mAh	240	1000	2000	2500
Pulse discharge current	15 mA	0.7 A	1.5 A	5 A
Operating time, days at 67 μA	149	622	1243	1554

In Table 1, average values of battery capacity and their pulse currents sufficient to operate the DA16200 transmitter with a peak consumption current of ~ 250 mA. The operating time is specified for maintaining a connection to the router with a response time of 1 second. When transmitting data, the operational time decreases proportionally to the increase in the current consumption during transmission.

However, even when addressing the issue of impulse current consumption at a level of 250 mA from a CR2032 battery using a supercapacitor, it is not possible to power the DA16200 for one year. To ensure the operation of the DA16200 transmitter for one year, it is necessary to use batteries with a capacity of at least 600-800 mAh, equivalent to two AAA batteries or more. However, with such initial capacities, the transmitter's runtime for a year will be limited. Increasing the battery capacity will extend the transmission time.

The conducted measurements indicate that DA16200 manufacturers have managed to introduce a high-speed and power-efficient Wi-Fi protocol into the battery-powered IoT domain.

Analysis of the oscillograms depicting the current consumption of the DA16200 has facilitated the elucidation of the underlying success of the new technology. The current consumption in the receiving mode is contingent upon:

- The current consumption of the radio circuits within the Wi-Fi receiver;
- The current consumption of the Wi-Fi protocol processing controller during data reception.;
- The duration of Wi-Fi protocol processing when receiving data.

The current consumption associated with the radio circuits of the Wi-Fi receiver is a nominal few milliamps, contributing insignificantly to the overall consumption of the DA16200. However, the pivotal elements of the technical solution by Renesas reside in the current consumption of the processing controller and the duration of its Wi-Fi protocol processing. Renesas engineers achieved a reduction in the duration of current consumption during reception to 2 milliseconds by employing a proficient and energy-efficient ARM Cortex M4F controller, complemented by the authorship of optimized code for Wi-Fi protocol processing. This reduction is of paramount importance, as the duration of current consumption during reception, particularly for IoT applications requiring response times on the order of several seconds, constitutes the primary (90%) contributor to the total power consumption. After this, the aforementioned metric will be juxtaposed with analogous parameters in competing solutions.

4 COMPARATIVE ANALYSIS OF POWER CONSUMPTION OF DA16200 AND ESP8266/ESP32 CONTROLLERS

In this section, a comparative analysis of the energy consumption of the DA16200 and ESP8266/ESP32 controllers will be conducted. The analysis will utilize data obtained during experimental research, as presented in Tables 2 and 3.

As seen from the provided data, the impulse current consumption during transmission for the DA16200 controller is almost identical to the impulse current of ESP8266/ESP32 controllers. The impulse current consumption during reception is

1.6-2.5 times less than that of ESP8266/ESP32 controllers. This is because the activity time during data exchange with the router for the DA16200 controller is less than 2 ms, while the activity time for ESP8266/ESP32 with the most efficient ESP-NOW algorithm is 130 ms.

Table 2: Comparison of current consumption for DA16200, ESP8266 and ESP32 controllers.

Controller	Pulse current consumption during transmission, mA	Pulse current consumption during reception, mA
DA16200	220	32
ESP8266	200	55
ESP32	240	80

Table 3: Comparison of average current consumption and operating time for DA16200, ESP8266 and ESP32 controllers.

Controller	Average current consumption	Operating time from batteries (2000 mah), days
DA16200	67 μA	1243
ESP8266	4.7 mA	11
ESP32	10.7 mA	7.7

Based on this data, it can be concluded that the DA16200 is the most energy-efficient controller among the mentioned ones. It exhibits lower average current consumption and longer battery runtime. However, it is essential to note that these data are based on nominal values of controller current consumption. Conducting independent measurements is necessary for obtaining more accurate values.

5 CONCLUSIONS

The DA16200 controller, designed for operation within a home environment and connectivity to the internet through standard Wi-Fi routers, represents a significant advancement in the development of the IoT. It addresses a key limitation of existing IoT technologies, such as Bluetooth Low Energy, which require the use of dedicated gateways for internet connectivity. This simplifies and reduces the cost of implementing IoT technologies in household devices, opening up new possibilities for their application.

Based on the research findings, it can be concluded that the DA16200 controller is the most energy-efficient Wi-Fi controller for IoT devices. It is capable of maintaining a constant connection to the router and providing a level of energy consumption that allows it to be powered by batteries for a minimum of one year and even longer.

In the IoTMark®-Wi-Fi test, the DA16200 received a score of 815, equivalent to 815 days of autonomous operation for an IoT sensor powered by two AA batteries. It is anticipated that the smart door lock will last for over three years without recharging, which is 50% longer than what the closest competitor can offer. For the first time in Wi-Fi history, it can provide autonomous operation time comparable to Zigbee and Z-wave.

REFERENCES

[1] G. V. Kumari, V. Sailaja, P. Sunitha, and B. V. Lakshmi, "Embedded systems and IoT: A theoretical approach," GCS PUBLISHERS, 2022.

[2] Y. Wang, J. Xu, and L. Wang, "Low-Power Wi-Fi for IoT Applications: A Survey," IEEE Wireless Communications, vol. 26, no. 5, pp. 152-161, Oct. 2019.

[3] A. Bel, T. Adame, and B. Bellalta, "An energy consumption model for IEEE 802.11ah WLANs", Ad Hoc Networks, vol. 72, pp. 14-26, 2018.

[4] S. Sarkar and A. Debnath, "Green IoT: Design Goals, Challenges and Energy Solutions," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 637-642, doi: 10.1109/ICCES51350.2021.9489167.

[5] A. Jangid and P. Chauhan, "A Survey and Challenges in IoT Networks," 2019 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 2019, pp. 516-521, doi: 10.1109/ISS1.2019.8908079.

[6] ESP8266 User Manual, Espressif Systems, 2023, [Online], Available: https://www.espressif.com/sites/default/files/documentation/esp8266-technical_reference_en.pdf.

[7] ESP32 User Manual, Espressif Systems, 2023, [Online], Available: https://www.espressif.com/sites/default/files/documentation/esp32_technical_reference_manual_en.pdf.

[8] D. Ma, G. Lan, M. Hassan, W. Hu, and S. K. Das, "Sensing, Computing, and Communications for Energy Harvesting IoTs: A Survey," in IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 1222-1250, Secondquarter 2020, doi: 10.1109/COMST.2019.2962526.

[9] W. B. Qaim, A. Ometov, A. Molinaro, I. Lener, C. Campopo, E.S. Lohan, and J. Nurmi, "Towards Energy Efficiency in the Internet of Wearable Things: A Systematic Review," in IEEE Access, vol. 8, pp. 175412-175435, 2020, doi: 10.1109/ACCESS.2020.3025270.

[10] UM-WI-023 DA16200 ThreadX Evaluation Kit, [Online], Available: https://www.renesas.com/eu/en/document/oth/um-wi-023-da16200-threadx-evaluation-kit-user-manual.

Measuring the Economic Value of Investment Activities: A Case Study of Ukrainian Telecommunications Companies

Galyna Otlyvanska[1], Iryna Stankevych[2], Inna Yatskevych[2], Hanna Sakun[2],
Oksana Vasylenko[3] and Eduard Siemens[3]

[1]*Electronic Engineering Department, Maynooth University, Maynooth, W23F2H6 Co. Kildare, Ireland*

[2] *Department of management and marketing, Odesa State Academy of Civil Engineering and Architecture, Didrikhsona Str. 4, 65029 Odesa, Ukraine*

[3]*Departments of Electrical Engineering, Mechanical Engineering and Engineering Management, Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*

otlyvanskaya.g@gmail.com, arnika@ukr.net, innav2018@gmail.com, sakun.anna@ukr.net, {oksana.vasylenko, eduard.siemens}@hs-anhalt.de

Keywords: Value Metrics, Measurement, Stakeholder, Distance from Benchmark Method, Integrated Score, Economic Value of Investment Activities, Telecommunications Companies.

Abstract: The research in the companies' Economic Value of Investment Activities (EVIA) show the most important problems in creating value for Ukrainian mobile providers are customers' value and differentiation of financing. These problems have appeared not only because of the companies' management mistakes but also there are general conditions on how to provide investment activities in the Ukrainian telecommunications market. A new approach to measure of a company EVIA, considering stakeholders' value metrics, and based on the Distance from a Benchmark Method (DBM) is proposed by authors. The DBM provides an opportunity to quantify such a category as EVIA and compare the key characteristics of several companies' activities simultaneously with the identification of factors that affect the efficiency of investment activities, since this process is based on a significant set of criteria and a multidimensional space. The application of the DBM approach to EVIA assessment, in the presented conditions, on the considered example of Ukrainian mobile operators, allows to identify the main problematic aspects of the level of efficiency of invested capital, customer value and differentiation of financing.

1 INTRODUCTION

On average, a company's investment activities are the defined set of its investment decisions. Every decision has different conditions by investment volume, interest rate, level of risk, and term of implementation.

Results of investment activities from capital and financial investments are implemented in a company's operating and financial activities. It influences the complexity of development tasks that are solved in economic activities of a company. This multi-aspect complicates the management of a company's investment activities, including the measure of investment activities efficiency.

To solve this problem, it is necessary to define and research the wide set of characteristics of economic activities. The paper proposes to do this by measuring a company's Economic Value of Investment Activities (EVIA). Why is this necessary?

In economic theory, Value is a balance between the benefit (result) and the costs associated with obtaining a good. Economic Value is the result of purposeful economic activity that combines the usefulness of a good for the consumer and the costs associated with obtaining it. While remaining a subjective category, the economic value of any object (process or phenomenon) largely depends on its usefulness and has a monetary expression in its market value or in the potential value that parties interested in this object of value are willing to pay. In this context, investment activity through investment costs is an object of management that forms and lays the basis for the development of economic activity of the enterprise on the basis of value.

Currently, Value is a key characteristic of a business. The measure of Value is the basis of a management decisions. An investor has many criteria to determine Value when making a choice on how to invest.

The relevant management concepts that are widespread in practice – Process Approach, Value-Based Management, Total Quality Management, and Balance Score Card are all based on Value.

A new view on the measurement of a company's EVIA opens new opportunities to raise the efficiency of investment activities management.

2 LITERATURE REVIEW

Research in economic value has become relevant over time. There is research that investigates many aspects of creation economic value, including the place of Value Theory in Economics [1]; the process how to create economic value of IT [2]; money's value [3]; the value creation on macro level based on functions between GDP and human and capital resources [4]; as well as the value of faculty working with eye tracking [5]; the value of strategy [6]; the value of artificial intelligence that provides the potential of technology for the economy and society [7] etc. Stefan F. Dieffenbacher [8] defines that in the process of creating value, enterprises should apply a comprehensive approach that focuses on key strategies and practices that ensure their continued competitiveness and relevance in a rapidly changing environment.

In addition, there are other attempts to measure economic value. Tsagkarakis K.A. & Katsikis I.N. [9] consider possible methods to measure the value in different ways. Hejazi R. et. al. [10] propose to find the monetary equivalent of the natural resources in response to a challenge between the environment and development. The article [11] notes that economic value is subjective and dependent on a person's intentions it cannot be directly measured. However, the question how to obtain the integrated score of the economic value has always been relevant.

Previous research [12] defines the theoretical and methodological basis of measuring a company's EVIA based on Value-Based Management (VBM), the Stakeholder Theory, and Benchmarking methods.

Value-Based Management is the strategic management approach. Its goal is to raise the business value for the owners (shareholders) over time [13].

At the same time, the Stakeholder Theory also has a significant following to raise a company's market value but not to measure a company's EVIA. The main position of the Theory is if a company is oriented to increasing value for the stakeholders, it has more chances to reach permanent growth [14, 15]. In fact, the term "stakeholder" first appeared an internal memorandum at the Stanford Research Institute in 1963 [14]: "those groups without whose support the organization would cease to exist."

As a criticism of Stakeholder theory, scientists [16, 17, 18] state that a main function of management is transformed from one factor (shareholders' value) to a multifactor function (stakeholders have different value measurements). This aspect makes the management process more complicated.

Measuring a company's EVIA provides for the estimation of qualitative and quantitative performances of investment, operating and economic activities. These are different by content and measurement units. Thus, it is proposed to measure a company's EVIA using the Benchmarking method.

According to the Benchmarking Theory, there are internal, competitive, functional, and generic benchmarking. Competitive benchmarking is the most useful type to measure a company's EVIA. It is a process of comparing a company's activities to that their primary competitors have been observed to demonstrate superiority in some important elements of performance [19, 20]. The methods of comparative economic analysis as a specific method group of competitive benchmarking are chosen with the aim of providing the possibility for the assessment of the different measurement's units of the metrics. The questions of how to apply these methods to measure a company' activities and to perform the complex economic analysis of economic systems are highlighted in the research [21, 22]. Their research forms the basis of measuring a company's EVIA.

3 MATERIAL AND METHODS

Methods that allow both evaluating results and making management decisions aimed at achieving certain goals include Strategic Management, Strategic Alliances, Benchmarking, the Balanced

Score Card (BSC), the Sum-of-Digits Method, the Sum of Rankings Method, the Taxonomy Method, etc. All these methods operate with multidimensional and complex structures, but not all of them use quantitative methods to make optimal decisions. And given that the management of an enterprise's investment activity is the management of a multidimensional and multicriteria system, we should focus on methods that largely determine the choice of the best solutions based on the economic value and the best result in a multi-stage investment process. And if we consider investment activity as a multidimensional cybernetic system, then the assessment of its effectiveness should be chosen to meet the criteria for optimal search for the best solutions.

One of the peculiarities of companies' investment activities is the dynamism and multi-stage nature of the investment process. It is the utilization of dynamic programming method that allows for optimizing decisions at each stage of the investment process because, as a rule, simple summation of results from individual stages does not yield the optimal solution. The control vector of dynamic programming systematically conducts step-by-step maximization of the total sum of the objective function, for instance, quarterly expenses are optimized to achieve the best outcome by the end of the period.

To address applied economic problems, particularly those related to measuring the EVIA, which reflects the multifaceted nature of investment activities, a modified method of goal programming is applied. This method belongs to quantitative methods and doesn't take into account the preferences of the person making management decisions. It is based on a combination of linear programming and the method of sequential narrowing of the distance from the benchmark. This method is mostly used to find optimal solutions in managing company activities, as the decision-making process is based on a significant set of criteria, the consideration of which requires appropriate compromises in defining a multidimensional space. Therefore, there are certain grounds to apply it in our research, under the generalized name - Distance from a Benchmark Method (DBM). All other methods based on the evaluation of preference functions, such as: a priori, a posteriori, and interactive (based on pairwise comparison of solutions) methods, involve introducing a linear programming problem into a multicriteria optimization system.

The application of DBM for measuring the economic value of a company's investment activities requires the development and justification of a system of quantitative indicators, the involvement of which will provide the possibility of comprehensive evaluation of value and formation of a multi-criteria vector – the results of the company's investment activities. By comparing such vectors in the set of vectors of other evaluation objects, the search for the optimal vector – benchmark is ensured, which consists of maximum values of criteria that collectively represent the optimal economic value of the company's investment activities according to defined objectives. The substantive content of quantitative indicators applicable for evaluation should consider the specificity of evaluation objects and modern requirements for evaluating the efficiency of a company's investment activities, particularly its economic value. Processing the results of quantitative evaluation and finding the benchmark among the set of objects requires managerial decisions aimed at improving the efficiency of the company's investment activities by determining deviations (distances) from the benchmark, thereby demonstrating an increase in economic value.

It is important to notice that the method can be used in both measuring a current score of an object and to get a forecast score of the object.

Despite the advantages, there are some disadvantages. There is a possibility to get contradictions of results without previous data analysis. It is necessary to have a high level of knowledge in the appropriate field to apply the method [22, 23].

The procedure of applying the DBM for measuring a company's EVIA can be defined in three stages:
1) Establish of a company's EVIA indexes and set a benchmark at the selected company.
2) Transform the indexes to the standard score.
3) Calculate the integrated score of a company's EVIA.

The establishment of a company's EVIA indexes is grounded on the characteristics and valuable aspects of a company's stakeholders.

The transformation of the indexes to the standard score. The calculation of the *standard score* for every single index. In this case, it is necessary to define the correspondence between the level of a factual singular index and its benchmark.

$$I_{EVi} = {SI_{EVi}}/{BI_{EVi}}, \qquad (1)$$

where I_{EVi} – a standard score for a singular index of economic value; SI_{EVi} – a factual singular index of economic value; BI_{EVi} – the benchmark for the singular index of economic value.

The benchmark for every singular index is the highest economic value that equals 100%. Therefore, its standard score is one. Every singular index of economic value receives its benchmark measure. The benchmark reflects the best level of an index that selected company reached during past seven years.

The *integrated score of a company's EVIA* is defined as a sum of the standard scores of the singular indexes of economic value, considering the weight of the singular index in the integrated score:

$$EVIA = \sum_{i=1}^{n} I_{EVi} \times w_{I_{EVi}}, \qquad (2)$$

where n – number of a singular index of a company's EVIA; $w_{I_{EVi}}$ – the weight of the singular index in the integrated score.

The weight of the singular economic value index in the integrated score defined by experts.

4 RESULTS AND DISCUSSION

The essence of EVIA is elaborated through links between its need, utility, and cost. These are the key notions of economic value [12].

EVIA reflects the significance of investment activities to save and develop a company's economic activities through an influence on its results, thus creating Economic Value Added (EVA).

A company's Need of Investment Activities (NIA) is the ability to create the basis to save and / or increase the economic activities value by the activation of value drivers, which are capital and financial investments

A company's Cost of Investment Activities (CIA) is a monetary extent of its economic value. That can be characterized in absolute terms by a total amount of factual or planned investment activities for a certain period of time, and in a relative measure as the share of

investment activities in Income (Capex Ratio).

EVIA as an internal characteristic is studied through the conditions and efficiency of investment decisions, indexes of development of operating and financial activities, and summarizing performances of the economic activities by a stakeholders' value chain (Figure 1).
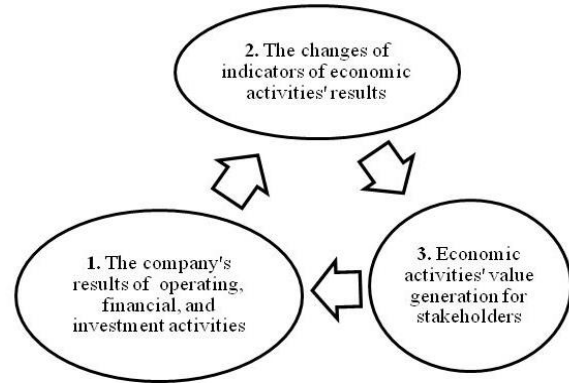
Figure 1: The stakeholder value chain of a company's economic activities.

This paper gives an example of the application of the DBM for measuring EVIA for telecommunications provider.

The NIA for a telecommunications provider is a scaling up of capital investment in new technologies and communications standards to provide the growth of a telecommunications network's capacity. This necessary growth stems from Information and Communication Technology (ICT) spreading through society.

The NIA for a telecommunications provider forms the basis of its existence. The Utility of Investment Activities (UIA) can be reflected in qualitative indexes of the development of capital and intangible assets of operating activities. There are indexes of an accessibility of services to customers such as a network coverage, technical indicators of capacity, bandwidth, and transmission speed. In addition, this utility can be reflected in energy efficiency, environmental friendliness of a network, etc (Figure 2).
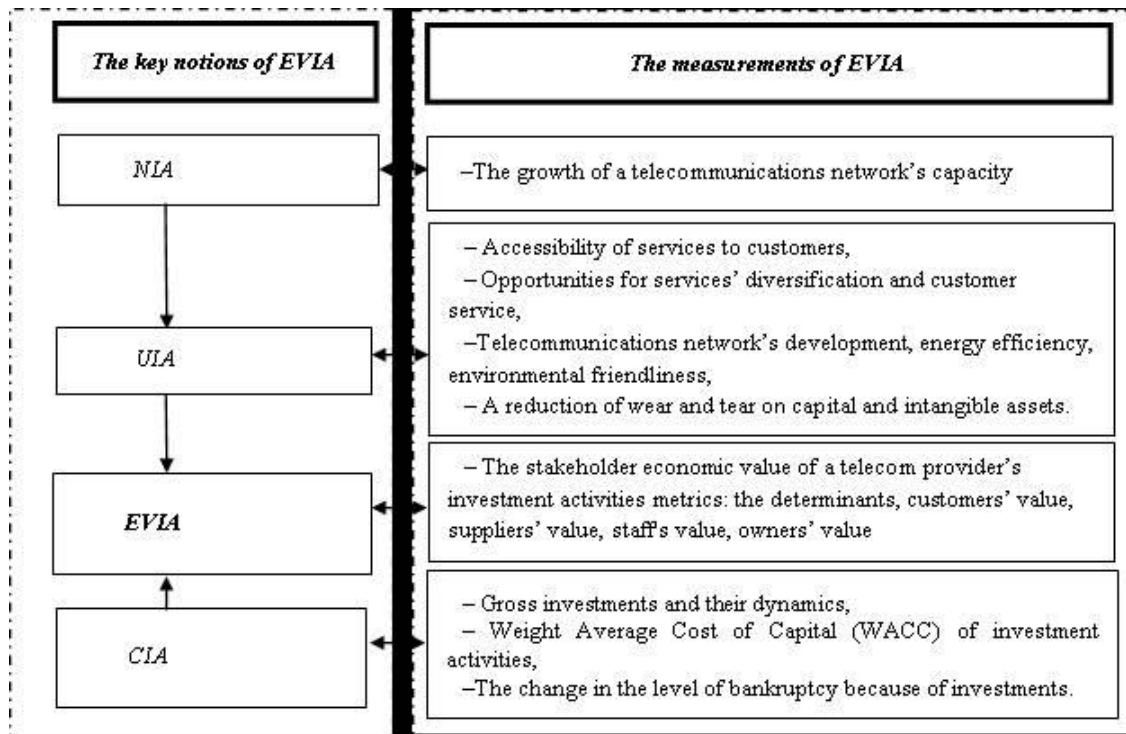
Figure 2: The links and content of key notions of a telecommunications company's EVIA.

Technical indexes of a telecommunications network development are indicators for stakeholders that characterize future prospects of the stakeholders' value changes. These value changes are connected with different facets of investment decision implementation and their influences on a telecommunications provider's economic activities. The important significance for UIA is the wear and tear on capital and intangible assets. It reflects whether or not there are enough investments in the renewal of assets.

The summary of the result of previous research [12] provides an opportunity to establish the list of EVIA indexes of a telecommunications provider:

1) The determinants: technological leadership, differentiation of services, return on invested capital (ROI), capex ratio, differentiation of financing; payback period.
2) Customers' value: saving of the subscriber base, upturn in the subscriber base, upturn in the average revenue per user (ARPU).
3) Government's value: increase in the internet-traffic, tax payments.
4) Suppliers' value: amount and change of material expenses, the share of the permanent suppliers.

5) Staff's value: wage Increase, staff's productivity.
6) Owners' value: dividends increase, competitiveness; EVA.

The indicators provided by the authors of the article have been chosen as those that most effectively reflect the efficiency of the investment activities of a telecommunications provider in terms of economic value. They reflect the technical and economic efficiency of a telecommunications provider's investment activities results. In addition, the features of telecommunications activities are reflected in the list of the singular value indexes. However, this list of indicators is not exhaustive; the number of indicators and metrics can be expanded or narrowed depending on the goals/object of comparison. Therefore, the authors of the article recommend considering the indicators as a conditional example set of EVIA indices for research purposes and demonstrating the application of DBM.

In order to set a benchmark, it is necessary to select a representative company and define benchmarks measurements. According to the data cited from Complex Statistical for the period 2015-2022 [24], as well as taking into account the

previous investigation [13] has shown that big non-Ukrainian companies that provide permanent investment police obtain better general economic results. Among Ukrainian telecommunications providers has approximately the same investment police only Private Stock Company (PrSC) "Kyivstar" [25]. It is also a leader in the Ukrainian telecommunications. For that reason, PrSC "Kyivstar" was chosen as the Benchmark Company.

The best level of PrSC "Kyivstar" EVIA indexes for 7 years and denote by convention the periods P1 - P8 are used as the benchmark. The

list of the EVIA benchmark measures of a telecommunications provider is given in Table 1.

The measurement of a telecommunications provider's EVIA is taken from two the largest mobile providers of Ukrainian communications – PrSC "Vodafone Ukraine" and PrSC "Kyivstar".

Some aspects of a telecommunications provider's EVIA are considered next. Among the characteristics of an economic value, there are particular singular indexes. They require the used a special consideration as how to clarify an economic value benchmark standard of the telecommunications company's investment activities.

Table 1: The EVIA benchmark measures of a telecommunications provider.

The singular value index	Characteristic of standard benchmark measure	Benchmark Measure	Benchmark Standard	Weight
The determinants of EVIA				
1. Technological leadership, stations	Number of 4G base stations	7139	1	0,1
2. *ROI*, %	The efficiency of invested capital as a ratio between Net Operating Profit (*NOPAT*) and Capex	42	1	0,04
3. Capex Ratio, %	The ratio between Capex and company's Revenue	16	1	0,04
4. Differentiation of financing, %	The minimum *WACC*	8,3	1	0,05
5. Payback Period, years	The period of time between technology 3G to 4G in Ukraine	3	1	0,05
6. Differentiation of services, (the content of service list)	The maximum service types that are accessed to use at a mobile provider in Ukraine	1	1	0,09
Customers' value				
7. Saving of the subscriber base, subscribers	Subscribers loss of	0	1	0,05
8. Upturn in the subscriber base, %	The maximum upturn in the subscriber base	3	1	0,05
9. Upturn in *ARPU, %*	The maximum upturn in *ARPU*	14	1	0,04
Government's value				
10. Increase of the internet-traffic, %	The maximum yearly upturn	200	1	0,04
11. Tax payments, UAH billion	The place in the taxpayer list (first)	5 047,7	1	0,06
Suppliers' value				
12. Amount and change of material expenses, %	The increasing of material expenses	1	1	0,04
Staff's value				
13. Wage Increase, %	The increase of the company's average salary greater than the average in Ukraine	20	1	0,08
Owners' value				
14. Dividends Increase, %	The year upturn of the dividends	176	1	0,09
15. Competitiveness, %	The maximum share in the telecommunications market	50	1	0,08
16. *EVA*, UAH billion	*NOPAT – WACC*	3724	1	0,1

For example, according to the official data PrJSC "Kyivstar" Disclosure of Information: PrSC "Kyivstar" is the leader by the number of 4G base stations in Ukraine with 7139 stations in P_8 [25]. PrSC "Vodafone Ukraine" had 6058 4G base stations in P_8 [26]. So, the *singular index "Technological leadership"* of PrSC "Vodafon Ukraine" ($I_{TL.VF}$) is calculated thusly:

$$I_{TL.VF} = \frac{NBS_{VF}}{NBS_{BM}}, \qquad (3)$$

where NBS_{VF} – the number of 4G base stations of PrSC "Vodafone Ukraine"; NBS_{BM} – the benchmark measure, the number of 4G base stations of PrSC "Kyivstar".

The singular index "Differentiation of services" is a value index in the metrics "The determinants of the economic value" because capital investments in new technologies and equipment provide modern and qualitative services. The index shows the opportunities of a mobile company's customers to get any mobile service that is accessible in Ukraine.

The singular index "Saving of the subscriber base" can be measured through the existence or absence of the loss of customers. Therefore, if the telecommunications provider in the year of measured has the same or bigger subscriber base, the loss of customer is zero, and the benchmark standard of the index "Saving of the subscriber base" is one.

The singular index "Wage Increase" has a percent of average wage annual increase rate in Ukraine as a benchmark measure. Ukraine has a 20% average wage annual increase rate in P_8. If company has the same increase in wages or it is higher the singular index "Wage Increase" is one.

Table 2 presents the results of the economic value measure of PrSC "Vodafone Ukraine" and PrSC "Kyivstar" investment activities.

The results of measuring the companies' EVIA allow some conclusions. The EVIA integrated score of PrSC "Kyivstar" is 0,84. It is a high economic value level. However, such result shows that there are some problematic aspects. The most problematic aspect is customers' value. The indexes "Saving of the subscriber base" and "Upturn in the subscriber base" got a zero-standard score. It is a problem because they represent 10% of telecommunications provider's EVIA. In addition, a challenging characteristic is "Differentiation of financing". It is general for Ukrainian enterprises to have a difficulty in obtaining sufficient and timely funding. The company has enough high level of WACC and continues to apply traditional tools of financing.

Table 2: The EVIA measuring of PrSC "Vodafone Ukraine" and PrSC "Kyivstar".

The singular value index (measurement units for the singular value indexes see Table 1)	PrSC "Kyivstar"			PrSC "Vodafone Ukraine"		
	FM	SM	SMw	FM	SM	SMw
1. Technological leadership	7139	1	0,1	6058	0,85	0,09
2. *ROI*	39,5	0,9	0,036	10,6	0,25	0,01
3. Capex Ratio	16	1	0,04	21	1	0,04
4. Differentiation of financing	13,1	0,1	0,005	13,5	0,00	0,00
5. Payback Period	1,86	1	0,05	3,17	0,94	0,05
6. Differentiation of services	1	1	0,09	1	1	0,09
7. Saving of the subscriber base	-100	0	0	-1100	0	0,00
8. Upturn in the subscriber base	-	0	0	-	0	0,00
9. Upturn in the *ARPU*	18	1	0,04	17	0,94	0,04
10. Increase of the internet-traffic	150	0,75	0,03	200	1	0,04
11. Tax payments	5047,7	1	0,06	2439,1	0,48	0,03
12. Amount and change of material expenses	1	1	0,04	- 2	0	0,00
13. Wage Increase	29	1	0,08	29	1	0,08
14. Dividends Increase	176	1	0,09	- 29	0	0,00
15. Competitiveness	54,5	1	0,08	36	0,65	0,05
16. *EVA*	4356	1	0,1	-414,6	0	0,00
Integrated score of EVIA	-	12,75	0,84	-	7,46	0,51

FM – *a factual measure of a singular value index;* SM – *a standard measure of a singular value index;* SMw – *a standard measure of a singular value index considering the weight.*

The EVIA integrated score of PrSC "Vodafone Ukraine" allows the definition of the main problematic aspects of the level of invested capital efficiency, customers' value, and differentiation of financing. As a result, the company has a much lower level of EVIA.

5 CONCLUSIONS

This paper considered were approach to measure a company's EVIA. It presents a new understanding of the content of economic value that reflects relationships between investment activities and the financial results of a company's economic activities, considering the value characteristics of the stakeholders.

The developed approach to measure a company's EVIA is universal and, unlike existing ones, is based on six blocks of indicators of economic value (the determinants; Customers' value; Government's value: Increase in internet traffic, Tax payments; Suppliers' value; Staff's value: Wage Increase, Staff's productivity; Owners' value), each of which includes a range of indicators depending on the goals and the object of comparison and involves a combination of linear programming methods with the method of sequential narrowing of the distance from the benchmark, referred to in the paper as the DBM.

The measurement of the EVIA, demonstrated in the article using Ukrainian telecommunications companies as an example, proves the possibility of quantitatively assessing the subjective category of value as a dependence of the outcome (the utility of effective investment activities) on the costs incurred (invested capital, customers' value, and differentiation of financing). Employing the approach utilizing DBM, proposed by the authors of the article, will facilitate solving an applied economic problem related to measuring EVIA, reflecting both the multifaceted nature of investment activities and the ability to compare key characteristics of the activities of several companies simultaneously and identify factors influencing the efficiency of investment activities.

ACKNOWLEDGMENTS

REFERENCES

[1] W. Hamilton, "The Place of Value Theory in Economics," Journal of Political Economy, vol. 26, no. 3, pp. 217-245, 1918, [Online]. Available: https://www.jstor.org/stable/pdf/1821866.pdf.

[2] H. Maoz, H. Tsipi, D. Gefen, and N. Pliskin, "From IT Assets to Business and Economic Value," MCIS 2010 Proceedings, p. 58, [Online]. Available: http://aisel.aisnet.org/mcis2010/58.

[3] A. Orlean, "What is Economic Value Made of?", [Online]. Available: https://doi.org/10.13140/RG.2.1.1990.0648.

[4] J. Malmaeus, "Economic Values and Resource Use," Sustainability, vol. 8, no. 5, p. 490, 2016, [Online]. Available: https://doi.org/10.3390/su8050490.

[5] J. Bluhm and W. Dirk, "Increasing the Economic Value from Digitalisation through Eye-tracking," Proceedings from CHIRA 2019, Vienna, Austria, [Online]. Available: https://doi.org/10.5220/0008364001990205.

[6] R. Shivakumar, "The economics of strategies," California Management Review, Oct. 6, 2020, [Online]. Available: https://cmr.berkeley.edu/2020/10/economics-of-strategies.

[7] "The economic potential of generative AI: The next productivity frontier," McKinsey Insights, June 14, 2023, [Online]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-AI-the-next-productivity-frontier#introduction.

[8] F.S. Dieffenbacher, "Value Creation Definition, Model, and Examples in Business," Digital Leadership, Jan. 15, 2024, [Online]. Available: https://digitalleadership.com/blog/value-creation.

[9] K. Tsagkarakis and I. Katsikis, "The Measurement of Economic Value of Cultural & Environmental Resources: A Critical Review of Evaluation Methods," Proceedings from First International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE), Skiathos, Greece, 2008, [Online]. Available: https://www.researchgate.net/publication/305755427_The_Measurement_of_Economic_Value_of_Cultural_Environmental_Resources_A_Critical_Review_of_Evaluation_Methods.

[10] R. Hejazi, et al., "Measuring the economic values of natural resources along a freeway: a contingent valuation method," Journal of Environmental Planning and Management, vol. 57, no. 4, pp. 629-641, 2014, [Online]. Available: https://doi.org/10.1080/09640568.2012.75862.

[11] E. Estevez, "Economic Value: Definition, Examples, Ways To Estimate," Investopedia, Sept. 1, 2023, [Online]. Available: https://www.investopedia.com/terms/e/economic-value.asp.

[12] G. Otlyvanska, "Investytsiina Diialnist Telekomunikatsiinoho Pidpryiemstva: Teoriia Ta Praktyka," Odesa: Bondarenko M.O., 2019.

[13] F. Fabozzi and J. Grant, "Value-based Metrics: Foundations and Practice," Hoboken, New Jersey: John Wiley & Sons, Inc., 2000.

[14] R. Freeman, "Strategic Management: A stakeholder approach," Boston: Pitman, 1984.

[15] R. Mitchell, B. Agle, and D. Wood, "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts," The Academy of Management Review, vol. 22, no. 4, pp. 853-886, 1997, [Online]. Available: http://www.jstor.org/stable/259247?seq=6#page_sc an_tab_contents.

[16] R. Freeman and J. Mcvea, "A Stakeholder Approach to Strategic Management," SSRN Electronic Journal, 2014, [Online]. Available: https://doi.org/10.2139/ssrn.263511.

[17] D. Volkov, "Teoriya cennostno-orientirovannogo menedzhmenta: finansovyj i buhgalterskij aspekty," 2nd ed., Saint Petersburg, 2008.

[18] T. Donaldson and L. E. Preston, "The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications," The Academy of Management Review, vol. 20, no. 1, pp. 65-91, 1995, [Online]. Available: https://www.jstor.org/stable/258887?seq=24#metad ata_info_tab_contents.

[19] J. Moriarty, "A Theory of benchmarking," Lincoln University, 2008, [Online]. Available: https://pdfs.semanticscholar.org/ec4f/6da18711171 c4dbea74435ad25fa380d5b4e.pdf.

[20] S. Sekhar, "Benchmarking," African Journal of Business Management, vol. 4, no. 6, pp. 882-885, 2010, [Online]. Available: http://www.academicjournals.org/AJBM.

[21] R. Chuhumbaiev, "Metody Kompleksnogo Sravnitelnogo Analiza Effektivnosti Agroformirovanij s Uchetom Specifichnosti Aktivov" (Doctoral dissertation), 2003, [Online]. Available: http://www.dslib.net/?fbclid=IwAR1wBchAxXaGe ZpTigNfz6HZbzKPfS4Ux5OnyysXrFnnKfWd_ujj T9NvYMU.

[22] E. Vyborova, "Osobennosti primeneniya metodov sravnitelnogo ekonomicheskogo analiza pri ocenke finansovogo sostoyaniya organizacii," Ekonomicheskij analiz: teoriya i praktika, vol. 38, no. 389, pp. 22-28, 2014.

[23] C. Scott and O. Vasylenko, "Mathematical and Statistical Methods of Analyzing the Successful Implementation of German-Ukrainian Projects," Proceedings of the 11th International Conference on Applied Innovations in IT, 2023, pp. 151-160.

[24] Complex Statistical Publications, [Online]. Available: https://ukrstat.gov.ua/druk/publicat/kat_u/publ1_u. htm, [Accessed: 25 Dec 2023].

[25] PrJSC "Kyivstar" disclosure of information in the stock market, [Online]. Available: https://kyivstar.ua/uk/about/about/partners/issuers, [Accessed: 12 Jan 2024].

[26] PrJSC "VF Ukraine" information disclosure at the Stock Exchange, [Online]. Available: https://www.vodafone.ua/en/company/investors, [Accessed: 12 Jan 2024].

Forecasting Indicators of the Region Intellectual Potential

Pavlo Hryhoruk, Inna Chaikovska, Oksana Proskurovych, Kateryna Gorbatiuk, Oleksandr Valkov
and Maksym Chaikovskyi

Department of Economics, Analytics, Modeling and Information Technologies in Business, Khmelnytskyi National
University, Instytuts'ka Str. 11, 29016 Khmelnytskyi, Ukraine
hryhorukpm@khmnu.edu.ua, chaikovskaii@khmnu.edu.ua, proskurovycho@khmnu.edu.ua, horbatiukk@khmnu.ua,
valkovo@khmnu.edu.ua, max.chaikovskyi@gmail.com

Abstract: This study established that in the conditions of Russian aggression, one of the most important factors for the recovery of Ukraine's economy is the effective use and multiplication of the existing intellectual potential of the regions. It was determined that the main structural elements of the intellectual potential of the region are educational, scientific and innovative potential. Fractal analysis and forecasting of key indicators of the intellectual potential of the region were carried out. Within the framework of the educational potential of the region, the following indicators were studied: the coverage of children in preschool education institutions, the number of graduate students, the number of students in general secondary education institutions, the number of teachers in general secondary education institutions, the number of students and trainees in vocational and vocational education institutions. As part of the scientific potential, the indicator of the number of employees involved in the performance of scientific research and development was considered. Within the framework of innovation potential - the specific weight of enterprises that introduced innovations and the number of implemented new technological processes. As a result of the conducted fractal analysis, it was established that all indicators belong to the anti-persistent type of time series. Therefore, the method of exponential smoothing and moving average was used as forecasting. A negative trend in the development of the component indicators of the intellectual potential of the region was revealed. This requires appropriate decisive actions on the part of the authorities in order to activate and grow the educational, scientific and innovative potential, which will lead to the improvement of the indicator of the intellectual potential of the region and its positive impact on the indicators of economic efficiency.

1 INTRODUCTION

The modern stage of society's development is characterized by the formation of an innovative economy, in which the role of knowledge and information is growing significantly. This leads to the need for continuous updating of knowledge, accelerated learning, transformation of intellectual resources into a key factor of the region's competitiveness. In this connection, the role of intellectual potential is increasing, and the creation of conditions for its reproduction is becoming a priority area of Ukrainian policy. Many different factors determine the economic growth of the developed countries of the world. One of the most important is the effective realization of the intellectual potential of each country, the recognition of intellectual potential as an indispensable element of social well-being.

Therefore, the issue of forecasting the key indicators of the intellectual potential of the region in order to identify their development trends in the future is quite relevant.

2 LITERATURE REVIEW

A significant number of scientists devoted their research to the assessment and forecasting of intellectual potential and its components [1-8].

In work [1], a system of indices was created based on subject views of experts to assess regional scientific and technical innovation opportunities and collect empirical data for comparison of individual municipalities in Taiwan using the CFPR-VIKOR hybrid approach. This approach is a combined sequential fuzzy preference relationship (CFPR) and

the VIKOR model aimed at prioritizing criteria and decision alternatives. The work [2, 3] predicts the innovative development of regions using correlation analysis, the Bartlett method, and the formation of an integral indicator.

The paper [4] develops a model based on genetic programming to solve the problem of forecasting the effectiveness of regional innovations.

The work [5] predicts the level of development of innovative potential of enterprises using factor analysis, K-means cluster analysis, taxonomy method, discriminant analysis.

The work [6] builds a model of the scenario of mixing in the conditions of innovative changes and market fluctuations and the evolutionary process of the development of the resort and recreation system.

The work [7] is evaluated of enterprise knowledge management system.

In the work [8], the indicators of the innovative attractiveness of the region were forecast.

3   METHODOLOGY

By the intellectual potential of the region, we will understand the possibility of realizing the educational, scientific and innovative potential of the region with the aim of transforming it into an intellectual basis of new quality for solving problems arising in the future.

Therefore, we will analyze the dynamics of indicators of educational, scientific and innovative potential using the example of Khmelnytskyi region in order to forecast them.

Reasoned choice of the forecasting method allows to reflect the existing trends of the process of intellectualization of the economy, identify problematic aspects and, accordingly, make management decisions that will improve the situation. A significant amount of research by scientists is devoted to the study of the issue of dynamic processes in the economy, finance, technology from the point of view of their fractal analysis.

In the general case, the time series can be viewed as a sum of various components: the system component of the trend function – average values over long averaging intervals; cyclic components with a certain repetition period; local features; anomalies of different order; higher-order fluctuations (noise) around all the listed components. The presence of trend and cyclical components requires the use of appropriate approaches to determine the self-similarity of the studied process. The most important concept, from which it is

necessary to start the study of the fractal structure of a random process, is the verification of the hypothesis of the presence of self-similarity (long-term dependence) [9, 10].

We will use Hurst's method to perform a fractal analysis of indicators of intellectual potential according to the algorithm. Having determined the Hurst index for the investigated series of observations, it is possible to choose a method of forecasting the values of the indicator, taking into account the persistence (anti-persistence) of the series.

Algorithm for determining the Hurst indicator (fractal method based on R/S analysis or normalized swing method), which affects the choice of indicator forecasting method, consists of the following stages:

Stage 1. Determination of the accumulated deviation from the average value for the studied indicator:

$$X_{t,N} = \sum_{u=1}^{t} (e_u - M_N)$$

In this formula, N is the length of the period, which varies from 2 to <length of the time series>; t is a variable whose value ranges from 1 to N-1; M_N is the average of N elements; e is a specific element of the time series.

Stage 2. Obtaining N-1 values of $X_{t,N}$ at each iteration:

$$R = Max(X_{t,N}) - Min(X_{t,N})$$

In this formula, R is the range of X deviations.

Stage 3. Normalization of the range by dividing by the standard deviation S, which is found by N values.

Stage 4. We logarithmize R/S and N and build a graph of the function of the dependence of the R/S value in a logarithmic scale on the period in a logarithmic scale based on the data obtained.

Stage 5. On the graph of the function Ln (R/S) from Ln(t), we find the slope by linear approximation. The tangent of the angle of this slope is Hurst's index.

The Hurst index can be in the range from 0 to 1. To characterize the subjects, it can be explained as follows:

1) $(0 \leq H < 0.5)$ – anti-persistent or ergodic time series ("pink noise"), a countertrend is observed, the tendency of the economic system to constantly change the trend (growth is replaced by decline, and vice versa). The stability of such anti-persistent behavior depends on how close H is to zero. The closer its value is to zero, the more variable or volatile the series. This type of system is often called "reversion to the mean";

2) (H = 0.5) – the numerical series is completely random or stochastic ("white noise"), random behavior of the studied indicator is observed;

3) (0.5 < H ≤ 1) – persistent time series ("black noise"), a trend is observed.

Therefore, the deviation of the value of the H indicator from 0.5 indicates the fractal properties of the processes that cause the time series.

The following forecasting methods were also used: the moving average method and the exponential smoothing method.

The moving average model used is as follows:

$$SMA = \frac{\sum_{i=1}^{n} y_i}{n},$$

where SMA is the value of the moving average; n is the smoothing length; y_i is the current value of the dynamics series.

The exponential smoothing model used is as follows:

$$S_t = \alpha \cdot y_t + (1 - \alpha) \cdot S_{t-1},$$

where S_t is the value of the exponential average at time t; y_t is the current value of the dynamics series; S_{t-1} is the value of the exponential average at time t-1; α is the smoothing (damping) coefficient.

4 RESULT AND DISCUSSION

We will apply the Hurst index and the R/S method for selected indicators of the intellectual potential of the region. The following indicators of the educational potential of Khmelnytskyi region were chosen for the study: the coverage of children in preschool education institutions, percent to the number of children of the corresponding age (x_1); the number of graduate students (x_2); the number of students in general secondary education institutions, thousands of people (x_3); the number of teachers in general secondary education institutions, thousands of people (x_4); the number of students and trainees in vocational and vocational education institutions at the end of the year, thousands of people (x_5) (Table 1).

Figure 1 shows a graph of the function of the dependence of the value of R/S in a logarithmic scale on the period in a logarithmic scale for the indicator x_1.

According to Figure 1, it is obtained:

$Ln(R/S)=0,2538\ Ln(t)+0,3389$.

Table 1: Indicators of the educational potential of Khmelnytskyi region, selected for research.

Year	x_1	x_2	x_3	x_4	x_5
1995	44	96	217,5	22,5	16,5
1996	44	131	217,8	21,7	15,7
1997	39	138	218,0	21,5	15,8
1998	39	163	216,3	21,4	15,9
1999	44	184	214,8	21,4	16,0
2000	43	201	212,0	22,0	16,1
2001	45	231	207,8	21,4	15,2
2002	50	258	201,5	21,1	14,7
2003	53	270	194,2	20,6	14,6
2004	55	274	185,9	20,1	14,8
2005	59	316	176,7	20,4	15,3
2006	60	368	168,4	20,8	15,2
2007	61	392	159,9	20,7	14,7
2008	61	436	151,7	20,5	15,0
2009	60	417	144,7	20,5	15,1
2010	61	416	137,7	20,1	15,4
2011	62	411	131,5	19,8	14,8
2012	62	412	128,9	19,8	14,3
2013	64	379	127,3	19,1	13,3
2014	64	367	127,9	18,6	12,6
2015	65	394	128,3	17,9	12,9
2016	66	390	129,2	17,3	10,9
2017	68	411	131,2	17,1	10,1
2018	71	412	134,3	16,8	9,5
2019	72	422	136,8	16,4	8,9
2020	72	393	138,7	16,3	9,1
2021	74	375	138,8	15,9	9,3
2022	68	568	140,3	15,6	9,5
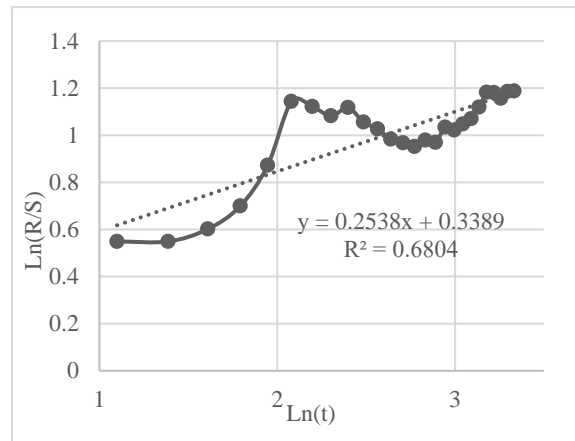
Figure 1: Graph of function Ln (R/S) from Ln(t) for x_1.

Therefore, H=0.2538. The investigated indicator x_1, according to the Hurst indicator ($H < 0.5$), belongs to anti-persistent time series.

According to the above algorithm, intermediate calculations in the R/S analysis are displayed (Table 2).

Table 2: Intermediate calculations in the *R/S* analysis for the indicator (x_1).

Year	x_1	t	Ln(t)	$x_{t,N}$	R	S	R/S	Ln(R/S)
1995	44	1	-	-14,0	-	-	-	-
1996	44	2	0,69	-14,0	0	0,00	0,00	-
1997	39	3	1,10	-19,0	5	2,89	1,73	0,55
1998	39	4	1,39	-19,0	5	2,89	1,73	0,55
1999	44	5	1,61	-14,0	5	2,74	1,83	0,60
2000	43	6	1,79	-15,0	5	2,48	2,01	0,70
2001	45	7	1,95	-13,0	6	2,51	2,39	0,87
2002	50	8	2,08	-8,0	11	3,51	3,14	1,14
2003	53	9	2,20	-5,0	14	4,56	3,07	1,12
2004	55	10	2,30	-3,0	16	5,42	2,95	1,08
2005	59	11	2,40	0,9	20	6,54	3,06	1,12
2006	60	12	2,48	1,9	21	7,30	2,87	1,06
2007	61	13	2,56	2,9	22	7,88	2,79	1,03
2008	61	14	2,64	2,9	22	8,23	2,67	0,98
2009	60	15	2,71	1,9	22	8,36	2,63	0,97
2010	61	16	2,77	2,9	22	8,49	2,59	0,95
2011	62	17	2,83	3,9	23	8,64	2,66	0,98
2012	62	18	2,89	3,9	23	8,72	2,64	0,97
2013	64	19	2,94	5,9	25	8,88	2,81	1,03
2014	64	20	3,00	5,9	25	8,99	2,78	1,02
2015	65	21	3,04	6,9	26	9,12	2,85	1,05
2016	66	22	3,09	7,93	27	9,26	2,92	1,07
2017	68	23	3,14	9,93	29	9,47	3,06	1,12
2018	71	24	3,18	12,93	32	9,81	3,26	1,18
2019	72	25	3,22	13,93	33	10,13	3,26	1,18
2020	72	26	3,26	13,93	33	10,38	3,18	1,16
2021	74	27	3,30	15,93	35	10,69	3,27	1,19
2022	68	28	3,33	9,93	35	10,67	3,28	1,19

In a similar way, it was established that the indicators x_2-x_5 belong to anti-persistent time series.

Consider the indicator of the number of employees involved in the implementation of scientific research and development - in total, persons with scientific potential as x_6 (Table 3).

Table 3: The number of employees involved in the performance of scientific research and development - total, persons (x_6).

Year	x_6
2010	950
2011	925
2012	460
2013	434
2014	439
2015	368
2016	321
2017	380
2018	348
2019	373
2020	233

The following indicators were selected for the study of innovation potential: the specific weight of enterprises that introduced innovations (x_7) and the number of implemented new technological processes (x_8), which are shown in the table 4.

Table 4: The specific weight of enterprises that introduced innovations (x_7) and the number of implemented new technological processes (x_8).

Year	x_7	x_8
2007	6,4	8
2008	3,9	16
2009	5,1	14
2010	16,0	42
2011	20,9	44
2012	22,2	32
2013	17,9	24
2014	9,9	11
2015	12,3	9
2017	5,7	7
2019	6,6	7

Figure 2 shows a graph of the function of the dependence of the value of *R/S* in a logarithmic scale on the period in a logarithmic scale for the indicator x_6, Figure 3 - for the indicator x_7, Figure 4 - for the indicator x_8.

According to Figure 2, it is obtained:

$$Ln(R/S)=0,2929\ Ln(t)+0,7227.$$
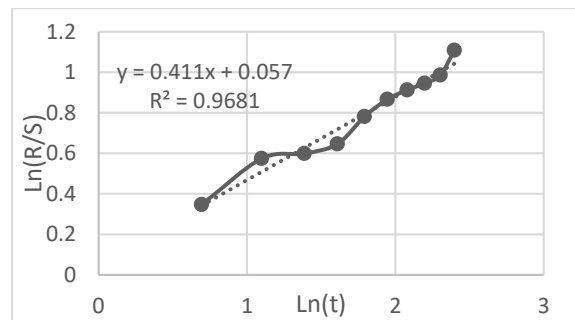
Figure 2: Graph of function Ln (*R/S*) from Ln(t) for x_6.
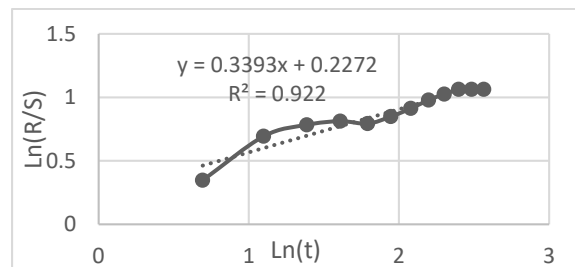
Figure 3: Graph of function Ln (*R/S*) from Ln(t) for x_7.
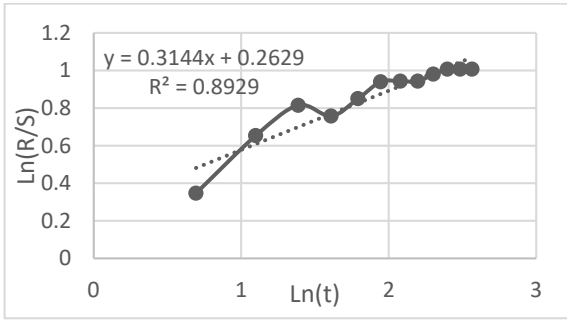
Figure 4: Graph of function Ln (R/S) from Ln(t) for x_8.

According to Figure 3, it is obtained:

$$Ln(R/S)=0,3393 \; Ln(t)+0,2272.$$

According to Figure 3, it is obtained:

$$Ln(R/S)=0,3144 \; Ln(t)+0,2629.$$

Let's summarize the results of the fractal analysis of indicators of intellectual potential (table 5). Therefore, all indicators refer to anti-persistent time series.

Table 4: Results of fractal analysis of indicators.

№	Indicator	H	Time series type
1	x_1	0,2538	anti-persistent
2	x_2	0,2204	anti-persistent
3	x_3	0,1125	anti-persistent
4	x_4	0,3129	anti-persistent
5	x_5	0,2929	anti-persistent
6	x_6	0,4110	anti-persistent
7	x_7	0,3393	anti-persistent
8	x_8	0,3144	anti-persistent

Since it was determined that the studied indicators belong to anti-persistent time series, this type of system is often called "reversion to the mean", so exponential smoothing and moving average should be chosen as forecasting methods.

Moving average models for 3, 4, 5, 6, 7, 8, 9, 10 periods and exponential smoothing models with α equal to 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9 were built for indicators x_1 - x_5.

Figure 5 presents real and model values (moving average models) for x_1.

Figure 6 presents real and model values (exponential smoothing models) for x_1.

For indicators x_6 - x_8, in connection with the published data for a smaller number of years, moving average models for 2, 3, 4, 5, 6, 7, 8 periods and exponential smoothing models with α equal to 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9 were built.
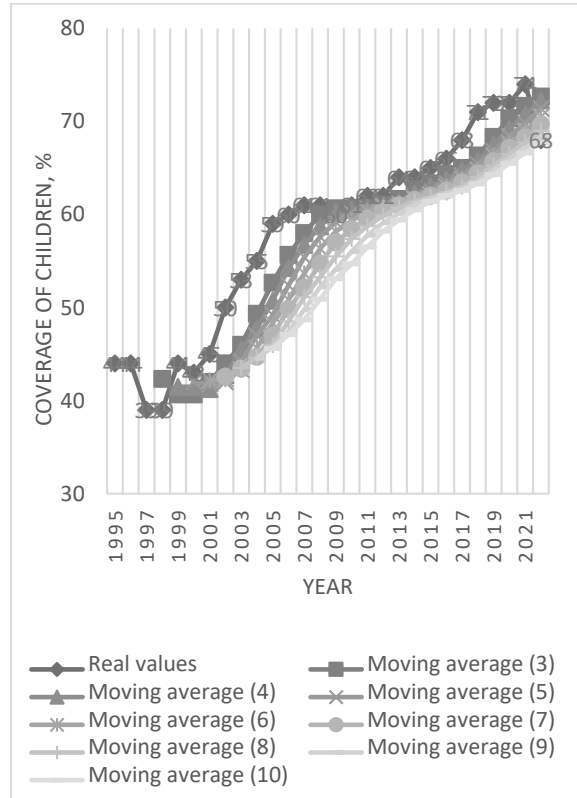
Figure 5: Real and model values (moving average models) for x_1.
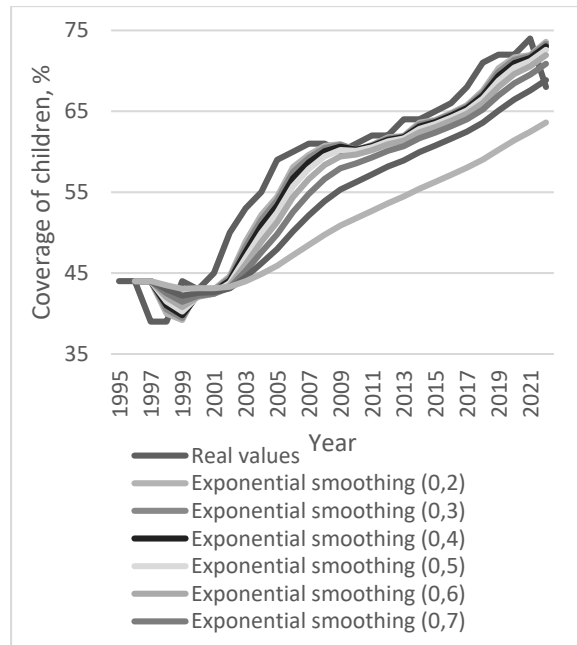
Figure 6: Real and model values (exponential smoothing models) for x_1.

Let's summarize the simulation results in table 5, which shows the R^2 indicator (coefficient of determination) for each of the constructed models (moving average (MA), exponential smoothing (ES). It is the criterion for choosing optimal models for forecasting.

Table 5: R^2 indicators of constructed models.

Model	Indicator							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
MA(2)	-	-	-	-	-	0,43	0,22	0,30
MA(3)	0,92	0,80	0,95	0,95	0,90	0,44	0,01	0,11
MA(4)	0,92	0,75	0,92	0,94	0,87	0,41	0,00	0,12
MA(5)	0,88	0,68	0,88	0,93	0,85	0,08	0,00	0,17
MA(6)	0,87	0,59	0,84	0,93	0,82	0,04	0,00	0,15
MA(7)	0,85	0,49	0,79	0,92	0,80	0,35	0,01	0,36
MA(8)	0,82	0,40	0,16	0,07	0,24	0,56	0,52	0,33
MA(9)	0,78	0,25	0,65	0,91	0,77	-	-	-
MA(10)	0,76	0,11	0,54	0,92	0,78	-	-	-
ES(0,2)	0,94	0,86	0,98	0,97	0,94	0,60	0,41	0,42
ES(0,3)	0,94	0,86	0,97	0,96	0,93	0,60	0,36	0,96
ES(0,4)	0,93	0,86	0,96	0,96	0,93	0,59	0,29	0,88
ES(0,5)	0,92	0,85	0,95	0,96	0,92	0,58	0,21	0,76
ES(0,6)	0,91	0,85	0,93	0,96	0,91	0,56	0,12	0,70
ES(0,7)	0,89	0,83	0,90	0,95	0,90	0,54	0,04	0,70
ES(0,8)	0,86	0,80	0,84	0,95	0,88	0,60	0,00	0,71
ES(0,9)	0,81	0,74	0,74	0,95	0,88	0,60	0,01	0,74
Max R^2	0,94	0,86	0,98	0,97	0,94	0,60	0,52	0,96

According to Table 5, the optimal model for forecasting for x_1, x_3, x_4, x_5, x_6 is exponential smoothing (0.2); for x_2, x_8 - exponential smoothing (0.3); for the x_7 indicator, none of the proposed models is suitable for forecasting according to the low R^2 indicator.

Table 6 summarizes information about the selected models for forecasting the studied indicators, their real values of the last research period and forecast values for the next period, as well as the forecast trend of the indicator to increase or decrease is displayed.

Therefore, the indicator x_2 (number of graduate students, people) will decrease from 568 people in 2022 to 512 people in 2023. Indicator x_3 (number of students in general secondary education institutions - total, thousand people) will decrease from 140.3 thousand people in 2022 to 140.0 thousand people in 2023. Indicator x_4 (number of teachers in general secondary education institutions, thousand people) will increase slightly from 15.6 thousand people in 2022 to 15.7 thousand people in 2023. Indicator x_5 (number of students, students in professional (vocational and technical) education institutions at the end of the year, thousands of people) will remain the same in 2023 as in 2022 at the level of 9.5

thousand people. Indicator x_6 (the number of employees involved in scientific research and development - total, people) will increase from 233 people in 2020 to 266 people in 2021. The x_8 indicator (the number of introduced new technological processes, units) in 2020 will remain at the level of 2019 - 7 units.

Table 6: The results of forecasting the investigated indicators.

№	Indicator	Selected model	Real value	Predictive value	Trend
1	x_1	ES (0,2)	2022 y. 68	2023 y. 69,1	↑
2	x_2	ES (0,3)	2022 y. 568	2023 y. 512	↓
3	x_3	ES (0,2)	2022 y. 140,3	2023 y. 140,0	↓
4	x_4	ES (0,2)	2022 y. 15,6	2023 y. 15,7	↑
5	x_5	ES (0,2)	2022 y. 9,5	2023 y. 9,5	-
6	x_6	ES (0,2)	2020 y. 233	2021 y. 260	↑
7	x_7	-	2019 y. 6,6	-	not established
8	x_8	ES (0,3)	2019 y. 7	2020 y. 7	-

5 CONCLUSIONS

Development of intellectual potential is the result of successful management of economic, ecological, social and cultural life processes of territorial communities and society as a whole. Thus, for the continuous and successful development of the region, it is necessary to constantly increase its intellectual potential, which will contribute to increasing the competitiveness of the resource provision of communities and the country's economy as a whole.

Indicators of innovative potential of Ukraine as a component of intellectual potential were analyzed; the indicators of the scientific potential of Ukraine as a component of the intellectual potential were analyzed; indicators of the educational potential of the Khmelnytskyi region were analyzed. The conducted analysis shows negative trends in the processes of formation and utilization of intellectual potential and its components.

Fractal analysis and identification of trends in the development of indicators of the innovative potential of Ukraine and predicted indicators of the intellectual potential of the Khmelnytsky region were carried out.

As a result of the fractal analysis, it was established that the constituent indicators of the intellectual potential of the Khmelnytsky region are anti-persistent, and therefore it is advisable to use the method of moving average and exponential smoothing for their forecasting.

According to the selected model of exponential smoothing (2), the forecast value of the indicator x_1 of the coverage of children in preschool education institutions (percentage to the number of children of the appropriate age) for 2023 will be 69.1% compared to 2022, in which this indicator was 68%, so the situation is insignificant will improve indicator x_2 (number of graduate students, people) will decrease from 568 people in 2022 to 512 people in 2023. Indicator x_3 (number of students in general secondary education institutions - total, thousand people) will decrease from 140.3 thousand people in 2022 to 140.0 thousand people in 2023. Indicator x_4 (number of teachers in general secondary education institutions, thousand people) will increase slightly from 15.6 thousand people in 2022 to 15.7 thousand people in 2023. Indicator x_5 (number of students, students in professional (vocational and technical) education institutions at the end of the year, thousands of people) will remain the same in 2023 as in 2022 at the level of 9.5 thousand people. Indicator x_6 (the number of employees involved in scientific research and development - total, people) will increase from 233 people in 2020 to 266 people in 2021. The x_8 indicator (the number of introduced new technological processes, units) in 2020 will remain at the level of 2019 - 7 units.

It has been established that the forecast values demonstrate an insufficient level of development of innovative processes at industrial enterprises of Ukraine and the Khmelnytskyi region and a high probability of maintaining similar trends in the future, which will lead to a deterioration of the economic situation both in the region and in the country as a whole. It is possible to come to a conclusion regarding the disappointing trend of the constituent indicators of the intellectual potential of the Khmelnytskyi region, which requires appropriate decisive actions on the part of the authorities in order to activate and grow the educational, scientific and innovative potential, which will lead to the improvement of the indicator of the intellectual potential of the region and its positive impact on the indicators of economic efficiency.

Time to invest in human resources will reveal new priorities for the region. Any development requires investments - intellectual investments justify themselves with the quality of life and longevity of dividends in the form of inexhaustible qualities of intellectual potential.

The obtained research results can be used for: forecasting and improving indicators of the intellectual potential of the region. The models built in the work can be adapted to the operating conditions of any region.

REFERENCES

[1] Y.F. Dai, P.Y. Chu, S.T. Lu, W.T. Chen, and Y.C. Tien, "Evaluation of Regional Innovation Capability: an Empirical Study on Major Metropolitan Areas in Taiwan," Technological and Economic Development of Economy, vol. 28(5), pp. 313-1349, 2022.

[2] K. Boiarynova, O. Popelo, S. Tulchynska, S. Gritsenko, and I. Prikhno, "Conceptual Foundations of Evaluation and Forecasting of Innovative Development of Regions," Periodica Polytechnica Social and Management Sciences, vol. 30(2), pp. 167-174, 2022.

[3] O. Popelo, S. Tulchynska, N. Lagodiienko, M.A. Radin, and A. Moskalenko, "Methodical Approach to Forecasting the Intensification of Innovative Development of Regions Using the Mathcad Program International Journal of Circuits," Systems and Signal Processing, vol. 15, pp. 1591-1601, 2021.

[4] P. Hajek, R. Henriques, M. Castelli, and L. Vanneschi, "Forecasting Performance of Regional Innovation Systems using Semantic-based Genetic Programming with Local Search Optimizer," Computers & Operations Research, vol. 106, pp. 179-190, 2019.

[5] V. Shcherbak and P. Puzyrova, "The Forecasting the Level of Development of Innovative Potential of Textile Enterprises," Management, vol. 2 (32), pp. 65-83, 2020.

[6] P. Zakharchenko, Y. Glazova, S. Zhvanenko, S. Kucher, and V. Mukhin, "Models for forecasting innovative development of the economy of resort-recreational sphere," CEUR Workshop Proceedings, vol. 2927, 2021.

[7] I. Chaikovska, "Evaluation of enterprise knowledge management system," Actual problems of the economy, vol. 10 (172), pp. 221-229, 2015.

[8] L. Lisovska, O. Mrykhina, A. Terebukh, and N. Kozmuk, "Modeling of financial results of subjects of the innovation process in the region", Financial and credit activity: problems of theory and practice, vol. 3(44), pp. 122-131, 2022.

[9] H.E. Hurst, "The Long-Term Storage Capacity of Reservoirs," Transactions of the American Society of Civil Engineers, vol. 116, pp. 770-799, 1951.

[10] H.E. Hurst, "Methods of using long-term storage in reservoirs," Proceedings of the Institution of Civil Engineers, p. 1, pp. 519-577, 1955.

SECTION 3

Statistical Modelling and Analysis and their Applications in Life Sciences

MARCH 2024

Computer Simulation of Critical Phenomena in Materials of Cyber Systems Elements

Hennady Shapovalov, Anatoly Kazakov and Gleb Ksendziuk

Institute of Information Security, Radioelectronics and Telecommunications, Odesa Polytechnic National University,
Shevchenko Avenue 1, 65044 Odesa, Ukraine
shapovalov@op.edu.ua, kai@op.edu.ua, ksendziuk.8142006@stud.op.edu.ua

Keywords: Phase Coexistence, Multicomponent Solid Solutions, Computer Simulation, Cyber Systems Elements.

Abstract: The safety and reliability of the operation of materials on which modern cyber systems are developed and new autonomous cyber devices are created that are capable of performing tasks in modern combat conditions was considered. Mathematical modeling of critical phenomena in semiconductor solid solutions of elements of groups III – V of the periodic system under various operating conditions was performed. The properties of the $A_xB_{1-x}C_yD_{1-y}$ type semiconductor solid solutions were predicted. The spaces, in which the conditions of stable and unstable phases are satisfied, as well as areas of coexistence of several phases simultaneously, were determined on sections of phase diagrams. A computer modeling technique for the formation of critical spaces, based on the use of a differential topological approach in complex multicomponent systems, was used. A method for determining the total derivatives of the free energy of quaternary solid solutions with mixing in two sublattices from the third to the eighth inclusive using the regular solution model and standard thermodynamic functions for binary compounds was applied to describe the interaction of atoms in multicomponent solid solutions. Sections of the phase diagram of the In-Ga-As-P system, critical spaces and phase coexistence spaces in four-dimensional space for various temperatures were calculated. The possibility of forming regions of coexistence of phases of different orders in solid solutions $In_xGa_{1-x}As_yP_{1-y}$ is shown by the obtained modeling results. The results obtained are consistent with the available experimental data, in which spatial modulation of the composition of the solid solution was observed.

1 INTRODUCTION

The safety and reliability of the operation of modern cyber systems and the creation of new autonomous devices largely depends on the properties of the materials on the basis of which they are developed. The scope of use of modern automated systems has expanded significantly, especially in modern combat conditions. In this regard, interest in predicting the properties of semiconductor compounds under various operating conditions is growing. Epitaxial layers of semiconductor solid solutions grown in unstable regions exhibit, at certain temperatures and compositions, a tendency to form structures with a periodic composition distribution [1–5]. Such phase changes are extremely difficult to identify under operating conditions, since this requires complex laboratory equipment. As a consequence, the parameters of devices created on the basis of multicomponent solid solutions can change

significantly and lead to errors in the operation of the systems. At the same time, new functional electronic materials that are stable under critical conditions are currently attracting significant interest. New areas of application of multicomponent semiconductor compounds for modern micro- and nanoelectronics generate interest in understanding the problem of self-organized formation of ordered structures. The study of these phenomena is important, first of all, for modern technologies of microstructures and nanoobjects. These are, first of all, multicomponent semiconductor solid solutions with a non-uniform distribution of component concentrations, such as periodic structures, superlattices, systems of "quantum dots", etc. To analyze the processes of the emergence of self-organized ordered structures, it is possible to use computer modeling based on thermodynamic approaches. The calculation of multidimensional phase diagrams, taking into

account the possibility of the existence of bifurcation spaces, critical spaces and spaces of coexistence of phases of different orders, should be used for such an analysis [6–8]. The equation of state of matter specifies some m-dimensional manifold in the corresponding space in the thermodynamic description of phase transitions. The equation of state of matter in the thermodynamic description of phase transitions specifies some m-dimensional manifold in the corresponding space, which is assumed in this article. Thom's catastrophe theory in the case of using one order parameter is considered as a generalized form of the Ginsburg-Landau theory of phase transitions for assessing the features of the potential function of a self-organizing system. However, to describe possible phase transitions in multicomponent systems, an approach is used that allows one to analyze the features of potential functions of several order parameters, allowing one to estimate the compositions of coexisting phases.

2 PURPOSE OF THE WORK AND RESEARCH METHODS

The purpose of the work is to further develop methods for mathematical modeling of phase equilibrium in multi-component systems to predict the reliability of electronic components in modern cyber systems. Certain provisions of the theory of bifurcations and the theory of Thom's catastrophes were used to analyze phase states, allowing one to analyze the features of potential functions of several order parameters [7 - 9]. Four-component solid solutions of elements of groups III-V of the periodic table with mixing of atoms in two sublattices of the form $A_xB_{1-x}C_yD_{1-y}$ were chosen as the object of study. It is these materials that are widely used in modern microelectronics and integrated optics, in the practice of creating large and high-speed integrated circuits, as well as optoelectronic devices, and form the basis of modern optical radiation detectors (photodetectors) for a wide spectral range. Such materials have broad prospects for creating elements of cyber systems operating in critical conditions. Modeling the process of formation of critical spaces and spaces of coexistence of phases in multicomponent and multiphase systems is possible based on the use of a differential topological approach [9]. The space of the phase coexistence occurs when one stable state coexists with another stable state. The first-order phase transition corresponds to the appearance of such a space and is

determined by Maxwell's principle when two (or more) global minima of the potential function have the same depth. At some points in the space under study, a stable phase may become unstable, forming a bifurcation subspace. Two phases in a certain region can be identical for certain values of the order parameter and form a critical space of order two. In the presence of three or four identical phases, critical spaces of order 3 or 4 are formed, respectively. The conditions for the existence of singularities in this case are determined by the free energy of the system G [7, 9]:

- stable phase:

$$\frac{dG}{dx} = 0 \quad ; \quad \frac{d^2G}{dx^2} > 0 , \qquad (1)$$

- instability space or bifurcation space:

$$\frac{dG}{dx} = \frac{d^2G}{dx^2} = 0 \quad ; \quad \frac{d^3G}{dx^3} > 0 , \quad (2)$$

- condition for the existence of a second-order critical space:

$$\frac{dG}{dx} = \frac{d^2G}{dx^2} = \frac{d^3G}{dx^3} = 0 \quad ; \quad \frac{d^4G}{dx^4} > 0, \quad (3)$$

- condition for the existence of a third-order critical space:

$$\frac{dG}{dx} = \frac{d^2G}{dx^2} = . \quad . \quad . = \frac{d^5G}{dx^5} = 0 \quad ; \quad \frac{d^6G}{dx^6} > 0, \quad (4)$$

- condition for the existence of a fourth-order critical space:

$$\frac{dG}{dx} = \frac{d^2G}{dx^2} = . \quad . \quad . = \frac{d^7G}{dx^7} = 0 \quad ; \quad \frac{d^8G}{dx^8} > 0, \quad (5)$$

where $\frac{d^iG}{dx^i}$ are the matrices of derivatives of the free energy of the system with respect to the corresponding components $x(x^1, x^2, x^3, x^4)$.

3 PRESENTATION OF THE MAIN MATERIAL

Generalization of (1-5) to the case of m-dimensional concentration space allows us to analyze possible critical spaces and spaces of coexistence of phases in multicomponent solid phases based on the study of higher derivatives with respect to order parameters xi from the free energy of the system. An analysis of the existing principles of using multidimensional matrices and obtaining matrix derivatives was carried out [9, 12]. The possibility of expanding the free energy of multicomponent systems into a

multidimensional Taylor series was considered. The method of matrix-vector differentiation of multidimensional systems using the method of direct sums was used to obtain higher derivatives of the free energy of a four-component system. Derivatives of the free energy expression with respect to the corresponding concentration parameters were calculated from the third to the eighth inclusive.

Four-component solid solutions with a mixture of atoms in the cationic and anionic sublattices of the form $A_xB_{1-x}C_yD_{1-y}$ were presented as a mixture of four hypothetical binary compounds with concentrations X_{ij}. Accordingly, the free energy of such a solid solution will be a function of four concentration parameters:

$$G=G(X_{AC}, X_{BC}, X_{AD}, X_{BD}), \qquad (6)$$

and the matrix of second derivatives of free energy with respect to these parameters will have a dimension of 4 x 4. The model of a strictly regular solution (6) was used in the work, based on the assumption of a random distribution of dissimilar atoms over the sites of the corresponding sublattices. The equation for the equilibrium number of binary clusters for modeling phase states according to the chosen model was taken in the form [6]:

$$\frac{X_{AD}X_{BC}}{X_{AC}X_{BD}} = \exp\left(\frac{1}{4kT}\left(\Delta G_1 + \Delta G_2\right)\right), \quad (7)$$

where ΔG_1 (7) is the change in free energy in the substitution reaction $AD + BC = AC + BD$, taking into account the interaction in the first configuration sphere, and ΔG_2, respectively, taking into account the interaction in the first and second coordination spheres:

$$\Delta G_1 = \mu_{AC} + \mu_{BD} - \mu_{Ad} - \mu_{BC},$$
$$\Delta G_2 = \alpha_{AC-BC} \frac{X_{BC} - X_{AC}}{X_{BC} + X_{AC}} + \alpha_{AC-AD} \frac{X_{AD} - X_{AC}}{X_{AD} + X_{AC}} \qquad (8)$$
$$+ \alpha_{AD-BD} \frac{X_{AD} - X_{BD}}{X_{AD} + X_{BD}} + \alpha_{BC-BD} \frac{X_{BC} - X_{BD}}{X_{BC} + X_{BD}}$$

On sections of phase diagrams, expressions for the concentrations of binary components to display the modeling results were written through the concentration parameters x and y:

$$X_{AC} = (1-x)(1-y); \ X_{AD} = (1-x)y;$$
$$X_{BC} = x(1-y); \ X_{BD} = xy. \qquad (9)$$

The determination of higher derivatives of the free energy of the system to analyze the possibility of the existence of critical subspaces was carried out

as follows. Symbolic elements of the matrix of second derivatives of the free energy of a solid solution of the form $A_xB_{1-x}C_yD_{1-y}$ were used as initial relations. The result of Kroniker multiplying the matrix of second derivatives of the free energy by differentiation operators with respect to the corresponding arguments was used to obtain the third derivative of the function of m arguments. Next, block matrices of the complete third derivative using the direct sum method were obtained containing m blocks of partial derivatives of size (m × m). Calculations of derivatives of the free energy of a multicomponent system of higher orders were carried out using a similar algorithm. Symbolic differentiation and numerical calculations of derivatives were performed using the Maxima computer mathematics system [11].

4 RESULTS

Analytical expressions for the total derivatives of the Gibbs free energy of four-component homogeneous solid solutions of the type $A_xB_{1-x}C_yD_{1-y}$ were obtained in the work starting from the third derivative to the eighth inclusive. Numerical calculations of the position of the regions of existence of derivatives and zero contours of derivatives from the second to the eighth inclusive were carried out on the cross section of the existence of solid solutions of the phase diagram of the studied system $In_xGa_{1-x}As_yP_{1-y}$ for the temperature range 773K – 1023K. Thermodynamic parameters used in the calculations (8) are given in the Table 1.

Table 1: Parameters of interaction in the solid phase for quasi-binary systems (sources [12, 13, 14]).

Quasi-binary system	Interaction parameter α_{ij-jk}^s J/mol
GaP - InP	14665
GaAs - InAs	12570
GaP – GaAs	1676
InP - InAs	1676

The contours of the spaces of coexistence of phases of orders two, three and four were constructed on the cross section of the existence of solid solutions in the phase diagram of the system under study based on the calculations performed for different temperatures. The obtained results of calculating the position of the zero contours of the

derivatives of the Gibbs free energy of the analyzed system according to the concentration parameters are presented in Figures 1 – 4, starting from the second derivative and up to the eighth derivative inclusive.
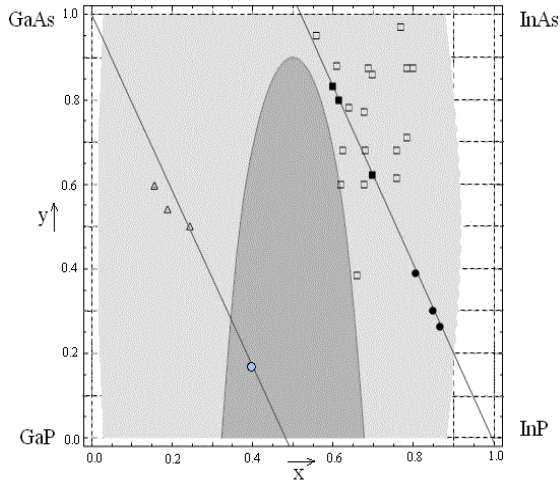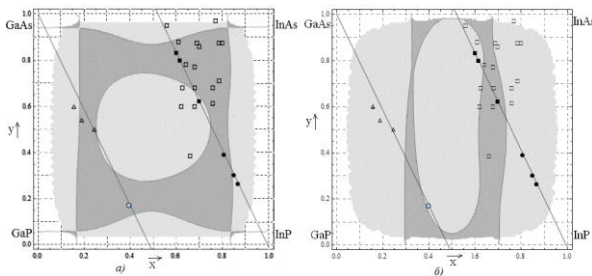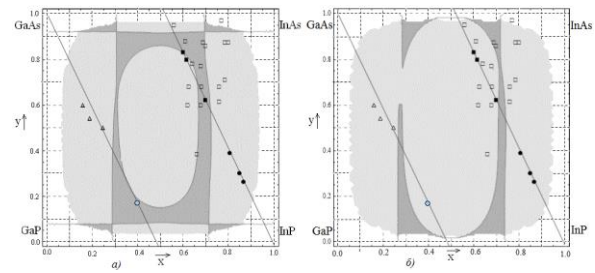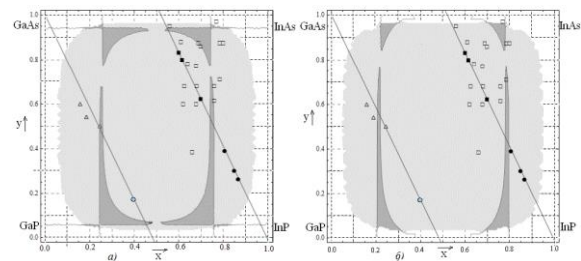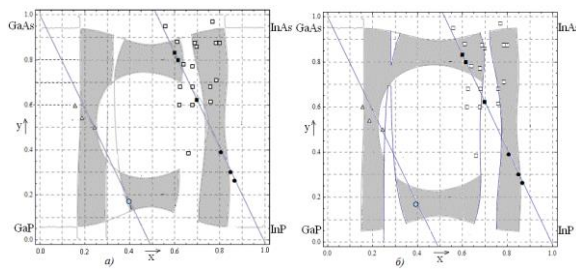

Figure 1: Cross section for the existence of solid solutions in the phase diagram of the In-Ga-As-P system. The results of numerical calculations of the second derivative of the free energy of the system are presented. Isoperiodic composition lines for GaAs and InP substrates are shown. The space of negative derivative values is shown in dark color. The space of positive derivative values is shown in light color. Temperature 773K. Sources of experimental data are: ■-[1], ● - [2], ○ - [3], △ - [4], □ - [5].

Figure 2: Cross section for the existence of solid solutions in the phase diagram of the In-Ga-As-P system. The results of numerical calculations of a) the third; b) fourth derivatives of the free energy of the system are given. The space of negative values of the corresponding derivative is shown in dark color. The space of positive derivative values is shown in light color. Temperature 773K. Sources of experimental data are: ■-[1], ● - [2], ○ - [3], △ - [4], □ - [5].

Thus, the position of space in which the conditions of the stable phase are satisfied in accordance with (1, 9) is shown in Figure 1. The obtained positions of the zero contours for the third

and fourth derivatives of the free energy of the system with respect to the concentrations of binary components for the $In_xGa_{1-x}As_yP_{1-y}$ solid solution (Figure 2) show that most of the experimental points for which composition modulation is observed fall into the region of negative values of the third derivative. These points are located in the region of positive values of the fourth derivative, which corresponds to the fulfillment of conditions (3) for the formation of spaces of coexistence of phases of order two.

Figure 3: Cross section for the existence of solid solutions in the phase diagram of the In-Ga-As-P system. The results of numerical calculations of a) the fifth; b) the sixth derivatives of the free energy of the system are shown. The space of negative values of the corresponding derivative is shown in dark color. The space of positive derivative values is shown in light color. Temperature 773K. Sources of experimental data are: ■-[1], ● - [2], ○ - [3], △ - [4], □ - [5].

Figure 4: Cross section for the existence of solid solutions in the phase diagram of the In-Ga-As-P system. The results of numerical calculations of a) the seventh; b) the eighth derivatives of the free energy of the system are shown. The space of negative values of the corresponding derivative is shown in dark color. The space of positive derivative values is shown in light color. Temperature 773K. Sources of experimental data are: ■-[1], ● - [2], ○ - [3], △ - [4], □ - [5].

The calculation results for a temperature of 773K are shown in Figures 1 – 4. Lines of isoperiodic compositions for GaAs and InP substrates and experimental compositions of epitaxial layers $In_xGa_{1-x}As_yP_{1-y}$, in which the formation of periodic modulated structures was observed, are shown.

These compositions were obtained using liquid-phase epitaxy on InP [1, 2] and GaAs [3, 4] substrates and gas transport epitaxy on InP substrates [5]. In works [1-5], cases of the formation of structures with simple one-dimensional modulation [1, 3], complex modulation of composition are described, such as the presence of two different concentration wavelengths in different crystallographic directions [2] and the formation of a system of alternating solid solution domains of two different compositions [4, 5]. The results of studying the structure of $In_xGa_{1-x}As_yP_{1-y}$ epitaxial layers obtained by vapor deposition on InP substrates at a temperature of 973K are presented in [5]. The formation of quasi-periodic structures of composition concentrations with a modulation wavelength from 10 – 20 nm along the {110} direction to 80 nm along the {001} direction for the concentration ranges of the resulting solid solutions (0.2 < x < 0.53 and 0.37 < y <1) was observed. As a result of additional experimental studies, it was shown that these composition modulations are not a consequence of the instability of the growth process [5]. The fact that the growth conditions of epitaxial layers do not affect the composition modulation period has also been shown [1].

Figure 5: Cross section for the existence of solid solutions in the phase diagram of the In-Ga-As-P system. Compositions for which the conditions for the formation of phase coexistence spaces of order two are met for temperatures a) 773 K; b) 1050 K are shown in dark color.

5 CONCLUSIONS

The obtained modeling results based on the proposed approach make it possible to explain the available experimental data on the spontaneous formation of two-phase modulated structures in epitaxial layers of the $In_xGa_{1-x}As_yP_{1-y}$ solid solution. Figure 5 shows that the regular solution model used to construct the Gibbs free energy of a solid solution allows for a fairly correct assessment of the position of coexistence spaces of order 2 in phase diagrams.

An increase in temperature leads to a decrease in the resulting phase coexistence space. Based on the obtained results of modeling the position of regions of coexistence of phases of order 2, it can be assumed that the proposed model makes it possible to predict the spaces of possible appearance of periodic modulated structures in the considered solid solutions. The possibility of such a phenomenon should be taken into account in documentation on the safe use of the studied materials in modern cyber devices operating in extreme conditions.

REFERENCES

[1] P. Henoc, A. Izrael, M. Quillec, and H. Launois, "Composition modulation in liquid phase epitaxial InxGa1-xAsyP1-y layers lattice matched to InP substrates", Appl. Phys. Let., vol. 40, pp. 951-963, 1982.

[2] S. Mahajan, B.V. Dutt, and H. Temkin, "Spinodal decomposition in InGaAsP epitaxial layers", J. Crystal Growth, vol. 68, pp. 589-595, 1984.

[3] N.Kuwano, K.Funuka, and K. Oki, et. al., "Electron microscope study of modulated structures and heterointerfaces in LPE-grown GaInAsP layers lattice matched on GaAs", J. Crystal Growth, vol. 98, pp. 82-89, 1989.

[4] L.S. Vavilova, V.A. Kapitonov, and A.V. Murashova, "Spontaneously formed periodical InGaAsP-structures with a modulated composition", Fizika i tekcnika poluprovodnikov, (Physics and technique of semiconductors), vol. 33, pp. 1108-1110, 1999.

[5] S. Chu, S.Nakahara, K. Strege, and W.Johnston, "Surface layer spinodal decomposition in In1-xGaxAs1-yPy and In1-xGaxAs grown by hybrid transport vapor phase epitaxy", J.Appl.Phys., vol. 57, pp.4610-4616, 1985.

[6] K. Onabe, "Thermodynamics of the type A1-xBxC1-yDy, III-V quaternary solid solutions" J.Phys.Chem.Solids, v. 43, pp. 1071-1086, 1982

[7] A.I. Kazakov, G.V. Shapovalov, and P.P. Moskvin, "Computer simulation for formation of critical spaces in II–VI solid solutions", Journal of Crystal Growth, vol. 506, pp. 201-205, 2019.

[8] P.P. Moskvin, S.I. Skurativskyi, O.P. Kravchenko, G.V. Skyba, and H.V. Shapovalov, "Spinodal decomposition and composition modulation effect at the lowtemperature synthesis of semiconductor solid solutions", Journal of Crystal Growth, vol. 510, pp. 40-46, 2019.

[9] K. Okada and I. Suzuki, "Classical calculations on the phase transition I. Phase diagram in four-dimensional space for the system with one order parameter", J. Phys. Soc. Jap., vol. 51, no 10, pp. 3250-3257, 1982.

[10] I. Traat, "Matrix calculus for multivariate distributions", Transactions of Tartu State Univ., no 733, pp. 64-85, 1986.

[11] Maxima, a Computer Algebra System, [Online]. Available: http://maxima.sourceforge.net.

[12] R Sankaran and G. Antypas, Journal of Crystal Growth, vol. 36, pp. 198-206, 1976.

[13] M. Panish and M. Ilegems, Progres in Solid State Chemistry, Pergapon, Oxford, vol. 7, pp. 39-47, 1972.

[14] K Sakai and A. Ariba, Japan Journal Appl. Phys., vol. 16, pp. 2043-2051, 1977.

Scenario Modelling of University International Activities Based on Fuzzy Cognitive Maps

Anastasia Krivtsun and Daria Kushchiy

Department of Information Technologies and Management, Platov South-Russian State Polytechnic University (NPI),
Prosveshchenie Str. 132, 346428 Novocherkassk, Russia
anastasia.srstu@gmail.com, dkushchiy@rambler.ru

Keywords: International Activities, Fuzzy Cognitive Maps, Scenarios, System Analysis, Cognitive Modelling.

Abstract: In this article the concept of managing the effectiveness of international activities of universities based on fuzzy cognitive maps is discussed. The focus on the international relations currently plays an important role in almost all universities, as well as in the general context of higher education policy. However, there is a significant difficulty in analyzing the international activities of a university and making management decisions in this area due to the multifaceted nature of the processes occurring in international activities and their interconnectedness, due to the complexity of studying individual phenomena, and the lack of sufficient quantitative information about the dynamics of processes. In this article for analyzing the international activities of universities, as a weakly structured system, used cognitive analysis. A cognitive model for managing the international activities of universities has been developed. To build this model, we took into account the quality criteria for formalizing experts' ideas about the international activities of universities, as well as statistical data from universities. Using the developed cognitive model for managing the effectiveness of international activities of universities, it is possible to forecast the self-development of the situation; it is possible to forecast the situation with fixed control; it is possible to implement control that ensures the implementation of the required or desired scenario. Several Scenarios for managing the international activities of a university based on the developed cognitive model using the FCMapper tool are considered.

1 INTRODUCTION

The focus on the international relations currently plays an important role in almost all universities, as well as in the general context of higher education policy. However, there is a significant difficulty in analyzing the international activities of a university and making management decisions in this area due to the multifaceted nature of the processes occurring in international activities and their interconnectedness, due to the complexity of studying individual phenomena, and the lack of sufficient quantitative information about the dynamics of processes. Due to these features, the system of international activities of a university can be classified as weakly structured structures.

Several works devoted to study of internationalization and problems of development of international activities of universities [1, 2], consider issues of information support for managing international activities [3, 4] and decision-making systems in managing academic mobility using expert systems technologies, knowledge engineering and processing semi-structured information, as well as reasoning based on precedents [5]. A comprehensive study of the international activities of universities has been done [6]; indicators and indicators used to assess international activities in German universities are considered [7]; modeling of international activities was carried out and the degree of influence of internationalization on the strategic development of universities was determined [8].

Issues of integrated management of international activities are not sufficiently developed. A large number of factors that must be taken into account when assessing the effectiveness and forecasting the international activities of a university requires the use of intelligent methods of information processing and its presentation.

In this article, cognitive modelling is used to analyze and study the international activities of a university.

Cognitive (scenario) modelling is a tool that allows you to make flexible long-term plans and

think through how an organization should respond to future events

There are different approaches to developing scenarios:

- Descriptive approach.
- Morphological analysis.
- Mathematical, simulation modelling.

The first approach involves a verbal description of events and the development of the situation.

The second approach is to compile a table in which morphological chains are selected that make it possible to clarify the qualitative characteristics of the system corresponding to various scenarios for the development of the situation.

The concept of the third approach based on a situation model under conditions of risk and uncertainty, the dynamics of output parameters are studied based on changes in input parameters, connections and structural characteristics of the system [9].

One of the most effective approaches is to construct scenarios based on cognitive modelling. It combines the features of all approaches, namely, it contains a comprehensive description of the situation inherent in qualitative methods, as well as clarity, the ability to conduct simulation analysis and forecasting, inherent in formalized methods.

Cognitive modelling in scenario planning processes thus allows one to take into account subjective and objective factors both under conditions of certainty and under conditions of risk and uncertainty. The methodology proposed below for modelling scenarios for the development of the international activities of the university situation ensures the implementation of a systems approach and system analysis and thereby ensures the integrity of the approach to modelling, the connection of theoretical principles with practical problems in economics and management.

Cognitive modeling contributes to a better understanding of the problem situations, identifying contradictions and qualitative analysis of the system. Target modeling consists in forming and clarifying the hypothesis about functioning of the object under study, considered as a complex system, which consists of separate, but still interconnected elements and subsystems. Based on the above, we can conclude that the construction cognitive maps and models for analyzing the international activities of a university as a weakly structured system is the optimal solution.

2 COGNITIVE MAPS AS A TOOL FOR SCENARIO MODELLING

A cognitive map (CM) belongs to a family of models for representing expert knowledge in the form of a structure of cause-and-effect influences of factors characterizing the object of research, its external environment (for example, economic, state, etc.) and the interests of active subjects of the situation [10].

The influence of a factor on a factor is established using linguistic variables that indicate the nature and strength of the influence (for example, consumer demand can be "weak", "moderate", "exciting" and the like). The construction of the CM is carried out taking into account the substantive interpretability of the connections between factors entered into the CM, that is, to build a cognitive model using the CM, the verbal states of the factors are mapped to numerical ones from the interval [0, 1] (or [−1, 1]), and the verbal representations of influences are to numbers from the interval [−1, 1].

In modern research [11, 12, 13], to describe changes in a factor, a dynamic equation is written, the structure of which reflects the direct influence on this factor of other factors, as well as factors of the "external environment" (sometimes cognitive models use functional graphs in which the "power of influence" of one factor on the other is a known function). The set of such equations for all QC factors represents a cognitive model of the situation under study.

At the initial stage of constructing a CM, it is essential to identify important factors that influence and determine the situation under study. In works [14, 15], devoted to the use of models based on cognitive maps for the study of weakly structured structures, it is noted that four approaches are used to construct maps:

1) Identification of factors and connections through content analysis of documents;
2) Identification of factors and relationships based on conceptual frameworks (usually PEST analysis and/or SWOT analysis);
3) Identification of factors and connections through analysis of expert views;
4) Identification of factors and relationships through the analysis of quantitative data, for example, regression analysis of time series of parameters.

A distinctive feature of the development of scenarios based on   s is the possibility of

representing, along with objective factors, subjective opinions of experts regarding a particular situation, as well as combining them to create an integral fuzzy cognitive map.

2.1 Fuzzy Cognitive Maps

Fuzzy cognitive maps (FCM) are a way of representing real dynamic systems in a form that corresponds to human perception of such processes [16]. This is the main reason for their widespread use in various spheres of life.

FCM represents a system as a combination of concepts and the various relationships that exist between the concepts. The FCM consists of nodes (N_1, N_2, …, N_n), which represent the important elements of the system being mapped, and directed arcs (e_{ij}), which represent the cause-and-effect relationships between two nodes (N_i, N_j).

Directed arcs are assigned fuzzy values in the interval [-1, +1], which show the "strength of influence" between factors. A positive value indicates a positive cause-and-effect relationship between two factors, a negative value indicates a negative cause-and-effect relationship between two factors, and a zero value corresponds to the absence of connections between the factors under consideration [17].

The adjacency matrix provides insight into the cognitive map. Based on it, you can calculate the measure of concept centrality, as well as obtain information about direct and indirect cause-and-effect relationships in the map.

FCM, first proposed by Bartolomew Kosko in 1986, best reflect uncertainty, the dynamics of the states of concepts and connections between them.

FCM is an extension and improvement of the cognitive map with the additional ability to model complex chains of cause-and-effect relationships through weighted cause-and-effect relationships [18].

The FCM model describes the behavior of the system, and each concept represents a factor characteristic of the system.

Modelling using fuzzy cognitive maps is a combination of fuzzy logic and cognitive modelling. This is a way to imagine a system under conditions of uncertainty and complexity, when formal logic does not work.

Although fuzzy cognitive maps are computationally similar to an artificial neural network, there are differences between the two models. Fuzzy cognitive maps can be purely expert in nature (although they can also be trained) and

correspond to a "white box" model, while an artificial neural network is fundamentally oriented towards learning (a "black box" model).

2.2 Modelling Tool Based on Fuzzy Cognitive Map

Fuzzy cognitive map modelling can be done using the Excel spreadsheet-based tool FCMapper.

This program allows to calculate all significant indices (incoming, outgoing, overall centrality, density), the number of concepts, connections between them, determine the type of factors (sender, recipient, regular), calculate their number, thereby giving a general characteristic of the network. It is possible to conduct a simulation analysis of the behavior of a specific system and to see its development trends. FCMapper creates a file that, using special network analysis programs, e.g.Pajek, allows to visualize a fuzzy cognitive map [19].
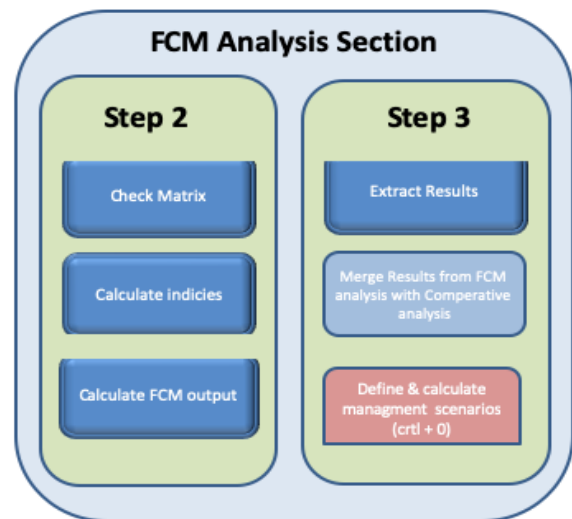
Figure 1: FCMapper interface.

2.3 Basic Scenario Modelling Algorithm Based on Fuzzy Cognitive Maps

Figure 2 shows an FCM-based scenario modelling diagram that includes six main steps. Let's take a closer look at them.
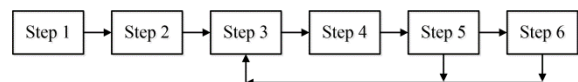
Figure 2: FCM-based scenario simulations.

Step 1. Preparing the scenario. Determining the purpose, timing and boundaries of scenarios.

Step 2. Data collection. Determining the relevant concepts of the cognitive map through studying the literature, interviewing experts; combining cognitive models of various experts; transformation of the combined model into a fuzzy cognitive map, on the basis of which scenarios will be developed.

Step 3. Scenario Modelling. Simplification of cause-and-effect relationships, determination of their weights, selection of a transformation function for assessing the state of concepts.

Step 4. Develop scenarios. Calculation of parameters of a fuzzy cognitive model for various vectors reflecting possible combinations of initial states of concepts.

Step 5.Selection and refinement of scenarios. The scenarios obtained in the fourth step are further evaluated and refined.

Step 6. Strategic decisions. The developed scenarios are used to make strategic decisions.

If the input of additional knowledge from experts or the results of further analysis lead to a new understanding of the situation, then the fuzzy cognitive models built in the third step can be adjusted and recalculated in the fourth step to identify a possible future. Thus, FCM-based scenario modelling becomes a dynamic process.

3 ASSESSMENT OF INTERNATIONAL ACTIVITIES OF THE UNIVERSITY BASED ON FUZZY COGNITIVE MAPS

Let us consider the international activities of the university, paying special attention to external environmental factors at the macro and micro levels, as well as how changes in their condition affect the dynamics of the enterprise's profit.

Factors determined at the meso level make it possible to assess the international activities of each individual university, and factors determined at the macro level make it possible to assess the impact of the country's social and economic activities on the international activities of the university in question.

By the concept of "major factors" we will consider those that characterize the effectiveness of the international activities of a university, and by "minor factors" we will consider those on the variation of which changes in the main factors depend.

3.1 Collection and Processing of Experimental Data

During the research, several experts in the field of international education were interviewed (director of the Institute of International Education of South Russian State Polytechnical University, E.V. Kirievsky; head of the sector of the Institute of International Education of South Russian State Polytechnical University, Tokmakov G.E.; head of the department of international activities of Kazan National Research Aviation University, Gilmetdinova A.M.; deputy head of the department of international activities of Kazan National Research Aviation University, Snegurenko A.P.; head of the international relations department of Volgograd State Technical University, Boyko G.V.).

The survey was conducted using a developed application that allows the expert to establish the significance of the influence of factors on the international activites of a university (major, secondary or insignificant factor), to establish the degree of interaction between factors (definition of the matrix $A = \|a_{kl}\|_{M \times M}$,where elements a_{kl} [-1, 1] reflect direct impact l-th factor to k-th factor, as well as ensure the fulfillment of formalization reliability criteria for models based on cognitive maps and the conditions for their use for protection against risks (criterion of cognitive clarity of a mathematical model, criterion of adequacy of translation templates, criterion of completeness of influences of factors, criterion of proportionality factors according to the generality of concepts).

To confirm or refute the identified cause-and-effect relationships, as well as determine the degree of interaction between factors using regression analysis, statistical data from four universities located in the Rostov region were analyzed: South Russian State Polytechnic University, Rostov State Economic University and Southern Federal University. Statistical data was taken from 2008 to 2015.

Using the estimated knowledge of experts and the analysis of quantitative data, the degree of influence of one factor on another was determined on a scale from -1 to 1. As a result of the expert and regression analysis, 9 factors were identified as major, 18 as secondary, and 3 factors (V6, V7 and V8) are determined to be insignificant for the cognitive model of international activity of university.

The cognitive map of international activity of university is given in Table 1. The name of factors is given below:

- V1 – Foreign trade turnover;
- V2 – Average dollar to ruble exchange rate;
- V3 – Gross regional product;
- V4 – Gross regional product per capita;
- V5 – Number of people employed in education;
- V6 – Number of unemployed;
- V7 – Population of the region;
- V8 – Population of the region by age (from 15 to 24 years);
- V9 – Average cost of studying at universities in the region;
- V10 – Average per capita living expenses in the region;
- V11 – Number of foreign students from CIS countries;
- V12 – Number of foreign students from non-CIS countries;
- V13 – Number of bilateral and multilateral agreements with foreign organizations;
- V14 – The number of international associations and communities in which the educational organization is a member or participates;
- V15 – Number of joint international projects financed by grants (Erasmus+, Horizon, etc.);
- V16 – Position in the international ranking (QS, THE, or others);
- V17 – Amount of funding for international activities at the university (internationalization programs);
- V18 – Number of courses taught in English;
- V19 – Number of educational programs implemented in English;
- V20 – Number of programs leading to a double degree, implemented in partnership with foreign universities;
- V21 – Number of professors, teachers and researchers who have received a master's

degree, PhD, etc. in foreign educational and scientific organizations;
- V22 – Number of professors, teachers and researchers working in an educational organization abroad for at least 1 month (outgoing academic mobility);
- V23 – Number of foreign leading professors, teachers and researchers working in an educational organization for at least 1 month (incoming academic mobility);
- V24 – Number of foreign graduate students from CIS countries;
- V25 – Number of foreign graduate students from non-CIS countries;
- V26 – Number of educational programs developed and implemented in partnership with leading foreign universities;
- V27 – Number of articles prepared jointly with foreign organizations;
- V28 – Number of students who received a double diploma in an educational program implemented in partnership with leading foreign universities;
- V29 – Number of foreign students who received scholarship support for studying during a semester in an educational organization (incoming student mobility);
- V30 – Number of students who received scholarship support for studying for a semester at an educational organization abroad (outgoing student mobility).

After the cognitive map has been built, it is possible to simulate the situation. First of all, the final, long-term, stable state of the system is revealed while maintaining the current state of the concepts and connections between them. To do this, the vector of the initial state of concepts, consisting

Table 1: Cognitive map of the university's international activities.

International activity	V1	V2	V3	V4	V5	V9	…	V25	V26	V27	V28	V29	V30
V1	0,00	0,30	0,30	0,30	0,40	0,60	…	0,10	0,10	0,10	0,20	0,20	0,00
V2	-0,40	0,00	-0,30	-0,20	0,30	0,30	…	0,10	0,30	0,20	0,10	0,40	-0,40
V3	0,40	-0,40	0,00	0,70	-0,30	-0,40	…	0,30	0,30	0,40	0,40	0,50	0,40
V4	0,40	-0,40	0,30	0,00	-0,30	-0,40	…	0,20	0,30	0,30	0,30	0,30	0,40
V5	-0,30	0,20	-0,20	-0,20	0,00	0,40	…	-0,30	-0,30	-0,30	-0,30	-0,20	-0,30
V9	-0,30	0,10	-0,20	-0,20	0,60	0,00	…	-0,40	-0,30	-0,40	-0,40	-0,40	-0,30
…	…	…	…	…	…	…	…	…	…	…	…	…	…
V25	0,30	-0,10	0,20	0,20	0,30	0,20	…	0,00	0,80	0,80	0,60	0,50	0,30
V26	0,20	-0,10	0,20	0,20	0,30	0,10	…	0,80	0,00	0,80	0,60	0,60	0,20
V27	0,10	-0,10	0,20	0,20	0,30	0,20	…	0,80	0,80	0,00	0,70	0,50	0,10
V28	0,20	-0,10	0,20	0,20	0,20	0,20	…	0,60	0,60	0,70	0,00	0,30	0,20
V29	0,20	-0,10	0,20	0,20	0,30	0,20	…	0,50	0,60	0,50	0,30	0,00	0,20
V30	0,00	0,30	0,30	0,30	0,40	0,60		0,10	0,10	0,10	0,20	0,20	0,00

of ones, is multiplied by the adjacency matrix. The resulting result is converted using logistic function of the form (1) into a vector of values in the interval [0;1]:

$$C_j(x) = 1/(1 + exp(-c*x)), \qquad (1)$$

where $C_j(x)$ - degree of concept activation N_j in the moment of time x;

This positive transformation allows us to better understand and imagine the level of activation of concepts and compare the final states of concepts. This procedure continues until a stable, unchanging state of concepts is achieved. This usually requires less than 30 steps. Theoretically, the system may not come to a fixed state, but to a cycle or a chaotic attractor [20].

3.2 Scenario Modelling

As a result of calculating the first scenario, corresponding to a stable state of the system, fairly close numerical characteristics were obtained.

Table 2: Results of three scenarios.

Results – no change (scenario 1)	Results – scenario 2	Results – Scenario 3
0,99887	1,00	0,9990309
0,0273243	0,5	0,02472257
0,9991667	0,9990398	0,99924687
0,9996941	0,9996638	0,99972348
0,9999797	0,9999824	0,5
0,9997978	0,9998245	0,99975292
0,9999184	0,9999386	0,99994246
0,999910	0,999939	0,99993981
0,999455	0,999589	0,99961538
0,999399	0,999568	0,99957601
0,999955	0,999969	0,999970
0,999836	0,999876	0,999890
0,999634	0,999711	0,999714
0,999799	0,999841	0,999843
0,999853	0,999904	0,999901
0,999879	0,999921	0,999919
0,999900	0,999925	0,999922
0,999890	0,999913	0,999910
0,999950	0,999961	0,999957
0,999890	0,999913	0,999914
0,999900	0,999925	0,999922
0,999989	0,999989	0,999990
0,999985	0,999987	0,999987
0,999963	0,999966	0,999968
0,999877	0,999882	0,999894
0,999818	0,999849	0,999835
0,998873	0,998638	0,999031

Table 2 is an example of the calculation of three scenarios for the cognitive map presented in Table 1. In the second scenario, the value of concept V2 (as the most different in value) is assumed to decrease by 2 times compared to the current state.

In the third scenario, it is assumed that the values of concepts V5, V25, V26 (as those most influencing the state of the system) will be reduced by 2 times compared to the current state.

The analysis of three scenarios allows us to conclude that when making strategic decisions, the main attention must be paid to adjusting the value of the V2 indicator, as well as preventing a decrease in indicators V5, V25, V26.

It is worth noting that reducing the value of V2 is a higher priority than maintaining the values of V5, V25, V26.

4 CONCLUSIONS

This paper describes an aproach for assessing the international activities of a university using fuzzy cognitive maps. As a result of the study, a cognitive model for managing the effectiveness of international activities of universities was built. To build this model, we took into account the quality criteria for formalizing experts' views on the international activities of universities, as well as statistical data. The main purpose of the model is to help the expert to develop the right decision. This model displays and organizes information about international activities of university, taken in account a large number of influencing factors, on different levels. The cognitive model not only allows to systematize and "clarify" the expert's knowledge, but also helps to identify the most advantageous points applications of control actions of the subject of management.

It is important here to foresee what the consequences will be have certain management strategies. To develop such forecasts, a scenario modelling is used within the framework of cognitive analysis. After the cognitive map was built, several scenarios simulated the situation using the FCMapper tool. Three scenarios for the development of the university's international activities were reviewed and analyzed based on a fuzzy contive map.

Future work is connected with development of the decision support system based on the developed congintive map, which will allow experts to set up the most important factors and to simulate the different scenarios.

REFERENCES

[1] T. Schubert, E. Baier, and C. Rammer, "A behavioural approach. Journal of International Business Studies," vol. 49(1), pp. 70-95, January 2018, doi: 10.1057/s41267-017-0101-0.

[2] H. de Wit and P. G. Altbach, "Internationalization in higher education: global trends and recommendations for its future," Policy Reviews in Higher Education, vol. 5(1), pp. 28-46, September 2020, doi: 10.1080/23322969.2020.1820898.

[3] A. Kolomiets, D. Grinchenkov, and K. Vodenko, "PEST- and SWOT-analysis of university internationalization factors," Journal of Physics: Conference Series, vol. 1415, p. 012003, October 2019, doi: 10.1088/1742-6596/1415/1/012003.

[4] G. Sivertsen, "Patterns of internationalization and criteria for research assessment in the social sciences and humanities," Scientometrics, vol. 107, pp. 357-368, May 2016, doi: 10.1007/s11192-016-1845-1.

[5] P. B. Keenan and P. Jankowski, "Spatial Decision Support Systems: Three decades on," Decision Support Systems, vol. 116, pp. 64-76, January 2019, doi: 10.1016/j.dss.2018.10.010.

[6] P. Morone, P. M. Falcone, and A. Lopolito, "How to promote a new and sustainable food consumption model: A fuzzy cognitive map study," Journal of Cleaner Production, vol. 208, pp. 563-574, Jan. 2019, [Online]. Available: https://doi.org/10.1016/j.jclepro.2018.10.075.

[7] E. Bakhtavar, M. Valipour, S. Yousefi, R. Sadiq, and K. Hewage, "Fuzzy cognitive maps in systems risk analysis: a comprehensive review," Complex & Intelligent Systems, vol. 10.1007/s40747-02000228-2, Nov. 2020, [Online]. Available: https://doi.org/10.1007/s40747-020-00228-2.

[8] G. Felix, G. Nápoles, R. Falcon, W. Froelich, K. Vanhoof, and R. Bello, "A review on methods and software for fuzzy cognitive maps," Artificial Intelligence Review, vol. 52, no. 3, pp. 1707-1737, Aug. 2017, [Online]. Available: https://doi.org/10.1007/s10462-017-9575-1.

[9] P. Szwed, "Classification and feature transformation with Fuzzy Cognitive Maps," Applied Soft Computing, vol. 105, p. 107271, Jul. 2021, [Online]. Available: https://doi.org/10.1016/j.asoc.2021.107271.

[10] István Harmati, M. F. Hatwágner, and L. T. Kóczy, "Global stability of fuzzy cognitive maps," Neural Computing and Applications, vol. 35, no. 10, pp. 7283-7295, Dec. 2021, [Online]. Available: https://doi.org/10.1007/s00521-021-06742-9.

[11] G. Sivertsen, "Patterns of internationalization and criteria for research assessment in the social sciences and humanities," Scientometrics, vol. 107, no. 2, pp. 357-368, Feb. 2016, [Online]. Available: https://doi.org/10.1007/s11192-016-1845-1.

[12] Jose A. Fernandez Leon and Gerardo G. Acosta. J. A. Fernández-León and G. Acosta, "Uncovering the Secrets of the Concept of Place in Cognitive Maps Aided by Artificial Intelligence," Cognitive Computation, Oct. 2022, [Online]. Available: https://doi.org/10.1007/s12559-022-10064-w.

[13] N. D. Pankratova, G. Gorelova, and Vladimir Pankratov, "Study of the Plot Suitability for Underground Construction: Cognitive Modelling," Advances in intelligent systems and computing, pp. 246-264, Jul. 2020, [Online]. Available: https://doi.org/10.1007/978-3-030-54215-3_16.

[14] P. Yang, "China in the global field of international student mobility: an analysis of economic, human and symbolic capitals," A Journal of Comparative and International Education, pp. 1-19, May 2020, [Online]. Available: https://doi.org/10.1080/03057925.2020.1764334.

[15] T. Schubert, E. Baier, and C. Rammer, "Firm capabilities, technological dynamism and the internationalisation of innovation: A behavioural approach," Journal of International Business Studies, vol. 49, no. 1, pp. 70-95, Sep. 2017, [Online]. Available: https://doi.org/10.1057/s41267-017-0101-0.

[16] S. Soliman, J. Anchor, and D. Taylor, "The international strategies of universities: deliberate or emergent?," Studies in Higher Education, vol. 44, no. 8, pp. 1-12, Mar. 2018, [Online]. Available: https://doi.org/10.1080/03075079.2018.1445985.

[17] J. Aguilar, "Multilayer Cognitive Maps in the Resolution of Problems using the FCM Designer Tool," Applied Artificial Intelligence, vol. 30, no. 7, pp. 720-743, Aug. 2016, [Online]. Available: https://doi.org/10.1080/08839514.2016.1214422.

[18] Holman Bolívar Barón, Rubén González Crespo, Jordán Pascual Espada, and Óscar Sanjuán Martínez, "Assessment of learning in environments interactive through fuzzy cognitive maps," Soft Computing, vol. 19, no. 4, pp. 1037-1050, May 2014, [Online]. Available: https://doi.org/10.1007/s00500-014-1313-x.

[19] Z. Kahvandi, E. Saghatforoush, A. ZareRavasan, and C. Preece, "Integrated Project Delivery Implementation Challenges in the Construction Industry," Civil Engineering Journal, vol. 5, no. 8, pp. 1672-1683, Aug. 2019, [Online]. Available: https://doi.org/10.28991/cej-2019-03091362.

[20] T. Mansouri, A. ZareRavasan, and A. Ashrafi, "A Learning Fuzzy Cognitive Map (LFCM) Approach to Predict Student Performance," Journal of Information Technology Education: Research, vol. 20, pp. 221-243, 2021, [Online]. Available: https://doi.org/10.28945/4760.

Artificial Intelligence in Forecasting Demographic Processes

Stepan Mezhov and Maxim Krayushkin'

Altai State University, Lenin Str. 61, 656049 Barnaul, Russia
megoff@mail.ru, kramaks-97@mail.ru

Keywords: Forecasting, Artificial Intelligence Models, Artificial Neural Networks, Model Error, Population Size, One-Year Old.

Abstract: At the moment, there are no universal tools for forecasting indicators of socio-economic development in general and demographic in particular. However, the amount of budget allocations that are directed to solving personnel issues, creating social facilities and implementing other activities significant for economic, social and infrastructural development, development vectors depend on the forecast of demographic indicators by one-year-olds. In this article, a meaningful analysis is carried out an analysis of the researchers' work, based on the results of which it is determined that artificial intelligence models, in particular the most adaptive neural networks, are almost not used in predicting demographic indicators, forecasts based on models of artificial neural networks for all ages or by age groups, have almost no significance from the point of view of managing socio-economic development. The result of the research is a neural network methodological approach, methodology and tools for forecasting demographic processes for one-year-olds. The developed methodological approach to forecasting and tools are universal in the field of forecasting socio-economic indicators. In addition, the results described in the article can be used in other research in the field of forecasting, planning of any indicators.

1 INTRODUCTION

The amount of budget allocations that are directed to solving personnel issues, creating social facilities and implementing other activities significant for economic, social and infrastructural development, development vectors depend on the forecast of demographic indicators by one-year-olds.

At the moment, there are no universal tools for forecasting indicators of socio-economic development in general and demographic in particular.

A meaningful analysis of the researchers' work has shown that the most common methodological approaches to demographic forecasting are:

- Time series modeling.
- Building regression models.

This thesis is determined based on a constructive analysis of the work of more than 50 researchers.

The problems of constructing demographic forecasts as an essential element of the socio-economic development planning system have been sufficiently worked out in terms of tools and modeling. Conceptual approaches to the construction of demographic forecasts were developed by such

researchers as Graunt J. [1], Malthus T. [2], Pearl R. [3], Reed L.[3], Welton P. [4], Vishnevsky A. G. [5] and others. A significant contribution to the development of tools for building computer models of demographic processes has been made: Meadows D. [6], Simon G.A. [7], Forrester J. [8, 9], Tsybatov, V.A. [10] and others. In their work, the above-mentioned researchers also focused on the possibility of predicting demographic processes using regression models and time series models. However, the main disadvantage of these models is their weak adaptability to changes in the socio-economic situation, which negatively affects the accuracy.

A constructive analysis of studies, including the latest ones from 2023-2024, in which models of artificial neural networks were used to build a forecast of demographic indicators, showed that forecasts using neural networks are built either for all ages or for individual age groups. In turn, such neural network forecasts have almost no value in managing socio-economic development [14, 15, 24, 25]. Only demographic forecasts based on one-year ages have significant value in this process.

In turn, the scientific problem can be determined as follows: the use of subjective, non-adaptive methodological approaches to forecasting indicators of socio-economic development leads to a high forecast error and, as a result, ineffective planning of development. This has a negative impact on economic development and the standard of living of the population.

Therefore, the purpose of the study is to develop theoretical and methodological provisions and tools based on them to improve the planning system for socio-economic development in the context of forecasting demographic processes by one-year-olds based on artificial intelligence.

2 MATERIALS AND METHODS

When developing tools for the formation of an information base, autocorrelation and cluster analysis are used, and artificial neural networks are used for modeling.

When building neural networks in the framework of the study, neural network training was chosen as a paradigm: with a teacher; training rule – error correction; architecture – a multilayer neural network; learning algorithm (optimizer) – BFGS.

When using training models of artificial neural networks with a teacher, when correcting an error, the parameters of the neural network are determined in such a way that the actual data are as close as possible to the values obtained from the model, that is, so that the model error is minimal (it is on the test sample). To solve the forecasting problem, it is important that the accuracy of the model be as high as possible, because the effectiveness of socio-economic development management depends on the accuracy of the forecast. Therefore, neural network training with a teacher and the error correction training rule were chosen as the paradigm.

To solve the forecasting problem, only multi-layered neural networks are needed. After all, when taking into account a large number of parameters, the probability of increasing the accuracy of modeling increases.

The BFGS algorithm was chosen as the optimizer because it is one of the most effective ways of optimization. Numerous publications on the effectiveness of using this algorithm to solve complex problems not only in forecasting, the use of this algorithm in modern neural network simulation systems prove the expediency of its application in the framework of the study [12].

The structure of an artificial neuron can be represented as follows (Figure 1).
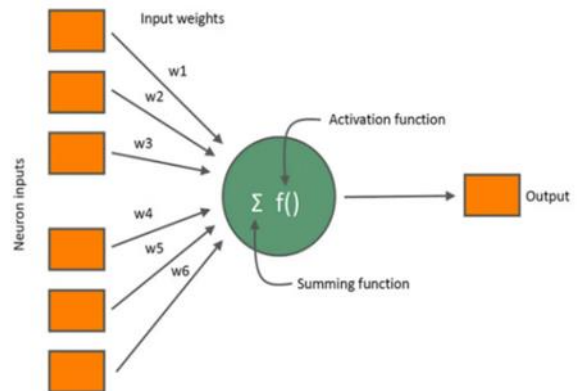
Figure 1: The structure of an artificial neuron.

The scheme of operation of the BFGS algorithm is shown in Figure 2.
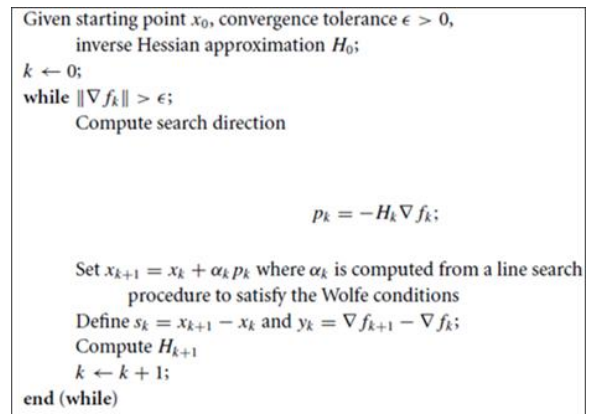
Figure 2: Diagram of the BFGS algorithm.

Meaningfully, the algorithm works as follows: at the first iteration, the initial weighting coefficients of the connections in the neural network are determined. In the following iterations, the weight coefficients change until the error value on the test sample becomes minimal [12].

When forming the information base, building neural network models, and forecasting based on them, the STATISTICA 13 version software package.

3 RESULTS

Modern neural network models are much more accurate than time series models and regression models, they have a wider degree of applicability and they are more adaptive. Using specialized software

products, they can be built by a specialist without having in-depth knowledge of neural network modeling [13].

Accordingly, in order to increase the accuracy of forecasting demographic processes by one-year-olds, a methodological approach based on the construction of artificial intelligence models, namely artificial neural networks, is proposed.

The characteristics of the neural network approach to demographic forecasting by one-year age proposed by the authors of the article are:

- Building models that conceptually work in a similar way as an expert who develops a forecast.
- Using only adaptive algorithms for building neural network models.

The authors of the article have developed a methodology for constructing a demographic forecast, it is presented in Figure 3.
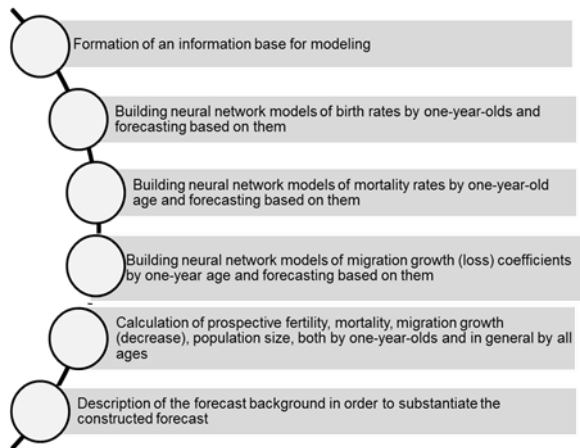
Figure 3: Methodology for constructing a demographic forecast.

Let's consider the developed methodology (Figure 3) in detail.

At the first stage, retrospective information is collected about:

- Age - related birth rates by one- year- olds.
- Age-related mortality rates by one-year age, both among men and women.
- The coefficients of migration growth (decrease) by one-year age, both among men and women.
- The population of each age as of January 1, both among men and women.
- Factors that significantly affect the demographic dynamics.

To collect information, you can use open [11].

Further, for the purpose of modeling, it is necessary to process the data in the appropriate way, that is, to divide it into representative groups. To do this, it is necessary to conduct a cluster analysis. First, it is necessary to determine the optimal number of clusters using the tree clustering method [16].

After that, it is necessary to conduct a cluster analysis using the method, for example, "k-means".

After the data is properly collected and processed, you can proceed to the next steps related to building models and forecasting directly. At subsequent stages, models of fertility rates, mortality, migration growth (loss) by one-year-olds and forecasts based on them are being built. It is already advisable to use artificial neural networks here.

In general, the procedure for building an artificial neural network can be represented as the following steps:

- Choosing a paradigm.
- Selection of the training rule.
- The choice of neural network architecture.
- Choosing the algorithm for building a neural network.
- Selection of the criterion for stopping learning.

Based on a meaningful analysis of the work of researchers in the field of neural network modeling, it was revealed that to solve the problem of forecasting, including demographic processes by one–year–olds, when building neural networks in the framework of the study, neural network training was chosen as a paradigm: with a teacher; training rule – error correction; architecture – a multilayer neural network; learning algorithm (optimizer) – BFGS [12, 17-22].

At the last stage, the stop criterion is selected. It is usually understood as achieving the minimum error value in the test sample.

Based on the results of a comparative analysis of packages, systems and technologies in the field of network neurosimulation in forecasting tasks (STATISTICA, Matlab, Loginom and others), it follows that the most suitable software package is STATISTICA due to its advantages (it is free software, it is possible to analyze and visualize big data, there are packages of statistical analysis, access rights management, predictive analytics, it is possible to download the program code of the model in different programming languages) and ease of implementation for a user who does not have high qualifications in the field of neural network modeling and forecasting [23].

At the penultimate stage of the methodology, the following is carried out:

- Calculation of the number of births by age of one year. The calculation of the number of births for each age is carried out by multiplying the average annual value of the population of the corresponding age in the prospectus period by the corresponding forecast value of the fertility rate.
- Calculation of the number of deaths by age of one year. The calculation of the number of deaths for each age is carried out by multiplying the average annual value of the population of the corresponding age in the prospectus period by the corresponding forecast value of the mortality rate.
- Calculation of migration growth (loss) by one-year age. The calculation of the number of migration growth (loss) for each age is carried out by multiplying the average annual value of the population of the corresponding age in the prospectus period by the corresponding forecast value of the migration growth coefficient (loss).
- Calculation of the population for each age. The calculation of the population of the corresponding age is carried out as follows: from the value of the population of the corresponding age of the previous year, the number of deaths and the number of retirees in the current year are subtracted. However, if there is a migration increase rather than a decrease, then the number of arrivals in the current year is added to the population of the corresponding age of the previous year, minus the number of deaths. The number of newborns is equal to the projected birth rate.

At the final stage of the methodology, a description of the forecast background is carried out in order to substantiate the constructed forecast within the framework of planning socio-economic development.

Using the estimated values of factors that significantly affect demographic processes, the justification of the forecast calculations is carried out. The description of the forecast background is carried out both for the forecast of fertility, and for mortality and migration processes. The predicted value of the demographic variable and the factors influencing it is taken. Then the description of the change in the variable from the corresponding changes in the factors is carried out.

In addition, a description of the forecast background is carried out for the total number and population size by one-year age, depending on changes in fertility, mortality and migration.

Then, based on the methodology developed within the framework of the study (Figure 3), models of artificial neural networks and a forecast based on them were built.

At the first stage, an information base was formed. For this purpose, statistical data were collected on the rates of fertility, mortality, and migration growth (loss) by one-year age from 2012 to 2022. Then an autocorrelation analysis was carried out, the results of which revealed that it is advisable to build models using data from 2017 to 2022. After that, a cluster analysis was performed using tree clustering and "k-means" methods to determine the data arrays for further modeling.

At the second stage, complexes of predictive models of the birth rate by one-year age were built for the corresponding clusters.

When building neural networks in the framework of the study, neural network training was chosen as a paradigm: with a teacher; training rule – error correction; architecture – a multilayer neural network; learning algorithm (optimizer) – BFGS.

When forming the information base, building neural network models, and forecasting based on them, the STATISTICA 13 version software package was used due to its advantages mentioned earlier.

The absolute verification of the constructed models is shown in Figures 4, 5.

Figures 4, 5 show the actual data of the fertility rate by one-year age and obtained using the constructed neural network model. The value of the average absolute error in the percentage of models (MAPE) on training data is less than 2%, on test data – 2-2.5%, retrogression – 3-3.5%. Neural networks have 2 hidden layers of 7 neurons in each of them.

The graph of the neural network of the fertility rate is shown in Figure 6.

The constructed neural network models have 7 neurons in 2 hidden layers. The activation function of neurons is a sigmoid.

The weight coefficients are shown in Table 1.

Similarly, the construction and analysis of the accuracy of models for 2, 3, 4 clusters were carried out.

At the third and fourth stages, neural network models were built in a similar way for the corresponding clusters to predict the mortality rate for men and women, as well as the coefficient of migration growth (loss) for men and women.

All built models, including those for verification based on data from 2022, are quite accurate, according to some models the accuracy reaches 99,5 percent.

Table 1: Weight coefficients.

The weighting coefficients of the connections of the inputs with the neurons of the first hidden layer	The weighting coefficients of the connections between the neurons of the first layer and the neurons of the second layer	The weighting coefficients of the connections of neurons of the second hidden layer with the output
-0,126	0,66	0,132
-0,232	-0,306	-0,576
-0,148	0,372	0,101
-0,47	-0,189	0,511
-0,007	0,651	-0,54
-0,014	0,64	0,415
0,06	0,125	0,136
0,056	-0,013	-
0,096	0,6	-
0,12	0,74	-
0,2	-0,128	-
0,45	0,504	-
-0,05	-0,144	-
-0,165	-0,03	-
0,4	0,008	-
-0,13	-0,08	-
-0,546	-0,042	-
0,198	0,155	-
0,013	-0,044	-
0,232	-0,26	-
-0,104	-0,035	-
-	0,33	-
-	-0,006	-
-	-0,002	-
-	-0,09	-
-	-0,029	-
-	0,132	-
-	-0,06	-
-	0,007	-
-	0,008	-
-	0,008	-
-	0,008	-
-	0,03	-
-	0,052	-
-	0,04	-
-	0,07	-
-	-0,06	-
-	0,01	-
-	-0,015	-
-	0,011	-
-	-0,042	-
-	0,01	-
-	-0,222	-
-	0,033	-
-	-0,018	-
-	-0,168	-
-	0,06	-
-	-0,028	-
-	0,032	-

Figure 4: Absolute verification of the neural network for the first cluster (retrospective forecast for 2020).
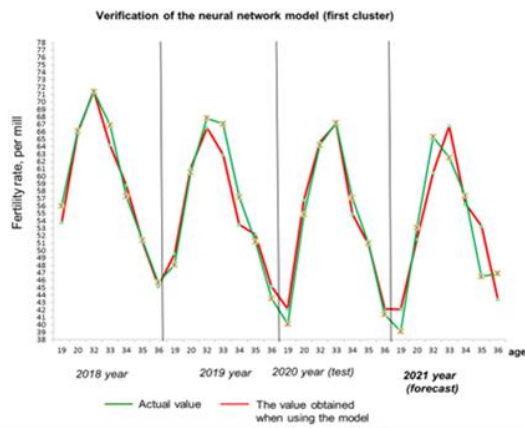
Figure 5: Absolute verification of the neural network for the first cluster (retrospective forecast for 2021).
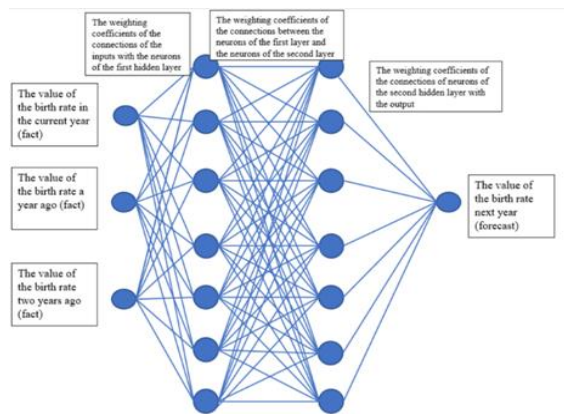
Figure 6: The graph of the neural network of the fertility rate.

At the fifth stage, a population forecast was built and its verification was carried out.

The population forecast is shown in Figures 7, 8.
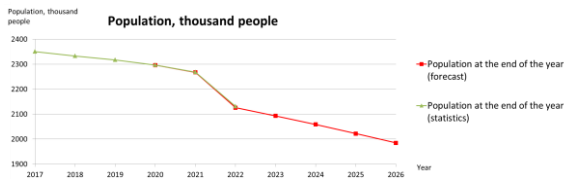
Figure 7: Population forecast.
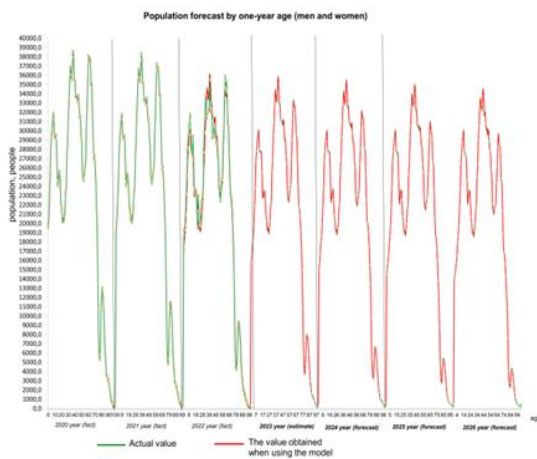
Figure 8: Population forecast (by one-year age).

The use of a neural network approach to forecasting demographic processes by one-year-olds has led to an increase in its accuracy, which will entail increased opportunities to solve socio-economic problems and the need to adjust development plans.

4 CONCLUSIONS

The effectiveness of socio-economic development planning depends on the forecast of demographic processes by one-year age.

During the analysis of the researchers' work, it was revealed that the most adaptive models of artificial neural networks are practically not used in predicting demographic processes.

Therefore, a methodological approach, methodology and tools have been developed for predicting demographic processes by one-year-olds in based on artificial neural network models.

Based on the results of the analysis of the accuracy of the constructed models, it was found that their use is advisable in the process of planning of socio-economic development.

REFERENCES

[1] J. Graunt, "Natural and political observations listed in the attached table of contents and made on the basis of bulletins on mortality. In relation to governance, religion, trade, growth, air, diseases and other changes of the named city. The work of John Graunt, citizen of London," London, 1662.

[2] T.R. Malthus, "Experience of the law on population," L. : Petrokom, 1793, 292 p.

[3] A. G. Volkov, "Pearl Raymond Demography," M. : Soviet Encyclopedia, 1975, p. 556. (The Great Soviet Encyclopedia, ch. ed. A.M. Prokhorov, vol. 19. pp. 1969-1978).

[4] P.K. Whelpton, "An empirical method of calculating future population," Journal of the American Statistical Association, 1936, no. 31, pp. 457-473.

[5] A.G. Vishnevsky, "Prospects for the development of Russia: the role of the demographic factor," M.: Russian National Library, 2003, 51 p.

[6] D. Meadows, "Limits of growth, 30 years later," translated from English, M.: Akademkniga, 2007, 342 p.

[7] H.A. Simon, "Rationality as a process and product of thinking," translated from English by K. B. Kozlova and M. A. Blanco, [Online]. Available: https://www2.econ.iastate.edu/tesfatsi/RationalityAsProcessAndProductOfThought.ElyLecture1978.Hsimon.pdf.

[8] J. Forrester, "World dynamics," translated from English, M.: AST; St. Petersburg: Terra Fantastica, 2003, 379 p.

[9] J. Forrester, "Dynamics of city development," translated from English, M.: Progress, 1974, 281 p.

[10] V.A. Tsybatov, "Macroeconomic modeling, forecasting and planning of regional development: a textbook," Samara: Publishing House of Samara State University, 2016, 316 p.

[11] Federal State Statistics Service: Official website, [Online]. Available: https://rosstat.gov.ru/folder/12781, [Accessed: 12 Jan 2024].

[12] The BFGS method or one of the most effective optimization methods, [Online]. Available: https://habr.com/ru/post/333356, [Accessed: 04 Dec 2023].

[13] S. Pattanayak, "Pro Deep Learning with TensorFlow: A Mathematical Approach to Advanced Artificial Intelligence in Python," Apress, 2019, 480 p.

[14] S. Lutfiani, M. Sri, and R. Hendrisman, "Forecasting the mortality rates of Indonesian population by using neural network," International Conference on Mathematics: Pure, Applied and Computation, IOP Conf. Series: Journal of Physics: Conf. Series, pp. 974-012030, 2018.

[15] N. Bheemanna and M.N. Megeri, "Prediction of India's demographic and economic variables using the neural network auto-regression model," International Journal of Statistics and Applied Mathematics, vol. 11, pp. 44-55.

[16] A.A. Khafalyan, "Industrial statistics. Quality control, process analysis, experiment planning in the STATISTICA package: textbook," M.: Rusains Publishing House, 2019, 384 p.

[17] T. Theophili, "Deep learning for search engines: textbook," 2020, 318 p. .

[18] O. Geron, "Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems: textbook," Jupiter 2020, 688 p. .

[19] S. Haikin, "Neural Networks: textbook," 2019, 1408 p.

[20] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach: textbook," Pearson, 2019, 1104 p.

[21] D. Practical, "Artificial Intelligence with Python," Williams, 2019, 448 p.

[22] M. Bernard, "Artificial Intelligence in Practice," Wiley, 2019, 605 p.

[23] Statistica software package, [Online]. Available: https://statsoft.ru, [Accessed: 04 Dec 2023].

[24] M. F. H. Mohammad, "Comparison Markov Chain and Neural Network Models for forecasting Population growth data in Iraq," University of Kirkuk Journal For Administrative and Economic Science, 13(4), pp. 1-14, 2023.

[25] W. Rajesh, "Neural Network Based Mortality Prediction in Covid-19 Dataset," University of Kirkuk International Conference on Artificial Intelligence of Things, 2024.

Anticrisis Marketing of IT Sector Companies in the Conditions of Economic Activity`s Informatization and Intellectualization

Olha Polous, Inna Mykhalchenko, Hanna Radchenko and Olena Lysenko

National Aviation University, Liubomyra Huzara Avenue 1, 03058 Kyiv, Ukraine

polousolgavictorovna@gmail.com, inna.mykhalchenko@npp.nau.edu.ua, hanna.radchenko@npp.nau.edu.ua, olena.lysenko@npp.nau.edu.ua

Keywords: IT Sector, Company, Crisis, Anticrisis Marketing, Marketing Tools, Informatization, Intellectualization.

Abstract: The research actuality of the study is due to the fact that in fast-paced world of IT, the ability to adapt, innovate, and proactively address challenges is key to sustaining success and anticrisis marketing should be seen as an ongoing, dynamic process that evolves alongside the industry it serves. Aim of the paper is to justify the necessity of forming a model of anticrisis marketing of IT companies on the example of the adaptation of Ukrainian IT companies to the crisis conditions of their activities. In order to achieve the research goal, general scientific methods of analysis and synthesis, logical and situational analysis and specific methods inherent in management sciences (subjective-objective approach, economic and statistical analysis, method of ranking, grouping, comparison, classification) were used. The practical and theoretical relevance of this study is that the proposed and substantiated ways of IT companies' anticrisis marketing model forming can be of practical interest to managers and leaders of these companies when avoiding or mitigating the consequences of crisis situations. Research materials can be used on practice when solving the problems of increasing the effectiveness of marketing activities of both Ukrainian and other IT companies during the crisis period in order to increase their competitiveness on the national and global markets.

1 INTRODUCTION

The era that we are experiencing at the beginning of the 21st century undoubtedly has features of progressiveness and intellectualization, will play a key role in the course of the history of human development in all its manifestations, including as a time of active development of companies that provide IT services and are capable of optimizing the functioning of business process of companies in other sectors of the global economy, allowing them to be more effective in competing for resources and consumers. As a rule, modern researchers attribute the characteristics and features described above to a new phase of human life, known as the "information society" [1, 2], which creates a new type of economy – the information economy [3, 4]. In turn, the effective functioning of the information economy of a separate state is impossible without the creation and optimal business activities of companies that provide information and telecommunications services, and they, in turn, are rarely limited to serving only domestic consumers and often operate in several countries, which makes them vulnerable to global negative economic, political, social, legal and technological factors. Anticrisis marketing, as an important component of the activities of a modern IT company, can help solve a number of problem situations that are caused by different groups of negative factors, but its tools must be constantly reviewed and improved, since the IT sector is the area of business that is developing most dynamically.

2 PREREQUISITES FOR THE NEED TO IMPROVE ANTICRISIS MARKETING OF UKRAINIAN IT COMPANIES

The active use of information, knowledge, and various communication technologies, as well as information in electronic form, significantly complicates the activities of companies that do not use modern technologies for processing them, making them less competitive. Currently, the rapid development of information technology, computer engineering, and the Internet makes it possible to implement a variety of methods for transmitting and processing a huge amount of information, and it is no

longer possible to refuse these advantages. This is what determines the modern popularity of IT companies that develop both universal products and unique solutions based on customer requests.

On Figure 1 we have systematized the factors that influence the activities of modern Ukrainian IT companies.

```
┌──────────────────────────────────────────┐        ┌──────────┐
│ Economic and political factors:          │        │          │
│ - war (unstable political and economic   │   ⇒    │          │
│ situation for doing business);           │        │          │
│ - globalization;                         │        │          │
│ - innovations;                           │        │          │
│ - competition;                           │        │          │
│ - the need to diversify products         │        │          │
└──────────────────────────────────────────┘        │          │
┌──────────────────────────────────────────┐        │   IT     │
│ Social factors:                          │        │ company  │
│ - education;                             │        │          │
│ - emergence of new professions;          │   ⇒    │          │
│ - recognition (availability of information│       │          │
│ about products);                         │        │          │
│ - flexibility of approaches to consumers;│        │          │
│ - population migration                   │        │          │
└──────────────────────────────────────────┘        │          │
┌──────────────────────────────────────────┐        │          │
│ Technological factors:                   │        │          │
│ - Artificial Intelligence;               │        │          │
│ - improvement of Internet services;      │        │          │
│ - convergence of computer equipment;     │   ⇒    │          │
│ - availability of a large number of software│      │          │
│ products to support activities;          │        │          │
│ - techniques of collecting, storing,     │        │          │
│ processing and transmitting information  │        │          │
└──────────────────────────────────────────┘        └──────────┘
```

Figure 1: Factors that influence the activities of modern Ukrainian IT companies.

Along with the above factors, the activities of Ukrainian IT companies are also influenced by approaches to determining various directions for the development of the information society. In our opinion, in the context of the Ukrainian environment for doing this type of business, the following directions should be considered:

- pessimistic view (techno-totalitarian society). All applications and databases will be developed taking into account the fact that people will be increasingly tied to the computer, and the main goal will be to gain control over them. In such a society, human functioning will be deprived of privacy and freedom, which can lead to physical and psychological isolation of workers, deepening inequality in access to information, increased crime rates, etc. [5];
- optimistic view (network society). An optimistic vision, endorsed by authors such as Gilder and Castells, emphasizes the positive role of information and communication technologies in economic development and the improvement of human capabilities. This point of view mainly highlights the positive impact of social networks

and modern information technologies, viewing them as a means to build a better civilization and establish better relationships between companies, customers and other economic actors [6];
- evolutionary view (logical continuation of the development of society). Proponents of this trend suggest that information and communication technologies do not have a dominant influence over humans, and changes will be exclusively evolutionary in nature. In other words, new technologies will effectively support existing social, economic and cultural systems [7].

The information technology sector became one of the few economic spheres that demonstrated growth during 2022 and contributed 16% more funds to the state budget of Ukraine compared to 2021 [9]. However, along with such positive statistics, a number of alarming signals were also recorded, which once again forced the management of Ukrainian IT companies to think about the need to develop and improve anticrisis measures.

According to the study "Do IT like Ukraine" by the IT Ukraine Association [8], all Ukrainian IT companies in one way or another faced the challenges associated with a full-scale war. Only 34.3% of Ukrainian IT companies were able to fully adapt to the new conditions, which took measures such as equipping office premises with generators in case of power outages, moving key systems to the "cloud", paying for the relocation of leading specialists to another region of the country or the world, or providing funding for secure co-working services, as well as diversifying the network of providers to ensure uninterrupted product creation [8].

The executive director of the IT Ukraine Association Kostyantyn Vasyuk claims that during the crisis, which was caused by the start of a full-scale war, two key groups of customers of Ukrainian IT companies can be distinguished:
- those who are more vulnerable to the challenges associated with the need to move key parts of the project outside of Ukraine and are reluctant to make contact;
- those who consider concluding new agreements with Ukrainian companies as some kind of social mission and continue to cooperate with Ukrainian IT teams [8].

Therefore, it can be said that large and medium-sized companies that operate with large amounts of data and serve a significant number of customers show a high degree of caution regarding cyber

security issues and are extremely wary of the need to transfer data outside the state, while small businesses are somewhat more likely to accept risks in exchange for a quality and efficient IT product. At the same time, those companies that already cooperated with Ukrainian IT companies as customers (both Ukrainian and foreign) during the war and did not feel the negative impact of political and economic factors are mostly interested in continuing cooperation in the future. However, there is also a category of customers who, along with the above-mentioned crisis factors, carefully analyze the dependence of the Ukrainian IT sector on the state of the global IT market and the recession of the world economy, which affects their final decision on ordering products and services.

It is quite difficult to assess and measure the number of IT companies operating in Ukraine, because different sources report different numbers depending on their access to analytical data. We are invited to consider the dynamics of the number of Ukrainian IT companies based on official data provided on the website of the State Statistics Service of Ukraine (Figure 2) [9].
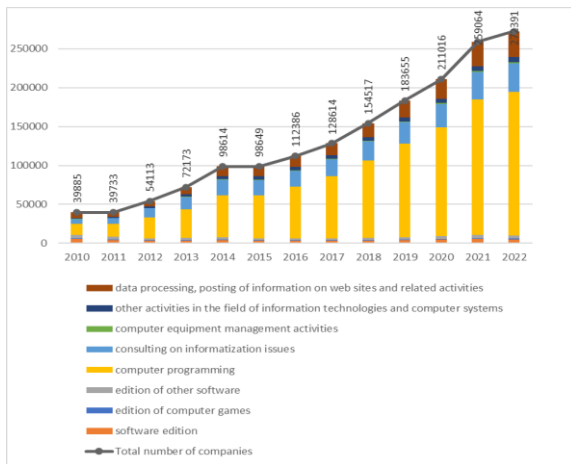
Figure 2: Dynamics of the Ukrainian IT companies' number, units, 2010-2022 [9].

At the end of 2010, there were 39,885 IT companies in Ukraine, while at the end of 2022, their number increased to 272,391, that is, almost seven times, while the growth of the number of companies did not stop even in war conditions. The largest number of Ukrainian IT companies is engaged in computer programming, which can also be seen from the given statistical information. It is worth noting that when compiling statistical reports, liquidated companies are also taken into account, so it is possible to determine the real number of operating companies only approximately. Along with

this, it is characteristic of the Ukrainian IT sector that one company can have several legal entities, which also somewhat affects the official statistics, which reflect exactly the number of legal entities in this sector of the economy.

The majority of top Ukrainian IT companies exhibit a net profit margin below 15% (Figure 3). However, two local market product companies stand out with unusually high margins: Fintech Band, LLC (60%) and UAPROM, LLC (18%). Similarly, both payment processing firms, Ukrainian Processing Center, PJSC (28%) and "Visa Ukraine" (18%), demonstrate notable profit margins [10].
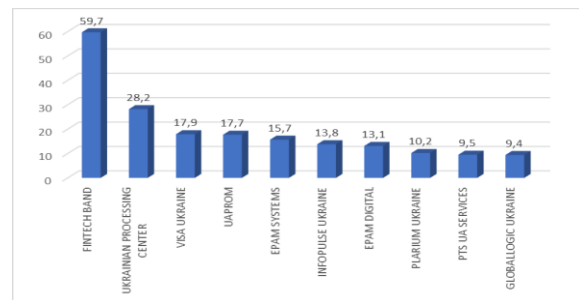
Figure 3: Net profit margin of top Ukrainian IT companies, %, 2022 [10].

In 2021, the highest growth in the volume of IT exports over the last 10 years (+38%) was observed, but already in 2022, the smallest growth for this period was noted - only +5.9% per year. Despite all the crisis phenomena in the country's economy and at the global level, the volume of IT exports in Ukraine still continues to grow and in 2022 it reached the mark of 7.3 billion US dollars, which is explained by the presence of a significant demand for this type of service, which is not always can satisfy only national companies (Figure 4) [9].
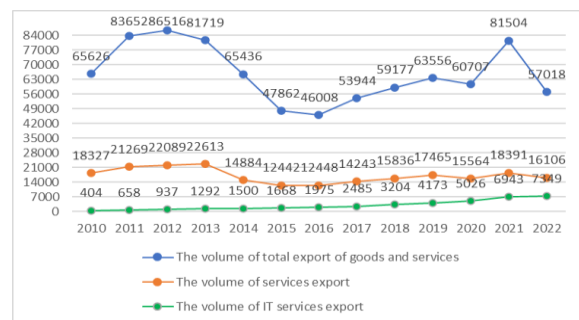
Figure 4: Comparison of the dynamics of total export of goods and services, export of services and export of IT services of Ukraine, million dollars, 2010-2022 [9].

The share of IT in the export of services in 2022 reached 45.6%, and in the total export of Ukraine - 12.9%. It can also be seen that it was constantly growing for the years 2010-2022.

Despite the generally positive trends in the IT services market of Ukraine and the rather successful anticrisis management of Ukrainian IT companies, there are still challenges that need to be overcome. The main one is related to a possible further reduction in the export of Ukrainian IT services and the transition of key customers to the use of the services of Asian, American and other European IT companies. Against the background of some decline in interest in digital and information services after the COVID-19 pandemic and the complete exit from the lockdown, this may have a noticeable impact on the profits of Ukrainian IT companies. In the context of the above, a more detailed consideration of anticrisis marketing of IT companies and ways of its adaptation to the changing needs of the national and global markets is being updated once again.

3 SCIENTIFIC AND PRACTICAL SUBSTANTIATION OF THE IT SECTOR COMPANIES ANTICRISIS MARKETING MODEL FORMATION

There is some difference in understanding the essence of activity of IT-companies according to the product or service they produce/provide [11]. IT-product companies are oriented on end users of their product, service IT-companies or outsource companies are focusing on clients that need some services which help further production or providing of product/services. IT-products include apps, social networks, online-games, antiviruses, online games, websites, operation systems.

80% of the top ten companies by capitalization in the world are IT-product companies. Among them are Apple, Microsoft, Amazon, Alphabet and others. The products of Apple Inc. include Mac, iPhone, iPad, iOS operating system, Apple Watch, Apple TV, iCloud, Apple Pay, iTunes Store etc. Microsoft Corp. products include Xbox, Microsoft 365 (Word, Excel, PowerPoint, Outlook on Microsoft Windows, OneDrive, Microsoft Teams), Windows operating system. Samsung Electronics products, like one most famous IT-product companies, include Smartphones, Tablets, Smart watches, Televisions, Audio and video players, Monitors, SSD etc. As of 2023, these corporations are among the world's largest companies by turnover.

IT-product companies operate in different niches, such as FinTech, HealthTech, EdTech, LegalTech, AITech, SportsTech (Table 1).

According to research of Genesis, Ukrainian co-founding IT company, the most common features of IT-product companies in Ukraine include:
- the most valuable part of IT-product company is the team and expertise;
- earn on the final (intellectual) product of Ukrainian specialists, instead of selling hours of their work mainly abroad;
- they have strong brand and help build the brand of Ukraine;
- they have greater influence on the development of innovations - companies work on innovative solutions and innovative technologies for the development of their own product;
- there is greater need for employees, so they create more jobs;
- they have less linear and predictable development;
- in order to grow, product companies need not only to hire and resell employees, but also to build an interesting product for the market;
- they have higher revenue per 1 employee;
- demand for people with higher professional level (top talent);
- the team in IT-product company is the alive organism that strives for self-organization;
- the team members have expertise not only in their field;
- responsibility not only to business goals but also to other team members.

Table 1: Niches and examples of IT-product companies.

FinTech: EPAM Ukraine, SoftServe PayPal, Stripe, Square	HealthTech: Zocdoc, NIX, HealthTap, Nudge Health, Babylon	EdTech: Duolingo, Coursera, Khan Academy
LegalTech: DocuSign, Clio, DoNotPay	AITech: Grammarly, OpenAI	SportsTech: Strava, STATS Perform, Opta Sports
AgriTech: Infopulse, Trinetix, Granular, FarmerAI, Infarm	GovTech: Diia, Intecracy, ELEKS, OpenGov, SmartCity	CybersecurityTech: Sigma Software, CrowdStrike, Fortinet, CyberArk

Crises may be caused by internal or external factors. In case of internal imbalance or errors, which could be made by employee, the gold standard of actions include investigation the situation in detail; acceptance the fault if it lies on the company or the

product; sincerely apologizing; taking measures to restore public confidence.

But it`s necessary to act as before as during crisis (Table 2). Choosing a team for communication during crisis, the best option in this situation will be C-level representatives, however, it`s necessary to be sure that they will be able to cope with the tasks. Also, it`s important to be sure the whole team knows who will communicate about the crisis. Values matrix can be used in developing communication strategy with determining the communication channels through which communication with target audience will be set. The prepared media kit is useful to refer to when communicating with the media.

Table 2: The algorithm of actions of the company's communicators during the crisis situation.

Before crisis	During the crisis
Determine which situations pose the greatest threat to product and can become a crisis; to make a list of such events; to determine which operational processes are the most important for the functioning of product.	Collect and analyze all information about the crisis situation that occurred.
	Be proactive: communication with the public as soon as possible about what happened and what next steps are.
For each crisis situation, collect a list of potential questions from the public and prepare clear answers that contain facts.	If company is to blame for a crisis situation - admit the mistake and ask for forgiveness, tell about your next steps.
Choose a specific person or people who will communicate with the public during the crisis.	Communicate with your target audience with key messages.
Define the key target audience with whom it will be necessary to communicate.	Control the flow of information about the crisis event and your brand.
Develop a step-by-step communication strategy and key messages.	Stay calm, take care of condition and try not to react to irritants.
Create a media kit in advance.	Record all communication

In crisis situations, reaction has to be immediately: 10% of communication is about the mistake, 90% is about how and what the company does next. During the crisis focusing have to be on response to media inquiries with quick and appropriate comments. It`s useful to make a note of all the media and the requests - in the future work with it may help deepen anticrisis communications plan.

The key development vectors of IT-product companies are orientation on employee self-improvement, quick carrier growth because of colleges experience, the value of ideas. Communication is the key factor during any type of crises that`s why performance

Anticrisis marketing deeply correlates with behavior economics which is transformed from the 2000s onwards, from a centralized research program to a multipolar specialty [12]. Among different behavioral tactics in crises case most effective may be personalization, social proof, limitations and cost avoidance effect.

During crisis adaptation of the content of the application, site, platform to the interests of each individual user should be done. People are more likely to adopt a certain behavior if they see others already engaging in it. One of the examles of using this technique is the distribution of UGC (or user-generated content) – content created by other users of the product. External crisis factors cause limitations. In this case the perception of value increases as soon as it becomes limited or unavailable. And off cause, cost avoidance effect is becoming more important because consumers have limited abilities to buy.

Anticrisis marketing model is basing on modification process of value proposition caused by crisis (Figure 5). Unique value proposition of a product has to be changed and satisfy modified needs of users, i.e. meet market demand (product market fit). So communication messages and positioning are changes too. Such components of ideal customer profile as budget, pain points, factors affecting decision-making are changing: budget is becoming limited, customer pain points are growing, factors are diversified. Such crises as pandemic, war in Ukraine have demonstrated these modifications.

The example of modification process of value proposition clearly may be seen on the activity of such governmental services as Diia in the niche of GovTech which is digital government of Ukraine. It includes portal of online public services, where everything is fast, clear and understandable; mobile application with electronic documents and data about a person from registers; portal to help small and medium-sized businesses; Diia City – special legal regime for the IT industry [13]. At March of 2022 new temporary digital document for the period of martial law appeared in Diia – eDokument. It`s very flexible and contains passport data, RNOKPP (taxpayer card). The document has become automatically available to all users after updating to the latest version. The value matrix of Diia has become more expanded during the crisis. In general value matrix consists of segment or common group of users, their pain points, product value and key messengers.

Anticrisis marketing of IT-products includes performance, brand, product marketing which are
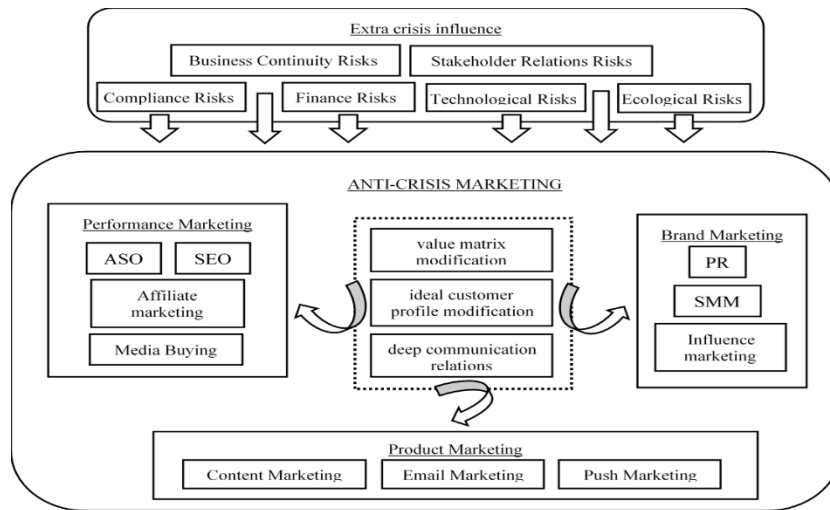
Figure 5: Anticrisis marketing model for IT sector companies.

focusing on communications with consumers. Generally, performance marketing is aimed on effectively attracting new users at the lowest cost but in cases of crisis it`s focusing on reducing business risks. Types of performance marketing include ASO, SEO, Affiliate Marketing and Media Buying. Brand marketing aimed at the use of tools that make it possible to build relationships not only between the product and the user, but also between the entire brand and the user. Types of brand marketing include PR, SMM and Influence Marketing. Product marketing focuses on improving the product based on the user experience, retaining and returning users. Types of product marketing include content, email and push marketing.

The results of highlighting specific points of anticrisis decisions may be clearly approved on Ukrainian IT-companies which nowadays operates in high crisis environment because of unstable conditions. The most obvious is the activity of of such governmental services as Diia in the niche of GovTech which is digital government of Ukraine. It includes portal of online public services, where everything is fast, clear and understandable; mobile application with electronic documents and data about a person from registers; portal to help small and medium-sized businesses; Diia City – special legal regime for the IT industry [13].

At March of 2022 new temporary digital document for the period of martial law appeared in Diia – eDokument. It`s very flexible and contains passport data, RNOKPP (taxpayer card). The document has become automatically available to all users after updating to the latest version. The value matrix of Diia has become more expanded during the

crisis. Such actions of this company demonstrate quick reaction on changed environment. Such approval result demonstrates activating of performance marketing. The number of Diia' users has increased by 20% since the beginning of the full-scale invasion and now are 18.6 million [13]. Moreover, content is extending each time. There are a lot of anticrisis decisions on website to help businesses in Ukraine to recover (more than 2.5 thousand publications on Instagram) (Figure 6).
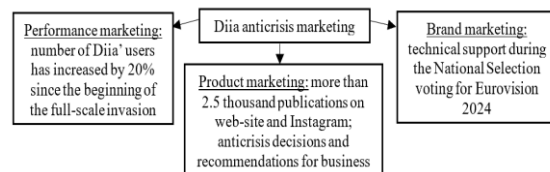
Figure 6: Anticrisis marketing of Diia IT-company.

The evidence of current using of brand marketing instruments is technical support by Diia during the National Selection voting for Eurovision 2024. Ukrainians had an opportunity to make the choice as easily as possible just using Diia application. But during the process of voting the internal crisis situation was appeared: the system crashed and it was impossible to make a choice for some period of time. The rules of anticrisis reaction were used (which were discussed above): quick and appropriate comments were made in media. A lots of media resources were used to explain the situation. Minister of Digital Transformation of Ukraine, Mykhailo Fedorov, announced the record number of users for the entire history of the project. Instead of 800 requests per

second, as was the case during last year's Eurovision voting, there were 15,000 requests per second.

4 CONCLUSIONS

So, during a crisis, adapting the content of applications, websites, or platforms to individual user interests is crucial. This can be achieved through personalized recommendations, tailored communication, or customized user experiences. Personalization helps users find relevant information more efficiently, making the overall experience more valuable to them.

People are more likely to adopt a certain behavior if they see others engaging in it. Leveraging social proof, such as showcasing user testimonials, ratings, or the number of users engaged, can positively influence the behavior of new users. The distribution of User-Generated Content is a powerful tool. Users are often more trusting of content created by their peers, which can enhance the perceived value of a product or service.

External crisis factors may lead to limitations in the availability of products or services. The scarcity principle suggests that the perception of value increases when something becomes limited or unavailable. Creating a sense of urgency or scarcity can drive consumer action, encouraging them to take advantage of an opportunity before it's no longer accessible.

During a crisis, consumers may have limited financial resources, making cost avoidance a significant factor in decision-making. Emphasizing cost savings, discounts, or highlighting the value proposition of a product/service becomes more crucial in such situations.

Overall, these principles highlight the importance of understanding human behavior and psychology in times of crisis and ese them in anticrisis marketing model for IT sector companies. Adapting strategies to address individual needs, leveraging social influence, and emphasizing value and affordability can contribute to better user engagement and resilience during challenging times.

During times of crisis, effective marketing strategies for IT products often involve a combination of performance, brand, and product marketing with a strong focus on communication. Combining these marketing approaches can create a comprehensive strategy that not only promotes IT product but also addresses the unique challenges presented by a crisis. The key is to remain agile, empathetic, and responsive to the changing landscape.

The results of solutions approval are presented on example of Diia anticrisis marketing which includes as external crisis factors (full-scale invasion) and internal crisis (technical crashing during Eurovision National Selection). Its effective marketing policy has minimized negative influence on its activity.

REFERENCES

[1] R. Arcos and H. Smith, "Digital communication and hybrid threats presentation", Icono14, vol. 19(1), pp. 1-14, 2021, doi: 10.7195/RI14.V19I1.1662.

[2] G.P. Siroli, "Considerations on the cyber domain as the new worldwide battlefield", International Spectator, vol. 53(2), pp. 111-123, 2018, doi: 10.1080/03932729.2018.1453583.

[3] S.R. Clegg, M.P.E. Cunha, A. López, E. Sirage, and A. Rego, "Tackling sustainable development goals through new space", Project Leadership and Society, vol. 5, no. 100107, 2024 [Online]. Available: https://doi.org/10.1016/j.plas.2023.100107.

[4] O. Arefieva, O. Polous, V. Arefiev, Yu. Kopcha, and S.K. Gupta, "Intellectualization of human capital development in digital economics", International Journal of Advanced Science and Technology, vol. 29, no. 8s, pp. 2297-2303, 2020.

[5] I. Al-Amoudi and E. Lazega, "Post-Human Institutions and Organizations: Confronting the Matrix", London, Routledge, 2019 [Online]. Available: https://doi.org/10.4324/9781351233477.

[6] A. Dabbous and K.A. Barakat, "The road towards environmental sustainability: Investigating the role of information and communication technologies and green technology innovations", Journal of Cleaner Production, vol. 432, no. 139826, 2023 [Online]. Available: https://doi.org/10.1016/j.jclepro.2023.139826.

[7] T. Chin, W. Zhang, and I.M. Jawahar, "Intellectual capital and employee innovative behavior in cross-border e-commerce enterprises: the moderating role of career sustainability", Journal of Intellectual Capital, vol. 24(6), pp. 1532–1549, 2023 [Online]. Available: https://doi.org/10.1108/JIC-10-2022-0193.

[8] IT Ukraine Association. Do IT like Ukraine. [Online]. Available: https://itukraine.org.ua/files/reports/2022/DoITLikeUkraine2022.pdf.

[9] State Statistics Service of Ukraine. [Online]. Available: https://www.ukrstat.gov.ua/.

[10] Better Regulation Delivery Office (BRDO) and YouControl. Overview of the TOP 30 Ukrainian IT companies. [Online]. Available: https://brdo.com.ua/en/wp-content/uploads/sites/5/2023/07/Top-30-UA-IT-companies-overview-EN.pdf.

[11] Genesis Press kit. Genesis co-founding IT-company. 2023. [Online]. Available: https://www.gen.tech.

[12] A. Truc, "Forty Years of Behavioral Economics", The European Journal of the History of Economic, vol. 29, no. 3, pp. 393-437, 2022.

[13] Digital Government. What is Diia? [Online]. Available: https://plan2.diia.gov.ua.

Statistical Modeling of Determinants Influencing Economic Security in the Context of Sustainable Development and National Security

Olha Kovalchuk[1], Kateryna Berezka[1], Mariia Masonkova[2] Nataliia Chudyk[1],
Vasyl Ukhach[1] and Yuriy Pilyukov[1]

[1]*West Ukrainian National University, Lvivska Str. 11, 46009 Ternopil, Ukraine*
[2]*International Affairs Department, Kherson State Maritime Academy, Ushakova Avenue 20, 73009 Kherson, Ukraine*
o.kovalchuk@wunu.edu.ua, k.berezka@wunu.edu.ua, masyonkova@gmail.com, n.chudyk@wunu.edu.ua,
v.ukhach@wunu.edu.ua, yu.pilyukov@wunu.edu.ua

Abstract: The aggravation of the political crisis and the increase in geopolitical conflicts destabilize the world economy and create serious challenges for the national security of all countries of the world. Today, a new theoretical and methodological approach to understanding economic security, which is an important dimension of national security, is needed. Our research uses an approach using multivariate statistical modeling based on a unique set of generally recognized macroeconomic indicators for the year 2022 in most countries of the world. We created a factor model for detections of determinants of influence on national economic security and a model of canonical correlations for evaluating complex two-way relationships between selected determinants and generally recognized reliable indices that can serve as approximate estimates of the level of economic security in the context of the concept of sustainable development (Human Development Index) and national security (GPI – Global Peace Index). The obtained results are not without flaws and are only approximate estimates. However, they can improve the understanding of the problems of strengthening international coordination and cooperation in strengthening economic security, ensuring sustainable development and implementing the tasks of national security strategies of various states of the world.

1 INTRODUCTION

Geopolitical competition has changed the world economy and complicated national security issues. Today, there are no unified methodological approaches to defining the concepts of economic security, national security and their determinants [1]. However, the economic component is present in all proposed systems of national security indicators [2–4]. It is generally recognized that economic security is closely related to both national and global security [5]. Economic security is an important dimension of national security. This is one of the defining strategic priorities for many states of the world.

Economic leverage is often used as a weapon in international conflicts and for political purposes. Ensuring economic security today is one of the mandatory requirements not only for sustainable development but also for the survival of the nation. This is one of the basic aspects of understanding national security [6], solving modern global challenges [7] and ensuring the goals of sustainable development [8]. Economic security is defined as the country's ability to resist threats and maintain its sustainable development and is a measure of the state's competitiveness in the world economy [9]. Ensuring the economic security of the state is a set of measures aimed at protecting national economic interests from external and internal threats and creating conditions for the stable and effective functioning of the national economic system [10]. The main components of ensuring economic security are the protection of the economic sovereignty of the state from external economic threats, the stability of the national currency and the prevention of excessive inflation, the promotion of sustainable economic growth and increasing the competitiveness of the economy, the protection of strategic sectors of the economy, food security, the creation of favorable conditions for business development, and the fight against the shadow economy, social protection of the population [11].

Economic security is one of the basic dimensions of national security and is interconnected with all its other components [4]. There is a close relationship between a country's economic security and geopolitical conflicts. Economic weakness can provoke external aggression and conflicts. Countries with weak economies are more vulnerable to encroachment and expansion. Economic sanctions and restrictions are often used as a pressure tool in international conflicts. They can significantly harm the economic security of the country under sanctions. Military conflicts cause significant damage to economic security through the destruction of infrastructure, disruption of supply chains, outflow of investment, etc. Cyber attacks and other hybrid threats can be aimed at undermining the opponent's economic security [12]. Countries may resort to economic espionage or theft of intellectual property to gain a competitive advantage that threatens the security of others. Economic claims and disputes over resources, territories, borders are often a source of conflicts. Strengthening economic security is an important factor in preventing conflicts and strengthening the country's international position. Today, a clear understanding of the determinants of national economic security and a multi-faceted qualitative analysis of assessments of the impact of various factors on sustainable economic development and national security of the countries of the world is necessary.

Assessing the state of economic security of the countries of the world is a complex problem that has been studied by scientists from all over the world for many years. To date, a unified system of indicators that would reliably and fully identify the real level of economic stability has not yet been developed. It is also practically impossible to directly measure all the values that assess the volumes and characteristics of the determinants of economic security. Often, the number and content of factors that have a significant impact on economic stability are unknown. However, models built based on empirical data provide an opportunity to obtain new information important for decision-making in the field of economic security. The purpose of this study is to improve understanding of the risks, challenges and opportunities for economic security, which is a significant component of national security. The obtained results can provide additional information regarding the determination of priority directions for the economic national security sector and the development of effective national security strategies.

2 RELATED WORK

Economic security is a complex debatable issue. Many modern researchers update the problems of the importance of ensuring economic security in the context of geopolitical transformations and growing global threats [11]. L. Retter et al. investigated the relationship between the economy and national security in the context of globalization and economic integration [4]. I. Gryshova et al. proved that the security of the national economy characterizes the country's ability to resist the destabilizing effects of various internal and external threats and ensure its sustainable development [13]. J. Jekl et al. developed the theoretical foundations of the possible application of economic and geographic metrics for assessing national security based on well-known metrics and actual data [14]. J. Li et al. explored the advantages of economic security and environmental protection from the point of view of sustainable development and the scientific and technical ecological environment. They proposed a system of indicators for the coordinated development of eco-technological innovations and the economic environment [15]. O. Solosich et al. proposed a methodological toolkit for assessing the economic security of regions as a possible potential-forming space in terms of intellectualization [16]. G. Yuan et al. studied the regularities of economic security and economic fluctuations as an important component of ensuring economic stability and scientific development. Based on cloud computing technology and intelligent data analysis, they developed a model that can adaptively assess and warn about the state of economic security [17]. M. Mogyorós et al. created a structural equation model for measuring and monitoring economic insecurity based on macroeconomic data of individual EU countries for 2005-2020 [18]. A. Sunduk et al. investigated the peculiarities of ensuring the economic security of regions in the event of risks and threats in the information sphere. They received assessments of the level of economic security based on the system of indicators of national economic security and its regions [19]. The problem of ensuring economic security, which determines the state's ability to resist various internal and external threats, to ensure sustainable development and competitiveness in the global economic system, is becoming more and more urgent in the context of the aggravation of the global political crisis. Research on the application of statistical modeling to improve the understanding and assessment of the determinants of economic security in the context of ensuring sustainable development and strengthening national

security today has acquired a new meaning and requires a more detailed multifaceted study and transparent understanding.

3 METHODOLOGY

Today, there are various theoretical and methodological approaches to monitoring and assessing the state of economic security of the world's countries and the level of national security. Different scientists use different sets of macroeconomic indicators, aggregate indicators, and indices. Representatives of economic science use a set of indicators that reflect the most important areas of economic security: gross domestic product (GDP), the share of fixed capital investment in GDP, the share of defense expenditures in GDP, the share of expenditures on social sciences in GDP, the share of innovative products in total industrial output, etc. [20]. Researchers in the field of social sciences calculated the coefficient of economic security based on a system of the following indicators: gross product, revenues of the budget, budget expenditures, amount of actual revenues from taxes and fees to state trust funds, capital investments, foreign direct investment, volume of exports of goods, and cost of research [19]. Individual researchers identified threats to economic security at the macroeconomic level (median income, GINI coefficient, risk of poverty, social government expenditures, Internet access, number of deaths in mental disorders, Democracy Index) and microeconomic level (unemployment, sickness disability, widowhood, divorces, old age) [18]. Empirical comparisons of the economic security of the world's countries are complicated by the fact that not all countries of the world have reliable estimates of many generally accepted indicators.

The economic system is extremely complex. Thousands of economic indicators have been developed that characterize its development. The redundancy of such information creates significant difficulties in assessing the real state of national economic security. We offer an approach based on the development of statistical models based on empirical data, which can provide information about the most important indicators that determine the state of economic security of the state and significantly affect the level of national security. Based on literature review, we formed the empirical base of our research from the following reliable indicators of the state of various areas of development of national economies,

which can be potential significant determinants of the impact on national economic security: economies by corruption levels; informal economy (size as a percentage of GDP) [21]; GDP per capita, unemployment rate, inflation rate [22]; poverty rate by country [23]; external debt [24]; Gini coefficient (measures economic inequality) [25]; WB – well-being (measures feeling healthy and happy), WBG – well-being gaps between the top and bottom halves of each country's populations [26]; Index of Economic Freedom [27]; Global Food Security Index [28]; foreign direct investment, exports of goods and services, imports of goods and services, military expenditure (percentage of GDP) [29]; social spending public [30]; e-security (aggregated measure of Internet affordability, Internet quality, E-infrastructure, E-security, E-government) [31].

According to the concept of sustainable development, GDP is not an adequate assessment of economic development and economic security of countries, but only an averaged expression of the level of economic security of citizens. Important in this regard are measures of well-being, equal opportunities, poverty level, economic freedom, access to benefits, etc. Most of these dimensions are taken into account when evaluating the Human Development Index (HDI) [32]. In our research, we chose the HDI as the optimal approximate measure of the level of economic security of the countries of the world.

While there is no universal aggregate measure, different countries use different theoretical approaches to national security assessment [2, 3, 6, 11, 35]. For example, the Safety Index is calculated based on the level of crime in the country [33]. The modern interpretation of national security assumes the protection of the country's citizens primarily from external threats to the geostrategic environment of the state and threats to personal security in the context of sustainable development. For this study, the Global Peace Index was chosen as an approximate measure of national security. He evaluates peaceful countries by the following domains: safety and security, ongoing conflict, and militarization [34].

To understand the state of national security, a clear idea of the aspects vital for the sustainable functioning of society is required. The assessment of the state of national security from the side of economic factors can be understood by highlighting the interaction between the determinants of the impact on economic security and the measures of national security of the countries of the world.

4 RESULTS AND DISCUSSION

The data set for empirical analysis consisted of the values of indicators selected for the study of 207 countries of the world for the year 2022, obtained from reliable statistical sources [21-32, 34].

The variables used in the empirical study were: EXP (Exports of goods and services), IMP (Imports of goods and services), GINI (Gini coefficient), ESEC (e-security), WB (well-being), WBG (well-being gaps between the top and bottom halves of populations), IEF (Index of Economic Freedom), UE (unemployment rate), FDI (foreign direct investment, net inflows), IE (informal economy), ECR (economies by corruption levels), GDP (GDP per capita), IR (inflation rate), PR (poverty rate by country), ED (external debt), ME (military expenditure), GFSI (Global Food Security Index), GPI (Global Peace Index), HDI (Human Development Index).

Empirical results obtained by Statistica software.

Factor analysis was used to aggregate the initial data and determine the significant determinants of the impact on economic security. By the principal components method, two factors are selected. Table 1 shows the calculated factor loadings, which are the correlation coefficients between the original variables and the selected factors.

Table 1: Factor loadings (Varimax normalized).

Extraction: Principal components Marked loadings are > 0.7					
Variable	F_1	F_2	Variable	F_1	F_2
GINI	-0.80	0.07	IE	-0.81	-0.42
ESEC	0.82	-0.01	ECR	0.92	0.18
WB	0.86	0.20	GDP	0.88	0.21
WBG	-0.92	-0.13	IR	-0.26	-0.20
EXP	0.15	0.92	PR	-0.78	-0.12
IMP	0.13	0.96	ED	-0.73	0.01
IEF	0.91	-0.01	ME	-0.01	0.35
UE	-0.71	-0.13	GFSI	0.87	0.28
FDI	0.04	0.90			

The higher the modulus load, the closer the factor is to the input variable. A positive sign of the factor loading indicates a direct relationship between the variable and the factor, and a negative sign indicates an inverse relationship. Factor loadings reflect the most important information for the interpretation of the obtained factors.

The variables ECR (0.92), WBG (–0.92), IEF (0.91), GDP (0.88), GFSI (0.87), and WB (0.86) are closely correlated with the first selected factor. The variables IMP (0.96), EXP (0.92), and FDI (0.90) are closely correlated with the second factor. The variables WBG, GINI, and IE have a negative impact on national economic security. IR and ME variables did not enter into any of the extracted factors.

The following factor model was constructed:

$$F_1 = 0.92{\times}ECR - 0.92{\times}WBG + 0.91{\times}IEF + \\ - 0.88{\times}GDP + 0.87{\times}GFSI + 0.86{\times}WB + \\ + 0.82{\times}ESEC - 0.81{\times}IE - 0.80{\times}GINI - \\ - 0.78{\times}PR - - 0.73{\times}ED - 0.71{\times}UE, \quad (1)$$

$$F_2 = 0.96{\times}IMP + 0.92{\times}EXP + 0.90{\times}FDI. \quad (2)$$

As a result of building the factor model, 2 factors were identified: the first of them is a linear combination of endogenous factors influencing the economic security of the countries of the world, and the second forms exogenous indicators of economic security.

So, in the context of the concept of sustainable development, the endogenous determinants of the impact on the economic security of most countries of the world are corruption, inequality of well-being between different strata of the population within the country, promotion of economic opportunities and prosperity, GDP, food security, well-being, which combines welfare and prosperity. The modern meaning of endogenous determinants of influence on economic security is not in a satisfactory average level of material well-being of the country's population, but in ensuring equal financial opportunities for all and guaranteeing safe and stable well-being. Import, export and foreign direct investment are exogenous determinants of the impact on the economic security of most modern countries of the world.

Canonical correlation analysis was conducted to assess complex two-way relationships between selected determinants and generally recognized reliable indices that can serve as approximate estimates of the level of economic security in the context of the concept of sustainable development (Human Development Index) and national security (GPI – Global Peace Index). This method determines the interdependencies between two sets of variables.

The set of dependent variables (left set) consists of the following variables:

- HDI – Human Development Index, which is the optimal approximate measure of the state of economic security of the countries of the world;
- GPI – Global Peace Index, which is an approximate measure of national security.

As can be seen from Figure 1, there is a direct relationship between the HDI, which we chose as an approximate measure of economic security, and the GPI, which is an approximate estimate of the level of
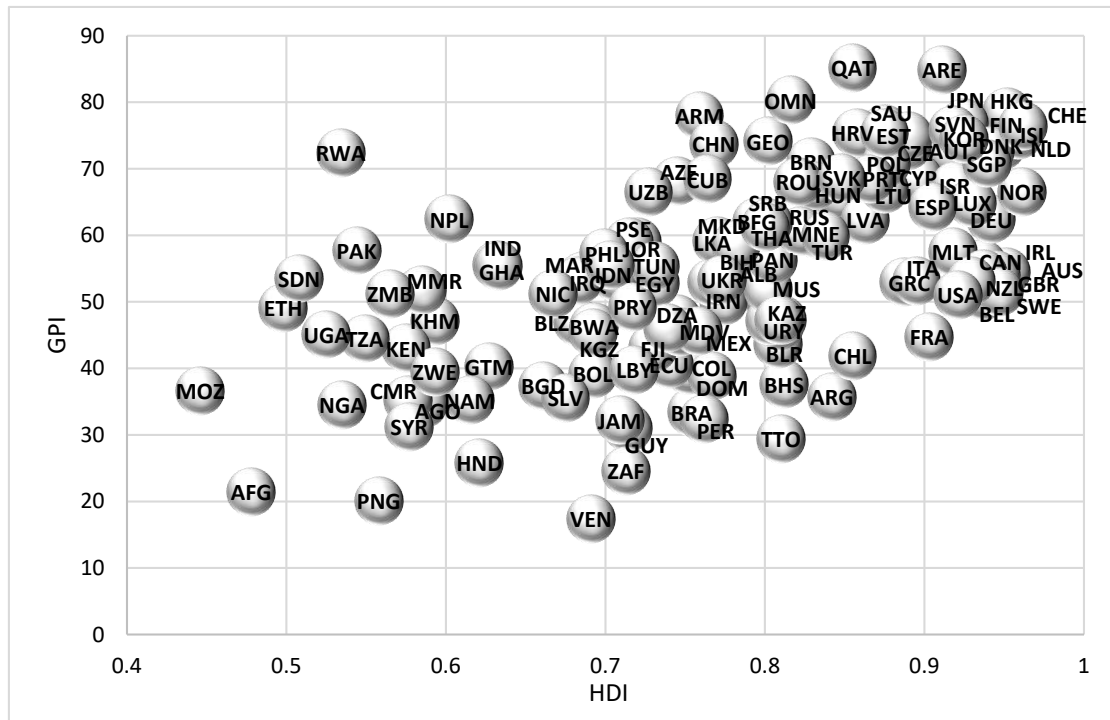
Figure 1: Plot of HDI and GPI dependence for most countries of the world.

national security, for the vast majority of analyzed countries. The set of independent variables (right set) consisted of all the indicators included in the factor model (1), (2).

Table 2 presents the overall results of the canonical analysis.

Table 2: Canonical analysis summary.

Canonical R: .96124 Chi2(30) = 126.18 p = 0.0000			
	Left Set	Right Set	
No. of variables	2	15	
Variance extracted	100.00%	54.68%	
Total redundancy	83.66%	48.95%	
Variables:	GPI	GINI	IMP	ECR
	HDI	ESEC	IEF	GDP
		WB	UE	PR
		WBG	FDI	ED
		EXP	IE	GFSI

As can be seen from Table 2, the obtained results are statistically significant ($p < 0.01$). The canonical R-value is (0.96). Hence, there is a close correlation between the variables in the left and right sets. The total share of the variance of the features of the first

group, allocated with the help of the first canonical variable of the right set, is 100.00%. The total share of the variance of the features of the second group (right set) of the canonical variable of the left set is 54.68%.

The total loss for variables of the first group equals 83.66%, and for variables of the second group – 48.95%. Therefore, almost 84% of the variation in the indicators of the economic component of national security is determined by changes in the analyzed determinants of national economic security. At the same time, the indicators of the economic security of the world's countries (the left set) determine almost 49% of the variation of the studied determinants of national economic security (the right set).

The obtained results indicate the sufficient accuracy of the constructed canonical model – less than 17% of the variance of the variables of the left set depends on factors that were not included in the model due to the limitations of the canonical analysis.

As can be seen from Table 3, the first root is statistically significant ($p < 0.01$). The resulting canonical R^2 value is sufficiently large (0,92), $\chi_e^2 = 126.18$. $\chi_0^2(0.01;25) = 50.9$. $\chi_e^2 > \chi_0^2$. The first canonical root will be considered in more detail.

Table 3: Chi-square tests with successive roots removed.

Root Removed	Canonical R	Canonical R2	Chi-sqr.	df	p
0	0.96	0.93	126.18	30	0.00
1	0.78	0.60	33.42	14	0.03

The calculated canonical weights are correlations of the canonical root with the variables in the initial sets (Table 4, 5). The HDI variable is highly correlated with the canonical root. Therefore, the HDI is an adequate proxy for national economic security and the GPI is an acceptable measure of national security in the context of our study.

Table 4: Factor structure, left set.

Variable	Root 1
GPI	0.68
HDI	0.99

Table 5: Factor structure, right set.

Variable	Root 1	Variable	Root 1
GINI	-0.66	FDI	0.37
ESEC	0.85	IE	-0.86
WB	0.90	ECR	0.89
WBG	-0.92	GDP	0.83
EXP	0.30	PR	-0.60
IMP	0.29	ED	0.63
IEF	0.82	GFSI	0.95
UE	-0.26		

The following canonical model was obtained:

$$Z_y = 0.99 \times HDI + 0.68 \times GPI, \quad (3)$$

$$\begin{aligned} Z_x = {} & 0.95 \times GFSI - 0.92 \times WBG + 0.90 \times WB + \\ & + 0.89 \times ECR - 0.86 \times IE + 0.85 \times ESEC + \\ & + 0.83 \times GDP + 0.82 \times IEF - 0.66 \times GINI + \\ & + 0.63 \times ED - 0.60 \times PR + 0.30 \times EXP + \\ & + 0.29 \times IMP + 0.37 \times FDI - 0.26 \times UE. \end{aligned} \quad (4)$$

As can be seen from Figure 2, there are no sharp outliers and deviations from the regression line. For example, formations in the form of U or S. Therefore, it can be concluded that no noticeable violations of the main assumptions of canonical analysis are observed.

The built model (3), (4) gives reason to state that the most important determinants of the impact on economic security in the context of sustainable development and national security of the countries of the world are food security, inequality of well-being between different layers of the population, well-being, corruption, informal economy, e-security, GDP, Index of Economic Freedom and economic inequality.
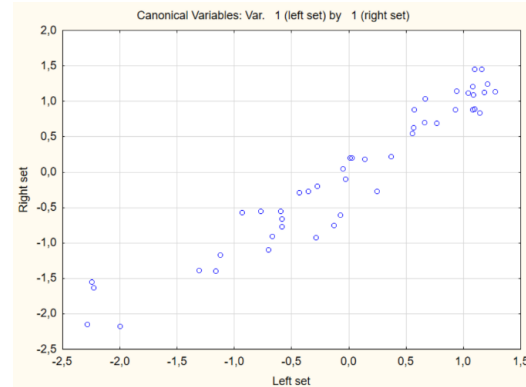
Figure 2: Graphic representation of canonical values.

5 CONCLUSIONS

We hope you find the information in this template useful in the preparation of your submission.

The concept of economic security in the modern context is extremely complex and multifaceted and includes a wide range of aspects related to ensuring the stability and sustainability of national economies in the context of sustainable development. Economic security is one of the basic dimensions of national security and is interconnected with all its other components. This paper proposes an approach based on multivariate statistical modeling to determine and evaluate the determinants of the impact on the economic security of the countries of the world in the context of sustainable development and national security. Based on reliable statistical indicators of 207 countries of the world for 2022, a factor model was built. 2 factors are identified, which are latent indicators, calculated as linear combinations of initial indicators. For most countries of the world, the endogenous determinants of the impact on economic security are corruption, inequality of well-being between different strata of the population within the country, promotion of economic opportunities and prosperity, GDP, food security, well-being, which combines welfare and prosperity, ensuring equal financial opportunities for everyone and guaranteeing safe and stable well-being. The factor of endogenous determinants of influence was formed by import, export, and foreign direct investment.

The authors developed a model of canonical correlations for evaluating complex two-way relationships between the determinants selected as a result of factor analysis and approximate estimates of the level of economic security in the context of the concept of sustainable development (HDI) and national security (GPI). The obtained results are only

approximate. However, they can be used to improve the understanding of the problems of maintaining sustainable economic growth, ensuring economic security, developing anti-crisis strategies of states and providing useful information for making managerial decisions in the field of strengthening the economic and national security of most countries of the world. The issue of our next research will be the assessment of the impact of modern international military conflicts on national economic security.

REFERENCES

[1] L. G. Voronovska, "The phenomenon of security in philosophical discourse," Actual problems of philosophy and sociology, vol. 33, 2021, pp. 23-29.

[2] National Security Strategy, The White House, Washington, 2022, 48 p.

[3] A. Shull and W. Wark, Reimagining a Canadian National Security Strategy, CIGI Special Report, 2022, 33 p.

[4] O. Kovalchuk, M. Karpinski, L. Babala, M. Kasianchuk, and R. Shevchuk, "The canonical discriminant model of the environmental security threats, Complexity, vol. 2023: 5584750, 2023, 15 p.

[5] L. Retter, E. Frinking, S. Hoorens, A. Lynch, F.Nederveen, and W. Phillips, "Relationships between the Economy and National Security: Analysis and Considerations for Economic Security Policy in the Netherlands", RAND Corporation: Santa Monica, CA, USA; Cambridge, UK, 2020, 182 p.

[6] 2022 National Security Public Survey Report, Ipsos, 2022, 57 p.

[7] World Economic Forum, "The Global Risks Report," 2023, 98 p.

[8] L. Shi, L. Han, F. Yang, and L. Gao, "The Evolution of Sustainable Development Theory: Types, Goals, and Research Prospects". Sustainability, 11(24): 7158, 2019.

[9] M. Minke and M. Merel, "Economic Security. Methodology and Approach", Netherlands Institute of International Relations, 2019.

[10] D.H. McCormick, C.E. Luftig, and J.M. Cunningham "Economic Might, National Security, and the Future of American Statecraft", Texas National Security Review, 3 (3), pp. 50-75, 2020.

[11] A. Posen and D. Tarullo, "Report of the Working Group on Economics and National Security", Princeton Project on National Security, 2017.

[12] O. Kovalchuk, M. Shynkaryk, M. Masonkova, and S. Banakh, "Cybersecurity: technology vs safety", ACIT-2020, Deggendorf, Germany, pp. 765–768.

[13] I. Gryshova, M. Kyzym, I. Hubarieva, V. Khaustova, A. Livinskyi, and M. Koroshenko, "Assessment of the EU and Ukraine economic security and its influence on their sustainable economic development", Sustainability, 12 (18), 7692, 2020.

[14] J. Jekl and J. Jánský, "Security challenges and economic-geographical metrics for analyzing safety to achieve sustainable protection", Sustainability, 14 (22), 15161, 2022.

[15] J. Li and Q. Hai, "Evaluation of economic security and environmental protection benefits from the perspective of sustainable development and technological ecological environment", Sustainability, 15 (7), 6072, 2023.

[16] O. Solosich, O.Popelo, O. Nusinova, M. Derhaliuk, and S. Tulchynska, "Ensuring economic security of regions as a potential-forming space in the conditions of intellectualization", Academy of Entrepreneurship Journal, 27(6), 2021, pp. 1-8.

[17] Y. Guanghui, F. Xie, and H. Tan, "Construction of economic security early warning system based on cloud computing and data mining", Comput. Intell. Neurosci., vol. 2022: 2080840, 2022, 12 p.

[18] M. Mogyorósi, K. M. Bella, I. R. Kazimir, and T. Cseph, "Measurement of economic insecurity in the European Union between 2005 and 2020", 37th IARIW General Conference, Luxembourg, August 22-26, 2022.

[19] A. Sunduk, Yu. Zhukova, P. Shvets, and V. Loiko. "Economic security parameters in the conditions of increasing information risks and threats", CEUR Workshop Proceedings, CPITS-II-2021: Cybersecurity Providing in Information and Telecommunication Systems, Kyiv, Ukraine, October 2021.

[20] V. Hnatenko, "Economic Security in The State National Security System", Polish Political Science Yearbook, 2020, pp.: 96-105.

[21] World Economics, 2023, [Online]. Available: https://www.worldeconomics.com.

[22] Trading Economics, 2023, [Online]. Available: https://tradingeconomics.com.

[23] Poverty Rate by Country. Wisewoter, 2023, [Online]. Available: https://wisevoter.com.

[24] List of countries by external debt. Wikipedia, 2023.

[25] Gini coefficient. Our Word in Tada, 2023, [Online]. Available: https://en.wikipedia.org.

[26] J. F. Helliwell, R. Layard, J. D. Sachs, J.-E. De Neve, L. B. Aknin, and S. Wang. World Happiness Report, 2023. 162 p.

[27] B. Kim, 2023 Index of Economic Freedom, The Heritage Foundation. DC 20002, 418 p., 2023.

[28] Global Food Security Index 2022, Economist Impact, 2023, [Online]. Available: https://impact.economist.com.

[29] The World Bank, 2023, [Online]. Available: https://data.worldbank.org.

[30] Social spending. OECD Data, 2023, [Online]. Available: https://data.oecd.org.

[31] Which Country Has the Best Digital Well-Being in 2023? [Online]. Available: https://www.visualcapitalist.com.

[32] Human Development Index, Human Development Report, 2022, [Online]. Available: https://hdr.undp.org.

[33] Safety Index by Country 2023, Numbeo, 2023. Available online: https://www.numbeo.com.

[34] Institute for Economics & Peace, Global Peace Index 2023, [Online]. Available: https://www.economicsandpeace.org.

[35] K. Berezka and O. Kovalchuk, "Modelling factors connected with the effect of international migration for security and economy", Econometrics. Advances in Applied Data Analysis, vol. 23(4), 2019, pp. 30-42.

Control and Automation

Proceedings of the 12th International Conference on Applied Innovations in IT (ICAIIT)

An Innovative Approach of API Automation Testing Implemented on Cloud Environments Using Container Management Services

Tanja Dimova[1], Igor Kalendar[1], Daniel Denkovski[2], Danijela Efnusheva[2] and Marija Kalendar[2]

[1]*MIR! Software Solutions, Londonska Str. 6, 1000 Skopje, N. Macedonia*
[2]*Computer Technologies and Engineering Department, Faculty of Electrical Engineering and Infomation Technologies, University "SS. Cyril and Methodius" in Skopje, Rugjer Boshkovikj Str. 18, 1000 Skopje, N. Macedonia*
{tanjadimova8, igor.kalendar}@gmail.com,{danield, danijela,marijaka}@feit.ukim.edu.mk

Keywords: Automation Testing, Cloud Platforms, Container Services, Docker, Kubernetes, Google Cloud Platform, Services Monitoring.

Abstract: This research paper focuses on developing a complete system for daily automation testing of comprehensive web applications implemented on cloud environments, encompassing the execution of automated API tests, real-time monitoring and results visualization of the testing environments. Despite the tools for developing automated API tests, the study uses containerization tools as Docker and Kubernetes, showcasing their integration into a cohesive testing framework. Furthermore, the implementation leverages the potential of the Google Cloud Platform (GCP) to demonstrate the usage of cloud computing services, emphasizing scalability and efficiency. Additionally, the paper details the integration of monitoring tools, specifically Elasticsearch, to assess and visualize the health and performance of the underlying Kubernetes cluster. Through a comprehensive approach, encompassing a wide variety of tools, the research establishes a continuous and automated testing environment essential for cutting-edge software applications. Results showcase the successful orchestration of all the technologies, highlighting their collective impact on achieving a robust and efficient system for continuous automation testing and monitoring.

1 INTRODUCTION

In today's world of technology, every successful organization needs to be connected and accessible over the Internet. This would require setting-up an organizational data center. The traditional data center is on-premises, performing all of its functions in a physical location within the enterprise's office space, and usually managed by an in-house IT team. Thus, the digitalization of business processes in a traditional way would entail: a physical data center, buying and managing hardware (servers), software and licenses, creating a physical network, building the entire infrastructure and hiring a team of experts to manage or maintain this data center.

On the other hand, cloud computing is the current state-of-the-art technology for delivery of computing resources and services such as servers, data storage, databases, networking, software, analytics, and intelligence over the Internet to offer flexible resources, faster innovation, and economies of scale. Using this paradigm, instead of owning data centers, organizations can rent access to the service provider infrastructure (storage, computer servers, databases, networks) and pay only for the resources they use.

Consequently, the transition from traditional to cloud computing data centers is a very current topic. Our research focuses on a very specific topic, automation of API testing of a modern web-based application in a cloud environment, encompassing, researching and evaluating all aspects of building a complete, comprehensive, scalable and manageable automation testing cloud solution.

The rest of the paper is organized as follows. Section 2 elaborates the related work in the area of cloud computing and automation testing. Section 3 elaborates the design of the system for automated API testing. Section 4 presents the container elements incorporated into the automated testing environment, while Section 5 considers the cloud elements comprising the system. Section 6 showcases the performance characteristics of the Kubernetes cluster system deployment in the cloud, thus comparing all system elements and technologies. Section 7 concludes the paper pointing out the main results and benefits of the system.

2   RELATED WORK

Research in the domain of automated API testing, containerization, and cloud computing has provided valuable insights that enhance the understanding of these technologies. Authors in [1] examine DevOps technologies, offering perspectives on continuous integration and deployment, laying a foundation for our research on automating testing services. The research on API testing using Postman [2] contributes insights into automated API tests execution, while [3] presents considerations for secure implementation of the automated testing processes. Authors in [4] offer insights into building modern clouds by integrating Docker, Kubernetes, and cloud platforms, aligning with our research context. Additionally, [5] presents a study on the Grafana visualization tool, using an Influx DB data source. The authors in [6] provide a foundation for understanding the role of monitoring tools in the system health and performance evaluation, while the study on observability using Kubernetes operators [7] complements our goal of improving system monitoring and visualization by automating deployment, visualization, and monitoring within the cloud system and the Kubernetes cluster.

Our research builds upon these foundations, introducing a novel approach that integrates regression end-to-end API testing, containerization,

and cloud computing. The solution optimizes daily testing processes, improves test coverage, and ensures quick error detection, making it a significant advancement in the field of automated testing. The comprehensive integration of these technologies positions our system as a more scalable, reliable, and efficient solution for automated testing services in modern cloud environments.

3   DESIGNING AN INNOVATIVE SYSTEM FOR AUTOMATED API TESTING

Automation testing is crucial in the process of modern software development for its efficiency, accuracy, and speed. It ensures comprehensive test coverage, particularly for regression testing, and integrates seamlessly into *Continuous Integration/Continuous Deployment* pipelines. By leveraging new tools and optimizing the testing ecosystem, automatic testing facilitates early bug detection and reduces overall testing costs for the software development company.

Illustrated in Figure 1 is the proposed design of an innovative holistic system enabling automated API testing utilizing cloud computing resources and advanced containerization technologies.
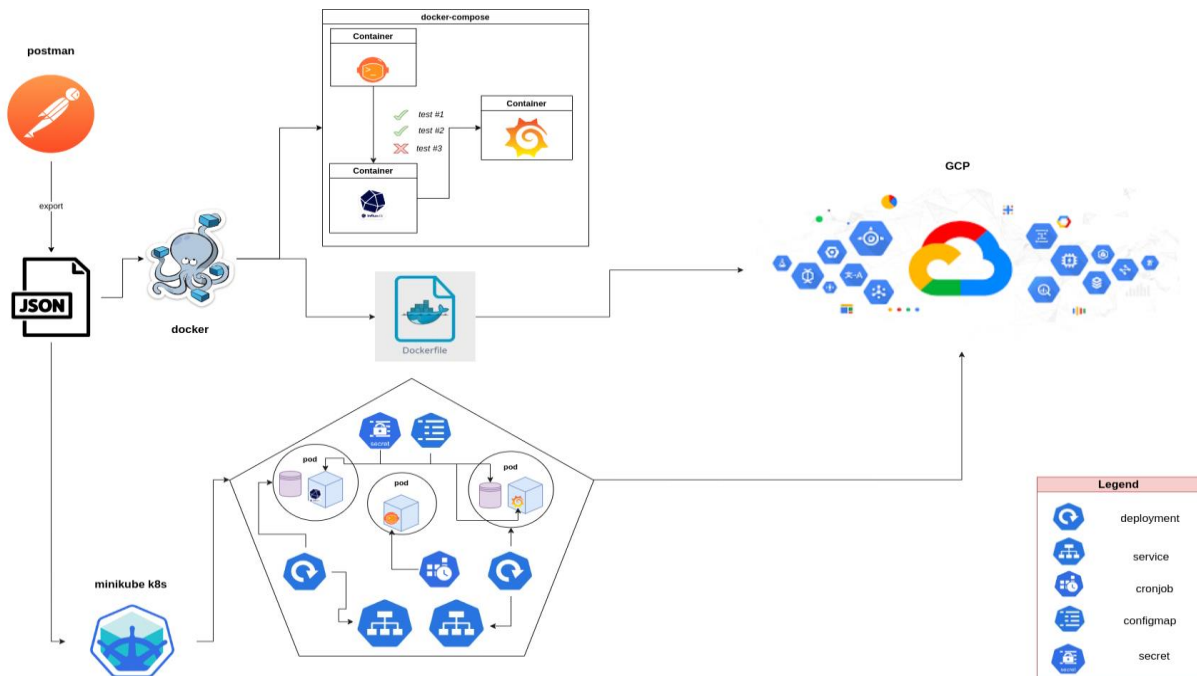
Figure 1: Components of an automated API testing system using containers and cloud computing.

The innovative system comprises many elements to achieve the goals for automation API testing in the cloud, managed by container orchestration:

1) *Postman* - a tool for testing API access points and creating and documenting automated tests.
2) *Newman* - command line tool allowing collections to be executed directly from Postman.
3) *InfluxDB* – a time series database used to store time-stamped sub-data (e.g. the execution time of a single query).
4) *Grafana* – a tool for visualization and analysis of data stored in the InfluxDB database.
5) *Docker* – system automation component (using Docker images and containers for Newman, Grafana and InfluxDB).
6) *Kubernetes* – an open source container orchestration system for automating software deployment, scaling and management.
7) *Google Cloud Platform (GCP)* - cloud computing services available and managed by Google.

The backend application intended to be tested can be represented as a group of API access points. Postman tool enables the environment for creating and saving simple and complex HTTP/s requests, as well as for reading their responses. Each application module is represented in Postman as a collection of requests (tests) organized as multiple folders that correspond to specific API access points.

A general scenario for testing the functionality of an API usually consists of the following steps:

1) Create Object (POST).
2) Verify previously created object (GET by ID).
3) Update object (PUT).
4) Get a list of all existing objects and check if the newly created object is in the list (GET All).
5) Delete object (DELETE).
6) Get a list of all objects again and verify that the deleted object is no longer in the list (GET All, deleted object not present).

The test script associated with a request will be executed after the request has been sent and a response has arrived. These tests verify that the APIs work as expected, determine that integrations between services work reliably, and that new changes have not broken the existing functionality.

Newman is the next (command line) tool in the toolchain, used after the Postman collections are set-up and configured. It allows collections to be executed directly using scripting. Its extensibility enables easy integration with continuous integration servers, and at the same time providing a rich suite of performance customization options.

InfluxDB, a time series database, is the next tool of choice that enables storing the reports and obtained test results using the influxDB Newman report tool. This database enables storing detailed statistical information and results from each test run.

To visualize the gathered data, the system uses Grafana, an open-source platform used for visualization, monitoring and analysis of data from the Influx DB database. Grafana enables creating dashboards with panels, each representing specific metrics over a specific time frame, thus presenting visual interpretation of the test runs, and the occurring failed tests.

The entire process presented in Figure 2 allows a well-defined and mostly automated process for API access points testing, and enables a simple further analysis and diagnostics from the saved data.

Upon detecting errors by the automated system, final check needs to be done manually, with two possibilities:

1) Either there was a change in the API not reflected into the test (the test needs to be corrected-fixed).
2) Or, the error cannot be detected into the test itself, thus the API request is broken, and an actual bug in the code has been detected.
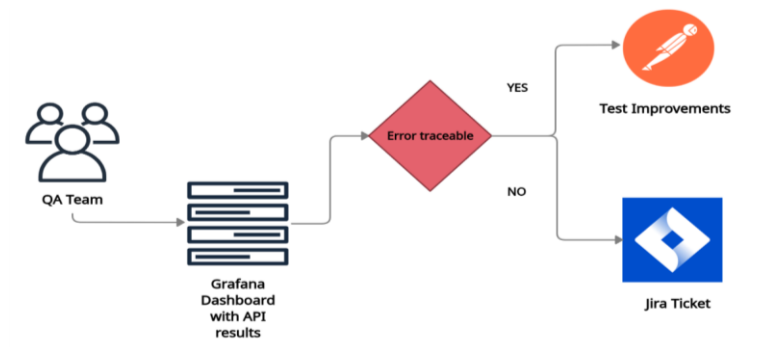
Figure 2: Analysis and diagnostics of test results.

4 CONTAINER TECHNOLOGIES FOR AUTOMATION TESTING

Today, containerization plays a crucial role in software development due to its efficiency, portability, and scalability. By using tools like Kubernetes, Docker, and GCP services, the testing workflow is streamlined, scalability is enhanced, and resource management is optimized. Figure 1 illustrates the core tools essential for automating testing processes and achieving efficient and reliable testing outcomes.

4.1 Containers for Automation Testing

Container technology presents an advantageous choice for integration within automation systems due to its inherent characteristics: lightweight design, minimal memory footprint, and efficient resource utilization, thus offering unparalleled portability and scalability. Furthermore, their isolation properties ensure secure deployment, and parallel coexistence on a single server without interference. The swift creation and destruction of containers, enhance operational speed, supporting dynamic workloads. Leveraging containerization facilitates seamless software deployment and management in cloud environments, abstracting applications from their underlying infrastructure.

Docker technology simplifies software delivery by providing a universal tool for creating, deploying, and running applications using containers, ensuring portability and consistency across different environments.

Four services are established using Docker container technology. The first service, InfluxDB's setup includes a Docker image, container name, and a port. Additionally a database and user are configured. However, the creation of the service alone does not facilitate storing of Newman results to the database, prompting the establishment of a second service to notify InfluxDB when it becomes operational. Similarly, Grafana is configured with its image, port, and container name.

The services defined with docker-compose need to be connected to enable data exchange and communication among themselves in order to provide a seamless system pipeline. Configuring Grafana to use InfluxDB data enables visual representation additionally facilitating analysis through preconfigured and modified dashboards and panels.

4.2 Orchestration of Container Services

In order to enable a more suitable usage of resources, and to simplify the management of the elements, the proposed systems encompasses the most widely used Kubernetes service orchestration tools. Kubernetes provides a highly elastic infrastructure with zero downtime deployment, automatic rollback, self-healing, and container scaling features, enabling seamless management of the entire lifecycle.

Before transitioning to cloud computing, the initial step involves setting up a local cluster. This is accomplished by utilizing Minikube, a lightweight implementation of Kubernetes. Compatible with all major OSs, Minikube aims to excel as a development tool for Kubernetes applications, ensuring comprehensive support for all Kubernetes features.

4.2.1 Establishing a Kubernetes Cluster

Establishing a Kubernetes cluster incorporating InfluxDB and Grafana involves several key steps. First, configuration files must be created for InfluxDB deployment, service, secret, and persistent volume claim. These files, specify essential details like the Docker image, exposed ports, and environment variables needed for database configuration. For instance, the deployment file outlines metadata and specifications, including the image source and container port, while the service file connects the service to the deployment through selectors. Additionally, a secret configuration file is essential for securely storing sensitive data, such as passwords, required for database configuration. This secret file uses Kubernetes secrets to securely store and inject the necessary data into the deployment as environment variables.

Moreover, enabling persistent storage for the InfluxDB database requires a persistent volume claim (PVC) configuration file. This file describes the type and details of the storage space required by InfluxDB, enabling Kubernetes to allocate or provision the appropriate volume to meet the database's storage needs. Setting up Grafana components follows a similar process to InfluxDB. ConfigMaps, serving a similar purpose to Secrets, are utilized to add configuration files to pod containers. Grafana requires three configuration files to be written to the running container. These files are added to a ConfigMap and mounted in different locations within the container for proper configuration.

This ample setup ensures proper functioning and configuration of both InfluxDB and Grafana, facilitating efficient data storage, management, and access within the Kubernetes environment.

4.2.2 Test Execution Schedule

The Kubernetes CronJob acts like a traditional cron utility, scheduling tasks within containers. Its purpose in the system is orchestrating API collection tests with Newman and sending results to InfluxDB. By setting a daily schedule, Kubernetes automates these tests.

As Minikube clusters lack access to local images, image construction takes place directly within the cluster using Minikube based command. Following image construction, the CronJob specification is created, including a jobTemplate that outlines the executing task. The schedule parameter dictates the recurring execution of tasks, ensuring periodic test runs.

After each run, the associated pod terminates until the next scheduled execution, allowing test results to be displayed in Grafana within a 24-hour window.

Although Minikube deployments do not inherently assign external IP addresses, accessing the Grafana service is made feasible via Minikube command, which seamlessly integrates with InfluxDB and automatically presents test results for analysis.

5 CLOUD IMPLEMENTATION

Introducing our system's move to the cloud involves tapping into the powerful advantages of cloud technology for refining our testing infrastructure. By transitioning to the cloud, scalability, reliability, and resource management are optimized. Cloud technology enables flexible scaling of testing resources, ensuring top-notch performance during peak demand. Furthermore, cloud platforms provide robust infrastructure, built-in security, and global accessibility, bolstering the resilience and efficiency of our testing environment.

5.1 Choosing Google Cloud Platform

Google Cloud Platform (GCP) is selected as the platform for several reasons. Firstly, it offers a comprehensive suite of cloud computing services integrated with Google's global network infrastructure. Additionally, GCP seamlessly integrates with the existing container technologies ecosystem, simplifying Docker and Kubernetes management. Its

commitment to innovation, reliability, and security aligns with the system's requirements, making it an ideal choice for the cloud implementation for the automated testing system deployment.

5.2 Kubernetes as a Service

Kubernetes as a Service (KaaS) provides a streamlined approach to managing Kubernetes clusters in the public cloud, with Google Kubernetes Engine (GKE) being a prominent solution within GCP. The setup of a Kubernetes cluster on GCP is initiated through the user-friendly interface of the Google Cloud console's Kubernetes Engine section.

Utilizing Cloud Shell, a versatile tool provided by GCP, a seamless connection is established between the Google Cloud console and the Kubernetes cluster. This connection is pivotal for effectively managing and overseeing the cluster's operations, including deployment, monitoring, and maintenance tasks.

To deploy containerized applications on GCP, we use Google Container Registry (GCR). Storing Docker images in GCR ensures they are accessible and available across different platforms and environments.

With the foundational setup in place, the necessary components for applications are created and managed using familiar Kubernetes configuration files. These configurations are applied through standard procedures, orchestrating the deployment and scaling of applications on the Kubernetes cluster within GCP.

The seamless integration of Kubernetes with GCP empowers users to harness the scalability, reliability, and flexibility of cloud-native technologies for their applications and services. Through intuitive interfaces and robust infrastructure, GCP simplifies the complexities of managing Kubernetes clusters, enabling efficient and effective cloud operations.

6 PERFORMANCE CHARACTERISTICS OF THE KUBERNETES CLUSTER SYSTEM DEPLOYMENT

Finally, exploring the performance characteristics of the final Kubernetes cluster deployment for the system is crucial for ensuring the reliability and efficiency of the automated testing processes. Monitoring and analyzing the performance of the Kubernetes cluster within the system plays a pivotal role in optimizing resource utilization, identifying

potential bottlenecks, and enhancing overall performance. Through the examination of key metrics and the implementation of monitoring solutions, the objective is to achieve optimal performance and reliability.

6.1 Kubernetes Cluster Monitoring

Operating a Kubernetes cluster introduces complexities due to the distributed nature of its components and the vast array of available metrics. One crucial aspect of monitoring Kubernetes clusters is tracking the state of objects within the cluster, including deployments, nodes, and pods. kube-state-metrics (KSM) addresses this need by providing a simple, yet effective, mechanism for generating metrics directly from the Kubernetes server API.

KSM offers insights into the current state of the Kubernetes objects, enabling a detailed assessment of the health and performance of the cluster. Unlike traditional monitoring tools focusing on individual components, KSM collects various metrics related to the state and resource usage of objects within the cluster: the number of running pods, the available CPU and memory resources, and the overall health status. These metrics are crucial for delivering the operational efficiency and performance of the cluster.

Figure 3: Kubernetes cluster health state.

Deploying KSM within the Kubernetes cluster involves creating essential Kubernetes objects, such as Service accounts, Cluster Roles, and Cluster Role Bindings, along with the kube-state-metrics deployment itself. These objects are crucial for enabling KSM to access and monitor Kubernetes API objects seamlessly. KSM exposes metrics via the HTTP /metrics endpoint, providing real-time visibility into the cluster's state. By analyzing these metrics, valuable insights into anomalies, resource utilization,

and cluster performance can be gained as illustrated in Figure 3.

6.2 Cloud Environments Log Management

The ELK Stack, comprising Elasticsearch, Logstash, and Kibana tools, plays an important role in log management within cloud-based environments. It offers a centralized way for tracking and analyzing various issues across system infrastructure, including performance monitoring and node failure detection.

Elasticsearch, serving as the storage engine, efficiently stores and retrieves log data. Logstash handles log delivery, processing, and storage, ensuring seamless data handling. Kibana, the visualization tool, provides an intuitive interface for visualizing log data and conducting advanced data analysis.

Deploying ELK Stack involves creating essential Kubernetes objects and resources: Elasticsearch clusters and Logstash configurations. These components aggregate, process, and visualize log data, helping effective monitoring and analysis. Kibana completes the ELK Stack with visualization features, empowering users to gain insights into system performance and troubleshoot issues efficiently.

6.3 Results from Kubernetes Monitoring

In the deployed system, the Elastic Stack serves as a fundamental component for Kubernetes monitoring, offering essential tools such as Filebeat and Metricbeat to collect monitoring data. These lightweight agents, deployed as Daemon Sets in Kubernetes, capture both system and application-level metrics and logs. Filebeat is responsible for collecting logs from pods, containers, and applications running on Kubernetes.

Filebeat dynamically detects components within pods and applies logging modules accordingly, thus providing real-time access to log data, enabling efficient log analysis and troubleshooting. Metricbeat, on the other hand, collects and preprocesses system and service metrics, including CPU, memory, disk, and network data. Deployed on each node in the cluster, Metricbeat gathers metrics from the Kubelet API, offering insights into the state of nodes, pods, containers, and other Kubernetes resources. Additionally, Metricbeat accesses cluster-wide metrics directly from the kube-state-metrics service.

Deployment of Filebeat and Metricbeat in our Kubernetes cluster involves configuring YAML files provided by Elastic. These files define deployment settings and specify connection details, ensuring seamless integration with existing Elasticsearch deployments and Kubernetes environments. Connecting Filebeat to Elasticsearch entails configuring index patterns to define indexed data for efficient retrieval and visualization in Kibana. Filebeat utilizes predefined index patterns to seamlessly transmit data to an existing Elasticsearch deployment, establishing direct integration. Address specification and TLS certificate inclusion in the Filebeat configuration ensure secure communication with Elasticsearch. For effective container log collection, Filebeat instances require access to the local log path mounted by the host, enabling comprehensive log data collection from Kubernetes pods, containers, and applications.

Once deployed and configured, Filebeat's integration with Elasticsearch and Kibana facilitates efficient log analysis and troubleshooting, empowering users with valuable insights into system performance and health. Accessing log data collected by Filebeat is straightforward, as it automatically becomes available for exploration in Kibana's Logs application, as demonstrated in Figure 4.
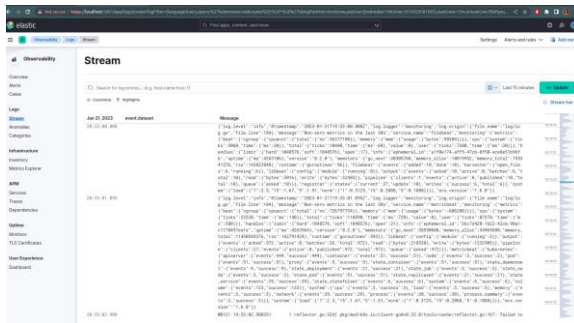
Figure 4: Collecting Kubernetes logs using filebeat.

In the Kibana Logs Stream captured from the Kubernetes cluster logs, we observe a real-time display of log events, including timestamps, log messages, and associated metadata. These may include error messages related to failed deployments, application crashes, resource constraints, network issues, security breaches, and other operational challenges. Additionally, anomalies such as sudden spikes or drops in log activity, unusual patterns in resource consumption, or unexpected behavior in application logs may indicate underlying issues requiring investigation and remediation.

Similarly, Figure 5 illustrates the performance and health metrics collected by Metricbeat, showcased in Kibana's Observability > Metrics

section, providing an overview of the containers and pods within the Kubernetes environment.
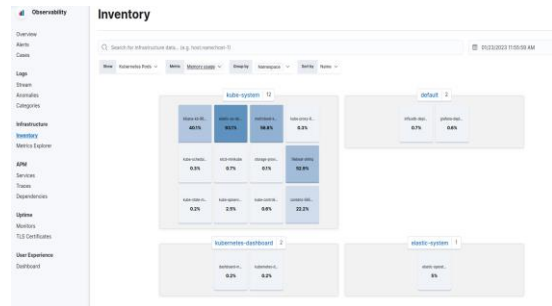
Figure 5: Overview of pods and containers.

The configuration of Filebeat and Metricbeat deployments within our Kubernetes cluster allows for the direct visualization of our Kubernetes resources. The use of ready-made panels is highlighted, which inherently pull information about our configured Kubernetes resources, as shown in Figure 6.
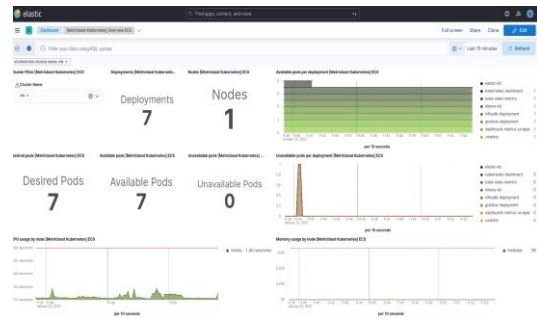
Figure 6: Kubernetes cluster overview.

Utilizing pre-built Kibana dashboards provided by Metricbeat, we gain insights into various aspects of our Kubernetes environment, including node, deployment, and pod overviews. These dashboards, shown in Figure 6, offer valuable insights into the performance and health of our Kubernetes cluster, enabling us to identify potential issues, optimize resource utilization, and ensure the reliability and efficiency of our system.

For instance, if the dashboard indicates a sudden increase in CPU usage across multiple pods, we promptly investigate potential resource bottlenecks and scale up affected pods to ensure optimal performance. Similarly, consistent patterns of pod failures within a deployment prompt analysis of root causes, adjustments to configurations, or implementtation of auto-recovery mechanisms to enhance system reliability. This approach, facilitated by comprehensive insights from Metricbeat dashboards,

allows us to effectively address issues, optimize resource allocation, and sustain overall cluster efficiency.

7 CONCLUSIONS

This research paper explores the development of an automated API testing system with a focus on real-time monitoring and visualization. By integrating containerization tools like Docker and Kubernetes, the study establishes a cohesive testing framework for efficiency and scalability. Using Google Cloud Platform (GCP) further enhances the system's scalability and performance. Additionally, the integration of monitoring tools, particularly Elasticsearch, enables the assessment and visualization of the health and performance of the Kubernetes cluster underlying the testing environment.

By adopting a continuous and automated approach, the research successfully orchestrates these technologies to create a robust and efficient system for daily automation testing. The results highlight the collective impact of these integrated technologies in achieving reliable and effective automated testing processes, ultimately contributing to the advancement of software development.

REFERENCES

[1] P. Agrawal and N. Rawat, "Devops, 'A New Approach To Cloud Development & Testing'," Proceedings of the 2019 Int. Conf. on Issues and Challenges in Intelligent Computing Techniques (ICICT), India, 27-28 Sep. 2019, pp. 1-4.

[2] P.P. Kore, M.J. Lohar, M.T. Surve, and S. Jadhav, "API Testing Using Postman Tool," Int. Journal for Research in Applied Science & Engineering Tech. (IJRASET), 2022, doi: 10.22214/ijraset.2022.48030.

[3] Dh.K. Sharma, "Security Testing of API using Postman and Swagger tools and its use in Internet of Things (IOT)," Journal of Emerging Technologies and Innovative Research, Feb. 2019, vol. 6, no. 2.

[4] J. Shah and D. Dubaria, "Building Modern Clouds: Using Docker, Kubernetes & Google Cloud Platform," IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), 2019, doi: 10.1109/ccwc.2019.8666479.

[5] S. Kumar and C. Saravanan, "A Comprehensive study on Data Visualization tool - Grafana," Journal of Emerging Technologies and Innovative Research, ISSN:2349-5162, vol.8, no. 5, page no.f908-f914, May-2021.

[6] N. Kathare, O.V. Reddy, and V. Prabhu, "A Comprehensive Study of Elasticsearch," International Journal of Science and Research (IJSR), vol. 10, no. 6, June 2021, doi: 10.21275/SR21529233126.

[7] P. Shenoy, S.V. Soudri, R. Kumar, and S. Bailuguttu, "Enhancement of observability using Kubernetes operator," Indonesian Journal of Electrical Engineering and Computer Science, 2022, doi: 10.11591/ijeecs.v25.i1, pp. 496-503.

[8] S. Kaiser, M.S. Haq, A. Tosun, and T. Korkmaz, "Container technologies for ARM architecture: a comprehensive survey of the state-of-the-art," IEEE Access, 2022, doi: 10.1109/ACCESS.2022.3197151.

[9] G. Ambrosino, G.B. Fioccola, R. Canonico, and G. Ventre, "Container mapping and its impact on performance in containerized Cloud environments," IEEE Int. Conf. on Service Oriented Systems Engineering (SOSE), 2020, doi: 10.1109/SOSE49046.2020.00014.

[10] S. Garg and S. Garg, "Automated Cloud Infrastructure, Continuous Integration and Continuous Delivery using Docker with Robust Container Security," Conference: 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2019, doi: 10.1109/MIPR.2019.00094.

[11] A.M. Potdar, D.G. Narayan, S. Kengond, and M.M. Mulla, "Performance Evaluation of Docker Container and Virtual Machine," Third International Conference on Computing and Network Communications, 2020, doi: 10.1016/j.procs.2020.04.152.

Modular Robotic Reinforcement Learning Platform for Object Manipulation

Vishnudev Kurumbaparambil, Subashkumar Rajanayagam and Stefan Twieg

Department of Electrical, Mechanical and Industrial Engineering, Anhalt University of Applied Sciences,
Bernburger Str. 55, 06366 Köthen, Germany
vishnudev.kurumbaparambil@student.hs-anhalt.de, subashkumar.rajanayagam@hs-anhalt.de, stefan.twieg@hs-anhalt.de

Keywords: Reinforcement Learning, Robot Arm, Object Manipulation, OpenAI Gym, Stable Baselines, Robot Operating System.

Abstract: The field of robotics and autonomous systems has witnessed significant advancements in recent years, with an increasing focus on enhancing the capabilities of robotic agents through RL. This project centres around applying Reinforcement Learning techniques to do object manipulation tasks with a specific focus on making a robot arm reach a target object. Using a combination of Gazebo and Robot Operating System (ROS) environments, the robot arm is trained using custom OpenAI Gym environments to simulate the task. The primary objective involves positioning the end-effector of the robot arm close to a designated object and overcoming challenges such as self-collisions during movements. Various iterations of RL training, including different reward logics, curriculum learning approaches, and fine-tuning parameters, are explored to refine the decision-making capabilities of the agent. The training process of Curriculum Learning involves a phased approach, starting with basic movements and progressing to more complex tasks, demonstrating improved performance. However, challenges such as prolonged training times and uncertainties in arm behaviour persist. The project highlights the complexities inherent in designing effective RL strategies for robotic systems and stresses the need for further research to enhance computational efficiency and reliability in real-world applications.

1 INTRODUCTION

1.1 Motivation

Over the years, machines have played a significant role in reducing human workload. With technological advancements like artificial intelligence, machines have become more efficient and intelligent. Exploring the broad spectrum of applications, robots, especially robotic arms, hold immense potential in transforming various industries. This project aims to leverage Reinforcement Learning (RL) fundamentally based on [1], a machine learning technique facilitating experiential learning for robots, to enhance object manipulation capabilities, unlocking diverse possibilities for applying robotic arms across industries.

1.2 Problem Statement

The effectiveness of robotic systems often relies on their ability to perform tasks with precision [2]. This project aims to teach a robot arm to reach for an object using RL. The primary objective involves positioning the end-effector of the robot arm close to a designated object and overcoming challenges such as self-collisions during movements. The idea is to start with a simpler goal, like reaching the object, and later move to more complex tasks, like grabbing. The model's performance can be initially refined through simulation. In the future, this can be applied to a real robot arm. Continuous adjustments and fine-tuning based on simulation outcomes are crucial for achieving the best results in practical situations. This approach leads to a robot system that can apply knowledge because it already learned how to do it (in a virtual environment/ OpenAI Gym) [3].

2 LITERATURE REVIEW

2.1 Reinforcement Learning

In simple terms, RL is the study of agents and how they learn from their interaction with the environment to fulfil desired goals by trial and error. As shown in Figure 1, based on the action (a_t) and reward (r_t) hypothesis, calculated by comparing the desired goal and observed state (S_t), facilitates learning the desired outcome. The Environment is the world where the agent resides and interacts. After each interaction, the agent gets information about the environment's new State. Based on this new state, it decides its next step. A Reward is a signal that an agent gets from the environment which tells how good or bad the state is and successful the goal is [4]. The goal of the agent is to maximize these rewards called "return". So, using RL methods the agent can learn to achieve its goal [5]. In our case, the robot arm (Agent) will decide on its next best joint angles (Action a_t) to reach a desired end effector location based on Reward r_t, which is calculated using the distance between the previous end effector location (State S_t) and the desired location. This is repeated until a satisfactory level of rewards is achieved which corresponds to a high success rate or accuracy of the robot arm reaching the desired end effector location.
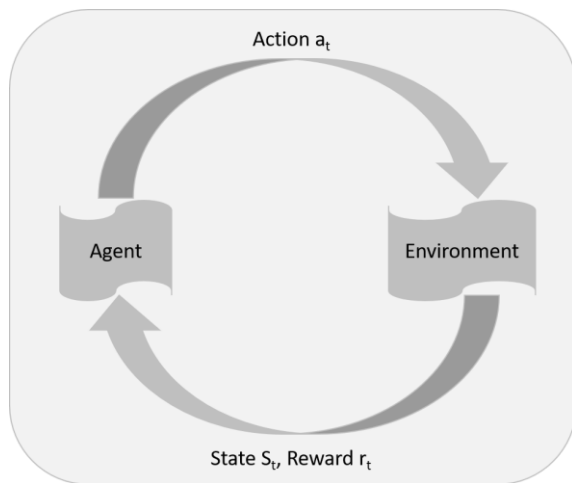
Figure 1: RL Agent-Environment interaction loop.

Mapping the states to the actions are called policies. These policies optimize themselves by choosing actions that maximize the rewards. These policies are categorized as value, policy, and model-based algorithms. In value-based algorithms such as Deep Q-Networks (DQN), policies are derived indirectly by learning the value function [6]. In policy-based algorithms such as Proximal Policy Optimization (PPO), the policies are directly improved [7]. While model-based algorithms such as Dyna-Q focus on learning the model of the environmental dynamics and thereby being able to learn policy functions [8].

2.2 Robot Arm Manipulation

A few research studies have been reviewed. It gave insight into different reinforcement methods. Open-World Object Manipulation using Pre-Trained Vision-Language Models [9]: The researchers utilize pre-trained vision-language models to extract object-related details from the textual instructions and images. By leveraging this information, the robot becomes capable of identifying unknown objects and performing the required task, such as picking up the object. A Survey on Deep RL Algorithms for Robotic Manipulation [10]: This review provides an overview of recent advancements in deep RL for robotic manipulation tasks. Review papers [11-16] provided applications in environmental perception, path planning, behaviour decision and path planning. This provided a good idea of the potential of RL when applied appropriately specific to the domain. In addition, it provided a good overview of the available state-of-the-art RL algorithms.

2.3 Limitation

Implementation based on RL in robotics is mostly implemented on simulation and then transferred to the real hardware via transfer learning. In simulation, the environment can be reset easily, but on real hardware, human intervention is necessary to reset the environment [17]. Utilizing RL algorithms also has the potential to learn much faster and more efficiently in solving real-world problems [18, 19]

However, to address these limitations, it is important to develop a robotics RL platform that provides flexibility and appropriate evaluation criteria, with which integration and experimentation with available state-of-the-art or new novel algorithms with various robots can be done. During the literature review, we only identified platforms that are general to reinforcement learning or specific to certain robots.

3 DEVELOPMENT PLATFORM

After the literature review, it became evident that the ROS is the primary choice for robot control. When it comes to RL, OpenAI Gym [20] is a widely adopted framework. However, OpenAI Gym is designed for general RL and lacks specific features required for robotics applications. Bridging the gap between these two systems posed a significant challenge. The final goal is to combine all these, creating a stable training environment for our RL agent.
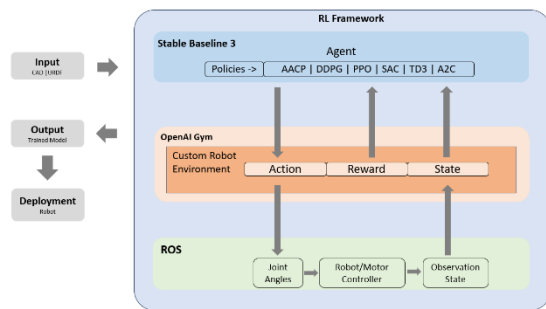
Figure 2: Platform architecture for robotics.

Figure 2, shows the overview and interconnection of the developed robotic RL platform.

3.1 Integrating OpenAI Gym and ROS

The initial step involves the comprehensive analysis of translating the problem statement into Python code while considering the integration of Gym and ROS. The logical step is to develop a custom environment class using OpenAI Gym's APIs for path planning, employing RL to guide the robot arm to desired locations. Within this environment, integrate ROS to manage the arm's initialization, control, and feedback acquisition. The computation of rewards is based on the feedback generated by various ROS nodes, which are continually publishing data. The core framework for the reward system, data gathering from sensors, and state management are encapsulated within the step function of this custom OpenAI Gym environment. Additionally, the procedures to be executed at the end of each episode, such as resetting variables and the robot's state, are delineated in the reset function of the custom OpenAI Gym environment.

Integrating OpenAI Gym and ROS provides a robust solution for robot arm manipulation, combining their strengths to develop and train RL agents for precise and effective movements.

3.2 Designing a Custom OpenAI Gym Environment

OpenAI Gym package provides templates for creating a custom environment which should mirror the real-world application. Understanding the real world provides insights into the factors influencing the model and this can define how the agent can predict the actions that can be taken by the model. Then, we need to calculate rewards based on how the environment changes due to the agent's actions. ROS APIs can be used to read the environment state like the robot's joint positions. The reward calculation is crucial because it tells the agent if it's actions are good or bad. The reward system inherently steers the learning process of the agent, shaping its understanding of the optimal course of action in pursuit of the defined objectives.

To create a custom OpenAI Gym environment, we need to define several key functions required by the class template. These functions include:

1) **__init__**: This function initializes the environment and sets up any necessary parameters.
2) **step**: It defines what happens during each step in the environment. It calculates the next state, reward, and whether the episode is done.
3) **reset**: This function resets the environment to its initial state at the beginning of each episode.
4) **render**: If necessary, this function can be used to visualize the environment.

3.3 Robot Arm Models

During the initial phase, a simple robot arm model was custom-designed using the Unified Robot Description Format (URDF) [21]. This model consisted of three cylinders stacked on top of each other, with two movable joints of type 'revolute'. In this early stage, our primary goal was to control these joints using ROS from Python, and subsequently and subsequently integrating it with OpenAI Gym to utilize RL.

The initial ROS and OpenAI Gym framework was established using this basic three-cylinder model shown on the left side of Figure 3. A CAD model can be converted to URDF format using the "sw_urdf_exporter" plugin in SolidWorks. Our robot arm features five revolute joints for arm movement and two prismatic joints for the end effector, as shown on the right side of Figure 3.
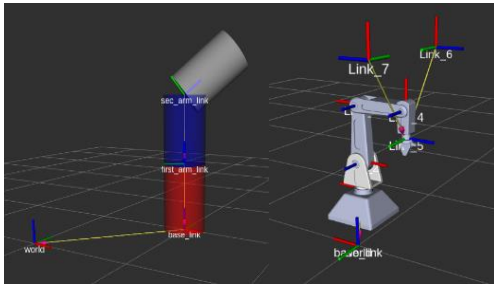
Figure 3: Initial simple cylindrical model (left) and CAD to URDF model (right).
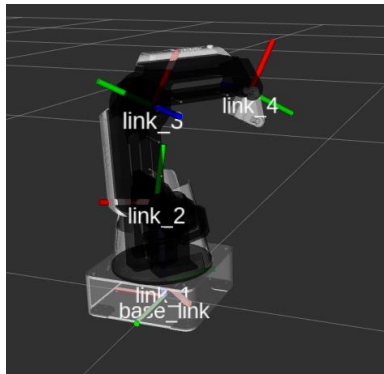
Figure 4: Dobot Magician robot arm.

The next step was get insights into the performance difference between the simulation and the real world. For this purpose, we chose Dobot Magician robotic arm available at our laboratory and its URDF model was obtained from the manufacturer's website as shown in Figure 4. This URDF model would be used for further training and experimentation. The model featured four revolute joints for arm movement, with no end effector. One notable issue encountered with the URDF of the Dobot Magician is its lack of self-collision constraints, that's why we implemented "selfCollision" tags and a contact sensor, working together to detect self-collisions.

4 TRAINING OF THE RL AGENT

During the training process, we aim to refine the RL Agent's performance. This involves adjusting simulations and enhancing the reward logic, all to improve the agent's decision-making capabilities.

4.1 Varying Simulation Parameters

In our simulation setup, the parameters `<max_step_size>`, `<real_time_factor>`, and `<real_time_update_rate>` in the <physics> tag of world file play crucial roles in controlling the dynamics of the simulation. The `<max_step_size>` determines the maximum time step for simulation updates, influencing the trade-off between accuracy and speed. The `<real_time_factor>` sets the ratio of simulation time to real-world time, allowing for control over the simulation's temporal behaviour. Finally, `<real_time_update_rate>` dictates how frequently the simulation updates in real-time, influencing responsiveness and computational load. Fine-tuning these parameters is essential to achieve the desired balance between simulation fidelity and computational efficiency in our experimentation.

The optimal simulation rate parameters are found out to be 0.01 for MSS and 1000 for real_time_update_rate.

4.2 Varying Environment Parameters

4.2.1 Scaling Episode Iteration Steps

For each training episode, the maximum number of tries or episode length can be configured.

Table 1: Training results after varying the Episode Length.

Episode Length	TT (s)	C/E/BF	AS/ AU	ELM	ERM	G R
50	1860	3337/0/103	930/5415	44.3	-901	0
75	1859	4222/0/111	724/4776	51.3	-1440	0
100	2082	5072/0/134	646/4001	48.5	-1750	1

Table 1 consists of TT (Time Taken), Collisions/Errors/Bottle Fell (C/E/BF), Episode Length Mean (ELM), Episode Reward Mean (ERM) and Goal Reached (GR) of the experiment. Based on the obtained results, setting the episode length to no more than 50 yields better outcomes in terms of performance and stability. This is seen in Table 1 at the BF for the Episode Length of 50.

4.2.2 Observation Space

Exploring an alternative approach is to expand the observation space by adding the positions of both the object and the robot arm end effector. While it remains experimental, this adjustment aims to evaluate the impact on the learning process and determine if a more comprehensive observation space enhances the agent's understanding of the environment.

Table 2: Training results with different Observation Space (OS). D – Distance, OP – Object Distance and GP – Gripper Distance.

OS	C/E/BF	AS/ AU	ERM	GR
D	3337/0/103	930/5415	-901	0
D + OP	2398/0/94	934/6375	-818	1
D + OP + GP	2787/0/111	953/5942	-876	2

Table 2 consists of Collisions/Errors/Bottle Fell (C/E/BF), Action Success/ Action Unsuccess (AS/AU), Episode Reward Mean (ERM), and Goal Reached (GR) of this experiment. As can be seen in above Table 2, the goal is reached when the OS is more comprehensive and hence is the better approach.

4.3 Varying Reward Logic

4.3.1 Reward Logic v1

In defining the reward strategy, the central objective is to guide the robot arm toward the object while steering clear of collisions. The proposed approach involves assigning an exponentially increasing penalty for consecutive collisions, focusing on the impact of repeated impacts. Conversely, as the robot arm approaches the object, the reward scales proportionally, underlining the importance of proximity. To provide a stronger incentive, the reward for nearing the object is intentionally set slightly higher. Importantly, a distinct reward is introduced for transitioning from a collision state to a favourable movement state, emphasizing the priority of avoiding collisions over mere distance alterations. The reward system guides the agent to minimize the distance to the object while avoiding collisions effectively. Adjusting these reward components will be crucial for shaping the agent's behaviour for optimal task performance. Table 3 below illustrates the results after training, where the Iteration Per Episode (IPE) is assigned in the first column.

Table 3: Training results of the Reward Logic v1.

IPE	TT (s)	C/E (%)	AS/ AU	ELM	ERM	GR
50k	8953	0/2.45	11.93/5 2.29%	28.8	-414	66
50k	8635	0/2.21	11.57/5 5.86%	20.4	-89600	117

The obtained results fell short of perfection, revealing a notable issue in the potential for rewards to exhibit significant fluctuations, particularly with very high negative values attributed to penalties for self-collisions. The learning process of the agent lacks stability, evident in the mean reward that varies sporadically between negative hundreds and occasional spikes into the thousands.
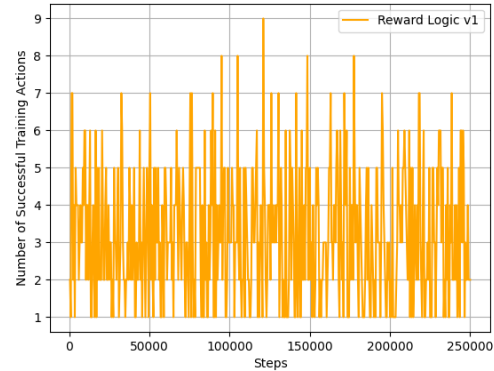
Figure 5: Number of Successful Training Actions with Reward Logic v1.

Furthermore, the number of actions demonstrating improvement has not shown any significant enhancement as seen in Figure 5. Looking at the average reward in Figure 6, we see that it often goes up and down a lot. This is mainly because of the harsh penalty for consecutive collisions. While this penalty is needed, we should adjust the rewards to avoid extremely negative values. This should help stabilize the learning process.
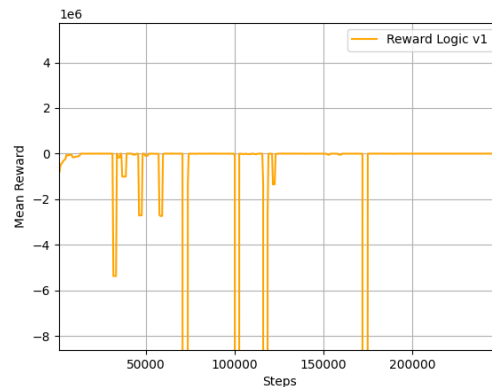
Figure 6: Reward Mean with Reward Logic v1.

4.3.2 Reward Logic v2

The updated logic adopts a human-centric approach, aligning with how a person would be motivated and penalized in reaching the object. This strategy aims to

create a more realistic and balanced learning experience. The key considerations for this version reward system are as follows:

1) Max Steps and Object Reach Reward: The highest reward should occur when the robot reaches the object in the fewest steps, with diminishing rewards as the step count increases. Even if the maximum steps are taken, a positive but reduced reward should encourage continued attempts.

2) Collision Penalties: The penalty for collisions should be set in a way that, while giving higher penalties for discouraging continuous collisions, it does not plunge into extremely negative values. This ensures that the agent perceives the task as challenging but not impossible.

3) Safe Movement Reward: Transitioning from a collision to a safe movement should be positively rewarded. However, this reward should not be too high to discourage frequent switching between collision and safe movements.

4) Distance-Based Rewards and Penalties: Penalties for moving away from the object and rewards for proximity should be balanced to motivate the agent without overly emphasizing proximity.

5) Balanced Reward Values: Reward values for positive and negative actions must be finely tuned to motivate goal achievement while discouraging collisions and moving away from the object. Striking a balance ensures that the agent remains engaged and does not lean towards inactivity due to disproportionately high penalties.

This approach mirrors human motivation and helps stabilize the learning process for the agent. Below Table 4 shows the comparison among versions.

Table 4: Training results with Reward Logic v2.

Version	TT (s)	C/E/BF	AS/AU	ELM	GR
v1	8953	15465/0/1223	5965/26143	28.8	66
v2	10319	17807/0/2004	6907/22086	10.6	129

Upon comparing the plots of the two versions in Figure 7, a significant improvement is observed in the number of successful episodes where the robot arm successfully reached the target.
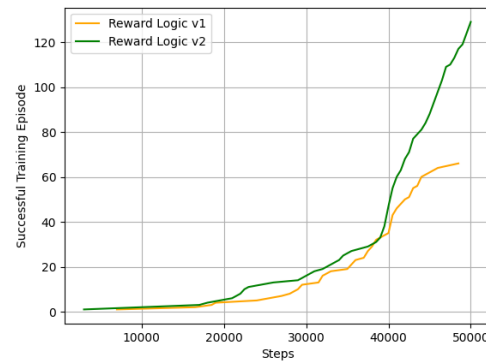
Figure 7: Successful Training Episodes with Reward Logic v2.

Other parameters such as the number of collisions and successful training steps don't show a significant improvement in the observed results as seen in Figure 8 below.
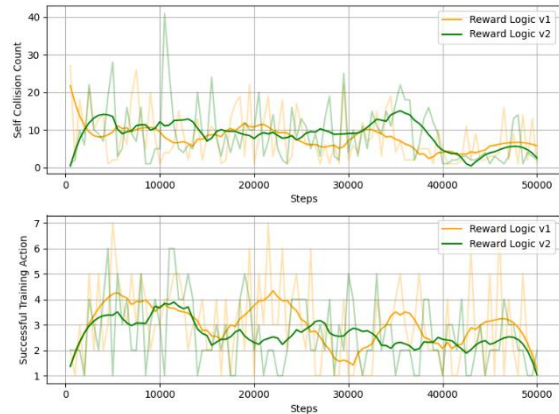
Figure 8: Reward Logic v2 indicating no improvement in Collisions and Successful Training Action.

As seen in Figure 9, the rewards appear to be within an acceptable range, steadily increasing with no irregular negative peaks.
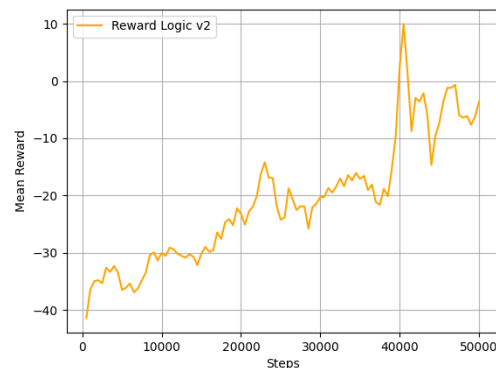
Figure 9: Reward Mean with the Reward Logic v2.

To gain further insights, we conducted testing for 250,000 steps in both versions, and the results are as follows in Table 5.

Table 5: Training results with Reward Logic v2 (250k IPE steps).

Ver	TT (s)	C/E/BF (%)	AS/AU (%)	ELM	GR
v1	44111	20.78/0/3.01	12.99/59.03	-168000	1003
v2	47154	34.69/0/2.91	12.53/46.69	-11.2	497

Surprisingly, the improvement in the number of successful episodes is not as significant as in version 1. Additionally, the number of collisions continues to increase. Analyzing the mean reward plot in Figure 10 reveals prominent negative peaks at multiple points. This problem emerged when the bottle was thrown far away, leading to an excessively high distance calculation and subsequently resulting in a very negative reward based on distance.
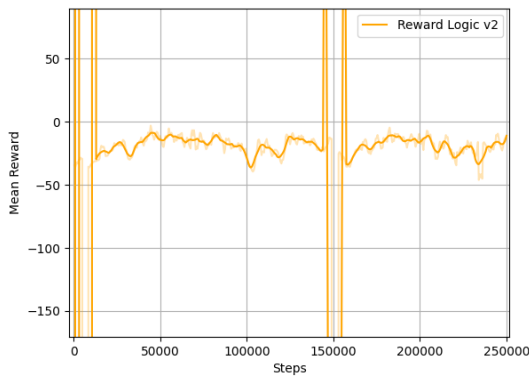
Figure 10: Reward Mean with Reward Logic v2 after running for 250k IPE steps.

4.3.3 Reward Logic v3

To address the excessive negative values of reward in v2, extra safeguards have been incorporated in this version. If collisions persist beyond a certain threshold, the training is terminated with a penalty. Additionally, a check has been implemented to ensure the bottle's proximity, accounting for scenarios of potential displacement due to rapid arm movements.

Upon analyzing Table 6, a decline in performance is seen. The current threshold for successive collisions is set at 5, potentially terminating many training episodes prematurely and limiting the agent's learning. In the next training iteration (version 3.1), the self-collision threshold will be increased to 10.

However, there is no apparent improvement, and the results even show a slight negative trend. The mean reward exhibits no occurrence of negative peaks, suggesting the success of safeguard mechanisms.

Table 6: Training results with Reward Logic v3.

Version	TT (s)	C/E/BF	AS/AU	ELM	GR
v1	8953	15465/0/1223	5965/26143	28.8	66
v2	10319	17807/0/2004	6907/22086	10.6	129
v3	9298	17376/0/746	6004/24022	26.6	38
v3.1	10472	19693/0/751	5563/22697	33.5	13

4.3.4 Reward Logic v4

Despite numerous iterations, most versions of the reward logic performed poorly compared to the initial one. To enhance the reward system, reward shaping was considered, in which rewards are based on the quality of action by evaluating deviations from ideal actions. Consequently, penalties are applied, providing the agent with feedback and influencing its behaviour. This strategy aims to guide the agent towards actions that align more closely with the desired outcomes, potentially accelerating the learning process and improving overall task performance. The collision threshold also has been updated to 25. The shaping reward based on the action is implemented specifically whenever the valid distance measured is not improved. The training has been executed for 50,000 steps, and the outcomes are presented in Table 7 below. The successful training episode is also shown in Figure 11.
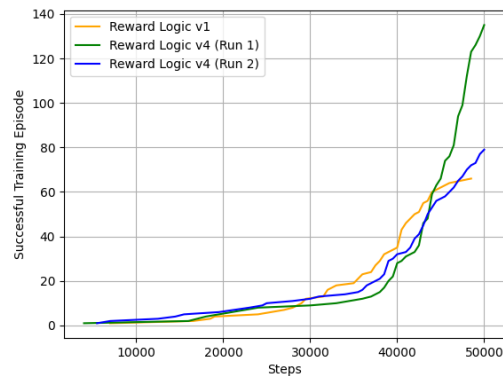
Figure 11: Successful Training Episode with the Reward Logic v4 (50k IPE steps).

There is also a reduction in the number of collisions, and successful actions show a slight improvement as seen in Figure 12.

Table 7: Training results with Reward Logic v4 after executing for 50k steps.

Version	TT (s)	C/E/BF	AS/AU	ELM	GR
v1	8953	15465/0/1223	5965/26143	28.8	66
v2	10319	17807/0/2004	6907/22086	10.6	129
v3	10472	19693/0/751	5563/22697	33.5	13
v4	9301	12640/0/1583	6782/27941	-	135
v4	8799	12967/0/1238	6266/28488	-192	79
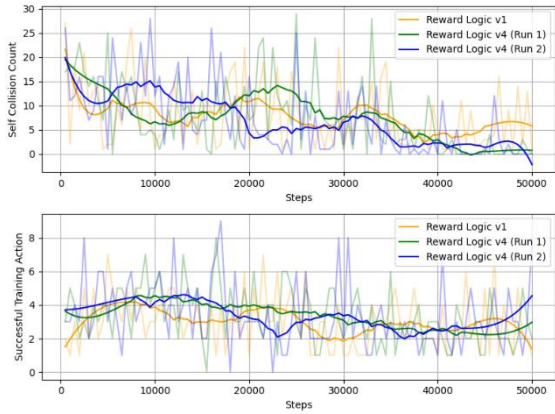
Figure 12: Plots of Self-Collision Count and Successful Training Action with Reward Logic v4.

Table 8: Training results of Training with Reward Logic v4 after running 250k IPE steps. Ver stands for Version.

Ver	TT (s)	C/E/BF (%)	AS/AU (%)	GR
v1	44111	20.78/0.0012/3.0	12.99/59.03	1003
v4	45692	21.92/0/3.25%	14.11/58.07	1069
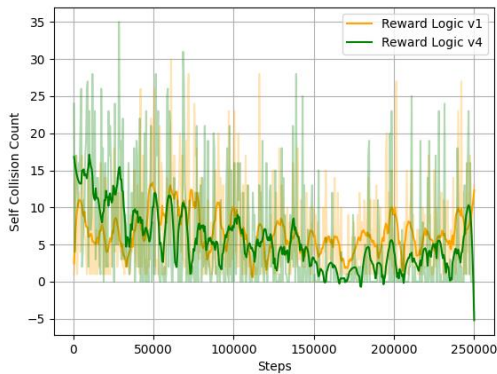
Figure 13: Self-Collision Count with Reward Logic v4 running for 250k IPE steps.

To delve deeper, additional testing was conducted for 250,000 steps and summarized in Table 8. A noticeable improvement in the number of successful episodes is evident, and there is a decreasing trend in the number of collisions toward the end of training as seen in Figure 13. Overall, this version exhibits slightly better performance than the initial reward logic and will be used for further testing.

4.4 Varying Policies

Stable Baselines provides a diverse range of RL policies such as Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradients (DDPG), Trust Region Policy Optimization (TRPO), and more. The applicability of each policy depends on the action and observation types specific to the environment. The possible policies for the current task are tried. These are Advantage Actor-Critic (A2C), Deep Deterministic Policy Gradient (DDPG), Proximal Policy Optimization (PPO), Soft Actor-Critic (SAC) and Twin Delayed DDPG (TD3).

4.4.1 Advantage Actor-Critic Policy

Up to this point, all training sessions have been conducted using this A2C policy. The results are summarized below in Table 9.

Table 9: Training results with A2C Policy.

TT (s)	C/E/BF (%)	AS/AU (%)	GR
8799 (50k steps)	25.93/0/2.48	12.53/56.98	79
45692 (250k steps)	21.92/0/3.25	14.11/58.07	1069

4.4.2 Policy Comparison

Results obtained with DDPG, PPO and SAC policies highlight a much lower number of successful episodes as can be seen in Table 10. This is also evident in Figure 14 when DDPG is compared with A2C.

Table 10: Training results with PPO Policy.

RL	TT (s)	C/E/BF (%)	AS/AU (%)	GR
DDPG	8995	16.97/0/1.73	11.48/67.74	16
PPO	8697	24.63/0/0.87	9.58/62.80	5
SAC	11884	15.51/0/1.09	11.26/67.44	4
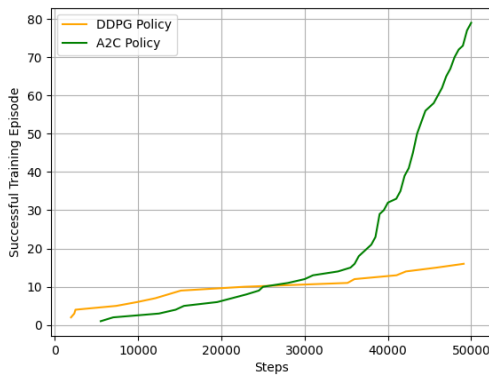
Figure 14: Comparison of Successful Training Episode between A2C and DDPG Policies.

4.4.3 Comparison of TD3 and A2C Policy

The TD3 policy has demonstrated a notable number of successful actions and achieved goals as shown in Table 11.

Table 11: Training results with TD3 Policy.

TT (s)	C/E/BF	AS/AU	GR
10274 (50k steps)	8476/0/1860 16.95/0/3.72%	7083/31565 14.17/63.13%	45
50951 (250k steps)	43062/0/10308 17.22/0/4.12%	45894/144635 18.36/57.86%	785
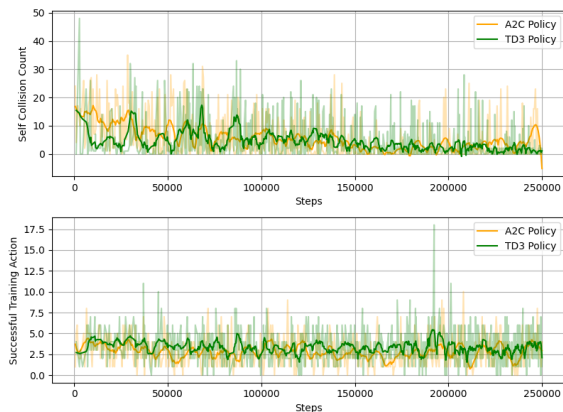
Figure 15: Plot of Self-Collisions Count and Successful Training Actions with A2C and TD3 Policies.

To further assess its performance, it was executed for 250,000 steps, and the results are detailed in Table 11. Upon analyzing the plots, it is evident that the TD3 Policy exhibits fewer collisions compared to the A2C policy. Additionally, the number of successful actions is higher as seen in Figure 15.

Although the A2C policy achieves more goals, the TD3 policy shows an exponential increase in successful goals towards the end of training, indicating a potential for superior performance as shown in Figure 16.
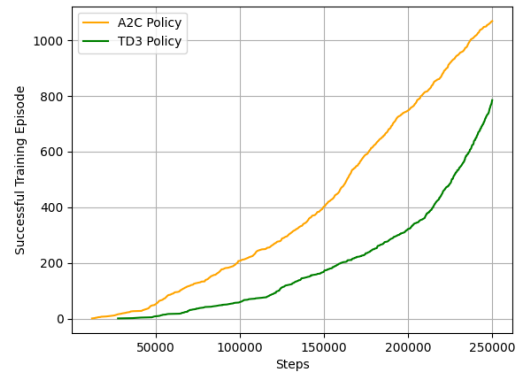
Figure 16: Plot of Successful Training Episodes with A2C and TD3 Policies.

Based on these findings, the decision has been made to proceed with the TD3 policy for subsequent testing.

4.5 Curriculum Learning

Curriculum learning is a training strategy in machine learning where the model is exposed to a series of tasks or goals of increasing complexity. The approach gradually introduces more challenging scenarios, mirroring human learning by starting with simpler concepts before advancing to more intricate ones. Curriculum learning can enhance the overall learning efficiency and generalization capabilities of a model, enabling it to better handle a diverse range of situations. An initial experiment involved breaking down the task into two stages. In the first stage, the robot was trained to reach a specific point instead of the bottle, overlooking the scenario where the bottle might fall. Following this, the model was loaded and retrained with the comprehensive goal of reaching the bottle.

In Figure 17 above, it is evident that the process of retraining the model is functioning effectively. The loaded model retains and builds upon previous training knowledge, resulting in improved performance during subsequent training sessions. Given the objective of having the robot arm reach the bottle, we encounter two primary challenges: collisions between the arm and the risk of the bottle toppling. To systematically address these challenges at different stages, the task is partitioned into three distinct goals, each trained separately:

- The primary objective is to guide the robot arm to move without collisions. Penalties are given exclusively for collisions, while all valid movements receive rewards, with greater weight assigned to getting closer to the fixed point.
- In this stage, additional penalties are introduced for movements that take the robot farther away from the fixed point. There is no condition check for the possibility of the bottle falling.
- The final stage involves the robot attempting to reach the bottle. Extra penalties are imposed if the bottle falls during this stage.

Figure 17: Training the agent in two stages.

Figures 18 and 19 below illustrate the successful episodes and self-collisions. Each stage undergoes training for a duration of 100,000 steps.
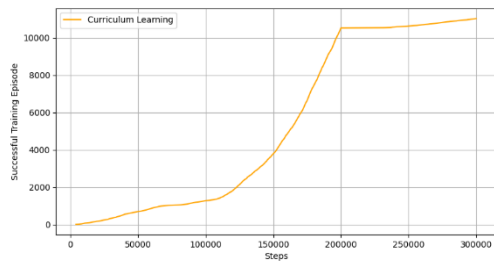
Figure 18: Plot of Successful Training Episodes with Curriculum Learning.

As seen in Figure 19, the total count of self-collisions has considerably decreased. The stage with the highest number of successful episodes is stage 2, where the objective is to move to a fixed point without considering the possibility of the bottle toppling over while penalizing collisions and movements away from the designated point. However, once it moves to stage 3, the rate of successful episodes is reduced significantly. But the total number of successful episodes is still good as seen in Table 12. Table 12

consists of Collisions/Errors/Bottle Fell (C/E/BF), Action Successful/ Action Unsuccessful (AS/AU) and Goal Reached (GR) of the experiment.
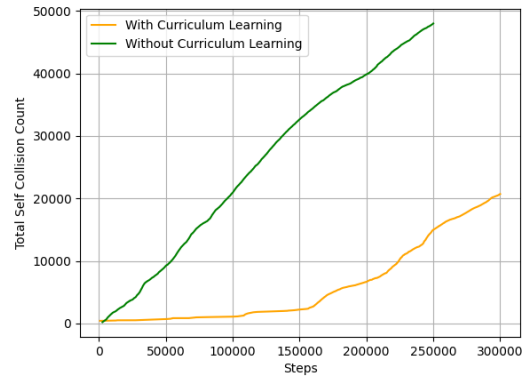
Figure 19: Plot of Total Self-Collision Count with Curriculum Learning.

Table 12: Training results with Curriculum Learning.

Stage	C/E/BF (%)	AS/AU (%)	GR
1	1.03/0/0	10.72/85.32	1275
2	5.42/0/0	26.74/57.69	9253
3	12.23/48.75	16.85/63.33	501

The understanding from this curriculum learning is that the robot can be trained initially for simple goals and then trained on top of this for sophisticated goals.

4.6 Results

In all experiments, total training time emerges as the key factor influencing performance. Regardless of policy and parameters, better results are consistently observed with increased training steps for both A2C and TD3 algorithms. The introduction of curriculum learning indicates potential for enhanced performance. The extended training duration for each stage facilitates better learning by the robot. However, inherent variability in machine learning outcomes presents a challenge to achieving consistent results across trials. The best model obtained through the Curriculum-based approach is evaluated in two scenarios. In the first approach, the final position is fixed, without any condition regarding the bottle getting toppled. In the second approach, the condition of the bottle getting toppled is considered. Two trials are conducted for each approach, and the results are

presented in Tables 13 and 14 below. The observed trend reveals that the robot arm tends to topple the bottle during its attempts to reach it, resulting in a lower success rate. However, when tasked with reaching a fixed point without considering the risk of the bottle falling, the success rate is notably higher.

Table 13: Testing of the best model with a Scenario of fixed position.

Scenario 1: Fixed Position	Count of final states [Episode Counts: 50]		
	BottleFell	End_Success	End_NoSuccess
Trial 1	0	32	18
Trial 2	0	22	28

Table 14: Testing of the best model with a Scenario of reaching a bottle.

Scenario 2: Bottle Position	Count of final states [Episode Counts: 50]		
	BottleFell	End_Success	End_NoSuccess
Trial 1	43	3	4
Trial 2	34	11	5

Moreover, the variation in outcomes between the two trials in both scenarios underscores the dynamic nature of the model's performance across different instances. This highlights the importance of assessing the confidence level of the model for real-life scenarios.

5 LIMITATIONS

Executing the project posed a significant time challenge, with optimal results achieved through extended training steps, but with the trade-off of higher execution time. Training for 50k steps alone took approximately 2.5 hours, while the recommended training times for better performance range from 500k to 3000k steps. Attempts to run for such prolonged durations were also hindered by computational constraints, occasionally resulting in script stalling. The current agent predicts actions within the motion range, occasionally leading to abrupt arm movements, causing objects in its path to topple. An incremental update to actions could mitigate this issue, with the initial action being randomly predicted and subsequent actions featuring incremental changes. This approach aims to prevent sudden, unpredictable movements. Despite extended training durations, the performance of the arm cannot be guaranteed or predicted with absolute confidence. Occasional collisions and abrupt arm movements persist, posing challenges when transferring the learned behaviour to real hardware. Rigorous analysis of results after extended training periods is essential to reaffirm confidence in the model's performance.

6 CONCLUSIONS

This project has explored the application of RL in the domain of object manipulation. As the field of artificial intelligence continues to evolve, the integration of advanced RL techniques in object manipulation will undoubtedly play a pivotal role in shaping the future of automation and robotics.

The successful integration of Gazebo, ROS, and custom OpenAI Gym environments provided an efficient development environment for simulating realistic robotic tasks. The primary goal of positioning the robot arm near the target object involved distance calculation and collision management within the robot arm simulation, utilizing Gazebo plugins and Python scripts. Throughout the training process, we optimized environment parameters and iteratively refined reward logic, transitioning from basic logic to more sophisticated versions, and incorporating reward shaping and tuning approaches. The improvement of the reward system, especially the shift towards a more human-centric approach, highlighted the complexity and challenges inherent in designing effective RL strategies. Implementing curriculum learning, mirroring human learning progression significantly enhanced the model's efficiency. This is a promising approach and better results are expected when executed for more training steps. However, computational limitations were observed, suggesting the need for more powerful devices in future research for enhanced performance. The project deepened our understanding of object manipulation with RL, emphasizing the intricate dynamics in robotic systems.

ACKNOWLEDGMENTS

REFERENCES

[1] R.S. Sutton and A.G. Barto, "Reinforcement Learning: An Introduction," IEEE Transactions on Neural Networks, vol. 9, no. 5, pp. 1054-1054, 1998, doi: 10.1109/TNN.1998.712192.

[2] Z. Zhu et al., "High precision and efficiency robotic milling of complex parts: Challenges, approaches and trends," Chinese Journal of Aeronautics, vol. 35, no. 2, pp. 22-46, 2022, doi: 10.1016/j.cja.2020.12.030.

[3] H. Ju et al., "Transferring policy of deep reinforcement learning from simulation to reality for robotics," Nat Mach Intell, vol. 4, pp. 1077-1087, 2022, doi: 10.1038/s42256-022-00573-6.

[4] A. K. Shakya et al., "Reinforcement learning algorithms: A brief survey," Expert Systems with Applications, vol. 231, p. 120495, 2023, doi: 10.1016/j.eswa.2023.120495.

[5] A. Ray et al., "Spinning Up OpenAI," 2018, [Online]. Available: https://spinningup.openai.com/en/latest/spinningup/rl_intro.html#.

[6] V. Mnih et al., "Human-level control through deep reinforcement learning," 2015.

[7] J. Schulman et al., "Proximal Policy Optimization Algorithms," 2017.

[8] A. Plaat, "Model-Based Reinforcement Learning," 2022.

[9] A. Stone et al., "Open-World Object Manipulation using Pre-trained Vision-Language Models," ArXiv, /abs/2303.00905, 2023.

[10] D. Han et al., "A Survey on Deep Reinforcement Learning Algorithms for Robotic Manipulation," Sensors, 2023, doi: 10.3390/s23073762.

[11] S. Chen and Y. Li, "Active vision in robotic systems: a survey of recent developments," The International Journal of Robotics Research, vol. 30, no. 11, pp. 1343-1377, 2011, doi: 10.1177/0278364911410755.

[12] R. S. Pol and M. Murugan, "A review on indoor human aware autonomous mobile robot navigation through a dynamic environment: survey of different path planning algorithm and methods," 2015 International Conference on Industrial Instrumentation and Control (ICIC), IEEE, 2015.

[13] M. Foukarakis et al., "Combining finite state machine and decision-making tools for adaptable robot behavior," International conference on universal access in human-computer interaction, Heraklion, Crete, Greece, pp. 625–635, Springer, 2-27 June 2014.

[14] R.-E. Precup and H. Hellendoorn, "A survey on industrial applications of fuzzy control," Computers in Industry, vol. 62, no. 3, pp. 213-226, 2011.

[15] O. Boubaker, "The inverted pendulum benchmark in nonlinear control theory," International Journal of Advanced Robot Systems, vol. 10, no. 233, pp. 1-9, 2013.

[16] L. Sciavicco and B. Siciliano, "Modelling and control of robot manipulators," Springer Science & Business Media, Berlin/Heidelberg, 2012.

[17] A. Sharma et al., "Autonomous Reinforcement Learning: Formalism and Benchmarking," 2021.

[18] H. Nguyen and H. La, "Review of Deep Reinforcement Learning for Robot Manipulation," 2019.

[19] M. Towers et al., "Gymnasium," 2023.

[20] Stanford Artificial Intelligence Laboratory et al., "Robotic Operating System, Noetic," 2018, [Online]. Available: https://www.ros.org.

[21] Ageofrobotics, "urdf_tutorial," 2023. [Online]. Available: https://github.com/ageofrobotics/urdf_tutorial.

Paradoxes of the Multi-Chain Critical Paths as the Dissipative Structures

Viktor Nazymko[1], Liudmila Zakharova[2] and Denis Boulik[2]

[1]*Brahch of Physics of Mining Processes, Institute of Geotechnical Mechanics, NAS Ukraine,*
Simferopol Str. 14, 95000 Dnipro, Ukraine
[2]*Department of Management, Donetsk National Technical University, Potebni Str. 56, 43003 Lutsk, Ukraine,*
viktor.nazymko@gmail.com , mila2017ma@gmail.com, boulikdenis@gmail.com

Keywords: Project Scheduling, Simulation, Multiple Critical Paths, Multi-Chain Critical Path, Parametric Uncertainty, Structural Uncertainty.

Abstract: Parametric and structural uncertainties complicate the project management processes. The critical path is one of the pivotal parameters, which helps to control the project schedule and is used to determine the criticality of the tasks and activities that are the most decisive and should be treated during a project expediting or controlling. There may be a set of the critical paths in uncertain environment. Therefore, the main question is which of the critical paths to select. The aim of this paper is to answer to this question. We used Monte Carlo simulation to investigate the multiple critical paths. We revealed and explained several paradoxes that emerged as results of the multiple critical paths occurrence. They are inevitable late bias of the project duration under uncertainty, the tasks probability and their correlation effects, the impact of concurrent chains of the tasks on their criticality, multiplicity of the critical paths and especially multi-chain critical paths. We demonstrated that multiple critical paths are not negative effect. On the contrary, they play extraordinary useful role and are the reliable criterion of the project robustness and stability.

1 INTRODUCTION

Operational management is an ongoing process whereas project management has final deliverable and a finite timespan [1]. That is why timely completion of the project within the planned budget, the scope of the project results and their quality is of particular importance [2].

Unfortunately, the majority of the projects are deployed in uncertain environment [3]. Therefore, almost half of the projects are completed late, over budget, and the scope of planned services and deliverables is incomplete or does not meet the declared quality [4]. Wauters and Vanhoucke demonstrated that the uncertainty induces cost deviation in complex projects and leads to deadline overruns and, consequently, to losses and penalties [5].

One of the most critical parameters in the project schedule control is a critical path. It is a sequence of project network descendent activities. They adds up to the longest overall duration, regardless of whether that longest duration has a float or not. The critical path determines the shortest time possible to complete the project.

Deterministic approach consider a single critical path, which is sufficient to determine the project span. However, there may by several critical paths in a project schedule in uncertain environment. This peculiarity has been noted since the very inception of the critical path method [6], [7]. Nevertheless, the literature analysis has shown that experts try to find the most critical path neglecting the other minor critical paths. Other authors consider that multiple critical paths pose a threat to the project, as the most critical path is hidden among others and may appear in the future, when it is too late and a significant share of resources will be spent on reducing 'faux or miner' critical paths [8], [9]. Nazimko, Zakharova and Smith demonstrated that multiple critical paths in a project schedule as an oriented stochastic activity network are extremely useful information playing an important role in project schedule management under parametric and structural uncertainty. They showed in an example of a project expediting that the most critical path and near critical paths are equitable items or concepts that deserve equal treatment and attention.

Furthermore, Nazimko and Zakharova revealed multi-chain multiple critical paths, which emerge in the uncertain setting [10]. These authors demonstrated that multiple critical paths could bifurcate facilitating extra alternative multi-chain paths, which boost interaction among the tasks and create the synergetic nonlinear effects that multiply the efficiency of the schedule crashing and facilitate the fast-tracking process. The number of the single chain critical paths may be as many as the number of the tasks chains, whereas multi-chain critical paths quantity can exceed them two order of magnitude. Ignoring the multiple critical paths may lead to a 14-23% bias in the modeling results [10].

Despite such an important role of the multi-chain critical paths, their parameters have not been investigated thoroughly so far. Therefore, the aim of this paper was to explore parameters of these paths and their behaviour during the project deployment in an uncertain environment.

First, we described the method of stochastic simulation in this presentation and selected a relevant benchmark to investigate the multiple critical paths behaviour. In the next sections, we analysed the multiplicity of the critical paths, their multi-chain form, duration, effect of the tasks probability variation and correlation between the tasks. Then, we discussed some paradoxes associated with the multiple critical paths and limitations of the research scope and results, made conclusion and proposed practical recommendations.

2 METHODOLOGY OF THE MULTIPLE CRITICAL PATHS SIMULATION

2.1 Selection of a Computer Model

We used the Monte-Carlo Simulation method [11] to investigate the multiple critical paths, which emerge in a project schedule.

We presented a project schedule as an acyclic antisymmetrically oriented multigraph. Nodes in the graph stood for events, whereas arcs played the role of activities or tasks. As the activities duration varied, we simulated it using predetermined distributions. We employed field data, experience, or assigned the distributions by experts according to PMBoK [12]. Because numerical methods have been used, any popular type of distribution could be programmed, namely, uniform, normal, triangular, PERT, and arbitrary distribution that might be designed from discrete bins.

In addition, we took into account symmetric and asymmetric correlation between input parameters. The structural uncertainty has been modeled by assigning a probability to some activities according to [10]. When an activity disappeared eventually, a dummy arc with zero duration has been used.

An initial input record in a stochastic procedure presented every task. The record consisted of task number, task probability, and task position relatively a descendant (from one node of the multigraph to another). Next, the type of task duration distribution and its parameters were input.

Besides, we assigned correlation (yes or no) and the coefficient of correlation between the activities. For example, task A correlated with task B that means asymmetrical correlation when A depends on B. Symmetrical correlation provides other effects, thus this reciprocal correlation was indicated by the mutual dependence. To this end, we used the same correlation coefficient and indication of the companion both in the record of task A and the record of task B.

A current critical path has been detected during every run. When the critical path was structurally identical to previously identified critical paths, the number of such paths was incremented by unit. If there was no such critical path in the formerly identified set, then the critical paths' list has been upgraded.

The closeness of the activities correlation was less then unit because the duration of the activities was selected as a random value. Besides, the structure of the network was uncertain, hence the critical paths and their duration varied during the simulation. The model counted the number of the critical paths and determined the frequency of every structurally identical path at the end. As a result, a set of all possible critical paths has been identified because of a large number of runs.

Any type of the tasks duration distribution has lower and upper boundaries, whereas the dimension of the durations was integer and measured by days, weeks, etc. Thus, theoretically, the number of possible permutations and runs is limited. The real number of the permutations is even essentially less taking into account the correlations and physical restrictions. For example, ten thousands of runs are sufficient to discover almost all critical paths [7].

In addition, a limit rate of the critical paths number increment was assigned to check when the increment of the critical paths becomes negligible and simulation should be stopped.

2.2 Substantiation of a Benchmark

The critical paths diversity occurs especially in a schedule that consists of many serial and parallel chains of tasks.
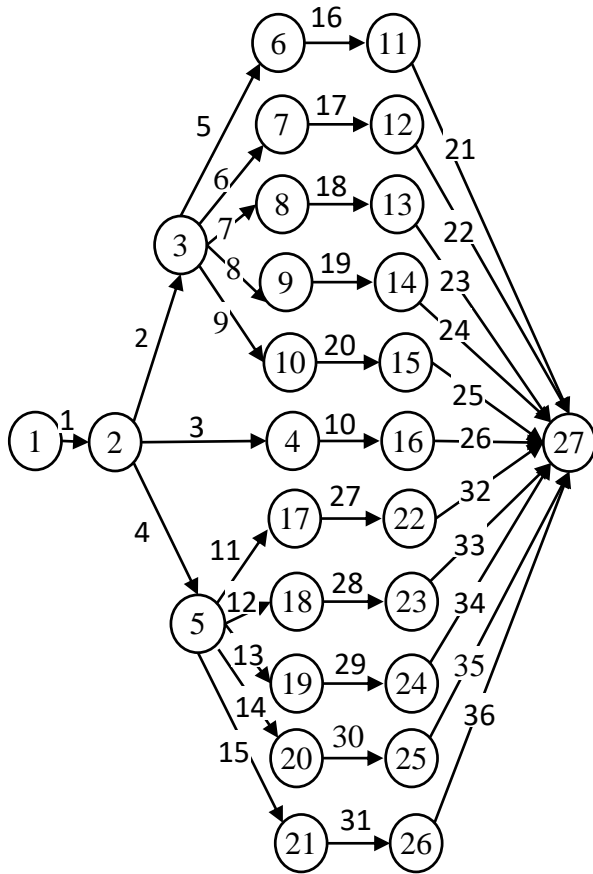

Figure 1: Project duration distribution.

That is why a project schedule composed by Tolentino Rena, R.A. [13] has been selected as the benchmark because it comprised a big number of serial and parallel chains of the tasks (Table 1). This schedule has been selected as the benchmark by several authors [10], [14].

The structure of the benchmark schedule is depicted in Figure 1. The benchmark schedule consisted of 36 tasks that were structured in 11 concurrent chains composed by 4-5 serial tasks. The tasks durations were presented by triangular distributions that have been characterized by minimum, maximum possible duration and the mode.

Table 1: Initial data of the benchmark project schedule composed by Tolentino Rena [13].

Task number	Probability	From	To	Type	Mode	Min	Max
1	1	1	2	3	10	5	15
2	1	2	3	3	20	10	30
3	1	2	4	3	30	20	40
4	1	2	5	3	20	10	30
5	1	3	6	3	15	10	20
6	1	3	7	3	15	10	20
7	1	3	8	3	15	10	20
8	1	3	9	3	15	10	20
9	1(0.5)	3	10	3	15	10	20
10	1	4	16	3	22	17	27
11	1	5	17	3	15	10	20
12	1	5	18	3	15	10	20
13	1	5	19	3	15	10	20
14	1	5	20	3	15	10	20
15	1	5	21	3	15	10	20
16	1	6	11	3	15	10	20
17	1	7	12	3	15	10	20
18	1	8	13	3	15	10	20
19	1	9	14	3	15	10	20
20	1	10	15	3	15	10	20
21	1	11	27	3	15	10	20
22	1	12	27	3	15	10	20
23	1	13	27	3	15	10	20
24	1	14	27	3	15	10	20
25	1	15	27	3	15	10	20
26	1	16	27	3	15	10	20
27	1	17	22	3	15	10	20
28	1	18	23	3	15	10	20
29	1	19	24	3	15	10	20
30	1	20	25	3	15	10	20
31	1	21	26	3	15	10	20
32	1	22	27	3	15	10	20
33	1	23	27	3	15	10	20
34	1	24	27	3	15	10	20
35	1	25	27	3	15	10	20
36	1	26	27	3	15	10	20

3 RESULTS OF SIMULATION

3.1 Multi-Chain Multiple Critical Paths

A specific feature of stochastic activity network is that it yields approximate or not exact result because the random nature. Put other words, the results of computer simulation are fluctuated relatively certain mathematical expectation which was 77 time units in this case. Seventy seven units duration has been computed and the crisp sum of the tasks' duration modes. The average duration of the project was 82.7±4.1 (Figure 2). Noticeably, this fluctuation biased to the right (82.7>77) or towards higher value relatively the crisp mathematical expectation.

The project may be completed during 84 time units with 95% confidence (Figure 2, bottom).
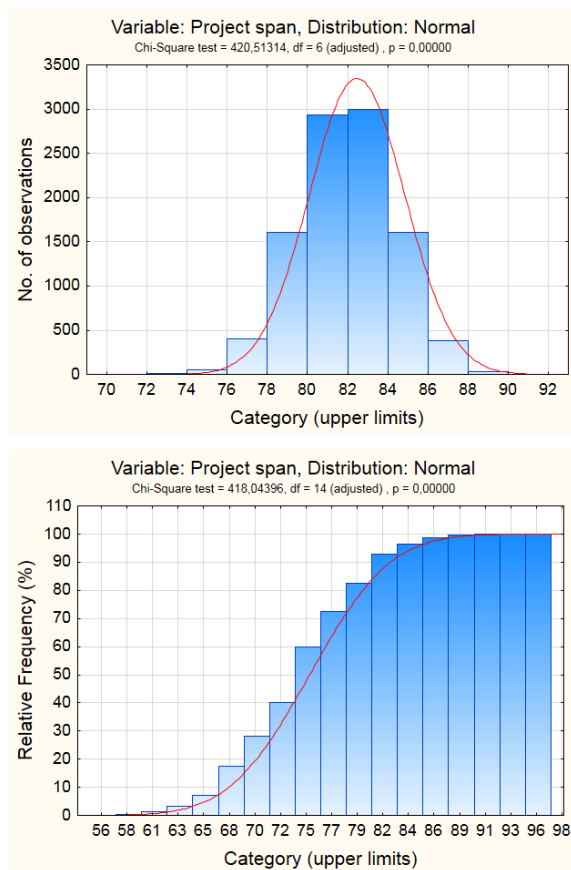
Figure 2: Project duration distribution.

Computer simulation produced 209 critical paths that is by order more than the number of the tasks chains in the project schedule. That is why we classified the critical paths as multiple. Eleven single-chain critical paths (Table 2) corresponded to eleven

tasks sequences in the project schedule. Critical path 1-3-10-26 occurred the most frequently. The other 10 critical paths represented 10 parallel series of the tasks. The frequency of the every path variated in the range of 0.059-0.065. Thus, number of the single chain critical paths corresponded to the number of the task chains. The other one hundred and ninety six critical paths were multi-chain trails.

Table 2: Typical example of the multiple multi-chain critical paths.

#	Critical path	Frequency
7	1-3-10-26	0.207
2	1-2-7-18-23	0.059-0.065
4	1-4-12-28-33; 1-4-14-30-35	0.028-0.047
11	1-2-9-20-25; 1-3-10-26; 1-4-2-28-33	0.005-0.008
155	1-2-6-17-22; 1-2-9-20-25; 1-4-12-28-33; 1-4-15-31-36; 1-2-6-8-19-24; 1-2-6-17-21; 1-2-8-19-24	< 0.001
207	1-4-11-27-32; 1-4-12-28-33; 1-4-13-29-34-; 1-4-14-30-35	0.001
173	1-2-5-16-21; 1-2-7-18-23; 1-2-8-19-24; 1-4-14-30-35; 1-4-15-31-36;	0.001
203	1-2-6-17-22; 1-2-8-19-24; 1-2-9-20-25; 1-3-10-26; 1-4-11-27-32; 1-4-13-29-34	<0.001

Criticality of the major paths is presented in the Table 3. Sequential task 1 has maximal criticality that is natural. Despite the tasks 3 belongs to the most critical path, its criticality is almost two times less than those of tasks 2 and 4, belonging to the critical paths (Table 2) having lover frequency.

Table 3: Distribution of the tasks criticality.

Task	1	2,4	3,10,26	5-9	11-15
Criticality	1.0	**0.42**	**0.25**	0.09	0.09
Frequency of the critical paths, which contains this task	1.0	**0.059-0.065**	**0.207**	0.005-0.008	0.001

3.2 Duration of Multi-Chain Multiple Critical Paths

Another unusual feature of the multiple critical paths is their duration and composition. Let us recall that

we identified the critical paths that go straight the same sequence of the tasks. However, similar critical paths may have different duration.
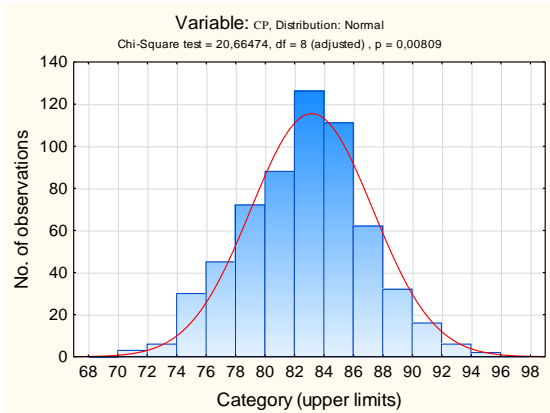
Figure 3: Distribution of duration for critical path 1-2-6-12-22-27.

For example, path 1-2-6-12-22-27 duration ranged from 68 time units up to 98, and average was 83.2±4.1 (Figure 3). Therefore, one should not only identify a relevant multiple critical path but also take into account the distribution of its duration. This particularity imposes additional restrictions and requirements on the possible range decisions.
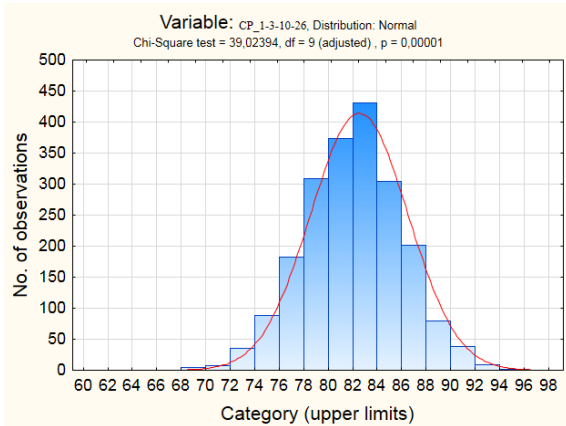
Figure 4: Critical path 1-3-10-26-27 duration distribution.

Figure 4 shows similar distribution of critical path 1-3-10-26-27. Expected value of path duration is 82.5±4.0 that is some more than former critical path. However, the distributions should be compared using certain criterion, for example Kolmogorov-Smirnov [15].
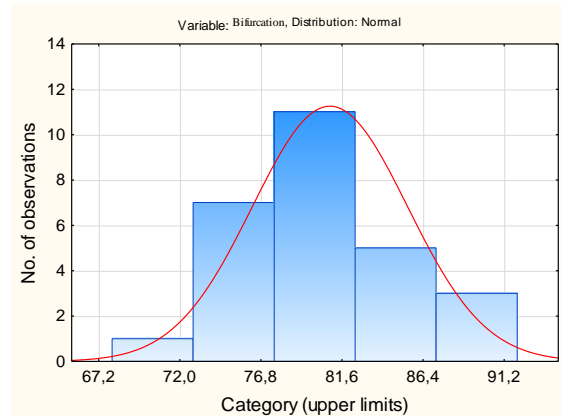
Figure 5: Critical path 1-3-10-26-27.

The duration of the bifurcated critical path 1-3-10-26;1-4-11-27-32 is presented in Figure 5. Range of its duration varied from 67 to 92 time units whereas the expected average was 80.9±4.6. This was another paradox: how may the multi-chain critical path have less duration in comparison the single-chain critical path?

3.3 Effect of a Task Probability and Tasks Correlation

The probability of the task 9 has been reduced down to 0.5 to simulate structural uncertainty (look in Table 1 position 9 in parenthesis). Table 4 demonstrates the main effect of tis modification. Previous frequencies of the critical paths are indicated in the parenthesis.

Table 4: Redistribution of the critical paths frequency after decrease of the task 9 probability.

#	Critical path	Frequency
7	1-3-10-26	0.219(0.207)
2	1-2-7-18-23	0.060-0.0692 (0.059-0.065)
21	1-2-**9**-20-25	0.028(0.062)

The single-chain critical path in which task 9 was arranged reduced the frequency more than twice whereas adjacent critical paths increased their incidences. All multi-chain critical paths or their branches which had task 9 dropped their frequencies essentially. Generally, redistribution of the frequencies occurred for the benefit of the neighboring chains of the schedule's tasks.

Reciprocal relation between the tasks was simulated indicating the correlation of the task 9 and task 24. The correlation coefficient was selected as 0.8.

Table 5: Redistribution of the critical paths frequency after introduction of the correlation.

#	Critical path	Frequency
7	1-3-10-26	0.201(0.207)
2	1-2-7-18-23	0.057-0.069 (0.059-0.065)
21	1-2-**9**-20-25	0.058(0.062)
11	1-2-8-18-**24**	0.07(0.065)

There was not apparent effect of the correlation between the tasks although some redistribution of the frequencies was registered. The frequency of the chain having task 9 slightly dropped whereas the frequency of the sequence where task 24 was located increased (Table 5).

3 DISCUSSION

The general effect of the uncertainty manifested first of all in the bias of the project duration to the more later deadliness. This bias of the tasks sequences and project duration is natural effect of the fundamental thermodynamic law, which postulates excess entropy production from the chaos and uncertainty [16]. Despite the task duration variation occurs symmetrically in initial data, the final results inevitably connect with the entropy production.

The next paradox was presented by illogical effect when the tasks that were located in the serial sequences having evidently less frequencies manifested essentially bigger criticality. The explanation of this paradox lays in the parallel structure of the schedule. The parallel chains of the tasks arrangements took over the flow of resources and information creating alternative channels. Such a parallel system played a role of an open dissipative structure that absorb the energy, material and information fluxes [10].

The most unexpected feature of the project schedule as a stochastic activity network was multiplicity of the critical paths and especially their multi-chain structure. The computer simulation has demonstrated that such structure induced essential bias and put forward the problem of a structural uncertainty that plays an important role in processes of the project management. This finding proved the importance of the critical paths multiplicity and even their versatile multi-chain forms. We demonstrated that all critical paths are important equally and should be concerned carefully. There is no 'the most critical' path or 'minor' path. The alternative paths may change each other which induces useful competitive process of a schedule expediting or enhancing.

We investigated effects of the tasks probability and their correlation demonstrating that these structural parameters may essentially impact the critical path frequency redistributing it between the tasks' sequences. These structural parameters may be involved more actively in the project management process to provide its sustainability and robustness.

4 CASE STUDY

Underground coalmines operate in highly uncertain geologic environment. Moreover, their activities exhibit evident characteristics of the project. For example, any panel extraction occurs in unique geologic conditions. Hence, an underground coal mine consistently takes shape within a structured program framework. There are strategic programs of coal extraction for a big block of coal resources. These programs have a substantial duration, often spanning 10 to 20 years. In contrast, tactical or current programs are generally organized for shorter intervals, typically covering a one-year period.

We analyzed a year program of a big Ukrainian coalmine "Pokrovs'ke". The program consisted of 118 tasks, namely extraction of 12 longwall panels and driving of 106 underground entries. Computer simulation registered 218 critical paths. The most critical tasks were development of Northern conveyor entry (horizon 780 m) and assembling of the equipment in Second northern panel (bock 10). These tasks were integrated in several critical paths of the year program that increased their frequency.

The reason was subjective. Technologists used a constant rate of the entry development. Furthermore, the coalmine management has stipulated a stringent requirement for the assembly of longwall equipment to be completed within a two-month timeframe. However, our examination of experimental data indicated that the actual average duration for this task is three months, with the possibility of its extending even further. Additionally, our findings revealed an exponential attenuation pattern in the driving of entries, correlating with an increase in their length. These distinctive characteristics were taken into consideration during program planning, enhancing its overall reliability.

To offset the delay in commencing the Second northern panel, a coal pillar was extracted from another operational panel. Additionally, a road-header underwent refurbishment to increase the Northern conveyor entry rate and align with the scheduled program.

We employ this Monte Carlo Simulation code to assess project schedules and programs within the construction and mining industries. The insights derived from the analyses have led to recommendations that significantly improve both planning processes and the ongoing control of risks.

5 CONCLUSIONS

Multiplicity of the critical paths generates structural uncertainty of the project schedule having casual variation of input parameters, first the tasks' uncertain durations. Furthermore, the critical paths can branch out which provides more effective dissipation of energy and information.

Having multiple critical paths is advantageous for effective project control, as it enhances the ability to manage and maintain the project's timeline within the scheduled limits.

Reducing the probability of a task from 1.0 to 0.5 can lead to a notable increase in the frequency of the critical path by a factor of 2.8. Furthermore, the correlation between the durations of tasks in distinct critical paths has the potential to redistribute their frequencies. This redistribution could result in an 8% shift, with one path experiencing an increase in frequency while the other undergoes a decrease.

ACKNOWLEDGEMENTS

REFERENCES

[1] APM Body of Knowledge, 7th edition, Association of Project Management, Buckinghamshire, 2022.

[2] S. Monghasemi and M.R. Nikoo, "A Novel Multi Criteria Decision Making Model for Optimizing Time-Cost-Quality Trade-off Problems in Construction Projects," Expert Systems with Applications, 2014, doi: 10.1016/j.eswa.2014.11.032.

[3] G.N. Stock, J. Chia-An Tsai, and G. Klein, "Coping with uncertainty: Knowledge sharing in new product development projects," International Journal of Project Management, 10 October 2020, doi: 10.1016/j.ijproman.2020.10.001.

[4] J. Bröchner, "Project tragedies," International Journal of Project Management, 11 April 2022, doi: 10.1016/j.ijproman.2022.04.001.

[5] M. Wauters and M. Vanhoucke, "A study on complexity and uncertainty perception and solution strategies for the time/cost trade-off problem," Project Management Journal, vol. 47, no. 4, pp. 29–50, 2016, doi: 10.1177/875697281604700404.

[6] J.E. Kelly and M.R.M. Walker, "Critical-path planning and scheduling," in Proc. of the Eastern Joint Computer Conference, pp. 160–173, 1959, doi: 10.1145/1460299.1460318.

[7] R.M. Van Slyke, "Letter to the editor-monte carlo methods and the PERT problem," Operations Research.

[8] A. Bakó, "All paths in an activity network," Mathematische Operationsforschung und Statistik, vol. 7, no. 6, pp. 851-858, 1976, doi: 10.1080/02331887608801343.

[9] Z. Cai, X. Li, and J. Gupta, "Critical path-based iterative heuristic for workflow scheduling in utility and cloud computing," Lecture Notes in Computer Science, 2013, doi: 10.1007/978-3-642-45005-1_15.

[10] V. Nazimko and L. Zakharova, "Project Schedule Expediting under Structural and Parametric Uncertainty," Engineering Management Journal, 2022, doi: 10.1080/10429247.2022.2030179.

[11] D.P. Kroese, T. Brereton, T. Taimre, and Z.I. Botev, "Why the Monte Carlo method is so important today," WIREs Computational Statistics, vol. 6, no. 6, pp. 386–392, 2014, doi: 10.1002/wics.1314.

[12] Project Management Institute, "A guide to the project management body of knowledge (PMBOK® guide)," 6th ed., 2017.

[13] R.A.T. Rena, "Project management: A simulation-based optimization method for dynamic time-cost tradeoff decisions (M.S. thesis)," Rochester, NY: Rochester Institute of Technology, 2009.

[14] M.K. Floyd, K. Barker, C.M. Rocco, and M.G. Whitman, "A Multi-Criteria Decision Analysis Technique for Stochastic Task Criticality in Project Management," Engineering Management Journal, vol. 29, no. 3, pp. 165-178, 2017, doi: 10.1080/10429247.2017.1340038.

[15] M.A. Stephens, "EDF Statistics for Goodness of Fit and Some Comparisons," Journal of the American Statistical Association, vol. 69, no. 347, pp. 730-737, 1974, doi: 10.2307/2286009.

[16] P. Glansdorff and I. Prigogine, "Thermodynamic theory of structure, stability and fluctuations," J. Willey & Sons, New York, 1971, doi: 10.1002/bppc.19720760520.

Power Engineering

Proceedings of the 12th International Conference on Applied Innovations in IT (ICAIIT)

Advancement and Assessment of Power-to-X Strategies as a Significant Contribution for the De-Fossilization of Economies

Halina Falfushynska[1,2], Vladyslav Zhadan[3] and Markus Holz[1]

[1]*Anhalt University of Applied Sciences, Strenzfelder Allee 28, 06406 Bernburg, Germany*
[2]*ENERTRAG SE, Gut Dauerthal, 17291 Dauerthal, Germany*
[3]*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",* Beresteiskyi Avenue 37, 03056 *Kyiv, Ukraine*
galyna.falfushynska@hs-anhalt.de, markus.holz@hs-anhalt.de

Abstract: In order for the EU to achieve its 2050 climate objectives, transport-, power-, and industry sector, which heavily relies on fossil fuels, must significantly decrease their emissions. Green hydrogen and sustainable fuels, among them e-methanol, e-kerosene, and green ammonia, produced using the Power-to-X approach, have been recognised as viable alternatives to reduce emissions and support decarbonization. However, there are some techno-economic challenges in adopting sustainable fuels, particularly related to the higher costs compared to fossil fuels. The economic feasibility of sustainable fuels depends on the cost of green hydrogen and carbon capture technology. Large-scale deployment, improvements in electrolysis modes, and intensive implementation of point-of-source CO_2 capture technologies could make sustainable fuels cost-effective and competitive with fossil fuels by 2050. However, it seems improbable that e-fuels would become widely available at low cost in the near future. The prospective environmental performance of adopting e-methanol, e-kerosene, and e-ammonia is examined, highlighting the necessity of conducting a thorough investigation into the possible negative impacts on human health and ecosystems.

1 INTRODUCTION

Nowadays global concerns surrounding climate change have significantly escalated, with greenhouse gas emissions reaching ceiling. The combustion of fossil fuels contributes to elevation CO_2 and NOx concentrations, leading to the entrapment of thermal energy from the sun. The resultant global warming poses a critical threat to ecosystems, animals, and humanity. In response to this issue, the 2015 Paris Agreement has outlined pivotal objectives, aiming to limit the temperature rise to 1.5 °C and prevent it from exceeding 2 °C above pre-industrial levels[1]. Recognizing the importance of the situation, Germany has acknowledged the imperative need for mitigating carbon emissions and therefore decided in 2021 to strengthen the Climate Action Law in order to reach net greenhouse gas neutrality by 2045[2].

Various potential solutions have been deliberated, with a significant emphasis on transitioning to renewable energy sources for a sustainable future. Solar, hydro, wind, biomass, and biogas represent renewable energy alternatives that stand out for their abundance and continuous renewal [1,2], without affecting long-term environmental damage. There is strong negative correlation between rate of renewable energy usage in EU countries and CO_2 emission (r<-0.9, p<0.001, Figure 1). However, despite of the numerous advantages of solar energy and wind power, their intermittent nature presents a considerable challenge in maintaining a consistent and reliable energy supply [1]. The fluctuations in their output pose obstacles to achieving a stable energy provision. Addressing these challenges requires innovative solutions that go beyond mere energy generation.

To effectively harness the potential of renewable energy, surplus power generated by renewable energy systems must be efficiently stored or transformed into storable substances. Among the promising technologies addressing this need is Power-to-X (PtX), an emerging energy storage solution designed to alleviate the imbalances inherent in renewable energy sources [2]. PtX employs electrochemical
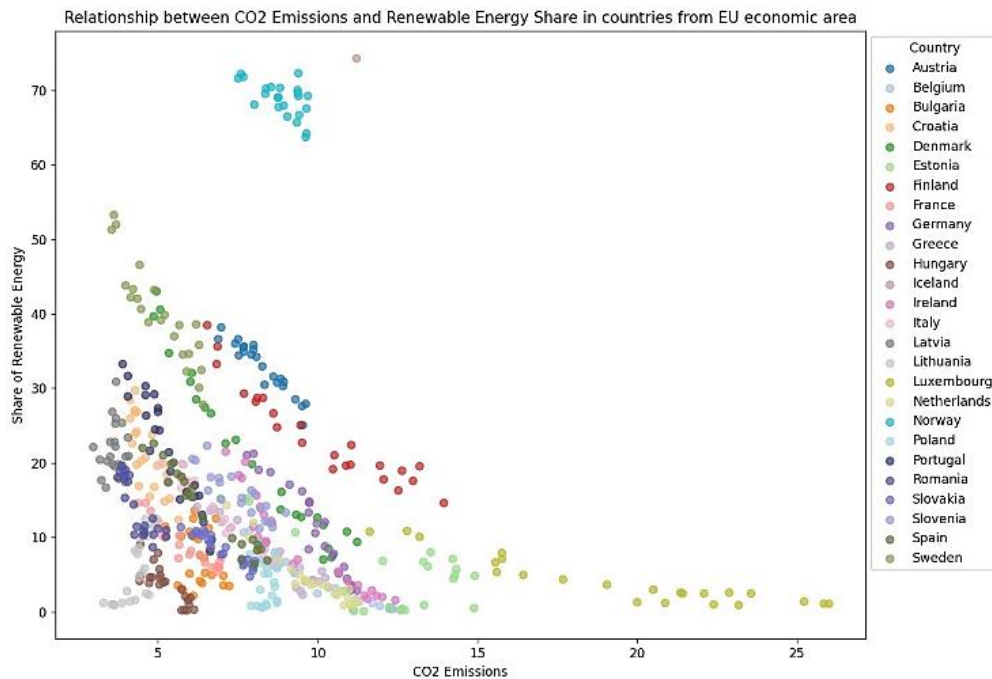

Figure 1: Clustering of relationship between CO_2 emission and rate of renewable energy implementation in EU countries[3].

reactions to convert excess renewable power into hydrogen, which is then combined with carbon molecules to generate a versatile substance referred to as "X". It is worth noting that the number of publications regarding PtX within the last 10 years was increased in 5-10 times depends on the pathway, with the most promising power-to-gas (PtG) and power-to-liquids (PtL), based on the Scopus and WoS database.

Despite their promising potential, PtX technologies face challenges come from diverse pathways, economic uncertainties, inconsistent policies, and spatial and temporal variations [2, 3]. This analytical review explores these challenges, offering insights into the complexities of PtX technologies and their role in shaping the future of renewable energy storage. While the majority of research on e-fuels primarily concentrates on their economic aspects and disregarding their environmental impact and related expenses, we pay attention these critical issues as well.

2 METHODOLOGY

The methodology employed in this study involved a comprehensive search across the Scopus database and Google Scholar, using key terms such as "green

hydrogen", "wind energy", and "Power to X", with a primary focus on "'title, abstract, and keywords". To refine the search and ensure relevance, filters were applied to restrict results to specific research disciplines, namely "Chemical Engineering", "Energy", "Environmental Sciences". Further refinement included filtering by document type, with an emphasis on highly cited publications and review papers to track hydrogen trends over the preceding three-five years.

Data acquisition predominantly relied on the International Energy Agency (IEA), International renewable energy agency (IRENA) website, Statista, which recently updated its information on hydrogen policy and economy for nations worldwide. Official reports from entities such as the Germany's Renewable Energy Agency (AEE), German Energy Agency (DENA), provided additional insights into regional policies. Insights into the current perspectives of hydrogen specialists were derived from news sources and press conferences.

To carried out meta-analysis of existed data Meta-Essentials macros for Excel was used [4]. Tornado and population graphs were plotted using GraphPad 9.0 in order to compare economic- and environment-related traits of conventional and e-fuels. Python was used to estimate the relationship between CO_2 emission and progress in renewable energy implementation.

---

[3] based on open source data, https://ourworldindata.org

3. POWER-TO-X

3.1 E-Methanol

Methanol and its derivates serves as initial compounds for the manufacturing of several everyday products, including paints, carpets, plastics, and pharmaceuticals. Its versatility expands worldwide, being utilized in innovative ways to address the increasing energy demands, among them its implementation as environmentally friendly energy carrier and storage as well as a renewable fuel that can replace the fossil-derived analogues [5]. Also, switching to e-methanol might help significantly cut down on CO_2 emission (by 59%) [5] because production of grey and blue methanol is responsible for ~ 0.3 Gt CO_2 annually which accounts for almost 10% of the overall emissions from the chemical sector [6].

3.1.1 Techno-Economic Challenges and Putative Solution in E-Methanol Adoption

Power-to-Methanol employs a synergistic combination of hydrogen and CO_2 in 3:1 equimolar ratio over $CuO/ZnO/Al_2O_3$ at elevated pressure and temperature (200-300 °C) to synthesize methanol. This resulting methanol can be conveniently stored in liquid form and harnessed for energy generation through various processes, including hydrogen breakdown, direct utilization in fuel cells, or combustion or even transform into dimethyl ether, diesel, or jet fuel. Detail analysis of previous findings and predictions proved that the average expenses are in the range of 38 and 62 $€GJ^{-1}$ which is almost two folds higher than for conventional methanol and, therefore, still remains economically uncompetitive, particularly in case of energy produced using onshore wind turbines [6]. It is anticipated that these prices

will become economically feasible with a significant decrease in the cost of green hydrogen which significantly depends on the type of electrolysis (30–35% of the total levelized production costs) and energy mode (50–55% of the overall levelized production costs) and, on the other hand, carbon capture technology (IRENA). e-MeOH has the potential to be an effective rival to traditional MeOH derived from fossil fuels in the market, as long as the cost of H_2 is 2.5 $€·kg^{-1}$ and the price of CO_2 reaches 120 $€·t^{-1}$ [6]. A reduction of renewable electricity cost might be reach in terms of economies of scale, an increase in wind turbine (from MW to GW transition) and wind onshore/offshore capacity (IRENA).

When it comes to electrolysers, alkaline electrolysis and proton exchange membrane electrolysers are the most favoured options to produce green H_2, with a majority of projects choosing AEC. Nevertheless, SOE and PEM electrolysers exhibit a discernible disparity in terms of development compared to AEC systems, notwithstanding their optimistic prospects, which proved by projecting of the average levelized costs of e-methanol. Based on meta-analysis it will have been by the 2030 around 4.2±0.68 $€·kg^{-1}$ with the low level of publication biased (Figure 2).

However, this price is still far from the price for blue hydrogen, which costs ~1.6 $EUR·kg^{-1}$ ([4] [9]). This price holds economic viability within specific sectors but lacks suitability for large-scale industrial supply. Moreover, some developers tend to believe that the costs of green hydrogen unlikely to fall dramatically in the near future because of increase in capex and shortages in electrolyser manufacturing capacity as was announced on the *Investing in Green Hydrogen* conference (London, 2023)[5]. It dictates the necessitates additional technological advancements and paramount importance of permanent intense search of smart solutions.
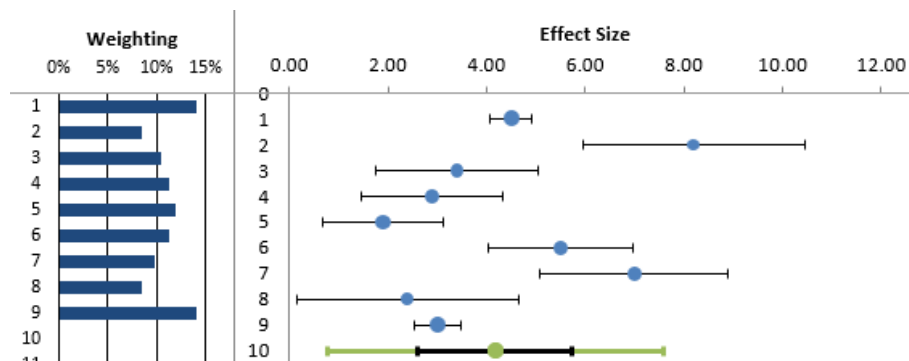
Figure 2: Meta-analysis of levelized costs for green hydrogen production ($€·kg^{-1}$) based on [5-8].
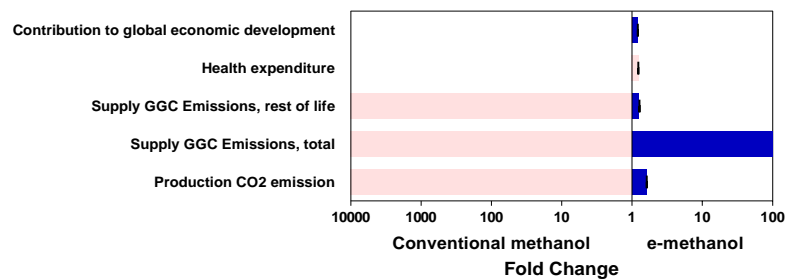
---

[4] BloombergNEF

[5] www.hydrogeninsights.com

Figure 3: Comparative analysis of conventional and e-methanol related to environment and human health effects (based on data from [5-8, 10, [6]]).

3.1.2 Prospects for Environmental Performance Regarding E-Methanol Adoption

If conventional methanol production supports sky-rocketing CO_2 emission, e-methanol brings itself closing to zero-emission CO_2 [6], around 4.4 g CO_2 eq/MJ (Figure 3). This indicates a reduction in emissions of almost 90% compared to conventional methanol. Green methanol also connected to a decrease in NOx emissions by up to 80% and totally eliminates SOx and particulate matter emissions. It might help humankind to decrease carbon footprint and tackle a problem of constantly increasing greenhouse gas emission and related global warming.

However, there is the trend to further transform e-methanol into dimethyl ether, which would be used as a sustainable diesel fuel replacement because it needs only minor engine modifications. PtX derived dimethyl ether in the case of partial oxidation might form formaldehyde, reactive carbonyl species, a recognised carcinogen. The emissions of formaldehyde during DME fuel production and combustion should be identified as the most important health-related issue in the whole life cycle of this fuel, and, therefore, has to be tightly controlled.

Also, social life-cycle portrait of e-methanol respect for forced labour and healthcare costs among others, shows some disadvantages when compared to conventional methanol [10]. The heightened complexity of the supply chain inside the green methanol system might be the fundamental reason of these issues. Unfortunately, environmental aspects and social life-cycle portrait of e-methanol are still overlooked and, therefore, have to be deep analyse.

3.2 Sustainable Aviation E-Fuels

By 2022, aviation contributed 2% of the world's energy-related CO_2 emissions, counted 800 Mt CO_2, seeing a more rapid growth compared to rail, road,

and shipping in recent decades[7]. Therefore, the projected carbon emissions from 2021 to 2050, assuming no changes in current practices, amount to around 21.2 gigatons of CO_2[8]. Following medium and long-term objectives, ReFuelEU Aviation advocates for the utilization of e-fuels, known as well as sustainable aviation fuel, e.g. e-kerosene, blended with conventional fuels at a rate of 2% in 2025, gradually escalating to 70% by the year 2050 in order to significantly decrease the emissions intensity of aviation or even reach net zero emissions by 2050.

Nevertheless, in 2019, their share in the overall aviation fuel market was less than 0.01% according to the International Energy Agency (IEA, 2019). In 2021, Germany established a goal to utilise e-kerosene for 0.5% of aviation fuel by 2026, with a further increase to 2% by 2030 [11]. To favour these goals, Lufthansa has indicated its intention to procure a minimum of 25000 L e-kerosene per year over the next five years[9]. However, the costs of these fuels are currently three to six times higher than those of traditional fossil-fuel-based aviation fuel.

3.2.1 Particular Techno-Economic Challenges of E-Fuel Production Related to CO_2 Capture Technologies

E-fuel, consists of high-molecular alkanes, is produced by green hydrogen, generated through water electrolysis utilizing renewable energy, with CO_2 sourced from industrial outlets, biogenic sources, or directly captured from the atmosphere or just before its emission to atmosphere in terms of the Fischer-Tropsch synthesis or Methanol-to-Jet synthesis technique. Because of main steps of manufactured process, it is predicted that a substantial proportion of the cost of e-fuel is attributed to the expenditure related to renewable power follow by > synthesis and conversion > CO_2 supply > electrolysis > H_2 storage [12]. It is worth mentioning, that the cost fraction of green energy is pretty the same for production of e-fuel and e-methanol (see 3.1.) [IRENA, 12] and

---

[6] www.methanol.org

[7] https://www.iea.org/energy-system/transport/aviation

[8] https://www.iata.org/en/programs/environment/flynetzero
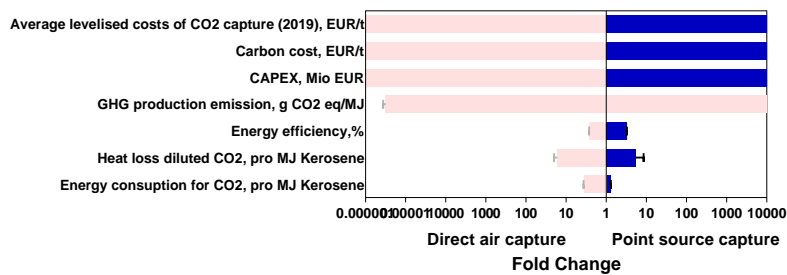
[9] Lufthansa Group, 2021

Figure 4. Comparative analysis of direct air capture and point source capture techniques of CO_2 takeover based on data from [12]. [10, 11]

the activities intended to decrease its price, have been mentioned in 3.1, might be universal, we would pay more attention to CO_2 supply.

Although Direct air capture (DAC) appears as a highly promising way of removing CO_2 from the atmosphere because of its space effectivity, it belongs to the most cost-demand technology in comparison to alternative methods, e.g. point source capture[10].

The cost-effectiveness of collecting CO_2 from industrial systems, among them cement manufacturing and particularly natural gas coal to chemicals processing, exceeds that of DAC because of the greater volume fraction of CO_2 involved, leading to a more economical value. Furthermore, the gaseous stream produced by industrial emissions is found to have a higher purity of CO_2 compared to the surrounding air. This contributes to a more efficient and economical method of concentrating CO_2. It is projected that point of source capture might allow not only optimise energy consumption during the production pathway of e-kerosene, but also support an increase in energy efficiency (44% vs 52%) and environment decarbonization alongside lower CAPEX and levelized cost (up to 300%) (Figure 4). When consider industrial and renewable energy capacities some Asia and South America countries attract attention on that purpose. As an example Taiwan and Brazil possess leading position into cement manufacturing around the world[12] and at the same time have abundant green power resources, among them high off- and onshore wind capacities what makes them secure players on the e-fuel marketplace.

Alongside with DAC and point source capture technique, there are some others, such as Visible-Light-Driven CO_2 Reduction [13], that might be useful in CO_2 capture and processing, however they are on the juvenile stage and its large-scale implementation is pretty murky.

___

[10] www.iea.org

[11] https://kth.diva-portal.org/smash/get/diva2:1520216/FULLTEXT01.pdf

[12] https://www.globalcement.com/magazine/articles/822-top-75-global%20cement%20companies

3.2.2 Prospects for Environmental Performance Regarding Green Kerosene Adoption

Presently, the most reliable approach involves the use of emission factors that establish the relationship between the quantity of fuel consumed and the pollutants produced (e.g., in grams of pollutant per gram of fuel burned). However, the standards for incorporating fossil kerosene and e-kerosene into a blend are often disregarded, as evidenced by numerous references, among them [14]. While e-jet fuels typically exhibit a lower global warming potential [15, 16] and are less scarce in comparison to fossil jet fuels, they might adversely affect human health and ecosystem quality. This implies a shift in the balance of negative effects from resource scarcity towards concerns related to human health and the quality of ecosystems. It is believed that the environmental impacts of Power-to-Liquid kerosene varied significantly across the production pathways and the layout using high-temperature electrolysis, low-temperature Direct Air Capture and electricity from wind power pretends to be the most environmental-friendly [16]. Although various materials that were suggested for Direct Air Capture, including metal hydroxide-based absorption and amine-supported materials, they might provoke specific toxicity due to potential leakage of target chemicals into milieu, among them pH effect (alkalisation), hypercalcemia, skin and respiratory track irritations. Eliminating environmental threats by using green sorbents, in other words bioregenerative materials and Ionic liquids, e.g. mesoporous chitosan-SiO_2 nanoparticles with a capacity of 193.16 mg of CO_2 per gram of sorbent and superhydrophobic PVDF/Si-R hollow fiber membrane [17]. A scarcity of findings regarding the environmental impacts of e-kerosene, as a desirable fuel option in the future within the context of potential hazards for humankind, underscores the importance for an in-depth study into its plausible adverse effects.
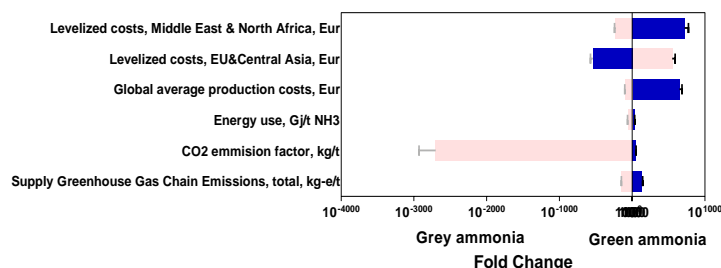
Figure 5: Comparative analysis of grey and green ammonia based on data from [19].[13, 14]

3.3 Green Ammonia

Approximately 70% of global ammonia production is conveyed for fertilizer synthesis, among them ammonium nitrate, with the remaining proportion finding applications across diverse industrial sectors, including the production of plastics, explosives, and synthetic fibers. Nowadays, the USA, China, Russia belong to the main producers of ammonia. The current trajectory of the ammonia industry is considered unsustainable because of unstable geopolitical situations, massive CO_2 emission (~450 Mt CO_2, 1.8% overall global CO_2 emissions) and energy (~4% of the sector's energy inputs) demand (ww.iea.org), since predictions suggest a nearly 40% increase in output by 2050 under the Stated Policies Scenario, driven by economic development and population expansion. While the potential of ammonia as a clean energy source is acknowledged, its current integration into energy systems remains in its nascent stages.

3.3.1 Techno-Economic Challenges of Green Ammonia Production

Ammonia should be synthesized using iron-based catalytical Haber-Bosch reaction. Although ammonia possesses a lower heating value of 18.6 $MJ \cdot kg^{-1}$, which is about 40% of the energy density of petrol, it pretends to be the important player on the marketplace. Ammonia is particularly noteworthy due to its remarkable gravimetric hydrogen content of 17.8%, which places it among the energy carriers with the greatest levels, transportability and potential for decarbonization of NH_3. The existing infrastructure of a comprehensive liquid ammonia transport network, mostly used in the fertilizer sector, offers a readily available basis for its global transportation. Once delivered, ammonia can be catalytically decomposed back into hydrogen or used directly as a fuel, thus making a substantial contribution to the worldwide effort to achieve sustainable energy solutions. To this end it is worth

mentioning that its production and decomposition of ammonia require less energy than liquefying hydrogen, favouring its use as an efficient method of storing hydrogen.

The main differentiating aspect among modern ammonia manufacturing methods is the hydrogen generation pathway. Currently, more than 96% of the hydrogen used worldwide in the production of ammonia comes from fossil fuel sources, what, however, in turn related to releasing 2.5-2.9 kg of CO_2 eq$\cdot kg^{-1}$ of ammonia produced (Figure 5) [18]. In order to follow Green Deal politic, grey/blue hydrogen needs to be substitute with green H_2. However, because of high price for green hydrogen, the present price of green ammonia produced by an integrated plant that combines renewables, hydrogen, and ammonia variated in the range of 540- 810 EUR per t vs an average cost for convential ammonia of less than 200 US $ based on the natural gas price before 2022. However, considering the peak price of natural gas in 2022, the grey ammonia price reached sky-high 1.500 US $, what bears witness to the favour of green NH_3[15].

It is expected, however, that in some countries, among them Australia, Central Asia and Middle East [20], with a favourable combination of solar and wind resources, as well as economic conditions that support low capital costs (capital cost < 8%), the projected cost of producing green ammonia will have been 360 Euro/t NH_3 by 2030 (Figure 5) [19]. It is worth noting, that in particular cases the cost of producing green ammonia keep up with the cost of producing grey ammonia, placing them in direct competition.

It is believed that the synthesis of green ammonia might be a feasible and economically advantageous substitute for the conventional Haber-Bosch method. This belief holds true when considering improvements in the energy efficiency of electrolyzers, reductions in the prices of water electrolyzers, and decreases in the levelized cost of electricity for renewable energy. Some findings in the frame of Techno-Economic Analysis and Life Cycle

---

[13] https://www.futurebridge.com/industry/perspectives-energy/green-ammonia-an-alternative-fuel

[14] https://www.mckinsey.com/industries/agriculture/our-insights/from-green-ammonia-to-lower-carbon-foods

[15] https://www.eia.gov/todayinenergy/detail.php?id=52358

Assessment have emphasized the potential advantages of Solid Oxide Electrolysis for producing green H_2 as the feedstock for ammonia synthesis. When compared with conventional Alkaline Water Electrolysis and promising Proton Exchange Membrane electrolysis of water, Solid Oxide Electrolysis demonstrated significant cost rate of 0.9372 \$/s with an exergy efficiency of 13.15 % and the least amount of CO_2 emissions [21, 22]. This is mainly due to its reduced energy requirements compared to other electrolysis methods. Meanwhile, all these methods demonstrate more valuable traits related to energy demands and CO_2 emission in the future than conventional ammonia (Figure 5). The Ammonia Energy Association has projected the production costs for NH_3 in 2030 (0.48 USD per kg) and 2050 (0.32 USD kg^{-1})[16], which in relation to CO_2 tax appear more economically valuable than the conventional Haber-Bosch NH_3 prices (with and without a CO_2 tax: 0.3 and 0.6 USD kg^{-1} NH_3, respectively) [21].

Considering the projected 37% increase in global ammonia production by 2050 and the associated concerns about CO_2 emissions, the adoption of green ammonia emerges as a valuable contributor to global economic growth, thanks to its low carbon footprint. While the substantial costs linked to ammonia make it unlikely to become a primary player in the global energy supply, there are niche applications where it can prove beneficial. One such use is as a transitional fuel in existing equipment, facilitating the transition to renewable energy sources. Moreover, ammonia may find relevance in scenarios where the feasibility of renewable energy is limited, such as in the shipping industry. In these specialized fields, green ammonia can play a strategic role in advancing sustainability goals and mitigating environmental impact.

3.3.2 Environmental Challenges Regarding Green Ammonia Adoption

Green ammonia has the potential to serve as a substitute for fossil fuels on a large scale in challenging-to-decarbonize regions of the energy and transportation sectors. Nevertheless, it is important to acknowledge potential drawbacks associated with its adoption, which should be prevented. The potential reduction in CO_2 emissions achieved by using ammonia may be completely negated by the release of nitrogen oxides (NO_X) and nitrous oxide (N_2O) emissions during its burning. It is tightly connected to potential adverse outcomes (Figure 6), among all respiratory system disorders, inflammation and immune disorders in human and acid rain and ozon depletion in frontline in environment [23].

It is worth noting that only in China NOx emission will have reached 45.5 Mt by 2030 under baseline scenario with an average annual growth rate of 1.84% [24]. It is however expected that this amount should be exceeded because of additional emission of NOx by e-ammonia adoption. To prevent this, it is necessary to decrease flame temperature or employ highly efficient post-treatment technologies, e.g. denitrification technology, to reduce the release of unintended emissions.
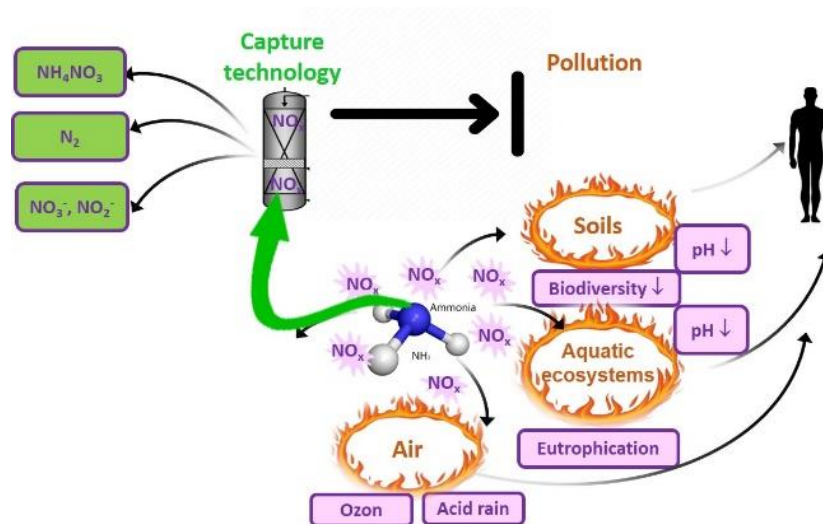
Figure 6: Exploring the potential adverse outcomes of NO_x emissions in relation to Green Ammonia and proposed solutions for mitigation.

---

[16] https://www.ammoniaenergy.org/articles/the-cost-of-co2-free-ammonia

The selective catalytic (using Mn-based catalyst) or selective non-catalytic reduction [25] appear promising that allow to convert NO_X to N_2 which is actually the initial reagent of Haber-Bosch synthesis of ammonia and NH_4NO_3, fertilizer and just NH_3, correspondingly (Figure 6). Due to our knowledge, there is no approved NOx reduction emission strategy but only Nitrogen Oxides (NOx) Control Regulations by US EPA. However, the growing emphasis on NOx-related issues encourages proactive efforts to prune down on emissions and support the smooth adoption of green ammonia and sustainable development of the environment.

4 CONCLUSIONS

In conclusion, the adoption of green fuels, including e-methanol, e-kerosene, and green ammonia, synthesized through Power-to-X technique, holds promise in addressing environmental challenges and reducing carbon emissions in various sectors. These green fuels present opportunities for a sustainable energy transition, but a balanced approach is crucial to address economic, environmental, and social considerations. Effectively integrating e-fuels and PtX approach into various economic sectors requires addressing paramount challenges related to techno-economic aspects and environmental factors. In the foreseeable future, extending beyond 2035, there seems to be no defined role for storage of the mentioned derivatives. This is due to the ongoing high cost of Green Hydrogen, making it a precious primary energy source. Any re-transformation, such as converting green NH_3 back into green Hydrogen, results in notable efficiency and energy yield losses.

ACKNOWLEDGMENTS

REFERENCES

[1] P.A. Owusu and S. Asumadu-Sarkodie, "A review of renewable energy sources, sustainability issues and climate change mitigation," Cogent Engineering, vol. 3, 2016.

[2] A.R. Dahiru, A. Vuokila, and M. Huuhtanen, "Recent development in Power-to-X: Part I - A review on techno-economic analysis," J. Ener. Stor., vol. 56, pp. 105861, 2022.

[3] B.S. Zainal, P.J. Ker, H. Mohamed, H.C. Ong, et. al., "Recent advancement and assessment of green hydrogen production technologies," Renew. Sustain. Energ. Rev., vol. 189(A), pp. 113941, 2024.

[4] R. Suurmond, H. van Rhee, and T. Hak, "Introduction, comparison and validation of Meta-Essentials: A free and simple tool for meta-analysis," Res. Synth. Meth., vol. 8, pp. 537-553, 2017.

[5] S. Sollai, A. Porcu, V. Tola, F. Ferrara, and A. Pettinau, "Renewable methanol production from green hydrogen and captured CO2: A techno-economic assessment," J. CO2 Utiliz., vol. 68, pp. 102345, 2023.

[6] Y. Rahmat, S. Maier, F. Moser, M. Raab, et. al., "Techno-economic and exergy analysis of e-methanol production under fixed operating conditions in Germany," Applied Energy, vol. 351(C), pp. 121738, 2023.

[7] S. Pratschner, F. Radosits, A. Ajanovic, and F. Winter, "Techno-economic assessment of a power-to-green methanol plant," J. CO2 Utilization, vol. 75, pp. 102563, 2023.

[8] N. Heimann, R.-U. Dietrich, M. Raab, N. Wulff, R. Moser, and T. Francisco, "Standardized techno-economic analysis for the production of e-fuels in Germany," 11th International Conference "Fuel Science: From Production to Propulsion" 2023, Deutschland, 2023.

[9] F. Ueckerdt, P. Verpoort, R. Anantharaman, C. Bauer, F. Beck, T. Longden, and S. Roussanaly, "On the Cost Competitiveness of Blue and Green Hydrogen," Joule, 2024.

[10] D. Iribarren, R. Calvo-Serrano, M. Martín-Gamboa, A. Galán-Martín, and G. Guillén-Gosálbez, "Social life cycle assessment of green methanol and benchmarking against conventional fossil methanol," Sci. Total Environ., vol. 824, pp. 153840, 2022.

[11] K. Appunn, "CO2 reduction and biofuels in Germany's transport sector-Implementing the RED II directive," Clean Energy Wire, 2021.

[12] E-Fuels: A technoeconomic assessment of European domestic production and imports towards 2050, Concawe, 2022, [Online]. Available: https://www.concawe.eu/wp-content/uploads/Rpt_22-17.pdf.

[13] S. Roy, and E. Reisner, "Visible-Light-Driven CO2 Reduction by Mesoporous Carbon Nitride Modified with Polymeric Cobalt Phthalocyanine," Angewandte Chemie, vol. 58, pp. 12180-12184, 2019.

[14] F.D. Ordóñez, T. Halfdanarson, C. Ganzer, N. Shah, N.M. Dowell, and G. Guillén-Gosálbez, "Evaluation of the potential use of e-fuels in the European aviation sector: a comprehensive economic and environmental assessment including externalities," Sustain. Energy Fuels., vol. 6(20), pp. 4749-4764, 2022.

[15] Deutsche Energie-Agentur (Publisher) (dena, 2022) "E-Kerosene for Commercial Aviation, From Green Hydrogen and CO2 from Direct Air Capture – Volumes, Cost, Area Demand and Renewable Energy Competition in the United States and Europe from 2030 to 2050".

[16] M. Micheli, D. Moore, V. Bach, and M. Finkbeiner, "Life-Cycle Assessment of Power-to-Liquid Kerosene Produced from Renewable Electricity and CO2 from Direct Air Capture in Germany," Sustainability, vol. 14(17), pp. 10658, 2022.

[17] Y. Abdullatif, A. Sodiq, N. Mir, Y. Bicer, et. al., "Emerging trends in direct air capture of CO2: a review of technology options targeting net-zero emissions," RSC Adv., vol. 13, pp. 5687-5722, 2023.

[18] P. Mayer, A. Ramirez, G. Pezzella, B. Winter, S.M. Sarathy, et. al., "Blue and green ammonia production: A technoeconomic and life cycle assessment perspective," iScience, vol. 26, pp. 107389, 2023.

[19] D. Saygin, H. Blanco, F. Boshell, J. Cordonnier, K. Rouwenhorst, et. al., "Ammonia Production from Clean Hydrogen and the Implications for Global Natural Gas Demand," Sustainability, vol. 15, pp. 1623, 2023.

[20] J. Egerer, V. Grimm, K. Niazmand, and P. Runge, "The economics of global green ammonia trade – 'Shipping Australian wind and sunshine to Germany'," Appl. Energy., vol. 334, pp. 120662, 2023.

[21] B. Lee, L.R. Winter, H. Lee, D. Lim, H. Lim, and M. Elimelech, "Pathways to a Green Ammonia Future," ACS Energy Lett., vol. 7, pp. 3032-3038, 2022.

[22] M.M. Nejadian, P. Ahmadi, and E. Houshfar, "Comparative optimization study of three novel integrated hydrogen production systems with SOEC, PEM, and alkaline electrolyzer," Fuel., vol. 336, pp. 126835, 2023.

[23] W. de Vries, "Impacts of nitrogen emissions on ecosystems and human health: A mini review," Cur. Opinion Environ. Sci. Health, vol. 21, pp. 100249. 2021.

[24] L. He, B.B. Wang, W. Xu, Q. Cui, and H. Chen, "Could China's long-term low-carbon energy transformation achieve the double dividend effect for the economy and environment?" Environ. Sci. Pollut. Res. Int., vol. 29, pp. 20128-20144, 2022.

[25] Z. Zhu and B. Xu, "Purification Technologies for NOx Removal from Flue Gas: A Review," Separations, vol. 9, pp. 307, 2022.

Results of Study of Photovoltaic Thermal Battery Based on Thin-Film Module by Modeling and Computational Methods

Islom Jurayev, Isroil Yuldoshev and Zukhra Jurayeva

Tashkent State Technical University named after I.Karimov, University Str. 2, 100095 Tashkent,Uzbekistan
nauka-jir@mail.ru, yuldashev.i2004@gmail.com, zuxra_199@mail.ru

Abstract: This article presents the results of model studies of a photovoltaic thermal battery (PVTB) based on a thin-film cadmium telluride (CdTe) module depending on the influence of environmental parameters: solar radiation, air temperature, wind speed on the heating temperature of PVTB elements. A heat transfer scheme in PVTB elements, a mathematical model of heat balance processes and a program in the algorithmic Python language for calculating thermal and energy parameters were developed. The equations of thermal balances were solved in the static state of the installation using the numerical method of Gauss-Seidel iterations. To validate the model, the calculated simulation results were compared with experimental data. A comparison of the values of the corresponding parameters is presented in graphical images. The adequacy of the model was verified by the method of "assessment of agreement and standard deviation". The following values of RMSE and determination coefficients were calculated for the parameters, respectively, for the temperature of the photovoltaic cell, liquid and electrical efficiency R2=0.88, for thermal efficiency and power R2=0.99, for power RMSE=2.56 W, RMSE deviation 5.72%. Using the SolidWorks Flow Simulation software was performed CFD PVTB modeling, simulating the temperature dynamics of glass, PV and absorber depending on environmental factors and the temperature dynamics of cooling water circulating in a copper heat exchanger. The values of the parameters obtained from CFD modeling correspond to the experimental data. The considered modeling methods make it possible to calculate the thermal and energy parameters of PVT installations at the engineering design stage.

1 INTRODUCTION

Solar energy is a potential source for the production of environmentally electrical and thermal energy. Usually photovoltaic modules (PV) convert about 15-20% of solar energy into electrical energy the remaining 80-85% of the energy is dissipated into heat. As a result of heat accumulation, the PV will begin to lose efficiency and productivity. The use of photovoltaic thermal batteries (PVTB) technology is an acceptable way for cooling of PV and using the accumulated heat to generate heated water [1], [2], [3], [4], [5], [6]. PVTB, by removing heat through the circulation of liquid through a thermal battery, cooling of PV and increase the efficiency of solar energy conversion and also simultaneously produce electrical energy and heated water. PVT research focuses on technology development for improve design solutions and the efficiency of solar energy conversion [7], [8], [9]. Reducing the cost of installations and the cost of a unit of energy produced is an important task of photovoltaic thermal installation technology. To perform such tasks, we have developed a PVTB design based on CdTe thin-film modules. In this paper is considered mathematical and CFD modeling of the calculation of thermal and energy parameters of the PVTB.

2 METHODOLOGY

The thermal and energy parameters of PVT can be calculated using modeling. For this will develop a scheme of thermal processes in PVT battery layers and a mathematical model. Will write the heat balance equations for each PVTB layer depending on environmental factors: solar radiation, ambient temperature, wind speed, geometric and thermo-physical properties of materials, liquids. The assumptions and limitations made during the development of mathematical and CFD modeling as follows:

- heat is not lost at the edges of PVTB;
- the thermo-physical properties of the materials remain constant over the area;
- the effect of dust or partial shading is not taken into account;
- the fluid flow rate is constant, 0.002 kg/s.

The initial temperature of the liquid at the inlet at time t=0 is T_{in} =20°C. The PVTB installation consists of a CdTe-based thin-film PV technology and a thermal battery in the form of a "sheet-tube" structural assembly. An absorber made of aluminum sheet in the form of a box is attached to the back side of the PVM, inside which heat-exchange copper tubes covered with thermal insulation material are placed. The "sheet-pipe" design of the PVTB is shown in Figure 1.
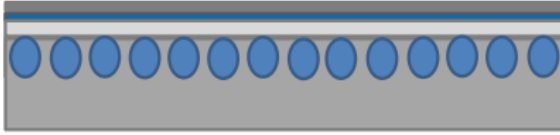
Figure 1: Cross-section of the PVTB "sheet-tube" design.

The research of the authors of these works relates to the development of a model of the PVTB, which can assess the suitability of various design parameters in any environmental conditions, which can lead to improved performance of the modules. Various configurations of PVTB have been developed and their performance characteristics have been studied depending on the optical, geometric and thermo-physical properties of materials, type of liquid, wind speed and direction, solar radiation and ambient temperature.
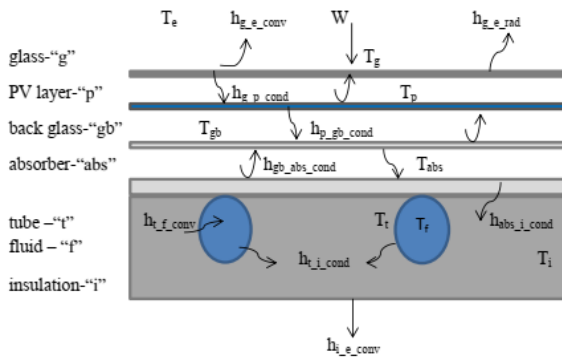
Figure 2: Scheme of heat transfer coefficients in PVTB layers according to the "sheet-pipe" design in the section.

The name of the designations in the scheme in Figure 2 as follow:

$h_{g\_e\_conv}$ – coefficient of convective heat transfer from the front glass to the environment, $h_{g\_e\_rad}$ – coefficient of radiative heat transfer between the front glass and the environment, $h_{g\_p\_cond}$ - heat transfer coefficient between the front glass and the photovoltaic layer, $h_{p\_gb\_cond}$ - heat transfer coefficient between the photovoltaic (PV cell) layer and the back glass, $h_{gb\_abs\_cond}$ - heat transfer coefficient between the rear glass and the absorber, $h_{abs\_i\_cond}$ - heat transfer coefficient between absorber and insulation, $h_{t\_f\_conv}$ – coefficient of convective heat transfer between tube and fluid, $h_{t\_i\_conv}$ – heat transfer coefficient between absorber and insulation, $h_{i\_e\_conv}$ - coefficient of convection between insulation and the environment, W – the density of the solar radiation flux incident on the front surface of the glass, T_e – temperature of environment or ambient-Ta, T_g -front glass temperature, T_p - temperature of the photoelectric layer, T_{gb} – rear glass temperature, T_{abs} – absorber temperature, T_t – tube temperature, T_f – fluid temperature, T_i – insulation temperature, T_{sky} – equivalent radiation temperature of the sky.

2.1 Mathematical Modeling

The mathematical model is based on the heat transfer processes described in Figure 2. The equations of thermal balances for the elements of the PVTB will write down. Similar processes were modeled in [10], [11], [12]. As is known from thermal engineering and solid state physics, when heat is applied to a body, its internal energy changes, which is described by the following:

$$Mc\frac{dT}{dt} = \frac{dU}{dt}.$$

M- body weight, kg; c – body heat capacity, J/kg, K; T - body temperature, K; U - internal energy of the body, J.

Thermal balance between the front glass and the photovoltaic (PV) layer:

$$M_g c_g \frac{dT_g}{dt} = h_{g-e-conv} S_g (T_e - T_g) + h_{g-e-rad} S_g (T_{sky} - T_g) + h_{g-p-cond} S_g (T_p - T_g) + + S_g \alpha_g W. \quad (1)$$

From (1) it is possible to express the temperature of the front glass (2):

$$T_g = \frac{h_{g-e-conv} T_e + h_{g-e-rad} T_{sky} + h_{g-p-cond} T_p + \alpha_g W}{h_{g-e-conv} + h_{g-e-rad} + h_{g-p-cond}}, (2)$$

where:

$$h_{g-e-conv} = \begin{cases} 5,7 + 3,8 \cdot v_w \text{ ,для } v_w < 5 \text{ м/}c \\ 6,47 + v_w^{0,78} \text{ ,для } v_w \geq 5 \text{ м/}c \end{cases}$$

$$h_{g-e-rad} = \varepsilon_g \cdot \sigma(T_g^2 + T_{sky}^2)(T_g + T_{sky}),$$

where:

$$T_{sky} = 0,00552 \cdot T_e^{1,5}$$

$$h_{g-p-cond} = \frac{1}{\frac{\delta_g}{k_g} + \frac{\delta_p}{k_p}}$$

ε_g- glass emissivity coefficient, 0,88; δ_g –thickness of glass, 0,0032 m and δ_p - thickness of PV, 0,0004 m, σ – Stefan-Boltzmann constant, $5,67 \cdot 10^{-8} \, W/m^2 \cdot K^4$, $S_g, S_{gb}, S_p, S_{abs}$ - areas of the front, rear glass, PV layer, absorber, 0,72 m^2,

$$h_{g-p-rad} = \frac{1}{\frac{1}{\varepsilon_g} + \frac{1}{\varepsilon_p} - 1} \cdot \sigma(T_g^2 + T_p^2)(T_g + T_p),$$

where: ε_p- the emissivity coefficient of the PV layer, 0,10; $(\alpha\tau)_p$-absorption and transmission coefficient of solar radiation of the PV layer, $\alpha = 0,9$; $\tau = 0,93$; η_{el} – electrical efficiency of the PV layer;

$$\eta_{el} = \eta_{stc}[1 - \beta_p(T_p - T_{stc})].$$

The equation of the thermal balance between the PV layer and the rear glass:

$$M_p c_p \frac{dT_p}{dt} = h_{p-g-cond}S_p(T_g - T_p) + \\ + h_{p-gb-cond}S_{gb}(T_{gb} - T_p) + (\alpha\tau)_p \cdot S_p \cdot \quad (3) \\ W(1 - \eta_{el}).$$

From (3) express:

$$T_{gb} = \frac{h_{g-p-cond}T_p + h_{gb-abs-cond}T_{abs}}{h_{gb-p-cond} + h_{gb-abs-cond}}, \quad (4)$$

$$h_{p-g-cond} = h_{p-gb-cond},$$

$$h_{p-g-cond} = h_{p-gb-cond} = \frac{1}{\frac{\delta_{gb}}{k_{gb}} + \frac{\delta_p}{k_p}}.$$

The equation of thermal balance between the rear glass and the absorber:

$$M_{gb} c_{gb} \frac{dT_{gb}}{dt} = h_{gb-p-cond}S_{gb}(T_p - T_{gb}) + \\ h_{gb-abs-cond}S_{gb}(T_{abs} - T_{gb}). \quad (5)$$

From (5) express:

$$T_{gb} = \frac{h_{g-p-cond}T_p + h_{gb-abs-cond}T_{abs}}{h_{gb-p-cond} + h_{gb-abs-cond}}, \quad (6)$$

where: $h_{gb-abs-cond} = \frac{1}{\frac{\delta_{gb}}{k_{gb}} + \frac{\delta_{abs}}{k_{abs}}}$,

k_g, k_{gb}, - thermal conductivity of glases, 1,1 W/m·K and k_p - thermal conductivity of PV,140 W/m·K.

Heat balance equation between absorber and tube:

$$M_{abs} c_{abs} \frac{dT_{abs}}{dt} = h_{abs-t-cond}S_{abs-t}(T_t - \\ - T_{abs}) + h_{abs-i-cond}S_{abs-i}(T_a - T_{abs}) + \quad (7) \\ + h_{abs-gb-cond}S_{abs}(T_{gb} - T_{abs})$$

From (7) express:

$$T_{abs} = \frac{h_{abs-t-cond}S_{abs-t}T_t + h_{abs-i-cond}S_{abs-i}T_a + h_{abs-gb-cond}S_{abs}T_{gb}}{h_{abs-t-cond}S_{abs-t} + h_{abs-i-cond}S_{abs-i} + h_{abs-gb-cond}S_{abs}} \quad (8)$$

$$h_{abs-t-cond} = \frac{1}{\frac{\delta_{abs-t}}{2k_{abs-t}}}$$

$$\delta_{abs-t} = \frac{X - D_{ext}}{4}$$

$$S_{abs-t} = \delta_{abs}L$$

k_{abs-t} - thermal conductivity of the absorber, 237 W/m·K; δ_{abs}- thickness of the absorber, 0,0006m; L - the length of the absorber,10,5 m; X is the pitch of the tube arrangement, 0,058 m,

$$h_{abs-i-cond} = \frac{1}{\frac{\delta_{abs-i-cond}}{2k_{abs-i-cond}}}$$

$$S_{abs-i} = S_{abs}\left(\frac{W1 - D_{ext}}{W1}\right)$$

$$h_{abs-gb-cond} = \frac{1}{\frac{\delta_{gb}}{k_{gb}} + \frac{\delta_{abs}}{k_{abs}}}$$

The equation of the thermal balance between the tube and the coolant-fluid:

$$M_t c_t \frac{dT_t}{dt} = h_{t-f-conv}S_{t_{int}}(T_f - T_t) + \\ + h_{t-abs-cond}S_{abs-t}(T_{abs} - \\ - T_t) + h_{t-i-cond}S_{t\_i}(T_i - T_t) \quad (9)$$

$$h_{t-f-conv} = \frac{Nu_f k_f}{D_{t.in}},$$

From (9) express:

$$T_t = \frac{h_{t-f-conv}S_{t\_int}T_f + h_{t-abs-cond}S_{abs-t}T_{abs} + h_{t-i-cond}S_{t\_i}T_i}{h_{t-f-conv}S_{t\_int} + h_{t-abs-cond}S_{abs-t} + h_{t-i-cond}S_{t\_i}} \quad (10)$$

where Nu_f – Nusselt number; $D_{t.in}$ – inner diameter of the tube, m.

$$Nu_f = 4.36, \ \ Re_f \leq 2300,$$
$$Nu_f = 0.023(Re_f^{0.8}Pr_f^{0.4}), Re_f > 2300,$$

Re - Reynolds number; Pr- Prandtl number.

$$h_{abs-t-cond} = \frac{1}{\frac{\delta_t}{2k_t}},$$

$h_{t-i-cond} = \frac{1}{\frac{\delta_i}{2k_i}}$ by, $\delta_i \gg D_{t.ext}$, δ_t, δ_i – tube wall

thickness and insulation layer, m; k_t and k_i- thermal conductivity coefficients of the tube and insulation material; $D_{t.ext}$ – external diameter of the tube, m.

$S_{t\_i} = (\frac{\pi}{2} + 1)D_{t.ext}L$ –internal area of the tube, m²

The equation for the coolant (fluid):

$$M_f c_f \frac{dT_f}{dt} = h_{t-f-conv}S_{t\_int}(T_t - T_f) - \dot{m}_f c_f(T_{f_{out}} - T_{fin}), \quad (11)$$

$$T_f = 0.5(T_{f_{out}} + T_{fin}),$$

$$h_{t-f-conv} = \frac{Nu_f k_f}{D_{t.in}},$$

Thermal balance equation of the insulation layer:

$$M_i c_i \frac{dT_i}{dt} = h_{abs-i-cond}S_i(T_{abs} - T_i) + h_{t-i-cond}S_{t\_i}(T_t - T_i) + h_{i-e-conv}S_i(T_e - T_i) \quad (12)$$

From (12) express:

$$T_i = \frac{h_{abs-i-cond}S_iT_{abs} + h_{t-i-cond}S_{t\_i}T_t + h_{i-e-conv}S_iT_e}{h_{abs-i-cond}S_i + h_{t-i-cond} + h_{i-e-conv}S_i} \quad (13)$$

$$h_{i-e-conv} = \begin{cases} 5,7 + 3,8 \cdot v_w \text{ ,для } v_w < 5 \ m/s \\ 6,47 + v_w^{0,78} \text{ ,для } \ v_w \geq 5 \ m/s \end{cases}$$

$$h_{t-i-cond} = \frac{1}{\frac{\delta_{abs-i-cond}}{2k_{abs-i-cond}}}$$

$$h_{abs-i-cond} = \frac{1}{\frac{\delta_{abs-i-cond}}{2k_{abs-i-cond}}}$$

$$S_{t\_i} = (\frac{\pi}{2} + 1)D_{t.ext}L \ ,$$

The thermal efficiency of the installation can be calculated by the following expression:

$$\eta_{th} = \frac{\dot{m}_f c_f(T_{f_{out}} - T_{fin})}{S_{pvtb}W},$$

where, $\dot{m}_f$ – fluid flow rate, 0,002kg/s; c_f – heat capacity of the fluid, 4180J/kg K; $T_{f_{out}}$ – temperature of the fluid at the inlet to the battery, K; T_{fin} - temperature of the fluid at the outlet of the battery, 273,15 K. According to formulas respectively, the changes in the electrical and thermal efficiency of the PVTB were calculated, the values of which are derived in the form of graphs of graphs depending on T_r - the specific temperature decrease relative to the solar radiation flux density, which is expressed as follows.

$$T_r = \frac{T_{fin} - T_a}{W},$$

where, T_a – temperature of ambient.

3 RESULTS AND DISCUSSIONS

According to this mathematical model, a program was compiled in the algorithmic Python language (Spider 3.8.10-version) to calculate the corresponding parameters of thermal processes of HTP. Using the values of environmental parameters and some thermal parameters, geometric dimensions, the program solved the system of equations from (1) to (13) numerically using the Gauss Seidel iteration method and found the temperature values T_g, T_p, T_{abs}, T_f of the corresponding layers and the energy parameters of PVTB. According to the calculated simulation data, graphs of the corresponding parameters were plotted in Figure 3-7.
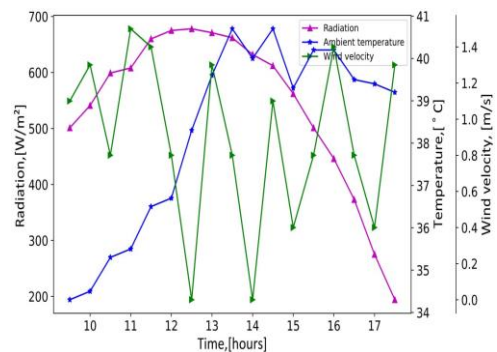
Figure 3: Ambient parameters.
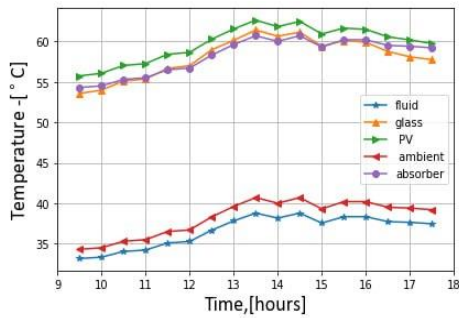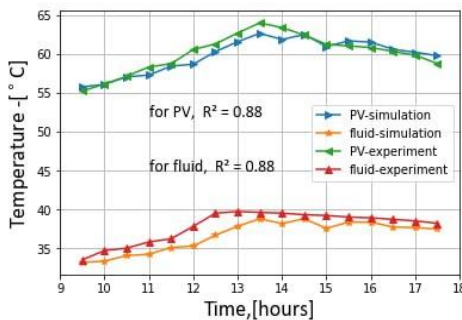
Figure 4: Temperature dynamics.

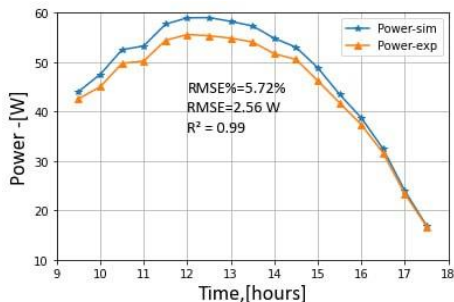Figure 5: Comparison of values of modeling and experimental temperatures.

Figure 6: Comparison of values of modeling and experimental powers.
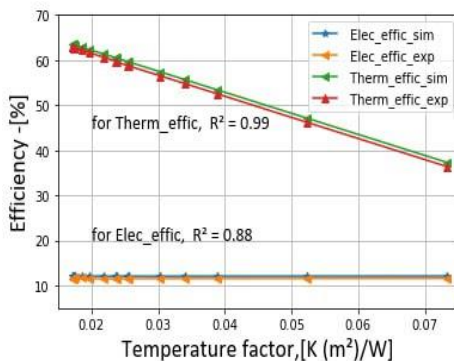

Figure 7: Graphs of efficiency.

Model were out validate with the experimental data by the temperatures, powers, efficiency of the PVTB of July 27, 2022. The accuracy of the proposed model at each verification level was evaluated using the methods of "agreement and Root Mean Square Error (RMSE) estimation". Therefore, we use a general analysis based on the coefficient of determination (R^2) and RMSE.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_{e_i} - P_{m_i})}{\sum_{i=1}^{n}(P_{e_i} - \bar{P}_{e_i})}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(P_{e_i} - P_{m_i}\right)^2}{n}} \text{ or}$$

$$RMSE\% = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(P_{e_i} - P_{m_i}\right)^2}{n}}}{\frac{\sum_{i=1}^{n} P_e}{n}} \cdot 100\%$$

where: n - number of measured data; P_e and P_m- electric power of the PVTB, experimental and simulation-based model; $\bar{P}_e$ - the average power value determined by the experiment. Comparison of results were out by RMSE and coefficient of determination and is R2 =0,88…0,99.

4 CFD MODELING

For researching and engineering estimates of PVT installations is we execute a CFD(Computational Fluid Dynamics) simulation for analysis of the temperature distribution in the elements of the same PVTB, which studied by experiments and modeling. Using the SolidWorks.2020.SP1.0, Flow Simulation software, CFD calculation of the temperature of the layers of the PVTB structure and the coolant-water at the exit from the thermal battery was performed. This program makes it possible to predict the performance of the structure, allows calculations at different values of input parameters. The program makes it possible to avoid numerous experiments and make accurate measurements. CFD calculation data can be used to create prototypes of the PVT structure and evaluate these structures at the project implementation stage and save time and finances. CFD calculation also allows optimization of the design parameters and operation of the prototype during its testing. The figures show the results of the CFD analysis in solar flux density 800W/m², mass flow rating of fluid 0.002 kg/s.
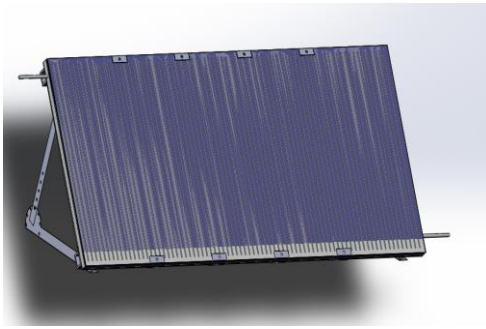
Figure 8: 3D drawing of PVT battery.
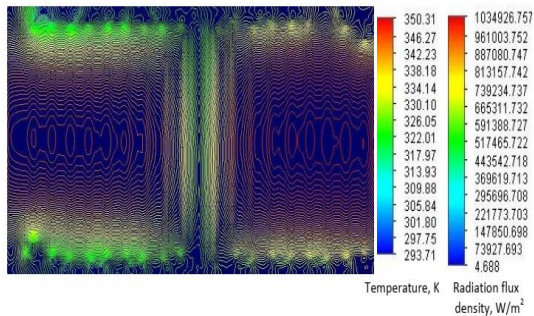
Figure 9: Temperature dynamics of glass.
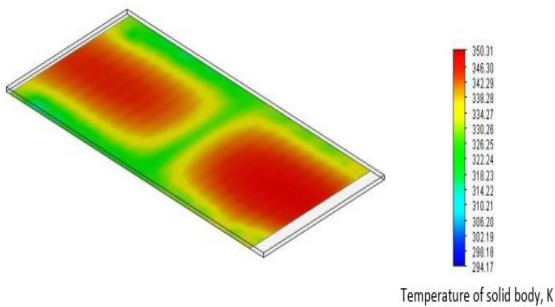
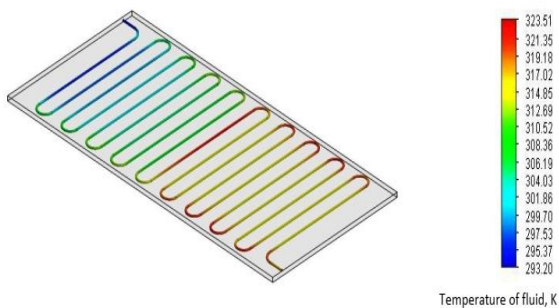Figure 10: Temperature dynamics of absorber.

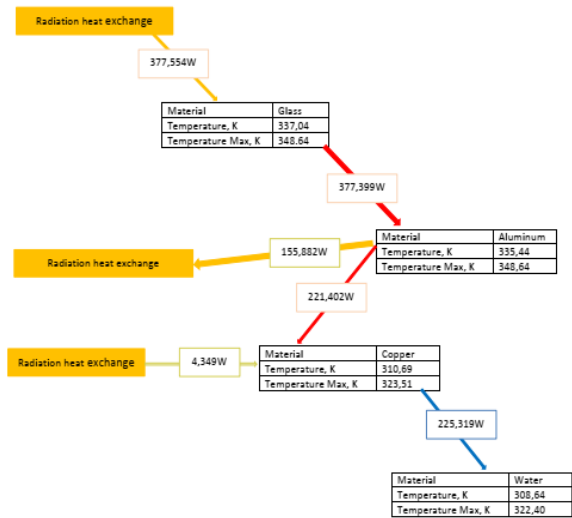Figure 11: Temperature dynamics of water.

Figure 12: Temperature distribution in PVTB elements.

In Figures 8-12 presented dynamics of temperatures in elements of PVTB. Changes of values of temperature by materials give in Table 1. Average values of the temperature of the elements of the PVTB at the peak time of the sunshine obtained by the calculation correspond to the temperature values obtained by the mathematical model and experimental results.

Table 1: Temperature values of PVTB elements.

Materials	Temperature		Temperature Max,	
	K	^{0}C	K	^{0}C
Glass of PV	337,04	63,89	348,64	75,49
Aluminum plate (absorber)	335,44	62,29	348,64	75,49
Copper (tube exchanger)	310,69	37,54	323,51	50,36
Fluid (water)	308,64	35,49	322,40	49,25

5 CONCLUSIONS

According to the mathematical model, a program was written in the algorithmic Python language, according to which calculations were carried out by the numerical method of the corresponding parameters. The calculated values of temperature changes across the layers of the PVTB structural elements, power, thermal and electrical efficiency

of the installation are shown graphically. Comparison the values of a few parameters calculated by simulation with experimental data showed the reliability of the model with a determination coefficient by power, thermal efficiency is 0,99, temperature, electrical efficiency and fluid is 0.88. This model will be used to calculate the energy and thermal parameters of the PVTB for climatic conditions of different geographical areas. CFD model also easy for calculate of temperatures of elements of PVT installations.

ACKNOWLEDGMENTS

REFERENCES

[1] R. Schmalensee, S. Alsaqoor, A. Alqatamin, A. Alahmer, Z. Nan, Y. Al-Husban, and H. Jouhara, "The impact of phase change material on photovoltaic thermal (PVT) systems: A numerical study," International Journal of Fluids and Thermal Engineering, 2023, [Online]. Available: https://doi.org/10.1016/j.ijft.2023.100365.

[2] O. Badran, A. Alahmer, F.A. Hamad, Y. El-Tous, G. Al-Marahle, and H.M.A. Al-Ahmadi, "Enhancement of solar distiller performance by photovoltaic heating system," International Journal of Thermofluids, vol. 18, 2023, [Online]. Available: https://doi.org/10.1016/j.ijft.2023.100315.

[3] A. Alahmer and S. Ajib, "Solar cooling technologies: state of art and perspectives," Energy Conversion and Management, vol. 214, 2020, [Online]. Available: https://doi.org/10.1016/j.enconman.2020.112896.

[4] M.H. Ahmadi, M. Ghazvini, M. Sadeghzadeh, et al., "Solar power technology for electricity generation: A critical review," Energy Science & Engineering, vol. 6, pp. 340-361, 2018 , [Online]. Available: http://doi.org/10.1002/ese3.239.

[5] M. Barbu, G. Darie, M. Siroux, "Analysis of a residential photovoltaicthermal (PVT) system in two similar climate conditions," Energies, vol. 12, pp. 1-18, 2019 , [Online]. Available: http://dx.doi.org/10.3390/en1219359.

[6] A. Ramos, I. Guarracino, A. Mellor, D. Alonso-Alvarez, P. Childs, et al., "Solar-Thermal and Hybrid Photovoltaic-Thermal Systems for Renewable Heating," no. 22, Imperial College London, May 2017, [Online]. Available: http://doi:10.13140/RG.2.2.10473.29280.

[7] N. Aste, F. Leonforte, and C. Del Pero, "Design, modeling and performance monitoring of a photovoltaic-thermal (PVT) water collector," Solar Energy, vol. 112, pp. 85-99, 2015, [Online]. Available: https://doi.org/10.1016/j.solener.2014.11.025.

[8] R. Liang, C. Zhou, Q. Pan, and J. Zhang, "Performance evaluation of sheet-and-tube hybrid photovoltaic/thermal (PVT) collectors connected in series," Procedia Engineering, vol. 205, pp. 461-468, 2017, [Online]. Available: https://doi:org/10.1016/j.proeng.2017.10.411.

[9] H. Haloui, K. Touafek, M. Zaabat, et al., "The Copper Indium Selenium (CuInSe2) thin films solar cells for Hybrid Photovoltaic Thermal Collectors (PVT)," Energy Procedia, vol. 74, pp. 1213-1219, 2015, [Online]. Available: https://doi.org/10.1016/j.egypro.2015.07.765.

[10] S. Bhattarai, J.-H. Oh, S.H. Euh, and G. Kafle, "Simulation and model validation of sheet and tube type photovoltaic thermal solar system and conventional solar collecting system in transient states," Solar Energy Materials and Solar Cells, vol. 103, pp. 184-193, August 2012 , [Online]. Available: https://doi.10.1016/j.solmat.2012.04.017.

[11] M. Barbu, M. Siroux, and G. Darie, "Numerical model and parametric analysis of a liquid based hybrid photovoltaic thermal (PVT) collector," Energy Reports, 2021 , [Online]. Available: https://doi.org/10.1016/j.egyr.2021.07.058.

[12] H. Haloui, K. Touafek, M. Zaabat, and A. Khelifa, "Comparative Study of the Hybrid Solar Thermal Photovoltaic Collectors Based on Thin Films Solar Cells in South of Algeria," [Online]. Available: https://doi.oa.mg/work/2787765823.

Effectiveness of Photoelectric Systems Against Intensive Gardening and Desertification

Muhammadjon Tursunov[1], Khabibullo Sabirov[1], Tohir Axtamov[1], Maxamadi Chariyev[2],
Umirbek Abdiyev[2], Boysori Yuldoshov[2] and Sirojiddin Toshpulatov[2]

[1]*Physical-Technical Institute, Chingiz Aytmatov Str. 2B, 100084 Tashkent, Uzbekistan*
[2]*Termez State University, Barkamol avlod Str. 43, 190111 Termez, Uzbekistan*
muhammadtursunov54@gmail.com, sabirovhabibullo@gmail.com, axtamovtz@uzsci.net, chariyevmaxamadi@gmail.com,
umr79@mail.ru, b.yuldoshov10@mail.ru, *toshpulatovs@tersu.uz*

Abstract: The increase in the efficiency of photoelectric panels from year to year is the basis for the large-scale use of this type of renewable energy source. Many subsidies are allocated for the use of renewable energy sources in our country. The use of solar energy, which is one of the renewable energy sources, is one of the most optimal solutions for creating intensive gardens in desert areas. In the article, the efficiency of the photoelectric system (PES) was compared with the portable generator (PG) used to supply water pumps with energy. It was determined that the payback period of PES is equal to 3.5 years compared to PG. It was found that there is an opportunity to use long-term green energy at almost no cost for the remaining 15-20 years. Instead of a system with a total power of 5kW, a PES system of 9.9kW was connected as an energy supply for irrigation works. 4.2 tons of CO_2 emissions from PG are avoided annually. This practical work proved that the most problematic desertification, water supply and green energy sources can be solved in one place.

1 INTRODUCTION

Today, critical situations are occurring all over the world. These are global warming, air pollution all over the world, especially in cities, desertification, water supply and many other factors. Among the projects aimed at mitigating such processes at least a little, the work done using the photoelectric system is commendable. The application of photoelectric systems in the necessary directions is bearing fruit all over the world. Uzbekistan has set a plan to increase the amount of energy obtained from renewable energy sources to 40% by 2030 [1]. It should be noted that this was a move typical of European countries in this direction. One of the most effective solutions to desertification is the creation of intensive gardens. In Uzbekistan, as well as in the whole world, strategic goals are being set against desertification and measures are being taken to prevent it. Since more than 70% of the territory of Uzbekistan corresponds to desert and waterless areas [2], it is clear that it is necessary to carry out

specific targeted work in this regard. It is appropriate to use drip irrigation in the organization of intensive gardens in the desert region [3]. The durability and degradation properties of polymer materials used in drip irrigation are also important [4]. In addition, a source of energy is needed to manage water in the desert and to draw and distribute water from wells through pumps.

An AC power source is cost-effective and requires minimal maintenance when there is a power grid nearby. However, in places where there is no grid, they are forced to use diesel or similar fuel generators to power water pumps. The design of this type of power source is compact, convenient, easy to install and portable. One of their major disadvantages is constant refuelling and maintenance work. In addition, fuel prices are high and unavailable in many rural and desert areas. One of the problems is the constant supply of fuel and the constant increase in fuel prices. For this reason, the rational use of solar energy in such areas is widely promoted worldwide. Photoelectric systems (PESs)

are a safe, reliable, maintenance-free, noise-free and environmentally friendly power source in desert areas [5]. There are also works comparing the efficiency of diesel and PES mentioned above [6]. Of course, when installing photoelectric systems (PES), the power of all pumps and other consumers is considered. Based on the maximum daily energy consumption and the suitability of the system for the purpose, the optimal PES is set.

Degradation of many materials is high in desert areas due to high temperatures and radiation. Therefore, it is recommended to make high-quality backup systems.

2 METHODS AND MATERIALS

Our experiment was conducted in the Surkhandarya region, located in the southernmost part of Uzbekistan. The main goal of the work is to evaluate the technical and economic efficiency of PES, which is being used as water supply on 10 hectares of land, which is being built in the desert area. It is planned that this area will become an intensive garden in the future, and a suitable part of the desert area was selected and a well was dug for the irrigation system at a depth of 35-37 meters. This water was subjected to technical expertise and it was found that it can be used for watering plants. Information about the general view of the studied area is given in Figure 1.

Figure 1: General view of the area.

Capillary hoses for drip irrigation have been laid for irrigating this 10 hectares of land, and they can be recycled at the end of their service life. 3 water storage tanks were dug (20m^3, 20m^3, 25m^3). The main two irrigation reservoirs are potentially above the irrigated land. The reason is to irrigate oneself in the evening without spending energy. Due to the high temperature during the day, watering has a bad effect on the plant. Bacterialization and heating of water in the open air are very beneficial for the plant. Water is pumped from a well to the reservoir in the lower part of the reservoir, and two reservoirs located above this reservoir are filled with water during the day using PES energy. PES overview with water tanks Figure 2.

Figure 2: PES system overview with water reservoirs.

The total power of PES is 9.9kW. Ipvisola panels were used in the system. The panels are installed at an angle of 18 degrees facing south. Panel characteristics are shown in Table 1.

A frequency inverter is used in the system. The reason is that water reservoirs are filled using daytime PES energy using three pumps (total power 3.5kW, 1.7kW, 1.1kW, 0.7kW). During the day, the pumps work using maximum PES energy. During irrigation, at night, irrigation works are carried out using the potential energy of water without energy

consumption. Before the PES was used, a Portable Generator (PG) was used and this PG was continuously supplied with fuel. PG has an average maximum power of 5 kW and consumes 1 litre of petrol per hour. The main aim of our work is to compare the efficiency of the systems in these two cases and to calculate the economic payback periods. In addition, it is to prove that desertification can be stopped at least a little by creating intensive gardens with water extraction.

Table 1: Geometric size and physical parameters of photoelectric battery.

Model Type	YH550W-36MN
Rated Maximum Power (P_{max})	550W
Power Tolerance	± 3 %
Open circuit voltage, U_{oc}	50.1V
Short circuit current, I_{sc}	13.9 A
Maximum Power Voltage (Vmp)	42.13V
Maximum Power Current (Imp)	13.06A
Dimensions, mm	2279·1134·35
FIK, %	21.51 %
PEB weight, kg	28,4kg
Maximum System Voltage Nominal	1500VDC
Maximum Series Fuse Rating	25A
E=1000W/m2, T_{PEB}=25ºC, AM=1,5 STC conditions	

Table 2: The total cost of PES and PG systems.

PES (Photoelectric System)		PG (Portable Generator)	
550W BEBs	$3960	PG	$800
Frequency inverter	$1000	Installation of connecting wires	$100
Construction and connecting wires	$1000		
Total cost for PEB:$5960		Total cost for PG: $900	

It should be noted that the irrigation and reclamation season begins in April and ends in early November at the highest level of solar radiation flow density. During this period, along with the high level of solar radiation flow density, the air temperature also rises to a high level, especially in the desert regions. Since there are few cloudy days during the irrigation season, the use of photoelectric systems in water extraction is effective, but the main drawback of photoelectric water extraction systems is the decrease in system efficiency due to the loss of PES power at high temperatures [7]. We will consider the energy efficiency of drip irrigation works in the case of PES and PG. In addition, we will dwell in detail on the payback period of these systems and their

impact on the environment. Table 2 shows the total cost of PES and PG systems.

For the PES system, the initial cost is almost 6.5 times higher than that of the PG system.

3 RESULTS AND DISCUSSION

Pumps used to irrigate 10 hectares of land in a desert area far from a power station definitely need a source of energy. Initially, the farmer used a 5kW PG as this energy source, but now 9.9kW PES is installed considering the scale of the system. Initial costs for PES are high. However, the 20-year warranty for the panels and almost negligible maintenance costs ensure the long-term operation of this system. PG consumes 1 litre of petrol per hour and spends $6 on petrol during an average working time of 6 hours per day. This average value is for spring, summer and autumn months. In addition, there is a problem of constant transportation of fuel for PG. In this case, the average cost of fuel for the envisaged 8-month period will be $1,440. Costs for maintaining the technical condition of PG and PES are not taken into account at this time. The economic payback period of PES compared to PG was calculated as follows (1)

$$N = \frac{PES_{Total\_cost} - PG_{Total\_cost}}{P_{Fuel\_cost}} \quad (1)$$

In this case, the PES system can justify itself in 3.5 years (Figure 3). The farm has also started a greenhouse project on 1 hectare of land for the winter season. Construction of an efficient greenhouse has been started to grow crops even on cold days. The aim is to use green energy for the electricity supply for this greenhouse.

The new garden and these green energy plants, which are being built against today's ever-expanding desertification, are certainly not without benefits for the environment. PES performance depends on radiation, PV panel surface cleanliness and module temperature [7-9]. For this purpose, the PEB station is installed in a convenient position for cleaning, 0.5 m above the ground at 18^0 degrees for convective cooling. The power obtained from the PEB is enough to ensure the energy consumption of 3 pumps with a total power of 3.5 kW for full operation during the day. In addition, there is reserve capacity for consumers who may be connected in the future. If it is planned to be used by consumers even at night, an accumulator battery and an inverter are

necessary. And for this change, there is no need for a change in the power of PES. Table 3 presents the energy values that can be obtained from PES for months.

Figure 3: Greenhouse construction.

Table 3: The energy values that can be obtained from PES for months.

Winter	Spring	Summer	Autumn	Total annual energy
3.74 MW·h	5.54 MW·h	6.24 MW·h	5.54 MW·h	21.06 MW·h

So, based on theoretical calculations, when used in the maximum state, it can be said that such a PES provides 21 MW•h of energy per year. In addition, 0.7 kg of CO2 gas released into the environment due to PES is avoided for every 1kW•h of energy [10]. When one litre of petrol is burned, it creates approximately 2.31 kg of carbon dioxide [11]. This prevents 4.2 tons of CO2 from escaping into the atmosphere per year.

4   CONCLUSIONS

Against the background of technology development, PEB electric efficiency increased by 22%, and PESs became an energy supply system in various fields. The role of such energy systems in providing energy to water-releasing devices in the agricultural sector is extremely important. In the farm analyzed above, replacing the energy supply with PES is an effective solution, which avoids the cost of daily fuel delivery and fuel prices. The payback period of PES compared to PG is 3.5 years. In addition, 4.2 tons of CO2 released into the atmosphere per year will be avoided. Using PES in energy supply while reducing desertification by creating a garden supports green energy capacities and guarantees long-term performance. It was found that there is an opportunity to use long-term green energy at almost no cost for the remaining 15-20 years. Instead of a system with a total power of 5kW, a PES system of 9.9kW was connected as an energy supply for irrigation works and an additional power source was also created. Our practical work proved that the most problematic desertification, water supply and green energy sources can be solved in one place.

REFERENCES

[1] Renewable energy sources, [Online]. Available: https://lex.uz/uz/docs/6600413.

[2] Sh.S. Sayfiddinov and S.I. Ashurmahmatov, "Problems of desertification: dynamics, assessment, solutions," in International Conference, Samarkand, December 13-14, 2019, pp. 289-291.

[3] M.R. Goyal, "Management of drip/trickle or micro irrigation," Oakville, California: Apple Academic Press, p. 104, ISBN 978-1926895123, [Online]. Available: https://books.google.co.uz/books?id=c8TFYZd9cw4C.

[4] S.Y. Inagamova, U.A. Asrorov, and E.B. Xujanov, "Structure and physico-mechanical properties of polyelectrolyte complexes based on sodium carboxymethylcellulose polysaccharide and polyacrylamide," East European Journal of Physics, vol. 4, pp. 258-266, 2023, https://doi.org/10.26565/2312-4334-2023-4-32.

[5] L.M. Musanga, W.H. Barasa, and M. Maxwell, "The Effect of Irradiance and Temperature on the Performance of Monocrystalline Silicon Solar Module in Kakamega," Physical Science International Journal, vol. 19, no. 4, pp. 1-9, 2018; Article no. PSIJ.44862, ISSN: 2348-0130.

[6] D.H. Muhsen, T. Khatib, and H.T. Haider, "A feasibility and load sensitivity analysis of photovoltaic water pumping system with battery and diesel generator," Energy Conversion and Management, vol. 148, pp. 287-304, 2017.

[7] M.N. Tursunov, Kh. Sabirov, T.Z. Axtamov, U.B. Abdiyev, M.M. Chariyev, B.A. Yuldoshov, and S.F. Toshpulatov, "Capacity Utilization Factor (CUF) of the 70kW on-grid solar station in the dry climate of Termez," E3S Web of Conf. 401 02059 (2023), doi: 10.1051/e3sconf/202340102059.

[8] A. Nahar, M. Hasanuzzaman, N. Rahim, and S. Parvin, "Numerical investigation on the effect of different parameters in enhancing heat transfer performance of photovoltaic thermal systems," Ren. Ener., vol. 132, pp. 284-95, 2019, [Online]. Available: https://doi.org/10.1016/j.renene.2018.08.008.

[9] B. Yuldoshov, E. Saitov, J. Khaliyarov, S. Toshpulatov, and F. Kholmurzayeva, "Effect of Temperature on Electrical Parameters of Photovoltaic Module," Proceedings of International Conference on Applied Innovation in IT, 2023, vol. 11, no. 1, pp. 291-295.

[10] M. Nizami and W.W. Purwanto, "Solar PV based power-to-methanol via direct CO2 hydrogenation and H2O electrolysis: Techno-economic and environmental assessment," Journal of CO2 Utilization, vol. 65, 102253, 2022.

[11] Utilitysmarts, [Online]. Available: https://www.utilitysmarts.com/automobile/gasoline/how-much-co2-does-1-litre-of-petrol-produce/#google_vignette.

Advancing Solar Irradiation Prediction in Extreme Climates: A LASSO Regression Analysis in Tomsk

David Akpuluma[1,3] and Aleksey Yurchenko[1,2]

[1]*National Research Tomsk Polytechnic University, Lenin Avenue 30, 634050 Tomsk, Russia*
[2]*National Research Tomsk State University, Lenin Avenue 36, 634050 Tomsk, Russia*
[3]*Centre for Nuclear Energy Studies, University of Port Harcourt, East-West Road,*
PMB 5323 Choba, Port Harcourt, Nigeria
{aa06, reaper}@tpu.ru, david.akpuluma@uniport.edu.ng, akpoebi@gmail.com

Abstract: This study presents a robust approach to predicting solar irradiation in the challenging climatic conditions of Tomsk using LASSO regression, with a particular emphasis on interpretability and climatic variability. Two distinct models were developed: Model 1, integrating specific humidity at 2 meters, and Model 2, excluding this variable to assess the impact of a wider range of meteorological factors. The comprehensive meteorological dataset from NASA's POWER database underpinned the analysis. The models' efficacy was demonstrated by impressive R-squared values: 0.843 for Model 1 and 0.813 for Model 2, indicating a substantial proportion of variance in solar irradiation was captured. Notably, Model 1's RMSE of 0.0353 and Model 2's RMSE of 0.0386 affirm the precision of the predictions. The study advances the predictive modeling of solar power output, offering valuable contributions to renewable energy forecasting literature and operational practices by providing a methodological framework that is both accurate and comprehensible, even amidst the complexities of extreme weather patterns.

1 INTRODUCTION

In the wake of pressing global challenges such as climate change and resource scarcity, the United Nations' Sustainable Development Goal 7 (SDG7) underscores the imperative of universal access to sustainable energy, emphasizing renewable sources and energy efficiency as pivotal elements in the achievement of this goal [1].

Concurrently, transition engineering emerges as a crucial discipline, orchestrating the shift towards sustainability in energy systems [2]. This field meticulously blends engineering principles with long-term strategic planning to ensure that future energy demands are met in harmony with the environment, thus facilitating a robust and sustainable societal infrastructure [3].

The intersection of SDG7 and transition engineering's targets encapsulates the essence of modern energy strategies that are environmentally sound, economically feasible, and socially inclusive [4], [5].

Against the current backdrop, the utilisation of advanced machine learning techniques, particularly LASSO (Least Absolute Shrinkage and Selection Operator), marks a notable advancement in the field of predictive analytics for photovoltaic (PV) power generation. This paper introduces a pioneering case study in Tomsk, leveraging these sophisticated methods to enhance the accuracy of PV power predictions. LASSO, with its proven effectiveness in feature selection and its ability to manage overfitting in complex datasets, provides a robust analytical framework. This approach adeptly addresses the challenges associated with the variability and complexity of PV power output, thereby facilitating more precise and reliable solar energy forecasting. This methodological choice is especially pertinent in contexts requiring clear and transparent model-driven decisions, aligning with the increasing need for comprehensible and accountable predictive models in various sectors.

The purpose of this study is to showcase the efficacy of LASSO in precisely predicting PV power generation [6]. The use of LASSO, an interpretable

model, is particularly significant in contexts where legal considerations demand transparency and justification in decision-making processes. By employing LASSO, this research not only adheres to the core principles of SDG7 but also aligns with the objectives of transition engineering. This alignment enhances the reliability and efficiency of renewable energy resources, ensuring that the methods used for forecasting and analysis are both legally compliant and transparent. The importance of such interpretable models is underscored in environments where the decisions and predictions of AI-driven models must be clear and justifiable, especially in light of increasing regulatory scrutiny in the use of complex algorithms in various sectors.

The future scope of this research is expansive. It will delve into the integration of real-time data feeds to enhance model responsiveness, the exploration Explainable Artificial Intelligence (XAI) tools for broader interpretability spectrum, and the scalability of the proposed framework to other regions and renewable energy forms. By doing so, it aims to contribute significantly to the literature on renewable energy forecasting and the operational optimization of PV systems, thereby supporting the global endeavour towards a sustainable and resilient energy future.

2 METHODOLOGY

2.1 General Methods

In the field of photovoltaic power prediction, various methods are commonly used, each, with its strengths and weaknesses. Statistical techniques like the integrated moving average (ARIMA) model are known for their ability to estimate solar power effectively by examining the linear relationships in past data [7]. However, these approaches may not fully grasp the complexities in solar energy data. To address this, machine and deep learning methods have been increasingly applied, with a focus on data processing, feature extraction, and uncertainty evaluation[12], [13].

Machine learning methods such as Random Forest (RF) and Support Vector Machines (SVM) are lauded for their capability to identify linear connections between multiple input variables and solar energy production [8]. These algorithms excel at handling datasets with numerous variables, offering a deeper understanding of solar energy systems [14].

Nevertheless, their performance relies heavily on the quality and quantity of training data available which can be a limitation. Artificial Neural Networks (ANNs) including variants, like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have emerged as tools in solar power forecasting [9]. Their ability to model patterns and adapt to forecasting tasks is unmatched.

However, the lack of transparency, in Artificial Neural Networks (ANNs) can make it difficult for practitioners to understand how the models make decisions. Physical models, which are based on photovoltaic principles and solar geometry use factors like irradiance, panel positioning and temperature to predict energy output [10].

While these models provide insights into power generation, they may not be able to account for all real-world variables accurately leading to potential prediction errors. Hybrid models aim to combine the strengths of modeling approaches by integrating machine learning with methods. This combination seeks to improve prediction accuracy by harnessing the capabilities of machine learning and the foundational principles of physical models [11].

Although hybrid models show promise in enhancing forecasting accuracy, their complexity and reliance on data sources can be challenging. Ultimately choosing a power forecasting method depends on balancing factors like accuracy, interpretability and computational requirements based on task needs. The ongoing advancements in techniques and data availability are driving the improvement of these methodologies, for more reliable and efficient solar power forecasting.

2.2 LASSO Framework

The approach followed in this work focuses on two models, Model 1 and Model 2, for predicting solar irradiation in Tomsk using LASSO regression. Model 1 examines the relationship between all-sky surface UVB irradiance (allsky_sfc_uvb) and various independent meteorological variables. It highlights the significant influence of specific humidity at 2 meters (qv2m) on the predictive model. The performance of Model 1 is evaluated using metrics like RMSE, MSE, and R-squared, indicating a strong predictive capability.

Model 2, in contrast, explores the same dependent variable (allsky_sfc_uvb) but excludes the qv2m variable from its analysis. This model assesses the impact of other meteorological variables, including wind direction at 10 meters (wd10m) and wind speed at 10 meters (ws10m), on solar irradiation

predictions. Similar to Model 1, Model 2's effectiveness is gauged using RMSE, MSE, and R-squared metrics, which provide insights into its predictive accuracy. Both models demonstrate the utility of LASSO regression in handling complex datasets for solar irradiation forecasting, emphasizing the importance of specific meteorological factors in predicting solar power output. The flow diagram of the process can be seen in Figure 1.
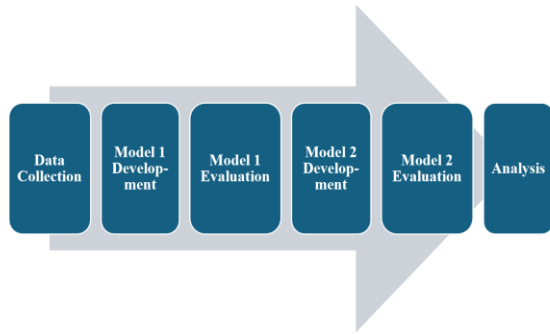

Figure 1: Flow diagram of the approach.

Loss function = OLS loss function

$$+ \lambda * \sum_{i=1}^{n} |a_i|. \qquad (1)$$

Equation (1) describes the loss function used in LASSO regression. It consists of two parts: the Ordinary Least Squares (OLS) loss function, which is the sum of the squared differences between the observed and predicted values, and a penalty term. The penalty is applied to the absolute values of the regression coefficients (represented by a_i), summed up across all coefficients (from i=1 to i=n). The term λ is a non-negative regularization parameter that controls the strength of the penalty. By increasing λ, the LASSO method can shrink less important coefficients to zero, effectively performing feature selection. This penalty encourages the model to maintain simplicity and prevent overfitting, leading to more interpretable models.

2.3 Data Source

Data for this work was accessed from NASA's POWER (Prediction of Worldwide Energy Resources) database. Monthly meteorological data for Tomsk from January 2001 to December 2020 was collected and prepared for input in the LASSO models. Tomsk, a city in Siberia, with the coordinates, approximately 56.5010 degrees latitude North and 84.9924 degrees longitude East was chosen

as the case study location due to its distinctive climatic and geographical characteristics, which present a unique opportunity to study solar irradiation patterns in a region with significant seasonal variations. The choice of Tomsk adds to the diversity and comprehensiveness of solar irradiation research, particularly in areas with extreme climatic conditions.

2.4 Feature Selection Analysis

In our research we conducted a feature selection analysis to identify the variables that impact how well our predictive model works. We looked at the dataset sourced from NASA POWER, with 260 data points across 9 variables as outlined in Table 1. This thorough examination was crucial in capturing the essence of our dataset, ensuring that only important variables with significant predictive power were included in the model. Our goal with this feature selection process was to improve the accuracy and clarity of the model setting a groundwork, for further analysis.

3 RESULTS AND DISCUSSION

3.1 Model 1

In Model 1, we analysed allsky_sfc_uvb vs other independent variables. The model shows in Figures 2 and 3 that qv2m was the most influential variable, followed by y2m.

Figures 4 and 5 are plots for visualisation of the results. Normal Q-Q (Quantile-Quantile) plots are graphical tools used to assess if a dataset follows a particular distribution, usually a normal distribution. If the points in the plot fall approximately along a straight line, it suggests the data are normally distributed. Series residual plots are used in regression analysis to visualize the residuals (differences between observed and predicted values) across the data series. These plots help to identify any patterns in the residuals, suggesting issues with the model, such as non-linearity, heteroscedasticity, or outliers.

The metrics used to evaluate the performance of a LASSO regression model are usually the root mean squared error (RMSE), the mean squared error (MSE) and the r squared (r^2). Here's an interpretation of each metric:

Delving into the performance of our Lasso regression model, we observe that it boasts a promising precision in its predictions. An RMSE of 0.0353 suggests that our model's forecasts are, on

average, only a small fraction off from the actual figures, which is quite commendable.

Further affirming its accuracy, the MSE stands at a minimal 0.00125, reflecting minor average errors in the predictions squared. Moreover, an R-squared value of 0.843 is noteworthy, indicating that the model can explain over 84% of the variability in the dependent variable—pointing to a robust model that captures the essence of the data well. These indicators collectively point towards a model that performs reliably and can be trusted for its predictive insights even in extreme weather climates like Tomsk.

Normal Q-Q Plot

Figure 4: Normal Q-Q plot of model 1.

Lasso Coefficients

Figure 2: LASSO analysis from model 1.

Series resid(lm(allsky_uvb ~ ., data = reg_data))
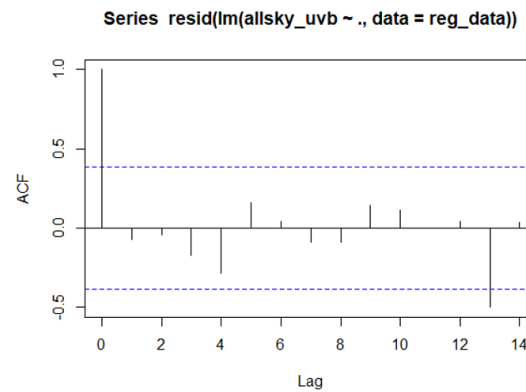
Figure 5: Series residual plot of model 1.

3.2 Model 2

In model 2 we excluded the specific humidity at 2 meters (qv2m) from the analysis and instead assessed the impact of other variables such as the temperature in degrees Celsius at 2 meters, wind direction at 10 meters (wd10m), and wind speed at 10 meters (ws10m) on solar irradiation forecasts. Computations were done for allsky_sfc_uvb vs other independent variables except qv2m. The model shows that y2m was the most influential variable, followed by ws10m and wd10m. The effectiveness of model 2 is evaluated using RMSE, MSE, and R-squared metrics, which provide insights into its predictive accuracy.

Lasso Coefficient Paths

Figure 3: LASSO coefficient paths for model 1.

Table 1: Characteristics and description of data set.

| Short name	Unit	Type
y2m	merra-2 temperature at 2 meters (c)	Independent
qv2m	merra-2 specific humidity at 2 meters (g/kg)	Independent
rh2m	merra-2 relative humidity at 2 meters (%)	Independent
wd10m	merra-2 wind direction at 10 meters (degrees)	Independent
ws10m	merra-2 wind speed at 10 meters (m/s)	Independent
allsky_sfc_uvb	ceres syn1deg all sky surface uvb irradiance (w/m^2)	Dependent
clrsky_sfc_par_tot	ceres syn1deg clear sky surface par total (w/m^2)	Dependent
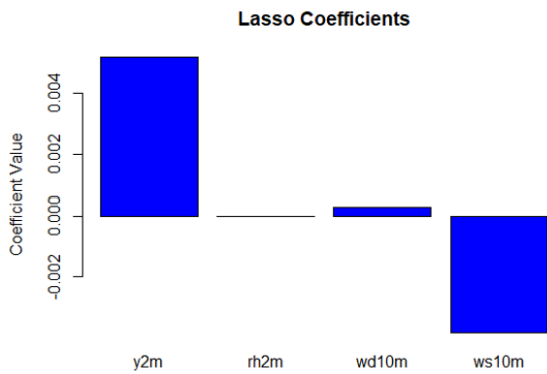
Figure 6: LASSO analysis from model 2.

In Figure 6, using LASSO regression we highlight the most important meteorological factors required for predicting power output. We can see the coefficients assigned to four variables: y2m, rh2m, wd10m and ws10m. The y axis displays the coefficient values selected by the LASSO model. This plot illustrates how each variable influences the prediction outcome. Specifically, the positive impact of temperature (y2m) on the prediction is evident. However, relative humidity at 2 meters (rh2m) does not strongly predict outcomes in this model. The small coefficient for wind direction at 10 meters (wd10m) suggests a minor relationship with the predicted value. Conversely a substantial negative coefficient for wind speed at 10 meters (ws10m) indicates an inverse relationship with the target variable.
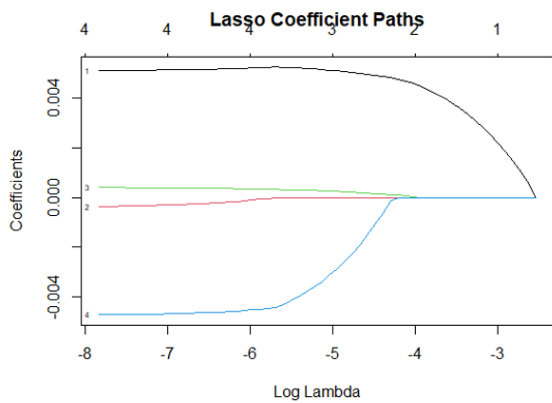
Figure 7: LASSO coefficient paths for model 2.

Figure 7 illustrates the LASSO coefficient paths for model 2, providing insight into the feature selection process as regularization is applied. Here, the x-axis, which represents the logarithm of lambda

(the regularization parameter), shows the trajectory of each feature's coefficient as the model complexity is varied. LASSO simultaneously performs feature selection and regularization to enhance the prediction model's robustness, particularly under conditions of multicollinearity or when a parsimonious model is desired [15].

The plot shows several paths corresponding to different coefficients, with their values converging towards zero as the log lambda increases. This convergence is indicative of the LASSO method's ability to shrink less important coefficients down to zero, effectively eliminating them from the model. Notably, the coefficient paths can also inform us about the relative importance and stability of the features across different regularization strengths; features with paths that quickly converge to zero are less robust, whereas those that remain non-zero at higher lambda values are more influential.

For a renewable energy application, such as solar irradiation prediction, the interpretability of this model is crucial. Figure 7 suggests that some features have a more consistent influence on the model's predictions, maintaining a non-zero coefficient across a range of lambda values. In contrast, features whose coefficients drop to zero more quickly are deemed less relevant. The key advantage of this approach is that it reduces the model's complexity and potential overfitting, leading to a model that is both interpretable and generalizable.
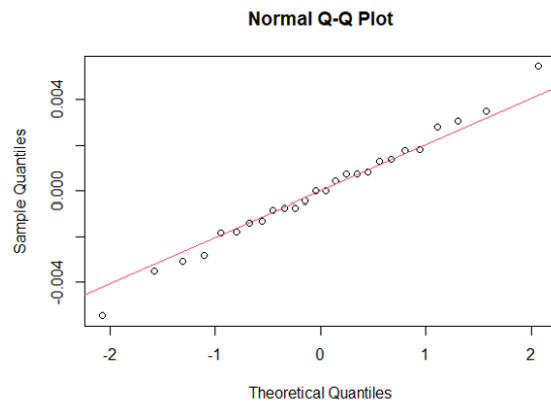

Figure 8: Normal Q-Q plot of model 2.

Figure 8 shows the Normal Q-Q plot of model 1 which, as mentioned earlier, is a tool used to check the normality of residuals, in regression analysis. This plot compares the quantiles to a standard normal distribution with the theoretical quantiles of the normal distribution on one axis and the sample quantiles of residuals on another. In a situation where

residuals follow a distribution the data points should align closely with a reference line. The alignment of points in Figure 8, along the line indicates that model 2's residuals adhere closely to a distribution suggesting that the model is well tuned, and its predictions are trustworthy.
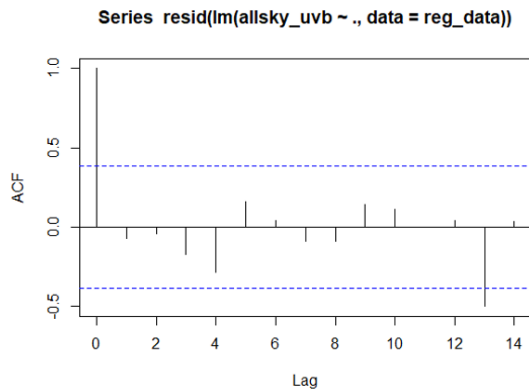

Figure 9: Series residual plot of model 2.

Figure 9 is the Autocorrelation Function (ACF) plot, for the residuals from model 2. It illustrates how the series correlates with itself at time lags. In a scenario where the model fits well, we would anticipate the autocorrelations at all lag points to fall within the confidence bounds (indicated by the dashed lines). This indicates no autocorrelation. It also shows that the model's residuals are random. The randomness signifies that the model has effectively captured all information from the data.

Our LASSO regression model's predictive accuracy is quantified through several statistical measures. The RMSE of 0.0386 highlights that our predictions deviate from the actual values by this small margin, indicating a tightly fitted model. The MSE, at 0.00149, reaffirms this, showing a minimal average of squared errors.

Impressively, the model accounts for approximately 81.3% of the variability in the target variable, as suggested by an R-squared of 0.813. Such a high R-squared value reflects the model's robustness in capturing the underlying data patterns.

To round up, these metrics underscore a strong predictive capability, suggesting the model is well-tuned to the nuances of our data.

4   CONCLUSIONS

The aim of the study was achieved through a systematic approach involving LASSO regression to forecast solar irradiation in Tomsk, which is a method particularly well-suited to handle the challenges posed by climatic variability. The study created two distinct models to interpret the complex relationships between meteorological factors and solar power output. Model 1 included specific humidity at 2 meters as an independent variable, while Model 2 excluded it, thus allowing for the analysis of other significant meteorological factors.

The effectiveness of these models was demonstrated by robust statistical metrics: Model 1 showed an R-squared value of 0.843, indicating that it could explain over 84% of the variability in the dependent variable, while Model 2 had an R-squared value of 0.813, accounting for approximately 81.3% of the variability in the target variable. These high R-squared values signify that both models have strong predictive capabilities and can reliably capture the underlying data patterns.

Furthermore, the use of the NASA POWER database provided a comprehensive set of meteorological variables, ensuring that the models had a solid data foundation to work from. The choice of Tomsk, with its distinctive climate, offered a unique case for examining solar irradiation patterns, contributing to the literature on renewable energy forecasting and the operational practices in the field.

In essence, the study succeeded in contributing to predictive modeling by developing interpretable models that effectively address climatic variability and demonstrate strong predictive performance, thus advancing the field of solar power output forecasting.

To further enhance the research presented, one could explore the integration of real-time data feeds to improve model responsiveness, use Explainable Artificial Intelligence (XAI) tools to broaden the interpretability spectrum, and test the scalability of the proposed framework across different regions and forms of renewable energy.

REFERENCES

[1]   "Transforming our world: the 2030 Agenda for Sustainable Development," UN Doc. A/RES/70/1, Sept. 25, 2015.

[2]   S. Krumdieck, Transition Engineering: Building a Sustainable Future. Taylor & Francis, 2020.

[3]   J. Stephenson, B. Barton, G. Carrington, A. Doering, R. Ford, D. Hopkins et al., "The energy cultures framework: exploring the role of norms, practices and material culture in shaping energy behaviour in New Zealand," Energy Research & Social Science, vol. 7, pp. 117-123, 2015, [Online]. Available: https://doi.org/10.1016/j.erss.2015.03.005.

[4] B. Sovacool, M. Burke, L. Baker, C. Kotikalapudi, and H. Wlokas, "New frontiers and conceptual frameworks for energy justice," Energy Policy, vol. 105, pp. 677-691, 2017, [Online]. Available: https://doi.org/10.1016/j.enpol.2017.03.005.

[5] D. McCollum, W. Zhou, C. Bertram, H. Boer, V. Bosetti, S. Busch et al., "Energy investment needs for fulfilling the Paris agreement and achieving the sustainable development goals," Nature Energy, vol. 3, no. 7, pp. 589-599, 2018, [Online]. Available: https://doi.org/10.1038/s41560-018-0179-z.

[6] N. Tang, S. Mao, W. Yu, and R. Nelms, "Solar power generation forecasting with a lasso-based approach," IEEE Internet of Things Journal, vol. 5, no. 2, pp. 1090-1099, 2018, [Online]. Available: https://doi.org/10.1109/jiot.2018.2812155.

[7] S. Pasari and A. Shah, "Time series auto-regressive integrated moving average model for renewable energy forecasting," Sustainable Production, Life Cycle Engineering and Management, pp. 71-77, 2020, [Online]. Available: https://doi.org/10.1007/978-3-030-44248-4_7.

[8] C. Voyant, G. Notton, S. Kalogirou, M. Nivet, C. Paoli, F. Motte et al., "Machine learning methods for solar radiation forecasting: a review," Renewable Energy, vol. 105, pp. 569-582, 2017, [Online]. Available: https://doi.org/10.1016/j.renene.2016.12.095.

[9] F. Niu and Z. O'Neill, "Recurrent neural network based deep learning for solar radiation prediction," Building Simulation Conference Proceedings, 2017, [Online]. Available: https://doi.org/10.26868/25222708.2017.507.

[10] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The national solar radiation data base (nsrdb)," Renewable and Sustainable Energy Reviews, vol. 89, pp. 51-60, 2018, [Online]. Available: https://doi.org/10.1016/j.rser.2018.03.003.

[11] V. Sharma and S. Chandel, "Performance and degradation analysis for long term reliability of solar photovoltaic systems: a review," Renewable and Sustainable Energy Reviews, vol. 27, pp. 753-767, 2013, [Online]. Available: https://doi.org/10.1016/j.rser.2013.07.046.

[12] S. Sobri, S. Koohi-Kamali, and N. Rahim, "Solar photovoltaic generation forecasting methods: a review," Energy Conversion and Management, vol. 156, pp. 459-497, 2018, [Online]. Available: https://doi.org/10.1016/j.enconman.2017.11.019.

[13] R. Ahmed, V. Sreeram, Y. Mishra, and M. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization," Renewable and Sustainable Energy Reviews, vol. 124, p. 109792, 2020, [Online]. Available: https://doi.org/10.1016/j.rser.2020.109792.

[14] C. Yen, H. Hsieh, K. Su, M. Yu, and J. Leu, "Solar power prediction via support vector machine and random forest," E3S Web of Conferences, vol. 69, p. 01004, 2018, [Online]. Available: https://doi.org/10.1051/e3sconf/20186901004.

[15] Y. Kim and J. Kim, "Gradient lasso for feature selection," Twenty-First International Conference on Machine Learning - ICML '04, 2004, [Online]. Available: https://doi.org/10.1145/1015330.1015364.

Ensure the Reliability of the Heating System of the Autonomous Helio-Greenhouse of the Trench Type in Emergency Situations

Bozorbek Botirov [1], Isroil Yuldoshev [1,2], Sanjar Shoguchkarov [1,2], Shahzod Rahmatillayev[1] and Afzalkhoja Kudratov[1]

[1]*Tashkent State Technical University named after Islam Karimov, Universitet, Str. 2, 100095 Tashkent,Uzbekista*
[2] *Tashkent University of Applied Sciences, Gavkhar Str. 1, 100081 Tashkent, Uzbekistan*

bbozorbek@bk.ru, yuldashev.i2004@gmail.com, sanjar42422@gmail.com, shahzodrahmatillayev12@gmail.com, afzalkhojakudratov@gmail.com

Abstract: In the article, the results of the temperature of the autonomous helio-greenhouse of the trench type in the abnormal cold climatic conditions of Uzbekistan are discussed. During the winter heating season at ambient temperatures from -40 °C to -21 °C, experiments were carried out, which showed that at an outdoor air temperature of less than -10 °C there was no need to use emergency heating sources, and the temperature in the autonomous greenhouse dropped to 5 °C. A two-circuit combined system based on an electric boiler and a solar water heating system has been proposed and applied as an emergency heating source. At an ambient temperature of more than -200 °C, the air temperature wizened at ~8-8,5 °C with the use of the additional water heating system (AWHS) and the photovoltaic system in the autonomous heating unit. The actual shares of the density of the flow of solar radiation within the greenhouse are established when compared with the total densities of the flux of sun radiation falling on the transparent surface of the object under investigation, covered with snow of varying thickness. In the course of the studies, the analysis of the temperature gradients of panel radiators with the help of a thermal visualizer and the content of the air environment of the internal part of the object under investigation with the aid of a gas analyzer was carried out.

1 INTRODUCTION

Oil, natural gas and coal are commonly used as fuel in heating systems, which exacerbates the environmental impact of the greenhouse industry. Depending on the dimensions, thermal characteristics, cultivated plants, the energy needed to heat 1 m^2 of greenhouse area varies from 500 to 2700 MJ/m^2 per year [1-3].

Modern greenhouse complexes are divided into two types: seasonal and year-round. Usually seasonal greenhouses are used from November to March. Seasonal greenhouses allow for harvesting in the cold period of the year and can be either over ground or trench [4-6]. It should be noted that the surface greenhouses are not protected from winds, and variations in temperature, which leads to a significant loss of energy on their heating in the winter [5]. One of the main advantages of trench type greenhouses is their low energy consumption [7-14].

In the winter of 2023, the temperature in Uzbekistan dropped to - 21^{0}C, in many greenhouses the whole crop froze [15]. It is clear that emergency heating systems must be installed in greenhouses in order to avoid the freezing of crops [16].

The aim of the study is to ensure the reliability of the heating system of the autonomous helio-greenhouse of the trench type in cold climates.

The design has been developed and an autonomous helio-greenhouse of the trench type (Figure.1) with a depth of 1,2 m, with an area of 105 m^2 has been built. The main source of heat for the heating of the greenhouse is solar radiation, coming through the transparent roof. To reduce heat losses, the walls of the greenhouse are made of concrete. Additionally, the outside walls of the greenhouse and the roof are covered with transparent polycarbonate film. 14 lemon trees were planted in the greenhouse [17].

Figure 1: General type of autonomous helio-greenhouse of the trench type.

In this work [18], for the first time, a mathematical model is proposed for studying the behavior of the dependence of the air temperature in a solar-fuel trench-type greenhouse. Since, at present, there is no thermal technical assessment of the microclimate of a solar-fuel trench-type greenhouse for the climatic conditions of the regions of Uzbekistan.

To check the reliability of the proposed mathematical model, experimental measurements of the air temperature inside the greenhouse were carried out, built on the experimental site of the Tashkent state technical university. The accuracy of the proposed mathematical model of a solar-fuel trench-type green-house was estimated using the methods of "standard deviation" and the square of the correlation coefficient. As the results show, the standard deviation is equal to 1,5 °C, the standard deviation in percent is equal to 7,2% and the correlation coefficient is 0,86.

2 MATERIALS AND METHODS

An autonomous photovoltaic plant of 3 kW was used as a source of electric and thermal energy. LED lamps were used at night to illuminate the greenhouse. For the heating of plants in the greenhouse used incandescent lamps of 150 W, with different spectrum of radiation, which were installed at a distance of 80-100 cm from the surface of the soil six sensors were installed to monitor the temperature of the soil outside its surface. The total power consumption of the lamps is 2,1 kW. The reserve emergency power source (Figure 2, 3) was a 3,5-kW electric boiler combined with a solar water heating system with forced hot water circulation in the greenhouse batteries, the average temperature of which was 60 ^{0}C.

Figure 2: General type of autonomous greenhouse electric heating system.

From Figure 3 it can be seen that the basic scheme of the combined system (electric boiler and solar water heater) consists of: 1 - solar heating water collector (SHWC); 2 - heat exchanger; 3, 7 - expansion tank; 4 - circulation pump; 5 - hot water battery; 6 - automatic mixing valve; 8 - emergency valve with manometers; 9 - electric boiler; CW and HW - cold and hot water; rad.1, rad.2, rad.3, rad.4 - heating radiators.
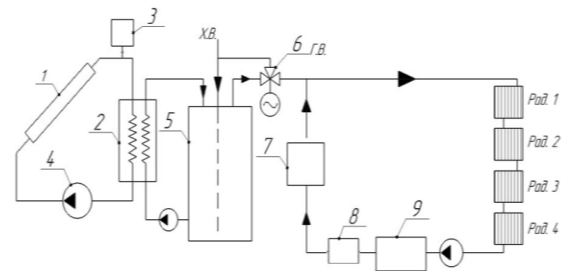

Figure 3: Principal scheme of two-circuit combined system (electric boiler and solar water heating system).

The experiments were carried out from 06.12. to 10.12.2022 at the beginning of the winter season. Daily values of solar radiation, ambient temperature, soil temperature, relative humidity outside and inside the room, indoor ambient temperature values, etc. were measured.

List of instruments used for measurement:
1) FLIR E5 - to measure the temperature on the surface of the radiators;
2) anemometer AS856 - for measurement of wind speed and air temperature;
3) pyranometer Solar Power Meter Di-LOG SL101 - for solar radiation;
4) pressure manometer - for water pressure;
5) termometer with output sensor - for temperature measurements.

In particular, a multi-channel digital thermometer of the brand DS18B20 with portable temperature sensors, with a temperature range of -55°C to +125°C, was used to measure the temperature of the air space of the autonomous helio-greenhouse of the trench type. The temperature measurement accuracy in the range from --10°C to +85°C is ±0.5°C.

3 RESULTS AND DISCUSSIONS

On the basis of the experimental studies carried out in the autonomous helio-heating plant of the tranche type, graphs reflecting the daily values of the ambient temperature and airspace inside the helio-greenhouses are constructed. Figure 4 shows the daily thermograms of the outdoor air and the air environment inside the heliothermal plant. On the day of 6.12.2022 the weather was cloudy, clear clouds, during the night there was snow. The ambient temperature (Figure.4, curve 1) in the afternoon varied from -7 ^{0}C to -4 ^{0}C. The thickness of the snow on the roof of the transparent surface of the autonomic greenhouse was 4÷6 cm.
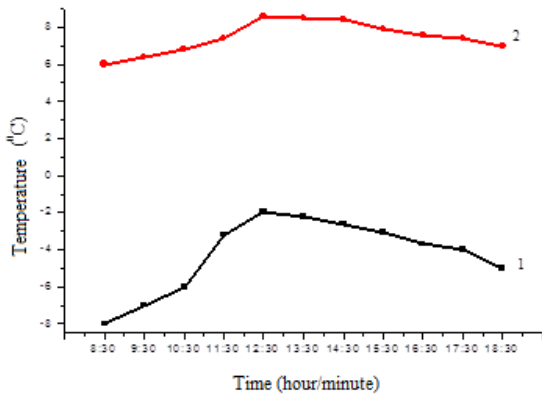
Figure 4: Daily thermo-grams of the outdoor air (curve 1) of the air environment inside the helio-greenhouse (curve 2) of the trunk type (06.12.2022).

The experiments continued on 10.12.2022 in the afternoon in clear weather. On the roof of the greenhouse until noon lay snow cover up to 1.2 cm thick. Environmental temperature values in the afternoon ranged from -10 ^{0}C to -4 ^{0}C (Figure. 5, curve 1).

From the graphs (curve 2) on Figure 4-5 it follows that the minimum values of the temperature of the air inside the greenhouse was reduced to 5 ^{0}C and 6^{0}C, while the outside air temperature varied from -10 ^{0}C to -7 ^{0}C (curve 1).
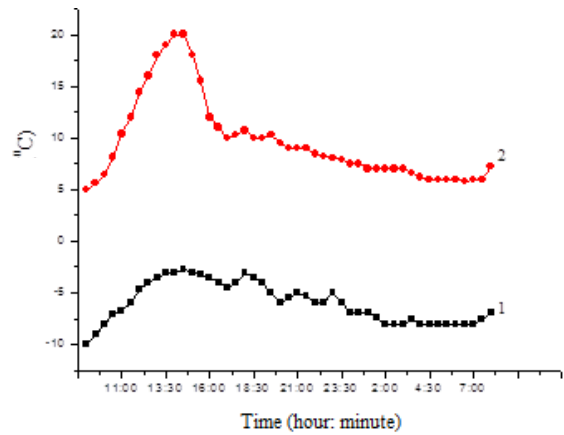

Figure 5: Daily thermograms of the outdoor air (curve 1) of the air environment inside the heliothermal plant (curve 2) of the trunk type (10.12.2022).

Figure 6 shows the dynamics of changes in solar radiation during the day. Analysis of the results of experiments indicate that with the increase of the snow thickness on the roof of the autonomous greenhouse, the transmittance coefficient of transparent multi-channel polycarbonate decreases, and vice versa, impedes the exchange of circulation processes of cold and warm air between the solid and air space.

The proportion of solar radiation (Figure. 6) taking into account the thickness of the snow cover ~1,2 cm and the consistent cloudiness inside the greenhouse is 7 ÷ 30% of the total value measured outside the research facility.
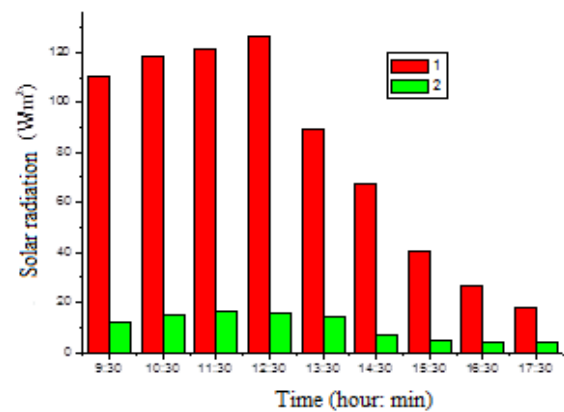

Figure 6: Dynamics of change of solar radiation during the day (1 - outside the greenhouse; 2 - inside the green house), (06.12.2022).

Figures 8-9 shows the results of experimental studies to determine the day and night changes in the temperature of the outdoor air and in the greenhouse

from 11 January to 14 January 2023 using a two-circuit combined system: an electric boiler with a solar water heating system.
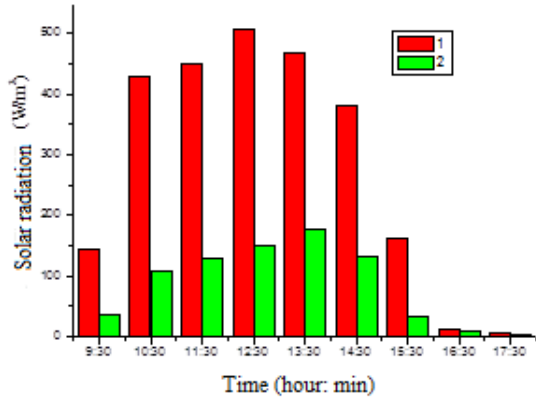

Figure 7: Dynamics of change of solar radiation during the day (1 - outside the greenhouse; 2 - inside the green house), (10.12.2022).
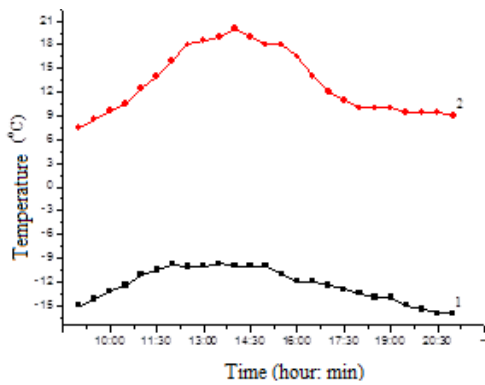

Figure 8: Daily thermograms of the outdoor air (curve 1) of the indoor air of the greenhouse (curve 2) of the tranche type (11.01.2023).

On 11.01.2023 the ambient temperature varied from -16 ^{0}C to -9,7 ^{0}C, and the intensity of solar radiation ranged from 6÷548 W/m^2. The transparent surface of the greenhouse was covered with snow, and its thickness was up to 10 cm. The average relative humidity in the greenhouse was 86%. The average temperature and humidity of the soil in the structure were ~16 ^{0}C and 76%, respectively.

As shown from the graphs on Figure 8 (curve 1, 2) the temperature in the greenhouse dropped to 10 ^{0}C at the outside air temperature -16 ^{0}C. When the ambient temperature dropped to -21 ^{0}C (Figure 9), the temperature inside the greenhouse was positive 8 - 9 ^{0}C. These abnormal frosts also include a lighting system based on a light bulb powered by a photovoltaic station.
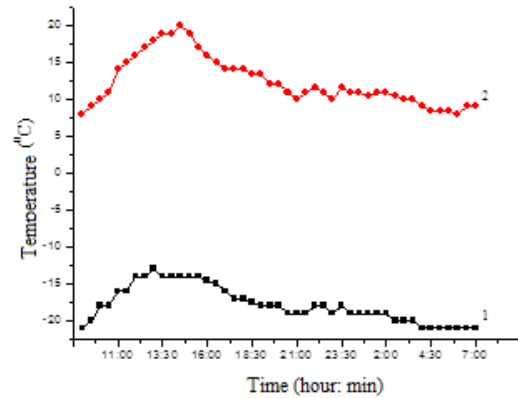

Figure 9: Daily thermograms of the outer air (curve 1) and the air environment inside the helio-greenhouse (curve 2) of the trunk type (13-14.01.2023).

Figures 10-11 shows the density of the flow of solar radiation during the day in Tashkent.
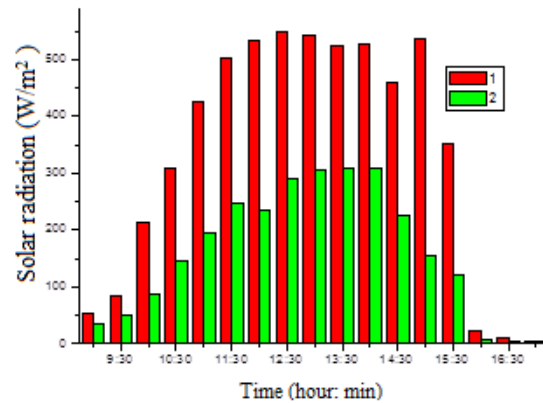

Figure 10: Dynamics of changes in solar radiation during the day (1.Outside the research facility, 2. Inside the research facility), (11.01.2023).
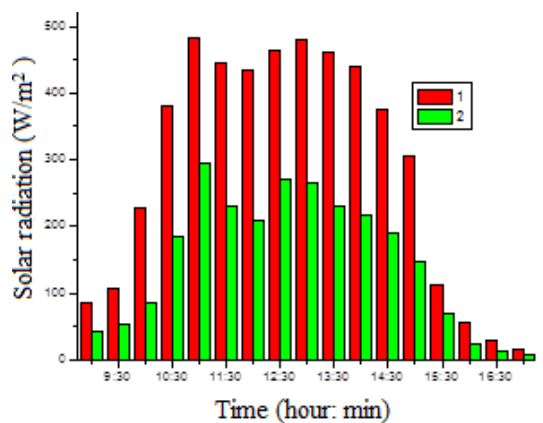

Figure 11: Dynamics of changes in solar radiation during the day (1.Outside the research facility, 2. Inside the research facility), (13.11.2023).

After analyzing Figures 11-12, it can be said that with a snow cover thickness of ~10 cm on the transparent surface of the autonomous greenhouse, the actual contribution to the maximum value of solar energy inside the structure is ~60% of the falling total solar radiation outside the object being studied.

In the framework of the study, infrared images (Figures 12, 13, 14, 15) of panel radiators obtained by the device (FLIRE63900) installed on the fencing of the greenhouse compartment were studied. The number of heating panel radiators 4 pcs and has a single circuit consecutive connection in the heating system.

As the results of Figure10 show, the temperature of the water on the radiators with a single-circuit sequential connection will begin to decrease. In the last radiator the water temperature differs by 12 ^{0}C when compared to the first radiator.
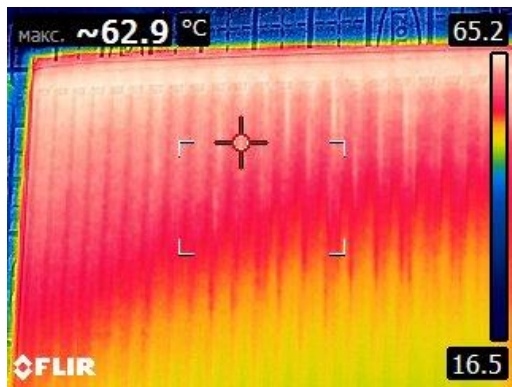

Figure 12: Dynamics of temperature change in greenhouse heating systems with sequential connection of radiators, Time: 14:00, Radiator №1, $T_{radiat.}$~62,9 ^{0}C.
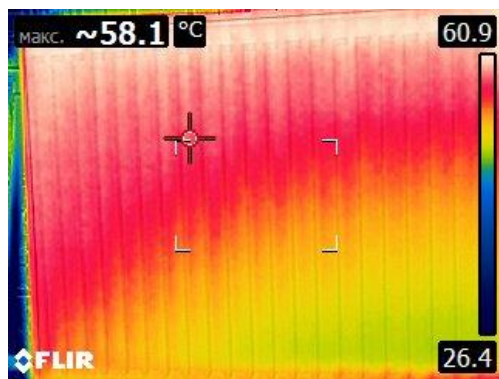

Figure 13: Dynamics of temperature change in greenhouse heating systems with sequential connection of radiators, Time: 14:01, Radiator №2, $T_{radiat.}$~58,1 ^{0}C.
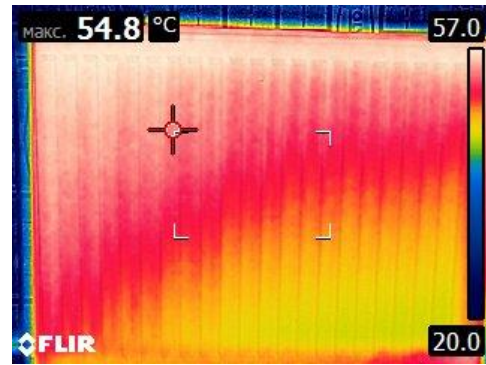

Figure 14: Dynamics of temperature change in greenhouse heating systems with sequential connection of radiators, Time: 14:02, Radiator №3, $T_{radiat.}$~54,8 ^{0}C.
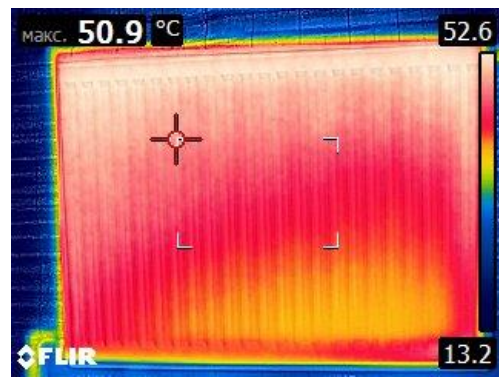

Figure 15: Dynamics of temperature change in greenhouse heating systems with sequential connection of radiators, Time: 14:03; Radiator №4, $T_{radiato}$~50,9 ^{0}C.

4 CONCLUSIONS

The following conclusions can be drawn from the results of the studies:

1) during the winter heating season at ambient temperature up to -10 ^{0}C and when using an autonomous photovoltaic system, the air temperature inside the autonomic heating facility dropped closer to +5 ^{0}C, and at an external temperature of more than -10 ^{0}C it was necessary to turn on emergency heating sources;

2) a basic scheme of a two-circuit combined system based on an electric boiler and solar water heating system is proposed;

3) the actual share of solar energy within the greenhouse is determined when compared with the total solar radiation outside the object under investigation, taking into account the thickness of the transparent surface covered with snow.

REFERENCES

[1] O. Korner, M.J. Bakker, and E. Heuvelink, "Daily Temperature Integration: A Simulation Study to Quantify Energy Consumption," Biosystems Engineering, vol. 87, no. 3, pp. 67-77, 2004, [Online]. Available: https://doi.org/10.1016/j.biosystemseng.2003.11.003.

[2] M. Teitel, A. Segal, A. Shklyar, and M. Barak, "A comparison between pipe and air heating methods for greenhouses," Journal of Agricultural Engineering Research, vol. 72, pp. 259-273, 1999, [Online]. Available: https://doi.org/10.1016/S0021-8634(98)90370-590370-5).

[3] H.J. Tantau, "Energy saving potential of greenhouse climate control," Mathematical and Computer Modelling, vol. 48, pp. 93-101, 1998, [Online]. Available: https://doi.org/10.1016/S0378-4754(98)00145-100145-1).

[4] G.N. Tiwari, M. Din, N.S.L. Srivastava, D. Jain, and M.S. Sodha, "Evaluation of solar fraction (Fn) for north wall of a controlled environment greenhouse: an experimental validation," International Journal of Energy Research, vol. 26, pp. 203-215, 2002, [Online]. Available: https://doi.org/10.1002/er.776.

[5] C.P. Kothandaraman, "Fundamentals of Heat and Mass Transfer," Revised Third Edition, New Age International (P) Limited, Publishers, 2015, [Online]. Available: https://edisciplinas.usp.br/pluginfile.php/5464110/mod_book/chapter/23393/Heat%20and%20Mass%20Transfer%20by%20kothadaraman.pdf.

[6] M.V. Shcherbakov, A. Brebels, N.L. Shcherbakova, A.P. Tyukov, T.A. Janovsky, and V.A. Kamaev, "A survey of forecast error measures," World Applied Sciences Journal, vol. 24, pp. 171-176, 2013, [Online]. Available: https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032.

[7] K. Kobayashi and M.U. Salam, "Comparing simulated and measured values using mean squared deviation and its components," Agronomy Journal, vol. 92, no. 2, pp. 345-352, 2000, [Online]. Available: https://doi.org/10.1007/s100870050043.

[8] N.T. Kottegoda and R. Rosso, "Applied Statistics for Civil and Environmental Engineers," Blackwell Publishing Ltd., Oxford, United Kingdom, p. 718, 2008, [Online]. Available: https://www.academia.edu/26724680/N_T_Kottegoda_R_Rosso_Applied_Statistics_for_Civil_and_Environmental_Engineers_Blackwell_Pub_2008_Copy.

[9] Y.V. Kozenko, A.V., Katayev, T.M. Katayeva, N.W. Likholetov, E.L. Makarov, L.V, Sharonin, "Collective monograph," Ufa: "OMEGA SCIENS", 108 p., 2016, [Online]. Available: http://www.aup.ru/books/m1527.

[10] A.M. Penjiyev, "Energy efficiency of energy resources and climate zoning of solar greenhouses," Aeconomics: Economy and Agriculture, no. 9 (21), 2017, [Online]. Available: http://aeconomy.ru/science/agro/energoeffektivnost-energeticheskikh.

[11] A.M. Penjiyev, "Agrotechnics of cultivation of pine tree (Carica papaya L.) in protected soil in Turkmenistan," Autoref. Ph.D. of agricultural sciences, Moscow, pp. 54, 2000, [Online]. Available: h https://search.rsl.ru/ru/record/01000259201.

[12] A.M. Penjiyev, "Mathematical model of sheet temperature mode calculation in hot-house conditions," International Journal "Alternative Energy and Ecology", no. 11(91), pp. 65-68, 2010, [Online]. Available: https://cyberleninka.ru/article/n/matematicheskaya-model-rascheta-temperaturnogo-rezhima-lista-v-usloviyah-solnechnoy-teplitsy/viewer.

[13] A.M. Penjiyev, " Mathematical modelling of the microclimate in solar hot-house deep of type". International Journal of Alternative Energy and Ecology, no. 7, pp. 59-66, 2010, [Online]. Available: https://cyberleninka.ru/article/n/matematicheskoe-modelirovanie-mikroklimata-v-solnechnoy-teplitse-transheynogo-tipa/viewer.

[14] Z.S. Akhatov and A.S. Khalimov, "Numerical calculations of heat engineering parameters of a solar greenhouse dryer," Applied Solar Energy, vol. 51, pp. 107-111, 2015, [Online]. Available: https://doi.org/10.3103/S0003701X15020024.

[15] Sinoptiki, "Utochnili, kak dolgo v Uzbekistane proderzhitsya anomalnyy moroz," [Online]. Available: https://kun.uz/ru/news/2023/01/10/sinoptiki-utochnili-kak-dolgo-v-uzbekistane-proderjitsya-anomalnyy-moroz, [Accessed: 25 Nov 2023].

[16] A. Abdullaev and Sh.K. Niyazov, "The IR-radiation transmission coefficient of a water condensate film on internal solar greenhouse surfaces and hotbed transparent covers as a function of thickness," Applied Solar Energy, Tashkent, 2001, vol. 37, no. 4, pp. 63-65, [Online]. Available: https://www.scopus.com/record/display.uri?eid=2-s2.0-77950044145&origin=resultslist.

[17] I.A. Yuldashev, B.M. Botirov, N.S. Kholmirzayev, and Y.M. Qurbanov, "About the Production of Lemons Grown in an Autonomous Gabled Solar Greenhouse," Applied Solar Energy, USA, 2023, vol. 59, no. 1, pp. 44-47, [Online]. Available: https://doi.org/10.3103/S0003701X23600431.

[18] B. M. Botirov, A.S. Khalimov, I.A. Yuldoshev, D.M. Pulatova, and Yu.M. Kurbanov, "Experimental Verification of a Mathematical Model for the Temperature Mode of a Solar-Fuel Trench-Type Greenhouse," Applied Solar Energy, USA, 2021, vol. 57, no. 6, pp. 510-516, [Online]. Available: https://doi.org/10.3103/S0003701X21060050.

Substantiating the Jet Agitator Parameters for Dredgers

Andrii Bondarenko[1], Viktor Kukhar[1] and Assel Nurmanova[2]

[1]*Department of Engineering and Design in Machinery Industry, Dnipro University of Technology,*
DmytroYavornytsky Avenue 19, 49005 Dnipro, Ukraine
[2]*Satbayev University, Satpaev Str. 22a, 050013 Almaty, Kazakhstan*
bondarenko.a.o@nmu.one, kukhar.v.yu@nmu.one, a.nurmanova@satbayev.university

Keywords: Jet Agitator, Mathematical Model, Dredger, Underwater Face.

Abstract: The paper presents the results of a theoretical study concerning jet flow-loose ground interaction in the underwater face. A mathematical model was developed to establish the values of a sufficient number of geometric parameters to develop the contact surface between the turbulent jetting with the ground being washed out. The adequacy of the developed mathematical model was confirmed while preparing and conducting a series of laboratory experiments aimed at granular soil jet agitating; the designed laboratory facility was applied. Numerical values of empiric coefficients were derived, and the experimental data have been processed statistically.The effectiveness of the use of jet agitators in underwater mining and construction works has been confirmed by the results of theoretical and experimental studies obtained in this work. The developed mathematical model to calculate jet agitators has been tested while developing designs of such ejector dredgers as *ЗНС 630-90* and *ЗНС 300-120*, and while modernizing soil suction heads of *ССБ 500/440*, *МЗ-16Э*, *МЗ-8*, *МЗ-11*, *Pechora*, and *ГПП-11* dredgers.

1 INTRODUCTION

Practices of using floating dredgers to mine underwater or waterlogged deposits of ore and nonmetallic sands suggest that excavation intensification can be achieved by using rationally designed jet agitators [1, 2]. The essence of the hydraulic disintegration and excavation method is as follows. A hydraulic disintegrator performs primary separation; slurry is sucked through a suction head, combined with an agitator in one draghead [3, 4]. The devices, used for hydraulic fragmentation of underwater face, are divided into two types: facilities for surface and deepwater washing out [5-7]. It is known that the use of jet agitators helps increase the dredger productivity by more than 25% while reducing the excavated soil cost [8-10]. In this regard, substantiation of the rational parameters of jet agitators of dredgers is the topical scientific and engineering problem.

The known previously developed mathematical models make it possible to determine some basic parameters of underwater and sub-bottom wash zones shaped by washing out jet, i.e. range, length, and width [1, 2]. However, the parameters are insufficient to describe the surface of the underwater face subjecting to the action of a turbulent washing jet [11-13]. In this regard, such models prevent from identification of rational design and technological parameters of the dredger soil suction head with a jet agitator.

The purpose of the work is to develop a mathematical model of the underwater face formation by jet agitators helping construct a contact surface of the turbulent washing jet with the granular mineral being washed out. The development of such a model makes it possible to substantiate rational structural parameters of the dredger soil suction head with a jet agitator.

2 MANUSCRIPT PREPARATION

A comprehensive research method has been applied. The procedure to calculate the rational parameters of hydraulic agitating involved both theoretical study and experiments concerning interaction between turbulent jet streams and underwater face of the loose ground. Standard approaches, criteria of the applied hydrodynamics methods as well as physical and mathematical modelling were applied [14-16]. To confirm the adequacy of the developed

mathematical model, a set of laboratory experiments of jet washing of granular soil was prepared and performed using the developed laboratory facility [17, 18]. The experimental data were processed and analyzed using standard methods of mathematical statistics and planning of experiments [19-21]. Analytical dependencies were derived through approximation of the experimental results using the least square procedure. Industrial full-scale experiments were carried out with the use of operating dredgers under the conditions of minefield mortar sand deposits. Technological and design parameters were measured with the help of standard metric equipment.

2.1 Mathematical modelling

The mathematical model has been constructed in accordance with the previously developed physical model [2, 3]. Thus, this paper considers an inclined water turbulent washing jet, separating particles of grain soil and shaping a bottom face (Figure 1). Operation of the vertical washing jet, as a special case of the inclined jet, is not considered, since the real operation of the dredge in the underwater face is accompanied exclusively by the operation of the inclined washing jet. To develop a model of the contact surface of the inclined washing jet with the underwater face, consider the scheme shown in Figure 1.
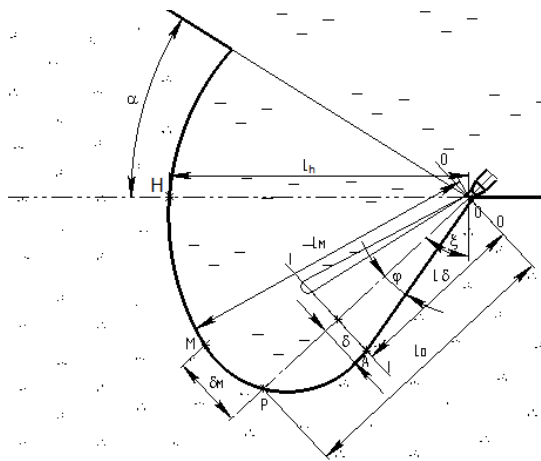

Figure 1: Scheme of an underwater face shaped by a turbulent washing jet.

Assume that a jet of incompressible fluid flows from a nozzle with **R** radius, into the grained soil formation where particle size is d_{av}. The nozzle is installed at the surface of the reservoir bottom; in this context, the axis of the jet passes through the

intersection of the forming surfaces of the bench and its toe.

The washing zone, formed by such a jet, is an underwater jet washing zone. Introduce some specifications: R is the nozzle radius, m; D_n is the nozzle diameter, m; φ is the angle of unilateral expansion of the outer boundary of the turbulent axisymmetric jet in the main section, deg.; L_δ is the length of the main section of the washing zone, characterizing the distance, singled out along the axis of the jet, between the nozzle and the main section boundary of the underwater jet washing zone, represented by the projection of point A on the axis, m; δ is the maximum half-width of the main jet section, characterizing the distance between point A and the jet axis, m; L_o is the jet range characterizing the distance being singled out along the axis of the jet between the nozzle and the point of the washing zone P spacing maximally from the nozzle, and lying on the jet axis, m; L_m is the furthest point of the washing zone characterizing the distance between the nozzle and the furthest point of the washing zone M, m; δ_m is the distance between M point and the jet axis, m; L_h is the horizontal range, characterizing the distance between the nozzle and the maximum point of the washing area H, and lying within the horizontal plane, m; u_0 is the initial velocity of liquid flow from the nozzle, m/s; $U_p^{\beta\gamma}$ is the analytical value of ground washing velocity at the point of the sloping surface of the underwater jet washing zone, m/s; Q_0 is the fluid flowrate through the nozzle, cross-section 0-0 m³/s; and Q_1 is the fluid flowrate through the cross-section 1-1, m³/s.

Due to the change in the live section of the flow along the length, fluid flow movement in the main section of the underwater jet washing zone is smoothly varied. It has the following characteristics:
a) centrifugal forces are not taken into consideration;
b) live flow sections are treated as the planar ones;
c) projection of the flow live sections on a plane perpendicular to the axis of the jet is a circle; and
d) vertical flows are not taken into consideration.

The description of the fluid flow in the main section of the underwater jet washing zone is performed using the Euler equation for an ideal (non-viscous) fluid [22-24] generally written as

$$\begin{cases} \dfrac{du_x}{dt} = \dfrac{\partial u_x}{\partial t} + u_x \dfrac{\partial u_x}{\partial x} + u_y \dfrac{\partial u_x}{\partial y} + u_z \dfrac{\partial u_x}{\partial z}; \\[2mm] \dfrac{du_y}{dt} = \dfrac{\partial u_y}{\partial t} + u_x \dfrac{\partial u_y}{\partial x} + u_y \dfrac{\partial u_y}{\partial y} + u_z \dfrac{\partial u_y}{\partial z}; \\[2mm] \dfrac{du_z}{dt} = \dfrac{\partial u_z}{\partial t} + u_x \dfrac{\partial u_z}{\partial x} + u_y \dfrac{\partial u_z}{\partial y} + u_z \dfrac{\partial u_z}{\partial z}. \end{cases} \quad (1)$$

Consider the flow movement in the main section of the jet underwater washing zone (Figure 1) along the O-X axis when it is expanding towards O-Y and O-Z axes. The flow rate in sections 0 and 1 is constant, i.e. $Q = const$. Hence,

$$Q_0 = Q_1 = s_0 u_0 = s_1 u_1 .$$

Then the flow velocity will be

$$u_0 = \frac{Q_0}{s_0} = \frac{Q_0}{\pi R^2} \qquad u_1 = \frac{Q_1}{s_1} = \frac{Q_0}{\pi R_x^2} ,$$

where R_x is the radius of the main section of the axisymmetric washing jet in the ith section.

Relying upon the theory of flowing of flooded jets [1,2,23] for the main section, the change in the parameters of the circular section along the coordinate x will be represented as

$$R_x = x\, tg\varphi. \qquad (2)$$

According to [25], $\varphi = 12^{\circ}30'$. The tangent of the expansion angle is commonly referred to as the expansion coefficient c; then, $c = tg\varphi = 0.22$.

Transversely (i.e. along y co-coordinate), the dependence of the flow velocity upon the longitudinal velocity (i.e. along x coordinate) will be

$$u_y = \frac{dR_x}{dt} = tg\varphi \frac{dx}{dt} = tg\varphi u_x = c \cdot u_x .$$

Due to the change in a cross-sectional area, the flow in the main section of the axisymmetric washing jet is uneven. Since the time-controlled velocity changes result from uniform flow expansion, the partial time velocity derivatives are very small compared with the other terms of the equation; so

$$\frac{\partial u_x}{\partial t} = 0 \quad \frac{\partial u_y}{\partial t} = 0 \quad \frac{\partial u_z}{\partial t} = 0 .$$

To solve simple problems, it is enough to consider the flow as a one-dimensional value and determine the maximum velocity and acceleration values. Then, the system of equations (1), taking into consideration the described dependencies of the flow

velocity in the transverse and vertical directions, can be written in the form

$$\frac{du_x}{dt} = u_x \frac{\partial u_x}{\partial x} + c u_x \frac{\partial u_x}{\partial y} + c u_x \frac{\partial u_x}{\partial z} .$$

In view of the assumptions about the insignificance of the curvature of flow live sections, equating the flow velocity at any point to its mean value, bringing the parameters of the flow live section to the axis O-X, as well as solving a flat problem, the system of equations (1) is reduced to the form

$$\frac{du_x}{dt} = u_x \frac{\partial u_x}{\partial x} . \qquad (3)$$

In the main section of the axisymmetric washing jet, the change of parameters of the circular section along x coordinate is described by expression (2). Hence, we obtain the flow velocity as the ratio

$$u_x = \frac{Q_0}{\pi (c \cdot x)^2} . \qquad (4)$$

Involving $Q_0 = \pi u_0 R^2$, private derivative $\dfrac{\partial u_x}{\partial x}$ will take the following form:

$$\frac{\partial u_x}{\partial x} = -\frac{u_0 R^2}{c^2 x^3} . \qquad (5)$$

Then, transforming (3) and taking into consideration the expression for flow velocity (4) as well as the type of partial derivative (5), we write flow acceleration in the main section of the axisymmetric washing jet as

$$\frac{du_{xp}}{dt} = -\frac{u_0^2 R^4}{c^4 x^5} . \qquad (6)$$

The main section of the jet is characterized by a regular decrease in velocity values in accordance with the increase in the cross-sectional area of the flow. The flow in the section under consideration is characterized by negative, acceleration values, approaching zero. Proceeding from the equality of productivity in sections 0 and 1 (Figure 1), we obtain

$$Q_0 = Q_1 = u_0 \pi R^2 = u_x \pi c^2 x^2 .$$

It is known that the velocity at the contact surface of the jet with the grained soil in section 1-1 is equal to the value of washing velocity $u_x = U_p^{\beta\gamma}$. According to the earlier research by the author

(see [1,2]), upper indices β and γ characterize the value of the slope angles of the surface formations of the jet-washing underwater zone. Thus, the equation to determine the main section length of the washing zone will be

$$L_\delta = \frac{R}{c}\sqrt{\frac{u_0}{U_p^{\beta\gamma}}} \qquad (7)$$

The correspondence of the mathematical model to the actual physical picture, taking place in the underwater face of the turbulent jet, becomes possible if empirical coefficient **k** is introduced into the dependence (7)

$$L_\delta = \frac{R}{kc}\sqrt{\frac{U_0}{U_p^{\beta\gamma}}} \qquad (8)$$

Analysis of the empirical curves, obtained as a result of earlier experimental studies using the developed laboratory facility [1, 2], helps formulate a type of dependence of the coefficient k upon the initial velocity of the liquid flow from the nozzle u_0 and the angle of inclination of the jet axis

$$k = \left(\frac{\alpha}{\xi}\right)^m (a - bu_0) \qquad (9)$$

where α is the natural slope angle of the underwater bench when mining by a dredger. The normative value of the angle for water-saturated medium-grained sand is α =32º [5, 10]; ξ is the inclination angle of the washing jet axis, deferred from the vertical one, degrees; and a , b , m are the empirical coefficients.

Consideration of a turbulent inclined water jet, forming sections of the turning and return flow (Figure 1), allows us to argue about the complexity of running processes to perform mathematical modelling with the theoretical model construction. Calculation of forces of the impact of jet flow on the grained soil, leading to relevant perturbations in the latter, is a very complex problem of fluid dynamics of jet streams. Therefore, the paper aim is to develop a mathematical model helping us set the values of a sufficient number of geometric parameters to design the contact surface of the washing jet with the grained soil based on empirical models.

Resulting from estimating processing of experimental data, empirical dependencies to determine the geometric parameters required to construct the contact surface of the washing jet with the grained soil were obtained [1, 2]:

- axial jet range

$$L_o = c_{L_o}\frac{\mathrm{Re}\,\nu_n}{U_p^{\beta_{90}\gamma_0}} - d_{L_o}\Delta_U d^{ls} + e_{L_o}d^{ls};$$

- the furthest point of the washing area

$$L_m = c_{L_m}\frac{\mathrm{Re}\,\nu_n}{U_p^{\beta_{90}\gamma_0}} - d_{L_m}\Delta_U d^{ls} + e_{L_m}d^{ls};$$

- distance between M point and the jet axis

$$\delta_m = c_{\delta_m}\frac{\mathrm{Re}\,\nu_n}{U_p^{\beta_{90}\gamma_0}} - d_{\delta_m}\Delta_U d^{ls} + e_{\delta_m}d^{ls};$$

- horizontal range

$$L_h = c_{L_h}\frac{\mathrm{Re}\,\nu_n}{U_p^{\beta_{90}\gamma_0}} - d_{L_h}\Delta_U d^{ls} + e_{L_h}d^{ls};$$

where $c_{L_o}, d_{L_o}, e_{L_o}, c_{L_m}, d_{L_m}, e_{L_m}, c_{\delta_m}, d_{\delta_m}, e_{\delta_m}, c_{L_h}, d_{L_h}, e_{L_h}$ are the empirical coefficients; Δ_U is the notional velocity; and d^{ls} is the near-mesh grain of the soil assumed to be 0.00016 m.

2.2 Laboratory Experiments

Numerical values of the empirical coefficients were obtained as a result of experimental data processing during the study of soil washing by an inclined turbulent jet in the underwater face. The procedure was studied using laboratory conditions when a turbulent water jet was applied to the sand of medium grain size d_{av} = 0.265 mm with slope angles ξ = 15, 30, 45, 60, 75, and 90 degrees, and the diameters of the washing jet nozzle being D_n =1.25; 2.0; 2.8; and 4.0 mm.

To provide visual control and the possibility to perform appropriate measurements of the washing zone formation for underwater surface using a water jet in a loose soil, a laboratory facility was designed (see Figure 2).

Following problems have been solved with the help of the laboratory facility:

- formation process of the underwater surface washing zone has been studied using a single water jet action on a loose ground in terms of 0-90º angles of its axis inclination from the vertical;
- behavioural processes of the loose ground, laid with the appropriate angle of the natural slope, has been analyzed as well as the solids suspended by the jet stream during its underwater washing by the water jet;

- rational operating modes of a water jet, have been defined while shaping the underwater surface mine working to determine rational design parameters of a soil suction head with a jet agitator.

Values of empirical coefficients $a =1.2$, $b =0.503$, and $m =0.18$ were obtained through experiments made for the underwater face shaped under the action of a washing turbulent jet with $\xi = $ 15º, 30º, 45º, 60º, 75º, and 90º inclination angles; on an average, 50 experiments have been carried out for each angle.

Figure 2: General view of the laboratory facility to study soil underwater washing processes by a water jet.

2.3 The Experimental Data Processing and Analyzing

Values of the empirical coefficients were obtained as a result of data processing using the method of least squares based on the Mathcad program as well as statistical data processing in the Microsoft Office Excel program. Statistical analysis of the experimental data showed standard deviation of individual measurements from the calculated values; for example, in the context of the main section length of washing jet it was 18.5 %, and 13.2 % in the context of jet range. Confidence interval for the mathematical expectation with 90% certainty was 4.5 % and 3.2 % of the analytical values, respectively. Confidence interval for the mathematical expectation of coefficient, calculated by (9), was 5.6% from the calculated values with 90% confidence.

Figure 3 represents graphically the empirical dependence and experimental values, for example, it concerns length of the washing jet main section from the initial velocity of the liquid flow from the nozzle. Analysis of the graphs has assured us that increase in the initial velocity of the flow results in the increase of the studied parameters; moreover, they have a linear function within the specified velocity range.

The proposed model, determining the length of washing zone of the main section, makes it possible to set the value of the maximum half-width of the jet main section in accordance with the theoretical dependence $\delta=cL_\delta=0.22L_\delta$.
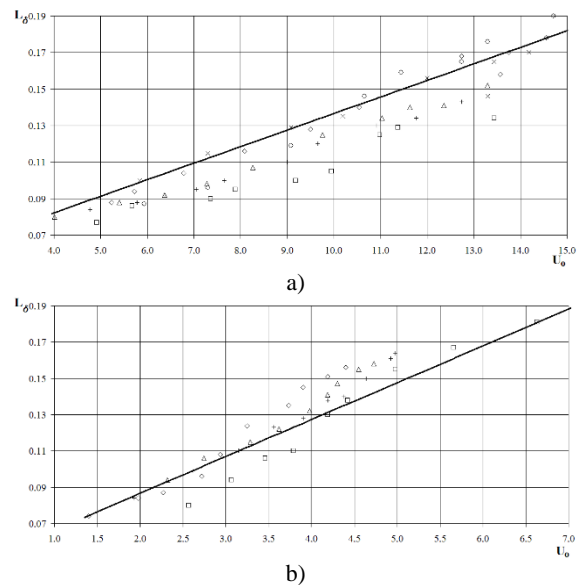

Figure 3: Dependence of jet range L_δ on the initial flow velocity u_0 depending upon a nozzle diameter, mm: a) $D_n = 2.0$; b) $D_n =4.0$.

Theoretical and empirical mathematical models obtained by the author are used in the design and modernization of dredging equipment with a jet agitator. In this case, it is important to take into consideration the rational balance between the geometric parameters of the underwater face, shaped by the turbulent water jet, and such parameters of the jet agitator as the radius of the washing nozzle; their number; direction; and the initial velocity of the fluid flow from the nozzle. Correct calculation of the rational parameters of jet agitators is an important factor in the process of soil washing velocity identification where slope angle of the washing zone surface is taken into consideration [1, 2].

Transformation of dependence (8) helps derive the analytical expression to determine the washing nozzle diameter

$$D_{\phi} = 2L_{\delta}kc\sqrt{\frac{U_p^{\beta\gamma}}{U_0}}.$$

2.4 Industrial Full-Scale Experiments and Implementations

The developed mathematical model to calculate jet agitators was tested while designing such ejector dredgers as *ЗНС 630-90*, and *ЗНС 300-120*, and updating dredge suction devices of *ССБ 500/440*, *МЗ-16Э*, *МЗ-8*, *МЗ-11*, *Pechora*, and *ГПП-11* dredgers. The detailed information about the projects by the author can be found on the site htmp.com.ua.

a)

b)

Figure 4: Soil suction head of the jet pump dredger *ЗНС 300-120* with a jet disintegration system: a) its general view; b) jet pump.

The ejector dredgers *ЗНС 630-90* and *ЗНС 300-120* are unique devices designed for extracting and dredging construction works under the difficult mining and geological conditions; if coarse gravel occurs in the ground; if off-shore operations are required; or operations take place at shallow depths [9].

Experimental tests of ejector dredgers *ЗНС 630-90* and *ЗНС 300-120,* equipped with suction heads, jet pumps, and agitators (see Figure 4), have shown their advantages and disadvantages. Industrial tests confirmed sufficient convergence of the analytical and experimental data, indicating relevance of the developed mathematical model. The mentioned ejector dredgers are widely used for dredging and mining operations in rivers and water-flooded quarries (Figure 5) [2, 9].

a)

b)

Figure 5: Jet pump dredgers: a) *ЗНС 630-90* ejector dredger at a mining site in Liberia; b) *ЗНС 300-120* ejector dredger in the flooded quarry of Ukraine.

Use of the developed system for soil preparation and hydro-transportation by turbulent jets under the specified conditions has helped:

- maintain sand mining productivity at the same level;
- reduce the cost by 15%;
- reduce current repair and maintenance of the equipment by 28%;
- develop soil with up to 200 coarseness mm; and
- reduce sand mining cost by 12%.

The obtained results offer us the possibility to recommend ejector dredger *ЗНС 630-90* for production. At present, *ЗНС 630-90* dredgers are applied successfully to develop nonmetallic river gravel deposits in Ukraine and a diamond deposit in Liberia (Figure 5) [2, 9].
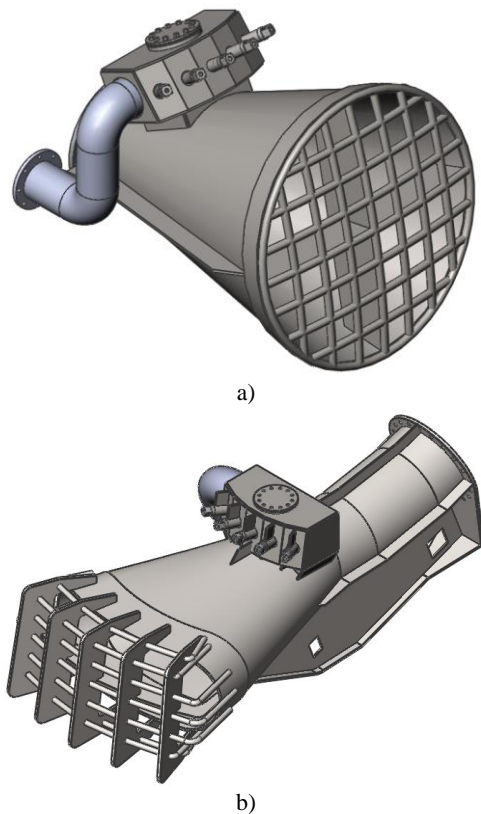

a)

b)

Figure 6: Modernized suction heads of the dredgers: a) *Pechora*, b) *ГПП-11*.

The developed mathematical model, calculating jet agitators, has also been used by the author to modernize a suction head of operating dredgers. Off-shore hopper dredger *Pechora* and river specialized sand loader *ГПП-11* are the examples [10]. Suction heads of the mining vessels have been upgraded by the jet agitation systems. Solid models of the upgraded suction heads with the jet agitation systems are shown in Figure 6. The detailed information about the projects can be found on the site htmp.com.ua.

Use of jet technologies to intensify soil washing in the underwater face has made it possible to modernize the dredge suction heads with minimal interference in the design of the facilities. It is also important that such modernization needs neither specialized equipment nor personnel qualification. Moreover, it can be performed by the mining company in accordance with the current relevant standards and regulations.

Pechora and *ГПП-11* updating has increased significantly the performance capability and efficiency of mining operations; quantitative factors are being processed and improved.

3 CONCLUSION

The result of mathematical modelling of the underwater surface face formation process by jet agitator was the obtaining of an analytical dependence to determine the diameter of the washing nozzle as a function being directly proportional to the length of the main section of the jet, the square root of *soil washing velocity-initial flow rate* ratio, and the coefficient, which is inversely proportional to the washing jet axis slope raised to the 0.5^{th} power. In addition, theoretical and empirical dependencies to define geometric parameters of the surface underwater face were obtained as a result of evaluating the effect of nozzle diameter and inclination angle of the washing jet axis on the formation of differently inclined surfaces of the turbulent water jet within the underwater face.

The developed mathematical model, calculating jet agitators, has been tested and successfully applied while designing *ЗНС 630-90* and *ЗНС 300-120* ejector dredgers as well as upgrading dredging equipment of *ССБ 500/440, МЗ-16Э, МЗ-8, Pechora,* and *МЗ-11* dredgers.

REFERENCES

[1] A. Bondarenko, "Mathematical modeling of soil dredger absorption processes in the underwater bottomhole," Metallurgical and Mining Industry, vol. 3, pp. 79-81, 2012.

[2] A. Bondarenko and Ye. Zapara, "Laws of determination of fine materials suction limits in submarine suction dredge face," Naukovyi Visnyk

Natsionalnoho Hirnychoho Universytetu, vol. 4, pp. 59-64, 2012.

[3] O. Medvedeva, "Development and exploitation of storages of enrichment process wastes as anthropogenic deposits," Theoretical and practical solutions of mineral resources mining, Taylor & Francis Group, London, 2015.

[4] O. Medvedieva, S. Kyrychko, N. Nykyforov, and N. Koval, "Substantiation of the boundary of the tailings storage core during the storing of the cleaning rejects by hydraulic method," E3S Web of Conferences, vol. 109, International Conference Essays of Mining Science and Practice, Dnipro, Ukraine, June 25-27, 2019.

[5] SOU-N MPP 73.020-078-3:2013, "Normy tehnolohichnogo proektuvannia hyrnychovydobuvnykh pidpryiemstv iz vidkrytym sposobom rozrobky rodovyshch korysnykh kopalyn," DPI Kryvbasproekt, Kryvyi Rih, 2013.

[6] E. Semenenko, N. Nykyforova, and L. Tatarko, "The method of hydraulic gradient and critical velocity calculation for hydrotransportation of particles with substantially different densities," in 15th International Freight Pipeline Society Symposium, June, 24-27, Prague, pp. 248-256, 2014.

[7] Y. Semenenko, S. Kril, N. Nykyforova, L. Tatarko, and O. Medvedieva, "Calculation of pressure loss and critical velocity for slurry flows with additive agents in vertical polyethylene pipelines," E3S Web of Conferences, vol. 109, International Conference Essays of Mining Science and Practice, Dnipro, Ukraine, June 25-27, 2019.

[8] A. Dryzhenko, S. Moldabayev, A. Shustov, A. Adamchuk, and N. Sarybayev, "Open pit mining technology of steeply dipping mineral occurrences by steeply inclined sublayers," in 17th International Multidisciplinary Scientific GeoConference SGEM 2017, pp. 599-605, 2017.

[9] A. Dryzhenko, A. Shustov, and S. Moldabayev, "Justification of parameters of building inclined trenches using belt conveyors," in 17th International Multidisciplinary Scientific GeoConference SGEM 2017, pp. 471-478, 2017.

[10] T. Al-Azab, F.M. Al-Ghathian, J.S. Haddad, et al., "Experimental Study of Drying Ratio and Humidity of Silica Sand Materials," J. Engineering. Thermophys, p. 550, 2019.

[11] Ye. Semenenko, N. Nykyforova, and L. Tatarko, "The features of calculations of hydrotransport plans of geotechnological systems," Theoretical and practical solutions of mineral resources mining, Taylor & Francis Group, London, 2015.

[12] O. Belov, O. Shustov, A. Adamchuk, and O. Hladun, "Complex Processing of Brown Coal in Ukraine: History, Experience, Practice, Prospects," Solid State Phenomena, vol. 277, pp. 251-268, 2018.

[13] O. Denyshchenko, A. Shyrin, V. Rastsvietaiev, and O. Cherniaiev, "Forming the structure of automated system to control ground heavy-type ropeways," Natsional'nyi Hirnychyi Universytet. Naukovyi Visnyk, no. 4, pp. 79-85, 2018.

[14] O. Anisimov, V. Symonenko, O. Cherniaiev, and O. Shustov, "Formation of safety conditions for development of deposits by open mining," E3S Web of Conferences forthcoming.USME 2018, 2018.

[15] V. Symonenko, J.S. Haddad, O. Cherniaiev, et al., "Substantiating Systems of Open-Pit Mining Equipment in the Context of Specific Cost," J. Inst. Eng. India Ser. D, p. 301, 2019.

[16] S. Moldabayev, Z. Sultanbekova, A. Adamchuk, and N. Sarybayev, "Method Of Optimizing Cyclic And Continuous Technology Complexes Location During Finalization Of Mining Deep Ore Open Pit Mines," in 19th SGEM International Multidisciplinary Scientific GeoConference EXPO Proceedings, Science and Technologies in Geology, Exploration And Mining, vol. 19, pp.407-414, 2019.

[17] O. Shustov, O. Bielov, T. Perkova, and A. Adamchuk, "Substantiation of the ways to use lignite concerning the integrated development of lignite deposits of Ukraine," Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu, no. 3, pp. 5-18., 2018.

[18] V. Nadutyy, J. Haddad, V. Sukhariev, and A. Loginova, "The results of experimental studies of influence of variable parameters on the performance indicators of shock-centrifugal disintegrator," Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu 169, no. 1, pp.42-47, 2019.

[19] V. Nadutyy, O. Tytov, and I. Cheberiachko, "Hereditary model of loose mined rock layer deformation in disintegrators," E3S Web of Conferences, vol. 60, 00033, Ukrainian School of Mining Engineering, 2018.

[20] O. Tytov, J. Haddad, and V. Sukhariev, "Modelling of mined rock thin layer disintegration taking into consideration its properties changing during compaction," in E3S Web of Conferences, vol. 109, International Conferences Essays of Mining Science and Practice, 2019.

[21] O. Chernyaev, "Systematization of the hard rock non-metallic mineral deposits for improvement of their mining technologies," Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu, no. 5, pp. 11-17, 2017.

[22] J. Bedrossian and V. Vicol, "The Mathematical Analysis of the Incompressible Euler and Navier-Stokes Equations: An Introduction," American Mathematical Society, 218 p., 2022.

[23] Ch. Wang, Zh. Chang, W. Chao, and Yu. Zhengand, "Local Well-Posedness and Break-Down Criterion of the Incompressible Euler Equations with Free Boundary," American Mathematical Society, 119 p., 2021.

[24] D. Stipp, "A Most Elegant Equation: Euler's Formula and the Beauty of Mathematics," Basic Books, 240 p., 2017.

[25] G. Abramovich, "The Theory of Turbulent Jets," Edited by Leon Schindel. The MIT Press, 684 p., 2003.

Features of Using Pumps in Turbine Mode

Kurbon Dzhuraev[1], Abdurauf Abduaziz uulu[2], Fotima Shadibekova[1] and Azamat Mambetov[1]

[1]Tashkent State Technical University, University Str. 2, 100095 Tashkent, Uzbekistan
[2]Namangan Institute of Engineering and Technology, Kosonsoy Str. 7, 160115 Namangan, Uzbekistan
djqs1983@gmail.com, abdurauf9222@mail.ru

Abstract: One of the possible directions of using renewable energy sources to save fuel and energy resources in Uzbekistan is the use of the hydropower potential of small rivers. The cost of electricity generated at micro and small hydroelectric power plant is already lower than the cost of electricity generated at traditional types of power plants, including gas turbine, wind, nuclear power plant and thermal power plant. In addition, due to constantly rising energy prices, the cost of electricity at traditional power plants is constantly increasing. The use of the hydropower potential of small rivers will contribute to the decentralization of the integrated energy system and improve energy supply to remote and hard-to-reach rural areas. Analysis of literature sources shows that the use of serial pumps as hydraulic turbines is a successful alternative to the use of specially designed hydraulic turbines for mini and small hydroelectric power plant. To reliably use pumps as hydraulic turbines, manufacturers need to obtain experimental performance of pumps in turbine mode, which can help expand their markets and better utilize available hydraulic potential.

1 INTRODUCTION

One of the possible directions of using renewable energy sources (RES) to save fuel and energy resources in Uzbekistan is the use of the hydropower potential of small rivers. The cost of electricity generated at micro and small hydroelectric power plant (HPP) is already lower than the cost of electricity generated at traditional types of power plants, including gas turbine, wind, nuclear power plant (NPP) and thermal power plant (TPP). In addition, due to constantly rising energy prices, the cost of electricity at traditional power plants is constantly increasing. The use of the hydropower potential of small rivers will contribute to the decentralization of the integrated energy system and improve energy supply to remote and hard-to-reach rural areas. This is the way the EU countries have gone. For example, in Switzerland, the percentage of electricity production at small HPP has already reached 8.3%, in Spain – 2.8%, in Sweden – almost 3%, and in Austria – all 10%. China, which has become a leader in the creation of small hydropower in recent decades (about 18-20% of all electricity in the country is supplied by more than 80 thousand small hydropower plants) and the former Soviet republics (Lithuania, Kyrgyzstan, Armenia, etc.) are moving in the same way. An important component of this process is the use of mini and microHPP for decentralized energy supply and thermal energy generation for consumers remote from industrial networks, for example, in mountainous areas, farms, etc. The increase in the cost of energy resources has stimulated interest in the development of new potential energy sources, which can serve as existing differences in irrigation canals, aqueducts, water supply and technological networks, and even transport oil pipelines.

As a first step in Uzbekistan to stimulate the use of renewable energy sources at the legislative level, one can note the adoption of so-called "green tariffs", which oblige the purchase of electricity produced from them at specially established prices, similar to the practice that exists in European countries.

One of the reasons hindering the widespread introduction of mini and microHPP in Uzbekistan is the high cost per kilowatt of installed equipment capacity, which is associated with the virtual absence of specialized companies performing the entire range of work - from equipment development, its manufacture and installation with subsequent commissioning and service. Abroad, there are more than 300 manufacturers of hydraulic turbine

equipment of various types of mini and micro hydroelectric power stations for various parameters, but the cost of the equipment is quite high and amounts to about 1,500-2,000 $/kW, which leads to significant payback periods for the equipment.

An alternative is the proposal to use commercially produced pumps as hydraulic turbines for microHPP, which in the power range of 5÷1000 kW can successfully compete with most used types of hydraulic turbines with significantly shorter payback periods.

Research into the possibility of using pumps as turbines to generate energy was carried out in the 90s in Germany [1], England [2], Iran [3], India [4, 5] and other countries. The recent increase in the number of publications on research on this topic in England, the USA, Russia, as well as on specific applications in developing countries [5, 6] confirms the need and prospects for using pumps in small-scale power generation.

This proposal is based on the proposition known from the theory and practice of hydraulic engineering that the hydraulic machine is reversible, and the pumps have fairly high efficiency in turbine mode and have some advantages over existing hydraulic turbines for mini and microHPP in the specified power range:

- there is no need to develop designs and establish small-scale production of equipment, since the pumps are already being produced, and the range is quite wide;
- serial pumps are usually supplied with an asynchronous motor, which can be used as a generator;
- high energy conversion efficiency, often corresponding to the level of specially designed hydraulic turbines;
- relatively low cost of equipment, which ensures quick payback for mini and microHPP, ease of maintenance and repair;
- the simplicity and compactness of the design makes it possible to consider various layouts of mini and microHPP units, which leads to a reduction in the cost of designing the construction part of mini and microHPP and construction time.

The disadvantages of using pumps as hydraulic turbines include:

- a steeply falling shape of the efficiency curve, close to the shape of a propeller hydraulic turbine, which requires either a narrow range of actuated pressures or the installation of a frequency converter that allows the generator to operate at a variable speed with subsequent conversion to an industrial frequency, which increases the cost of the installation;

- lack of experimental characteristics of pumps in turbine operating mode, which are necessary for the optimal selection of equipment for given parameters;
- the absence, as a rule, of regulatory bodies in the pump (except for a valve on the pressure pipeline, the use of which leads to a deterioration in operating conditions at partial loads).

In Figure 1 shows the areas of use for various types of hydraulic turbines used as mini and microHPP power equipment, and the area of pressures and powers in which serial pumps can be used [1]. From the above diagram it follows that serial pumps can replace mini and microHPP equipment for Banky type turbines, radial-axial Francis type and Pelton type active turbines in the pressure range H = 10-100 m.
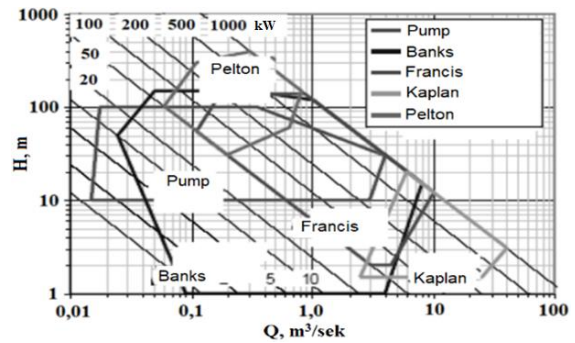

Figure 1: Areas of use of standard pumps as turbines according to [2].

To make a reasonable choice of the type of pump that provides the required parameters of the turbine mode in certain operating modes, it is necessary to have a characteristic in this mode. The lack of experimental characteristics of pumps in turbine mode leads to the need to establish, on the basis of a known pump characteristic, the relationship between the values of pressure $H_T\eta_{max}/H_P\eta_{max}$ and flow rate $Q_T\eta_{max}/Q_P\eta_{max}$ for the point with maximum efficiency in turbine and pump modes approximately using theoretical analysis [1-7], including using computational fluid dynamics (CFD) methods [8, 9].

Let us consider the flow features in pump and turbine modes for the pump impeller. These relationships can be approximately established based on the analysis of velocity triangles and the Euler equation

$$H_{Theory} = \frac{1}{g}(V_{U guide} \cdot U_{U guide} - V_{U\,BC} \cdot U_{U\,BC}). \quad (1)$$

Assuming that maximum efficiency values occur under the condition of flow supply to the impeller in

pump mode without preliminary swirling ($V_{UBC} = 0$) and axial flow exit from the blade system in turbine mode ($V_{UBC} = 0$), (1) taking into account hydraulic the efficiency of the flow path η_g will be for the pump mode

$$\frac{H_{pump}}{\eta_{g\,pump}} = \frac{1}{g}(V_{U\,pump} \cdot U_{U\,pump}) \quad (2)$$

and for turbine mode

$$H_{Turbine}\eta_{g\,turbine} = \frac{1}{g}(V_{U\,pump} \cdot U_{Turbine}). \quad (3)$$

If the circulations created or activated by the blade system are equal, it follows from (2) and (3) that for points with maximum efficiency, the ratio of the pressures of the turbine and pump modes

$$H_{Turbine}\,\eta_{max}/H_{Pump}\,\eta_{max} = 1/\eta_{gP}\cdot\eta_{gT}. \quad (4)$$

To determine the relationship between the flow rates $Q_{Turbine,\,max}/Q_{Pump,\,max}$ for the point with maximum efficiency in the turbine and pump modes, we consider the combined speed triangles in the turbine and pump modes on the pressure side of the impeller, shown in Figure 2.
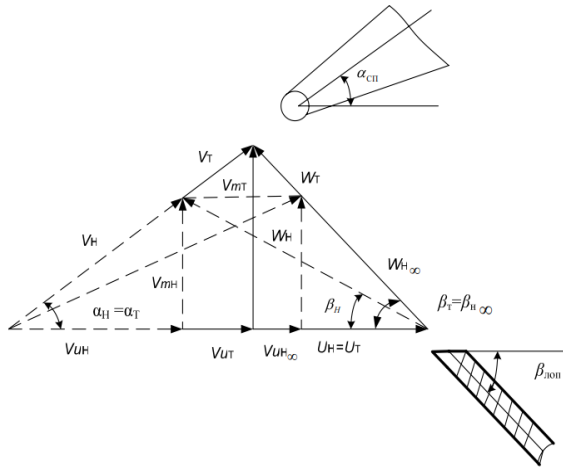

Figure 2: Combined speed triangles in turbine and pump modes on the pressure side of the impeller.

When calculating the geometry of the pump outlet elements, in order to achieve maximum efficiency of the flow path, the flow exit angle in absolute motion α_n must correspond to the outlet angle αsp. Consequently, in the turbine mode, the flow flows into the valve at an angle $\alpha_{sp} = \alpha_n$ in absolute motion, and the optimal mode is characterized by a minimum of shock losses at the inlet, i.e., the equality of the angles $\beta_t = \beta_{lop}$ in relative motion.

To achieve this, the flow rate V_{mT} must be higher in turbine mode than in pump mode. For the same

reason, a pump in turbine mode has maximum efficiency at high flow rates compared to discharge mode, which leads to approximately the same maximum efficiency value or, especially for low speeds, even higher in turbine mode than in pump mode [3]. The relative magnitude of volumetric and disk losses at the point of maximum efficiency is lower due to the greater power developed by the river. in turbine mode.

2 METHODS

For a scheme with an infinite number of blades, the flow angle in relative motion β_n at the outlet in pump mode coincides with the blade installation angle β_{lop}, which also occurs in turbine mode. Due to the flow deviation due to the influence of a finite number of blades in pumping mode, the flow angle β_p decreases while maintaining the value of the flow component of the velocity V_{mP} compared to the scheme of an infinite number of blades. Thus, the ratio of flow rates at the optimum characteristics $Q_{T,max}/Q_{P,max}$ mainly depends on the magnitude of the flow deviation, which in turn is determined by the geometric parameters of the blade system.

To determine the magnitude of the flow deviation within the framework of the jet theory, there are several types of corrections for a finite number of blades, specified by empirical dependencies proposed by a number of authors. However, this method is based on very rough assumptions, since the shape of the blade profiles, the geometry of the inlet and outlet, and spatial effects in the flow part of the pump are not taken into account.

In most of the reviewed works [1-7], this issue was studied on the basis of experimental data and theoretical analysis. In practice, in the absence of information about the parameters of the turbine mode, approximate calculation formulas proposed by a number of authors are used.

For example, the relationships between the values of pressure $H_T\eta_{max}/H_P\eta_{max}$ and flow rate $Q_T\eta_{max}/Q_P\eta_{max}$ for the point with maximum efficiency in turbine and pump modes are, according to Stepanov's monograph [10]:

$$h = \frac{H_T\eta_{max}}{H_P\eta_{max}} = \frac{1}{\eta_{gP}\cdot\eta_{gT}},$$
$$q = \frac{Q_T\eta_{max}}{Q_P\eta_{max}} = \frac{1}{\sqrt{\eta_{gP}\cdot\eta_{gT}}}, \quad (5)$$

which coincides with (4), and according to Sharma's data [11]:

$$h = \frac{H_{T\eta_{max}}}{H_{P\eta_{max}}} = \frac{1}{\eta_{maxP}^{0,8}},$$

$$q = \frac{H_{T\eta_{max}}}{H_{P\eta_{max}}} = \frac{1}{\eta_{maxP}^{1,2}}. \qquad (6)$$

The results obtained from these relationships, depending on the speed studied, the type and design of the pump, had a very large scatter of results - from ±2 to ±20% deviations from the experimental data [9].

Theoretical experimental studies conducted model studies of two versions of impellers I-1 and I-3 for turbine and pump operating modes in the flow part of a reversible hydraulic machine with a speed of 200. The flow part of the hydraulic machine model consists of a spiral chamber having a speed coefficient in the inlet section K = 0,87; stator with toroidal rings and number of columns Z_1 = 8 of constant height b_1 = 0,149 D_1; guide vane (GV) with 16 blades (Z_0 = 16) height b_0 = 0,149 D_1 and symmetrical profile, diameter of the rotation axes D_0 = 1,16 D_1; RK ORO 70/5215 (I-1) with diameter D_1 = 0,35 m and number of blades Z_1 = 6, diameter on the suction side D_{ss} = 0,6 D1; curved suction pipe with a turbine-type elbow.

In addition to the described flow part of the reversible hydraulic machine, theoretical experimental studies were carried out on two versions of the I-1 and I-3 in the flow part with the dismantled GV. In the absence of GV, the flow part is identical to the pump flow part, which makes it possible to use the results of theoretical experimental studies to obtain experimental characteristics and refine approximate formulas for calculating hydraulic parameters in turbine mode.

The experimental characteristics of the studied version of the flow path with the I-1 I for turbine (TM) and pump modes (PM) are shown in Figures 3–5. In pump construction, it is customary to use dimensionless coefficients of pressure ψ, supply - flow F and power π, which are calculated using $\psi = \frac{H}{n^2 \cdot D^2}$, $\psi = \frac{Q}{n \cdot D^3}$, $\pi = \frac{P}{n^3 \cdot D^5}$ and are related to the given values $n_1^!$ and $Q_1^!$, usually used in turbine construction, by the following relationships:

$$\psi = \left(\frac{60}{n_1^!}\right)^2, F = Q_1^! \cdot \sqrt{\psi}.$$

The research results confirm the main provisions obtained from literary sources and the analysis performed. The maximum efficiency value in turbine

mode occurs at higher pressures and flow rates than in pump mode, which leads to an increase in turbine power.
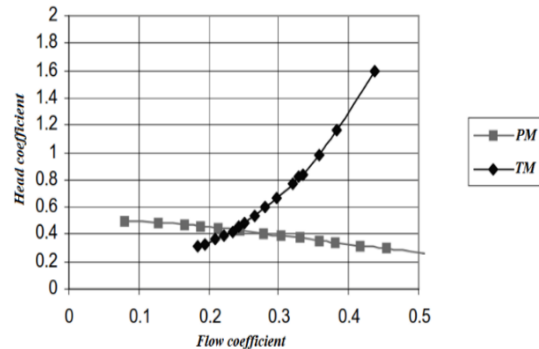

Figure 3: Dependence of the pressure coefficient in TM and PM on the flow coefficient.
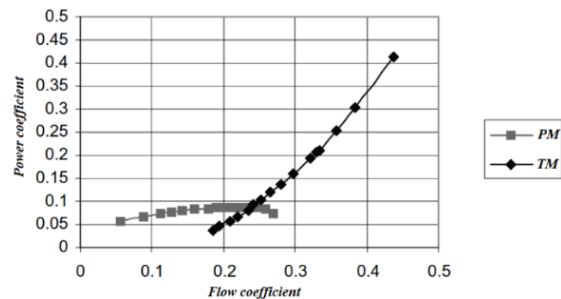

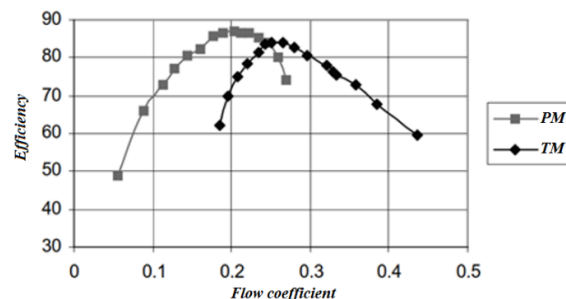Figure 4: Dependence of the power factor in TM and PM on the flow coefficient.

Figure 5: Dependence of efficiency in TM and PM on the flow coefficient.

3 RESULTS AND DISCUSSION

Experimental parameters in the optimum characteristics of turbine and pump modes for the studied flow parts in the absence of n. a. are given in Table 1.

Impeller type	Speed n	Efficiency		Experiment		Calculation by formula (5)		Calculation by formula (6)		Computational fluid dynamics calculation	
		η_{Pump}	$\eta_{Turbine}$	h	q	h	q	h	q	h	q
Pump [12]	297	0,91	0,84	1,25	1,379	1,263	1,238	1,2	1,119	-	-
Impeller - 1	205	0,91	0,882	1,238	1,208	1,245	1,116	1,1198	1,0783	-	-
Impeller - 3	195	0,934	9,894	1,09	1,137	1,195	1,0926	1,091	1,0597	-	-
Pump [5]	144	0,85	-	1,4	1,35	1,38	1,18	1,22	1,14	-	-
Pump [10]	56	0,89	-	1,518	1,25	1,52	1,23	1,15	1,097	1,55	1,219

Figure 6: Comparison of experimental and calculated values of h and q.

Table 1: Experimental parameters of impeller I-1 and I-3 at performance optimums.

Code of the flow part	Experimental parameters at performance optimums					
	Turbine mode			Pump mode		
	Q_i^t, m³/s	n_i^t, min⁻¹	η, %	Q_i^t, m³/s	n_i^t, min⁻¹	η, %
I-1	0,358	89,0	84,2	0,33	98,0	87,0
I-3	0,365	88,8	85,4	0,335	92,0	89,6

Comparison of experimentally obtained relationships between the values of pressure $h=H_T\eta_{max}/H_P\eta_{max}$ and flow rate $q=Q_T\eta_{max}/Q_P\eta_{max}$ at points with maximum efficiency in turbine and pumping modes for I-1 and I-3 with the results obtained from dependencies (5), (6), are given in Figure 6. For I-1, the values of h and q determined experimentally are closest to those determined by (4), which coincides with Stepanov's recommendations, and for I-3, they are closer to the values determined according to Sharma's recommendations. The indicated impeller have the same number of blades, values of b_0 and D_{vc} and blade installation angle β_{blade} on the pressure side, but differ in the geometry of the meridional section and the inter-blade channel. Optimization of the blade geometry leads to a significant increase in efficiency in turbine and especially in pump mode for I-3 and, accordingly, to a decrease in the values of h and q compared to the experimental ones. In addition, in Figure 6 shows a comparison of calculated and experimental data obtained from sources [3, 9] for pumps in the speed range n = 56-297. It also confirms the spread of the results of deviations of calculated and experimental data [9], reaching up to 5 in pressure and up to 12% in flow rate, and for the considered range of speeds Stepanov's recommendations turn out to be closer to the experimental ones. Therefore, statistical methods for processing data from experimental studies, which were used in [3, 6, 9, 19], cannot provide acceptable accuracy in predicting the hydraulic parameters of turbine mode. A promising way to determine the ratio of pressures and flow rates at the optimum characteristics is, along with experimental ones, the use of computational methods of computational fluid dynamics in a three-dimensional formulation, which is confirmed by the results obtained in [8-18].

4 CONCLUSIONS

Analysis of literature sources shows that the use of serial pumps as hydraulic turbines is a successful alternative to the use of specially designed hydraulic turbines for mini and small HPP.

To reliably use pumps as hydraulic turbines, manufacturers need to obtain experimental performance of pumps in turbine mode, which can help expand their markets and better utilize available hydraulic potential.

Subject to the availability of geometry and the accumulation of certain experience, it is possible to use methods for calculating three-dimensional viscous flow in the flow part to obtain the necessary hydraulic parameters of serial pumps in turbine mode with satisfactory accuracy.

REFERENCES

[1] S. Baumgarten, "Pumpen als Turbinen," Technik kompakt, KSB Pump company, July 2005, no. 11, pp. 2-9.

[2] A.A. Williams, "The turbine performance of centrifugal pumps: a comparison of prediction methods," Proc. IMech. Part A, vol. 208, pp. 59-66, 1994.

[3] A. Nourbakhsh and S. Derakhshan, "Mini and Micro Hydropower Stations for Production Inexpensive Energy," in HIDROENERGIA 2008-05-04, Intern. Conf. and Exhibition, SMALL HYDROPOWER, Bled-Slovenia, June 11-13, 2008.

[4] P. Singh, V. Ramasubramanian, and A. Rao, "Performance Evaluation of the Pump as Turbine based Micro Hydro Project in Kinko Village, Tanzania," in Himalayan Small Hydropower Summit, Dehradun, India, October 12-13, 2006, pp. 159-166.

[5] P. Maher, N.A. Smith, and A.A. Williams, "Assessment of pico hydro as an option for off-grid

electrification in Kenya," J. Renewable Energy, vol. 28, pp. 1357-1369, 2003.

[6] A. Rodrigues, P. Singh, and A.A. Williams, "Hydraulic Analysis of a Pump as a Turbine with CFD and Experimental Data," in IMech seminar Computational Fluid Dynamics for Fluid Machinery, London, Nov. 18, 2003.

[7] E. Kan, M. Mukhammadiev, K. Dzhuraev, and A. Abduaziz Uulu, "Assessment of economic efficiency of combined power plants based on renewable energies," in V International Scientific Conference "Construction Mechanics, Hydraulics and Water Resources Engineering" (CONMECHYDRO - 2023), E3S Web of Conferences, vol. 401, pp. 1-9, 2023, https://doi.org/10.1051/e3sconf/202340104005.

[8] E. Kan, M. Mukhammadiev, K. Dzhuraev, A. Abduaziz Uulu, and F. Shadibekova, "Methodology for determining technical, economic and environmental performance of combined power plants," in V International Scientific Conference "Construction Mechanics, Hydraulics and Water Resources Engineering" (CONMECHYDRO - 2023), E3S Web of Conferences, vol. 401, pp. 1-10, 2023, https://doi.org/10.1051/e3sconf/202340104036.

[9] V.V. Elistratov and M.M. Mukhammadiyev, "Integrated Use of the Hydropower Sources of Uzbekistan in the Context of Climate Change," Applied Solar Energy, vol. 59, no. 1, pp. 87-94, Feb. 2023.

[10] B. Bhandari, K.T. Lee, Y.M. Cho, and S.H. Ahn, "Optimization of hybrid renewable energy power system: a review," Renew Sustain Energy Rev, 2014.

[11] D.P. Navruzov, M.E. Madaliev, T.R. Pulatov, S.R. Juraev, and S.A. Boltayev, "Direct numerical simulation of flow in a flat channel with a rectangular step based on the nonstationary Navier-Stokes equations," in AIP Conference Proceedings 2612, 030007, 2023, Published Online: 15 March 2023, https://doi.org/10.1063/5.0135306.

[12] K.R. Allaev, "Modern energy and prospects for its development," Tashkent, 2021.

[13] Energy Storage Monitor, "Latest trends in energy storage 2019," www.worldenergy.org/assets/downloads/ESM_Final_Report_05-Nov-2019.pdf.

[14] Y.S. Vasilyev, V.V. Elistratov, I.G. Kudryasheva, M.M. Mukhammadiyev, and B.U. Urishev, "Use of maneuverable properties of hydroelectric power plants and hydro-accumulating power plants for improving reliability and operating efficiency of electric power systems of commonwealth countries (using example of power plants of Russia and the Republic of Uzbekistan)," E3S Web of Conferences, vol. 216, 01139, 2020, https://doi.org/10.1051/e3sconf/202021601139.

[15] Yu.S. Vasilev, V.V. Elistratov, and M.M. Mukhammadiev, "Integrated use of hydropower potential of Uzbekistan in the context of climate change," in International scientific and technical conference «New horizons for low-carbon development in the world and Uzbekistan», Sep. 23-24, 2022, Tashkent, pp. 62-64.

[16] M. Muhammadiev, B. Urishev, S. Juraev, and A. Boliev, "Detritus removal from a pumping-plant intake chamber with hydrajet pumps," in IOP Conf. Series: Materials Science and Engineering, vol. 883, 2020, https://iopscience.iop.org/article/10.1088/1757-899X/883/1/012123/pdf.

[17] G.I. Sidorenko, I.G. Kudryasheva, and V.I. Pimenov, "Economics of the installations of the non-traditional and renewable energy sources," in Technical and economic analysis: tutorial, ed. V.V. Elistratova and G.I. Sidorenko, SPB.: Politekhi Publisher, Un-ta, 2008.

[18] V.R. Elokhin, "Simulation modeling of energy systems," Vestnik Irkutsk State Technical University, 2013, no. 4 (75).

[19] V.V. Elistratov, M.A. Konishev, R.S. Denisov, and I.V. Bogun, "Arctic wind diesel power plant with intelligent automatic control system," Electricity, vol. 2022, no. 2, pp. 29-37.

Enhancing the Efficiency of a Gas Turbine Through the Integration of Solar Heat - a Case Study of Tashkent CHP

Jakhongir Normuminov and Zukhriddin Mukhiddinov

Tashkent State Technical University, University Str. 2, 100095 Tashkent, Uzbekistan

normuminovjakhongir@gmail.com, zukhriddinofficial97@gmail.com

Keywords: Solar Energy, Hybrid Solar-Gas Turbine Systems, Thermal Storage, Energy Efficiency, Exergy Analysis, Gas Turbines, Thermodynamic Modeling, Uzbekistan, Tashkent, Sustainable Energy.

Abstract: Solar power is currently a widely recognized energy source that has gained significant attention from numerous countries. Over time, two predominant solar technologies have emerged internationally: solar photovoltaics and solar thermal power. In the case of solar photovoltaics, sunlight is directly converted into electricity. However, despite advancements in thin-film solar photovoltaics, they still exhibit lower electricity output due to the uncontrollable and highly variable nature of rapidly changing weather conditions. Solar thermal power, on the other hand, harnesses solar heat in power plants to generate electricity. However, this type of power generation faces similar challenges. The lack of high-capacity thermal storage contributes to its inherent unpredictability and variability. In the coming years, the solar thermal power industry in Uzbekistan should prioritize harnessing the unique advantages offered by the combination of hybridization and thermal energy storage. With decreasing costs and the increasing adoption of renewable energies, solar thermal power plants have the capability to provide dispatchable power, making them highly suitable for a pivotal role in a future energy grid primarily powered by renewable sources.

1 INTRODUCTION

In Uzbekistan, the energy system has experienced stagnation over nearly a quarter of a century, with existing heat and electricity generating facilities constructed in the 1960s and 1980s struggling to meet the escalating demands of a growing population and expanding industrial sector. Consequently, the country has faced significant energy crises. Recent efforts have been undertaken to address this challenge, marked by decisions from the authority [1], including the establishment of a dedicated Ministry of Energy. This ministry represents the singular entity within the fuel and energy sector specifically tasked with ensuring the energy security of Uzbekistan, facilitating a stable supply of fuel and energy resources to economic sectors and the population, and advancing the promotion and development of renewable energy sources [2]. Notably, considerable emphasis has been placed on the construction of contemporary steam-gas stations, marking a departure from antiquated Rankine cycle devices characterized by low efficiency. However, it is imperative to acknowledge certain drawbacks associated with the adoption of steam-gas devices:

- High Fuel Consumption: Gas turbines, a key component of power plants, are acknowledged for their elevated fuel consumption. This characteristic poses the risk of increased operational costs and adverse environmental impacts.
- Low Part-Load Efficiency: A notable limitation of gas turbines is their diminished efficiency when operating at less than maximum capacity, commonly referred to as part-load conditions [3].

A potential solution to address these challenges involves the integration of solar power technology with the existing power plant infrastructure, which could result in improved efficiency and reduced fuel consumption.

2 STATE OF THE ART OF THE PROBLEM

In recent years, the utilization of gas turbines has experienced significant growth in industries such as petrochemicals, power generation, and offshore

operations. The power sector, in particular, has witnessed a surge in the adoption of combined cycle power plants, with the latest high-efficiency gas turbines playing a pivotal role in this expansion.

The increasing costs associated with natural gas have posed challenges for gas turbine plants originally designed for continuous base load operation. Many of these plants are now operating at varying loads, ranging from 50% to full load, and in some cases, they have to be shut down on weekends. The latest maintenance guidelines, including case studies, aim to support engineers working in the field who must operate these plants under non-ideal conditions that deviate from their original design parameters of base-loaded operation. Currently, numerous alternatives have been devised with a specific focus on reducing fuel consumption. Specifically, the integration of solar power into existing gas turbines is developing rapidly, and research is underway on many small-scale Solar Hybrid Gas Turbines (SHGT).

The outcomes of investigations in the realm of gas turbine hybridization have been announced by numerous researchers, accompanied by several authors reporting the findings of case studies focused on enhancing gas turbine performance through the utilization of solar power.

Faustino Moreno-Gamboa et al [3]. extensively documented the performance of an externally fired hybrid solar gas-turbine power plant in Colombia. Their study employed energy and exergy methods, taking into account local environmental conditions, such as ambient temperature and solar resource availability, as well as plant models and thermodynamic properties of the working fluid, utilizing the Modelica compiler Dymola. The results obtained from their investigation revealed that the efficiency of a solar concentration system is relatively less affected by solar radiation variations occurring around noon. However, the impact becomes more pronounced during sunrise and sunset periods. Notably, an escalation in radiation levels leads to heightened exergy destruction in both the heliostats field and the solar receiver, thereby causing a decline in the overall exergetic efficiency of the entire plant. Moreover, the effectiveness of the heat exchangers within the combustion chamber and the subsequent dissipation of heat into the environment have a direct and consequential influence on the performance of the hybrid solar plant. A decrease in the efficiency of these heat exchangers further exacerbates the detrimental effects, impairing the plant's overall functionality and efficiency.

The study conducted by Guzman [4] provides valuable insights into the operational characteristics and optimization potential of the parabolic trough solar power plant in Barranquilla. By employing the SAM, a widely recognized tool for analyzing renewable energy systems, the study contributes to the academic understanding of solar power plant performance in the specific geographical context of Barranquilla. The results underscore the plant's ability to generate substantial electricity output throughout the year, with May emerging as the peak production month. Moreover, the efficient utilization of solar radiation during peak periods obviates the requirement for supplementary stored energy, further enhancing the plant's operational efficiency and economic viability.

T. Prosin et al [5]. conducted an investigation on the initial modeling results of a centrifugal receiver system, specifically focusing on a turbine with a power output of 100 kW and a more efficient 4.6 MW unit. The study concludes that the payback time for the 100 kW microturbine is significantly low, indicating its economic feasibility. Moreover, the 4.6 MW unit demonstrates considerable potential and represents a promising market opportunity.

A study by Wujun Wang et al [6]. demonstrated the capabilities of load flexible combustors for hybrid Brayton applications. Using the KTH high flux solar simulator, the researchers experimentally investigated combustor performance at 30% workload. To test behavior under varying conditions, they abruptly halted the solar simulator and simultaneously increased fuel injection. This enabled thorough examination of the combustor's performance. Despite terminating solar input, the results showed that manually adjusting the main fuel injection, maintained air outlet temperatures within 914.5± 10 °C, preventing flashback. Overall, by demonstrating stable outlet temperatures even when solar input was suddenly lost and fuel supply increased, the study successfully showcased the load flexibility of the combustor for hybrid Brayton systems across challenging operational scenarios.

However, more appealing papers are shortly described below.

P. Klein et al [7]. conducted an experimental test program to obtain high-temperature heat transfer data for a packed bed operating within two temperature ranges: 350-900 °C and 600-900 °C. The researchers utilized flue gas from a 45kW LPG burner to heat a packed bed of Denstone ceramic pebbles. To achieve the desired temperature ranges, the system was preheated before testing. The study involved measuring fluid and solid temperature profiles in the

packed bed along both axial and radial dimensions. These measurements were then compared to a numerical model, showing a reasonable level of agreement. Additionally, the paper discusses potential modifications to the test facility and outlines plans for future testing.

R. Uhlig et al [8]. carried out an investigation to enhance the efficiency of Solar Hybrid Gas Turbine (SHGT) cavity receivers. The paper aimed to compare two different approaches to improve the receiver efficiency of a cavity receiver used in heating compressed air for a 4.7 MWel turbine. The desired temperature range was between 330°C and 800°C, with a mass flow rate of 15.9 kg/s at an absolute pressure of 10 bar. The researchers also analyzed the influence of various receiver sizes and explored a design option that involved incorporating a transparent covering for the aperture. The objective of this design option was to reduce convection losses. Furthermore, the study considered the impact of these enhancements on the levelized cost of energy (LCOE). The thermal and hydraulic layout was meticulously planned, accounting for a thermal heat input of 8.4 MWth and a 250 mbar pressure drop. Thermal finite element (FE) models were utilized for this purpose. These models took into consideration the local heat flux distribution, heat transfer to the working fluid, radiation exchange between components and the surrounding environment, as well as conduction and convection losses within the cavity.

The paper presents the groundbreaking Solugas project [9], which marks the introduction of a megawatt-scale solar hybrid gas turbine system. This innovative system combines a solar tower plant with a customized Mercury 50 gas turbine. At the heart of the system is the solar receiver, comprising 170 tubes crafted from nickel alloy. These tubes efficiently heat pressurized air to an impressive 800°C. The total solar input power generated by this setup amounts to approximately 7 MWth. Integration of the modified gas turbine with the solar receiver results in a net power output of 4.6 MWe. This remarkable achievement signifies the first-ever successful demonstration of solar hybrid gas turbine technology operating at a multi-megawatt capacity. The Solugas project serves as a testament to the technical feasibility and potential of solar hybrid gas turbines in generating cost-effective and dispatchable renewable power. The success of this project opens up exciting prospects for further capacity scaling in the future. With its maximum power output of 4.6 MWe, the Solugas project represents a significant milestone in the advancement of solar hybrid gas

turbine technology. It showcases the successful construction and testing of a cutting-edge system that paves the way for a cleaner and more sustainable energy future.

Despite the numerous results announced in the field of solar hybrid micro gas turbines, in this paper, we investigate the possibilities of integrating solar energy into the larger MW-scale Tashkent CHP gas turbine.

3 MODELLING OF GOVERNING EQUATIONS

3.1 Gas Turbine CHP Thermodynamic Model

To assess the electrical efficiency of a gas turbine, it's essential to take into account the various components of the turbine. Specifically, we need to calculate the actual work done by the gas turbine compressor.

This work calculation is represented in Figure 1.

3.1.1 Calculating Electrical Efficiency

Assuming that the compressor adiabatic thermal efficiency is η_c [10].

$$W_{ca} = \frac{m_a(h_2 - h_1)}{\eta_c},$$

where: m_a – mass flow of air (kg/s); h_1 – enthalpy of air at compressor inlet (kJ/kg); h_2 – air enthalpy at compressor outlet (kJ/kg).

Actual generated work of turbine is determined by equation [12]:

$$W_{ta} = (m_a + m_f)(h_3 - h_4)\eta_t,$$

where: m_f – mass flow of fuel (kg/s); h_3 – enthalpy at gas turbine inlet; h_4 – enthalpy at gas turbine outlet.

Thus, actual total work is:

$$W_{act} = W_{ta} - W_{ca}.$$

The overall internal efficiency of a gas turbine can be assessed using a specific equation:

$$\eta_c = \frac{W_{act}}{Q_F} = \frac{W_{act}}{m_f(LHV)},$$

where m_f – fuel mass flow to raise temperature and it is calculated by following equation:

$$m_f = \frac{h_3 - h_2}{(LHV)\eta_b},$$

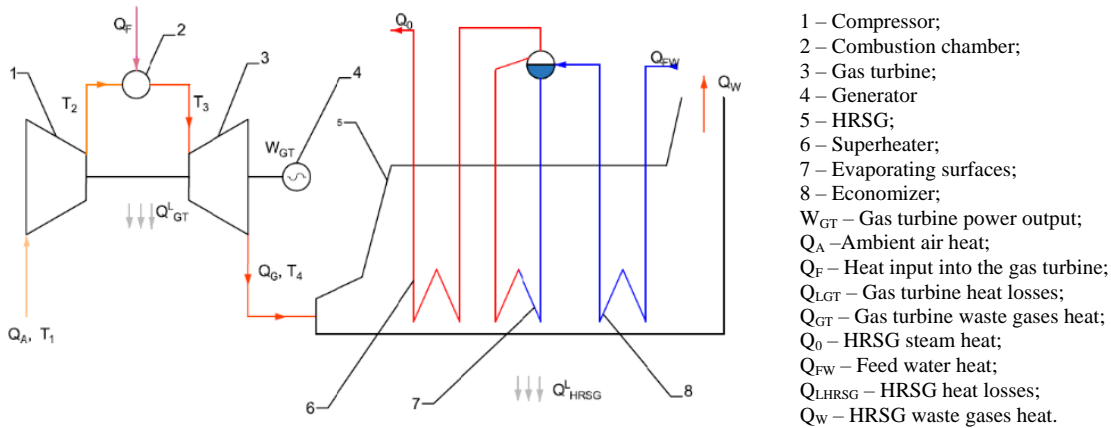where $h3$ – enthalpy at gas turbine inlet (kJ/kg).

Figure 1: Energy flow diagram of the gas turbine CHP without supplementary firing.

3.1.2 Heat Recovery Steam Generator

The heat output of the Heat Recovery Steam Generator (HRSG), which represents the heat absorbed by the water and steam within the HRSG, is calculated as follows:

$$Q_{HRSG} = Q_0 - Q_{FW.}$$

It also possible to determine the heat output of the HRSG by equation:

$$Q_{HRSG} = M_{GT}(I_{GT} - I_w) = D_0(h_0 - h_{fw}),$$

where: M_{GT} – the waste gases mass flow through turbine (kg/s); I_{GT} – gas turbine outlet enthalpy (kJ/kg); I_W – enthalpy of HRSG waste gases (kJ/kg); D_0 – H_{RSG} steam mass flow (kg/s); h_0 – H_{RSG} superheated steam enthalpy (kJ/kg); h_{fw} – H_{RSG} feed water enthalpy (kJ/kg).

The equation that defines the enthalpy of waste gases entering the evaporating surfaces at the inlet, denoted as IE, can be rephrased as follows:

$$I_E = c_{p.g} \times (t_s + \partial t_E),$$

where: t_s – saturated steam temperature in appropriate temperature; ∂t_E - Temperature difference between waste gasses and water evaporating surface inlet (°C).

The HRSG mass flow is defined by equation:

$$D_0 = \frac{G_{GT} \times (I_{GT} - I_E)}{(h_0 - h_E)}.$$

Heat balance equation for economizer:

$$G_{GT} \times (I_E - I_W) = D_0(h_0 - h_{FW}).$$

It is possible to define the HRSG waste gasses temperature:

$$t_W = t_E - \frac{D_0(h_E - h_{FW})}{c_{p.g} G_{GT}}.$$

where t_E - The waste gasses temperature at evaporating surface inlet (°C).

Waste gasses temperature at evaporating surfaces inlet t_E is defined according to the enthalpy IE. Now it is possible to define HRSG efficiency (utilization coefficient):

$$\eta_E = \frac{Q_{HRSG}}{G_H^E \times LHV}.$$

3.2 Gas Turbine with Solar Tower

To combine gas turbine and solar tower technologies for a hybrid solar power plant, Figure 2 illustrates the layout typically employed for utility-scale installations. Given the substantial power block size, the gas turbine must be located at the tower's base. Consequently, it becomes essential to transport the hot gases up and down the central tower, which introduces additional constraints related to temperature. To harness the sun's energy at elevated temperatures, it is necessary to employ high concentration ratios. These ratios help minimize radiation losses and ensure that the receiver maintains an acceptable level of efficiency.
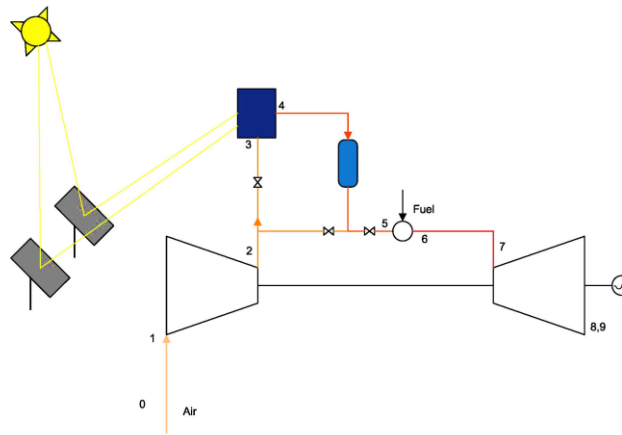
Figure 2: Schematic diagram of the simple-cycle hybrid solar gas-turbine: 1 – heliostats; 2 -solar receiver; 3- heat storage.

The amount of high-pressure compressor air diverted for cooling turbine blades and other purposes depends on the temperature at the combustion chamber's outlet. The hotter the combustion gases, the greater the volume of air needed for cooling. Some additional air is also taken for purging and sealing to prevent high-temperature gases from reaching unwanted areas. The remaining compressor airflow is directed to the solar subsystem, where it passes through the central tower's piping. Afterward, the preheated air from the solar receiver is sent to the combustion chamber, where a specific quantity of fuel is injected and burned to achieve the desired firing temperature. Solar preheating of compressor air significantly reduces fuel consumption, and adjustments to the combustion chamber's fuel flow can be easily and rapidly made to compensate for fluctuations in solar heat input, ensuring stable operation.

The extent of solar integration within a hybrid solar power system can be quantified using the solar share, denoted as f_{sol}. This parameter is defined in following equation as the ratio of the solar heat input (Q_{sol}) to the total heat input (Q_{tot}) for the entire cycle:

$$f_{sol} = \frac{Q_{sol}^+}{Q_{tot}^+}$$

The nominal solar share of a hybrid solar gas-turbine system is directly linked to key cycle temperatures, particularly the nominal receiver temperature (T_4) and the combustor outlet temperature (T_6). When the receiver temperature approaches the combustor outlet temperature, the solar share increases because less heat needs to be supplied through combustion. In the extreme case where T_4 equals T_6, the nominal solar share reaches 100%, and there is no need to burn any fuel. Neglecting losses in the tower piping, you can estimate the nominal solar share using (1), with T_5 representing the compressor discharge temperature and T_3 the receiver inlet temperature.

$$f_{sol,nom} \cong \frac{T_4 - T_3}{T_6 - T_5} \qquad (1)$$

The combustor outlet temperature exhibited by a contemporary industrial gas-turbine typically falls within the range of approximately 1400°C, whereas tower-mounted solar receivers presently encounter limitations with temperatures below 950°C. Consequently, the upper threshold for the maximum nominal solar fraction achievable by a conventional industrial gas-turbine is estimated to hover around 50%.

To mitigate the collective carbon emissions, a pivotal parameter for power plants revolves around the annual fuel-electric conversion efficiency. This metric is contingent not only upon the efficiency of the power block conversion and the nominal degree of solar integration but, more significantly, on the temporal extent of solar operation. For an idealized cycle, the annual fuel-electric efficiency can be approximated using following equation, drawing upon the conversion efficiency (η_{cycle}), the nominal solar share ($f_{sol,nom}$) of the power generation cycle, the total operational hours (hop), and the equivalent number of full-load solar operating hours ($h_{sol,eq}$) [13]. Once the conversion efficiency and nominal solar share of the hybrid solar gas-turbine power plant are established, the annual fuel-electric efficiency is contingent solely upon the fraction of total operating hours during which solar heat is furnished to the system. Augmenting the duration during which the solar collector imparts nominal heat to the system serves as a means to enhance the overall fuel-electric efficiency [11].

289

$$\eta_{fuel} = \frac{\eta_{cycle}}{1 - f_{sol,ann}} = \frac{\eta_{cycle}}{1 - f_{sol,nom} \times \frac{h_{sol,eq}}{h_{op}}}$$

The dimensions of the solar collector field can be articulated in relation to the solar multiple (SM), as delineated by (2), where it is defined as the quotient of the nominal thermal power dispensed by the field (Q_{field}) to the nominal power requisitioned by the receiver (Q_{rec}) [15]. The nominal yield from the heliostat field is commonly specified, accounting for a direct normal irradiation of 850 W/m2, precisely at solar noon during the Equinox (either on the 21st of March or the 22nd of September).

$$SM = \frac{Q^+_{field,nom}}{Q^+_{rec,nom}} \qquad (2)$$

The tower piping can be conceptualized as a counter-flow heat exchanger when considering heat transfer. The comprehensive heat transfer, denoted as Q, can be calculated using the equation below, based on the effectiveness-NTU (Number of Transfer Units) methodology. In this equation, ε represents the heat exchanger effectiveness, M_{main} signifies the mass flow rate through the tower, c_p denotes the specific heat capacity of the fluid, and T_2 and T_4 stand for the temperatures at the compressor discharge and receiver outlet, respectively.

$$Q = \varepsilon \times M_{main} \times c_p \times (T_4 - T_2)$$

Pressure losses within the tower piping arise from various phenomena. The total pressure drop, denoted as ΔP_{tot}, is determined by following equation, which sums up the losses occurring during the extraction of the main air-flow after the compressor (ΔP_{ext}), the friction losses within the piping itself (ΔP_{pipe}) during both ascent and descent of the tower, and the losses incurred during the return of the heated air to the gas turbine (ΔP_{ret}) [15].

$$\Delta P_{tot} = \Delta P_{ext} + \sum \Delta P_{pipe} + \Delta P_{ret}$$

4 ANALYSIS OF THE TASHKENT GAS TURBINE CHP

In the Tashkent Combined Heat and Power (CHP) facility, a Hitachi gas turbine has been installed, boasting a power output of 28.1 MW and an electrical efficiency of 34.2%. The flue gases emanating from the turbine are directed towards the Heat Recovery Steam Generator (HRSG) at a temperature of 553°C. The primary objective of the HRSG is to generate steam with specific parameters: 36 bar pressure, 416°C temperature, and a mass flow rate of 46 tons per hour. Subsequently, the produced steam is channeled to the steam turbine for further utilization.
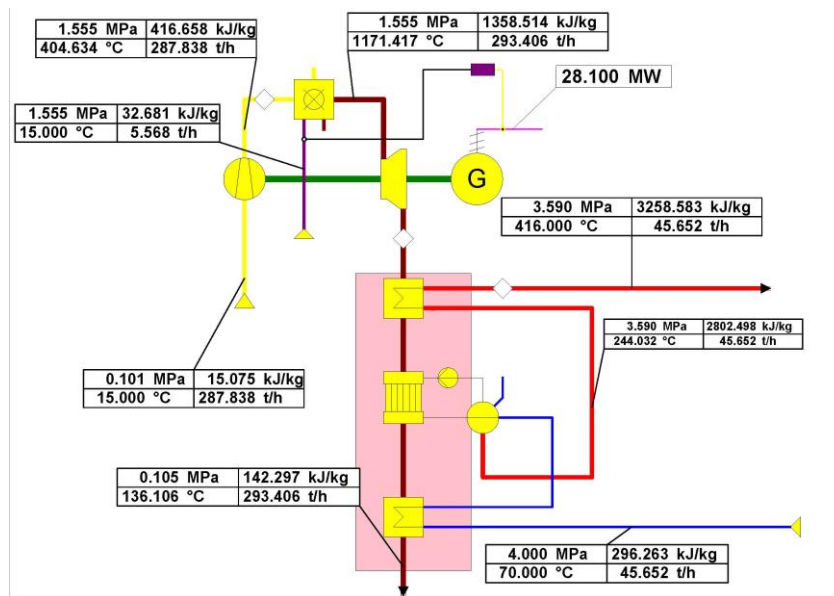

Figure 3: A simulation model of the H27 gas turbine combined heat and power (CHP) system has been created using Ebsilon®Professional software.

The feedwater introduced into the HRSG system initiates at a temperature of 70°C and a pressure of 40 bar. Initially, the feedwater undergoes a heating process within the economizer, elevating its temperature to 240°C, before being directed to the steam generator. The steam generator is responsible for generating saturated steam at a temperature of 244°C [15].

This comprehensive system ensures the efficient conversion of gas turbine exhaust heat into steam, subsequently harnessing the thermal energy to drive a steam turbine, exemplifying a cogeneration approach for enhanced energy utilization in the Tashkent CHP facility.

To investigate the correlation between CO_2 emissions and the solar share, a computational model of the extant H27 gas turbine combined heat and power (CHP) system was constructed using Ebsilon®Professional software. The simulation model, conducted at an ambient temperature of +15°C, is presented in Figure 3.

Additionally, the Carnot factor-energy diagram of the Heat Recovery Steam Generator (HRSG) at the same ambient temperature is illustrated in Figure 4.
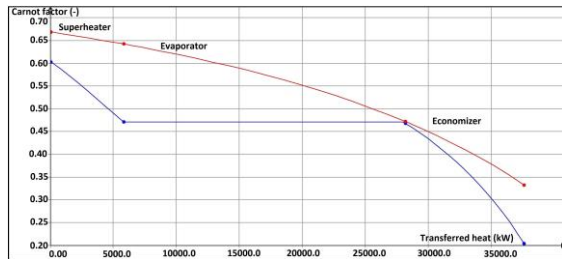

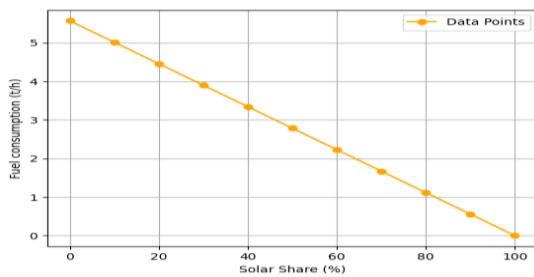Figure 4: Carnot factor-Energy diagram of HRSG.

Figure 5: Fuel consumption-Solar share of 28.1 MW SHGT.

Ebsilon software was employed for the comprehensive analysis of emissions emanating from an extant H27 gas turbine in correlation with its power output. The simulation model, presented in Figure 3, is representative of a scenario wherein the atmospheric temperature attains a specific value, specifically +15°C. Furthermore, Figure 4 elucidates the Carnot-factor-Energy diagram, offering insights into the thermodynamic efficiency of the system under consideration.

Figure 5 illustrates the changes in fuel consumption resulting from the integration of a solar tower device into a 28.1 MW gas turbine. Concurrently, as fuel consumption decreases, the emissions of harmful gases also decline proportionally

5 CONCLUSIONS

The integration of a solar tower into the 28.1 MW gas turbine, as simulated in the Ebsilon program, has yielded noteworthy outcomes with implications for both environmental sustainability and operational efficiency. The analysis of the hybrid system reveals a discernible reduction in fuel consumption, a critical metric indicative of improved resource utilization and decreased reliance on traditional fossil fuels. This decrease in fuel consumption, as depicted in Figure 5, correlates directly with a proportional reduction in the emission of harmful gases, showcasing the potential environmental benefits of incorporating renewable energy sources

Furthermore, the simulation results underscore the adaptability and efficacy of the hybrid system in response to varying solar shares. The correlation between the solar share and fuel consumption suggests that the system can be dynamically optimized to harness solar energy efficiently, thereby enhancing overall performance and minimizing the environmental footprint [16]. The Ebsilon simulation offers valuable insights into the intricacies of the integrated system, enabling a comprehensive understanding of its behavior under diverse operating conditions.

In conclusion, the integration of a solar tower into the 28.1 MW gas turbine, as elucidated by the simulation results, holds promise for advancing sustainable energy practices. The observed reductions in fuel consumption and emissions underscore the potential of such hybrid configurations in contributing to cleaner and more efficient power generation systems. These findings contribute to the growing body of knowledge aimed at fostering environmentally conscious and technologically innovative solutions within the realm of power generation.

REFERENCES

[1] Presidential Decree titled "On Measures to Organize the Activities of the Ministry of Energy of the Republic of Uzbekistan" (PD-4142).

[2] M.P. Boyce, Gas Turbine Engineering Handbook, Fourth Edition, 2012.

[3] F. Moreno-Gamboa, A. Escudero-Atehortua, and C. Nieto-Londoño, "Performance evaluation of external fired hybrid solar gas-turbine power plant in Colombia using energy and exergy methods," 2020.

[4] L. Guzman, A. Henao, and R. Vasquez, "Simulation and optimization of a parabolic trough solar power plant in the city of Barranquilla by using system advisor model (SAM)," 2013.

[5] T. Prosin, T. Pryor, C. Creagh, L. Amsbeck, and R. Uhlig, "Solar gas turbine systems with centrifugal particle receivers, for remote power generation," 2014.

[6] W. Wang, T. Pan, M. Swanteson, and T. Strand, "Experimental demonstration of a load flexible combustor for hybrid solar Brayton applications," 2023.

[7] P.Klein, T.H. Roos, and T.J. Sheer, "Experimental investigation into a packed bed thermal storage solution for solar gas turbine systems," 2013.

[8] R. Uhlig, R. Flesch, B. Gobereit, S. Giuliano, and P. Liedke, "Strategies enhancing efficiency of cavity receivers," 2013.

[9] M. Quero, R. Korzynietz, M. Ebert, A.A. Jiménez, A. del Río, and J.A. Brioso, "Solugas – Operation experience of the first solar hybrid gas turbine system at MW scale."

[10] E. Matjanov, "Gas turbine efficiency enhancement using absorption chiller: Case study for Tashkent CHP," 2020.

[11] J.D. Spelling, "Hybrid Solar Gas-Turbine Power Plants: A Thermoeconomic Analysis," 2013.

[12] K.A.B. Pathirathna, "Gas turbine thermodynamic and performance analysis methods using available catalog data," 2013.

[13] J. Normuminov, M. Tursunov, A. Unarov, and A. Kuchkarov, "Increasing the efficiency of the use of oil fuel in thermal power stations and boilers."

[14] J. Normuminov, A. Anarbaev, J. Tulkunov, R. Zakhidov, and B. Xurramov, "Rational Solutions for Automatic Control of a Solar Heating System."

[15] T.G.C. Veloso, U. Gampe, and S. Glos, "Optimization strategies of different SCO_2 architectures for gas turbine bottoming cycle applications."

[16] S. Rath, U. Gampe, and A. Jäger, "A Numerical Algorithm for Calculating Critical Points and Its Application to Predictive Mixture Models and Binary CO_2 Mixtures."