by Anhalt University of Applied Sciences and Perm National Research Polytechnic University

Proceedings of the 11th International Conference on Applied Innovations in IT

Volume 11

Issue 1

EDITION
Hochschule Anhalt

Proceedings of the 11th International Conference on Applied Innovations in IT

Volume 11 | Issue 1

Koethen, Germany
9 March 2023

Editorial boad:

Editor in chief

Prof. Dr. Eduard Siemens (Anhalt University of Applied Sciences)

Topical Editors

Prof. Dr. Halina Falfushynska (University of Rostock),

Prof. Dr. Hassan Bevrani (University of Kurdistan), Prof. Dr. Ralph Gottschalg (Fraunhofer IMWS),

Prof. Dr. Mykhailo Ilchenko (Igor Sikorsky Kyiv Polytechnic Institute),

Prof. Dr. Douglas Hall (St. Mary's University)

This jornal contains publications of the International Conference on Applied Innovations in IT (ICAIIT), which took place in Koethen March 9th 2023. 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.

ISBN: 978-3-96057-161-2 (Online)
ISSN: 2199-8876

All papers published in this jornal are full-open access:
immediately freely available to read and download.

Publisher: Anhalt University of Applied Sciences
Bernburg / Koethen / Dessau
http://www.hs-anhalt.de

Postfach 1458
D-06354 Koethen, Germany
Phone: +49 3496 67 2327
Email: eduard.siemens@hs-anhalt.de
Web: http://icaiit.org

Content

Section 1 Communication and Data Transport Technologies

Section 2 Data Analysis and Processing

Section 3 Statistical Modelling and Analysis and their Applications in Life Sciences

Section 4 Control and Automation

Management of Digital Streams of an Autonomous System by the Raw Socket Ethernet Channel Virtualization Method in Linux OS

Victor Tikhonov[1], Serhii Nesterenko[2], Olena Tykhonova[1], Oleksandra Tsyra[1], Olha Yavorska[1] and Vladyslav Hlushchenko[1]

[1]*Department of Computer Engineering and Information Systems, State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, Odesa, Ukraine*

[2]*Odesa Polytechnic National University, Shevchenko Avenue 1, Odesa, Ukraine*

{victor.tykhonov, elena.tykhonova}@suitt.edu.ua, sa_nesterenko@ukr.net, aleksandra.tsyra@gmail.com, yavorskayao7@gmail.com, vlad.gluschenko.97@gmail.com

Keywords: Autonomous System, Digital Stream Management, Channel Virtualization, Ethernet, Raw Socket, Linux.

Abstract: The issues of multimedia data transfer in packet-based networks have been considered in this work. Known approaches to digital streams management in autonomous systems studied with respect to the Internet of things traffic requirements and sensor networks real time telemetry provisioning. An original method of telecommunication channel virtualization presented based on Ethernet Raw Socket technique. An algorithm of digital flows integrating described for data multiplexing by conveyor transporting modules which payload the 802.3 Ethernet frame instead of conventional IP-packets. Computer model exhibited for simulation the process of joint multimedia data transmission via Ethernet frames in the form of four Python scripts under the Linux OS. Scripts SEND and RECEIVE implement the physical layer of Raw Socket interface. The scripts MUX and DEMUX perform multiplexing and de-multiplexing of integrated multimedia data on the data link layer. The results of the work intended for next generation networks application and Big Data distributed systems.

1 INTRODUCTION

The problem statement and the research purpose.

The global telecommunication industry has achieved outstanding success, which the main is the modern Internet. According to the International Telecommunication Unit (ITU) concept, a next generation network (NGN) should integrate multimedia services based on the Internet Protocol (IP) Multimedia Subsystem (IMS) platform. This concept considers a certain compromise between the existing information infrastructure of the Internet and new challenges of increasing quality of service (QoS) requirements, as well as the quality of content perception (QoE) [1]. One of the IMS concerns is ensuring the QoS aware transmission of real-time traffic (audio/video conferences, digital telemetry of sensor networks, etc.). This issue occurs due to stochastic data delay fluctuations while transporting IP-packets, as well as packet loss due to communication channel congestions. The above factors result in voice/video deterioration, end user connection instability, and slows down the data exchange rate in real-time control systems [2].

The improvement of NGN transport is carried out by coherent optical communication (COC) technologies and multiprotocol label switching (MPLS) transport profile (MPLS-TP), enhanced data flow control based on modern network operation researches. Among them, the software defined networking (SDN) concept has being actively developed in recent years. The SDN technology provides virtualization of given network physical topology within an autonomous telecommunication transporting system (AS) by network resources dynamic reconfiguration in accordance with the current data flow traffic. The overall data link capacity is virtually split to parallel channels by central controller of AS. The SDN networks studied by domestic and foreign scientists, and significant results have been achieved in this direction [3-10]. Enhancement the IMS-traffic transmission has been

also studied under the concept of multipath routing [11, 12]. Known approaches for telecom network and systems management mainly focus given potentials of embedded technologies and standard protocols. The QoS real-time data delivering means network resources reservation by virtual connection establishment among terminal network entities.

Though it goes beyond a distinct AS administration policy due to the fixed IP predetermined for IP-networks interoperability. Nevertheless, IP (in two standard versions IPv4 and IPv6) remains mandatory protocol for the Internet AS-to-AS interaction according to the NGN/ITU model, because of the most modern network services and applications are still based on IP. Thus, the real-time end-to-end data transport via the multilayer IP-packet encapsulation suffers an excessive overhead and the lack of efficiency. So, novel principles needed for multimedia IP packet-data transporting to overcome emerged challenges.

Given this fact, a fairly competitive mechanism to advance QoS aware end-to-end multimedia traffic transmission in the next few decades will be enhancement the local packet transfer function within any distinct autonomous transporting system, and first of all, at the L2/OSI data link layer [13].

The most common data-link-layer interfaces in modern digital network devices support either wired twisted pair Ethernet or/and Wi-Fi radio-Ethernet. To our opinion, exactly the Ethernet standards' utilization is one of the most promising direction for further development the Internet packet-based transport function for real-time multimedia applications.

The Ethernet/Wi-Fi technology operates with protocol data units (PDU) of three main types: line coding symbols (elementary protocol data units), block coding symbol groups (conventional "bytes" ranging in size from 2 to 8-12 or more bits) and, finally, frames channel coding. These three coding types and their corresponding protocol data units (PDU) define de facto three conditional data-link layer sublayers. The first two of them (line coding and block coding) are usually referred to as the so-called MAC sublayer - the access to the media-environment data transmission (twisted copper pair, fiber optic or radio broadcast). Ethernet frames are an object of the LLC sublayer - the control of data transfer from the data link to the network layer. There are a fairly large number of Ethernet technology standards supported by most network equipment manufacturers.

IEEE 802.3 is considered the basic standard that regulates the structure of an Ethernet frame. An Ethernet 802.3 frame, from the point of view of the LLC sublayer, has a header of 3 fields (the 6 octets frame destination MAC-address in the Ethernet local network; the 6 octets source MAC-address; the 2 octets "Type/Length" field); payload field of 46 octets minimal and 1500 octets maximal size; the 4-octets frame checksum field, which is often included into the frame header. In general, the overhead fields of an Ethernet frame at the LLC sublayer are 6+6+2+4=18 octets. If the frame payload is less than 46 octets, the payload field must be padded to 46 octets with non-content buffer information (padding).

In addition, two more service fields are added to the Ethernet frame at the MAC sublevel: preamble for physical synchronization of the frame receiver with the transmitter symbol sequence (7 octets - 56 bits of the form 101010...10); Frame Start Delimiter (FSD) in the form of octet 10101011. The appearance of the last 1 in the FSD octet means that the next character is the first bit destination MAC-address. Between any two consecutively transmitted Ethernet frames an Inter Frame Gap (IFG) must be inserted in the form of the signal absence for a period of at least 12 conditional octets; the physical size of the IFG depends on the nominal standardized bitrate in the Ethernet channel (starting from 10/100 Mbps and further up to 1/10/100 Gbps or more).

The typical payload of an Ethernet frame is a standard IP-packet. However, for the virtualization of communication channels at the L2 level, a non-trivial task arises of splitting an integral frame into separate virtual parts (transport containers) of variable length within the overall standardized frame payload size of 1500 octets. Some network operating systems (in particular, Linux) provide for the possibility of non-standard use of Ethernet frames and encapsulated protocols at different OSI levels using the so-called sockets (logical connections). In particular, an applied programmer is granted to artificially generate the TCP/UDP transport segments, IP-packet headers, and Ethernet frames as a whole. The non-trivial frames generation on the transmitting side of the channel (and the non-standard interpretation of these frames on the receiving side) corresponds to the lowest level of programming specialized application interfaces (the so-called Raw Ethernet sockets). At the same time, for the normal operation of Raw Ethernet sockets, it is necessary to correctly form the destination MAC-address (DMAC), as well as the source MAC-address (SMAC) in the first 12 octets of the frame header. The other two fields of the standard Ethernet frame (2 octets of the Type/Length field and 4 octets of the checksum field) may be freely formed and added to the frame payload [14].

The objective of this work is further developing of digital flows management based on multimedia data integration via the scalable LLC-sublayer multiplexing and routing-less IP-packets transportation through an autonomous system.

To achieve this, the task was set to jointly transfer data of different types with different sender and recipient addresses in one raw socket Ethernet frame. To experimentally verify this approach, a software package was developed in the python/Linux environment, which consists of four main modules: SEND – the raw socket cyclic frame transmission; RECEIVE – the raw socket periodic frame reception; MUX and DEMUX – the integrated stream segments multiplexing and demultiplexing. For testing the actual physical transmission of raw socket frames, two types of interaction are implemented: the master-slave Ethernet link via the 8-wire copper twisted-pair cross cable patch-cord, and radio-link via the Wi-Fi access point. The more detail results of this study are discussed below.

2 INTEGRATION THE MULTIMEDIA STREAMS IN VIRTUAL CHANNELS OF RAW SOCKET ETHERNET

The Figure 1 shows the principles of multimedia data integration over a packet network by virtualizing an IP-over-Raw Socket Ethernet link. The two OSI levels are presented here: the L3 network layer (IP-packets queue of three different multimedia streams F1, F2, F3), the L2 data-link layer （divided into two Ethernet technology sub-levels MAC and LLC). The data blocks of three different IP-packet's payloads are formed by the segments DS1, DS2, DS3 (each can include any arbitrary part of the corresponding IP-packet). Besides, the correspondent command segments CS1, CS2, CS3 are formed for individual treatment of these three data streams. The sizes of DS payload segments and CS command segments are variable.

Further, at the L2 LLC sub-level, a queue of segments (CS, DS) of individual streams is reshaped into payload transporting modules of a given size (Payload in Figure 1) within the allowable payload of an Ethernet frame (it was indicated above that this size can exceed 1500 octets due to the simplified Raw Header Socket Ethernet). The method of converting an IP-packets queue into a sequence of Payload_TM transport modules can be chosen arbitrarily depending on the individual requirements for the individual QoS requirements of multimedia streams. Finally, at the L2 MAC layer, the Payload_TM transport modules fit into Raw Socket Ethernet frames.

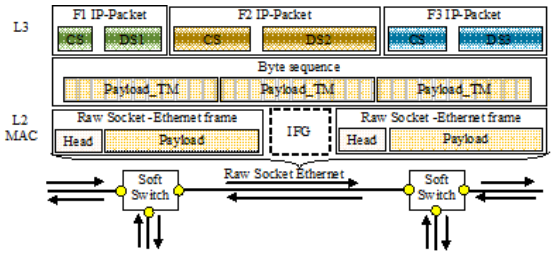


Figure 1: Transfer of multimedia traffic over the network IP/Raw Socket Ethernet.

The Raw Socket Ethernet frames circulate in both directions of a duplex link with 802.3 Ethernet network interfaces (or Wi-Fi radio-Ethernet); at the same time, instead of conventional Ethernet switches of the L2/OSI level, it is proposed to use modified flexible Soft-Switch switches to support the conditional L2.5/OSI level. The peculiarity of such a Soft-Switch is that an autonomous system enabled for faster interior IP-packets transfer by simple switching the separate DS units over the Raw Socket Ethernet virtual channels using the CS-labels; therefore, no routing function is needed more within the given autonomous system.

This way of transmitting IP packets is like MPLS technology (Multi-Protocol Label Switch) for the fast packet data transmission within an MPLS domain. However, the Ethernet over MPLS frame requires an additional header with a flow switch label (placed before the main frame header). Beyond, any conventional frame of Ethernet/MPLS solely contains a single IP-packet.

This fact is critical when transporting small IP-packets in real time applications and services (IP telephony, telemetry traffic in IoT sensor networks, etc.), since it significantly reduces the information efficiency of communication channels due to excessive header redundancy. The Pv6 packet redundancy with 40 octets header is even more sensitive.

Figure 2 shows structuring the overall data sequence into distinct command segments CS and data segments DS. For this, two 8-bit meta-commands are used: the command separator "11111111" (the "FF" byte in the hex-code); data separator "00000000" (byte "00" in hex-code). To recognize bytes "FF" or "00" in data segments, the byte-stuffing mechanism is used: two commands

("FF01" and "FF02") are reserved to replace the data characters "FF" and "00".

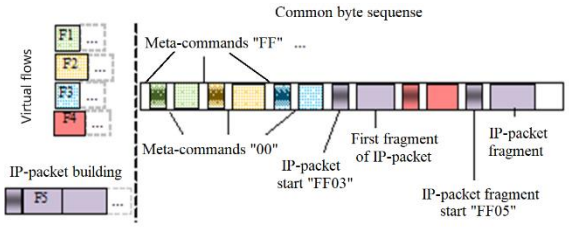


Figure 2: Structuring a data sequence by virtual channels of a frame Raw Socket Ethernet.

The packet fragmentation and defragmentation mechanism assume that any IP-packet of the incoming multimedia streams queue is divided into fragments of arbitrary size due to the three reserved 16-bit commands: packet start (byte "FF03" in the 16-bit code); packet end (byte "FF04"); packet fragment start (FF05). The packet fragment start command is removed after the packet arrives the input queue of the adjacent switch; the packet start/end commands are remaining. With these commands, the flow analyzer extracts individual packets from the queue.

3 COMPUTER SIMULATION THE INTEGRATED DATA STREAM TRANSMISSION

As mentioned above, the physical layer of the computer model for transmitting integrated multimedia streams is built since a special mode of operation of the standard Ethernet 802.3 technology (the so-called "raw sockets"). In this mode of operation, the frame structure, in particular its header, is formed in accordance with the requirements of application interface developer. To do this, each network adapter being interacting with any other one through the Raw Socket software interface must support the four software modules: SEND (transmitting Raw Socket frames to an Ethernet channel); RECEIVE (receiving/interpreting the frame); MUX (multiplexing the multimedia data flows); DEMUX (demultiplexing the multimedia data flows). The SEND and RECEIVE modules operate on the physical layer. Both modules can operate independently at one computer in two Linux terminals. The MUX/DEMUX operate on L2/OSI, All the 4 modules coded in Python/Linux [13, 14].

Figure 3 briefly shows the algorithm of SEND program module. The first block of the diagram imports the necessary modules (*time*, *socket*, *binaskii*) and creates the socket itself (*socket ()*).

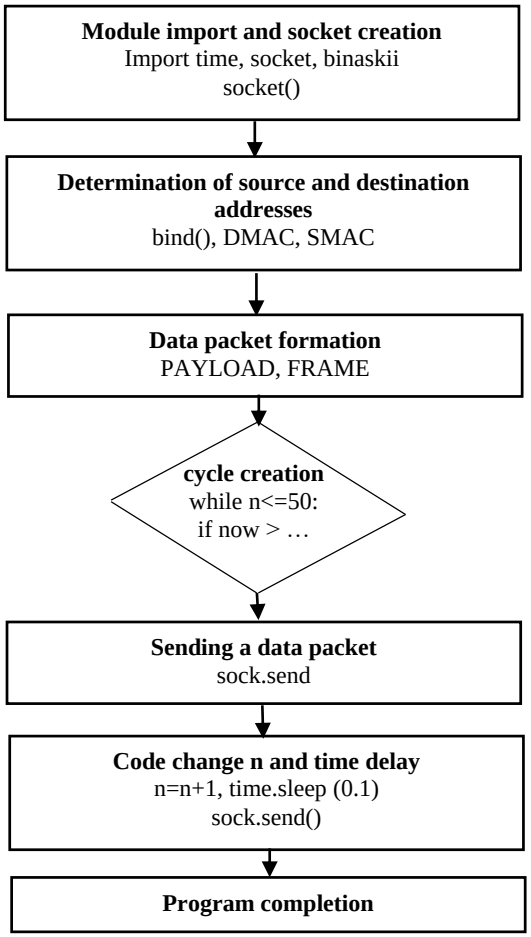


Figure 3: The algorithm of the SEND program module.

The second block of the algorithm determines frame transmission parameters (interface identifier, S/D MAC addresses). The third block forms the payload of the frame, as well as the structure of the entire frame. Next, a loop is created to repeatedly transmit the frame a certain number of times (for example, 50 times). This option is used exclusively at the stage of offline debugging and testing of the SEND/RECEIVE modules. For complete simulation the software model including the MUX/DEMUX modules, the SEND module performs a modified frame payload for each loop iteration.

The RECEIVE software module allocates the Raw Socket frame into memory buffer and extracts its Payload (which was generated on the transmitting side by the MUX module for multiplexing data streams). The payload content is further handled by the DEMUX module on the receiving side.

Briefly consider the key operators of the MUX software module. The first block of statements defines the parameters of the conveyor transporting modules (CTM) that is payload for the Raw Socket frame, as well as the initial value of the iteration variable *n*. Next, the operator

fin=open("/home/user/Desktop/InPack.txt",'r')

opens file *"/home/user/Desktop/InPack.txt"* for reading; this file contains a sequence of packets. The operator

fout=open("/home/user/Desktop/OutPack.txt",'w')

opens the file *"/home/user/Desktop/OutPack.txt"* for writing; this file contains CTM modules generated for output.

The operator *x=fin.read(1)* sets the read pointer in the file fin to the first position. The operator *while n<=6:* opens a loop of iterations over an iteration variable *n* from *n=1* (specified above) to 6; the iteration loop body is determined by the same offset of all first line loop operators.

The last statement in the loop is *n=n+1*. The *L=2* operator defines the initial length of the real-time data block *L=2*; this value is due to the fact that each CTM necessarily contains a two-character tag *"C0"* (end of real-time data block).

The operator *print '\n CTM='"%2d"%(n),* prints the current CTM value on the monitor. The script *\n* means jump to a new print line; the script *"%2d"%(n),* formats printing the module number *n* as an integer in two positions; comma at the end means that the next print statement will output the data at the same line.

The operator *fout.write('\n CTM='"%2d"%(n),)* prints CTM module number to the source file *OutPack.txt* with internal virtual name *"fout"*.

The Figure 4 shows the module DEMUX script for demultiplexing aggregated streams. The operator

f = open ("/home/user/Desktop/ OutPack.txt", 'r')

opens file *OutPack.txt* for reading; its internal is taken *"f"*.

The operator *x=f.read(1)* sets the file pointer of *"f"* to the first character of this file; the initial length of the character string *x* equals zero (i.e. the string of characters *x* is empty).

The operator *print ' '*, displays a space on the monitor screen (comma at the end means that the next print will be carried out on the same line through one space).

The operator *while len(x) > 0*: opens the main iteration loop until the size of the character string *x* is greater 0 (that is, until all non-empty characters in the open file *OutPack.txt*, denoted by *f* in the module, run out). The operator *x=f.read(1)* reads one character from the current position of the pointer (in this case, from the first position); the value of this character is assigned to variable *x*.

The operator *if (x == 'C'):* checks whether *x= "C"*; if so, then a following code executed:

```
f.seek(-1,1)
x = f.read(2)
if ((x == 'CA') or (x == 'C0')).
```

In this block, the operator rolls back the pointer by one position (i.e., sets the pointer to the first character of the file again). Next, operator *x = f.read(2)* reads two characters from this current pointer position. Finally, the operator *if ((x == 'CA')* or *(x == 'C0'))*: checks if these two characters are mark-up tags of type CA (beginning of packet) or C0 (end of real-time data block in CTM). If the "CA" or "C0" tag is indeed read, then this tag is printed on the monitor screen.

```
# Begin DEMUX
f=open("/home/user/Desktop/OutPack.txt", 'r')
x=f.read(1)          #set file pointer to the first symbol
print ' ',
while len(x)>0:
   #Do iteration cycle until len(f.read(1))>0,
   #e.g. reads    file to the end
   x=f.read(1)
   if (x=='C'):
      #print x,
      f.seek(-1,1)
      x=f.read(2)
      if ((x=='CA') or (x=='CO')):
         print x,
         for k in range (1,10):
         x=f.read(1)
         if x=='C':f.seek(-1,1); break
            print x,
```

Figure 4: Python script of the program module SEND.

Next, a built-in loop is introduced to read the following characters (which are the letters of the package): for *k* in range (1, 10): The increment *k* from *1* to *10* is taken arbitrary (consider the packet fragment size not exceed 10 characters).

The operator *if x == 'C': f.seek(-1,1);* checks whether the next character read is reserved letter *"C"*; if not, the next character is printed on the monitor screen (operator *print x,*); comma at the end means the following characters will be printed side by side on the same line. If the character *"C"* is read, then the file pointer is shifted for one position back, and the built-in loop for *k* in range (1, 10) ends by command *break*. Then the main loop continues (*while len(x) > 0:*). In this case, the operation of

reading characters from the file *f* is carried out with the current pointer (which gradually moves until the end of the file is reached); under these conditions, the main loop ends.

3 CONCLUSIONS

Novel principles proposed for digital streams management in autonomous telecommunication transporting systems at the OSI data link layer. This approach aims further improvement the information efficiency of network equipment and communication channels utilization, as well as to meet increased QoS requirements for real-time applications.

To achieve this, the original method of digital streams management in autonomous systems introduced on the base of multimedia data integration on Logical Link Control sublayer by virtualization the Raw Socket Ethernet channel.

The method provides scalable multiplexing and switching-mode IP-packets transportation via an autonomous system avoiding the interdomain routing processes. Computer model of integrated multimedia streaming presented on Linux Python. The results of the work intend for next generation networks application in the concept of the Internet of Things.

REFERENCES

[1] G. Bertrand, "The IP Multimedia Subsystem in Next Generation Networks," Gilles Bertrand, 2007. [Online]. Available: http://www.rennes.enst-bretagne.fr/ ~gbertran/files/ IMS_an_overview.pdf.

[2] V. Hol and M. Irkha, "Telekomunikatsiini ta informatsiini merezhi: navchalnyi posibnyk," Kyiv: ISZZI KPI im. Ihoria Sikorskoho, 2021, 250 p.

[3] M. Mousa, "Software Defined Networking concepts and challenges," 2016. [Online]. Available: https://www.researchgate.net/publication/312569297_Software_Defined_Networking_concepts_and_challenges.

[4] P. Vorobiienko, L. Nikitiuk, and P. Reznichenko, "Telekomunikatsiini ta informatsiini merezhi: Pidruchnyk [dlia vyshchykh navchalnykh zakladiv]," K.: SAMMIT-Knyha, 2010, 708 p.

[5] L. Kriuchkova, "Metodolohiia Upravlinnia infokomunikatsiŷnymy merezhamy zi zminnymy strukturamy v umovakh destruktyvnykh vplyviv," Dyssertatsyia na soyskanye uchenoi stepeny doktora tekhnycheskykh nauk po spetsyalnosty 05.12.02. Hosudarstvennyi unyversytet telekommunykatsyi, Kyiv, 2017.

[6] M. Klymash, O. Shpur, V. Bahrii, and A. Shvets, "Metod dyferentsiiovanoho multypotokovoho keruvannia trafikom u transportnykh prohramno-kerovanykh merezhakh," Natsionalnyi universytet Lvivska politekhnika. Radioelektronika ta telekomunikatsii, no. 796, 2014, pp. 60-68.

[7] O. Aouedi, K. Piamrat, and B. Parrein, "Intelligent Traffic Management in Next-Generation Networks," Future Internet 2022, 14(2), 44; [Online]. Available: https://doi.org/10.3390/fi14020044.

[8] M. Al-Jameel, T. Kanakis, S. Turner, A. Al-Sherbaz, and W. Bhaya, "A Reinforcement Learning-Based Routing for Real-Time Multimedia Traffic Transmission over Software-Defined Networking," Electronics, 2022. [Online]. Available: https://doi.org/10.3390/electronics11152441.

[9] M. Yanev, S. McQuistin, and C. Perkins, "Does TCP new congestion window validation improve HTTP adaptive streaming performance?" NOSSDAV'22, June 17, 2022, Athlone, Ireland, pp. 29–35. [Online]. Available: https://doi.org/10.1145/3534088.3534347.

[10] M. Seliuchenko, "Modeli ta alhorytmy pidvyshchennia yakosti obsluhovuvannia u telekomunikatsiinykh prohramno-konfihurovanykh merezhakh," Dysertatsiia na zdobuttia naukovoho stupenia kandydata tekhnichnykh nauk: 05.12.02 – telekomunikatsiini systemy ta merezhi; Lviv, 2016, 156 p.

[11] M. Dibrova, "Sposib bahatoshliakhovoi marshrutyzatsii v kompiuternykh merezhakh velykoi rozmirnosti," Avtoreferat dysertatsii, KhNURE, 2017.

[12] O. Lemeshko and K. Arus, "Model vidmovostiikoi marshrutyzatsii z realizatsiieiu riznykh skhem rezervuvannia resursiv merezhi v umovakh multypotokovoho trafiku," Materialy vseukrainskoi naukovo-praktychnoi konferentsii "Suchasni problemy Telekomunikatsii i pidhotovka fakhivtsiv v haluzi telekomunikatsii - 2014" SPTEL, 2014, Lviv: NU "Lvivska politekhnika", 2014, pp. 15-20.

[13] V. Tikhonov, O. Tykhonova, O. Tsyra, O. Yavorska, A. Taher, O. Kolyada, S. Kotova, O. Semenchenko, and E. Shapenko, "Modeling the conveyor-modular transfer of multimedia data in a sensor network of transport system," Eastern-European Journal of Enterprise Technologies, 2018, vol 2, no. 2 (98), pp. 6-14.

[14] V. Tikhonov, E. Tykhonova, O. Yavorskaia, V. Berezovskyi, A. Nehalchuk, and V. Hlushchenko, "Postroenye symuliatora dlia peredachy dannykh telemetryy cherez modyfytsyrovannyi ynterfeis Ethernet," Materialy 73 NTK profesorsko-vykladatskoho skladu, naukovtsiv, aspirantiv ta studentiv (Odesa, 12-14 December 2018), pp. 100-103.

Evaluation of LoRaWAN's Static and Dynamic Capabilities and its Limitations for IoT Applications

Simeon Trendov[1], Bohdan Zadoienko[1], Eduard Siemens[1], Dmitry Kachan[1], Marija Kalendar[2], Maksim Gering[1] and Sergii Maksymov[3]

[1]*Department of Electrical, Mechanical and Industrial Engineering, Anhalt University of Applied Sciences, Bernburger Str. 55, Köthen, Germany*

[2]*Faculty of Electrical Engineering and Information Technologies, "SS. Cyril and Methodius University" in Skopje, Rugjer Boshkovik Str. 18, Skopje, N. Macedonia*

[3]*Dexor Technology GmbH, Bernhard-Kellerman Str. 6K, Köthen, Germany*

{simeon.trendov, eduard.siemens, dmitry.kachan, maksim.gering}@hs-anhalt.de, bohdan.zadoienko@student.hs-anhalt.de, marijaka@feit.ukim.edu.mk, maksymov@dexor.de

Keywords: LoRaWAN, IoT, Wireless Communication, Network Performance, Network Performance Measurements.

Abstract: This paper is a study of connection parameters of Long Range Wide Area Networks (LoRaWAN) for reliable and robust data communication. It also seeks to highlight the limitations and shortcomings of the LoRaWAN technology, with the goal of identifying areas for improvement. The primary goal is to investigate how particular network parameters affect communication and message transmission. The experimental setup consists of a gateway and two end nodes. The maximum range between the end device and the gateway was tested in a free field in which a stable communication is possible. Furthermore, the connection between the gateway and the endpoints was evaluated at various movement speeds. Additionally examined in a laboratory environment were the roundtrip time, the uplink, delay and downlink, and the received signal strength indicator according to the used transmission power, packet size and the spreading factor. The maximum packet sizes for each spreading factor were also tested.

1 INTRODUCTION

One of the currently widespread Low Power Wide Area Network (LPWAN) technologies is LoRaWAN (Long Range Wide Area Network). A LoRaWAN-enabled end device is a sensor or actuator that is wirelessly connected to a LoRaWAN infrastructure network using radio gateways [1]. All of the data gathered by the end devices is transmitted through the low power LoRaWAN network to a listening gateway, where it is then sent to the network server and the application server [2]. The aim of this investigation was to explore the possibilities of expanding possible areas for the use of wireless communication channels in industrial environments based on the LoRaWAN technology. To achieve this goal, the Round-Trip Time (RTT) was measured, which is the sum of the time it takes for a data packet to be delivered or uplink time and for its acknowledgment to be received or downlink time together with the delay [3] in dependence of various transmission parameters used for data transmission. he achievable range, which is the distance between gateway and the end device, as well as the Received Signal Strength Indicator (RSSI), which is the relative quality of a received signal to a client device [4] were also measured.

2 STATE OF THE ART

There are several areas in which LoRaWAN technology can be used, such as: metre reading, street lighting, smart buildings, smart parking, cargo tracking, water leakage detection, water level monitoring, alerting about the occurrence of emergency situations, traffic management, smart energy systems, waste management, smoke detection, etc.

Modern LoRaWAN technology has experienced rapid development in recent years. Numerous studies have been conducted to investigate the

characteristics and performance of LoRaWAN technology [5, 6]. In this article, the operation range and packet loss rate have been tested by the authors using experimental measurements, but they have not investigated how parameter settings affect network performance of LoRaWAN.

Without changing complex network setups, the study in [7] examines the impact of modulation parameters on the connection between the end device and its gateway. The authors in [8] use a stochastic geometry model to obtain more trustworthy results. This allows the simultaneous study of time interference and frequency domain. It is found that the packet replay approach reduces the probability of failure while increasing redundancy, resulting in a decrease in average throughput.

The best combination of SFs to reduce the probability of frame overlap was determined in [9]. In addition, they have set up a plan to increase uniformity for endpoints that are far from the gateway by assigning the ideal SF value and increasing the transmission power across the endpoints to reduce packet error rates.

3 EXPERIMENTAL SETUP AND MEASUREMENT RESULTS

Two end devices and a gateway have been used in the experiments conducted in this research. The uplink, downlink, and round-trip duration were all evaluated using a LoRa Shield end-device as given in section 3.1. The end-device and the gateway were 10 metres apart during the laboratory testing for these measurements. The second group of measurements have been done in a free field with the second end-device and the gateway positioned also in a free field. So it shall be possible to see how the technology operates in a real-world environment and identify the dependencies between the parameters through testing in the laboratory. Measurements in the field made it possible to estimate the maximum distance at which a stable message exchange can be operated and how the distance between the gateway and the end device affects the transmission time.

3.1 Measurements of the Uplink, Downlink and Round Trip Time

For measurements in the laboratory, the MultiTech gateway [10] and an end-device based on Arduino UNO [11] and LoRa Shield [12] have been used. A host with the Ubuntu 20.04 operating system on which The Things Stack [13] was installed, was connected to the gateway. The end-device was connected to another laptop on which it was possible to see and collect information about uplink, downlink and RTT times using the Arduino IDE terminal. The settings of the end-device, namely the *size* of the transmitted packet, *transmission power* and *spreading factor* (SF) have been changed and performance parameters with different sets of those parameters have been measured.

3.1.1 Impact of the Transmission Power

In this section, the test results for round-trip time, and uplink/downlink transmission time changes, based on variations in transmission power are described. Four measurement sets have been performed with different transmission power on the LoRa Shield end-device. In all four cases, the end device sent 13-byte messages present an empty message with LoRaWAN service information attached [14]. In all cases, the device was located in the laboratory and the environmental conditions did not change. For these tests, only a spreading factor of 7 and a bandwidth of 125 kHz were used. The impact of power variation on the transmission time was investigated using transmit powers of 5, 10, and 14 dBm.

As expected, the performance parameter differences in this group of measurements in the laboratory are not noticeable, but during the measurement one message was missed and the spreading factor value was automatically changed to 8. Before switching to SF8, the message skips the first downlink window and then, after one second delay, the second downlink window. Figure 1 shows a spike in the measured time. This is due to the fact that the message transmission time for SF8 is longer.

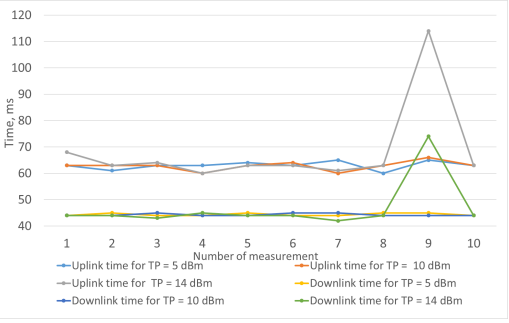


Figure 1: The relationship between uplink/downlink time and transmission power.

Also during the RTT measurement there were small changes, the graph behaves similarly to the uplink and downlink time graph at Figure 2. The message was sent again using SF8, adding extra time. The RTT measurements and the time jump is shown in Figure 2. During the tests in the laboratory, it was concluded that the transmission power does not affect the uplink, downlink time and RTT parameters.

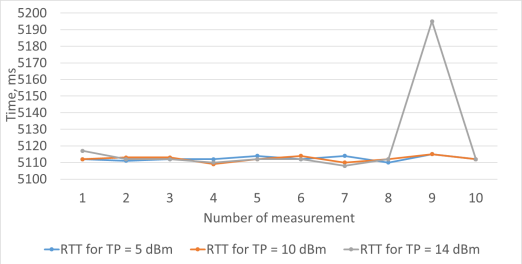


Figure 2: The relationship between RTT and transmission power.

3.1.2 Variation of Packet Size

In this part of the measurements, the *message size* has varied. The tests have been performed for three different message sizes from minimal to maximal possible sizes:- 13, 25 and 50 bytes which work with all spreading factors. These values were chosen to reveal the dependence of round-trip time on message size. Only a spreading factor of 7 and a bandwidth of 125 kHz were employed for these tests. The maximum permitted transmission power in the European Union, 14 dBm, was selected [15]. Increasing the packet size affected the message transmission time as expected. Values such as uplink time and RTT have increased. Figure 3 shows the dependence of uplink time on the packet size. It can be concluded that the time will increase in direct proportion to the packet size.

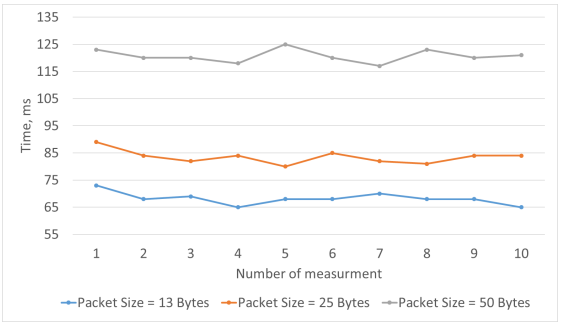


Figure 3: The relationship between uplink time and packet size.

The downlink time is unaffected by increasing the packet size since the acknowledgment packet's size is constant - 14 Bytes and independent of the transmitted packet size. downlink time ranges from 44 ms to 47 ms.

Although the difference in round-trip times between the three groups is small, it is noticeable, and as the size of the packet increases, the time gap tends to grow. A graphic in Figure 4 illustrates the measured data for better visualisation. Depending on the message that has to be transmitted to the gateway, the end node's data packet size may change when sending data. The results indicate that changes in packet sizes have no impact on message interference and collisions since there are relatively few messages.

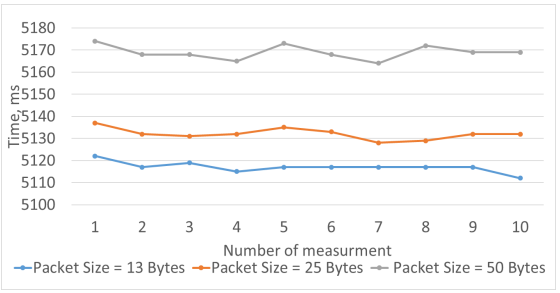


Figure 4: The relationship between RTT and packet size.

3.1.3 Variation of Spreading Factor

In this phase, there are six groups of measurements in which only the spreading factor is changed and all of the other parameters remain the same.

There is no doubt about the discrepancies between the six categories of measurements. The higher the spreading factor, the more time it takes for the message to reach the gateway and for an acknowledgment to be received by the end node. The uplink/downlink time measurements for the first four groups for spread factors 7, 8, 9 and 10 are closer to each other, compared to the measurements done on the last two groups for spreading factors 11 and 12. The last two groups of measurements show a significant increase in uplink/downlink time. This is because each spreading factor is associated with bits per second. For better visualisation the results are illustrated in Figure 5.

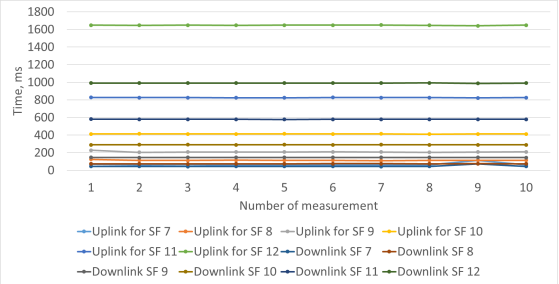


Figure 5: The relationship between uplink/downlink time and spreading factor.

For RTT the situation is similar to the uplink time measurements. The RTT significantly increases for the last two groups of measurements. The results are presented in Figure 6. The spreading factor affects the time of message transmission. By changing the spreading factor the transmission time of one symbol changes. The spreading factor affects the range of signal transmission. More details about this are described in paragraph 3.2.2.

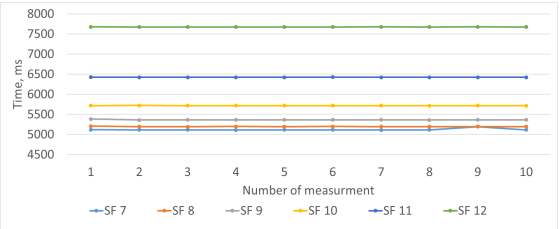


Figure 6: The relationship between round-trip time and spreading factor.

3.1.4 Variation of Spreading Factor with Respective Maximum Packet Size

The last group of laboratory tests were conducted to elaborate how other characteristics would change when the maximum message size is transmitted with different spreading factors. In these tests, the message size was maximally increased. Tests were performed for six different message sizes for all spreading factors. For spreading factors 7 and 8 a maximum message size of 248 bytes can be sent, which is 13 bytes more compared to the theoretical value of 235 bytes. But for SF 9 till 12 it was only possible to transmit a message less than the theoretical maximum value [16]. Maximum message sizes for spreading factor 9 till 12 of 113, 60, 57, 57 bytes can be sent, which is less compared to the theoretical values of 128, 64, 64, 64 bytes for each spreading factor respectively. For these tests, a bandwidth of 125 kHz and a transmission power 14 dBm were used since this is the maximum permitted power in that band for Europe. The maximum allowed transmission power for LoRaWAN for Europe is 14 dBm.

It is easy to see how the six measurement groups differ from each other. The larger the SF and the packet size are, the longer it takes for the message to reach the gateway and for the end-node to receive an acknowledgment. It can also be noticed that the transmission time at SF 8 and SF 10 look the same due to the fact that the message size was decreased. The results are shown in Figure 7.

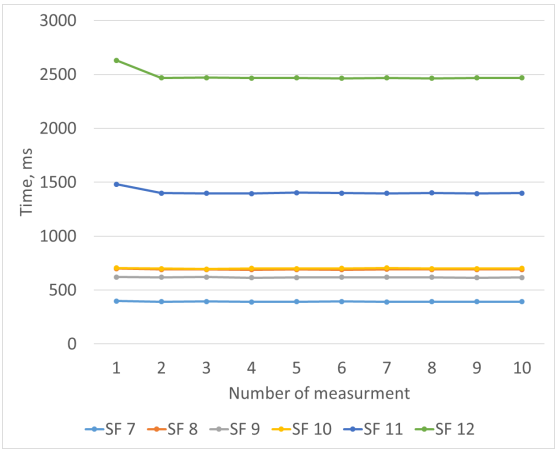


Figure 7: The relationship between uplink time and maximum packet size.

Because of the corresponding bits per second, the time needed for the last two groups increases drastically. A graph in Figure 8 illustrates the data.

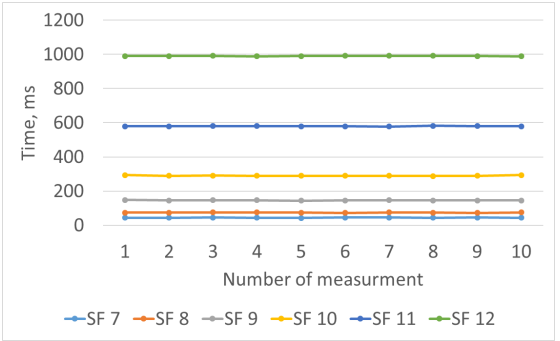


Figure 8: The relationship between downlink time and maximum packet size.

For RTT, the measurements are similar to the uplink time measurements. The difference between the transmission times at SF 8 and SF 9 is essentially undetectable. This is because when the

message size and SF were dropped, the uplink time decreased but the downlink time increased, practically causing the RTT to remain unchanged. The results are presented in Figure 9.

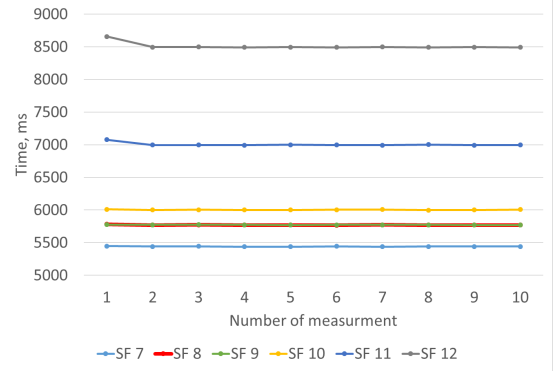


Figure 9: The relationship between round-trip time and maximum packet size.

3.2 Static Measurements of the Range in Free Field

Figure 10 shows the experimental setup used in this cycle of measurements. The gateway was placed on top of a 2.2 metre stick in order to maximise the range. The LoRa/GPS Shield was driven in a vehicle.

Figure 10: Free field device setup.

The end-device was plugged into a power bank and was being driven in a vehicle. A couple of car stops were chosen for all of the measurements. In each measurement series, the car stopped at nearly the same positions so the end-device can try to establish a connection with the gateway. Every 40 seconds, the end-device was sending LoRa signals that contained the end-device's most recent GPS coordinates. When the transmission power and spreading factor were changed, the behaviour of the range was observed. Only one parameter was altered at a time in each of these free field experiments, and the link between that parameter and the range may be seen below.

The gateway had two positions, chosen because of their higher ground compared to the surrounding area. The first position was on a small hill with free fields in front of it. The second position was on a higher hill compared to the first one. The only disadvantage of the second position was a couple of trees closely located to the gateway which were causing some link budget damage. There were a couple of small villages between the gateway and the end-device in some positions of the measurements, but because of their lower altitude they were not decreasing the link budget. The topology is explained in detail below, and can be seen in Figure 13 and Figure 16.

The range evaluation highly depends on the propagation environment. According to a LoRaWAN free field range calculator [17], if the gateway is placed on a position with a height of 30 metres and the end-device has a height of 1 metre, a range of 11100 metres should be possible for establishing communication between the gateway and the end-device.

3.2.1 Variation of the Transmission Power for the Static Measurements

The testing for the range's behaviour in relation to transmission power is covered in this section. Four measurements were conducted utilising four different transmission powers for the LoRa/GPS shield end device. In each of the four instances, the end-device was communicating with the gateway using 32-byte messages that contained the end-device's GPS coordinates. Since the device was always moved in the same direction, the environment wasn't radically changed. Only a spreading factor of 7 and a bandwidth of 125 kHz were employed for these tests.

The differences in the distances are clearly noticeable. Longer distances can be covered

between the end device and the gateway for data transmission as the transmission power is increased. Figure 11 displays the greatest distances that were achieved. When comparing the urban environment results from the paper [18] and the measurements done in the free field, shown in Figure 8, a clear increase in the range can be noticed. This is due to the number of objects in the Fresnel zone. A maximum range of 6700 metres between the end device and the gateway was achieved in which data transfer is possible.

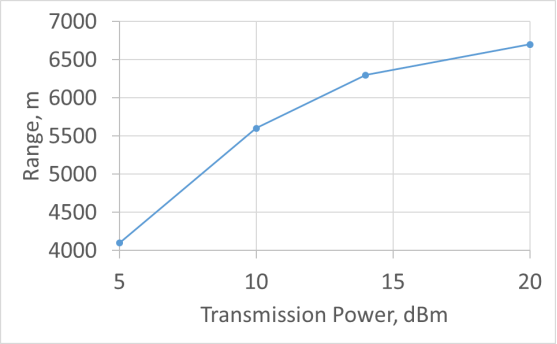


Figure 11: The relationship between the range and the transmission power in a free field.

In all these measurements the behaviour of the range according to the transmission power, the RSSI levels were also measured and noted. The chart in Figure 12 clearly shows the differences in the RSSI levels when the transmission power is changed. On each chosen car stop, the gateway receives a better signal when using a greater transmission power.

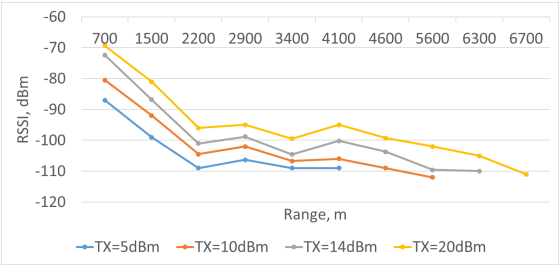


Figure 12: The RSSI levels when measuring the range with changing the transmission power in a free field.

The biggest distances reached with each tested transmission power are shown on the map in Figure 13. The gateway is represented with a red dot and the four biggest distances achieved with each transmission power are represented with blue dots. The gateway is located at an altitude of 94 metres. Additionally, it was placed on top of a 2.2 metres stick. This is beneficial for establishing better communication. The blue dots TX5, TX10, TX14 and TX20 which represent the last seen position of the gateway when using a transmission power of 5, 10, 14 and 20 dBm have an altitude of 75, 71, 73 and 70 respectively. The terrain between the gateway and TX5 has a low altitude dropping to 73 metres. Also, the terrain between the gateway and TX10 has a low altitude dropping to 61 metres. This is good for stable communication between the devices. The field between the gateway and TX14 has a lower altitude than the points at which the devices are loaded, dropping to 61 metres. In the middle between TX20 and the gateway there is a higher ground with an altitude of 73 metres compared to the position of TX20. This should not be a problem because of the higher altitude at which the gateway is located.

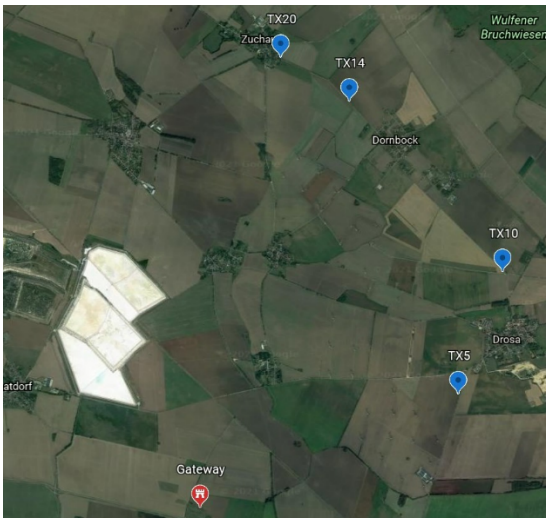


Figure 13: Map of static free field measurement of the range when changing the transmission power.

3.2.2 Variation of the Spreading Factor for the Static Measurements

Using spreading factors of 7 to 12, which correspond to data rates of 5469, 3125, 1758, 977, 537, and 293 bps, respectively, the behaviour of the range is investigated in this cycle of measurements. Only a transmission power of 14 dBm and a bandwidth of 125 kHz were employed for these experiments.

Figure 14 displays the biggest distances that were achieved on different spreading factors at which data transmission can be established between the end-device and the gateway. Data can be sent

across bigger distances with raised spreading factors. So, in this cycle of measurements, a maximum range of 10950 metres was achieved with the spreading factor 12. When comparing the urban environment results from the paper [18] and the measurements done in the free field, shown in Figure 10, a clear increase in the range can be noticed.

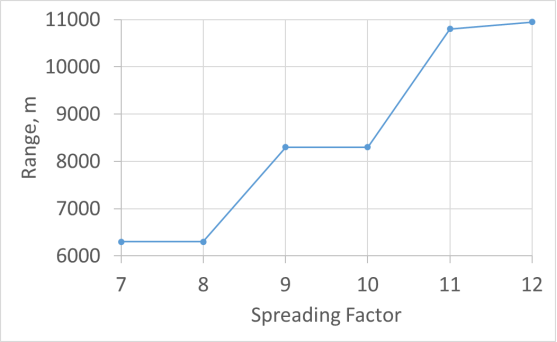


Figure 14: The relationship between the range and the spreading factor in a free field.

In all of these measurements of the behaviour of the range when varying the spreading factor, the RSSI levels of the received signals were also measured and noted. The chart in Figure 15 shows the RSSI level of each received signal when measuring the range. A clear decrease of the RSSI levels can be observed in the tendency with each increasing of the distance. However, there are a few points where RSSI is rising even on raised distance. This is because of the better line of sight in that particular position where the end device was located when transmitting the message.

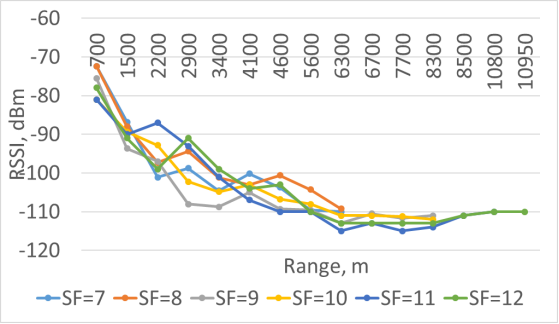


Figure 15: The RSSI levels when measuring the range with changing the spreading factor in a free field.

The biggest distances reached with each spreading factor are shown by the map on Figure 16. The first position of the gateway is again represented with a red dot, the second position with a yellow dot and all of the biggest achieved distances for each spreading factor is represented with a blue dot. The biggest ranges for the spreading factors 7, 8, 9 and 10 were measured when the gateway was located in the first position. Then the position of the gateway was changed and it was placed in the second position represented by the yellow dot. The biggest ranges for the spreading factors 11 and 12 were measured when the gateway was located in the second position.

The first position of the gateway has an altitude of 94 metres, and the second an altitude of 104 metres. Additionally to this, a 2.2 metres stick was used for the gateway. The points were specially selected because of their higher ground compared to the surrounding area. Dots SF7 and SF8 represent the last position of the end device when using a spreading factor of 7 and 8 have an altitude of 73 metres. The dots SF9, SF10, SF11 and SF12 representing the last seen position of the end device when using a spreading factor of 9, 10, 11 and 12 have an altitude of 69 metres. The altitude of the terrain between the dots SF7, SF8 and the first position of the gateway is lower than the altitude of their positions, with a minimum of 61 metres. There are two small villages in between SF9, SF10 and the first position of the gateway. The first village has an altitude of 63 metres and the second an altitude of 61 metres. The end device at dots SF9 and SF10 had an altitude of 63 metres. This can cause some link budget loss, but because of the height at which the gateway was located there with a line of sight. The altitude of the field between SF11 and the second position of the gateway reaches a maximum of 75 metres and the signal passes through one small village. The village has an altitude of 61 metres. The messages reach the gateway because of its higher position. The biggest accomplished range with this measurement was with spreading factor 12 and transmission power 14 dBm. The device was last seen at the position SF12. The altitude of the terrain between the second position of the gateway and SF12 reaches a maximum of 77 metres.

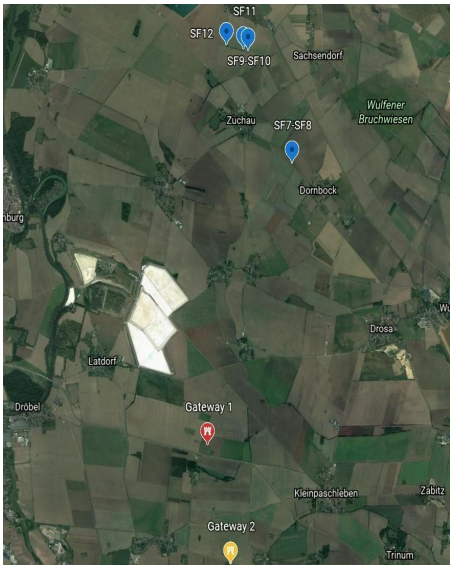


Figure 16: Map of static free field measurements of the range when changing the spreading factor.

3.3 Dynamic Measurements in Free Field

For the dynamic free field measurements, the LoRa/GPS shield end device was programmed to send messages every 20 seconds to the gateway. A 2.5 km straight road was chosen, so the car can reach the desired speed. The end device was located in a car, driven with speeds of 20, 50, 100 km/h respectively. The multitech gateway was placed at a height of 2.2 metre, in the middle of the road. The road together with the placement of the devices can be seen in Figure 17.

Figure 17. Map of the dynamic free field measurements.

3.3.1 Dynamic Measurements of the RSSI at Different Speeds

The RSSI behaviour was observed throughout this test cycle at selected speeds. The end device was communicating with the gateway in each case with 13 Bytes of data. Only a 14 dBm transmission power, a spreading factor of 7, and a 125 kHz bandwidth were used for these testing.

As shown in Figure 17 the car was following the road depicted by the yellow line. The gateway was placed in the middle of the route described by the red pointer. The end device was turned on for each group of measurements at the beginning of the road and switched off at the end.

Depending on the speed of the car, the end device reached the closest point to the gateway at a different moment. This is depicted with the peak of the three groups of measurements, shown in Figure 18. It can be clearly seen how the car started at a position that is further away from the gateway, came close to it and again moved away.

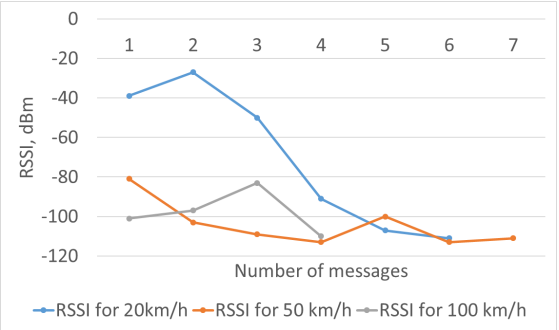


Figure 18. Dynamic measurements of the RSSI at different speeds.

3.3.2 Variation of the Transmission Power for the Dynamic Measurements

The tests for the behaviour of the RSSI according to the transmission power of 5, 10, and 14 dBm are covered in this section. Three measurements were made using three different transmission powers for the LoRa shield end device. In each of the three instances, the end device was communicating with the gateway using 13 Bytes of data. Only a spreading factor of 7 and a bandwidth of 125 kHz were employed for these testing. This cycle of measurements was only done with a constant speed of 100 km/h.

Figure 19 shows the RSSI levels while varying the transmission power. Because the starting point was the same for the three groups of measurements it can be clearly seen that the first message has a lower RSSI level compared to the other two. This is because of the lower used transmission power. When comparing the first messages for the other

two groups, it can be seen that their RSSI levels are very close to each other. This can be caused by a change in the fresnel zone in that particular moment, compared to the other two groups, or by a slight change at the starting position of the end device.

Three peaks can be again spotted, depicting the message sent when the end device was at the closest position compared to the other messages in the same group. It can be seen that the end device in the first group of measurements, represented with the blue line in Figure 19, disconnects after reaching the peak. This is caused by the small transmission power of 5 dBm.

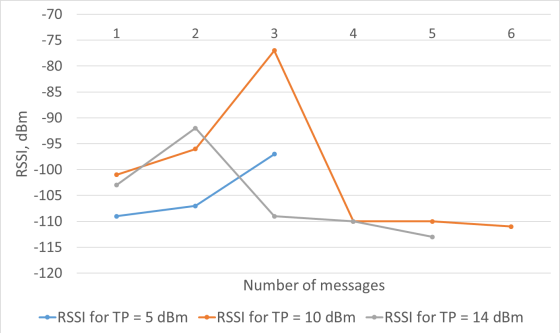


Figure 19. Dynamic measurements of the RSSI while varying the transmission power.

4 CONCLUSION

The advancements in the *IoT* field impose higher requirements for technologies that can provide long range and consume less energy. *LoRaWAN* is a good alternative to the technology of sensors and actuators that need to send and receive their data over long distances and don't require a lot of power.

The main goal of the paper is to evaluate the static and dynamic capabilities of LoRaWAN. The limitations of this technology were also tested in order to determine the best use cases for future IoT applications. The behaviour of the delay, uplink, downlink and round-trip time were tested according to the used transmission power, packet size and spreading factor. The maximum packet sizes were also tested for each spreading factor. This paper seeks to highlight the limitations and shortcomings of the LoRaWAN technology, with the goal of identifying areas for improvement.

With the help of the measurements a future user of the LoRaWAN technology can understand the influence of each parameter when changed, and its impact on the occupation of the channel and the time on air when creating a LoRaWAN end device for IoT applications. This knowledge should be implemented when following the strict limitations of the fair use policy [19].

The fair use policy limits the capabilities and use cases of LoRaWAN. For an actuator working as a Class C end device, the limitation of 10 downlink messages is strict. The uplink message policy is also very limited. In order to increase the applicability of LoRaWAN, the fair use policy should be modified. For limiting the data collision another method should be implemented. The devices should be also able to transmit the same message a couple of times, shortly separated known as time diversity and on different frequencies or frequency diversity. A third method, space diversity, or using a couple of antennas at a distance from each other can be also applied in order to ensure that the message reaches the gateway. In order to save the battery life of the end device, the space diversity should be implemented on the gateway.

LoRaWAN technology has a lot of potential applications where data needs to be collected. According to the results of the urban field tests conducted in [18], data transfer is feasible across short distances when necessary due to the presence of a large number of end devices. This will prevent the devices from interfering with one another and reduce the likelihood of data collision. The range can be restricted by lowering the transmission power and the spreading factor, which makes it ideal for applications in smart factories. The data collision can be also decreased by increasing the used unlicensed frequency bands as done recently in Europe.

The paper identifies the strengths and weaknesses of LoRaWAN. The study also highlights opportunities including the increase in battery life. By closely controlling the transmission power and the spreading factor, the life of the battery is extended. This is very useful for end devices located in remote places, which can not be often recharged. The battery life can be prolonged for a couple of years if the parameters are set right.

Data transfer is also possible on very big ranges as seen above in the free field measurements. The numbers can be controlled depending on the use case. As seen in the free field measurements in part 3.2.2, a maximum range of 10950 metres was reached with a spreading factor 12 and a transmission power of 14 dBm. This result is perfect for controlling sensors and actuators on big fields in smart farming or controlling a long street with smart lighting with LoRaWAN.

The stability of the communication was also tested at different moving rates. Stable communication is possible at 20, 50 and 100 km/h. Connection problems were encountered while testing the data transmission with a power of 5 dBm with a moving rate of 100 km/h. In this cycle of the measurements, the end device was able to establish communication but it was quickly disconnected. With a higher transmission power, there were no connection problems. The device was able to establish communication while the vehicle was moving at the tested speeds. The results of the dynamic measurements point out the opportunity for this technology to be used for predictive maintenance for industrial vehicles. With the increase of the amount of public gateways, this technology can also be used for tracking vehicles and goods as they travel to the end user.

There are many possibilities for future uses of the LoRaWAN technology that still need to be explored. Despite being quite young, this technology has a lot of potential. When discussing the wireless connection of sensors and actuators, it can become irreplaceable as the number of public gateways increases.

REFERENCES

[1] P. Lea, "Internet of Things for Architects," United Kingdom: Packt Publishing, 2018.

[2] B. Wiegmann, "IoT Networks with LoRaWAN," Leanpub British Columbia, Canada, 2019 [Online]. Available: https://leanpub.com/iot-networks-with-lorawan. [Accessed 15 Oct 2023].

[3] What is RTT (Round-Trip Time) and How to Reduce it. [Online]. Available: https://www.stormit.cloud/blog/what-is-round-trip-time-rtt-meaning-calculation. [Accessed 12 Dec 2023].

[4] Understanding RSSI Levels. [Online]. Available: https://www.metageek.com/training/resources/understanding-rssi/. [Accessed: 12 Dec 2023].

[5] J. Petajajarvi,, K. Mikhaylov, A. Roivainen, T. Hanninen, and M. Pettissalo, "On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology," 14th International Conference on ITS Telecommunications (ITST), pp. 55-59, 2015.

[6] A. J. Wixted, P. Kinnaird, H. Larijani, A. Tait, A. Ahmadinia, and N. Strachan, "Evaluation of LoRa and LoRaWAN for wireless sensor networks," IEEE Sensors, pp 1-3, 2016.

[7] Z. Li, S. Zozor, J.-M. Brossier, N. Varsier, andLampin, "2D Time Frequency interference modelling using stochastic geometry for performance evaluation in Low-Power Wide-Area Networks," IEEE International Conference on Communications (ICC), pp 1-7, 2017.

[8] G. Ferre, "Collision and packet loss analysis in a LoRaWAN network," 25th European Signal Processing Conference (EUSIPCO), pp. 2655-2659, 2017.

[9] B. Reynders, W. Meert, and S. Pollins, "Power and spreading factor control in low power wide area networks," IEEE International Conference on Communications, pp. 1-5, 2017.

[10] MultiTech Conduit AP, Access Point for LoRa Technology (MTCAP Series). [Online]. Available: https://www.multitech.com/documents/publications/data-sheets/86002212.pdf. [Accessed: 1 Dec 2023].

[11] Arduino UNO R3. [Online]. Available: https://docs.arduino.cc/static/85a18c19e252907a5fb428e7266bc70f/A000066-datasheet.pdf. [Accessed: 15 Dec 2023].

[12] Lora Shield - Wiki for Dragino Project. [Online]. Available: http://wiki1.dragino.com/index.php?title=Lora_Shield. [Accessed: 22 Oct 2023].

[13] What Is The Thing Stack? The Thing Stack For LoRaWAN. [Online]. Available: https://www.thethingsindustries.com/docs/getting-started/what-is-tts/. [Accessed: 15 Dec 2023].

[14] LoRa payload; AllThingsTalk Docs. [Online]. Available: https://docs.allthingstalk.com/tutorials/lora-payload/. [Accessed: 11 Dec 2023].

[15] LoRa - LoRa Documentation. [Online]. Available: https://lora.readthedocs.io/en/latest/. [Accessed: 8 Dec 2023].

[16] The Things Network; LoRaWAN airtime calculator", [Online]. Available: https://www.thethingsnetwork.org/airtime-calculator. [Accessed: 28 Oct 2023].

[17] LoraWAN Range Calculator. [Online]. Available: https://www.rfwireless-world.com/calculators/LoRaWAN-Range-calculator.html. [Accessed: 11 Dec 2023].

[18] S. Trendov, E. Siemens, and M. Kalendar, "Development and deployment of a LoRaWAN performance test setup for IoT applications", XV International Conference ETAI, pp. 312-317, 2021.

[19] LoraWAN Network Affiliate NetID Policy and Terms. [Online]. Available: https://lora-alliance.org/wp-content/uploads/2021/12/FINAL-LoRaWAN-Network-Affiliate-Policy-and-Terms-nonmembers-10-20-21.pdf. [Accessed: 1 Dec 2023].

[20] RSSI and SNR; The Things network. [Online]. Available: https://www.thethingsnetwork.org/docs/lorawan/rssi-and-snr/. [Accessed: 17 Oct 2023].

Dynamic Reconfiguration of Computing Resources to Support NaaS Technology

Larysa Globa, Svitlana Sulima, Oleksandr Romanov and Mariia Skulysh
Institute of Telecommunication Systems, Igor Sikorsky Kyiv Polytechnic Institute, Industrialna Lane 2, Kyiv, Ukraine
lgloba@its.kpi.ua, itssulima@gmail.com, a_i_romanov@ukr.net, mskulysh@gmail.com

Keywords: Virtualization, Network Function Virtualization (NFV), Machine Learning (ML), Network as a Service (NaaS), Traffic Management System, Resource Configuration, Network Reconfiguration.

Abstract: Future communication systems and networks, namely smart networks, are becoming the foundation of the human-centric Internet. They provide an energy-efficient and high-performance infrastructure where NGI (Next Generation Internet) and other digital services can be developed and deployed. The development of new services provided to the end user led to the technology of providing a network as a service NaaS (Network as a Service) implementation, which should be an intelligent, decentralized and highly automated network. The architecture of the intelligent network is already software-defined and provides functions that highly exceed the connectivity capabilities: multi-service and mobile edge computing will allow data to be stored and processed locally at the edges of the network to guarantee quick response and efficient use of network resources. But assignment the required amount of computing resources and determining their location remains an incompletely resolved issue. This paper proposes a method of dynamic reconfiguration of computing resources to support NaaS technology. The method allows you to manage the placement and determination of the required amount of network functions resources to optimize the amount of computing resources allocated to the network function in the telecom operator infrastructure, and also binds the performance of the network virtualized function with the amount of resources assigned to it, takes into account the hybridity of the communications environment.

1 INTRODUCTION

Network as a Service (NaaS) enables tenants and cloud users to connect their distributed services across multiple clouds without relying on their own networking services and resources [1]. Today, CSPs implementing NaaS are focused on increasing efficiency and flexibility in business and network operations while improving the service experience for business customers. The main goals are to reduce the complexity of the network and support system, to give customers more control over their services.

NaaS is a framework for building network services that decouples the customer's service request, along with the operational and business support systems (OSS/ BSS) used to create and manage the service, from the network resources required to provide that service. The decoupling occurs through an abstraction layer that hides the details and complexity of the network, which is central to simplifying the entire service lifecycle.

So, if a customer requests a service with a defined bandwidth with certain characteristics between specific points, the CSP provider's OSS/BSS system, as well as network systems and their elements, autonomously figure out how to perform the service, and how much network and computing resources are required to operate services.

To cope with a significant increase in signaling traffic, mobile operator is developing network virtualization and cloud computing technologies to build scalable and flexible mobile networks and offer them as cloud service.

Our research falls into the category of advanced placement algorithms, other placement algorithms range from ILP formulations, such as those described in [2], to evolutionary algorithms, like [3], and heuristic methods.

In the design of virtualized networks, the trade-off between the cost of resource deployment and the effective provisioning of services must be considered [4].

Current descriptions of NaaS services lack determining the location and capacity of reserved computing resources of virtual network functions, taking into account the state of computing and network resources, requirements to the Quality of Service for signaling flows and heterogeneity of the network. Due to the dynamic provision of resources proposed in the paper method can reduce the amount of resources used, implementing calculation of topology and capacity of resources at the same time.

2 PEOPLE-ORIENTED SERVICES

Telecommunication networks are critical and strategic infrastructures that must be protected with appropriate tools. TM Forum considers these main trends that significantly affect the development of modern technologies in various industries and citizens' lives, in particular:

1) Robotic automation, where advances in robotics and artificial intelligence accelerate the introduction of interconnected and autonomous machines in many sectors of human life [5–6].
2) Mass monitoring and remote control, which is necessary for life support processes to optimize productivity and increase their efficiency in the economy [7–8].
3) Autonomous and hyper-connected urban transport on demand [9].
4) Industrial IoT (IIoT) with cloud computing technologies that form industrial networks that must provide productivity, secure and reliable communication in real time in the enterprise where these technologies are implemented.
5) Tactile IoT, which will make human life more comfortable.

All these services can be provided as NaaS layers within one physical network using NFV technology. The challenge is to place these virtual layers on the network effectively.

2.1 Requirements for Parameters of Network Operation

Table 1 shows five abstract dimensions for evaluating the effectiveness of the functioning of services with the disclosure of a set of parameters that characterize the corresponding requirements for the network, namely: bandwidth, time, security, the level of application of artificial intelligence (AI) and the application of ManyNets technology (technology of inter-network interaction).

Table 1: Abstract dimensions with corresponding network requirements [10].

Abstract dimensions	Appropriate network requirements
Capacity	Bandwidth, QoE, QoS, flexibility
Time	Latency, timing, jitter, accuracy, scheduling, coordination, geolocation accuracy
Security	Security, confidentiality, reliability, sustainability, traceability (in the sense of monitoring) of authorized access, lawful interception
The level of application of artificial intelligence	Edge computing, storage, modeling, collection and analytics for network configuration and management
The level of application of inter-network interaction technology (ManyNets)	Addressing, Mobility, Network Interface, Multiple Radio Access Technologies, Heterogeneous Network and Computing Convergence

In the coming years, there is a clear trend towards the maximum increase in the level of automation in all processes and interactions (except for cases that may negatively affect the processing and interaction process).

Future networks will increase the level of requirements for real-time management of network services and the level of automation, which will be crucial for improving the economic efficiency of the work performed.

The emergence of the concept of Network Functions Virtualization (NFV) has opened up new opportunities for telecommunication systems, at the same time, new questions arise, that is, the need for new approaches, models and methods of organizing the process of servicing telecommunication flows in such systems.

The general idea behind this technology is that network functions are implemented as software entities that can run on a standard hardware platform. According to the ETSI specification, a network function is a functional unit within the network infrastructure that has well-defined external interfaces and well-defined functional behavior. Network functions in the LTE EPC network are, for example, MME, HSS, PGW, SGW, which for the NFV case will be deployed on server resources (CPU, memory, NIC).

Service delivery in the telecommunications industry has traditionally been based on network operators deploying static physical equipment that is located at fixed points and has fixed service characteristics.

Thus, the task of effective distribution of computing resources (CPU, storage, bandwidth) of the basic network based on NFV virtualization technology is relevant in order to increase the utilization rate of computing resources during the operation of the core network (LTE EPC) to guarantee the appropriate level of quality of service provision due to the dynamic redistribution of service flows in the network and reconfiguration computing resources of a service.

2.2 Provision of Services by a CSP-Provider in a Multi-Layered NaaS Architecture

Modern NaaS implementations provide clear benefits for CSP-providers and their customers. Customers benefit from a wide choice of data transfer options, taking into account the cost of such transfer, simplification of interaction processes, integration, control and delivery. These NaaS features are described as "smoothing at the edges" between the components needed to create end-to-end services. Customers receive a fully automated process of execution and activation of services. CSP providers gain efficiency and agility benefits from increased operational productivity due to a weaker connection of the network that interacts with the systems that support its operation (Figure 1, Figure 2).

2.3 Determining the Location and Required Capacity of Reserved Virtual Computing Resources

To solve the problem of dynamic reconfiguration of computing resources to support NaaS technology, it is necessary firstly to choose which of the possible data centers should be rented for the subsequent placement of virtualized networks on them, as well as to determine how much processor, memory and network bandwidth resources should be reserved for virtual machines for the worst case.

For this purpose, it is proposed to determine the number of virtual networks to be deployed, and to actually link each network function of the traditional network to the data center and the number of resources allocated to the corresponding virtualized network function, with the objective function of cost minimization [11].

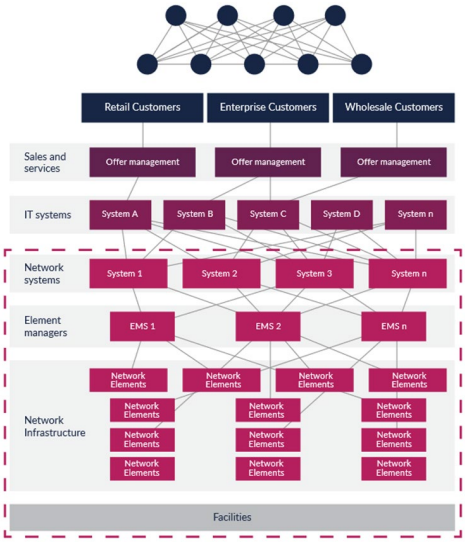


Figure 1: The network side of the services provision by a CSP-provider using a multi-layered NaaS architecture.

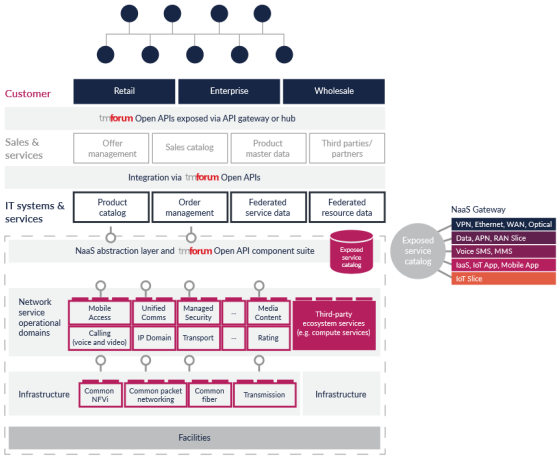


Figure 2: The client side of the services provision by a CSP-provider using a multi-layered NaaS architecture.

We will build a mathematical model of nonlinear optimization with integer variables, which takes into account the available resources on the nodes and their cost, the intensity of the excess service load, and permissible delays as constraints.

The method of solving such a problem is based on the joint location of individual virtualized services of the basic network on the physical network. We assume that the virtual network functions of the mobile core network have the same functionality and interfaces as the network

components of the 3GPP LTE Evolved Packet Core (EPC) architecture. The number of service chains must be determined in advance. A partial case is the consideration of one service chain per cell/eNodeB. Since realistic scenarios for mobile networks are 10000 eNodeBs, the resulting optimization model will be very large and require a long computational time to solve. In this regard, we choose reasonably large clusters of eNodeBs and assume that each of these eNodeB clusters addresses one service chain of the core network.

The case is considered when the provider of telecommunication services already has an existing topology of base stations. It is necessary to define a subset of network nodes where load aggregation units will be placed, which will form requests to one virtualized EPC service. After that, an aggregation node (Traffic Aggregation Point – TAP) is assigned to each base station site.

Let x_i be a binary decision variable that takes the value 1 if a TAP must be placed at point i and 0 otherwise. In addition, we define y_{ji} as a binary variable that takes the value 1 if base station j directs the load to the ith TAP, and 0 otherwise. It is necessary to determine the values of x_i and y_{ji} so as to find the optimal value of the objective function.

Objective function (1) seeks to minimize network delays. The objective function (2) represents the general cost of establishing traffic aggregation nodes and the cost of establishing links between base stations and the corresponding TAPs serving them. Objective function (3) aims to leave more free bandwidth on each physical channel. Residual bandwidth on all channels is maximized, as heavily loaded channels can lead to network congestion, so it is desirable to get a solution where more channel capacity remains.

These optimization goals can be useful for network operators to plan the best deployment strategy.

$$\min_{x_i,y_{ji}} \left(\sum_i \sum_j y_{ji} \cdot L_{ji}\right), \tag{1}$$

$$\min_{x_i,y_{ji}} \left(\sum_i x_i \cdot cost_i + \sum_i \sum_j y_{ji} \cdot costl_{ji}\right), \tag{2}$$

$$\max_{x_i,y_{ji}} \left(\sum_i y_{ji} \cdot (c_{ji} - B_{ji})\right), \tag{3}$$

where L_{ji} is the delay of the communication channel between site j and TAP i; $cost_i$ is a cost that consists of two parts: a fixed initial cost f_i, which is responsible for fixed investments such as space and equipment installation, and additional costs $costN_i$ per unit of processing power installed on a computing node, where d_i is the amount of computing resources used: $cost_i = f_i + costN_i \cdot d_i$;

$costl_{ji}$ is the cost of establishing a connection between site j and TAP i, defined as a linear combination of the initial fixed cost fl_{ji} and the variable part depending on the bandwidth B_{ji} required by the channel and the cost of a unit of bandwidth $costL_j$: $costl_{ji} = fl_{ji} + costL_i \cdot B_{ji}$; c_{ji} – available bandwidth.

It is possible to use a linear combination (4) of expressions (1)-(3) with weighting coefficients a, b, c, which can be used not only to set a greater weighting of one or another component, but also to scale the values of the expressions in order to reduce them to comparable values and have a meaningful composition:

$$\min_{x_i,y_{ji}} \begin{pmatrix} a \cdot \sum_i \sum_j y_{ji} \cdot L_{ji} + b \cdot \left(\sum_i x_i \cdot costN_i + \sum_i \sum_j y_{ji} \cdot costL_{ji} \right. \\ \left. -c \cdot \left(\sum_i y_{ji} \cdot (c_{ji} - B_{ji})\right) \right. \end{pmatrix}. \tag{4}$$

Constraints:

$$\sum_i y_{ji} = 1 \quad \forall j, \tag{5}$$

$$y_{ji} \le x_i \quad \forall j \ \forall i, \tag{6}$$

$$\sum_i x_i \le p, \tag{7}$$

$$\sum_j y_{ji} \cdot d_j \le p_i \quad \forall i, \tag{8}$$

$$\sum_i y_{ji} \cdot (c_{ji} - B_{ji}) \ge 0 \quad \forall j, \tag{9}$$

$$\sum_i y_{ji} \cdot L_{ji} \le T_j \quad \forall j. \tag{10}$$

Constraint (5) ensures that each base station will be attached to one TAP. Constraint (6) ensures that a link is established between base station site j and TAP i only if i has been placed.

Constraint (7) ensures that the maximum number of TAPs does not exceed the budget p, while (8) is a capacity constraint that ensures that the total processing requirements of all base stations assigned to a particular TAP do not exceed the actual installed physical resources p_i. Constraint (9) guarantees the sufficiency of channel resources for establishing channels, and (10) the admissibility of the delay value, i.e. not exceeding the threshold T_j.

The physical network is given in the form of a graph $SN = (N, NE)$, where N is the set of physical nodes and L is the set of channels. Each channel $l = (n_1, n_2) \in NE$, $n_1, n_2 \in N$ has a maximum bandwidth $c(n_1, n_2)$ and each node $n \in N$ is associated with certain resources c_n^i, $i \in R$, where R is the set of types resources (CPU, memory, disk space, network interface). The set of all Traffic Aggregation Points (TAP), i.e. eNodeB clusters, in the network is denoted by $K \subseteq N$. For each node n $\in$ N, $suit_n^{k,j}$ is a binary parameter that indicates whether it is administratively possible to deploy a function of type $j \in V$ on the node, where V is the set of types of network functions, k-th service, where $k \in K$.

The virtual basic mobile network is represented by a set of services (one service per TAP) that are built into the physical network.

The channel bandwidth requirements between two functions, *j1* and *j2*, *(j1,j2)* $\in E$ related to the TAP service $k \in K$ is denoted as $d_k^{(j1,j2)}$. $d_k^{j,i}$ is the amount of computing resource of type *i* allocated for network function *j* of service *k*. $s_{n,i}^{k,j}$ denotes the time of processing a request on resource type *i* of virtual network function *j* of service *k* by one resource unit of node *n*. The requirements for the allowable processing time of the request by the network function *j*, which is related to the service *k*, are denoted as P_k^j. T_k is the maximum delay for $k \in K$, $L(n_1,n_2)$ is the network delay for the channel $(n_1,n_2) \in NE$.

The goal of the optimization is to find the location of the virtualized services of the core network (that is, the placement of network functions and the distribution of resources, as well as the determination of traffic transmission paths between them), so as to minimize the conditional costs of the occupied resources of channels and nodes in the physical network, while satisfying the load requirements $\lambda_{k,j}$. We formulate the objective function (expression (11)) in the form of a linear combination of two cost expressions: the occupied amount of computing nodes resources, where the conditional cost of a unit of resource *i* on node *n* is denoted as *costN(i,n)*, and the occupied bandwidth of channels, where $costL(n_1,n_2)$ is the conditional cost of a physical channel bandwidth unit $(n_1,n_2) \in NE$.

The following expressions (11)-(20) represent the formulation of the nonlinear programming optimization problem. Boolean variables $x_n^{k,j}$ indicate whether network function *j* associated with service *k* is located on physical node *n*. For *j*=TAP, $x_n^{k,TAP}$ are not variables, but input parameters that indicate where TAP *k* is located, i.e.

$$x_n^{k,TAP} = \begin{cases} 1 \text{ if } k=n, \\ 0 \text{ otherwise} \end{cases}$$

Similarly, the boolean variables $f_{(n1,n2)}^{k,(j1,j2)}$ indicate whether the physical channel $(n1,n2) \in NE$ is used for the path between *j1* and *j2* for service *k*.

Expression (12) ensures that only one network function of each type is placed for each TAR/service. Expression (13) guarantees that the placement of resources is carried out on physical nodes that have the administrative ability to place the corresponding network functions. Expressions (14), (15) and (16) represent limitations on the available resources of physical nodes and channels. Expression (17) is a flow conservation constraint for all paths in the physical network. Expression (18) guarantees that the variables in the network function placement and path mapping problem are Boolean.

$$\min_{f_{(n_1,n_2)}^{k,(j_1,j_2)}, d_t^{j,i}} \left(\begin{array}{l} \sum_{k\in K}\sum_{j\in V}\sum_{n\in N}\sum_{i\in R} x_n^{k,j}\cdot d_k^{j,i}\cdot costN(i,n) + \\ + \sum_{(n_1,n_2)\in L} costL(n_1,n_2)\cdot \sum_{k\in K}\sum_{(j_1,j_2)\in E} f_{(n_1,n_2)}^{k,(j_1,j_2)}\cdot d_k^{(j_1,j_2)} \end{array} \right) \quad (11)$$

$$Constraints \sum_{n\in N} x_n^{k,j} = 1 \quad \forall k\in K, j\in V \quad (12)$$

$$x_n^{k,j} \le suit_n^{k,j} \quad \forall k\in K, j\in V, n\in N \quad (13)$$

$$\sum_{(w,n)\in NE}\sum_{k\in K}\sum_{(j_1,j_2)\in E} f_{(w,n)}^{k,(j_1,j_2)}\cdot d_k^{(j_1,j_2)} \le c_n^{bdw} \quad \forall n\in N \quad (14)$$

$$\sum_{k\in K}\sum_{j\in V} x_n^{k,j}\cdot d_k^{j,i} \le c_n^{i} \quad \forall n\in N, i\in\{R\backslash bdw\} \quad (15)$$

$$\sum_{k\in K}\sum_{(j_1,j_2)\in E} f_{(n_1,n_2)}^{k,(j_1,j_2)}\cdot d_k^{(j_1,j_2)} \le c(n_1,n_2) \quad \forall (n_1,n_2)\in NE \quad (16)$$

$$\sum_{w)\in NE} f_{(w,n)}^{k,(j_1,j_2)} - f_{(n,w)}^{k,(j_1,j_2)} = x_n^{k,j_1} - x_n^{k,j_2} \quad \forall k\in K, n\in N, (j_1,j_2)\in E \quad (17)$$

$$x_n^{k,j}, f_{(n_1,n_2)}^{k,(j_1,j_2)} \in \{0,1\} \quad \forall k\in K, j\in V, n\in N, (j_1,j_2)\in E, (n_1,n_2)\in NE \quad (18)$$

$$\sum_{(j_1,j_2)\in E}\sum_{(n_1,n_2)\in L} f_{(n_1,n_2)}^{k,(j_1,j_2)}\cdot L(n_1,n_2) \le T_k \quad \forall k\in K \quad (19)$$

$$\sum_{n\in N} x_n^{k,j}\sum_{i\in R}\left(\frac{1}{\frac{j,i}{d_k}\cdot\lambda^{k,j}}\right) \le P_k^j \quad \forall t\in T, j\in V \quad (20)$$

To limit channel delays, the delay constraint shown in expression (19) is also added. And in order to take into account the required performance of the virtual network function in the model, restrictions on the value of the application processing time defined in expression (20) are necessary.

The problem (11)-(20) is supposed to be solved offline at the initial stage. According to the solution, each network function is reserved a certain amount of resources of the virtual network function, based on the assessment of its greatest need for resources; the instantaneous needs of various network functions are dynamically met by activating the necessary configuration of virtual machines at runtime in such a way as to satisfy the guarantees provided for each network function.

The solution to the optimization problem can be found using a real coded genetic algorithm MI-LXPM.

2.4 Evaluation of the Method of Computing Resources Dynamic Reconfiguration

The cost of NaaS deployment consists of the cost of launching and using a server, using server resources, using communication channels and using migration resources. Quantitative and qualitative analysis (Figure 3) of the proposed method showed a reduction of costs associated with reserved resources by up to 15%, which contributes to increasing the efficiency of load maintenance and saving computing resources.

Thus, it is possible to draw a conclusion about the effectiveness of the proposed method, which

allows you to link the performance of the network virtualized function with the amount of resources allocated to it, as well as take into account the hybridity of the telecommunications environment. The method can be applied dynamically when managing the network functions deployment in a hybrid hardware environment to minimize the costs for the communication operator and improve the quality of service for subscribers.

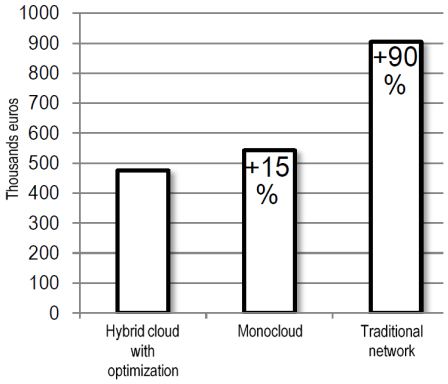


Figure 3: Costs of system resources using the NFV placement method and not using it.

3 CONCLUSIONS

This article considers: the classification of services proposed by TM Forum as the main factors that determined the emergence of NaaS technology, the level of requirements for managing network services in real time and the level of automation, which will be decisive for increasing the economic efficiency of providing modern services to end users.

A method of managing placement and determining the required amount of resources of network functions is proposed to optimize the amount of computing resources allocated to a network function in a telecommunications operator's network.

The proposed method allows you to bind the performance of the network virtualized function with the amount of resources allocated to it, and also takes into account the hybridity of the telecommunications environment.

The method can be applied in managing the deployment of network functions in a hybrid hardware environment to minimize the costs for the communication operator and improve the quality of service for subscribers.

Future research should be related to the development of technology for implementing the proposed method of dynamic reconfiguration of computing resources into the system architecture of the CSP-provider and its adaptation in real conditions..

REFERENCES

[1] Jerbi et al., "Enabling Multi-Provider Cloud Network Service Bundling," 2022 IEEE International Conference on Web Services (ICWS), Barcelona, Spain, 2022, pp. 405-414, doi: 10.1109/ICWS55610.2022.00067.

[2] M. C. Luizelli, L. R. Bays, L. S. Buriol, M. P. Barcellos, and L. P. Gaspary, "Piecing together the NFV provisioning puzzle: Efficient placement and chaining of virtual network functions," 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), Ottawa, ON, Canada, 2015, pp. 98-106, doi: 10.1109/INM.2015.7140281.

[3] P. T. A. Quang, Y. Hadjadj-Aoul, and A. Outtagarts, "A Deep Reinforcement Learning Approach for VNF Forwarding Graph Embedding," in IEEE Transactions on Network and Service Management, vol. 16, no. 4, pp. 1318-1331, Dec. 2019, doi: 10.1109/TNSM.2019.2947905.

[4] M. Masoumi et al., "Dynamic Online VNF Placement with Different Protection Schemes in a MEC Environment," 2022 32nd International Telecommunication Networks and Applications Conference (ITNAC), Wellington, New Zealand, 2022, pp. 1-6, doi: 10.1109/ITNAC55475.2022.9998347.

[5] R. Hussain and S. Zeadally, "Autonomous Cars: Research Results, Issues, and Future Challenges," in IEEE Communications Surveys & Tutorials, vol. 21, no. 2, pp. 1275-1313, Secondquarter 2019.

[6] S. Iranmanesh, R. Raad, M. S. Raheel, F. Tubbal, and T. Jan, "Novel DTN Mobility- Driven Routing in Autonomous Drone Logistics Networks," in IEEE Access, vol. 8, pp. 13661-13673, 2020.

[7] N. Mhaisen, O. Abazeed, Y. A. Hariri, A. Alsalemi, and O. Halabi, "Self-Powered IoT- Enabled Water Monitoring System," 2018 International Conference on Computer and Applications (ICCA), Beirut, 2018, pp. 41-45.

[8] H. T. Yew, M. F. Ng, S. Z. Ping, S. K. Chung, A. Chekima, and J. A. Dargham, "IoT Based Real-Time Remote Patient Monitoring System," 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Langkawi, Malaysia, 2020, pp. 176-179.

[9] H. Dia, “The real-time city: Unlocking the potential of smart mobility”, In Proceedings of the 38th Australasian Transport Research Forum (ATRF 2016), Melbourne, Australia, 16-18 November 2016.

[10] C. Babb, C. Curtis, S. McLeod, Sam, “The Rise of Shared Work Spaces: A Disruption to Urban Planning Policy?”, 2018, Urban Policy and Research, vol. 36., pp. 1-17, doi: 10.1080/08111146.2018.1476230.

[11] S.V. Sulima “Reconfiguration methods of the computing resources for the core network based on virtualization technology,” qualification scientific paper, manuscript.

Wireless Multiplayer Retro Gaming Console

Zlate Bogoevski, Kristijan Nelkovski, Danijela Efnusheva and Marija Kalender
Computer Technologies and Engineering Department, Faculty of Electrical Engineering and Infomation Technologies, "SS. Cyril and Methodius University" in Skopje, Rugjer Boshkovikj Str. 18, Skopje, N. Macedonia
{zlateb,ksiar212018,danijela,marijaka}@feit.ukim.edu.mk

Keywords: IoT System Design, Wireless Communication, Video Game Hardware, Handheld Gaming, Motion Sensor, PONG.

Abstract: Handheld video gaming has been a staple in the video game industry for over thirty years. While the current generation of handheld video game consoles are much more powerful than their predecessors, a substantial community of enthusiasts still enjoy the tactile feel and retro aesthetics of much older and more traditional hardware and media. Newer handheld devices that aim to replicate this through software emulation and contemporary hardware typically lack the wireless multiplayer connectivity that has since become a key feature for modern day handheld and mobile video game titles and devices. In this paper, we analyze these technologies and their software by developing our own basic handheld gaming console with wireless multiplayer capabilities for cooperative and competitive gaming between devices through the use of Wi-Fi ESP8266 based microcontrollers. The results demonstrate the development of two separate nodes capable of complete wireless interconnectivity for simultaneously running a proof of concept multiplayer recreation of the classic "Pong" arcade game.

1 INTRODUCTION

While processor and microcontroller architectures have seen a substantial increase in their performance and capabilities in conjunction with advancements in miniature display technologies such as color, pixel density and low power consumption, companies and engineers have been working on a multitude of new types of handheld gaming devices, combining the possibilities offered by electronic and electrical engineering with the art and format of graphic and interactive storytelling offered by the video game medium.

Although the current generation of handheld gaming consoles such as the Nintendo Switch, the Steam Deck, and even mobile Android devices (both regular smartphones as well as dedicated consoles) are vastly superior to the technologies of the eighties and nineties, many video game enthusiasts are still interested in older generations of video games that they remember from their own past.

What perhaps started as childhood nostalgia has been cultivated by consumer electronics companies through the phenomenon known as "retro handheld" devices which have seen increasingly more powerful electronics capable of emulating a wide range of gaming systems from the earliest examples such as 1980s consoles like the Atari 2600, to some more recent systems like the early 2000s PlayStation 2, typically in the form factor, look and feel of some of the most popular handheld devices of the past, most notably Nintendo's GameBoy line of products (original GameBoy, GameBoy Color, GameBoy Advance).

The rapid development of these modern emulation machines has been made possible in part thanks to the widespread availability and affordability of microcontroller and single board computer platforms such as the Arduino and the Raspberry Pi, that have enabled software emulation of older video games, but have also allowed for developers to make recreations (clones) as well as their own custom titles with similar gameplay and art styles without having to interact with any legacy hardware and software.

The aim of this paper is to design a retro handheld gaming device complete with wireless networked multiplayer features capable of interacting with multiple nodes of the same kind.

The paper is organized as follows: Section 1 provides an Introduction. Section 2 analyses the state of the art. Section 3 gives details about the methodology used in the design of the proposed solution. Section 4 discusses the practical implementation and results. The paper concludes with Section 5.

2 STATE OF THE ART

The proposed project falls into the second category mentioned above, prompted by devices such as the Arduboy and Makerbuino which are standalone 32 bit AVR microcontroller based single board miniature gaming consoles capable of running classic arcade style games in the form of open source code offered by the manufacturers of these devices as well as community software later developed by the end users themselves.

Using wireless networked microcontrollers our goal was creating a similar type of device that was capable of not just running custom-made arcade style video games, but also have on board wireless connectivity, capable of interacting with other units of the same type allowing for both competitive and cooperative gameplay between multiple users.

Wired multiplayer has been enabled with traditional handheld video game devices even on consoles such as the classic Nintendo GameBoy, although nothing compared to the wireless Internet capabilities of the Nintendo Switch and the Steam Deck, of course in conjunction with the server infrastructure that their owner companies possess today.

This work was prompted by research into the technologies behind the classic GameBoy, notably two works that aim to bring wireless capabilities to this thirty plus year old device.

The first work that inspired our research is an ESP8266 based wireless GameBoy cartridge [1] developed by physicist Dr. rer. nat. Sebastien Staacks capable of connecting and browsing the Internet while not necessarily allowing for multiplayer gameplay on the device right out of the box. This device fits the form of a typical GameBoy game cartridge, traditionally used to load individual games onto the device, allowing for the user to develop software clients for browsing the Internet and graphically displaying this information using the device itself as an interface.

The second project that inspired our work is a RaspberryPi Pico wireless gaming link cable adapter [2] for the GameBoy, developed by IT security researcher Thomas Roth which does enable online multiplayer through manipulating the GameBoy link cable protocol via the RaspberryPi Pico.

While these two projects build upon the existing hardware of old and hard to come by technologies, as well as legacy and unstandardized software, our goal was to develop both a hardware and software solution that reflects the modern IoT and maker scene while being both affordable and capable of performing our desired task.

We are further referencing two more projects that have aided us in the development of a classic Pong game as a prototype/proof-of-concept for our device although vastly differing in the hardware and later software of our final work. These two projects are the Arduino Pong game developed from scratch by Mark Geddes [3] without the use of graphical display libraries created for teaching and understanding Arduino display driving via SPI (serial peripheral interface) and the How to write a game instructable project by Hari Wiguna [4] developed as a non-networked basic multiplayer Pong game on a single Arduino based breadboard device.

3 METHODOLOGY

This system is developed using an ESP8266 which is a low cost 32-bit Wi-Fi microcontroller based on the Tensilica Diamond Standard 106 Micro CPU architecture with 32 KB of instruction memory and 80 KB user data memory supporting up to 16 MB of external flash while having no internal ROM. This module features integrated TCP/IP networking software as well as other communications protocols such as SPI and I2C/UART. This device can perform as either a standalone unit with 17 dedicated general-purpose input output pins or act as a slave device for a host chip, reducing communications overhead for a main applications processor.

For this project the ESP8266 chip is used as a main independent microcontroller in the form of the commercially available WeMos D1 mini development board which includes 11 digital GPIO pins and the full Wi-Fi IoT capabilities of the ESP8266 platform (Figure 1). Although there are many vendors of ESP8266 microcontroller boards, the WeMos D1 mini is a widely used product with extensive documentation, a small footprint as well as convenient breakout pin headers, onboard 3.3v voltage regulation through the ME6211 chip and integrated USB to serial programming capabilities through the CH340 USB bus converter IC.

Individual nodes in our system are built with identical hardware and are able to facilitate fast full-duplex peer-to-peer communication via Espressif's ESP-NOW protocol which is based on the IEEE 802.11 b/g/n Wi-Fi wireless communication network standard and works on a radio frequency of 2.4GHz. Using this protocol our nodes can wirelessly transfer game information between multiple player devices

and show correct positions, scores and flags to each individual device/player with extremely low latency without the necessity of a dedicated server or interfacing with the Internet.

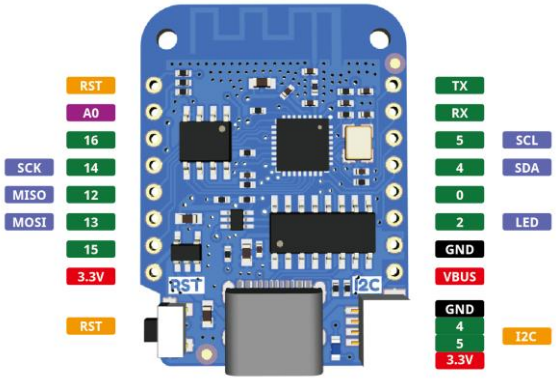


Figure 1: The WeMos D1 mini board and pinout.

3.1 Hardware

Each node in the proposed system is a standalone device created with a custom printed circuit board (Figure 2, Figure 3) and equipped with its own D1 mini board, a Nokia 5110 84 by 48-pixel graphic LCD, an ATtiny85 microcontroller, two piezoelectric buzzers, a GY-271 magnetic field sensor, a PCF8574 I2C remote 8-bit I/O expander, eight momentary input switches and two 2xAAA battery packs.

The Nokia 5110 LCD is a basic dot matrix graphic LCD display that uses the PCD8544 low powered CMOS controller driver which enables the control of 48 rows and 84 columns of pixels via SPI. This interface method offers synchronous serial communication for short distances primarily developed for embedded systems and extensively covered for the use and connectivity in numerous microcontroller platforms such as the ESP8266 boards. This display can only show 1-bit colors per pixel, has a diagonal of about 1.5 inches and features four LEDs as backlight, although not used for this system due to power efficiency.

Our system utilizes a singular ATtiny85 microcontroller as a simple solution for audio output processing. The ATtiny85 microcontroller developed by Atmel (now part of Microchip) is an 8-bit AVR RISC architecture device with 8 KB of ISP flash memory, 512 B of EEPROM memory and 512 B of SRAM memory. This module features six GPIO pins and has a clock speed of 8 MHz. In this project the ATtiny85 microcontroller is used to store multiple segments of MIDI (musical instrument digital interface standard) music and output it as background game music via a single piezoelectric buzzer.

This music is converted from original MIDI files into program code via Extramaster's online "MIDI to Arduino" tool and is stored within the software files of the ATtiny program capable of being activated by the ESP8266 board through two pins for simplex communication (ESP to ATtiny only) and a third pin for exiting the music loop via resetting the program on the ATtiny85 chip. The ATtiny code can have multiple user defined loops of music which contain different MIDI snippets and can be chosen through the one-way communication by the ESP8266 chip.

While the D1 mini is also capable of performing the same function of MIDI playback through a piezoelectric buzzer, the ESP8266 device integrates only a single core processor that makes it difficult to run both software loops of musical notes as well as game and wireless communications code concurrently. While this can be done through the use of multithreading or via connecting different libraries and overcomplicating the main system code which might impair speed and system memory, using a dedicated microcontroller for background game music is a simpler solution without overwhelming the software running on the main ESP8266 processor.

For singular tones of game sounds which require only one line of code per tone for playback however, this does not represent any problems which is why our system employs a second piezoelectric buzzer that is directly controlled by the ESP8266 chip and can activate singular notes at a time whenever a game function requires it.

One of the two control input methods for our system is the GY-271 magnetic field sensor. This device is a three-axis magnetometer module that uses the Honeywell HMC5883L digital compass IC. The chip itself includes multiple state of the art high resolution HMC118X series magneto-resistive sensors with a 12-bit ADC that enables compass heading accuracy within one to two plane angle degrees. For communications this chip uses the I2C protocol and is intended for use with consumer electronics devices such as phones, tablets and vehicles, while also offering sufficient documentation for the use with microcontrollers such as the ESP8266. In this system this sensor is used as a motion controller for the player through the detection of the change in a user's heading. This gives players the option to move their on-screen

characters via their own physical movements in the real world.

The second input method for controlling our system are eight standard momentary switch push-buttons with a clicky tactile feel laid out in two thumb operated directional pads with one button on each point similar to most modern and retro gaming console gamepad and controller layouts. The layout of the buttons is as follows: the left directional pad is marked with left, right, up and down arrows while the right directional pad is marked with the first four letters of the Cyrillic alphabet: A, B, V and G. These buttons are meant to be used as the primary method of interaction and selection within the main system menu as well as subsequent game menus. Furthermore, the buttons can be used as the primary method for moving the on-screen game characters or as a secondary method for movement combined with the use of the previously mentioned GY-2711 magnetometer.

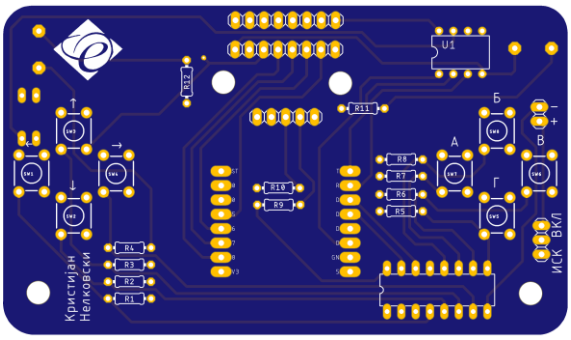


Figure 2: A CAD model of the front of the printed circuit board design.

Figure 3: A powered on fully soldered, functional node.

Because the WeMos D1 mini only has 11 digital general-purpose input/output pins, in order to facilitate the stable interconnectivity with all of the previously covered hardware, the system utilizes a GPIO expander IC for reading out the eight button inputs. This is done using the Texas Instruments PCF8574 8-bit input/output expander connected to the ESP8266 through the same I2C bus as the GY-271 sensor. Identifying each device is done through software via a fixed address for the sensor and a three-bit user adjustable (by connecting these bits to either ground or 3.3v) hardware address input for the GPIO expander.

3.2 Software

The WeMos D1 mini board can be programmed either using the Arduino C/C++ programming language or the microcontroller MicroPython implementation of the Python programming language. For our system we chose to use the Arduino C/C++ language because of the multitude of up to date, heavily documented and readily available libraries that are compatible with each other and capable of connecting all of our hardware elements. The programming environment used for this project is Microsoft's Visual Studio IDE in conjunction with the Visual Micro extension that uses Arduino's standard IDE in the background.

The software for our device is developed using miscellaneous default Arduino libraries such as SPI.h for facilitating communication via the SPI protocol and Wire.h for facilitating communication via the I2C protocol. The other two sets of libraries are specific Espressif as well as Adafruit libraries.

The first Espressif library used in this project is the ESP8266WiFi.h library based on Arduino's WiFi library used here for facilitating the connection between individual nodes. The second Espressif library is the espnow.h library which enables peer-to-peer communication without the need to connect to a server or the Internet. Using these two libraries our devices are capable of communicating with each other via previously known and software established MAC addresses through which they are capable of sending packets of information to specific destinations between nodes without the need for handshakes or packet receive acknowledgements.

Broadcast MAC address setting code in Arduino C:

uint8_t broadcastAddress[] = {0xA4, 0xE5, 0x7C, 0xB5, 0xDF, 0x9C};

This project also uses five Adafruit libraries developed by the open-source hardware company Adafruit, a popular hobby electronics manufacturer that offers free software and documentation for many Arduino and other microcontroller compatible hardware components. To begin with, our Nokia 5110 display is

controlled using the Adafruit_GFX.h and Adafruit_PCD8544.h libraries which allow for drawing text, shapes and images as well as reliable refreshing of the LCD through its onboard controller IC. These libraries also come with premade dot matrix fonts while offering support for using custom libraries such as our custom Cyrillic font for displaying text in the Macedonian language. The Adafruit_Sensor.h and Adafruit_HMC5882_U.h libraries facilitate the calibration and readout of the GY-271 magnetometer, and lastly the Adafruit_PCF8575.h allows for reading the input pins of our PCF8575 hardware expander.

4 FUNCTIONALITY, DEMONSTRATION AND RESULTS

For this project we have developed a custom printed circuit board (Figure 4) using the Eagle CAD electronic design automation (EDA) software in order to deliver a standalone, compact and aesthetic device that can be soldered together and assembled creating a single node that can connect with other nodes of the same kind and run complementing software for multiple users to play within the same competitive multiplayer game.

As a demo for this project we have decided to recreate one of the oldest retro arcade games – PONG. Pong is a two-player game with the screen set as a tennis or table tennis court where both players operate paddles on the opposite side of each other and are required to hit a moving ball to the other side of the screen avoiding it falling through their side of the screen. Traditionally the paddles in this game and its recreations are placed on the left and right sides, however due to the motion control functionality implemented in our device this demo has the players' paddles on the bottom and on the top of the screen for moving them by tilting the device left and right as opposed to moving the whole device up or down.

After both players have entered the game, only the host of the game is able to start it. When the game starts, both devices display the same image on the Nokia 5110 display. The code that does this perpetually draws the ball, paddles, court and score onto the screen using an X and a Y coordinate for each element of the game. While the court and score are stationary objects, the ball and paddles move while the value of the score updates for each missed ball hit.

The code between host and client varies in the source of these coordinates. For the host, the coordinates of its player come from either the GY-271 sensor or two of the push buttons (either left or right arrows or A and V buttons) while the coordinates for the other player come via information sent by the other device through the ESP-NOW protocol of its own GY-271 sensor or its own push buttons (again either left or right arrows or A and V buttons).

For the client both paddle coordinates are received from the host through the ESP-NOW protocol as well as the coordinates for the ball whose collision and direction are detected and operated only by the host device. The host device also updates the score and conveys this information along to the client.

This is done by sending a data structure filled with information such as coordinates, scores and flags which must be declared identically in both the host and client programs.

Example of the data structure in Arduino C:

```
typedef struct struct_message {
    byte ballX;
    byte ballY;
    byte paddle1X = 0;
    byte paddle2X = 0;
    byte control;
    byte score [2];
    byte checkStart;
} struct_message;
```

Throughout the code for both nodes, two data structures are constantly being updated, one being sent to while the other one being received from the other device. This is why for our project different nodes run different code and for functional operation one of the nodes must always be the host and others the clients essentially processing most of the gameplay mechanics only on one device and then streaming the same image on the second (or multiple) device.

Although it is possible to run the same code on both (or multiple) nodes at the same time, cross updating information between nodes and allowing them to individually detect collisions and calculate scores, latency between magnetometer or push-button readout cycles as well as LCD refresh rates can cause lag or disturb the synchronization between the ball and the paddles throughout both devices leading to mismatching object coordinates by one or

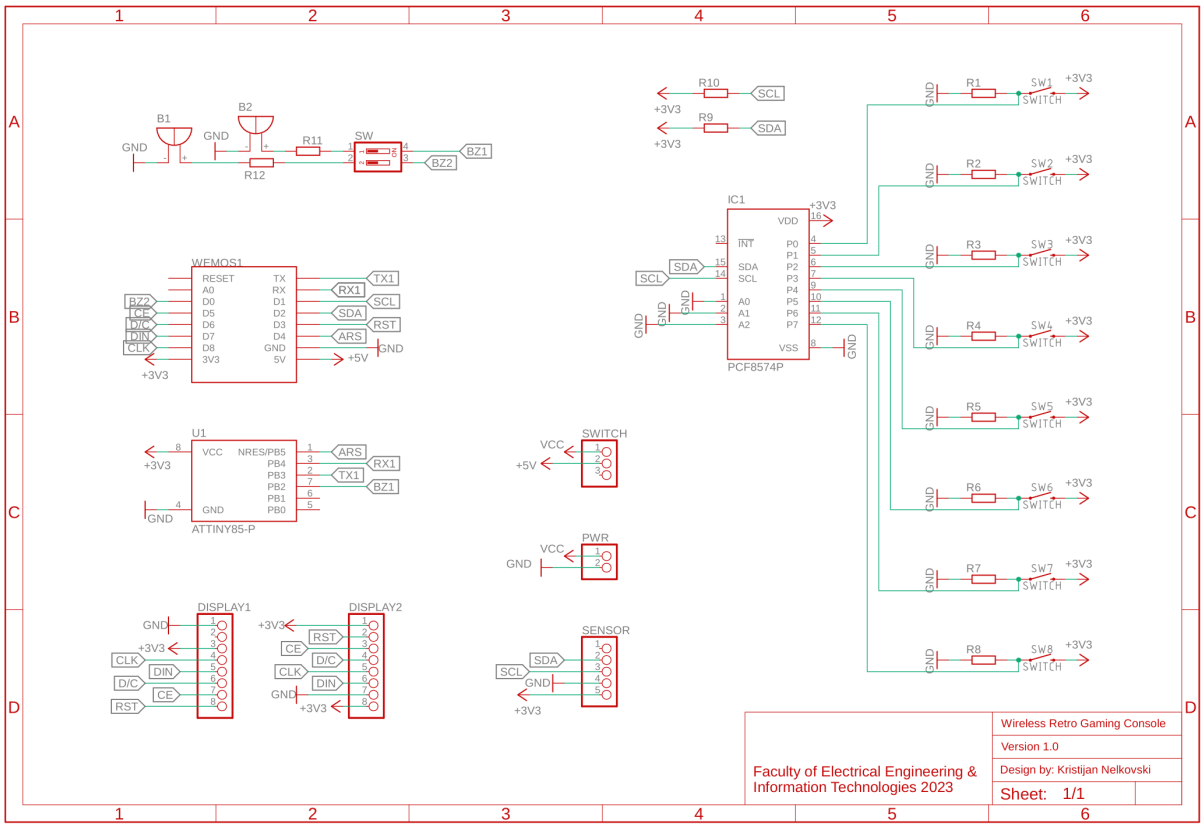


Figure 4: The circuit board schematic designed in Eagle CAD.

more pixels which renders gameplay useless as even a single pixel carries a lot of weight on a small matrix of only 84 by 48 pixels. While there might still be latency and inconsistency between data transfer and writing images on the displays by using our method within the range of a couple dozen milliseconds, for the user this is far less noticeable and will never lead to mistakes of mismatched object coordinates, not even by a single pixel.

5 CONCLUSIONS

In this paper a handheld, networked and multiplayer video gaming device has been developed. With the use of ESP8266 based wireless microcontroller hardware and Arduino based software development tools, a viable recreation of the classic Pong arcade game was created as a two-player competitive video game demonstration with optional motion sensing controls, piezoelectric audio outputs as well as 1-bit (black and white) dot matrix LCD graphics.

For future development this system can be expanded with creating different types of video games for more than just two players and integrating higher resolution color LCD displays for more complex graphics, venturing closer into the territory of late 1990s handheld and mid 2000s Java mobile phone games as opposed to simpler 1970s and 1980s style arcade titles.

REFERENCES

[1] S. Staacks, “There oughta be a WiFi Game Boy cartridge,” December 2021, [Online]. Available: https://there.oughta.be/a/wifi-game-boy-cartridge.

[2] T. Roth, “Online Multiplayer on the Game Boy,” May 2021, github.com, youtube.com, [Online]. Available: https://github.com/stacksmashing, https://www.youtube.com/watch?v=KtHu693wE9o.

[3] M. Geddes, “Arduino Project Handbook,” vol. 2, No Starch Press, San Francisco, California, 2017.

[4] H. Wiguna, “How to write a game,” June 2016, instructables.com, youtube.com, [Online]. Available: https://www.instructables.com/HariFun-136-How-to-Write-a-Game/, https://www.youtube.com/watch?v=Xzx-cp60T7k.

[5] K.J.M. Kiili., K. Devlin, and J. Multisilta, “Editorial: is Game-Based Math Learning Finally Coming of Age?”, int.l Journal of Serious Games, vol. 2, nr. 4., 2015, doi: 10.17083/IJSG.V2I4.109.

[6] Y.A. Badamasi, "The working principle of an Arduino", 11th Int. Conf. on Electronics, Computer and Computation (ICECCO), October 2014, doi: 10.1109/ICECCO.2014.6997578.

[7] D. Abbott, "Game-based learning for postgraduates: an empirical study of an educational game to teach research skills.", High Educ Pedagog. 2019, doi: 10.1080/23752696.2019.1629825.

[8] J. Hartley, "Toshiba progress towards sensory control in real time," Indust. Robot, vol. 14, no 1, pp. 50–52, 1987.

[9] R.L. Anderson, "A Robot Ping-Pong Player: Experiments in Real Time Control.", Cambridge, MA: MIT Press, 1987.

[10] E.R. Davies, "Machine Vision." New York: Academic, 1997.

[11] A. Rosenfeld and A. Kak, "Digital Picture Processing", vol. 1–2. New York: Academic Press, 1982.

Comparative Analysis of Local Positioning Methods in Wi-Fi/Indoor Networks

Irina Strelkovskaya[1], Irina Solovskaya[2] and Juliya Strelkovska[3]
[1]*Department of Information Technology, International Humanitarian University, Fontanska road Str. 33, Odesa, Ukraine*
[2]*Department of Computer Science, International Humanitarian University, Fontanska road Str. 33, Odesa, Ukraine*
[3]*Worthing college, Sanditon Way 1, Worthing, United Kingdom*
{i.strelkovskaya, i.solovskaya}@mgu.edu.ua, 4800632s@gmail.com

Keywords: Location, Wi-Fi/Indoor Network, Fingerprinting Method, Probabilistic Bayesian Method, Quadratic Complex Planar Spline, Positioning Accuracy.

Abstract: The rapid development of various LBS-oriented applications and services requires the development of new and improvement of known local positioning methods in order to increase the accuracy of determining the user's coordinates. First of all, it connects to methods for determining the location of users in rooms, subject to a high concentration of users and the presence of various difficulties in the propagation of radio signals. To determine the local location of a user in a Wi-Fi/Indoor network, the Fingerprinting method is considered. The comparison of the results of user positioning in the Wi-Fi/Indoor network based on the Fingerprinting method using various algorithms for determining user coordinates, such as the probabilistic Bayesian method and complex approximation using complex planar quadratic splines, was carried out. It has been established that the use of the Fingerprinting local positioning method using quadratic complex flat splines can improve the accuracy of determining the user's location coordinates, thereby ensuring the provision of LBS services and applications to users in the premises.

1 INTRODUCTION

The development of LBS (Location-based Services) services and applications is in demand today due to the active introduction of IoT (Internet of Things) technologies and devices, as well as further improvement of UE (User Equipment) mobile communication subscriber devices for determining the user's location, geolocation, local search and targeted advertising. LBS services are quite demanding in terms of positioning accuracy and appropriate choice of radio access technology, especially for IPS (Indoor Positioning System) indoor positioning. The most commonly used radio access technologies for local positioning are Wi-Fi (IEEE 802.11n/ac/ad), Bluetooth or BLE (Bluetooth low energy) (IEEE 802.15.1). Other technologies are also used, such as RFID (Radio Frequency IDentification) and NFER (Near-Field Electromagnetic Ranging), optical and inertial technologies, built-in smartphone sensors (gyroscope, compass, accelerometer, barometer, etc.) and LED lamps, magnetic sensors are used and many others [1-3]. Wi-Fi technology (IEEE 802.11n/ac/ad) has an unconditional priority for local positioning due to its widespread use on almost all user devices and indoor coverage, which allows using the existing infrastructure of already widely deployed radio access networks.

An important and urgent task today is the choice of a local positioning method that would allow achieving high positioning accuracy, and, given the need for implementation on smartphones, would satisfy the conditions of restrictions on power, memory, security, speed and the amount of necessary computing resources.

It is known [1-5] that the Fingerprinting method provides the highest positioning accuracy, which can use various algorithms to determine the location coordinates, such as: k-nearest neighbors KNN (K-Nearest Neighbor) and k-nearest neighbors methods using WKNN weight coefficients (Weighted K-Nearest Neighbor) [5],[7-8], probabilistic Markov and Bayesian methods [4-6], least squares [3-4] and neural network methods [9-10].

Using each of these methods to determine the user's coordinates in a Wi-Fi/Indoor network allows

achieving the required positioning accuracy results with varying degrees of error, computational complexity, and time costs.

Therefore, the authors consider it appropriate to consider the problem of positioning in a Wi-Fi/Indoor network for the Fingerprinting method based on the probabilistic Bayesian method and compare the results obtained with the positioning results using a complex spline approximation.

The purpose of this work is to compare the accuracy results of the local positioning of a user in a Wi-Fi/Indoor network using the Fingerprinting method based on a probabilistic Bayesian approach and complex spline approximation.

2 THE FINGERPRINTING METHOD IN LOCAL POSITIONING PROBLEMS

The Fingerprinting method is used for local positioning in a Wi-Fi/Indoor network and is known as the method with the highest positioning accuracy. The functioning of the Fingerprinting method is performed in two stages (Figure 1), one of which is implemented off-line and involves the formation of a Fingerprinting DataBase of the results of measuring the values of the RSSI signal strength level (Received Signal Strength Indicator) for the room in question (Measuring the original of the RSSI signal at the RPs) at predetermined reference points RP (Reference Point), which are distributed evenly throughout the room (Arrange Reference Point).

The Fingerprinting DataBase stores measurement data of RSSI power values and the MAC address of access points, AP_i, $i=\overline{1,3}$, known as "radio fingerprints" [11-12]. At the second stage on-line while positioning the user, measurements of the current RSSI values are performed (On-line measuring the RSSI of the estimation position).

The received data is matched against RSSI values stored in the DataBase. To determine the user's coordinates, we use the Bayesian approach and complex approximation based on flat quadratic spline functions [13-14].

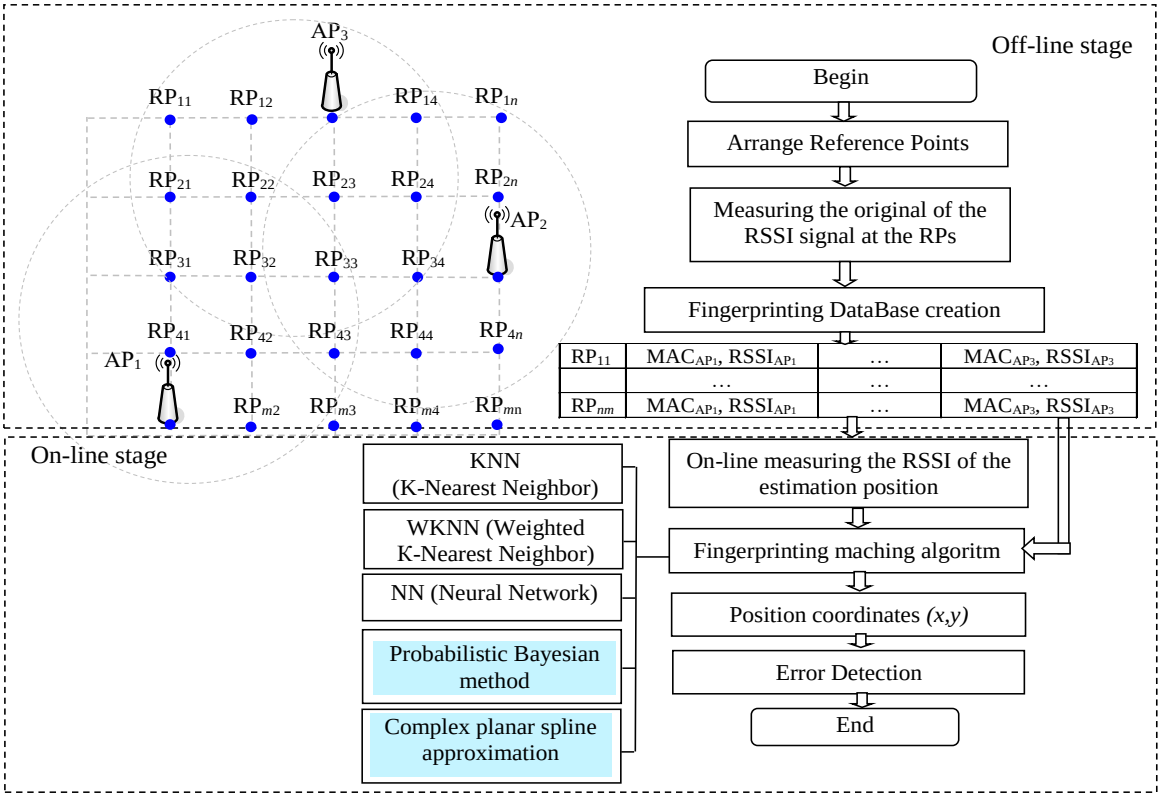


Figure 1: Algorithm for local positioning Fingerprinting of a user in a Wi-Fi/Indoor network.

2.1 Finding the User's Location using Complex Spline Approximation

To find the user's location in the Wi-Fi/Indoor network, consider a spline approximation. As an approximating function, we will use a complex plane quadratic spline [11-12]. Consider a Wi-Fi/Indoor network (Figure 1), which consists of a set of access points AP_i, where i is the number of access points AP_i, $i=\overline{1,3}$. Denoting the coverage area of the Wi-Fi/Indoor network as $\overline{G}$, for which $\overline{G} \subset Q$ (Figure 2), where $Q=[a,a+H]\times[b,b+H]$, with the side $H > 0$ and step $h_N = \frac{H}{N}$, N – natural number, $x_k = a + kh_N$, $y_j = b + jh_N$, $k, j = 0, 1,..., N$. Besides, $\overline{G} = G \cup \partial G$, where ∂G – domain boundary G. Let's divide the Wi-Fi/Indoor network coverage area into segments, then $Q = \bigcup_{k,j=0}^{N-1} Q_{k,j}$, where

$$Q_{k,j} = \left\{ z = x + iy : x \in [x_k, x_{k+1}] \ , y \in [y_j, x_{j+1}] \right\}.$$

We denote such a partition by Δ_N. We define the domain G_N as the union of all segments $Q_{k,j}$ for which $Q_{k,j} \cap G \neq \varnothing$.

Using the finite element method, we consider that at each point of the considered Wi-Fi/Indoor network, the user device UE (User Equipment) will be within the range of at least three reference points RP. To find the user's location, we use the Fingerprinting database, which stores statistical data on measurements of RSSI values at access points AP_i, $i=\overline{1,3}$, which we denote by $f(z_i)$, $i=\overline{1,3}$.

To find the coordinates of the user's location, consider one of the elements of the gap, which is included in the G_N domain with vertices P_1, P_2, P_3 and P_4 for which the following conditions are met (Figure 2):

$$\text{Re } P_1 = \text{Re } P_4, \text{ Re } P_2 = \text{Re } P_3, \text{ Im } P_1 = \text{Im } P_2,$$
$$\text{Im } P_3 = \text{Im } P_4. \qquad (1)$$

Based on the grid domain G_N, we construct a quadratic planar complex spline $S_\Delta(z)$ that interpolates the function $f(z)$ (or its continuation) at the vertices of the rectangles $Q_{k,j}$ included in the domain G_N, setting

$$S_\Delta(z) = a + bz + c\bar{z} + d(z^2 - \bar{z}^2), \qquad (2)$$

if $z \in Q_{k,j} \subset G_N$, where coefficients a, b, c, d are determined from interpolation conditions at points $z_{k,j} = x_k + iy_j$:

$$S_\Delta(z_{k,j}) = f(z_{k,j}), \quad S_\Delta(z_{k+1,j}) = f(z_{k+1,j}),$$
$$S_\Delta(z_{k,j+1}) = f(z_{k,j+1}), \ S_\Delta(z_{k+1,j+1}) = f(z_{k+1,j+1}).$$

Function $S_\Delta(z)$ – continuous in G_N [13].

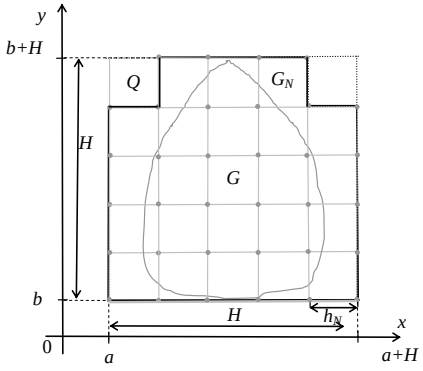


Figure 2. Domain G_N.

Lemma. Let the function $\varphi(z)$ be continuous in the cell domain $\overline{G_N}$, $S_\Delta(z)$ is a quadratic planar spline of the form (2),

$$|f(z) - S_\Delta(z)| \le h\omega(f, h), \qquad (3)$$

where $\omega(f, h)$ – is the modulus of continuity of the function f in G_N, symbol $\prec$ means for $a > 0$ and $b > 0$, $a \prec b$, if $a \le b$, where c – const.

2.2 Finding the User's Location Using Bayesian Method

Consider Bayesian method for determining the user's location in a Wi-Fi/Indoor network. To do this, we use the Fingerprinting DataBase database, which stores statistical data on measurements of RSSI values by access points AP_i, $i=\overline{1,3}$. Then

$$X = \left\langle (x_i, y_i), P_{r_i}[\lambda_i / (x_i, y_i)] \right\rangle, i = \overline{1,3},$$

where (x_i, y_i) – coordinates of the i-th point of the user's location from the Fingerprinting DataBase, $P_{r_i}[\lambda_i / (x_i, y_i)]$ – conditional probability of obtaining measurements of the RSSI signal of the transmitter of the user device with a distribution λ_i at a point with coordinates (x_i, y_i), N – number of access points AP_i, $i=\overline{1,3}$.

Using Bayes formula [14], we determine the posterior probabilities of the events:

$$P(H_i / A) = \frac{P(H_i)P(A/H_i)}{\sum_{k=1}^{n} P(H_k)P(A/H_k)}, \quad (3)$$

where $H_1,\ldots,H_n$ – a complete group of incompatible events for which the following conditions are met $H_i \bigcup H_j = \Omega$, Ω – space of elementary events, $H_i \bigcap H_j = \varnothing$, $i \neq j$, H_1 – an event that determines that the user's location is in the *i*-th element of the split, $P(H_i)$ – a priori probability of finding the user in the *i*-th element of the partition, $P(A/H_i)$ – conditional probability of the user being in the *i*-th element of the partition, $_i$, $P(H_i/A)$ – a posteriori probability of the user being in the *i*-th element of the partition after the experiment.

Then the coordinates of the user's location will be determined according to the expressions:

$$\bar{x} = \frac{1}{k}\sum_{i=1}^{k} \max_{x_i} P(H_i/A), \; \bar{y} = \frac{1}{k}\sum_{i=1}^{k} \max_{y_i} P(H_i/A) \quad (4).$$

2.3 Experiment

To conduct an experiment of positioning UE user devices in a Wi-Fi/Indoor network based on the Fingerprinting method, a fragment of a Wi-Fi/Indoor network in a room is considered, which consists of three access points AP_i, $i = \overline{1,3}$ with coordinates AP_1 (0,0), AP_2 (8,4), AP_3 (4,8). TP Link Tech access point equipment of IEEE 802.11n/ac standard uses dual frequency band 2.4/5 GHz (20/80 MHz), 2452-2472 MHz, 5170-5250 MHz.

During the first stage of the experiment, a Fingerprinting database was formed based on the measurement results of RSSI values obtained using the Open Source Wi-Fi Analyzer application for the considered room at evenly distributed points RP. The placement of RP points ($RP_{11}\ldots RP_{mn}$) is shown in Figure 1, and the RP coordinates and RSSI measurements for the access points AP_i, $i = \overline{1,3}$, which are used in the experiment, are shown in Table 1.

In the experiment, a number of restrictions were adopted that do not take into account the peculiarities of signal propagation inside the premises (reflection from walls and ceilings, diffraction, multipath and interference conditions, etc.).

At the second stage of the experiment for five UEs, measurements of the current RSSI values from three access points are performed AP_i, $i = \overline{1,3}$.

Table 1: Results of measurement of RSSI values.

RP coordinates		RSSI value, dBm		
		AP_1	AP_2	AP_3
RP_{12}	(2,8)	-71	-69	-52
RP_{13}	(4,8)	-68	-62	-42
RP_{22}	(2,6)	-67	-64	-51
RP_{23}	(4,6)	-70	-59	-42
RP_{24}	(6,6)	-74	-50	-45
RP_{31}	(0,4)	-45	-70	-64
RP_{32}	(2,4)	-43	-69	-68
RP_{34}	(6,4)	-74	-42	-74
RP_{42}	(2,2)	-48	-58	-60
RP_{43}	(4,2)	-56	-61	-60
RP_{45}	(8,2)	-70	-49	-72
RP_{54}	(6,0)	-59	-76	-56
RP_{55}	(8,0)	-72	-62	-74

The data of the current measurements of RSSI values of devices for UE users ($RSSI_{AP1}$, $RSSI_{AP2}$, $RSSI_{AP3}$) is shown in Table 2, for example, for the user UE-1 with coordinates (5;5), the RSSI values (-77, -63, -59) dBm are obtained.

Table 2: UE RSSI Values Measurement Results.

UE coordinates		RSSI value, dBm		
		AP_1	AP_2	AP_3
UE-1	(5;5)	-77	-63	-59
UE-2	(2,5;2,2)	-54	-71	-69
UE-3	(7,3;1,2)	-73	-61	-78
UE-4	(3.48;7.62)	-79	-76	-47
UE-5	(1.6;5.2)	-72	-72	-67

2.4 Accuracy of User Location Based on Complex Spline Approximation and Bayesian Method

To determine the accuracy of determining the user's location in the Wi-Fi/Indoor network using a complex spline approximation based on a flat quadratic spline, we use Lemma 1 (Section 2.1) expression (3). The results are summarized in Table 3.

To determine the positioning accuracy of a user device in a Wi-Fi/Indoor network using the Bayesian method, we use the value of the mean absolute error MAE (Mean Absolute Error) [6]:

$$MAE = \left(\frac{(x_{est} - x_{real})}{x_{real}} + \frac{(x_{est} - x_{real})}{x_{real}} \right) \cdot 100\%, \quad (5)$$

where x_{est}, y_{est} – user location coordinates determined during the experiment, x_{real}, y_{real} – real coordinates of the user's location.

The obtained results of positioning accuracy using the probabilistic Bayesian method are shown in Table 3.

Table 3: Comparison of local positioning results using complex spline approximation and Bayesian method.

User, coordinates	Complex spline approximation	Mean absolute error MAE, %
UE-1 (5;5)	(5.15;5.1)	5,1
UE-2 (2,5;2,2)	(2.7;2.31)	9,6
UE-3 (7;1)	(7.2;1.12)	10.0
UE-4 (3;7)	(3.1;7.2)	6,3
UE-5 (1;5)	(1.1;5.18)	8,6
User, coordinates	Bayesian method	Mean absolute error MAE, %
UE-1 (5;5)	(5.32;5.72)	20,8
UE-2 (2,5;2,2)	(2.75;2.25)	12,3
UE-3 (7;1)	(7.3;1.2)	24,1
UE-4 (3;7)	(3.48;7.62)	24,8
UE-5 (1;5)	(1.15;5.2)	19,1

3 CONCLUSIONS

The obtained results allow us to conclude that the greater accuracy of determining the user's coordinates in the Wi-Fi/Indoor network using the Fingerprinting method is achieved using complex spline approximation using flat quadratic splines. Moreover, it could be argued that regardless of the location of the user, the error is 10%. When determining the coordinates of the user's location, the use of the probabilistic Bayesian method gives a positioning error of up to 25%. Therefore, it is advisable to use the Fingerprinting method based on complex approximation using flat quadratic complex splines to find the user's location.

REFERENCES

[1] S.H. Hameedah, M. Hussein, M.S. Saad, and M.A. Mat Dzahir, “An Overview of Local Positioning System: Technologies, Techniques and Applications,” in International Journal of Engineering & Technology, vol. 7(3), 2018, pp. 1-5.

[2] F. Gu, X. Hu, M. Ramezani, and et al., “Indoor Localization Improved by Spatial Context - A Survey,” in ACM Computing Surveys, vol. 52(3):64, 2019, pp. 1-35, doi: 10.1145/3322241.

[3] A. Yassin, Y. Nasser, M. Awad, and et al., “Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications,” in IEEE Communications Surveys & Tutorials, 2016, 99 p., doi: 10.1109/COMST.2016.2632427.

[4] Q. Yang, S. Zheng, and M. Liu, “Research on Wi-Fi indoor positioning in a smart exhibition hall based on received signal strength indication,” in Journal on Wireless Communications and Networking, 2019, doi: 10.1186/s13638-019-1601-3.

[5] A.P. Rahmadini, P. Kristalina, and A. Sudarsono, “Optimization of Fingerprint Indoor Localization System for Multiple, Object Tracking Based on Iterated Weighting Constant - KNN Method,” in International Journal on Advanced Science Engineering Informational Technology, vol. 8, no. 3, 2018.

[6] P. Bráulio H.O.U.V., Horácio A.B.F. de Oliveira, and E. J.P. Souto, “Factor Optimization for the Design of Indoor Positioning Systems Using a Probability-Based Algorithm,” in Journal of Sensor and Actuator Networks, vol. 10, no. 1:16, 2021.

[7] G. Li, E. Geng, Z. Ye, and H. Zhu, “An Indoor Positioning Algorithm Based on RSSI Real-time Correction,” 14th IEEE International Conference on Signal Processing (ICSP), 2018.

[8] S. Gao and S. Prasad, “Employing spatial analysis in in-door positioning and tracking using Wi-Fi access points,” Eighth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, pp. 27-34, 2016.

[9] H. Mehmood, N.K. Tripathi, and T. Tipdecho, “Indoor Positioning System Using Artificial Neural Network,” in Journal of Computer Science, vol. 6 (10), 2010. [Online]. Available: https://doi.org/10.3844/jcssp.2010.1219.1225.

[10] B. Sulaiman, E. Natsheh, and S. Tarapiah, “Towards a better indoor positioning system: A location estimation process using artificial neural networks based on a semi-interpolated database,” in Pervasive and Mobile Computing, vol. 81, 2022, 101548.

[11] I. Strelkovskaya, I. Solovskaya, and J. Strelkovska, “Fingerprinting/Indoor positioning using complex planar splines,” in Journal of Electrical Engineering, vol. 72, no. 6, 2021, pp. 401-406, doi: 10.2478/jee-2021-0057.

[12] I. Strelkovskaya, I. Solovskaya, J. Strelkovska, and T. Grigoryeva, “Using quadratic complex planar splines in solving local positioning problems,” IEEE 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering **(**TCSET**)**, 2022, pp. 602-605, doi: 10.1109/TCSET55632.2022. 9766876.

[13] G. Opfer and M. Puri, “Complex planar splines,” in J. Approxim. Theory, vol. 31, no. 4, pp. 383-402, 1981.

[14] I.I. Gikhman, A.V. Skorokhod, and M.I. Yadrenko, “Probabilistic theory and mathematical statistics,” Kyiv. Publishing Association “Vishcha Shkola", 1979.

Extended Classification Model of Telemedicine Station

Roman Tsarov[1], Iryna Tymchenko[1], Vladyslav Kumysh, Kateryna Shulakova[2,1] and Liliia Bodnar[3]
[1]*State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, Odesa, Ukraine*
[2]*Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*
[3]*South Ukrainian National Pedagogical University, Staroportofrankyvska Str. 26, Odesa, Ukraine*
c4r1980@gmail.com, mentalitett@gmail.com,vlad.kumish@gmail.com, katejojo29@gmail.com, bodnarl79@pdpu.edu.ua

Keywords: Telemedicine, Telemedicine Station, Telemedicine Network, Classification Model, Facet.

Abstract: In the paper the relevance of the scientific problem of creation and development of telemedicine networks and telemedicine stations was proved by a systematic analysis of modern scientific research and regulatory sources. The conducted analysis revealed that there is no unified and system approach in place helping to solve this problem, especially on designing stage and during the technical implementation of these networks and their points. It results in complete or partial incompatibility of existing designs, construction documentation and blueprints as well, which leads to difficulties with various fragments of telemedicine networks and impedes international cooperation in this field. In order to solve this problem the definition of the telemedicine station is given, also its subsystems composition is defined, and an extended telemedicine station classification model is developed in the paper. The proposed classification model is based on the faceted classification method. It allows constructing classes by using any combination of characteristic features of telemedicine stations, while omitting and not using some of the characteristic features. Within the framework of this classification, the intersection of individual classes of telemedicine stations is allowed. A mathematical description of the classification model in the form of a faceted formula is provided.

1 INTRODUCTION

The modern communications industry is characterized by the penetration of information and communication technologies (ICT) in all spheres of human activities. ICT tools have become a key of any industry system. One of the urgent tasks of modern society, is to increase the availability of medical services, as well as their quality [1]. The solution to this problem is the introduction of modern ICT into the health care system and the creation of telemedicine networks on their basis. The main task of the telemedicine network (TMN) is to solve the problems of health care in providing access to specialized and high-tech medical care to anyone, regardless of their location and social status.

TMN allows organizing innovative interaction of various healthcare institutions of different levels with specialized and multidisciplinary national-level healthcare centres using advanced ICT and the intellectual potential of the industry's leading specialists when providing skilled medical care to the population remotely. In regards to COVID 19 pandemic outbreak, aggravation of local military conflicts, war in Ukraine, powerful earthquakes in Turkey and Syria, the point of designing telemedicine centres and networks is more relevant than ever.

The field of telemedicine development is relevant and largely researched, as evidenced by the large number of works [1-9, 11-19]. Modern scientists pay attention to such aspects as:

- architecture and design of telemedicine networks and stations;
- organization of the information component of telemedicine;
- telemedicine implementation and development strategies taking into account national peculiarities;
- implementation of information services and systems for the organization of telemedicine services.

In [2] an approach of creating the telemedicine infrastructure in South Sudan is proposed. The crore of the mentioned approach points to the autonomous medical clinic on the basis of a transport module (container) called Clinic-In-ACan (CIAN), which is ready for use in any location. In [3], the authors proposed the design and specification strategy for using backbone optical communication lines as the

wired terrestrial telecommunication component of the telemedicine network. Though [4] proposes a comprehensive approach to creating a network architecture for a nationwide telemedicine network that connects all regional hospitals and medical centres with the city hospitals and state hospitals. The approach considers the use of a web-based telemedicine system, which provides the basic services for medical teleconsultations. It is obvious that for the modern telemedicine networks and centres, the key network service is video conferencing. The peculiarities of organizing the telemedicine system based on videoconferencing are investigated in [5]. The proposed structure is based on a service-oriented architecture, taking into account the hierarchical relationships inherent in a medical organization. The authors of [6] investigate the use of cloud technologies and Grid-systems in the concept of telemedicine. Namely, they propose to use a computer network in a way of distributed sharing the Internet access to ensure the connection of remote telemedicine stations in rural areas. In [7] the issue of creating a telemedicine network based on a distributed peer-to-peer structure with all telemedicine stations as peers is investigated.

2 CURRENT STATE OF THE PROBLEM AND PAPER GOAL

The analysis of studies has shown that the need to create and to develop the telemedicine networks and telemedicine stations is undeniable, though there are a number of difficulties when implementing such projects, such as [1-8]:

- Lack of a unified regulatory framework governing the activities of telemedicine centers and the procedure for providing telemedicine services. This leads to inconsistency between the options for TMC construction and the requirements in standards, the characteristics of protocols adopted in medical diagnostics and MIS formats.
- Lack of unified approaches in development of technical designs when constructing and building TMNs. The requirements to TMN may be formed by several participants – the medical institution, the regulatory organization (ministry, health department), the project organization itself. It may cause the inconsistency and contradiction of these requirements.
- Incompatibility of designs used for different fragments of the telemedicine network.

In fact, it can be noted that research in telemedicine is not systematic - some tasks and problems are given considerable attention, while others, on the contrary, are little elaborated. All this has led to the fact that today mostly there are no unified approaches and system solutions for developing and creating the technical projects and designs of telemedicine networks and stations. Mostly to this moment telemedicine networks and stations were created separately, without coordination and planning for further integration with other similar networks and stations. As a result a significant number of the implemented projects are incompatible with each other, which in turn affects the development of national telemedicine networks.

As already mentioned, the construction of telemedicine networks and stations is an urgent scientific problem, but the lack of unified and systematic approach towards its solution has led to significant confusion, starting even with the terminology. This causes certain problems during the creation and implementation of international projects and initiatives in the field of telemedicine. For example, the term telemedicine itself has more than 100 different definitions [1, 8, 12-16].

The purpose of this work is to provide a basic definition of the term telemedicine station as the ground element of a telemedicine network and to develop a classification model of telemedicine stations. This classification model is called to become a foundation in order to create standard, unified technical and design solutions for constructing the telemedicine stations and networks.

3 TELEMEDICINE STATION CLASSIFICATION MODEL

In [11] it is stated that telemedicine station and telemedicine center are the basic objects of telemedicine network (Figure 1).

These two terms are fairly widely used. The term telemedicine center is more formalized, although different sources provide somewhat different definitions for this term. For example, [14] defines such concepts as a telemedicine consultation center and a specialized telemedicine center, [12, 13] defines the term telemedicine center, and [15] defines such a concept as a telemedicine resource center. On the

other hand, the term telemedicine station is not as formalized: it is widely used in various studies [10, 17, 18], but there is no official definition of this term. With this in mind, the following definition of the term is proposed:

Telemedicine station (TS) is an element of the telemedicine network that enables the clinical tasks handling, equipped with all the necessary facilities to provide and deliver the telemedicine services at the station.

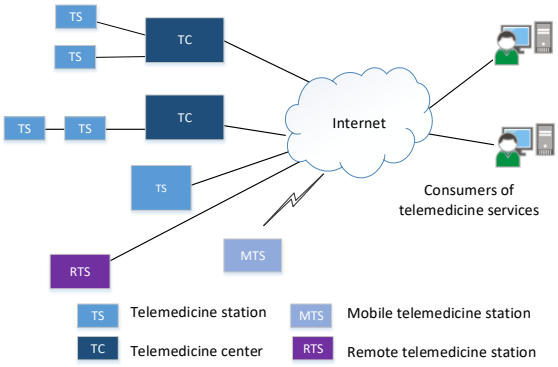


Figure 1: A generalized diagram of telemedicine network organization.

It follows from the proposed definition that telemedicine station should contain the following subsystems, namely:

A) Basic subsystems:

1) Telemedicine (information) subsystem.
2) Medical device subsystem.
3) Telecommunications subsystem.
4) Power supply and electric lighting subsystem.

B) Specialized subsystems:

1) Engineering subsystems (ventilation, water supply, sewerage, air conditioning, heating, etc.).
2) Security subsystems (access control, video surveillance, emergency and fire warning, etc.).
3) Transportation subsystem.

Telemedicine (information) TS subsystem includes hardware and software for automation of workflow and diagnosis, remote examinations, video consultations and videoconferences. Technically these achieved through the collection, input/output, digital processing, intelligent analysis, visualization, classification, storage, archiving, transfer to various media, transmission and provision of medical data and background information on request.

TS medical devices subsystem includes a set of instruments, devices, accessories, equipment, materials and other products used for medical purposes, whose functional purpose is not realized by pharmacological, immunological, genetic or metabolic effects on the human body. Medical devices can be used separately or in combination, as well as together with other accessories necessary for the intended use of these products, including special software.

Telecommunication subsystem of the TS is intended for connecting the medical devices (which have network interfaces) to the TS telemedicine information subsystem, as well as connecting TS itself to telemedicine centers, specialized stationary medical institutions or to external data transmission network (e.g., Internet). Telecommunication subsystem includes hardware and software, as well as communication lines.

Obviously, depending on the clinical tasks the telemedicine station will perform, the services it will provide, and on the external conditions - its characteristics and structure of its subsystems may vary. So, telemedicine stations can be classified in the space of their features, and in general, the entire set of telemedicine stations can be described on the basis of a classification model. Obviously, in this case, there will be no rigid classification structure, because the same telemedicine station may have functional characteristics that correspond to different classes. Moreover, a finite number of classes in the classification system cannot be clearly defined in advance since the new classes may appear in the process of analyzing the characteristics of different TS. With that said, we propose to create the TS classification model based on faceted classification technique [10].

The faceted classification technique enables the creation of a classification system from many independent subsets – facets. Each facet Φ_i contains a number of values taken by this feature or criteria. These values are mutually excluded.

In general, the classification system can be described by the following faceted formula

$$ClS = (\Phi_1, \Phi_2, ..., \Phi_n).$$

Let's consider some set of telemedicine stations - *PT*, over which the U_{PT} set of stations' features is defined:

$$U_{PT} = \{u_{PTi} \mid i = 1, ...k\},$$

here u_{PTi} stand for the elements of U_{PT} set.

Then the facetted set *ClS* can be defined over the *U* set as following:

$$ClS(U) = \{ClS_k \mid k = 1,...t\},$$

here ClS_k denotes the facet, which defines the k^{th} classification feature of the telemedicine station.

Each facet can take the defined set of possible values:

$$\phi^{ClS_k} = \{\phi_{k1},...\phi_{km}\},$$

here *m* denotes the number of possible values of i^{th} facet.

Each element of *PT* set should contain at least one value of the faceted feature:

$$\phi_j(u_{PTi}) = \{\phi_{ks}(u_{PTp}),\ p \in \{1,...k\}\} \subset \phi^{ClS_k}.$$

The set containing all the facetted features can be defined as following:

$$\phi_j(u_{PTi}) = \{\phi_{ks}(u_{PTp}), \forall p : p \in \{1,...k\}, \phi_j(u_{PTi}) \neq \varnothing\}.$$

The overall faceted formula *Ff* to classify the telemedicine station is following:

$$Ff\ (u_{PTi}) = \begin{Bmatrix} [ClS_j : \phi_j(u_{PTi})] \mid ClS_j \in ClS(U), \\ \phi_j(u_{PTi}) \in \phi^{FS_k} \forall j \in \{1,...,k\}, \phi_j(u_{PTi}) \neq \varnothing \end{Bmatrix}. \quad (6)$$

In [18], the following criteria are proposed as classification features (facets):

1) Type of communication TS (connection) with the telemedicine network.
2) Type of energy and power supply organization in TS.
3) Type of provided telemedicine service in TS.
4) Distance to the central or regional nodes of the telemedicine hub.
5) Type of placement.
6) Type of construction.
7) Type of ownership.
8) Subordination level.
9) Degree of mobility.

Further research allows to supplement this list with the following criteria, namely:

- Type of telemedicine services provision mode.
- Type of access mode.
- Integrity level.

Taking into account additional classification criteria and, grounding on the basic classification model of telemedicine stations [18], the following extended classification model can be formed (Table 1).

Table 1: Telemedicine station extended classification model.

Facet (feature/criteria)	Faceted value
Type of connection to TMN	Wired connection
	Wireless connection
Degree of mobility	Stationary
	Portable
	Mobile
Type of telemedicine service provided	Basic (only one type of telemedicine service is provided)
	Diagnostic (several types of diagnostic telemedicine services are provided)
	Universal (any telemedicine services provided)
Type of energy and power supply	"Green" (station is powered using the renewable sources e.g. solar energy)
	Hybrid (power supply designed by the classic scheme, and the backup using the renewable sources)
Distance to the central or regional telemedicine hub	Local
	Outlying
Type of placement	Station of walking accessibility (located in places of mass gathering of people e.g. airports, business centers, etc.)
	Establishment station (located on the territory of some medical institution)
	Separate station (separate institution with all the necessary infrastructure)
Type of construction	Open station
	Cabin
	Cabinet
Subordination level	National
	Oblast/Regional
	Urban
	District
Type of ownership	State
	Municipal
	Private
Type of telemedicine services provision mode	Station of provision of services in automatic mode (monitoring)
	Station of provision of services on demand
Type of access mode	Personal station
	Stations of common use
Integrity level	Distributed station
	Non-distributed station

A clear interaction between all participants of the TS design process can be achieved only if all parties have the same understanding of TS characteristics, which is impossible without the initial TS

classification. After implementing the classification system the further success of the design stage depends on the development of the unified approaches to solving 2 interrelated tasks:

- initial selection of the optimal option to implement a given class of TS;
- subsequent standard design at all stages of the technological process of building and deploying TS infrastructure.

To determine the class of TP based on the proposed model, it is necessary to get answers to some of questions regarding the structure of TP and related factors. This process can be formalized in the form of an algorithm, which is a further step of the research.

4 CONCLUSION

In this paper we define the telemedicine station term, determine the composition of its subsystems, and develop an telemedicine station extended classification. The advantage of the given classification model is the possibility of almost unlimited addition of the number of telemedicine station features and the use of already created classifiers as facets, which allows adding new classes to the model and developing the model. The proposed classification model allows one to determine the class of telemedicine station quickly, which will greatly simplify the task of telemedicine station design and will reduce the time-range required to develop the necessary design documentation.

It is obvious that the proposed classification model is not exhaustive and can be further supplemented and modified. Further research should be aimed at obtaining standard design solutions for constructing the telemedicine stations and networks in conditions of different types of input data.

REFERENCES

[1] World Health Organization, "Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth," 2010. [Online]. Available: https://apps.who.int/iris/handle/10665/44497.

[2] G. Sworo, M. Kam, and E. Juan, "Design of a Telemedicine-based system for Clinic-In-A-Can," in IEEE Global Humanitarian Technology Conference, 2012, pp. 265-270.

[3] M. Krishnan, D. Sheela, and C. Chellamuthu, "Design and dimensioning strategies for telemedicine backbone networks with optical links," in International Conference on Information Communication and Embedded Systems (ICICES), 2013, pp. 780-784.

[4] F. Lemma, M. Denko, and S. Kassegne, "Hierarchical Model Based LAN Architecture & VSAT-Based WAN for a National Telemedicine Network in a Developing Country: Case of Ethiopia," Addis Ababa, Ethiopia: Addis Ababa University, 2011.

[5] Y. Zhai and et al., "Design and application of a telemedicine system jointly driven by videoconferencing and data exchange: practical experience from Henan Province," Telemedicine and e-Health, vol. 26, no. 1, pp. 87-98, 2020.

[6] S. Saravanan, M. AnbuRajan, and A. Venkatraman, "Cloud and Grid Computing Based Telemedicine Using Computer Communication Network through Internet," in Global Trends in Computing and Communication Systems: 4th International Conference, ObCom 2011, Vellore, TN, India, December 9-11, 2011, Proceedings, Part I, Springer Berlin Heidelberg, 2012, pp. 434-446.

[7] S. Saravanan, M. Anbu Rajan, A. Venkatraman, N. Sriraam, and P. Thirusakthi Murugan, "Telemedicine using computer communication network," International Journal of Computer Science and Communication, vol. 2, no. 2, pp. 623-631, 2011.

[8] ITU-D Report on Question 6/2 Impact of telecommunications in health-care and other social services First Study Period, 1995-1998.

[9] R. Wootton, J. Craig, and V. Patterso, "Introduction to Telemedicine," 2nd ed. London: The Royal Society of Medicine Press Ltd, 2006, p. 206.

[10] B. C. Vickery, "Faceted Classification: A Guide to Construction and Use of Special Schemes," Front Cover, Aslib, 1960, p. 70.

[11] "Recommendations on the application of modern technical solutions in the design of e-health systems, including telemedicine networks," [Online]. Available: https://www.itu.int/en/ITU-D/Regional-Presence/CIS/ Documents/RI-WTDC17/ONAT_RI2 Recommenda-tions_Rev2.pdf.

[12] D. Rudel, M. Fisk, and R. Roze, "Definitions of Terms in Telehealth," Slovenian Medical Informatics Association (SIMIA), vol. 16, no. 1, p. 28, 2011.

[13] L. Otto, L. Harst, P. Timpel, B. Wollschlaeger, P. Richter, and H. Schlieter, "Defining and delimitating telemedicine and related terms-an ontology-based classification", Stud Health Technol Inform, vol. 268, pp. 113-122, 2020.

[14] S. Mishra, L. Kapoor, and I. Singh, "Telemedicine in India: current scenario and the future", Telemedicine and e-Health, vol. 15, no. 6, pp. 568-575, 2009.

[15] "Essential Telemedicine Terms Should Know", [Online]. Available: https://www.wheel.com/blog/ essential-teleme-dicine-terms-clinicians-should-know

[16] N. Sodhi, R. Weinstein, K. Stewart, and C. Doarn, "Analysis of Telehealth Versus Telemedicine Terminology in the Telemedicine and e-Health Between 2010 and 2020", Telemedicine Journal and E-health: the Official Journal of the American Telemedicine Association, Decaber 2022; vol. 28, no. 12, pp. 1861-1865, doi: 10.1089/tmj.2022.0073. PMID: 35417253.

[17] "Komplekt prakticheskih dokumentov dlya okazaniya pomoshi gosudarstvam-chlenam MSE v proektirovanii i vnedrenii telemedicinskih punktov, v

tom chisle na osnove vozobnovlyaemyh istochnikov energii," [Online]. Available: https://www.itu.int/en/ITU-D/Regional-Presence/CIS/Documents/ Publications.

[18] R. Tsaryov, I. Tymchenko, and V. Kumysh, "Basic classification model of telemedicine station," // Kolektivna monografiya "Problemy ta perspektyvy telemedytsyny v Ukraini," Kyiv, Vyshcha shkola, 2014, pp. 20-34.

Methodology for Evaluating the Impact of Sudden Failures on the Reliability Parameters of the Optical Cable Damaged Section

Boris Zelentsov[1], Viatcheslav Shuvalov[2], Irina Kvitkova[2] and Dmitry Kachan[3]
[1]*Department of Higher Mathematics, Siberian State University of Telecommunications and Information Science, Kirov Str. 86, Novosibirsk, Russia*
[2]*Department of Infocommunication Systems and Networks, Siberian State University of Telecommunications and Information Science, Kirov Str. 86, Novosibirsk, Russia*
[3]*Department of Electrical, Mechanical and Industrial Engineering, Anhalt University of Applied Sciences, Bernburger Str. 55, Köthen, Germany*
zelentsovb@mail.ru, shvp04@mail.ru, irin.creme@yandex.ru, dmitry.kachan@hs-anhalt.de

Keywords: Elementary Section of Optical Cable, Local and Global Failure, Wear-Out Failure, Sudden Failure, Markov Model, Reliability Measures.

Abstract: An optical cable is considered as a multicomponent system, which components are elementary sections of an optical cable. During operation, due to failures of these sections, they are replaced with new ones. The reason for the failures of elementary sections, which we will call local failures, are sudden failures. In addition to sudden failures, there are gradual failures that are a consequence of the fiber aging. There is a need to replace the optical cable as a whole. Such a failure is wear-out or global. The influence of sudden failures on the reliability measures of an optical cable in the conditions of its degradation is considered. To solve this problem, a Markov reliability model was developed, on the basis of which expressions were obtained for calculating the unavailability factor, the average time of operational and inoperable states during the degradation cycle of an optical cable. Methods for determining the probabilities of transitions from one state of degradation to another are proposed. It is shown that the unavailability factor of the optical cable and the average time of operational state on one degradation cycle increase with increasing intensity of sudden failures. The optical cable reliability model presented in this paper can be useful for predicting the operating time of an optical cable.

1 INTRODUCTION

The process of operation of the optical cable of the access network is accompanied by failures, which are the result of the rupture of the optical fiber. Failures follow at random intervals and are followed by the process of replacing the failed sections with new ones. During operation, the number of such replacements (repairs) is a random variable with the average time between MTBF failures (Mean Time Between Failure) and the average repair time (replacement of the site).

Thus, an optical cable can be considered as a multicomponent system consisting of sequentially connected cable sections (elements) with length l_e (elementary sections). The number of such sections on the length of optical cable is defined as L/l_e, where L is the total length of optical cable. The failure of one of the optical cable elements will call a local failure. The failures of elementary sections will be considered independent. After the failure of the elementary section, this section is repaired, which consists in replacing it with a new one. At the same time, the quality of repair can be considered perfect, i.e. after the repair, the section becomes "as good as new". After a certain number of local failures and repair of failed sections, there is a moment when it is necessary to start replacing the optical cable as a whole, since the cost of its repair becomes higher than the cost of replacement. Replacing the optical cable with a new one is a consequence of a failure, which will call global. This failure characterizes the service life of the optical cable and, of course, this period differs from the warranty period, which corresponds to about 25 years.

Global failure is a consequence of fiber degradation, i.e. it is a wear-out failure. This failure is caused by the natural processes of aging, wear,

corrosion and fatigue in compliance with all the rules and (or) standards of design, manufacture and operation [1]. Considering the wear-out failure as an optical fiber breakage, i.e. from a mechanical point, the strength of the optical fiber depends mainly on the appearance of microcracks on its surface [2].

Since the change in the depth of the crack is a non-decreasing function depending on the applied load, this process can be represented as a Markov process of pure reproduction or, if we consider the decrease in the strength of the optical fiber, as a process of pure doom.

Figure 1 shows the states of damage accumulation that occur due to the growth of the crack depth a. Let the first state be the initial one. It corresponds to the start time from the beginning of the operation of the optical fiber. The probability of finding a microcrack in the initial state is assumed to be 1. The state b is absorbing and at $t\to\infty$ the probability of finding a microcrack in the state b tends to 1.

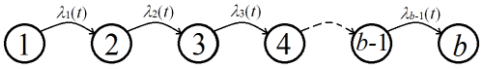


Figure 1: This caption has one line so it is centered.

The use of Markov processes for modeling the degradation phenomenon was first proposed and justified in [3-7]. The approach proposed in these works has found wide application. As an example, there are the works [8-10], where the theory of Markov processes is effectively used to describe the processes of systems.

Thus, in [8] the problem of maintenance and repair is considered from the position of degradation processes management. In [9], an optimal strategy for maintenance and repair is proposed in the conditions of degradation described by the Markov process of pure doom. Researches of degradation of thin-walled pipeline systems with defects are presented in [10].

In this article, unlike the works in which the processes of maintenance and repair in conditions of degradation are considered in general terms, the task is to obtain formulas for assessing the impact of sudden failures on reliability parameters in conditions of degradation of fiber.

2 CONCEPTUAL MODEL OF THE PROCESS

Over time, the optical cable degrades. At the same time, transitions between states occur. The state of degradation is determined by the depth of the microcrack. The depth of the microcrack varies from state to state according to the linear law: $a_j = a_0+j\cdot\Delta a$, where Δa is the increment of the depth of the microcrack during the transition from one state to another. In a certain limiting state the depth of the microcrack reaches a value at which a wear-out failure occurs.

After the onset of wear-out failure, the optical cable section is restored by replacing it.

The operating time of an optical fiber is divided into the same time intervals, which is called the operating intervals. At each operating interval a sudden failure may occur, which is detected at the moment of its occurrence, that is, the failure is explisit. After such a failure is detected, the optical cable section is restored by replacing it.

When forming the model, the following conditions are accepted.

1) The operating time of the optical fiber is divided into the same operating intervals T.
2) Each operating intervals is characterized by a state of degradation, which is determined by the depth of the microcrack.
3) When the depth of the microcrack increases by a certain amount, the transition to the next state of degradation occurs.
4) Sudden failure during the operation interval occurs with the same intensity and does not depend on the state of degradation.
5) The initial state of the degradation cycle is a crack-free state.

3 MATHEMATICAL MODEL OF THE PROCESS

The mathematical model is based on the theory of Markov and semi-Markov processes. Under the accepted conditions, a state-transitions diagram can be drawn up. To describe the approach to model development, the number of degradation states is assumed to be five, that is, the crack length corresponding to the wear-out failure is divided into five parts and each part is a characteristic of the degradation state. In real conditions, the number of states taken into account can reach several tens or even hundreds.

The degradation cycle is the duration of operation from the initial state to the inoperable state, which is the result of a wear-out failure. The degradation cycle includes being in states of degradation and in states of recovery after sudden and wear-out failures.

The state-transitions diagram is shown on Figure 2.

There are following signs on the diagram Figure 2: D_j − the state of degradation; p – the probability of transition to the next state of degradation at one operating interval; λ − sudden failure rate; R – restoration of the optical fiber section up state.

In the given model, transitions between states occur both in discrete time ($D_j \rightarrow D_{j+1}$) and in continuous time ($D_j \rightarrow R$). The dotted arrow indicates the last transition to recovery in a state of wear-out failure. These transitions can be described together using a semi-Markov process. Therefore, the main model is based on the theory of semi-Markov processes.

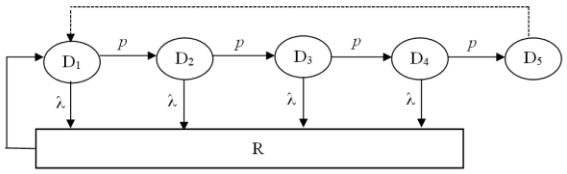


Figure 2: The state-transitions diagram of one section with transitions in discrete and continuous time.

In the states D1, D2, D3, D4, there is an increase in the depth of the crack in the optical fiber, and in the state D5 there is a wear-out failure. After replacing the optical cable due to wear-out failure, the degradation cycle ends. The initial state of the degradation process is D1.

Let's introduce the probabilities of transitions between the semi-Markov process states, which in [11] are called probabilities of passage: $\hat{p}_{ij}$ is the probability that a transition to the *j*-th state will occur, provided that there is an exit from the *i*-th state.

In continuous time, the transition to the recovery state occurs during the operation interval. The probability of this transition is $q = 1 - \exp(-\lambda T)$. For sufficiently small values of $\lambda \cdot T$ ($\lambda \cdot T << 1$), $q \approx \lambda T$ can be assumed.

The probabilities of passage are determined by the initial probabilities p and q, which do not depend on the state of degradation:

$$\hat{p} = \frac{p}{p+q}, \quad \hat{q} = \frac{q}{p+q}. \tag{1}$$

The state-transitions diagram of the semi-Markov process is shown on Figure 3.

The diagram Figure 3 shows: $\hat{p}$ is the probability of transition from one state of degradation to another; $\hat{q}$ is the probability of sudden failure during the operation interval. $\hat{p}$ and $\hat{q}$ are defined from (1).

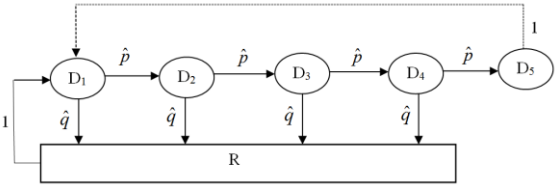


Figure 3: The state-transitions diagram of the semi-Markov process.

Matrix of transition probabilities of a semi-Markov process is:

$$\hat{P} = \begin{pmatrix} 0 & \hat{p} & 0 & 0 & 0 & \hat{q} \\ 0 & 0 & \hat{p} & 0 & 0 & \hat{q} \\ 0 & 0 & 0 & \hat{p} & 0 & \hat{q} \\ 0 & 0 & 0 & 0 & \hat{p} & \hat{q} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Let's form a subset of states where transitions occur on one degradation cycle. Let U = {D_1, D_2, D_3, D_4, R} be a subset of states in which there is no wear-out failure. Then the matrix of transition probabilities on the subset U is:

$$\hat{P}_{UU} = \begin{pmatrix} 0 & \hat{p} & 0 & 0 & \hat{q} \\ 0 & 0 & \hat{p} & 0 & \hat{q} \\ 0 & 0 & 0 & \hat{p} & \hat{q} \\ 0 & 0 & 0 & 0 & \hat{q} \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2}$$

The matrix (2) is used to calculate the matrix of relative frequencies of the subset U [11]:

$$\hat{N}_U = \left(\hat{n}_{ij}\right) = \left(E - \hat{P}_{UU}\right)^{-1}, \tag{3}$$

where $\hat{n}_{ij}$ is the average number of hits in the *j*-th state during one degradation period at the initial *i*-th state; E is a unit matrix.

Since the state D_1 is assumed to be initial, all information about the degradation process is contained in the first row of the matrix $\hat{N}_U$ from (3):

$$\hat{n}_1 = \left(\frac{1}{\hat{p}^4} \quad \frac{1}{\hat{p}^3} \quad \frac{1}{\hat{p}^2} \quad \frac{1}{\hat{p}} \quad \frac{1-\hat{p}^4}{\hat{p}^4} \right). \tag{4}$$

To determine the duration of being in states, convert string (4) into a string whose elements are the durations of states in units of time. After entering the states D1, D2, D3, D4, the process is in each of

them for an average of $1/(p + q)$ operating intervals. The average time spent in the state R after entering it is $1/\mu$, where μ is the recovery rate.

After making simple transformations of (4), get a string of average durations of states:

$$\overline{t_1} = T \cdot \left(\frac{(p+q)^3}{p^4} \quad \frac{(p+q)^2}{p^3} \quad \frac{p+q}{p^2} \quad \frac{1}{p} \quad \frac{(p+q)^4}{\mu \cdot T \cdot p^4} \right). \quad (5)$$

The average up state duration t_u is obtained by summing the elements of row (5), which correspond to the states D1, D2, D3, D4. The average down state duration t_d is equal to the element of row (5), which corresponds to the state R:

$$t_u = \frac{(p+q)^4 - p^4}{q \cdot p^4} \cdot T, \quad t_d = \frac{(p+q)^4}{\mu \cdot p^4}. \quad (6)$$

The average degradation cycle duration of an optical cable elementary section is determined by the sum of t_u and t_d from (6):

$$t_{dc} = \frac{T}{p^4} \cdot \left[\frac{(p+q)^4 - p^4}{q} + \frac{(p+q)^4}{\mu \cdot T} \right].$$

The obtained results can be generalized to the degradation process with an arbitrary number of states n. Formulas for calculating reliability measures for one degradation cycle are aggregated in Table 1.

Table 1: The expressions for calculating reliability measures.

Reliability measures	Expression for calculating
Average up state time per degradation cycle	$t_u = \frac{(p+q)^{n-1} - p^{n-1}}{q \cdot p^{n-1}} \cdot T$
Average down state time per degradation cycle	$t_d = \frac{(p+q)^{n-1}}{\mu \cdot p^{n-1}}$
Average degradation cycle duration	$t_{dc} = t_u + t_d$
Availability factor per degradation cycle	$F_{av} = \frac{t_u}{t_u + t_d}$
Unavailability factor per degradation cycle	$F_{un} = \frac{t_d}{t_u + t_d}$
Average wear-out failure frequency	$\omega_{dc} = \frac{1}{t_{dc}}$
Average number of sudden failures per degradation cycle	$n_f = \frac{(p+q)^{n-1} - p^{n-1}}{p^{n-1}}$
Average sudden failures frequency	$\omega_f = \frac{n_f}{t_{dc}}$

4 METHODOLOGY FOR CALCULATING RELIABILITY MEASURES

To carry out calculations, it is necessary to set the value of p and in the future for the selected value of p, changing the probability of sudden failures q, to assess the impact of sudden failures on the reliability measures of a degraded optical cable. The probabilities of transitions from state to state as a result of the optical cable degradation actually determine the degradation cycle time.

The option is considered when the cause of failures are sudden and wear-out failures. Sudden failures follow with rate λ. For a given value λ, find the probability q. The reason for the wear-out failure is the presence of a microcrack, the depth of which only increases over time.

There are three stages of the development of microcracks. The first stage is the stage of microcrack nucleation; the second is the stage of microcrack depth growth; the third stage lasts for seconds and ends with the rupture of the optical fiber. The second stage lasts for years, and it mainly determines the optical cable service life.

To describe the second stage of microcrack development under the influence of cyclic loading, Paris and Erdogan [12] proposed an equation called "Paris Law":

$$\frac{da}{dN} = C \cdot \Delta K^m, \quad (7)$$

where $\frac{da}{dN}$ is the rate of microcrack growth, [mm/cycle]; C is a constant determined by the material and depending on the frequency of the applied voltage change, $\left[\text{mm} / \left(\text{MPa}\sqrt{m} \right)^m \right]$; m is a constant determined by the material; ΔK is a stress intensity factor, $\left[\text{MPa}\sqrt{m} \right]$.

The range of the stress intensity factor can be calculated based on the values of the maximum (K_{max}) and minimum (K_{min}) stress intensity for the load cycle: $\Delta K = K_{max} - K_{min}$.

Using (7), it is possible to obtain an expression for the rate of decrease in the strength of an optical fiber depending on the depth of the crack [13].

By setting the value of the minimum required strength of the fiber, it is possible to determine the service life of the fiber at a known stress value. It

should be noted that if determining the stress applied to the fiber in the future does not cause much difficulty, then to determine the strength of the fiber, it is necessary to know the depth of the crack on the surface of the fiber. Determining the crack depth, in relation to optical fiber, seems to be quite a difficult task due to the lack of non-destructive methods for controlling the strength of optical fiber. Therefore, the prediction of the service life of an optical cable is usually carried out using the results of fiber testing for strength. So, in [14], an expression is proposed to determine the probability of fiber failure:

$$F = 1 - \exp\left[\frac{-L \cdot m_S \cdot B \cdot S_t^{n_g - 2}}{L_p \cdot (n_g - 2) \cdot \sigma_a^{n_g} \cdot t_a}\right], \qquad (8)$$

where F is the probability of a wear-out failure; L is the length of the elementary fiber section (equal to L_p); L_p is the length of the fiber that has been tested for rupture; σ_a is the load applied during degradation; m_s is the parameter of the Weibull distribution (equal to 2 - 5); B, n_g are the parameters characterizing the strength of fused quartz glass (n_g = 15-20); t_a is the γ- percentage service life, defined as the calendar duration from the start of operation of the optical cable, during which it will not reach the limit state with a given probability (1-F); S_t – is the fiber strength.

5 EXPERIMANTAL RESULTS

For certain values of the parameters included in (8), the service life of the optical cable t_a can be found for F. Let, as follows from [15], this service life is 30 years. Let's divide this service life into intervals of 6 years and calculate the dependence of $t_u(p)$ at $q \to 0$. Find the value of p^* at which $t_u \approx t_a$ and, finally, investigate the dependence of the reliability measures given in Table 1 from q.

The uptime probability (1-F) during the service life is chosen sufficiently large. So, according to [16], the γ-percent service life is chosen to be 0.95.

As a result of calculations, the value $p^* = 0.95$ was obtained.

Figure 4 shows the dependence of the average up state time on the degradation cycle from the probability of the optical cable transition to the degradation state $t_u(p)$ at the values of the sudden failures probability $q = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$.

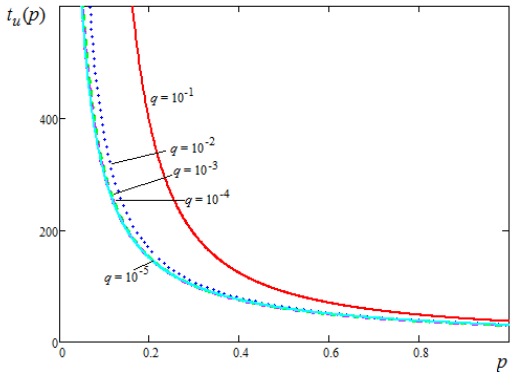


Figure 4: Dependence $t_u(p)$ at different values of the sudden failures probability.

Figure 5 shows the dependence of the unavailability factor $F_{un}(q)$ from the sudden failures probability at $q \in [10^{-6}; 10^{-1}]$. Figure 5 is constructed for the recovery rate μ = 1/3 of an hour [17].

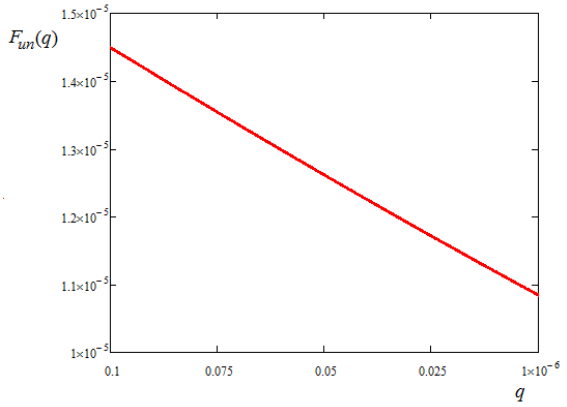


Figure 5: Dependence $F_{un}(q)$.

6 CONCLUSIONS

The average operating time for one degradation cycle increases with growth q due to an increase in the intensity of replacement of optical cable sections. At the same time, the unavailability factor of the optical cable increases with the growth q due to the increase in the average time of down state on one degradation cycle.

The above model allows calculating a number of reliability measures based on the initial characteristics, which are the probability of transition to the next state of degradation, the sudden failure rate, the operation interval duration, the recovery rate after sudden failures.

Based on the approach presented in the article, a methodology for collecting and processing statistical information about the reliability of an optical cable can be developed. During the observation, the following statistical parameters can be recorded: the number of wear-out failures, the number of sudden failures, and the recovery time after failures. According to these parameters, the initial characteristics of the degradation process can be calculated.

The data of the technical condition monitoring system can serve as a source of reliable statistical information.

REFERENCES

[1] GOST R 27.102-2021, “Nadezhnost' v tekhnike: nadezhnost' ob’ekta. Terminy i opredeleniya,” Rossijskij institut standartizacii, Moscow, Russia, 2021, 36 p.

[2] I.A. Ovchinnikova, “Issledovanie i razrabotka opticheskih kabelej special'nogo naznacheniya,” Dissertaciya na soiskanie uchenoj stepeni doktora tekhnicheskih nauk, VNIIKP, Moscow, Russia, 2021.

[3] J.L. Bogdanoff, “A new cumulative damage model, part 1,” Journal of Applied Mechanics, vol. 45, pp. 246-250, 1978.

[4] J.L. Bogdanoff, “A new cumulative damage model, part 3," Journal of Applied Mechanics, vol. 45, pp. 733-739, 1978.

[5] J.L. Bogdanoff and W. Krieger, “A new cumulative damage model, part 2,” Journal of Applied Mechanics, vol. 45, pp. 251-257, 1978.

[6] J.L. Bogdanoff and F. Kozin, “A new cumulative damage model, part 2,” Journal of Applied Mechanics, vol. 47, pp. 40-44, 1980.

[7] J.L. Bogdanoff and F. Kozin, “On stationary cumulative damage models, part 2,” Journal of Applied Mechanics, vol. 49, pp. 39-42, 1982.

[8] V.A. Kopnov, “Optimal'noe upravlenie processami degradacii elementov mekhanicheskih sistem: monografiya,” RGPPU, Ekaterinburg, Russia, 2017, 305 p.

[9] S-H. Sheu, Y-L. Chen, C-C. Chang, and G.Z. Zhang, “Extended optimal replacement policy for systems subject to non-homogeneous pure birth shocks,” Computer and Industrial Engineering, vol. 64, no. 2, pp. 573-579, 2013.

[10] S.A. Timashev and A.V. Bushinskaya, “Ocenka entropii truboprovodnyh sistem s korrozionnymi defektami na osnove markovskogo processa,” Izvestiya Samarskogo nauchnogo centra Rossijskoj akademii nauk, Russia, vol. 15, no. 4, pp. 159-164, 2013.

[11] B.P. Zelencov, “Matrichnye metody modelirovaniya odnorodnyh markovskih processov,” Palmarium Academic Publishing, 2017, 133 p.

[12] P.C. Paris and F. Erdogan, “A critical analysis of crack propagation laws,” Journal of Basic Engineering, vol. 8, pp. 528-533, 1963.

[13] Y. Yang, “Reliability assessment of optical fibers under tension and bending loads,” Graduate School of the University of Maryland, 2003.

[14] Y. Mitsunaga, Y. Katsuyama, H. Kobayashi, and Y. Ishida, “Failure prediction for long length optical fiber based on proof testing,” Journal of Applied Physics, vol. 53, no. 7, pp. 4847-4853, 1982.

[15] V. Burdin, V. Andreev, A. Bourdine, M. Dashkov, and A. Nizhgorodov, “Reliability and lifetime estimations for field-aged optical cable,” in Proceedings of 20th International Conference, NEW2AN 2020 and 13th Conference, ruSMART 2020 St. Petersburg, Russia, August 26-28, 2020, part II, pp. 364-372.

[16] GOST R 52266-2020, Kabeli opticheskie. Obshchie tekhnicheskie usloviya, Moscow, Russia, Standartinform, 2020, 65 p.

[17] E.B. Alekseev, V.N. Gordienko, and V.V. Kruhmalev, Proektirovanie i tekhnicheskaya ekspluataciya cifrovyh telekommunikacionnyh setej. Moscow, Russia, Goryachaya liniya - Telekom, 2008, 392 p.

Monitoring in the Physical Channels of Optical Access Networks

Igor Bogachkov[1], Nicolay Gorlov[2], Tatiana Monastyrskaya[3] and Natalya Medvedeva[4]
[1]*Communications and Information Security Department, Omsk State Technical University,* Mira Str. 11, *Omsk, Russia*
[2]*Communication Lines Department, Siberian State University of Telecommunications and Computer Science, Kirov Str. 86, Novosibirsk, Russia*
[3]*Department of Sociology, Siberian State University of Telecommunications and Computer Science, Kirov Str. 86, Novosibirsk, Russia*
[4]*Department of Foreign Languages for Technical Faculties, Novosibirsk State Technical University, Karl Marx Avenue 20, Novosibirsk, Russia*
bogachkov@mail.ru, gorlovnik@yandex.ru, t.monastyrskaya@mail.ru, n.medvedeva@corp.nstu.ru

Keywords: Monitoring, Passive Optical Network, Physical Stress, Brillouin Scattering, End-Reflection Technique.

Abstract: This research concerns the issues of monitoring in the physical channels optical access networks. This is relevant because automated diagnostic and monitoring systems in passive optical networks (PONs) are not available at the moment. The authors discuss the technique for monitoring physical channels of the tree topology in PONs by Brillouin optical time domain analysis (BOTDA) using end reflections. The paper analyses the main factors affecting the transmission and mechanical characteristics of the optical fiber, presents the results of the comparative analysis of optical fiber deformation measurement methods, and justifies the application of the technique based on the principles of Mandelstam-Brillouin scattering. Also, it provides the theoretical basis and the calculated relations, which allow us to evaluate the main functional capabilities of BOTDA reflectometry. The feasibility of monitoring the physical channels of the tree topology of the optical access network by Brillouin analysis in the time domain using end reflections is evaluated.

1 INTRODUCTION

To date the number of users of fiber to the home (FTTH) services is growing rapidly, and telecom operators are striving to reduce the maintenance costs related to a huge amount of optical equipment. To provide FTTH services in an optical access network, they commonly employ a passive optical network (PON) which decreases the requirements for the optical cable and the provider's maintenance equipment. However, there is a problem of diagnostics due to remote monitoring of PON parameters.

Conventional Optical Time Domain Reflectometer (OTDR) has already found its practical application as a remote testing technology. Nevertheless, when an optical splitter is installed in the access network, backscattering signals from the branched fibers behind the splitter overlap on the backbone fiber. Therefore, it is problematic to identify a malfunction in a single branch of a tree topology with several splitters and subscriber nodes from the central office using OTDR. In this regard, the reduction of maintenance operations carried out by remote OTDR testing remains limited. On the other hand, most commonly used FTTH networks use PONs with an external splitter. So, this substantiates the urgency to develop remote testing technology outside of the splitter for the needs of the telecommunications industry. This work aims to increase the scalability and reliability of fault detection technique in optical fiber, and make it compatible with a branched topology. In addition, we apply the basic principle to communication networks and refine the performance of monitoring systems in the mode of remote testing of PON subscribers' facilities. This paper is the first to describe the distributed optical sensing technology for a branched topology, designed to improve the scalability of connection and reliability of optical sensor systems. When conducting probing in an extended network, monitoring can be continued using other branches, even if one of them is broken. In addition, by adding optical splitters to the network, one can easily expand the detection area at any time.

2 BASIC REQUIREMENTS FOR THE MONITORING SYSTEMS

Telecom operators impose the following requirements for monitoring systems [1]:

- the ability to detect and locate faults in each branch of the distributed network;
- the ability to automatically collect information about the parameters of the physical environment, there is no need for technical specialists to visit the site and the user of telecommunication services is not involved;
- the demarcation function which allows the operator to distinguish between their responsibility and the responsibility of the client;
- no active components in the field between the central office and the client's facilities;
- system's reliability which is ensured by the system's ability to perform its necessary functions for a certain period of time;
- scalability as the ability of the monitoring method to smoothly process changes in the net-work infrastructure;
- the monitoring technique should be applicable to available deployed networks without changing the network infrastructure;
- cascading of remote nodes should not represent an obstacle for the monitoring system;
- notification time defined as the time between the occurrence of a fault and its detection should be as short as possible;
- the cost of the monitoring system, which is a critical characteristic for any service operator;
- complexity of feasibility and practical implementation;
- independence from the client as it facilitates service and increases clients' satisfaction.

Monitoring systems based on the principles of Mandelstam - Brillouin scattering meet the above requirements to a greater extent.

3 FUNDAMENTALS OF BRILLOUIN OPTICAL TIME DOMAIN ANALYSIS (BOTDA)

An optical measurement method using Brillouin scattering is a method that uses the linear dependence of the Brillouin frequency shift (BFS) on temperature change or deformation. To determine BFS, the peak frequency of the Brillouin gain spectrum (BGS), observed with the help of the fit function, is used. Brillouin scattering is the inelastic scattering of photons by acoustic phonons. The Brillouin frequency shift is determined by (1)

$$\nu_{\text{в}} = \frac{2nV_A}{\lambda_u}, \quad (1)$$

where V_B is the Brillouin frequency shift; n is the refractive index of the core material; V_A is acoustic wave velocity; λu is the wavelength of the pump light.

For an optical fiber made of silica (refractive index n = 1.45), at the speed of sound V_A = 5.96 km/s, the frequency shift caused by Brillouin scattering in the wavelength band 1.55 microns is approximately 11.1 GHz [2].

Brillouin measurement methods track the linear increase in BFS that occurs with an increase in temperature [3] or tensile deformation [4]. This dependence is due to the dependence of the acoustic velocity in the optical fiber on temperature and deformation.

Equations (2) and (3) express the dependence of BFS on temperature and tensile strain, respectively.

$$\nu_{\text{в}}(t) = \nu_{\text{в}}(t)[1 + C_t(t - t_r)] \quad (2)$$

$$\nu_{\text{в}}(\varepsilon) = \nu_{\text{в}}(0)[1 + C_s\varepsilon] \quad (3)$$

Here t is the temperature, t_r is the initial temperature, and ε is the tensile strain.

C_t and C_s are linear coefficients for temperature and strain changes of 1.10 MHz/°C and 0.0483 MHz/με, respectively, at a wavelength of 1553.8 nm [5].

Since temperature and strain changes cannot be measured separately, the change in BFS $\delta\nu_B$ is sometimes expressed by (4)

$$\delta\nu_{\text{в}}(t, \varepsilon) = C_t\delta t + C_\varepsilon\delta\varepsilon. \quad (4)$$

Here, δt and $\delta\varepsilon$ represent temperature changes and deformations, respectively. In practice, a reference optical fiber or reference data is prepared to compensate for changes in temperature or deformation that are not intended for measurement.

To determine BFS, a fit is applied to the resulting BGS. BGS has a form that can be approximated by the Lorentz function with a central frequency ν_B and a spectrum width at 0.5 relative to the maximum $\Delta\nu_B$. In the case of a quadrature fit, the measurement error of BFS can be estimated using (5) [6].

$$\sigma_\nu(z) = \frac{1}{SNR(z)}\sqrt{\frac{3}{4}\delta * \Delta\nu_{\text{в}}} \quad (5)$$

Here *SNR(z)* is the signal-to-noise ratio (SNR) of the BGS peak in the z coordinate, and δ is the frequency sampling step ($\delta << \Delta\nu_B$).

Equation (5) shows that the detectable minimum change in BFS decreases inversely with the SNR of the measured BGS distribution and increases as the square root of the frequency sampling step δ.

Brillouin Optical Time Domain Analysis (BOTDA) measures the intensity of Brillouin scattering as a function of time. When the photodetector response rate is high enough, the spatial resolution corresponds to the width of the pump light pulse.

The spatial resolution δz is expressed by (6) [7]

$$\delta z = \frac{\nu W_u}{2} \tag{6}$$

Here ν is the group speed of light in the optical fiber, and W_u is the pulse width of the pump light. The spatial resolution of a standard BOTDA is limited to 1 m.

This is due to the fact that the phonon lifetime is about 10 ns, and it is difficult to implement a pump pulse width of less than 1 m.

To measure BFS along a sensitive fiber, it is necessary to obtain the BGS distribution as a frequency dependence of the Brillouin gain intensity.

Figure 1 shows the diagram of a typical BOTDA system.

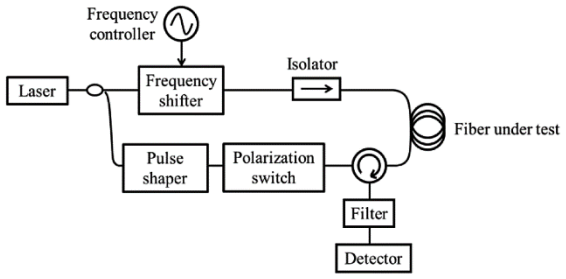


Figure 1: Diagram of a typical system of BOTDA.

The laser output signal is separated by a connector to form a pump light and a probe light. As a rule, the pump light is a pulsed light, and the probing light is launched from the opposite side using a continuous beam. The Brillouin interaction occurs during the collision of test beams in the tested fiber.

To generate a Brillouin interaction between the test beams, the frequency of the probing light is lowered using an optical frequency converter BFS (approximately 11 GHz at a wavelength of 1.55 microns). The pump light is generated by a pulse generator, such as an electro-optical modulator, an acousto-optical modulator, or a semiconductor optical amplifier.

To eliminate the polarization dependence of the Brillouin gain, a polarizing scrambler or polarization switch is turned on in the optical path of the pumping or probing light. Rayleigh scattering of pump light, which is an unnecessary component of the signal, can be eliminated by using an optical filter with a narrow line width, such as Bragg fiber lattice (FBG). The distribution of the gain spectrum by Brillouin can be obtained using the abovementioned measuring system.

4 FUNDAMENTALS OF END REFLECTIONS ASSISTED BRILLOUIN ANALYSIS (ERA-BA)

There has recently been proposed End-reflection-assisted Brillouin analysis (ERA-BA), which is able to measure the individual characteristics of branched fibers based on BOTDA [9]. ERA-BA measures the characteristics of branched fiber by analyzing the Brillouin gain caused by the collision between the pump pulse and the probing pulse reflected from the far end of the split. The ERA-BA branched fiber measurement technique allows remote testing of all optical access networks that include branched fibers, and significantly reduces the number of *in situ* operations. We can expect this technique to provide high-quality maintenance of extremely large network facilities at low cost and with a small workforce.

Figure 2 is a schematic illustration showing the principle of ERA-BA operation [8].

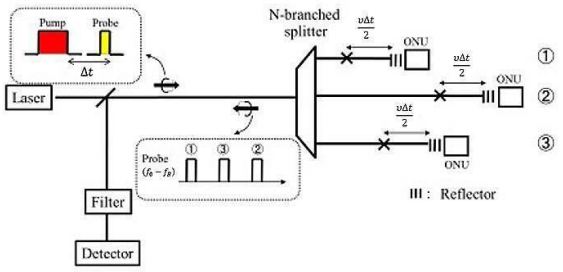


Figure 2: Schematic illustration of the ERA-BA principle.

ERA-BA is an optical measurement technique that measures the characteristics of each branched fiber from the backbone fiber side. Here, the characteristics are optical losses, as well as temperature changes and deformations. The characteristics of the branched fiber are measured using Brillouin gain analysis using a collision between a pump pulse and a probing pulse returning from the far end of the branch. The dependence of the Brillouin gain impact on the pump power is useful for measuring losses in

branches. Meanwhile, BFS which varies with temperature and strain changes is measured to identify an optical fiber with a branched fiber topology.

To provoke a collision of test pulses in the branched fibers, we need an optical filter that reflects the test pulse at the far end of the branched fiber. The pump pulse is triggered after a probing pulse with a time delay Δt for collision at the position $v\Delta t/2$ from the far end. The distribution can be measured by sequentially changing the time delay Δt. To generate long–range reflection in the available access network efficiently, we can use a test light cut-off filter, as recommended in ITU-T L.66 [10].

On the other hand, for probing with a branched fiber, we can use an FBG filter or a full reflection filter designed for the wavelength of the probe.

The difference in the time of the return of the flyback probing pulses caused by the different lengths of the branched fibers is used to determine the signals from the branched fiber. To identify the branched fibers, we at least should know the difference in the branch length δL, expressed by (7). This is referred to as branch identification resolution.

$$\delta L = \frac{vW_r}{2}, \tag{7}$$

where W_r is the width of the probing pulse.

For telecommunication networks, since the branched fiber lengths randomly differ by several meters, the target performance of fiber identification resolution should be equal to or greater than this value.

For branched fibers identification, the difference in the lengths of the branches can be arbitrarily set using an additional delay fiber. The spatial resolution can be determined by (7) using the pump pulse width, as in the conventional BOTDA.

To describe the configuration, let the lengths of the branches with respect to the shortest branch be $0<t_1<t_2, –, <t_{N-1}$. And the t_i values are written in terms of the roundtrip times to the end of each branch. The total branched sensing range is assumed to be T.

The technology proposed here uses a series of probe pulses and a single pump pulse instead of a single pair of probe and pump pulses. The probe beam consists of a series of pulses with a pulse width of $\Delta tpro$, a repetition interval of ΔT, and a train length that nearly equals the entire sensing range, T. A single pump pulse with a Brillouin frequency shift (BFS) accompanies the probe pulse train with an interval of Δt after the last probe pulse. The probe pulses are reflected by the end reflectors in each branch, and the m-th probe pulse collides with the pump pulse and acquires the Brillouin gain at a distance of $vg(\Delta t+m\Delta T)/2$ from the end reflectors in every branch, where vg is the light velocity and m is an integer (0). The entire range, T, can be analyzed by scanning Δt from 0 to ΔT (instead of scanning the entire range, T).

To determine the Brillouin gains yielded in the branches individually, the probe pulses, which are recombined at the splitter, must not overlap each other. In other words, the probe pulses from different branches must be located in different time slots. Probing pulses from the branches should be in different time intervals. We can implement this by changing the optical length of the fiber. To increase the data collection rate, we propose a fiber sensor design that employs different BFS fibers for the feeder and branched fibers. Since no Brillouin interaction occurs on the feeder fiber, the test beam repetition rate (data collection rate) is decided solely by the roundtrip time of the longest branched fiber (excluding the roundtrip time of the feeder fiber). Additionally, in terms of pump depletion, the effect does not occur in feeder fiber. To benefit from the use of different BFS fibers in the sensing fiber topology, the difference between the BFSs of the feeder and branched fibers should be designed to be sufficiently wider than the expected BFS change caused by temperature or strain.

5 CONCLUSIONS

Operators of telecommunication services still do not have at their disposal automated diagnostic and monitoring systems in passive optical networks, which would improve the maintenance efficiency, increase operational reliability and safety of company's optical networks. Nevertheless, the analyzed technique for monitoring physical channels of the tree topology in a passive optical network by BOTDA using end reflections sufficiently meets the criteria for fiber optic cable maintenance to test the fiber of access networks ITU–T L.66, (2007) in service. Here we have proposed a new measurement technique that provides high-speed monitoring with greatly simplified equipment settings. The possibility of practical application of this technique for in telecommunications is proved. It can be implemented as a loss distribution measurement technology for branched fiber PONs without making any changes to the network configuration. According to the authors, further development in the field of fiber optic sounding would be the study of limiting functionality, the metrological substantiation of technical requirements for the main components and the creation of a metrological support system for monitoring systems.

REFERENCES

[1] M. Rad, K. Fouli, H. Fathallah, L. Rusch, and M. Maier, "Passive Optical Network Monitoring: Challenges and Requirements," IEEE Communications Magazine, vol. 2, no. 49, pp. S45-S52, Feb. 2011.

[2] G.P. Agrawal, "Nonlinear Fiber Optics," 4th Edition, Academic Press, 529 p., 2007.

[3] T. Kurashima, T. Horiguchi, and M. Tateda, "Thermal Effects on Brillouin Frequency Shift in Jacketed Optical Silica Fibers," Appl. Opt., vol. 29, no. 15, pp. 2219-2222, 1990.

[4] T. Horiguchi, T. Kurashima, and M. Tateda, "Tensile Strain Dependence of Brillouin Frequency Shift in Silica Optical Fibers," IEEE Photon. Tech. Lett., vol. 1, no. 5, pp. 107-108, 1989.

[5] T. Horiguchi, T. Kurashima, and M. Tateda, "Tensile Strain Dependence of Brillouin Frequency Shift in Silica Optical Fibers," IEEE Photon. Tech. Lett., vol. 1, no. 5, pp. 107-108, 1989.

[6] M. Soto and L. Thévenaz, "Modeling and Evaluating the Performance of Brillouin Distributed Optical Fiber Sensors," Opt. Express, vol. 21, no. 25, pp. 31347-31366, 2013.

[7] T. Horiguchi, K. Shimizu, T. Kurashima, M. Tateda, and Y. Koyamada, "Development of a Distributed Sensing Technique Using Brillouin Scattering," J. lightw. Technol. vol. 13, no. 7, pp. 1296-1302, 1995.

[8] H. Takahashi, F. Ito, C. Kito, and K. Toge, "Individual Loss Distribution Measurement in 32-branched PON Using Pulsed Pump-probe Brillouin aAalysis," Opt. Express, vol. 21, no. 6, pp. 6739-6748, 2013.

[9] F. Ito, H. Takahashi, K. Toge, and C. Kito, "End reflection assisted Brillouin measurement for PON monitoring," OECC2013, pp. MS2-1, 2013.

[10] "Optical Fibre Cable Maintenance Criteria for In-service Fibre Testing in Access Networks," ITU-T L. 66, 2007.

Estimation of the Starlink Global Satellite System Capacity

Denys Rozenvasser[1] and Kateryna Shulakova[2,3]
[1]*Department of Computer Science, International Humanitarian University,Fontanska road Str. 33, Odesa, Ukraine*
[2]*Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*
[3]*Department of Computer Engineering and Information Systems, State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, Odesa, Ukraine*
denysrozenvasser@gmail.com, katejojo29@gmail.com

Keywords: Starlink, Link Budget, Modulation, Capacity, Signal-to-Noise Ratio, Spectral Efficiency, Informational Efficiency, Noise Immunity, Error-Correcting Code, Modelling.

Abstract: In this paper, the capacity of the Starlink global satellite system is considered and evaluated. We will evaluate using the link budget. The code gain due to the addition of error-correcting coding has been added to the calculation of the link budget. At high frequencies, the orbit height is limited to 340 km. Most of the satellites are planned to be located at this altitude. For lower frequencies, higher orbits can be used: 550 km and 1110 km. To date, the spectral efficiency of possible variations of Starlink is limited to 4.5 bits/Hz. At the time of this writing, there are approximately 3300 Starlink satellites in orbit, which theoretically provides a capacity of about 20 Tbps due to the reduced level of modulation and the use of only one polarization. At the end of the first stage of launching satellites into orbit, the system capacity will be 88.5 Tbps over most of the studied frequency band and it will increase as more satellites are launched into low-earth orbit. Working in the future in dual polarizations with 64QAM and with a planned 11 943 satellites, SpaceX will be able to provide Internet access anywhere in the world with high bandwidth and spectral efficiency.

1 INTRODUCTION

Capacity is one of the most important characteristics of any transmission system. The use of multiposition modulation methods makes it possible to provide high transmission rates by increasing the number of signal levels, but it is limited by the system capacity [1]. But increasing the number of signal levels leads to a decrease in the noise immunity of the signal, so there is a need to develop other ways to increase the capacity, which can cause less signal-to-noise loss. Increasing the capacity and spectral efficiency makes it possible to reduce the cost of using the allocated frequency band. However, with the growth of the transmission system capacity and the data transmission rate, the total cost of the system components increases and the probability of errors during data transmission increases. The optimal ratio between the values of these parameters is one of the most important factors for operators and equipment manufacturers.

Today, one of the world's largest operators and manufacturers of equipment for telecommunications services is SpaceX, which develops a network of Starlink satellites [2], through which it provides fast broadband Internet access services.

Initially, SpaceX planned to launch 4,425 satellites with Ka- and Ku-band transmitters (from 12 to 18 GHz and from 26.5 to 40 GHz, respectively) with a service life of 5-7 years into low Earth orbit in 83 orbital planes in the altitude range from 1110 up to 1325 kilometers. It is this placement, according to the company, should have eliminated the main drawback of satellite Internet - large delays in signal transmission [3-5]. In 2017, it became known that the SpaceX satellite constellation will consist of 11 943 satellites. In addition to the initial 4425 devices at altitudes of 550 and 1110 km, the constellation is planned to include 7518 satellites in lower orbits - from 335 to 346 kilometers. By design, all devices will be built in peer-to-peer connections. Each satellite in the network will be an independent unit, simultaneously performing the functions of both a client, that is, a network member, and a server that controls its segments. This should expand communication

channels and increase access speed in densely populated areas - a significantly lower location of satellites will reduce the signal delay to 25 ms [6,7]. For comparison, in cellular networks of the 4G standard, the delay is on average from 7-8 ms, but it can be more depending on the load on the network and the distance of the cell towers. At the moment, SpaceX has launched 3449 Starlink satellites into orbit during 65 Falcon 9 launches, of which approximately 3300 are operating normally.

The capacity [8,9] of the Starlink satellite will be 20 Gbit/s when operating in two polarizations and with 64QAM (Quadrature Amplitude Modulation) [10]. In this case, the rate of the error-correcting code reaches 0.95, and the spectral efficiency is 7.45 dB. This value is approximately 2 times higher than in other satellite methods of broadband Internet access. But as of today, the network can only use one polarization. To work with 64QAM modulation, it is necessary to have a signal-to-noise ratio of more than 17 dB. However, this parameter at the UT-1 terminal is between 11 and 12.5 dB, which corresponds to 16-32APSK (amplitude and phase shift keying) and has a spectral efficiency of 4.5 bit/Hz maximum.

The spectral efficiency of possible Starlink options is presented in Table 1 [11].

Table 1: Starlink spectral efficiency.

Modulation	Code rate	Spectral efficiency, bit/s/Hz	Spectral efficiency, dB
QPSK	0,5	0,989	-0,05
8PSK	0,75	2,228	3,48
8PSK	0,833	2,479	3,94
16APSK	0,666	2,637	4,21
16APSK	0,75	2,967	4,72
32APSK	0,9	4,453	6,49
64QAM	0,772	4,5234	6,55
64QAM	0,873	5,1152	7,09
64QAM	0,948	5,5547	7,45

2 LINK BUDGET

The characteristics of the system will be estimated using the link energy budget [10,12].

Power received at the user terminal

$$P_{rx} = P_{tx} + G_{tx} + G_{rx} - L_{tx} - L_{rx} - L_{atm} - L_p. \quad (1)$$

In this equation, P_{tx} is the power of the transmitter, G_{tx} and G_{rx} are the gains of the transmitter and receiver, respectively, L_{tx} and L_{rx} are the losses at the transmitter and receiver, L_{atm} is atmospheric losses, and L_p is the path loss introduced by the separation between the transmitter and receiver.

P_{tx} is the power of the transmitter is dependent on dynamic resource allocation η_a and satellite mass m.

Transmitter gain (G_{tx}) is dependent on the antenna technology (antennas diameter d, wavelength λ):

- Parabolic: $G_{tx} = 20 \cdot \log_{10}\left(\frac{\eta_a \pi d}{\lambda}\right)$;
- Phased Array: $G_{tx} = 20 \cdot \log_{10}\left(\frac{4\eta_a \pi \sqrt[3]{m^2}}{\lambda}\right)$.

Receiver gain on noise (G_{rx}) based on user terminal type:

- Parabolic: G_{rx} = 5.24 dB;
- Phased Array: G_{rx} = 10.8 dB.

Atmospheric losses L_{atm} is dependent on frequency band and atmospheric pressure P_{atm}, the averages are as follows:

- X-band L_{atm}= 0.61 dB/km;
- Ku/Ka-band L_{atm}= 0.97 dB/km;
- V-band L_{atm}= 7.42 dB/km.

Matlab's internal gaspl function for atmospheric losses simulation was used. The simulation results are presented in Figure 1.

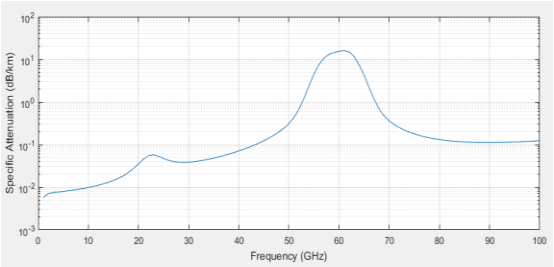


Figure 1: Atmospheric losses simulation.

We observe a burst of parameter L_{atm} at 60 GHz in fig. 1. To reduce the influence of atmospheric losses, it is desirable to use the frequency range up to 53 GHz, which corresponds to the frequency plan proposed by SpaceX (Table 2). For uplink transmission they use 14-52.4 GHz frequencies and for downlink transmission they use 10.7-42.5 GHz frequencies. Also in Table 2 the proposed

modulation types are shown. The modulation type can be changed from BPSK to 64QAM.

Table 2: Starlink frequency allocation and modulation type [10].

Characterictic	Uplink	Downlink
Frequency (GHz)	14.0-14.5 27.5-29.1 29.5-30.0 47.2-52.4	10.7-12.7 17.8-18.6 18.8-19.3 37.5-42.5
Modulation type	BPSK, MQAM	OQPSK, MQAM

Free-Space Path Loss in Decibels (L_p)

$$L_p = 20 \cdot \log_{10}\left(\frac{4\pi d}{\lambda}\right)$$

The following initial data were taken for calculations (Table 3).

Table 3: Link budget simulation parameters.

Parameter	Value	Parameter	Value
η_a	1	d	5 m
m	400 kg	P_{atm}	101.3 kPa
L_{tx}	0.5 dB	L_{rx}	0.5 dB

The results of modelling of Free-Space Path loss and atmospheric losses depending on the orbit altitude are shown in Figure 2.

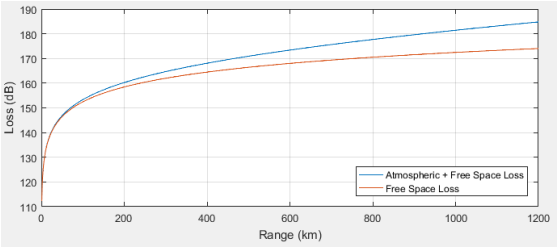


Figure 2: Free-Space Path loss and atmospheric losses.

Free-Space Path Loss is the main contributor to power loss (Figure 2). At altitudes calculated for Starlink satellites, this value is 160-175 dB. The total losses from Free-Space Path Loss and atmospheric losses are 165-185 dB.

The results of modelling and calculating the dependence of power received at the user terminal on the height of the satellite orbit for various frequencies (from 10 to 50 GHz) are shown in Figure 3.

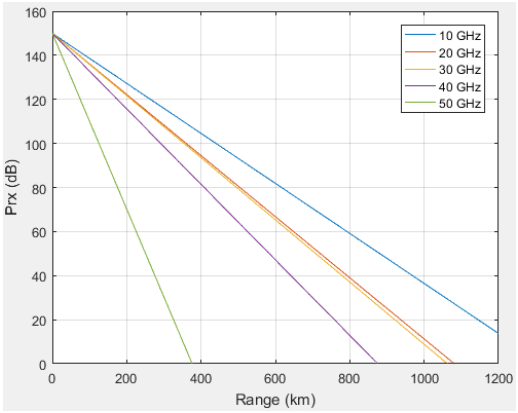


Figure 3: Link budget without ECC.

From Figure 3 it can be seen that the highest value of the signal power is 150 dB. As the signal passes through the atmosphere, it gradually attenuates. Attenuation also occurs on transmitting and receiving devices. The degree of attenuation depends on many parameters and it limits the height of the satellites orbit.

It is customary to use an error-correcting code to increase the noise immunity of systems. In this case, we use an error-correcting code to increase the link budget. We add CG to (1) - coding gain due to error-correcting code (ECC), its code rate options are shown in Table 1.

$$P_{rx} = P_{tx} + G_{tx} + G_{rx} - L_{tx} - L_{rx} - L_{atm} - L_{path} + \mathrm{CG} \quad (2)$$

The results of modelling and calculating the dependence of power received at the user terminal on the height of the satellite orbit for different frequencies (from 10 to 50 GHz), taking into account the use of an error-correcting code with a coding gain of 10 dB, are shown in Figure 4.

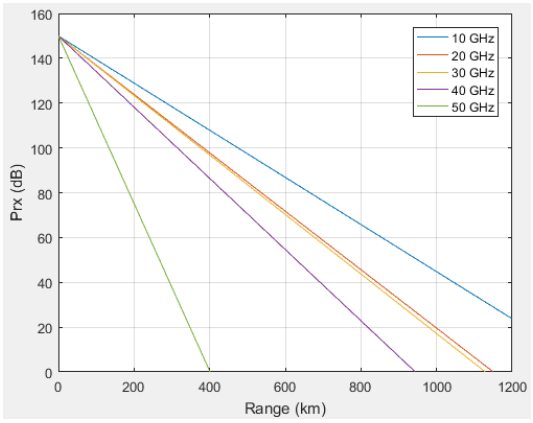


Figure 4: Link budget with ECC.

To ensure the minimum error probability, it is necessary to have a margin of the signal-to-noise ratio at the reception, which depends on the modulation type.

From Figure 4 it can be seen that at high frequencies the orbit height is limited to 340 km. Most of the satellites are planned to be located at this altitude. For lower frequencies, higher orbits 550 km and 1110 km can be used.

3 SYSTEM CAPACITY

System capacity is very important parameter. In accordance with Shannon's theorem, this is the maximum information transfer rate in the system.

The communication system is subject to noise, with Johnson noise contributing the most power to the bands of interest. Johnson noise power spectral density

$$N_0 = k \cdot T,$$

where $k = 1.3806 \cdot 10^{-23}$ – Boltzmann's constant; T – temperature.

Signal-to-noise ratio

$$\text{SNR} = \frac{P_{rx}}{P_n} = \frac{P_{rx}}{k \cdot T \cdot \text{BW}},$$

where P_n – noise power; BW – bandwidth.

We take power of signal P_{rx} from (2), bandwidth is based on the Starlink frequency plan (Table 2).

The information rate DR is calculated using the Shannon-Hartley:

$$\text{DR} = \text{BW} \cdot \log_2(1 + \text{SNR}).$$

The data rate per satellite is 10 Gbps by using only one polarization.

The capacity of the system as a whole [13] is calculated as the product of the information rate per satellite and the number of satellites n_{sat}:

$$C_{ch} = \text{DR} \cdot n_{sat}$$

In [14], it was shown that the capacity of the system has an optimum, that is, it is necessary to solve the problem of optimizing this parameter. An increase in the modulation level leads to an increase in capacity, however, when a certain level is reached, a further increase of capacity slows down significantly or even a decreasing is observed.

Dependence of system capacity on frequency band is situated in Figure 5.

System capacity is 88.5 Tbps over most of the studied frequency band and it will increase as more satellites are launched into low-earth orbit.

At the time of this writing, there are approximately 3300 Starlink satellites in orbit, which theoretically provides a capacity of about 33 Tbps. However, binary and quaternary modulation methods are used. The capacity of the constellation of satellites is 20 Tbps, which is about 4.5 times less than the declared capacity at the end of the first stage of launching satellites into orbit.

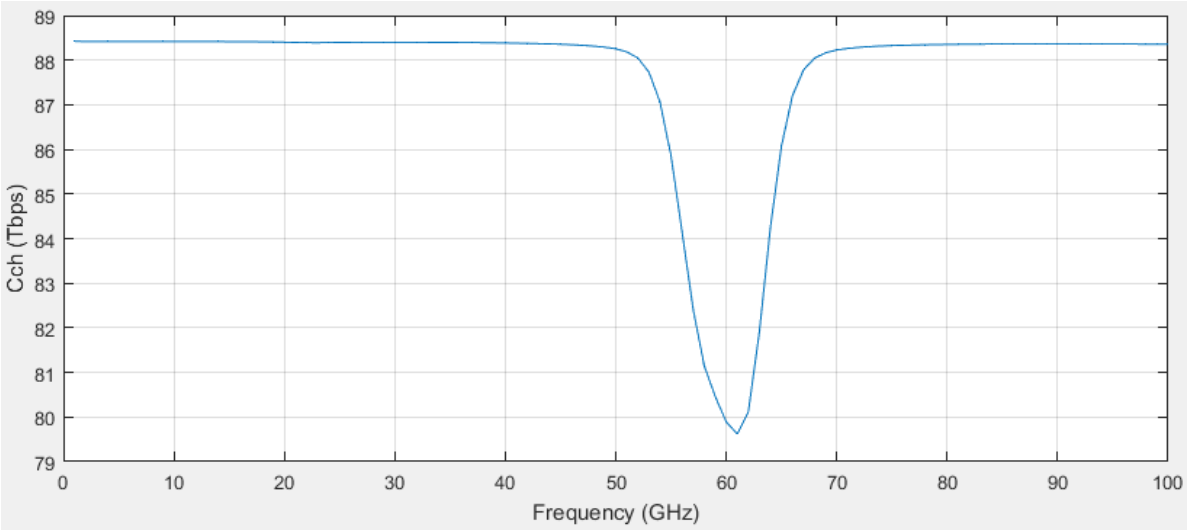


Figure 5: System capacity.

4 CONCLUSIONS

At the time of this writing, there are approximately 3300 Starlink satellites in orbit, which theoretically provides a capacity of about 33 Tbps. To ensure the minimum error probability, it is necessary to have a margin of the signal-to-noise ratio at the reception, which depends on the modulation type. However, binary and quaternary modulation methods are used. At high frequencies, the orbit height is limited to 340 km. That's why, the capacity of the constellation of satellites reaches 20 Tbps only, which is about 4.5 times less than the declared capacity at the end of the first stage of launching satellites into orbit.

Working in dual polarization with 64QAM modulation and with a planned 11943 satellites, SpaceX will be able to provide Internet access anywhere in the world with high capacity and spectral efficiency.

REFERENCES

[1] B. Sklar, “Digital Communications, Fundamentals and Applications”, 2nd ed. Prentice Hall PTR, 2001, ISBN: 0-13-084788-7.

[2] “World's most advanced broadband satellite internet,” Starlink, [Online]. Available: https://www.starlink.com/technology.

[3] T. E. Humphreys and et al., “Signal Structure of the Starlink Ku-Band Downlink,” arXiv, Oct. 2022, [Online]. Available: https://doi.org/10.48550/arXiv. 2210.11578.

[4] “FCC approves SpaceX’s plan to provide broadband services with Starlink satellites,” GeekWire, Feb. 2018. [Online]. Available: https://www.geekwire. com/2018/fcc-approves-spacexs-plan-provide-broadband-services-starlink-satellites/.

[5] A. Boyle, “SpaceX files FCC application for internet access network with 4,425 satellites,” GeekWire, Nov. 2016, [Online]. Available: https://www.geekwire.com/2016/spacex-fcc-application-internet-4425-satellites/.

[6] S. Cakaj, “The Parameters Comparison of the ‘Starlink’ LEO Satellites Constellation for Different Orbital Shells,” Frontiers in Communications and Networks, vol. 2, no. 7, 2021.

[7] T. G. Reid and et al., “Broadband LEO constellations for navigation,” Navigation, Journal of the Institute of Navigation, vol. 65, no. 2, pp. 205-220, 2018.

[8] T. G. R. Reid and et al., “Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications,” in Wiley-IEEE, vol. 1, 2020, ch. Navigation from Low Earth Orbit: Part 1: Concept, Capability, and Future Promise, pp. 1359-1380.

[9] Z. M. Kassas, “Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications,” in Wiley-IEEE, vol. 1, 2020, ch. Navigation from Low Earth Orbit: Part 2: Models, Implementation, and Performance, pp. 1381-1412.

[10] A. Aguilar, P. Butler, J. Collins, and M. Guerster, “Tradespace exploration of the next generation communication satellites,” in AIAA Scitech 2019 Forum, 2019.

[11] S. Pekhterev, “The bandwidth of the StarLink constellation and the assessment of its potential subscriber base in USA,” SatMagazine, Nov. 2021, pp. 54-57.

[12] M. G. Kim and H. S. Jo, “Performance Analysis of NB-IoT Uplink in Low Earth Orbit Non-Terrestrial Networks,” Sensors, vol. 22, no. 18, Sep. 2022, Art no. 7097, [Online]. Available: https://doi.org/10.3390/ s22187097.

[13] J. G. Proakis and M. Salehi, “Digital Communications”, 5th ed. McGraw-Hill, 2007.

[14] V. A. Breskin and D. M. Rozenvasser, “Optical transport network capacity optimization,” in 2nd International Conference on Information and Telecommunication Technologies and Radio Electronics, UkrMiCo 2017 - Proceedings, 2017, doi:10.1109/UkrMiCo.2017.8095407.

Persistent Homology in Machine Learning: Applied Sciences Review

Oleksandr Yavorskyi[1], Andrii Asseko-Nkili[1] and Nataliia Kussul[1,2,3]
[1]*Department of Mathematical Modeling and Data Analysis, Igor Sikorsky Kyiv Polytechnic Institute, Peremohy Avenue 37, Kyiv, Ukraine*
[2]*Department of Space Information Technologies and System, Space Research Institute National Academy of Science of Ukraine an State Space Agency of Ukraine, Glushkov Avenue 40, Kyiv, Ukraine*
[3]*Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*
yaotianjiu@gmail.com, a0494034@gmail.com, nataliia.kussul@gmail.com

Keywords: Algebraic Topology, Persistent Homology, Machine Learning, Physics, Healthcare, Topological Data Analysis, Chemistry, Biology, Material Sciences, Data Processing.

Abstract: Topological Data Analysis ('TDA') has become a vibrant and quickly developing field in recent years, providing topology-enhanced data processing and Machine Learning ('ML') applications. Due to the novelty of the field, as well as the dissimilarity between the mathematics behind the classical ML and TDA, it might be complicated for a field newcomer to assess the feasibility of the approaches proposed by TDA and the relevancy of the possible applications. The current paper aims to provide an overview of the recent developments that relate to persistent homology, a part of the mathematical machinery behind the TDA, with a particular focus on applied sciences. We consider multiple areas, such as physics, healthcare, material sciences, and others, examining the recent developments in the field. The resulting summary of this paper could be used by field experts to expand their knowledge on recent persistent homology applications, while field newcomers could assess the applicability of this TDA approach for their research. We also point out some of the current restrictions on the use of persistent homology, as well as potential development trajectories that might be useful to the whole field.

1 INTRODUCTION

Artificial Intelligence ('AI') is a fruitful and flourishing area that focuses on the development of algorithms that are capable of replicating human behavior. An important constituent of this area, which encompasses a variety of mathematical instruments developed for capturing, formalization, and optimization of methods that can help in the aims of AI, called Machine Learning ('ML')

Linear algebra, statistics, and probability theory, as well as functional analysis, comprise the list of the most widely-used mathematical instruments for Machine Learning. At the same time, more sophisticated mathematical machinery receives ever-growing attention from ML specialists [1]. Algebraic topology could be considered as one of the most important of such 'mathematical newcomers' to the field. The impact of topology on the current ML scene led to the emergence of a whole new area called Topological Data Analysis.

Algebraic topology raises basic questions about the shape of the object and is especially interested in the shape features that are invariant under deformation. A hole in S^2 sphere or torus is a typical example of such an invariant since it does not diminish up until we 'cut' the figure.

2 FUNDAMENTALS OF PERSISTENT HOMOLOGY

Informally, persistent homology ('PH') allows us to discover which features of the data set (called 'point cloud') are time-invariant. The latter notion gives researchers the right means to assess some constant geometrical (or, it is better to say, 'topological') features of the set. Below we give a more elaborate introduction to this concept.

2.1 Complexes, Filtrations and Persistence

A simplicial complex could be defined simply as a set consisting of points, lines, and *n*-order polytopes of

some point cloud defined on a manifold. The dimensionality of the simplex is defined through its vertices, i.e., a *k*-simplex is represented by a convex hull with *k*+1 independent vertices [2].

Now, we could induce a family of simplicial complexes out of the point cloud *X*. This family should be ordered by increasing inclusion, such that complex K_n is included in K_{n+1} and so on. As an analog to this idea, one can refer to the notion of filtration used in measure theory and probability theory [3].

One now could ask a question on how long a given structure, formed by the simplicial complex at step *n*, is preserved in the dataset throughout the filtration. In this case, we can define the birth and the death time of a given feature and use this information to describe the dataset which is given. The persistent homology is then just a way to quantify the geometrical features that are preserved at step *n*. A visual explanation of these concepts can be found below in Figure 1 and Figure 2.

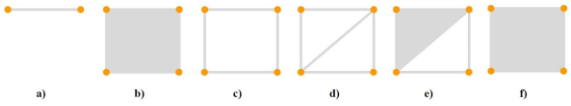


Figure 1: Different types of geometrical structures that possess different simplicial descriptions.

Figures a) and b) are the 1-simplex and 2-simplexes, respectively.

2.2 Persistent Homology as an Input

Before starting the discussion on the use of PH in Machine Learning, we should first consider how one could extract some meaningful information from it. By looking at the objects c) - f) in Figure 1, one could say that a 1-dim hole starts its existence at stage c) and lives up until stage e).

There are a number of ways to formalize and visualize this idea. Probably, one of the most common is the persistence diagrams akin to the one shown in Figure 2. On the diagram, each point represents a certain topological feature native to the point cloud analyzed (e.g., a hole). The coordinates on the XY plane of each point encode its birth and death times. Another popular way of showing this information is a barcode diagram, where instead of points, we use lines, and the length of each line denotes the lifetime of the feature. This information can be transformed [4] in order to get the persistence images, which can then be fed to the algorithm. Another option is to derive statistical information from the barcodes or diagrams, like the mean of the death time or average persistence length (which is *death time - birth time*).

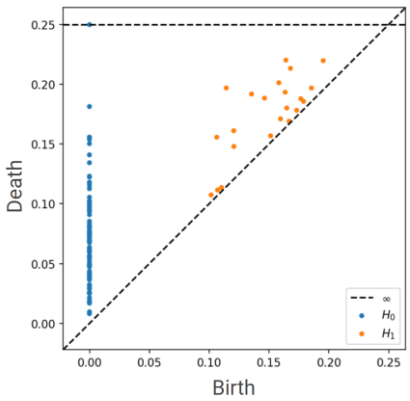


Figure 2: A persistence diagram of the random 2-dimensional point cloud. Features tracked in each dimension are shown in blue for H_0 and orange for H_1.

Normally, features that have a longer barcode (those located far from the diagonal on the diagram) describe features whose associated simplicial complexes are stable under various deformations (i.e., time-invariant from the filtration point of view). At the same time, those that die quickly tend to be treated as noise. The latter statement might be too expensive for areas where short-time transformations are important, such as chemistry, biology, drug design, physics, etc.

In this way, PH is able to provide two important insights about the data: does the data (as a point cloud) have a meaningful (useful) inner geometry and how to extract it; does the data (as a particular exemplar in the dataset) can be characterized based on its topological or geometrical features.

3 PERSISTENT HOMOLOGY IN MACHINE LEARNING

The opportunities for the use of PH information for Machine Learning lies in the concept of persistence of certain features of the dataset or its parts. This allows the characterization of the intrinsic geometrical features of the data available for the training. It becomes especially useful in cases when the geometry of the features might be put to use to crack the targeted problem. For example, in image recognition, where data is inherently geometrical [5-6].

At the same time, numerous data types can be transformed in a way that produces some unique geometry that corresponds to the data. The use of

Takens's theorem [7] to reconstruct the dynamical system behavior from the time series is one of the classic examples of such an approach.

Following this mindset, a great number of results were produced with the use of PH in Machine Learning for time series analysis [8-9], (computational) biology and chemistry [10-11], and even Natural Language Processing [12]. The general idea behind the scenes is to pre-process data using the PH-driven transformations and derive some characteristics of the objects (or the whole point cloud). The latter is then used for Machine Learning models. As a basic example, one can use the data on molecules or some materials' structure and apply PH analysis to obtain topological features of the molecules or materials and use this information for labels in classification.

A comparatively high area of applications of PH comes at a cost: most implementations are adjustments of some core ideas, which are hard to categorize and track; non-mathematicians might find it hard to cope with the nuances of these adjustments; the implementation of PH for Machine Learning requires a rather large amount of data preprocessing, which might create complications for experts outside of the industry. All these aspects related to the use of PH create a comparatively high 'entrance level', which leads to the low productivity of research. Below, we present an overview of the latest applications of persistent homology in Machine Learning in areas such as healthcare, physics, finance, and more. This coverage is intended to identify existing problems that are combated via PH, as well as the main techniques used during the problem-solving process.

3.2 Chemistry, Biology and Healthcare

This part is dedicated to the review of PH applications in Machine Learning problems that relate to Biology or Chemistry, including such practical applications as drug design and, generally, Healthcare.

As mentioned before, PH is a great tool when it comes to geometrically-enhanced data, such as images. A great example of how PH could be used for Magnetic Resonance Images ('MRI') analysis was proposed in [13]. The authors have proposed a DTA framework, which includes **D**ynamic Hierarchical Network Construction, Dynamic **T**opology Quantification, and Topological Pattern **A**nalysis. PH is used in multiple steps aiming to provide a topological description of the point cloud, which represents the MRI results. This description is then transformed and used along with a Balanced Random Forest (BRF) and Cost-effective Support Vector Machine (CE-SVM). The results of the detection of spatial patterns of multifocal lesions on clinical MRI were then compared to the existing approaches, showing a feasible increase in metrics. A great feature of the work is the authors' intention to additionally optimize their algorithm. The corresponding part might be used by other authors in their research.

Another Healthcare-related application can be found in [14]. Here, authors compare PH-induced metrics, such as (Slope of the) Betty Number Plot (BNP), BNP AUC, Mean of the area of persistent landscapes for components/holes, and others, with graph-induced metrics from the statistical point of view. In addition, the authors apply SVM for the classification comparison. The paper states that PH-induced metrics outperform their peers for the connectivity classification tasks, defined for Autism spectrum disorder conditions.

In [15], the authors use PH to compute coarsened topological features of atoms. After receiving the corresponding diagrams, they use Gaussian kernels for imagery transformation and then feed this information, along with other data, to a custom Neural Network in order to capture the protein dynamic information. The interested reader is strongly advised to look at the GitHub page provided by the authors to test the proposed approach.

As was shown in [16], functional cell description can be used in combination with PH to shed light on ionizing radiation-induced dysfunction in vascular endothelial cells. Akin to the approaches discussed before, the authors fuse the PH results with other data to get the molecular signatures of vascular dysfunction, which are then fed to a specific GANs-based algorithm to produce the results.

3.3 Physics and Geoscience

Persistence Homology receives ever-growing appreciation in material sciences due to its ability to work with shapes in a meaningful and deep way. At the same time, glass properties and structure is a long-lasting research field in material sciences, so there is no wonder that PH has found multiple applications in this area. A profound overview of this topic was published this year in [17].

Another research dedicated to the structural problem can be found in [18]. The paper discusses PH as a tool to analyze the structure of porous materials and Machine Learning approaches that can be used based on the results of PH in order to generate more efficient materials and model them in a more reasonable way.

A similar approach was followed in [19], here, the fuse of PH and Machine Learning is rather sequential than simultaneous. The paper discusses the filler morphology and how their properties can be analyzed using the combination of the aforementioned tools.

Developing the engineering trajectory of our discussion, an interesting paper covering slugging flow detection via the PH-induced techniques should be mentioned [20]. After Takens' theorem-based relevant time series transformation, authors use PH to track topological features of the obtained dynamical systems, This information is then used for the classification of the flow regimes.

An interesting example of how PH can be used even in the case of transfer learning is presented in [21]. In this paper, the authors examine the problem of chatter detection. In order to implement PH-induced metrics, authors have used a series of data transformation techniques, including FFT, which is a rare guest in PH and Machine Learning composition. The latter was presented by the use of SVM, logistic regression, random forests, and gradient boosting classification approaches.

In [22], authors implement another version of PH called zigzag persistence for crop insurance in agriculture. The key idea behind this concept is an ability to index the dynamics of the topological feature through their lifetime, providing a more comprehensive view of the point cloud structure. This approach was paired with the LSTM network, showing the reduction in the mean and variance of prediction error.

The sky is not the limit for PH applications. In [23], Large Scale Structures ('LSS') that refer to the patterns of galaxies or other objects of the universe were studied with the use of PH-related methods. More precisely, Copernicus Complexio's warm and cold dark matter models ('WDM' and 'CDM') were analyzed. It was shown that it is possible to develop a statistics-based approach, which uses the results obtained from the PH assessment, capable of distinguishing between the WDM and CDM. Moreover, the authors show that the scale at which differences occur is also trackable, which is especially important when dealing with LSS.

3.4 Selected Overview Papers

This section is intended to cover some important overview papers that came to light this year from each of the areas discussed above.

Chemistry experts can refer to [24] in order to examine the existing discussion on Materials Chemistry and how PH could be used for this purpose. Biology and drug discovery fields have recognized PH applications in [25-26], where a combination of Deep Learning and PH has become a major topic.

Healthcare topics were covered in [27-28], focusing on cancer detection, which is an important topic in AI-aided Healthcare and Precision Medicine.

4 DISCUSSION AND FUTURE WORKS

Persistent homology comprise a promising part of the Topological Data Analysis mainstream. As we have seen through the overview of the newly published papers dedicated to this approach, PH applications are mainly related to problems that have a geometrical nature. Biology and chemistry areas, including drug discovery as a joint field, tend to implement PH for cases where the problem is formulated around cellular, molecular or atomic structure, indicating the use of PH for mainly pre-processing purposes, deriving the characteristics of the objects being analyzed. Healthcare applications are, in turn, centred around image-driven data. PH methods tend to alleviate the same pain points, allowing a more meaningful approach for data extraction before the actual use of Machine Learning algorithms. At the same time, physics and material science enjoy PH in a similar way biology and chemistry do. These areas focus on geometrically-enhanced data to solve the emerging structural or dynamic problems implementing PH for data analysis as a part of the Machine Learning pipeline.

PH is rarely seen as a self-sufficient approach and tends to require an additional algorithmic pipeline. This implies an increasing production complexity for practical applications. Thus, a solution that automatically incorporates the PH-induced metrics and calculations would be highly appreciated by scientific and industry participants.

PH tends to be actively used for tasks that involve classification and seems to be flexible enough to be implemented with a variety of them. It's important to state that PH is used actively in combination with Deep Learning and generally Neural Networks-driven algorithms. Given an ever-increasing amount of data available, this is rather a good sign for PH, indicating a future development potential. At the same time, the scope of questions PH is capable of covering is not yet well-defined. For example, there is a limited conversation on how PH can be implemented for practical aspects of Meta Learning,

e.g., for synthetic data generation. It can be stated that expanded adoption of PH requires the development of new ways of data ‘ geometrization’, combining statistics (flash and blood of Machine Learning) and geometry/topology together. At the same time, the current paper covers most of the PH applications published throughout the year 2022, which poses a question on certain applications asymmetry compared to other areas, e.g., recommendation systems for marketing or Natural Language Processing for chat-bots.

In addition, there is a rather limited (if any) amount of self-sustainable PH data analysis pipelines that are capable of providing data insights. This brings another question of whether this type of algorithm can be formulated in an efficient and useful way.

We thus consider further development of the applicability analysis of PH. We will especially focus on the applications that relate to image processing problems, as well as synthetic data generation tasks, which are of great importance in the case of imbalanced datasets for both structured and unstructured data cases.

5 CONCLUSION

In this paper we have discussed the recent applications of the persistent homology-driven analysis in the context of Machine Learning. The overview shows a wide range of applicability of such methods. At the same time, it can be suggested that the place of such methods is still comparatively narrow, being mostly a preprocessing technique. Interestingly, despite the existence of topology-inspired neural networks [29], none of the recent PH applications implements such architectural solutions. Clearly, as stated in the previous section, the ease of use of PH techniques can and should be developed further. Yet, the absence of such a combination cannot be solely derived from this fact, which raises another important question on the optimal fusion of topological preprocessing and architectural solutions.

The current paper is limited in details and comparison of non-topological and topological preprocessing techniques, which is a valuable research topic we are hoping to cover in the future. In addition, we sincerely encourage our readers to acquaint themselves with the list of references provided below in order to investigate in greater detail whether PH methods can be used for their current or future research.

REFERENCES

[1] F. Hensel, M. Moor, and B. Rieck, “A Survey of Topological Machine Learning Methods,” Front. Artif. Intell., Sec. Machine Learning and Artificial Intelligence, May 2021.

[2] H. Edelsbrunner and J. Harer, “Computational Topology: An Introduction,” 2010.

[3] A. Klenke, “Probability Theory,” Berlin: Springer, 191 p., 2008.

[4] H. Adams et al., “Persistence Images: A Stable Vector Representation of Persistent Homology,” Journal of Machine Learning Research, vol. 18, pp. 1-35, 2017.

[5] P. Frosini and C. Landi, “Persistent Betti numbers for a noise tolerant shape-based approach to image retrieval,” Pattern Recognition Letters, vol. 34 pp. 863-872, 2013.

[6] G. Carlsson, T. Ishkhanov, V. Silva, and A. Zomorodian, “On the local behavior of spaces of natural images,” International Journal of Computer Vision, vol. 76, pp. 1-12, 2008.

[7] F. Takens, “Detecting strange attractors in turbulence,” Lecture Notes in Mathematics. pp. 366-381, 1981.

[8] A. Karan and A. Kaygun, “Time series classification via topological data analysis,” Expert Systems with Applications, vol. 183, November 2021.

[9] S. Majumdar and A.K. Laha, “Clustering and classification of time series using topological data analysis with applications to finance,” Expert Systems with Applications, vol. 162, December 2020.

[10] Z. Cang and G.W. Wei, “Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction,” International Journal for Numerical Methods in Biomedical Engineering, vol. 34, 2018.

[11] J. Townsend, C.P. Micucci, J.H. Hymel, V. Maroulas, and K.D. Vogiatzis “Representation of molecular structures with persistent homology for machine learning applications in chemistry,” Nat. Commun., vol. 11, 2020.

[12] X. J. Zhu, “Persistent homology: An introduction and a new text representation for natural language processing,” IJCAI, pp. 1953-1959, 2013.

[13] B.W. Xin, J. Huang, L. Zhang, and et al., “Dynamic topology analysis for spatial patterns of multifocal lesions on MRI,” Medical Image Analysis, vol. 76, 2022.

[14] A.T. Jafadideh and B.M. Asl, “Topological analysis of brain dynamics in autism based on graph and persistent homology,” Computers in Biology and Medicine, vol. 150, 2022.

[15] Y. Chiang, W.H. Hui, and S.W. Chang, “Encoding protein dynamic information in graph representation for functional residue identification,” Cell Reports Physical Science, vol. 3, July 2022.

[16] I. Morilla and Ph. Chan, “Deep models of integrated multiscale molecular data decipher the endothelial cell response to ionizing radiation,” iScience, vol. 25, January 2022.

[17] S.S. Sørensena, T. Du, C. Biscio, L. Fajstrup, and M.M. Smedskjaer, “Persistent homology: A tool to understand medium-range order glass structure,”

Journal of Non-Crystalline Solids: X, vol. 16, December 2022.

[18] D.P. Gao, J.H. Chen, Z.T. Dong, and H.W. Lin, "Connectivity-guaranteed porous synthesis in free form model by persistent homology," Computers & Graphics, vol. 106, pp. 33-44, 2022.

[19] T. Kojimaab, T. Washiob, S. Harab, and M. Koishia, "Search strategy for rare microstructure to optimize material properties of filled rubber using machine learning based simulation," Computational Materials Science, vol. 204, March 2022.

[20] S. Casolo, "Severe slugging flow identification from topological indicators," Digital Chemical Engineering, vol. 4, September 2022.

[21] M.C. Yesilli, F.A. Khasawneh, and B.P. Mann, "Transfer learning for autonomous chatter detection in machining," Journal of Manufacturing Processes, vol. 80, pp. 1-27, August 2022.

[22] T. Jiang, M. Huang, I. Segovia-Dominguez, N. Newlands, and Y.R. Gel, "Learning Space-Time Crop Yield Patterns with Zigzag Persistence-Based LSTM: Toward More Reliable Digital Agriculture Insurance," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, 2022.

[23] J. Cisewski-Kehe, B.T. Fasy, W. Hellwing, M.R. Lovell, P. Drozda, and M. Wu, "Differentiating small-scale subhalo distributions in CDM and WDM models using persistent homology," Phys. Rev. D, vol. 106, July 2022.

[24] D. Packwood et al., "Machine Learning in Materials Chemistry: An Invitation," Machine Learning with Applications, vol. 8, June 2022.

[25] Ch. Chinmayee, N.A. Murugan, and U.D. Priyakumar, "Structure-based drug repurposing: Traditional and advanced AI/ML-aided methods," Drug Discovery Today, vol. 27, pp. 1847-1861, July 2022.

[26] Y. Skaf and R. Laubenbacher, "Topological data analysis in biomedicine: A review," Journal of Biomedical Informatics, vol. 130, June 2022.

[27] C-E. Minciuna and et al., "The seen and the unseen: Molecular classification and image based-analysis of gastrointestinal cancers," Computational and Structural Biotechnology Journal, vol. 20, pp. 5065-5075, 2022.

[28] S. Prabhu, K. Prasad, A. Robels-Kelly, and X. Lu, "AI-based carcinoma detection and classification using histopathological images: A systematic review," Computers in Biology and Medicine, vol. 142, March 2022.

[29] C. Bodnar and et al., "Weisfeiler and Lehman Go Topological: Message Passing Simplicial Networks," Proceedings of the 38th International Conference on Machine Learning, PMLR, vol. 139, pp. 1-12, 2021.

Sustainable Development Goal 2.4.1 for Ukraine Based on Geospatial Data

Hanna Yailymova[1,2], Bohdan Yailymov[2], Yevhen Mazur[1], Nataliia Kussul[1,2,3] and Andrii Shelestov[1,2]
[1]*Institute of Physics and Technology, Igor Sikorsky Kyiv Polytechnic Institute, Peremohy Avenue 37, Kyiv, Ukraine*
[2]*Department of Space Information Technologies and System, Space Research Institute National Academy of Science of Ukraine an State Space Agency of Ukraine, Glushkov Avenue 40, Kyiv, Ukraine*
[3]*Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*
anna.yailymova@gmail.com, yailymov@gmail.com, cerpyps@gmail.com, nataliia.kussul@gmail.com, andrii.shelestov@gmail.com

Keywords: Geospatial Data Analysis, Machine Learning, Land Degradation, Remote Sensing, Land Cover, SDG 2.4.1.

Abstract: In this work, the indicator of sustainable development goal (SDG) 2.4.1 for Ukraine is calculated based on geospatial and satellite data. The generally accepted technology for calculating the given indicator cannot be applied for the territory of Ukraine due to the lack of systematic collection of the necessary indicators. Therefore, the authors have developed the complex method for land degradation estimation that uses different schemes for separate land cover and crop types at the country level based on satellite and modeling data using WOFOST model. The paper describes the sources of information used to create crop type classification maps and the data required for leaf area index (LAI) modeling for the WOFOST model. Calculated indicators from 2018 to 2022 for each of the regions of Ukraine. In 2022, the decrease of the indicator is monitored in almost all regions of Ukraine, which is a direct result of military actions on the territory of Ukraine.

1 INTRODUCTION

To monitor the sustainable development of the environment in the world the global indicator framework for Sustainable Development Goals was developed by the Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs) and agreed upon at the 48th session of the United Nations Statistical Commission held in March 2017 [1]. According to the proposed methodology, each country evaluates the indicators for its country, thereby receiving an assessment of the improvement or deterioration of the corresponding indicator for its country.

According to the study [2] geospatial data and data supplied by citizens are the most potential big data sources for SDG indicators assessment. Only a few SDG indicators may benefit directly from other large data sources such mobile phone data, web data (for example, data on prices or employment), postal data, and electricity data.

The climate change initiative land use data, data from the European Space Agency and the 30-meter global land cover dataset (GlobeLand30) [3] were included for multi-scenario simulation of land use [4]. The data source used to create multispectral photos included photographs from the China Environmental Disaster Mitigation Satellite, Landsat TM5, ETM+, and OLI, as well as HJ-1 multispectral images.

The relationship between agricultural interventions, dietary changes, and nutrition, which incorporates a few complex issues both within and outside of SDG-2, cannot be adequately captured by any one set of metrics (food production, diet diversification, biofortification, food safety). On the other hand, indicators 2.4.2 and 2.4.3 appear to be included into indicator 2.4.1 in the sense that sustainability has social, environmental, and economic components, which permeate every SDG and SDG-2 indicator [5]. However, access to high-quality assessment data is crucial for the SDG11 implementation to be successful [6]. Three SDG11 indicators were measured between 2013 and 2020 in Guilin's urban functional boundary. The main data sources employed were geospatial big data and high-resolution remote sensing pictures. After preprocessing, image fusion was used to combine the panchromatic and multispectral pictures.

In the analysis of urban growth processes, for example, remote sensing photos are combined with geoinformation systems and machine learning. In

work [7], [8] authors developed a technique for land productivity assessment and land cover classification using deep learning techniques and satellite data with medium and high spatial resolution.

Ukraine is one of the largest exporters of grain products in Europe, so regular analysis of the quality of agricultural land and its suitability for growing agricultural products, assessment of possible losses and yield [9] is an important task. With the onset of the war, the whole world was shaken by the problem of the possibility of a shortage of grain for importing countries, and as a result of the emergence of famine [10], as well as any other problems in the field of food security. The goal 2: End hunger, achieve food security and improved nutrition and promote sustainable agriculture is generally responsible for the agricultural sector. In particular Indicator 2.4.1: "Proportion of agricultural area under productive and sustainable agriculture", which is calculated in this study for the territory of Ukraine.

The generally accepted scheme for calculating indicators of sustainable development goals is set out in reference [11], in particular for SDG 2.4.1 in document [12] and it is recommended to be collected at least every three years. Through a consultative process that has lasted over two years, 11 themes and sub-indicators have been identified, which make up SDG 2.4.1 (Table 1).

Table 1: Themes and sub-indicators for SDG 2.4.1 assessment.

№	Themes	Sub-indicators
1	Land productivity	Farm output value per hectare
2	Profitability	Net farm income
3	Resilience	Risk mitigation mechanisms
4	Soil health	Prevalence of soil degradation
5	Water use	Variation in water availability
6	Fertilizer pollution risk	Management of fertilizers
7	Pesticide risk	Management of pesticides
8	Biodiversity	Use of agro-biodiversity-supportive practices
9	Decent employment	Wage rate in agriculture
10	Food security	Food Insecurity Experience Scale (FIES)
11	Land tenure	Secure tenure rights to land

Unfortunately, for Ukraine, there is no available data for calculating the indicator in a standard way, and the use of general global land use products is not accurate in terms of spatial resolution. Therefore, the authors of this study developed their own technology for land productivity assessment [13], which takes into account crops types information, soil parameters, and meteorological indicators during the growing season.

2 DATA USED

2.1 Land Cover / Crop Type Classification

The land cover and crop type classification maps based on own classification methodology [14] were used. For classification processing 2 bands (VV, VH) of SAR Sentinel-1 descending data with main preprocessing steps (correction of coordinates in orbit, specl-filtration, calibration, the Range-Doppler Terrain Correction, data transfer in decibel, creating a data stack, saving imagery bands, and merging of bands VV and VH within a single granule) with 10-meters spatial resolution are used. Also, for classification processing 4 bands (Red B4, Green B3, Blue B2, InfraRed B8) of Sentinel-2 data with preprocessing Level-2A and 10-meters spatial resolution are used. The revisit time of Sentinel-2 is 5 days, but due to high cloud cover, monthly composites was used obtained as the median value of all possible values for every 5 days in the respective bands. A Scene Classification Map (SCL) band with a spatial resolution of 20 meters is used to mask clouds from optical data. Optical composites are obtained in the Google Earth Engine cloud platform.

The multilayer perceptron (MLP) is used for training neural network. Compared to deep neural network algorithms, in particular convolutional neural networks, the MLP algorithm loses in accuracy by 1%, but requires much more powerful computing resources and time to obtain the final product. That is why we use MLP neural network algorithm.

As an input to the neural network model, we have a stack of rasters - a time series of satellite data. Together with the satellite data stack, training in-situ data is fed to the input of the neural network model, which is collected along the roads every year in the form of vector contours of the fields, indicating the corresponding land cover or crop type class. The output of the model is a raster georeferenced image, where each pixel contains the corresponding land cover or crop type class. The validation independent in-situ data set is used to obtain class accuracies and the overall accuracy of the resulting classification map.

2.2 Leaf Area Index Modelling

The simulation model for the quantitative analysis of the growth and production of annual field crops WOFOST (WOrld FOod STudies) [15] was used to simulate LAI for agricultural crops. The main input parameters for this model are soil profiles (including various soil characteristics), crop profiles (including sowing dates, flowering, maturity and other important characteristics), as well as meteorological indicators.

2.2.1 Soil Profiles

Soil parameters are important input data for estimating the model value of LAI. There are 40 main types of soils for Ukraine (Figure 1). For each type of soil parameters are recorded in profiles and used as input of the model. The soil parameters, in particular (soil moisture content at saturation, at wilting point, and at field capacity) are available according to European Soil Database [16].

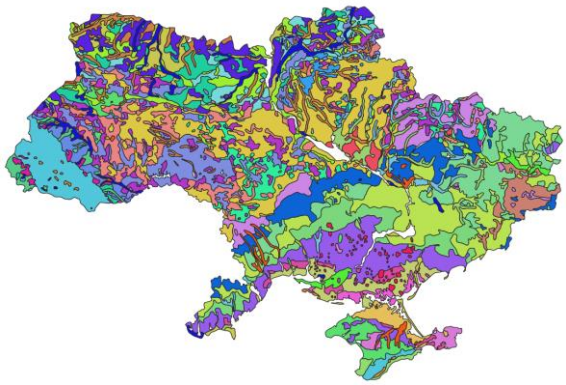


Figure 1: Soil map for Ukraine.

2.2.2 Crop Profiles

The most important characteristic in crop profiles is the accumulated sum of temperature from emergence to anthesis (TSUM1), as well as temperature sum from anthesis to maturity (TSUM2). For winter crops, an important characteristic is the sum of temperatures that exceed 4 degrees Celsius for the continuation of vegetation after wintering.

2.2.3 Meteorological Data

The WOFOST system uses daily meteorological parameters from NASA Prediction of the Worldwide Energy Resources (POWER) Project [17], in particular temperature, precipitation, irradiation, wind power and direction. That resolution is 1.0° latitude by 1.0° longitude for the radiation data sets and 0.5° latitude by 0.625° longitude for the meteorological data sets (or approximately 55.5 km x 69 km). The Figure 2 shows the geospatial ratio of meteorological data pixels to oblasts of Ukraine.

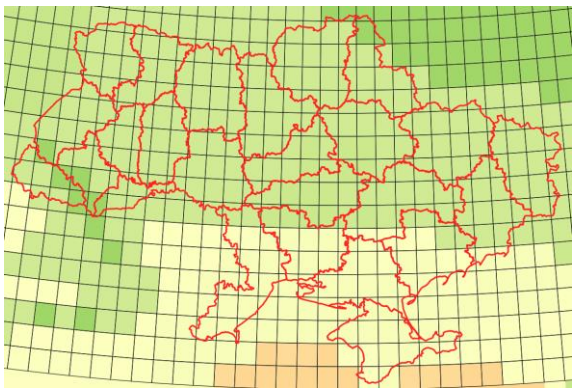


Figure 2: Soil map for Ukraine.

2.2.4 Grid Creation for Ukraine

Considering the rather low spatial resolution of meteorological data, as well as the fact that a single soil can extend over a long area, a new markup is created for the use of the WOFOST model, which is the intersection of meteorological data and soils (Figure 3). Thus, we increase the variability and taking into account the agro-climatic zones of Ukraine. For each polygon created, a point for which a soil profile, crop profile and meteorological indicators are assigned is set in accordance.

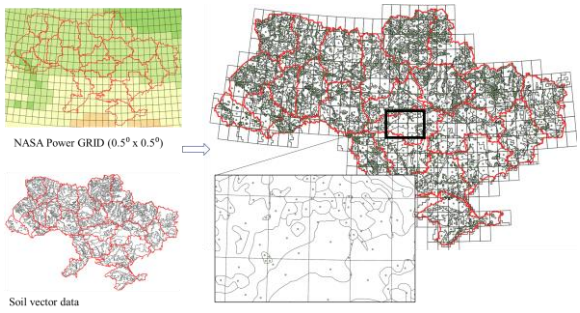


Figure 3: Grid for WOFOST model using for Ukraine.

3 METHODOLOGY

Our method of land productivity assessment is based on the land cover and crop type classification from satellite imagery and application of different schemes

of land degradation assessment for each of them [13] (Figure 4).

We consider forest cuts as land degradation for forests and assess them using deep learning models [18]. Land degradation for croplands is estimated by difference of real leaf area index (LAI) based on MODIS data [19] and ideal LAI, calculated with the biophysical crop development model WOFOST [15], which takes into account the biophysical characteristics of the soil (soil moisture content at saturation, at wilting point, and at field capacity), meteorological conditions (precipitation, temperature, wind direction and strength), altitude above sea level.

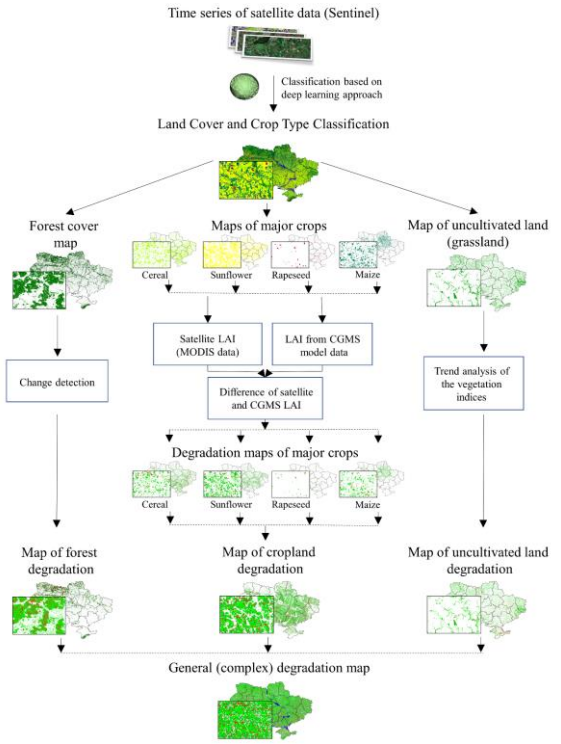


Figure 4: General scheme of the complex method of land degradation estimation [13].

The sustainability criteria are the distance from the 90th percentile of the national distribution [12]:

- Green (desirable or productive): Sub-indicator value is ≥ 2/3 of the corresponding 90th percentile.
- Yellow (acceptable or sustainable): Sub-indicator value is ≥ 1/3 and < 2/3 of the corresponding 90th percentile.
- Red (unsustainable or degradation): Sub-indicator value is < 1/3 of the corresponding 90th percentile.

The land degradation for grassland is determined with a traditional approach based on trend of vegetation index NDVI [20] extracted from satellite imagery.

The indicator 2.4.1 is defined by the (1):

$$SDG_{2.4.1} = \frac{Area\ under\ productive\ and\ sustainable\ agriculture}{Agriculture\ land\ area}, \quad (1)$$

where area under productive and area under sustainable agriculture calculated based on land degradation map, and agriculture land area based on crop type classification map.

4 RESULTS

According to the developed technology for indicator 2.4.1 calculation, the land degradation maps were calculated, which include three classes (productive, sustainable and degradation land) for the calculation of which the crop type maps for 2018 - 2022 were used. On the basis of the obtained maps, according to (1), the SDG indicator 2.4.1 was calculated for each region of Ukraine (Figure 5).

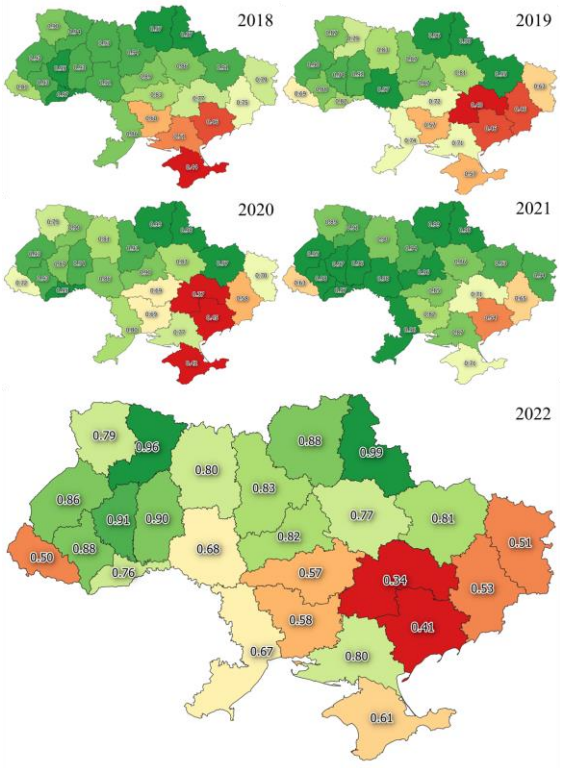


Figure 5: SDG 2.4.1 indicator for Ukraine (2018 – 2022).

During 2018 - 2021, a worse situation is observed in the southern regions in terms of the value of the indicator, which may be caused by the annual crop rotation violation in this zone [21].

The situation in 2022 has significantly worsened in almost all regions of Ukraine, in particular in the south-eastern regions, where hostilities continue. The worsening situation in other regions may also be due to shortages of fuel, resources for productive agriculture, and human resources, as many workers have gone to war.

The Table 2 shows the calculated SDG indicators 2.4.1 for 5 years, as well as their comparison with 2022 in percentage. The cases where the indicator decreased are marked in pink color, and the cases in which the indicator increased are marked in green color. From Table 2, it can be concluded that in almost all regions of Ukraine, compared to the previous 4 years, the condition of the land has deteriorated. Kherson region has higher indicators than other territories under occupation, as it contains the largest areas of irrigated territories in Ukraine, which helped it.

Table 2: Indicator SDG 2.4.1 for Ukraine (2018 – 2022).

	SDG 2.4.1					Difference (in %) with 2022			
Oblasts	2018	2019	2020	2021	2022	2018	2019	2020	2021
Vinnytska	0,92	0,97	0,86	0,98	0,68	-25,7	-29,3	-20,2	-30,5
Zakarpatska	0,81	0,69	0,72	0,63	0,50	-38,0	-26,7	-30,2	-20,3
Kirovohradska	0,83	0,72	0,69	0,85	0,57	-31,5	-21,7	-18,0	-33,6
Luhanska	0,79	0,63	0,70	0,94	0,51	-36,4	-19,8	-28,0	-46,3
Dnipropetrovska	0,77	0,40	0,37	0,71	0,34	-56,4	-14,9	-8,9	-52,8
Kharkivska	0,91	0,95	0,97	0,93	0,81	-10,5	-14,8	-16,2	-12,9
Zaporizka	0,46	0,46	0,45	0,54	0,41	-9,9	-10,8	-7,6	-23,5
Odeska	0,86	0,74	0,80	0,96	0,67	-22,6	-10,1	-16,9	-30,4
Volynska	0,90	0,87	0,79	0,89	0,79	-12,1	-9,1	-0,3	-11,3
Chernihivska	0,97	0,96	0,99	0,99	0,88	-8,9	-8,4	-11,0	-10,7
Chernivetska	0,97	0,82	0,96	0,97	0,76	-21,2	-7,6	-20,3	-21,6
Cherkaska	0,89	0,87	0,90	0,96	0,82	-7,4	-6,0	-8,3	-14,1
Lvivska	0,93	0,90	0,93	0,96	0,86	-7,9	-4,9	-7,1	-9,9
Kyivska	0,94	0,87	0,91	0,94	0,83	-11,5	-4,7	-8,3	-11,2
Poltavska	0,86	0,81	0,84	0,86	0,77	-10,3	-4,4	-7,4	-10,1
Zhytomyrska	0,93	0,83	0,81	0,90	0,80	-14,7	-4,0	-1,6	-11,3
Ternopilska	0,96	0,94	0,87	0,97	0,91	-4,8	-3,7	4,4	-6,7
Khmel'nytska	0,93	0,92	0,94	0,96	0,90	-3,1	-2,1	-3,7	-6,6
Ivano-Frankivska	0,93	0,88	0,93	0,98	0,88	-5,2	-0,7	-6,1	-10,4
Donetska	0,75	0,46	0,58	0,65	0,53	-29,2	14,7	-8,5	-18,2
Mykolaivska	0,60	0,57	0,69	0,82	0,58	-2,7	1,0	-16,4	-29,4
AR of Crimea	0,44	0,57	0,42	0,72	0,61	38,4	8,3	46,0	-14,3
Khersonska	0,51	0,71	0,77	0,87	0,80	56,8	13,0	3,1	-7,9
Sumska	0,97	0,98	0,98	0,98	0,99	2,4	0,6	0,6	0,5
Rivnenska	0,94	0,79	0,90	0,91	0,96	1,7	21,4	6,5	5,7
Ukraine	0,83	0,77	0,79	0,87	0,73	-12,5	-6,1	-8,1	-17,0

5 CONCLUSIONS

In this study, we calculated indicator 2.4.1 for 2018-2022 for the territory of Ukraine using a previously developed geospatial method for assessing land degradation based on remote sensing data, neural networks, and biophysical modeling [13]. It takes into account different land cover/land use classes and provides a specific way of assessing land degradation for each of them. Due to the high computational complexity of the method, it is implemented in the CREODIAS cloud environment, thanks to the resources within the EO4UA initiative. According to our research, most of the territory of Ukraine remains stable. The most land degradation is observed on arable lands in the southeastern regions due to environmentally unfavorable methods of farming and military operations.

The developed technology is flexible and applicable for different climatic zones, because during the biophysical simulation according to the WOFOST model, it takes into account precipitation, temperature, as well as the main stages of crop growth - seedlings, maturation, maturity. After receiving the annual maps of degradation according to the described methodology, the changes at the level of the regions of Ukraine were analyzed, and as a result it was concluded that in 2022 there was a significant deterioration compared to the previous 4 years. On the basis of the obtained result, it is possible to make appropriate management decisions regarding the prevention and regulation of land quality in Ukraine.

The proposed technology can be used for any country in the world. The only information required is a crop type classification map for the required area. In particular, the data technology is also applied to calculate the indicator of the sustainable development goal 2.4.1 for the territory of Germany within the framework of the Horizon 2020 e-shape project.

ACKNOWLEDGMENTS

The work was funding support from the European Commission through the joint World Bank/EU project ‘Supporting Transparent Land Governance in Ukraine’ (ENI/2017/387–093 and ENI/2020/418–654), carried out within the framework of the project 2020.01/0273 "Intelligent models and methods for determining land degradation indicators based on satellite data" (NRFU Competition “Science for the safety of human and society”), 2020.02/0284 “Geospatial models and information technologies of satellite monitoring of smart city problems” (NRFU Competition “Leading and Young Scientists Research Support”), Horizon 2020 e-shape project, EO4UA initiative and CloudFerro Company.

REFERENCES

[1] Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development. [Online]. Available: https://unstats.un.org/sdgs/indicators/indicators-list/.

[2] G. B. Van Halderen, “Big data for the SDGs: country examples in compiling SDG indicators using non-traditional data sources,” Working paper, United Nations, 2021. [Online]. Available: https://hdl.handle.net/20.500.12870/3442.

[3] J. Chen, X. Cao, S. Peng, and H. Ren, “Analysis and applications of GlobeLand30: a review,” ISPRS International Journal of Geo-Information, 2017, vol. 6(8), p. 230, doi: 10.3390/ijgi6080230.

[4] M. Cao, L. Chang, S. Ma, Z. Zhao, K. Wu, X. Hu, and M. Chen, “Multi-Scenario Simulation of Land Use for Sustainable Development Goals,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, vol. 15, pp. 2119-2127, doi: 10.1109/JSTARS.2022.3152904.

[5] J. D. Gil, P. Reidsma, K. Giller, L.Todman, A. Whitmore, and M. van Ittersum, “Sustainable development goal 2: Improved targets and indicators for agriculture and food security,” Ambio, 2019, vol. 48(7), pp. 685-698, doi: 10.1007/s13280-018-1101-4.

[6] L. Han, L. Lu, J. Lu, X. Liu, S. Zhang, K. Luo, and Q. Li, “Assessing Spatiotemporal Changes of SDG Indicators at the Neighborhood Level in Guilin, China: A Geospatial Big Data Approach,” Remote Sensing, 2022, vol. 14(19), p. 4985, doi: 10.3390/rs14194985.

[7] M. Brussel, M. Zuidgeest, K. Pfeffer, and M. van Maarseveen, “Access or accessibility? A critique of the urban transport SDG indicator,” ISPRS International Journal of Geo-Information, 2019, vol. 8(2), p. 67, doi: 10.3390/ijgi8020067.

[8] N. Kussul, M. Lavreniuk, A. Kolotii, S. Skakun, O. Rakoid, and L. Shumilo, “A workflow for Sustainable Development Goals indicators assessment based on high-resolution satellite data,” International Journal of Digital Earth, 2019, vol. 13:2, pp. 309-321, doi: 10.1080/17538947.2019.1610807.

[9] K. Deininger, D.A. Ali, and et al., “Quantifying War-Induced Crop Losses in Ukraine in Near Real Time to Strengthen Local and Global Food Security”, Food Policy, vol. 115, Feb. 2023, p. 102418, doi: 10.1016/j.foodpol.2023.102418.

[10] L. Shumilo, S. Drozd, and et al., “Mathematical Models and Informational Technologies of Crop Yield Forecasting in Cloud Environment,” Ilchenko, M., Uryvsky, L., Globa, L. (eds) Progress in Advanced Information and Communication Technology and Systems. MCiT 2021. Lecture Notes in Networks and Systems, vol 548. Springer, Cham. pp. 143-164, doi: 10.1007/978-3-031-16368-5_7.

[11] SDG Indicators Metadata repository. [Online]. Available: https://unstats.un.org/sdgs/metadata/.

[12] SDG 2.4.1 indicator metadata. [Online]. Available: https://unstats.un.org/sdgs/metadata/files/Metadata-02-04-01.pdf.

[13] N. Kussul, L. Shumilo, H. Yailymova, A. Shelestov, and T. Krasilnikova, "Complex method for land degradation estimation," 2nd International Scientific Conference on Environmental Sustainability in Natural Resources Management, 01 November 2022, Riga, Latvia, vol. 1126, no. 1, p. 012032, doi: 10.1088/1755-1315/1126/1/012032.

[14] A. Shelestov, M. Lavreniuk, V. Vasiliev, and et al., "Cloud Approach to Automated Crop Classification Using Sentinel-1 Imagery," IEEE Transactions on Big Data, 2020, vol. 6, no. 3, pp. 572-582, doi: 10.1109/TBDATA.2019.2940237.

[15] C. V. Van Diepen, J. V. Wolf, H. Van Keulen, and C. Rappoldt, "WOFOST: a simulation model of crop production," Soil use and management, 1989, vol. 5(1), pp. 16-24, doi: 10.1111/j.1475-2743.1989.tb00755.x.

[16] European Soil Database. [Online]. Available: https://esdac.jrc.ec.europa.eu/content/sinfo-esdb-data-adapted-mars-crop-yield-forecasting-system#tabs-0-description=1.

[17] T. Zhang, W. S. Chandler, J. M. Hoell, D. Westberg, C. H. Whitlock, and P. W. Stackhouse, "A global perspective on renewable energy resources: NASA's prediction of worldwide energy resources (power) project," In Proceedings of ISES World Congress 2007, vol. I–vol. V, pp. 2636-2640. Springer, Berlin, Heidelberg, doi: 10.1007/978-3-540-75997-3_532.

[18] L. Zhichkina, V. Nosov, K. Zhichkin, H. Aydinov, V. Zhenzhebir, and V. Kudryavtsev, "Satellite monitoring systems in forestry," Journal of Physics: Conference Series 2020, vol. 1515, no. 3, p. 032043. IOP Publishing.

[19] Y.Tian and et. all, "Multiscale analysis and validation of the MODIS LAI product: I. Uncertainty assessment," Remote Sensing of Environment, 2002, vol. 83(3), pp. 414-430, doi: 10.1016/S0034-4257(02)00047-0.

[20] M. H. Easdale and et all., "Trend-cycles of vegetation dynamics as a tool for land degradation assessment and monitoring," Ecological Indicators, 2019, vol. 107, 105545, doi: 10.1016/j.ecolind.2019.105545.

[21] N. Kussul, K. Deininger, L. Shumilo, M. Lavreniuk, D. Ayalew Ali., and O. Nivievskyi, "Biophysical Impact of Sunflower Crop Rotation on Agricultural Fields," Sustainability. 2022. no. 14(7):3965, pp. 125-132, doi: 10.3390/su14073965.

Comparative Analysis of Machine Learning Models for Diabetes Prediction

Zoran Stojanoski, Marija Kalendar and Hristijan Gjoreski
Computer Technologies and Engineering Department, Faculty of Electrical Engineering and Infomation Technologies, "SS. Cyril and Methodius University" in Skopje, Rugjer Boshkovikj Str. 18, Skopje, N. Macedonia
zoran_stojanoski@outlook.com, {marijaka, hristijang}@feit.ukim.edu.mk

Keywords: Machine Learning, Diabetes Prediction, Feature Analysis, ML Models Comparison.

Abstract: This paper focuses on analyzing the benchmark Diabetes dataset which consists of eight commonly measured characteristics. The goal of the study is to present comparative analysis of six machine learning models that predict diabetes, as well as various preprocessing techniques (under-over sampling, feature standardization). The study investigates various approaches and presents results demonstrating that machine learning algorithms can achieve high accuracy results for diabetes prediction, enabling early detection and better outcomes for patients. The paper shows that ensemble learning methods, such as Extra Trees Classifier and Random Forest Classifier, along with appropriate data pre-processing techniques, can lead to 86% accuracy in diabetes prediction classification problems. The paper highlights the potential for machine learning to play a valuable role in the prediction and management of diabetes, leading to improved quality of life and health outcomes for patients.

1 INTRODUCTION

Diabetes is a very common disease affecting millions of individuals in the world today. This disease may be detected and managed early, which could have a major positive impact on quality of life and health outcomes for patients. Thus, this research will focus on the possibilities to detect diabetes from easily measurable features and to enable prediction access to a wider audience. Machine learning, as a technique in common use today has the potential to completely transform prediction in general. Our research demonstrates that implementing machine learning in diabetes predications yields high accuracy results, proving that machine learning algorithms should be considered an important tool for medical practitioners in the early detection and management of the condition.

The main objectives of this research are to provide a simple classifier model solution for predicting diabetes from easily measurable feature variables, with the intention to be later included in a newly designed healthcare system together with a suitable mobile application for prediction of several possible diseases usable to both patients and medical personal.

The research on related work using the same dataset [1] reveals various approaches and different results. Comparing the achieved results through the accuracy metric gives us a better understanding on which methodologies to use in order to have satisfying accuracy, while maintaining efficient and effective models that can be used in a mobile application. One of the first studies showed that without data preprocessing using Naive Bayes and Decision Tree, the achieved accuracy is 79.57% [2]. In another study, a deep learning approach is used, where the authors achieve accuracy of 98% [3]. In this paper, the authors show results for diabetes prediction function using a Naive Bayes classifier of 90%. The Deep Learning approach is superior in this setting, however it has some disadvantages such as requiring significant computing resources for training and also using the model in practice.

The paper is organized as follows. Section two presents the dataset and it characteristics. Section three describes the used data pre-processing techniques, while Section four presents the used ML techniques and an overview of the results. In section five we conclude the paper.

2 DATASET ANALYSIS

The dataset was created by the National Institute of Diabetes and Digestive and Kidney Diseases, from the United States National Institutes of Health [1]. The dataset was devised with the purpose of diagnosing whether a patient has diabetes based on various diagnostic measurements. The measurements come from female Pima Indians who are at least 21 years of age. The focus on this particular population was decided due to the higher diabetes occurrences noticed in practice.

Table 1: Dataset features.

Feature Variables	Description
Pregnancies	Number of pregnancies
Glucose	Glucose level in blood
BloodPressure	Blood pressure measurement
SkinThickness	Patients' Skin Thickness
Insulin	Insulin level in blood
BMI	Body mass index
DiabetesPedigreeFunct.	Diabetes percentage
Age	Patients Age
Outcome	1 - positive, 0 - negative

The dataset includes eight independent medical predictor variables and was properly labelled with a target dependent variable. The dataset consists of entries from 768 patients. The measured features are described in Table 1. Most of the features in Table 1 are self-explanatory. Only the feature Diabetes Pedigree Function (DPF) [4] is a mathematical formula used in genetics to estimate the likelihood of an individual developing diabetes. The DPF takes into account factors, such as family health history and age, which may influence the development of the disease.

The distribution of positive and negative outcomes in the dataset is presented in Figure 1.

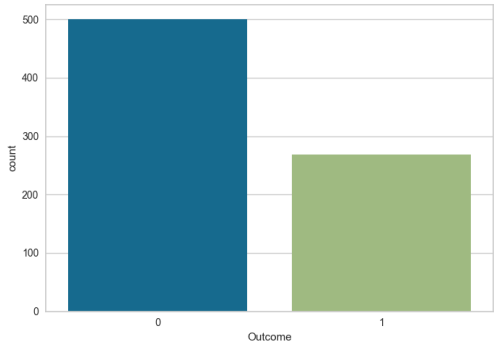


Figure 1: Distribution of positive and negative outcomes.

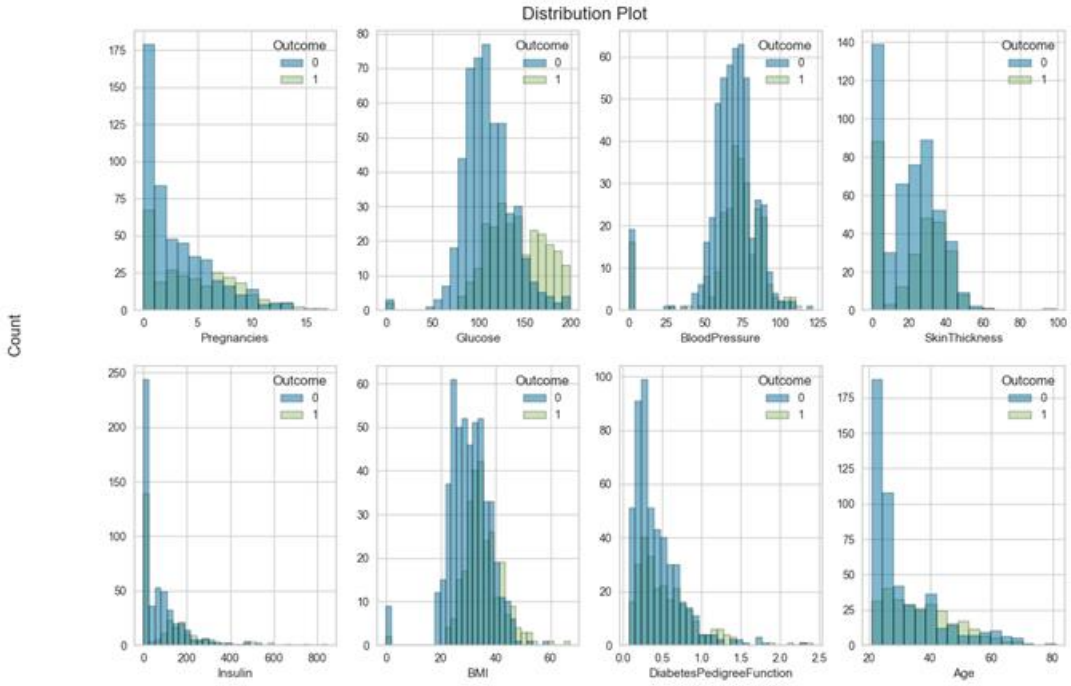


Figure 2: Ranges and distribution of dataset features.

It is easily noticeable there are more negative than positive outcomes in the dataset.

Figure 2 presents a summary of all the different feature ranges and distributions. We can also conclude that the features have varying scales.

2.1 Correlation Matrix

The correlation matrix of the dataset features provides insights into which features are most strongly correlated with the target variable, and which features are strongly correlated with each other (i.e., redundant). The correlation matrix of the used dataset (Figure 3) exhibits that our data feature variables have low correlation between each other.

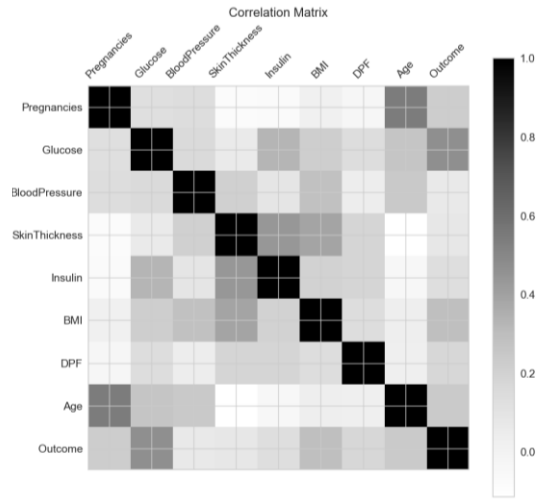


Figure 3: Correlation matrix of the dataset features.

The biggest correlation can be noticed between the age of the patients and their number of pregnancies; and the insulin level with their skin thickness. The glucose level has the biggest impact on the outcome, which is expected.

3 DATA PREPROCESSING

Handling null values in a machine learning dataset is an important pre-processing step, as many machine learning algorithms do not work well with missing values.

Furthermore, to make sure that the data is in a format that is suitable for the machine learning methods it is crucial to carry out additional pre-processing processes, such as feature scaling and normalization, even if there are no missing values. This step ensures that all features contribute equally to the analysis and prevents features with larger ranges from dominating the others.

3.1 Missing Values

The approach to handling missing values depends on the type and amount of missing data, as well as the specific problem we are trying to solve.

Common approaches to handling missing values in a machine learning datasets are dropping rows and interpolating missing values.

From the values analysis of the dataset it is clear that there are no null or missing values, which puts us in a strong position to move forward with the modelling process without having to deal with missing values.

3.2 Over-Sampling

Imbalanced classification is a common problem in machine learning where the target variable is unevenly distributed among the different classes. This can lead to a biased model that performs poorly in predicting the minority class. As presented in Figure 1, we are dealing with an imbalanced dataset.

Over-sampling is one of the techniques used to address this problem by creating synthetic samples of the minority class to balance the distribution of the target variable. Two most commonly used over-sampling techniques are Random Over-sampling [6] and Synthetic Minority Over-sampling Technique (SMOTE) [7].

Random Over-sampling involves duplicating random samples from the minority class in order to balance the distribution of the target variable. This increases the number of samples of the minority class, so that the classifier has a better chance of learning its pattern and making accurate predictions.

The SMOTE method is designed to balance the distribution of the target variable by generating synthetic samples of the minority class.

Both over-sampling methods will be applied for balancing the dataset target classes and the appropriate results are presented in the following sections.

3.3 Feature Standardization

The reason to use feature standardization is to ensure that the features have the same range and are not dominated by one feature with larger values. This allows the algorithm to give equal importance to all the features, rather than being biased towards features with larger values.

Two common techniques used in this research for feature standardization are Standard Scaler and Min Max Scaler. Standard Scaler rescales the input characteristics to give them a mean and standard deviation of 0 and 1, respectively. This is accomplished by dividing the result by the feature's standard deviation after deducting the mean of each characteristic from each data point. The standardized features that are produced have a standard deviation of one and a mean of zero. Min Max Scaler transforms the input features of a dataset to the range [0, 1]. This is done by subtracting the minimum value of each feature from each data point and dividing the result by the range (max - min) of the feature. The transformed features are then ready for input to a machine learning algorithm

In Figure 4 we present the distribution of the standardized features after applying Standard Scaler standardization.

4 CLASSIFICATION MODELS AND RESULTS

4.1 Classification Models

In this study we use six machine learning classification algorithms: Gaussian Naive Bayes [2], Random Forest Classifier [5], Extra Trees Classifier [8], Gradient Boosting Classifier [9] and XGB Classifier [10].

Gaussian Naive Bayes (GaussianNB) is a popular algorithm for classification problems which assumes that the distribution of the features is Gaussian (normal) and independent of each other. This makes it a good choice for problems where the features are continuous or real-valued, and the number of features is relatively small [2].

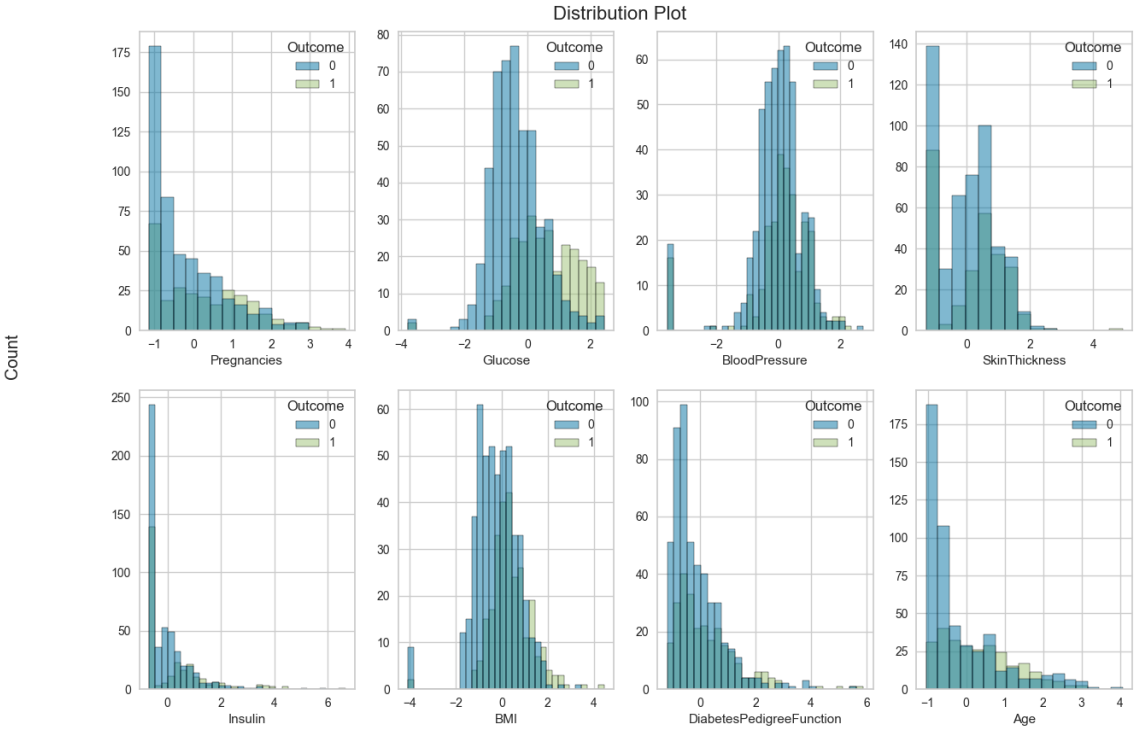


Figure 4: Ranges and distribution of dataset features after Standard Scaler standardization.

The Random Forest classifier (RandomForest) is usually a good choice for classification problems where the data has non-linear relationships, high-dimensional features, and categorical features and where overfitting is a concern [5].

Extra Trees Classifier (ExtraTrees) and Random Forest Classifier are both ensemble learning methods that use multiple decision trees to make predictions in classification problems. Extra Trees Classifier may be a better choice when dealing with many irrelevant features, imbalanced datasets, or where speed is a concern [5, 8].

Gradient Boosting Classifier (GradientBoosting) and XGB Classifier (XGB) are powerful algorithms that can handle a wide range of classification problems. If speed, scalability, or performance is a concern, XGB Classifier may be a better choice, while if ease of use and integration with scikit-learn is a concern, Gradient Boosting Classifier may be a better choice [11].

4.2 Results

Using the chosen models for training the data we can compare how different pre-processing methods perform on our dataset.

The accuracy is calculated as the ratio of the number of correct predictions to the total number of predictions made by the classifier, expressed in percentage. It is used as a simple and intuitive measure of how well the model is performing.

Table 2 presents the accuracies achieved for all the mentioned classifiers and both over-sampling techniques by applying Standard Scaler feature normalization to the data, as a more generally used method in pre-processing. From the results, we can come to the conclusion that for most of the models Random Over-sampling performs better than SMOTE, except for Gradient Boosting Classifier.

Table 2: Prediction accuracies with different over-sampling methods and standard scaler normalization.

Classifier	Rand. over-sampling	SMOTE
GaussianNB	0.76	0.74
RandomForest	0.86	0.85
ExtraTrees	0.88	0.82
GradientBoosting	0.82	0.83
XGB	0.86	0.81

Table 3, on the other hand, presents the accuracies achieved for all the classifiers and both feature normalization techniques by applying Random Over-sampling for solving the imbalanced classification problem in the dataset, again as it a more generally used method. We can conclude that both over-sampling methods work similarly, with slightly better results achieved when using Standard Scaler.

From the results in Table 1 and Table 2, we can conclude that classification models that are built on multiple decision trees give the best results of around 86% accuracy. This is because they avoid overfitting and capture complex relationships between the input features. Our further research will be focused on achieving better results with adjusting various parameters to improve the models' performance, such as the number of estimators and the depth of the trees.

Table 3: Prediction accuracies according to different feature standardization methods.

Classifier	Standard Scaler	Min Max Scaler
GaussianNB	0.76	0.76
RandomForest	0.86	0.85
ExtraTrees	0.88	0.88
GradientBoosting	0.82	0.83
XGB	0.86	0.85

Nevertheless, the results even from this preliminary research prove to have satisfying accuracy, while choosing relatively simple models that can be efficiently used in mobile or embedded applications on devices with limited resources.

5 CONCLUSION

The paper presented an analysis on the benchmark Diabetes dataset through comparison of six machine learning models that predict diabetes. It also presented comparison of different preprocessing techniques (under/over-sampling, feature standardization). The paper showed that ensemble learning methods, such as Extra Trees Classifier and Random Forest Classifier, along with appropriate data pre-processing techniques, can lead to at least 86% accuracy in diabetes prediction classification problems.

The work in this paper demonstrated that machine learning can be considered a valuable tool in the prediction of diseases, and in particular diabetes. The results of this study provide evidence that machine learning algorithms can be trained to identify patients

who are at risk of developing the condition, leading to early diagnosis and better outcomes.

Using Extra Trees Classifier, Random Forest Classifier, or other ensemble learning methods that use multiple decision trees, along with appropriate data pre-processing techniques, can often lead to high accuracy and performance in classification problems. Additionally, these ensemble models are based on Decision Trees, which can effectively be turned into simple decision rules and can run in real time even on devices with limited capacity and processing power.

ACKNOWLEDGEMENT

This work was partially supported by the WideHealth project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952279.

REFERENCES

[1] C.L. Newman, D.J. Blake, C.J. Merz, C.L. Blake, and C.J. Merz, "UCI repository of machine learning databases," 1998

[2] A. H. Jahromi and M. Taheri, "A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features," 2017 Artificial Intelligence and Signal Processing Conference (AISP), Shiraz, Iran, pp. 209-212, 2017, doi: 10.1109/AISP.2017.8324083.

[3] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset," Journal of diabetes and metabolic disorders vol. 19, pp. 391-403, Apr. 2020, doi: 10.1007/s40200-020-00520-5.

[4] M. Das, G. Bhattacharyya, R. Gong, and et all. "Determinants of gestational diabetes pedigree function for pima Indian females," Intern Med Open J. 2022, vol. 6(1), pp. 9-13, doi: 10.17140/IMOJ-6-121.

[5] A. Liaw and M. Wiener, "Classification and Regression by randomForest," R News, vol. 2, no. 3, pp. 18-22, 2002.

[6] M. Hayaty, M. Siti, and S. Ghufran, "Random and synthetic over-sampling approach to resolve data imbalance in classification," International Journal of Artificial Intelligence Research vol. 4.2, pp. 86-94, 2020.

[7] J. Wang, M. Xu, H. Wang and J. Zhang, "Classification of Imbalanced Data by Using the SMOTE Algorithm and Locally Linear Embedding," 2006 8th international Conference on Signal Processing, Guilin, China, 2006, doi: 10.1109/ICOSP.2006.345752.

[8] A. Géron, "Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow," O'Reilly Media, Inc., 2022.

[9] P. Sven, D. Ferran, F. A. Hamprecht, and B. Nadler, "Cost efficient gradient boosting," In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), New York, USA, pp.1550-1560, 2017.

[10] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," Proceedings of the 22nd ACM SigKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[11] C. Bentéjac, A. Csörgő, and G.A. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," Artificial Intelligence Review, vol. 54, 2021, pp.1937-1967, doi: 10.1007/s10462-020-09896-5.

Software Metrics Visualization

Vira Liubchenko
Department of Software Engineering, Odesa Polytechnic National University, Shevchenko Avenue 1, Odesa, Ukraine
lvv@op.edu.ua

Keywords: Metrics, Software, Visualization, Data, Diagrams, Analysis, Decision Making, Effectiveness.

Abstract: Software engineering is an empirical field of study. To support managerial and technical decision-making, the engineer needs numerical measures closely connected with different software metrics. Visual representation of numerical data improves the effectiveness of human data processing and shows insights that humans may miss. This paper aims to provide a systematic review of the approaches for software metrics visualization and define the possible recommendation for their use. The study is based on the literature review of the papers from two text collections – IEEE Xplore and ACM Digital Library – and the scientometric database Scopus. After merging and filtering, the final set of publications contains 16 papers. Our study showed that there were the metrics used significantly more often; among them are lines-of-code, cyclomatic complexity, coupling, and cohesion. We were not able to identify such leaders for visualization means. Instead, there was a tendency to combine different metrics on one chart or dashboard to provide the whole process picture. Based on the results of empirical studies reported in the literature, we offered an analysis of simple charts' properties and recommendations on their use for support decision-making in the software engineering process.

1 INTRODUCTION

It is well known that software engineering is an empirical field of study. Therefore, the numerical indicators are extremally essential and helpful for process monitoring and artifact evaluation.

Software metrics are an invaluable tool for measuring the progress and performance of software development projects. By providing a quantitative measure of the various aspects of a project, they can be used to inform decision-making, identify areas needing improvement, and track progress over time. Furthermore, software metrics are versatile and can be adapted to fit the needs of any given project. By carefully selecting the right metrics, software developers can gain insight into their project performance and make better-informed decisions.

However, each metric measures the distinctive feature or artifact of the development process. Therefore, the decision maker (e.g., architect, team lead, project manager) should simultaneously analyze the metrics set to get a realistic picture of the current situation in the development process.

To support decision-makers, we follow the idea of visualizing software metrics. Usually, the visualization can display a large amount of information in one chart, so the visualization can be used to present complex data dependencies typically found in software artifacts. As a result, the decision-maker can see the patterns, which allows the detection of known and unknown problems or opportunities in the software project.

This paper aims to provide a systematic review of the approaches for software metrics visualization and define the possible recommendation for their use.

To reach our goal, we focused on the two research questions:

RQ1: What are the most frequently used software metrics?

RQ2: What kind of visualization means are used most often?

The contributions of this study are twofold. First, we built a reference list of the most popular software metrics for visualization. Second, we gave a brief analysis of using visualization means.

The rest of the paper is structured as follows. First, in Section 2, we explain our study design and the methodology we followed. Section 3 outlines the publication trends of the papers we gathered. Section 4 presents the metrics and visualization means identified during our review and discusses our results. Some recommendations concerning chart use are considered in Section 5. Finally, we

close the study by summarizing the conclusions in Section 6.

2 STUDY DESIGN

The study aimed to answer RQ1 and RQ2 based on the relevant papers. The steps of the search and selection process we followed are depicted in Figure 1.

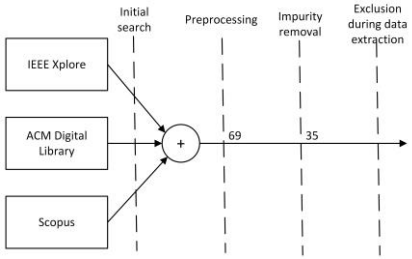


Figure 1: Papers selection methodology.

Initial search. We use two text collections – IEEE Xplore and ACM Digital Library – and the scientometric database Scopus as data sources for the study. The keywords for the queries were "software," "metrics," and "visualization" with connectors AND. We apply this query to all metadata at IEEE Xplore and ACM Digital Library and to "Article title, Abstract, Keywords" at Scopus.

Preprocessing. We formed the initial set of papers at this stage by merging the papers from three sources. Next, we removed the duplicates from the set. The issue of copies was caused by using the database Scopus; most of the papers found there had been collected in IEEE Xplore and ACM Digital Library. After the preprocessing, the papers set collected 69 papers.

Impurity removal. We performed an impurity removal per set of papers because many side papers matched the query but were irrelevant to the research purpose. We can distinguish irrelevant papers into three groups: concerned software structure visualization (tree structure hierarchy, software maps, trace messages, ontology-based visualization), presentation software statements and code coloring, and repository footprints. Some of these papers used the metrics but did not visualize them. We removed all these papers to make a coherent set of papers around our purpose. After removing them, we got 35 papers for the subsequent analysis.

Exclusion during data extraction. Working on the papers study, we found some papers that were not as relevant as expected. There were three types of such papers: presented the evolution of one framework and described the same concepts, introduced new metrics, and described an approach connected with building artificial entities (e.g., city, feather), which presented code and were based on metrics using. Therefore, we removed those papers from our final set. As one can see in the Reference Section, the final set of publications contains 16 papers.

3 PUBLICATION TRENDS

Now we present some publication trends found.

Figure 2 shows the distribution of publication years in the working set of papers obtained after impurity removal.

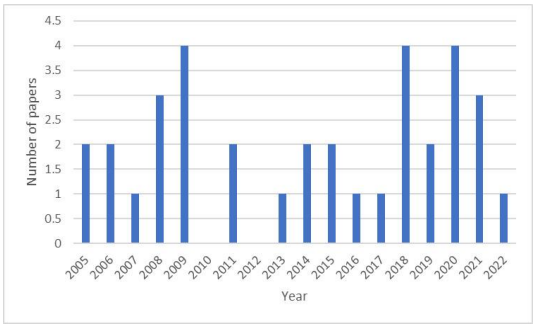


Figure 2: Distribution of publication years in the working set of papers.

For comparison, Figure 3 shows the distribution of publication years in the final set of publications.

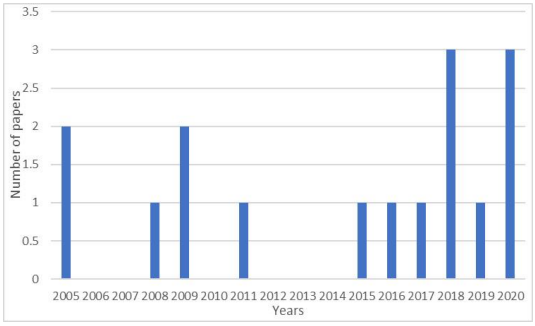


Figure 3: Distribution of publication years in the final set of papers.

We can see that the scientific interest in the topic was not very intensive. Although interest in empirical software engineering methods, based on metrics use, is consistently higher. In Figure 4, we

show the statistics of Google Trends on request "software metrics" for the last three years.

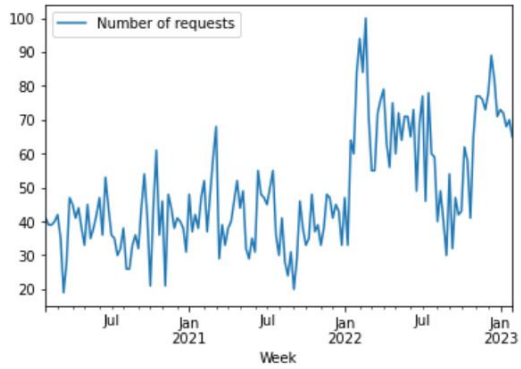


Figure 4: Interest over time to request "software metrics" (data from Google Trends).

However, we think the software metrics visualization issue was significantly underestimated. Different metrics reflect different aspects of the development process, and metrics understanding is a complex task for decision-makers. To get the whole picture, decision-makers should study many metrics simultaneously. Visualization can support and simplify the use and understanding of software metrics.

4 STUDY RESULTS

Our study showed that there were the metrics used significantly more often. We were not able to identify such leaders for visualization means. Let us present our results in detail.

4.1 Software Metrics

In the beginning, we should point out that often papers did not define the metrics precisely pointed only, e.g., "cohesion," when the figures were shown the concrete metrics. Some papers also presented a visualization approach in a metrics-independent way and did not mention concrete metrics. The distribution of metrics tackled in the papers is shown in Table 1.

Our study showed two prominent leaders of popularity: lines-of-code (LOC) and cyclomatic complexity. Many empirical studies demonstrated the dependency between LOC and the characteristics of the development process and software performance. Cyclomatic complexity is highly related to implementation and testing characteristics. We suppose these dependencies caused the high frequency of this metrics pair use.

Table 1: Software metrics tackled by the papers.

Metrics	Papers
Lines-of-code	[1] [2] [3] [4] [5] [6] [7] [8] [9]
McCabe's cyclomatic complexity	[2] [3] [5] [8] [9] [10]
Coupling and cohesion measures	[3] [4] [8] [10] [11] [12]
Fan-in, fan-out	[1] [3] [7] [8]
Depth of inheritance tree	[4] [8] [10]
Number of children	[4] [10]
Encapsulation measures	[4] [10]
Number of methods	[4]
Parameter number	[8]
Method length	[8]
Number of bugs	[6]
Passage rates of unit testing	[13]
Maintenance Index	[5]

We can see a focus on coupling and cohesion, as it is known that it is essential to keep low coupling and high cohesion in software design. But we should point out that both are generalizations of different metrics.

We can also see the attention to modularity connected with fan-in, fan-out, and encapsulation measures. As for coupling and cohesion, the generalization of different concrete metrics was used.

Most of the metrics relate to code features. However, researchers also pay attention to metrics specific to other stages of software engineering, e.g., testing and maintenance.

4.2 Visualization Means

Unlike software metrics, all papers clearly defined the type of visualization means they described. And unlike software metrics, there was no leader of popularity. The distribution of visualization means in the papers is shown in Table 2.

A large part of the working set of papers (the set after impurity removal) was dedicated to the issue of structure representations. Most often, the researchers used a dependency chart which reflects software structure with arc diagrams. To present the measures of software unit features, they used the size and color of nodes or provided the data in table form.

Table 2: Visualization means tackled by the papers.

Visualization means	Papers
TreeMap	[1] [6] [8]
Visual encoding	[4] [7]
Line chart	[5] [16]
Kiviat diagram	[14][15]
Rainbow-color mapped constant-size bars	[2]
Timeline and multilevel timeline	[3]
Metrics value distribution	[4]
HotSpot view	[8]
Scatter Plot	[8]
Toxicity Chart	[8]
Colored graph with variable size of nodes	[9]
Weighted digraph	[13]
Parallel coordinates plots	[10]
RadViz	[10]
Radial stacked bar chart	[11]
Zoomable circle packing graph	[12]
Bar chart	[16]
Punch card	[16]

A pretty exotic way of visualization was proposed in [2]. The approach combined the metric-lens technique (showing numerical values of metrics as colored bars) with UML diagrams (showing system structure). The authors claimed that the approach effectively helps understand the relations between metrics and structure at a finer level than the UML diagrams alone.

The significant issue concerning visualization is the need for simultaneous demonstration of different metrics. One of the proposed solutions was Kiviat diagrams. These diagrams are suited to present multivariate data, such as the feature vectors extracted from several source code releases and release history data. For similar purposes (visualizing source code metrics), Kiviat diagrams have also been used by related visualization approaches and tools.

The visual encoding approach combines different metrics as different parameters of geometry shapes. It supports the simplicity of the comparison between different software units. However, each researcher proposed their configuration. It does not look like a unified approach will be offered.

Simple line charts are reported not only as visualization means. For example, in [5], authors described line charts used for monitoring metrics from the time perspective and their prediction. In [10], the authors described using coordinates line plots multivariate software metrics (Figure 5) in parallel with the RadViz technique (Figure 6). Linking the two approaches was applied to detect outliers, which could indicate bad smells in software systems. They complement each other in identifying data patterns, clusters, and outliers.

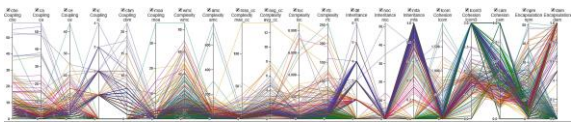


Figure 5: Parallel coordinates view [10].

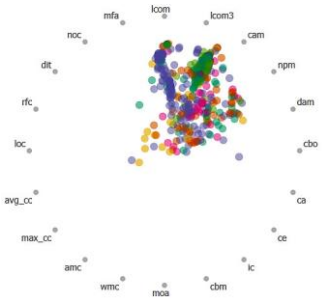


Figure 6: RadViz view of exploring noteworthy outlier patterns in detail concerning a focused set of metrics [10].

Finally, we should notice the idea of metrics dashboards in [8].

4.3 Discussion

There are many well-known software metrics. However, in most of the analyzed works, an attempt to combine some metrics was made. For example, two papers from the working set were devoted to studying the properties of indicators that combine several well-known metrics. As we already mentioned, many publications used visual encoding with geometric shapes, the parameters of which are determined by various metrics. In dependency charts, two metrics were also considered when designating vertices.

This confirms that a single metric is not valuable for analysis and decision-making because it reflects only one characteristic of an artifact or process. To provide a more comprehensive picture, multiple metrics need to be combined. It causes the search for different approaches to combine different metrics on one chart or dashboard.

Let us remember that analyzing software metrics is not an end; it supports engineering or management decisions. Software metrics are used in various applications, such as detecting low-quality code, finding design weaknesses, or estimating work progress. Therefore, the concept of a good visualization of software metrics depends mainly on their application.

In this context, it might be appropriate to consider structure diagrams in conjunction with metrics visualization. However, in our opinion, this approach does not cover all needs. Software types are becoming more and more diverse, which leads to the introduction of additional characteristics and, accordingly, additional metrics. For example, for software systems of artificial intelligence (AI-based software systems), such characteristics are interpretability, scalability, safety, fairness, staleness, etc. None of these metrics can be associated with the structural components of the system. However, for these characteristics, it can also be helpful to study them together and their behavior over time.

Software diversification leads to the introduction of new, specialized metrics requiring adequate visualization. For example, one uses Koopman Spaghetti Factor (*KSF*) for embedded software. *KSF* is calculated as

$$KSF = SCC + 5 \cdot Globals + SLOC / 20,$$

where *SCC* is the strict cyclomatic complexity, *Globals* is the global variables count, and *SLOC* is the number of non-comment source code lines.

As we can see, *KSF* combines three different metrics into a single metric. Therefore, it could be helpful while visualizing *KSF* to provide a granular representation of the three components of the metric.

Another example is model quality metrics in AI-based software systems. The common practice uses four metrics – accuracy, recall, precision, and F-score. Usually, their values for different models are shown in tabular form. It is necessary to compare the quadruples of metrics for different models to solve the problem of choosing the best or acceptable model. Under these conditions, using bar charts is more convenient than working with four numeric values from 0 to 1.

Summing up, we should note that metrics visualization is undoubtedly valuable for software engineering. Using metrics can make it possible to detect outliers and other deviations, predict the development of a process, and so on. Therefore, next, we provide an analysis of the charts' properties and recommendations on their use.

5 EFFECTIVE VISUALIZATION

Visualization capabilities of different charts may significantly affect the effectiveness of understanding and interpretation of the presented data by the decision-maker. Respectively, they also affect the quality and efficiency of decision-making. Note that software metrics are quantitative. Therefore, it is unnecessary to consider use cases for nominal and ordinal measurements.

In [17], there was pointed out that contextual information can serve as an essential input to developing and evaluating effective visualizations. The authors identified four principal contextual factors affecting visualization effectiveness: problem, stakeholder, purpose, and time.

The problem category concerns the problem situation to be supported and potential solutions. In the software metrics case, the problem is usually in the artifacts quality evaluation, efforts and bugs prediction, quality in use assurance, etc. It requires visualization to present the data clearly and comfortably. The stakeholder category involves any stakeholder-related aspects that affect the design of a visualization. In our case, the stakeholders are IT-friendly decision-makers who need easily readable and understandable information. The purpose category includes contextual information about what a visualization stakeholder is trying to achieve through applying the visualization in a particular domain. In our case, the purpose is to support decision-making by providing as much relevant information as can. The time category contains temporal information associated with decisional problems, stakeholders, and purposes. In our case, the interest in temporal information is restricted by prediction tasks only.

In the described context, using only the simplest charts, such as line charts, bar charts, scatterplots, or pie charts, is advisable. The focus of data presentation is mapping data values to graphical representations.

In [18], five recommendations were formulated. We analyzed their application for the software metrics visualization.

G1. Use bar charts for finding clusters. The bar charts have a better overall performance in terms of time, accuracy, and user preferences for finding clusters. Clustering is helpful, for example, for modularity evaluation. For such purposes, the alternative could be a scatterplot. But it restricts analysis to only two metrics simultaneously.

G2. Use line charts for finding correlations. The line charts performed better in terms of time, accuracy, and user preferences. As in the previous case, the possible alternative is the scatterplot, which restricts analysis by two metrics. Studying correlation with the line chart provides some additional effects; for example, [10] demonstrated

the identification of bad smells in code with line charts.

G3. Use scatterplots for finding anomalies. The scatterplots have high accuracy and speed and are highly preferred by users for this task. For purposes of software development, it could be helpful for outlier detection. The crucial important issue is the choice of metrics pair. We can apply dimensional reduction for scatterplot drawing. However, we should remember the importance of a clear interpretation of visual representation. If the decision-maker needs sense interpretation for the outliers, it could be impossible in the artificial feature space.

G4. Avoid line charts for tasks requiring readers to precisely identify a specific data point's value. The fact that the axes' values were drawn at uniform intervals makes it difficult to identify the value of a specific data point precisely. Anyway, such tasks are usually not relevant for decision-making in software engineering.

G5. Avoid using tables and pie charts for correlation tasks. As it was noticed in G2, for the correlation study, the more appropriate option is line charts.

In [19], a comparative study was realized for parallel coordinates, scatterplot matrices, and tabular visualization. The evaluation demonstrated that tabular visualization was familiar, accurate, and time-efficient for the retrieve value task. However, we recommend using the bar charts for the retrieve value task because it is very close to the clustering task in the software engineering context.

6 CONCLUSIONS

Decision-making in software engineering has become increasingly complex, spurring the need for effective decision-support tools. Numerical data visualization is a simple, fast, and effective way to enhance decision-making. In this paper, we conducted a literature analysis on software metrics visualization. As a result of the study, we identified the most frequent metrics used, the trend of simultaneous visualization of multiple metrics, and the criteria for choosing chart types.

Furthermore, the development of specialized software systems—such as embedded software and AI-based software—has created a demand for specialized metrics and, consequently, for visualization tools capable of making sense of these metrics. As such, it is of utmost importance to formalize and study the properties of various visualization tools for software metrics.

ACKNOWLEDGMENTS

This work was sponsored by the Federal Foreign Office through the Stibet I program of the German Academic Exchange Service (DAAD).

REFERENCES

[1] M. Balzer, O. Deussen, and C. Lewerentz, "Voronoi treemaps for the visualization of software metrics," in Proceedings of the ACM symposium on Software visualization (SoftVis '05), pp. 165-172, 2005.

[2] H. Byelas and A. Telea, "The metric lens: visualizing metrics and structure on software diagrams," in 15th Working Conference on Reverse Engineering, pp. 339-340, 2008.

[3] A. Gonzalez, R. Theron, A. Telea, and F. J. Garcia, "Combined visualization of structural and metric information for software evolution analysis," in Proceedings of the joint international and annual ERCIM workshops on Principles of software evolution (IWPSE) and software evolution (Evol) workshops (IWPSE-Evol '09), pp. 25-30, 2009.

[4] R. Francese, M. Risi, and G. Scanniello, "Enhancing software visualization with information retrieval," in 19th International Conference on Information Visualisation, pp. 189-194, 2015.

[5] B. Popović, A. Balota, and D. Strujić, "Visual representation of predictions in software development based on software metrics history data," in 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 352-357, 2016.

[6] T. Brunner and Z. Porkoláb, "Two-dimensional visualization of software metrics," in Proceedings of the Sixth Workshop on Software Quality Analysis, Monitoring, Improvement, and Applications, pp. 2:1-2:6, 2017.

[7] M. Alnabhan, A. Hammouri, M. Hammad, M. Atoum, and O. Al-Thnebat, "2D visualization for object-oriented software systems," in International Conference on Intelligent Systems and Computer Vision (ISCV), pp. 1-6, 2018.

[8] J. Slater, C. Anslow, J. Dietrich, and L. Merino, "CorpusVis - visualizing software metrics at scale," in Working Conference on Software Visualization (VISSOFT), pp. 99-109, 2019.

[9] G. Lacerda, F. Petrillo, and M. S. Pimenta, "DR-Tools: a suite of lightweight open-source tools to measure and visualize Java source code," in IEEE International Conference on Software Maintenance and Evolution (ICSME), pp. 802-805, 2020.

[10] H. Mumtaz, F. Beck, and D. Weiskopf, "Detecting bad smells in software systems with linked multivariate visualizations," in IEEE Working Conference on Software Visualization (VISSOFT), pp. 12-20, 2018.

[11] A. Yusuf and M. Hammad, “An approach to automatically measure and visualize class cohesion in object-oriented systems,” in International Conference on Decision Aid Sciences and Application (DASA), pp. 1174-1179, 2020.

[12] A. Yusuf and M. Hammad, “An automatic approach to measure and visualize coupling in object-oriented programs,” in International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), pp. 1-6, 2020.

[13] Y. Muto, K. Okano, and S. Kusumoto, “A visualization technique for unit testing and static checking with caller–callee relationships,” Journal of Convergence, vol. 2(2), pp. 1-8, 2011.

[14] M. Pinzger, H. Gall, M. Fischer, and M. Lanza, “Visualizing multiple evolution metrics,” in Proceedings of the ACM symposium on Software visualization (SoftVis '05), pp. 67-75, 2005.

[15] A. Kerren and I. Jusufi, “Novel visual representations for software metrics using 3D and animation,” in Software Engineering, J. Münch and P. Liggesmeyer, Hrsg. Bonn: Gesellschaft für Informatik e.V., 2009, pp. 147-154.

[16] S. R. Humayoun, S. M. Hasan, R. AlTarawneh, and A. Ebert, “Visualizing software hierarchy and metrics over releases,” in Proceedings of the International Conference on Advanced Visual Interfaces (AVI '18),” Article 40, pp. 1-5, 2018.

[17] X. Bai, D. White, and D. Sundaram, “Context adaptive visualization for effective business intelligence,” in 15th IEEE International Conference on Communication Technology, pp. 786-790, 2013.

[18] B. Saket, A. Endert, and Ç. Demiralp, “Task-Based Effectiveness of Basic Visualizations,” in IEEE Transactions on Visualization and Computer Graphics, vol. 25(7), pp. 2505-2512, 2019.

[19] G. J. Quadri and P. Rosen, “A survey of perception-based visualization studies by task,” in IEEE Transactions on Visualization and Computer Graphics, vol. 28(12), pp. 5026-5048, 2022.

Forecast of Yield of Major Crops in Ukraine in War Conditions 2022 Based on MODIS and Sentinel-2 Satellite Data

Nataliia Kussul[1,2,3], Sophia Drozd[1] and Hanna Yailymova [1,2]
[1]*Educational and Research Institute of Physics and Technology, Igor Sikorsky Kyiv Polytechnic Institute, Peremohy Avenue 37, Kyiv, Ukraine*
[2]*Department of Space Information Technologies and System, Space Research Institute National Academy of Science of Ukraine an State Space Agency of Ukraine, Glushkov Avenue 40, Kyiv, Ukraine*
[3]*Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*
nataliia.kussul@gmail.com, sofi.drozd.13@gmail.com, anna.yailymova@gmail.com

Keywords: Regression Analysis, Yield Forecasting, Random Forest, NDVI, MODIS, Sentinel-2.

Abstract: Ukraine was one of the main exporters of plant products. However, as a result of the aggression, the country's agriculture has suffered greatly, export volumes are decreasing, which may provoke a shortage of agricultural products on world markets. It is impossible to assess crop yield and forecast the harvest volume locally, as the collection of information has become difficult due to the active conduct of hostilities and the occupation of a large part of the territories. Therefore, it is necessary to use land remote sensing data to assess crop yield. In this research, we will build regression models based on a random forest for each region of Ukraine to estimate crop yield based on 16-day composites of the NDVI time series during the summer vegetation period from Sentinel-2 (10m) and MODIS (500m) satellites, involving in the calculation NDVI crop maps. The official yield of maize, sunflower, soybean, rapeseed, and wheat for the years 2016-2021 was used as training data. According to the results of the analysis, models based on NDVI from the MODIS satellite showed better accuracy (relative error within 8-18%), but models based on NDVI data from Sentinel-2 better described the variance of the predicted yield. During the research, we found a sharp drop in land productivity indicators compared to the productivity of 2021 for the territories of central, southern and eastern Ukraine. According to our estimates based on MODIS data, the average yield at the country level is expected to be 40.98 t/ha for wheat, 57.66 t/ha for maize, 23.57 t/ha for sunflower, 21.06 t/ha for soybeans, 21.15 t/ha for rapeseed. Estimates based on Sentinel-2 data: 43.22 t/ha for wheat, 71.93 t/ha for maize, 26.86 t/ha for sunflower, 22.94 t/ha for soybeans, 28.23 t/ha for rapeseed.

1 INTRODUCTION

For many years in a row, Ukraine has maintained the status of a leading exporter of agricultural products on world markets. According to the USDA, as of 2021, the state provided 46% of the world's sunflower oil exports, 9% of wheat exports, 17% of barley exports, and 12% of maize exports [1].

However, the country's agrarian well-being is now under threat. With the start of hostilities in February 2022, thousands of hectares of fields were damaged by rocket explosions and airstrikes [1]. The physical, chemical and biological characteristics of the soil were affected by shelling and explosions. Many territories, even those that managed to survive, remained unplowed, untreated from pests and weeds. Thus, the official information on sown areas throughout Ukraine as of the beginning of the year is no longer relevant. Previous forecasts regarding the volume of harvesting of the main agricultural crops are also not relevant, and it is very difficult to make new forecasts based on information from farmers and agronomists.

The yield forecasting process was previously based on a combination of data collected by local farmers directly on the ground with remote sensing data from satellites. Thanks to the combination and comparison of biophysical and satellite models [3], [3], [5], it was possible to achieve high accuracy of land productivity forecasts even in the early stages of vegetation [6] and to calculate the amount of harvest from certain land plots in advance. However, now, in wartime, the process of obtaining information about growth dynamics directly from local specialists has

become much more complicated. The problem of data availability is particularly acute in the occupied territories and territories of active hostilities in the east and south of Ukraine, which were the main suppliers of agricultural crops before the war.

Thus, the task of assessing the productivity of the country's agrarian sector has largely shifted to satellite monitoring and remote sensing data.

Many of the world's leading researchers have already begun to assess the volume of harvested crops in the conditions of war on the territory of Ukraine. Among them are NASA Harvest scientists [7]. Using data from Planet Labs satellites and the European Space Agency's Sentinel-2 mission, they set out to determine how the Russian-Ukrainian conflict is disrupting the global food system. To make the predictions, the scientists used vegetation index data from the Landsat-8 satellite, combining it with many factors of environmental conditions, such as precipitation, soil moisture and temperature during the growing season, obtained from the MERRA-2 NASA reanalysis data set. According to research results, according to NASA Harvest experts, the yield of winter wheat in Ukraine in 2021 will decrease to 4.1 t/ha instead of last year's 4.65 t/ha in 2021. Official data released by the Ministry of Agrarian Policy and Food of Ukraine as of October 2022 fully confirms NASA Harvest forecasts. However, these forecasts give an estimate of the yield at the level of the entire country, and do not allow tracing the situation in each region separately. In addition, the results of research are provided only on the yield of winter wheat, bypassing the spring agricultural crops, in the production of which Ukraine specializes, namely: maize, sunflower, soy, and rapeseed.

The Ministry of Agrarian Policy and Food of Ukraine provided its estimates of the yield of spring crops, but based mainly on data from controlled territories, and the question of the productivity of a large part of the country's cultivated areas remains open and needs an answer.

In this study, we want to estimate the yield of the main spring agricultural crops of Ukraine at the regional level according to the remote sensing data, covering all controlled and occupied agricultural lands of the state.

According to the experience of past studies, good yield predictions at a regional scale can be achieved using a vegetation index calculated from high spatial resolution satellite data such as Sentinel-2 in combination with Landsat missions [8], [9], [10]. However, when using the Sentinel-2 satellite, due to the low frequency of repeated visits of 5 days, there is often a problem of gloom and missing data. It is sometimes impossible to overcome this problem even with composites constructed from long time interval data. Therefore, yield forecasting is also practiced based on the data of a smaller spatial, but denser time step. For example, the Terra MODIS sensor [11], [12], with an average spatial resolution (from 250 m) and a daily frequency of repeated visits is successfully used to solve the problems of crop forecasting.

In this work, we decided to conduct experiments based on satellite data of both satellites - Sentinel-2 and MODIS, with further comparison and assessment of the reliability of the results of yields forecasts.

The purpose of the study is to obtain an accurate yields forecasting of Ukraine's land at the regional level for winter wheat and the main summer crops such as maize, sunflower, soybeans and rapeseed.

As the main method for conducting the research, we use regression analysis based on satellite data of the NDVI time series for the summer vegetation period and official statistical data on the yield of selected crops for previous years.

As a result of the successful conduct of the experiment, we will be able to provide accurate yield forecasts and estimate the volume of production of agricultural plant products in war conditions both in the controlled and non-controlled territories of Ukraine at the regional level.

2 DATA AND MATERIALS

To conduct an experiment on predicting crop yield as input parameters, we used time series of the NDVI index in 16-day composites for the summer growing season (June-August) 2016-2022 based on the data of the wide-area Sentinel mission satellites of the Copernicus program and Terra MODIS sensors.

The Multispectral Instrument S2 (MSI) provides high-resolution optical images from 13 spectral bands (10-20m), and captures atmospheric bands with a spatial resolution of 60m. For forecasting, we independently calculated NDVI from the products of the harmonized Sentinel-2 MSI, level A2, as the normalized difference of the red and infrared ranges.

Collections of 16-day Terra MODIS product composites (500m) already include ready-to-use NDVI products that generated based on near-field bands infrared and red range of each scene.

In preprocessing for input data analysis, NDVI binary masks were used for each studied culture, prepared based on classification maps of Ukraine [13], [14] (Figure 1).

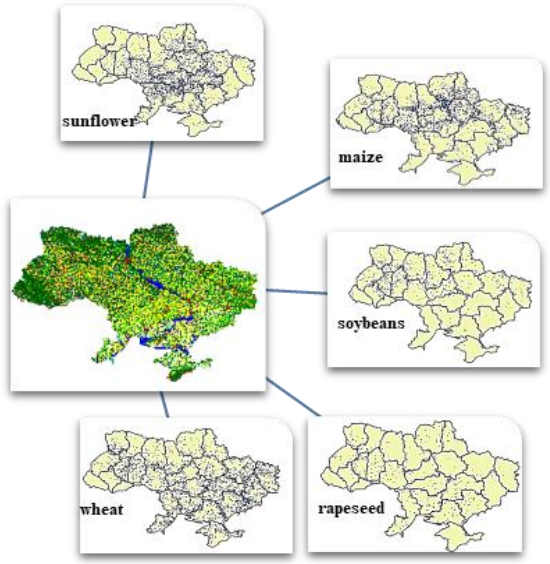


Figure 1: Binary maps of crops from the land cover classification map of Ukraine, 2022.

For each 16-day composite, an area-averaged NDVI was calculated, selecting for calculations only those vegetation index values that covered the target crop pixels. Due to this approach, the quality of input data was significantly improved. Thus, for each region of Ukraine, we obtained a time series that included six NDVI values, two for each summer month. In this time series, we determined the maximum and minimum for each region of Ukraine NDVI during the summer growing season. The total average NDVI for the entire study period was also calculated.

For each region, using correlation analysis, the period that most affects land productivity was also determined for each crop.

As a result, the maximum, average, minimum NDVI and the most correlated with land productivity NDVI for each region and each studied crop were selected as regressors.

3 METHODOLOGY

To predict crop yield for each region of Ukraine, regression models were built based on a random forest ensemble.

Crop yield was defined as the dependent variable, NDVI values as the independent variables.

The defining characteristic of the constructed regression models was the approximation reliability coefficient (1):

$$R^2 = 1 - \sum \frac{(y_t - y_f)^2}{(y_f - \Delta y_{ft})^2} . \quad (1)$$

where:

y_t – actual yield, according to official statistics,

y_f – predicted yield,

y_{ft} – the average value of all true values of crop yield.

Absolute (2), mean relative error (3), mean square error (4) and root mean square error (5) were the main metrics for assessing forecast accuracy.

$$\Delta y = y_t - y_f , \quad (2)$$

$$MRE = \frac{\sum \frac{\Delta y}{y_t}}{n}, \quad (3)$$

$$MSE = \frac{(\sum y_i - \bar{y}_i)^2}{n}, \quad (4)$$

$$RMSE = \sqrt{MSE} . \quad (5)$$

where:

n – the number of observations,

y_i – the i[th] observed value of actual yield,

$\bar{y}_i$ – the i[th] observed value of predicted yield,

Δy – absolute error,

MRE – mean relative error,

MSE – mean square error,

$RMSE$ – root mean square error.

When forecasting, we alternately took one year from 2016 to 2021 for validation, and others, by removing the validation year, for training the model, until we went through all the years. At each iteration, before changing the validation year, the accuracy and errors of the regression model were calculated.

The 2022 data were independently fed into the trained model to generate predictions.

4 EXPERIMENTAL RESULTS

4.1 Assessment of Model Accuracy

After performing the regression analysis, we conducted a general comparison of the accuracies of the models based on the NDVI data from the Sentinel-2 mission and the MODIS sensor, based on the regionally averaged results of the regression models when forecasting for 2021. Models built based on NDVI MODIS demonstrated better accuracy. In particular, MODIS-models showed smaller errors of MAE, RMSE, Relative error (Table 1) compared to models based on NDVI with Sentinel-2 (Table 2) for all crops except rapeseed (MODIS RMSE is 4.56 vs. 3.84 Sentinel-2 RMSE).

However, the MODIS models explained yield variances worse, with the agreement between

predicted and observed yield values for three of the five crops in 2021 (Figure 2).

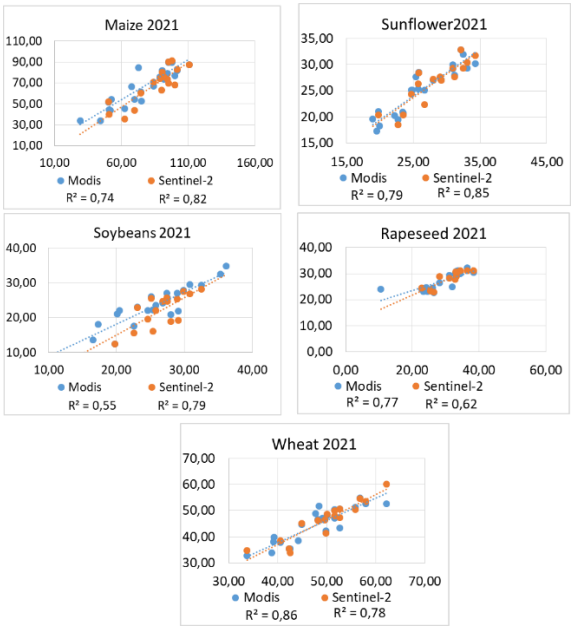


Figure 2: Comparative performance of Random Forests regression models based on Sentinel-2 NDVI and MODIS for 2021 yield test datasets. Plots are shown for five crops: maize, sunflower, soybean, rapeseed, wheat. Dashed lines represent the linear relationship between observations and predictions.

Table 1: Estimates of model errors based on NDVI data from MODIS satellites.

MODIS				
Crop	R2	MAE (c\ha)	RMSE (c\ha)	RE (%)
Maize	0,74	11,90	13,63	0,18
Sunflower	0,79	1,92	2,22	0,08
Soybeans	0,55	2,83	3,53	0,14
Rapeseed	0,77	3,38	4,56	0,16
Wheat	0,86	3,62	4,55	0,09

Table 2: Estimates of model errors based on NDVI data from Sentinel-2 satellites.

Sentinel-2				
Crop	R2	MAE (c\ha)	RMSE (c\ha)	RE (%)
Maize	0,8	17,14	19,17	0,22
Sunflower	0,9	2,15	2,51	0,09
Soybeans	0,8	4,54	5,45	0,2
Rapeseed	0,6	3,44	3,84	0,12
Wheat	0,8	3,54	4,33	0,09

The reliability coefficient of the approximation of forecasts based on MODIS data against forecasts based on Sentinel-2 data was 0.74 vs. 0.82 (maize), 0.79 vs. 0.85 (sunflower), 0.55 vs. 0.79 (soy). However, when forecasting the yield of wheat and rape, the coefficient of approximation of forecasts based on NDVI MSHDIS dominated (0.77 vs. 0.62 (rapeseed), 0.86 vs. 0.78 (wheat)).

In general, the forecasting results for each of the crops based on the data from both satellites were satisfactory with a maximum relative prediction error of 22% (Maize, models based on NDVI with Sentinel-2). Thus, the approach chosen by us is suitable for obtaining reliable yield forecasts and can be applied to calculate crop yield in 2022.

When constructing regression models based on Sentinel-2 satellites, we encountered the problem of missing data due to cloudiness caused by infrequent updates of Sentinel-2 data, even though we used a long time series. Therefore, when forecasting the harvest at the regional level for 2022, a greater weighting factor will be given to the results of model predictions based on NDVI from MODIS sensors.

4.2 Forecasting the Yield of the Land of Ukraine in 2022

The results of crop yield of each region of Ukraine, obtained by forecasting based on regression models based on NDVI MODIS and Sentinel-2, are provided in Table 3.

To assess the accuracy of the forecasts, we collected information on the average yield of the land of Ukraine and compared the official data with the yield averaged by region based on our predictions (Figure 3).

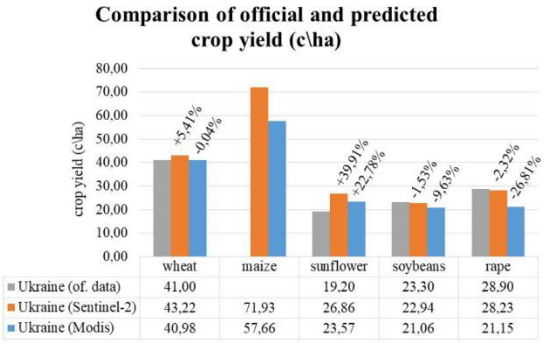


Figure 3: Comparison of official and predicted data on the average yield of land in Ukraine in 2022. The percentages in the chart show the relative deviation of the predicted crop yield from the MODIS and Sentinel-2 satellite data from the official statistics. There were no official data for maize at the time of the research.

Table 3: Predicted crop yield based on NDVI data from MODIS and Sentinel-2 satellites.

Region \ Crop	Wheat		Sunflower		Soybeans		Rapeseed		Maize	
	MOD	S2	MOD	S2	MOD	S2	MOD	S2	MOD	S2
Vinnytska	4,81	4,72	2,95	3,26	2,08	2,75	2,08	3,08	6,99	9,00
Volynska	4,55	4,46	2,69	2,90	2,69	2,64	2,69	3,23	8,42	8,98
Dnipropetrovska	3,27	NaN	1,98	NaN	1,52	NaN	1,53	NaN	2,96	NaN
Donetska	3,10	NaN	1,77	NaN	0,00	NaN	0,00	NaN	2,98	NaN
Zhytomyrska	4,31	4,30	2,46	2,46	2,29	2,64	2,31	2,84	6,83	6,80
Zakarpatska	3,34	3,24	2,03	2,15	2,25	2,31	2,28	NaN	5,10	5,27
Zaporizka	2,84	NaN	1,42	NaN	3,18	NaN	3,17	NaN	3,42	NaN
Ivano-Frankivska	5,02	4,68	2,79	2,77	2,41	2,76	2,42	NaN	6,51	6,78
Kyivska	3,90	4,06	2,49	2,78	1,64	2,16	1,62	2,69	5,62	7,77
Kirovohradska	3,53	4,43	2,03	2,50	1,18	1,20	1,22	2,40	3,99	4,95
Luhanska	3,25	NaN	1,74	NaN	0,00	NaN	0,00	NaN	3,15	NaN
Lvivska	4,97	4,78	2,73	2,73	2,66	2,70	2,69	2,95	7,31	7,33
Mykolaivska	3,01	3,11	1,61	1,86	1,05	1,35	1,06	2,08	3,23	3,82
Odeska	2,28	2,96	1,54	1,72	1,54	2,02	1,56	2,37	3,38	4,59
Poltavska	4,24	NaN	2,44	NaN	1,59	NaN	1,57	NaN	5,45	NaN
Rivnenska	4,69	4,72	2,69	2,77	2,56	2,58	2,55	NaN	7,35	8,39
Sumska	5,26	NaN	2,84	NaN	NaN	NaN	NaN	NaN	7,85	NaN
Ternopilska	5,53	5,36	2,82	3,23	2,71	2,92	2,74	3,31	8,17	9,48
Kharkivska	3,75	NaN	2,67	NaN	1,24	NaN	1,33	NaN	4,04	NaN
Khersonska	3,33	3,85	1,49	NaN	3,29	NaN	3,29	NaN	6,82	7,95
Khmelnytska	5,74	5,10	3,41	3,45	2,77	2,66	2,76	3,11	8,95	9,29
Cherkaska	4,16	4,58	2,56	2,82	1,59	1,67	1,53	2,86	5,65	6,84
Chernivetska	5,04	4,09	2,72	2,70	1,90	2,10	1,93	NaN	6,17	5,63
Chernihivska	4,44	5,04	2,70	2,89	2,17	2,24	2,17	2,97	8,04	9,39
Ukraine (mean)	4,10	4,32	2,36	2,69	1,93	2,29	1,94	2,82	5,77	7,19

According to the results of the comparisons, the largest forecast error was recorded in the assessment of sunflower yield. The predicted values exceed the official indicators by 22.78% (MODIS) and 39.91% (Sentinel-2). This can be explained by an incomplete harvest. The smallest error, which was 0.02 t/ha, was found when forecasting wheat yield based on MODIS data (Sentinel-2 – 2.22 t/ha). That is, the regression model absolutely confirmed the official data. Good results were found when predicting the yield of rapeseed. The error of the model based on Sentinel-2ya-2 data is 0.67 t/ha, which is about 2.32% yield. The MODIS-based model, on the contrary, had a high forecast error, which was -26.81% of the official data. Deviations from official data in predicting soybean yields were -1.53% and -9.63% for Sentinel-2 and MODIS-based NDVI models, respectively.

It was not possible to assess the accuracy of the maize yield forecast due to the lack of official statistics.

After comparing the predicted yields with the official data, maps of land productivity dynamics were constructed using the MODIS NDVI predictions to estimate the change in yields relative to the 2021 figures (Figure 4).

As can be seen from the maps, for each crop, the greatest yield decline is expected in the central, eastern and northern parts of the country. An extreme drop in yield is expected in Odesa region and is -46.09% for maize, -45.06% for soybeans, -34.48% for sunflower, -38.57% for rapeseed and -43.72% for wheat. In general, yield declines are observed for every crop in almost every region.

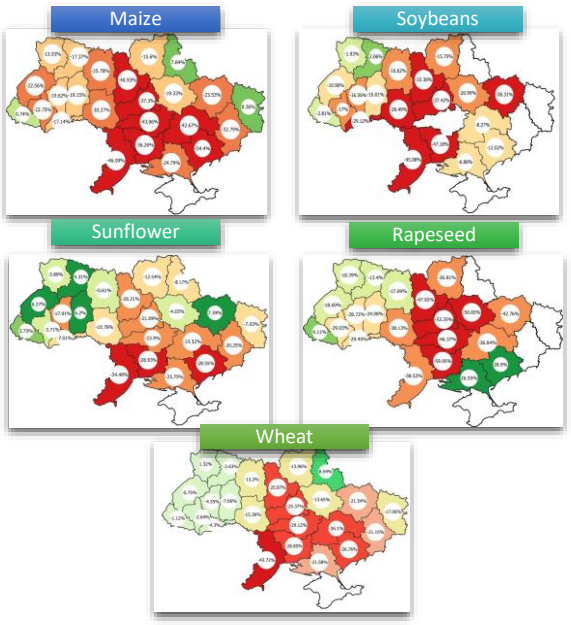


Figure 4: Assessing crop yield dynamics against 2021 yield data.

The exception, according to forecasts, is rapeseed in the Zaporizhzhia and Kherson regions (+28.9% and 26.55% productivity, respectively). Weak positive trends are also observed for the western part of Ukraine, which is certainly explained by the smaller number of active hostilities in the regions.

5 CONCLUSIONS

This study evaluated the effectiveness of crop yield forecasting based on the NDVI time series from MODIS and Sentinel-2 satellites, and forecasted crop yield of Ukraine at the regional level for 2022.

During the experiment, when validating the models on the 2021 test data, it was found that predictions based on NDVI from the Sentinel-2 satellite better describe the variance when comparing predictions with actual yield values, but models based on MODIS sensor data have better accuracy rates.

Although Sentinel-2 has better spatial resolution, due to the wide step of temporal updating of the data, it is difficult to correctly estimate the yield in certain areas, which were overcast during the survey period.

According to our estimates based on MODIS data, the average yield at the country level is expected to be 40.98 t/ha for wheat, 57.66 t/ha for maize, 23.57 t/ha for sunflower, 21.06 for soy, 21.15 t/ha for rapeseed. Estimated based on dataSentinel-2: 43.22 t/ha for wheat, 71.93 t/ha for maize, 26.86 t/ha for sunflower, 22.94 t/ha for soybean, 28.23 t/ha for rapeseed. These indicators are significantly lower than the indicators of crop yield in 2021.

Thus, according to the results of forecasts based on regression models for 2022, we found a sharp deterioration of crop yield in most regions of Ukraine, which is especially pronounced in central Ukraine, in the east and in the south. According to our estimates, the greatest drop in yield is expected in Odesa region (-46.09% for maize, -45.06% for soybeans, -34.48% for sunflower, -38.57% for rapeseed and -43.72% for wheat), as well as in Mykolaiv, Zaporizhzhia, Kherson, Kyiv, Kirovohrad, and Cherkasy. The negative dynamics of yield in these territories is explained by active military operations there, as a result of which large areas of agricultural fields were damaged and the territories were occupied. Thus, the war had an extremely negative impact on crop production in Ukraine, provoking a sharp deterioration in land productivity and, accordingly, decrease in the yield of agricultural crops.

ACKNOWLEDGMENTS

The authors acknowledge the funding support from the European Commission through the joint World Bank/EU project ‘Supporting Transparent Land Governance in Ukraine’ (ENI/2017/387–093 and ENI/2020/418–654), the project “Deep learning methods and models for applied problems of satellite monitoring” (2020.02/0292) within the competition of the National Research Foundation of Ukraine “Support research of leading and young scientists” from the State budget, Horizon 2020 project "Information technologies of geospatial analysis of the development of rural areas and communities", and Grant of Philipp Schwartz-Initiative Alexander von Humboldt Foundation.

REFERENCES

[1] Svit perebuvaie na pochatku prodovolchoi kryzy, abo yakyi vrozhai vanhuie NASA Harvest Ukraini [The globe is at the beginning of food crisis or what yield NASA Harvest forecasts for Ukraine], 2022, [Online]. Available: https://kurkul.com/ spetsproekty/1338-svit-perebuvaye-na-pochatkoviy-stadiyi-prodovolchoyi-krizi-yakiy-vrojay-vanguye-nasa-harvest-ukrayini.

[2] D. Rawtani, G. Gupta, N. Khatri, P. K. Rao, and C. M. Hussain, “Environmental damages due to war in Ukraine: A perspective,” Sci. Total Environ., vol. 850, no. 157932, p. 157932, 2022, doi: 10.1016/j.scitotenv.2022.157932.

[3] A. Kolotii and et al., “Comparison of biophysical and satellite predictors for wheat yield forecasting in Ukraine,” The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XL-7/W3, pp. 39-44, 2015, doi: 10.5194/isprsarchives-xl-7-w3-39-2015.

[4] F. Kogan and et al., “Winter wheat yield forecasting: A comparative analysis of results of regression and biophysical models,” J. Autom. Inf. Sci., vol. 45, no. 6, pp. 68-81, 2013, doi: 10.1615/JAutomat InfScien.v45.i6.70.

[5] F. Kogan and et al., “Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models,” Int. J. Appl. Earth Obs. Geoinf., vol. 23, pp. 192–203, 2013, doi: 10.1016/j.jag.2013.01.002.

[6] M. E. Holzman and R. E. Rivas, “Early maize yield forecasting from remotely sensed temperature/ vegetation index measurements,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 9, no. 1, pp. 507-519, 2016, doi: 10.1109/JSTARS.2015. 2504262.

[7] T. Ben Hassen and H. El Bilali, “Impacts of the Russia-Ukraine war on global food security: Towards more sustainable and resilient food systems?,” Foods, vol. 11, no. 15, p. 2301, 2022, doi: 10.3390/foods 11152301.

[8] S. Skakun and et al., “Winter wheat yield assessment from Landsat 8 and Sentinel-2 data: Incorporating surface reflectance, through phenological fitting, into regression yield models,” Remote Sens. (Basel), vol. 11, no. 15, p. 1768, 2019, doi: 10.3390/ rs11151768.

[9] S. Skakun and et al., “Winter wheat yield assessment using Landsat 8 and sentinel-2 data,” in IGARSS 2018 - IEEE International Geoscience and Remote Sensing Symposium, 2018, doi: 10.1109/IGARSS.2018. 8519134.

[10] S. Skakun, E. Vermote, J.-C. Roger, and B. Franch, “Combined use of Landsat-8 and sentinel-2A images for winter crop mapping and winter wheat yield assessment at regional scale,” AIMS Geosci., vol. 3, no. 2, pp. 163-186, 2017, doi: 10.3934/geosci. 2017.2.163.

[11] I. Becker-Reshef, E. Vermote, M. Lindeman, and C. Justice, “A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data,” Remote Sens. Environ., vol. 114, no. 6, pp. 1312-1323, 2010, doi: 10.1016/j.rse.2010. 01.010.

[12] O. Kussul, N. Kussul, S. Skakun, and et al., "Assessment of relative efficiency of using MODIS data to winter wheat yield forecasting in Ukraine," 2013 IEEE International Geoscience and Remote Sensing Symposium, 2013, doi: 10.1109/IGARSS. 2013. 6723516.

[13] A. Shelestov, M. Lavreniuk, V. Vasiliev, and et al., "Cloud approach to automated crop classification using Sentinel-1 imagery," IEEE Transactions on Big Data, 2019, vol. 6. no. 3, pp. 572-582, doi: 10.1109/TBDATA.2019.2940237.

[14] N. Kussul, L. Mykola, A. Shelestov, and S. Skakun, "Crop inventory at regional scale in Ukraine: Developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery," European Journal of Remote Sensing, vol. 51(1), pp. 627-636, doi: 10.1080/22797254.2018. 1454265.

War Damage Detection Based on Satellite Data

Andrii Shelestov[1,2], Sophia Drozd[1], Polina Mikava[1], Illia Barabash[1] and Hanna Yailymova[1,2]
[1]*Educational and Research Institute of Physics and Technology, Igor Sikorsky Kyiv Polytechnic Institute, Peremohy Avenue 37, Kyiv, Ukraine*
[2]*Department of Space Information Technologies and System, Space Research Institute National Academy of Science of Ukraine an State Space Agency of Ukraine, Glushkov Avenue 40, Kyiv, Ukraine*
andrii.shelestov@gmail.com, sofi.drozd.13@gmail.com, geor.polina@gmail.com, ilbar-ipt21@lll.kpi.ua, anna.yailymova@gmail.com

Keywords: Sentinel-2, Relative Difference of NDVI, Damaged Agricultural Fields Due to War, Validation, ACLED.

Abstract: As a result of the resolution of the armed military conflict on the territory of Ukraine on February 24, 2022, the agricultural infrastructure of the latter was marked by large-scale destruction. Thousands of hectares of fields, the harvest from which previously provided both domestic and world needs, were mined, destroyed, damaged by artillery shelling, explosions and movements of military equipment. To restore the affected areas to ensure food security of Ukraine and the world, the state government, with the support of international organizations, must correctly distribute financial resources between affected landowners and farmers. For this, there is a need for accurate identification of war-affected territories. This task can be effectively performed using remote sensing data. In this work, damage to agricultural fields due to military operations is searched for by calculating the relative difference of the vegetation indices based on Sentinel-2 satellite data. Cloud-free composites of normalized difference vegetation index (NDVI) are compared for the nearest period before and after active hostilities in a specific area (dates and locations are obtained from the ACLED source). Pixels whose relative difference exceeds a given threshold are considered damaged. The survey of the country's territories was conducted from February 24 to September 25, 2022, dividing the dates into biweekly periods. According to the results of the research, such damage to agricultural fields as craters from explosions and shelling, traces of machinery, burnt fields, etc., were found. The relative difference between the minimum and average values of vegetation indices in the affected areas averaged 25% versus 15% for the minimum period before and after the lesion. The detected damaged areas were validated using ACLED data. It was determined that more than 50% of the total number of areas identified as damaged were located within a radius of up to 5 km from the zone of combat activities.

1 INTRODUCTION

After the invasion of the enemy troops of the Russian Federation on the territory of Ukraine on February 24, 2022, a large part of the country was damaged, including not only cities and important infrastructure facilities, but also agricultural fields. In the territories where active hostilities are taking place, most of the agricultural land is mined and unsuitable for growing products. This was a big blow both for the agro-industrial complex of Ukraine and caused a shortage of food products on the European markets, because Ukraine was one of the leading exporters of agricultural products to Europe for decades in a row.

To mitigate the negative effects of the war on the agricultural sector and ensure food security, the government of Ukraine and the international community decided to allocate funds for the restoration of damaged agricultural lands. To do this, farmers whose lands were affected were asked to fill out questionnaires, and tens of thousands of landowners applied for compensation. However, it is not known for sure which areas were exactly affected by shelling and explosions, because not all farmers accurately declared the amount of their losses and the area of damage, and some did not declare at all. In order to properly distribute the allocated funds, the government needs reliable information and a thorough inspection of the condition of the fields. It is physically impossible to carry out this task in the usual way, by carrying out an actual survey of the territories, due to the excessively large area of land, constant hostilities

and occupation. In addition, this is an irrational use of human and material resources.

However, it is possible to solve the problem of detecting damaged areas using remote sensing data [1], [2]. Satellite data allow to identify land use and cover/land use changes using machine learning models [3]. The main advantage of providing remote analysis is safety. In article [4] authors created a framework that can be used to estimate the damage during the ongoing war and get information for restoration planning. As data were used Surface reflectance (SR) images from the Landsat-8 Operational Land Imager and several random forest algorithms were used for extracting needed information. As feature variables are used several indexes such as NDVI, MNDWI, NDBI, BSI. In work [5] authors tried assess damages using several types of optical images and SAR data. Multiple cities in the Kyiv region were chosen as study areas. All process of investigation contains 4 stages for both Sentinel-1 and Sentinel-2 data [6].

Also, according to the previous experience of world scientists, remote sensing data from satellites of medium and high spatial resolution are successfully used to analyze time series of vegetation indexes [7], [8], detect land degradation [9], monitor land use [10], predict land yield and productivity, records of fires, drought [8] and floods, etc. Satellite missions that provide open data for tracking over the territory of Ukraine are Landsat-8 (30m) and Sentinel-2 with a spatial resolution of 10m (data available since 2014).

Using indices calculated from satellite data, in particular the NDVI vegetation index, it is possible to detect abnormal changes in land cover that have occurred within a short period of time [11], and record a sharp decrease in plant growth rates or a drop in the amount of vegetation. These anomalies are usually provoked by weather phenomena, such as hail [7], [12], for example. However, for Ukraine, the changes caused by military actions can be detected by analyzing the time series of vegetation indexes, in particular NDVI [13]. The application of the technique of comparing NDVI in a fixed territory on specific specified dates can help to build an automatic recognizer capable of pixel-by-pixel identification of craters from bomb explosions, eruptions from artillery shelling, tracks from tank tracks, other military equipment, etc [14].

Thus, the topic of this work is the development of an automated tool for the identification of agricultural fields damaged as a result of military actions based on remote sensing data for the analysis of NDVI and spectral channels.

The main purpose of the paper is the accurate detection of damaged areas on the agricultural fields at the pixel level and an adequate assessment of the severity of damage.

The identifier developed in this way will make it possible to adequately assess the consequences of the war in agricultural area and will help to correctly allocate the budget for the restoration of the functioning of the agrarian industrial complex of Ukraine.

2 DATA AND MATERIALS

Data from a wide-area mission with high resolution (10m) and multispectral imagery supporting the Copernicus Land Monitoring study, Harmonized Sentinel-2 MSI level A2 were used to calculate the vegetation index and analyze spectral channels. The images of the collection contain 16 spectral bands. For the study, vegetation index NDVI was calculated, as well as four bands (Red, Green, Blue, and NIR).

Due to overcast areas during numerous satellite surveys, in order to obtain reliable data, there was a need to clear images from clouds and build composites in the shortest possible time.

An open source of information - the Armed Conflict Locations and Events Data (ACLED) project [15] was used to select appropriate dates for the construction of composites before and after potential field damage in order to detect damage to territories by comparing the values of vegetation indexes before and after damage. ACLED collects information on the dates, locations and types of all recorded events of political violence and protests worldwide, including information about the territories of Ukraine where active hostilities are currently taking place. Starting from February 24, the time was divided into biweekly intervals (periods), for each of which the damaged areas were determined (Figure 1). The parcels deliniation polygons developed by the Sinergise company for Ukraine within the EO4UA [16] initiative was used as field contours.

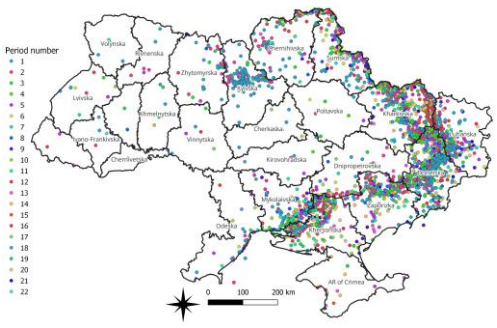


Figure 1: ACLED information about military action in Ukraine for 15 biweekly periods from 24th of February.

Thus, we selected the closest possible time period before the military attack, the date and location of which was provided by ACLED, and the closest period after the military attack to construct a relatively cloud-free satellite data.

3 METHODOLOGY

3.1 Damaged Detection Based on Comparison of NDVI Distribution

The method of detecting fields damaged by military actions on the territory of Ukraine is based on a pixel-by-pixel comparison of NDVI values during the narrowest time interval before and after the impact. NDVI – is a commonly used vegetation index, calculated by near infrared (NIR) and infrared (RED) bands of satellite observations with the (1):

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (1)$$

After obtaining the indicators of the vegetation index according to two composites - before and after military activities on a specific territory on a regional scale - the relative difference of NDVI in percentages is calculated according to the following (2):

$$dNDVI_t = \frac{NDVI_{t+1}-NDVI_t}{NDVI_t} * 100\%, \quad (2)$$

where: $dNDVI_t$ - is the relative percentage difference of NDVI at point t; $NDVI_t$– NDVI value before potential damage to the territory at point t; $NDVI_{t+1}$- NDVI value after potential damage to the territory at point t.

To calculate the NDVI only within the boundaries of agricultural crops the land cover classification map of 2022 was used. The next step is to set the maximum permissible (threshold) value of the relative difference of NDVI in order to filter out normal changes in vegetation cover and detect anomalies. To remove non-military NDVI changes, we thresholded within a field the number of pixels with a large NDVI drop of 60% of the field size using vector field contours. The previously given vector contours of the fields were reduced by 2 pixels along the perimeter inside, in order to avoid fixing the changes of NDVI in the territories between the fields.

3.2 Damaged Detection Based on spectral Values Distribution

Comparing pre-event and post-event satellite data shows good results in damage monitoring, but there are often situations where no pre-event satellite data is available, or it is all covered by clouds. For such cases, another algorithm has been developed that uses only one image, or rather, its spectral channels separately.

This algorithm works at the level of each individual field within the contours created by the Sinergise company. To begin with, the NDVI indicator is calculated within the contour of the field to determine the state of vegetation according the (1). In case the vegetation is low (NDVI < 0.3), Green and Blue spectral channels are used to damaged territories detection. Otherwise, the Green and NIR channels are used. Further, within each field, the distribution of the values of the corresponding spectral channels is analyzed and those of them that are not statistically average are cut off. In particular, for fields with high vegetation, the following (3) is used:

$$Damaged_{Hight\_vegetation} = \begin{cases} Value_{Green} \le \mu - 2\sigma \\ Value_{Blue} \le \mu - 2\sigma \end{cases}, \quad (3)$$

and for fields with low vegetation, the following (4) is used:

$$Damaged_{Low\_vegetation} = \begin{cases} Value_{Green} \ge \mu + 1.5\sigma \\ Value_{NIR} \ge \mu + 1.5\sigma \end{cases}, \quad (4)$$

where μ - mean channel's value at field level, σ - standard deviation of channel's value at field level. The general scheme of the developed algorithm is shown in the Figure 2.

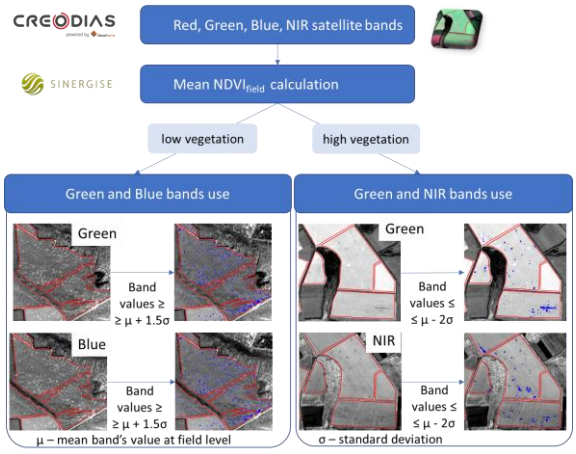


Figure 2: The general scheme for finding damage from bombs and missiles.

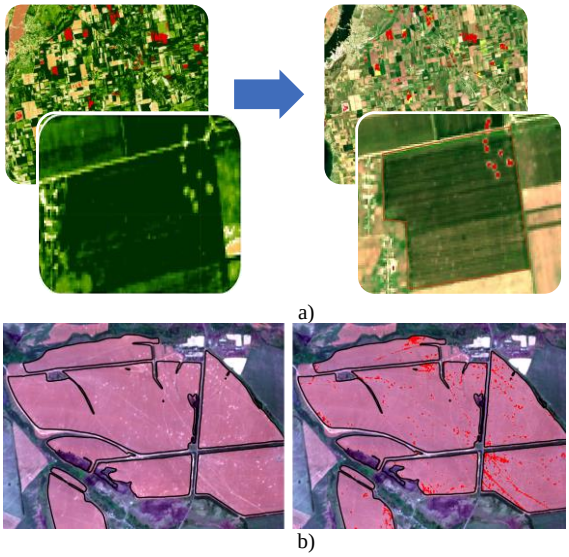
To confirm the assumptions about the possible damage to the territories, histograms of the distribution of NDVI in the cadastral boundaries of the target field are constructed according to the available composites in order to estimate the dispersion of the vegetation index and the deviations

of the minimum values of NDVI from the average value within the field before and after damage.

4 THE RESULTS

4.1 Damaged Fields Detection

With the help of calculations of the relative difference of the NDVI, we were able to identify damage from artillery fire (Figure 3a), as well as traces of the movement of equipment, burned fields and other numerous damages (Figure 3b).



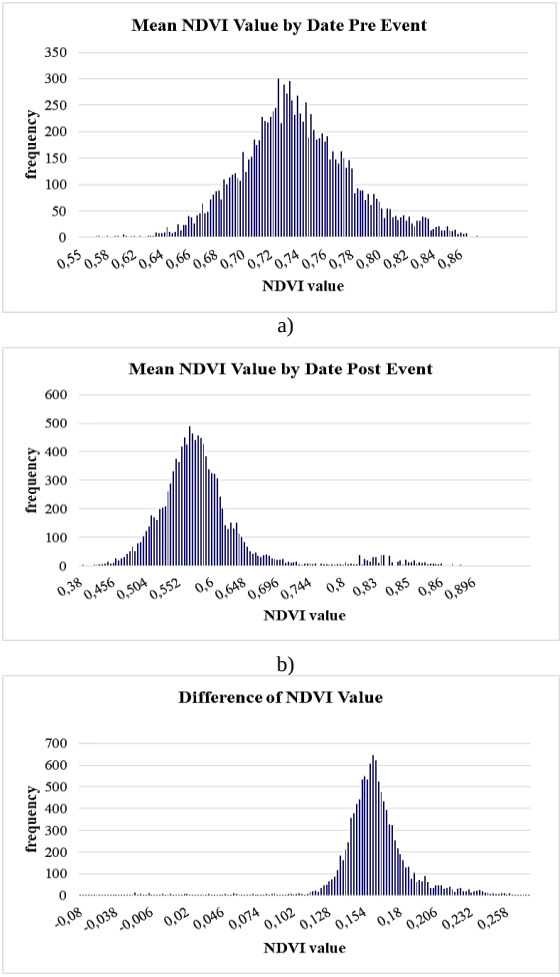
a)

b)

Figure 3: a) Damages detected based on relative difference of NDVI, Kherson region, (May 9 - May 2022). b) Detected damaged fields based on spectral channels, Donetsk region (02 July 2022).

To check the distribution of NDVI within the identified as damaged fields, histograms were constructed, an example of which is shown in Figure 4. The relative difference of the minimum and average NDVI in the affected field from the example (Figure 4b) was about 25% against 17% in the same field before the injury. The relative median difference between the minimum value of affected and uninjured areas on the image in which the damage was recorded was 13% on average (Figure 4c). In general, a similar situation was observed in other fields.

For comparison with the relative difference of NDVI in the unaffected area, we constructed a histogram of the relative difference of NDVI in the nearest neighboring undamaged field for the same date (Figure 5).



a)

b)

c)

Figure 4: Distribution of NDVI within the field. a) before the event of potential damage (June 1 - June 5); b) after the event of potential damage (June 6 - June 8) damage.; c) The difference in the distribution of NDVI before and after damage.

As can be seen in Figure 5, the largest difference of NDVI on an intact field does not exceed 0.07, while on a damaged field it exceeds 0.26.

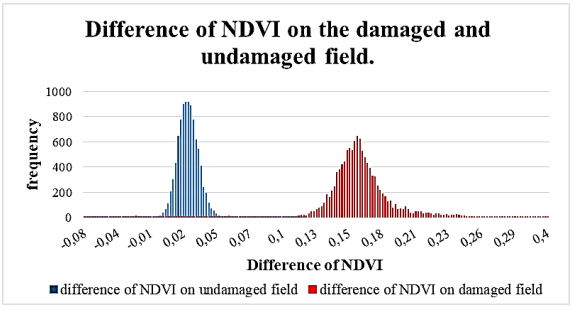


Figure 5: The difference of NDVI on the damaged and undamaged field.

4.2 Comparison of Automatically Recognized Damages and Damages Detected by the Experts

To validate the method of automatic field recognition, we conducted a visual inspection of the damaged areas, using data on the areas and dates of active hostilities for orientation from the ACLED source. To improve the quality of the verification, we will choose the 8th period of the war (6-19 June) as the period of active vegetation and the Donetsk region as the region of active hostilities.

According to the results of a comparison of automatic and "manual" methods for recognizing field damage (Figure 6), it was found that the method based on the calculation of the relative difference of the NDVI identifies many more fields than can be seen on a three-channel satellite image (Figure 7) (on cloudless composites, 2143 fields were automatically identified against 376 fields visually assessed as damaged).

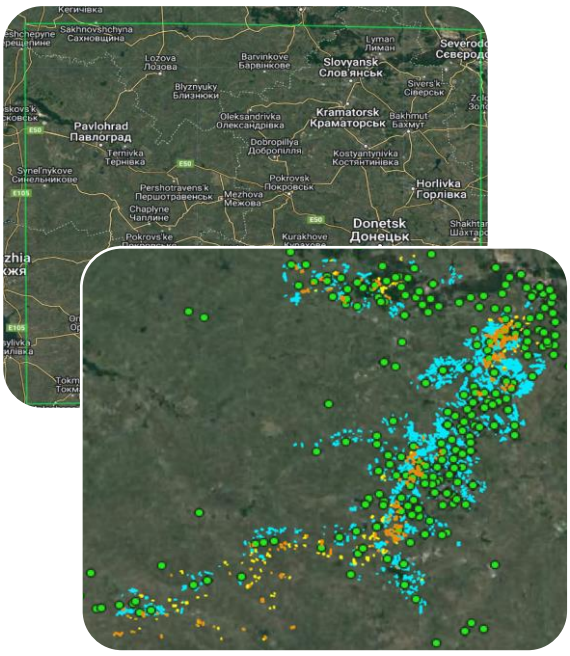


Figure 6: Comparison of fields identified as damaged by an automated method and fields marked as damaged by a "manual" method. Donetsk region, 8th period of the war (June 6 – June 19). Yellow polygons are manually marked fields. Blue polygons are fields recognized as damaged based on data on the relative difference of NDVI, Brown polygons are fields recognized as damaged by "manual" and automatic methods. Green dots show combat zones.

At the same time, the identification results coincided only for 193 fields, which is about 48% of visually recognized as damaged fields and 9% of the number of fields identified as damaged based on the relative difference of NDVI. Such a large difference in results may be explained by clouds, normal seasonal changes in NDVI, tillage, weather conditions, etc., which may have influenced the strongly negative relative difference in NDVI. On the other hand, there is the problem of the imperfection of the human factor, when the fields could not be recognized as damaged or, on the contrary, the surviving areas were marked as damaged. Therefore, for accurate damage recognition results, it is worth combining both methods, first calculating the relative differences of the NDVI, and then re-viewing the marked fields on the RGB satellite image.

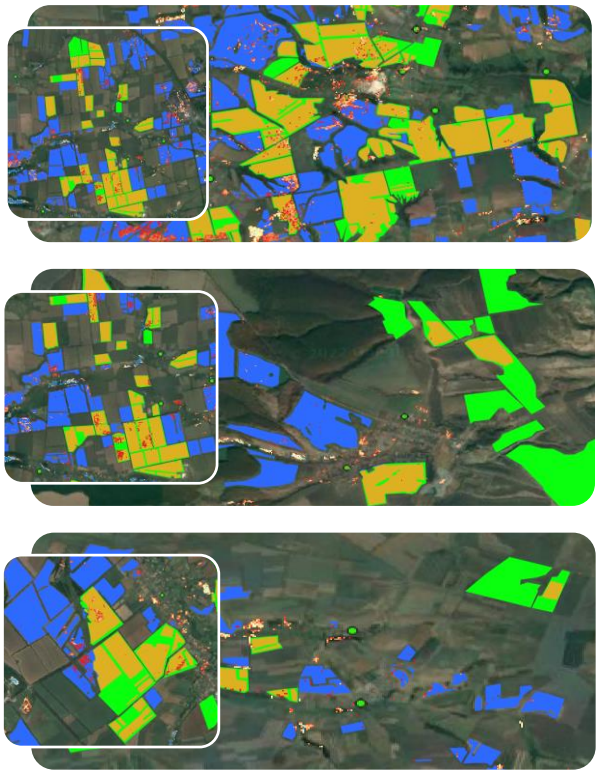


Figure 7: Comparison of fields recognized as damaged automatically (blue color) and manually (green color). Fields recognized as damaged by both methods are marked in brown. Red pixels describe the relative difference of NDVI.

4.3 Validation of Detected Damages by Experts

To check the accuracy of the visual "manual" method of identifying the fields as damaged, the obtained results were compared with the combat information provided by ACLED (Figure 8). For clarity, we

selected the 8th observation period as a period of high biomass growth rates during summer vegetation.

As can be seen from Figure 8, the damage detected using the satellite monitoring method coincides with the official data on the attacked territories. Particularly damaged areas were found in the Donetsk and Herson region (Figure 8). Thus, the method makes it possible to estimate losses even in occupied territories.

On the graph in Figure 9 shows the distribution of the number of damaged fields by the distance from the official points of military shelling for the 8 period of hostilities.

As we can see, in a small radius from the combat zone, the percentage of damage is higher. More than half of the damaged fields are located in a 5-kilometer zone from the official shelling areas, and about a quarter - in a 10-kilometer zone. In addition, the severity of attacks can be retrospectively estimated from the indicators of detected damage. Therefore, according to our estimates, the most affected areas were recorded in the 9th period of the war (Jun - 3 Jul, 75.4% in a radius of up to 5 km) and 11 (18 Jul - 31 Jul, 64.6% in a radius of up to 5 km). As a result of all periods, total damage in Ukraine is 54.3%, 27.8%, 9.3%, 3.8%, 4.8% in radii up to 5 km, 10 km, 15 km, 20 km, respectively.

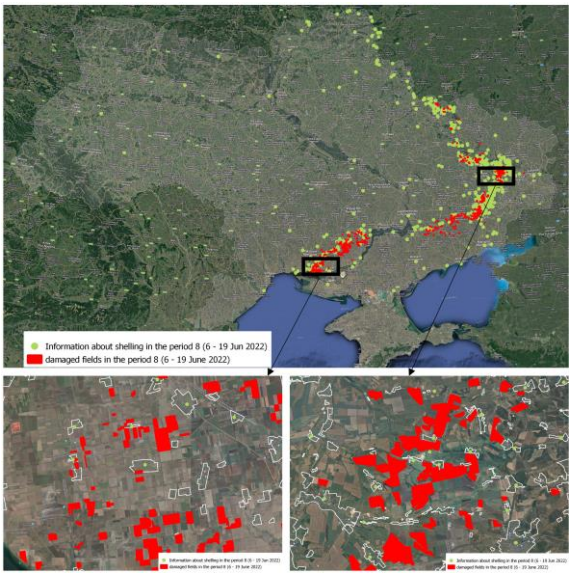


Figure 8: Geospatial location of identified damaged fields relative to official claims of military strikes for June 6-19, 2022. Examples of damages. Left - damaged fields Near Severodonetsk, Donetska oblast (June 6-19, 2022); Right - damaged fields near Herson city (June 6-19, 2022).

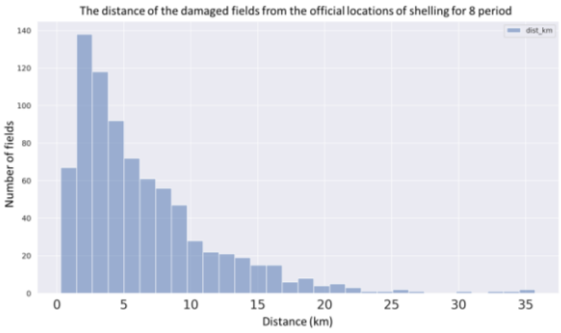


Figure 9: The distance of the damaged fields from the official locations of shelling for period 8 (June 6 – June 19, 2022).

5 CONCLUSIONS

As a result of this study, a method was developed for finding agricultural fields damaged as a result of military actions based on the relative difference of the NDVI index before and after active military actions in a specific territory and based of spectral channels. The method allows identification of point lesions of agricultural lands, such as explosions from bombs, traces of military equipment and the consequences of fires.

Damaged territories across Ukraine, including occupied lands, from the beginning of the armed conflict on February 24 to September 25, 2022, were identified. According to the results of the validation of the fields identified as damaged, it was established that the marked territories coincide with official data on the territories of military operations and are mostly located within a radius of up to 5-10 km from the zone of combat activities (up to 75.4% for the 9th period of the war, 20 Jun - 3 July). It was also found that Ukraine suffered the greatest area of damage during the 9th and 11th periods of the war (June-July).

As a next step it worth to test the proposed method for different stages of vegetation period and to discover informativeness of other indexes for detection the damages in the agricultural fields.

Thus, the method developed in this study to identify damage to agricultural fields due to war can be practically applied to help the government to accurately identify damaged land. This will provide an opportunity for the government of Ukraine and world representatives to correctly distribute financial resources among the affected landowners for the successful and effective restoration of the agricultural and industrial complex of Ukraine.

ACKNOWLEDGMENTS

The authors acknowledge the funding support from the European Commission through the joint World Bank/EU project 'Supporting Transparent Land Governance in Ukraine' (ENI/2017/387–093 and ENI/2020/418–654), Earth Observation for Ukraine (EO4UA) initiative, the project "Deep learning methods and models for applied problems of satellite monitoring" (2020.02/0292) within the competition of the National Research Foundation of Ukraine "Support research of leading and young scientists" from the State budget.

REFERENCES

[1] K. Deininger, D.A. Ali, N. Kussul, A. Shelestov, G. Lemoine, and H. Yailymova, "Quantifying War-Induced Crop Losses in Ukraine in Near Real Time to Strengthen Local and Global Food Security", Food Policy, vol. 115, Feb. 2023, p. 102418, doi: 10.1016/j.foodpol.2023.102418.

[2] D. Rawtani, G. Gupta, N. Khatri, P. Rao, and C. Hussain, "Environmental damages due to war in Ukraine: A perspective," Science of The Total Environment, 2022, vol. 850, 157932.

[3] N. Kussul, M. Lavreniuk, A. Shelestov, B. Yailymov, and I. Butko, "Land cover changes analysis based on deep machine learning technique," Journal of Automation and Information Sciences, no. 48.5 (2016), pp. 42-54, doi: 10.1615/JAutomatInfScien. v48.i5.40.

[4] Y. Ma, D. Lyu, K. Sun, S. Li, B. Zhu, R. Zhao, and K. Song, "Spatiotemporal Analysis and War Impact Assessment of Agricultural Land in Ukraine Using RS and GIS Technology," Land, 2022, vol. 11(10), p. 1810, doi: 10.3390/land11101810.

[5] Y. Aimaiti, C. Sanon, M. Koch, L. Baise, and B. Moaveni, "War Related Building Damage Assessment in Kyiv, Ukraine, Using Sentinel-1 Radar and Sentinel-2 Optical Images," Remote Sensing, 2022, vol. 14(24), p. 6239, doi: 10.3390/rs14246239.

[6] U. Haque, A. Naeem, S. Wang, J. Espinoza, I. Holovanova, T. Gutor, and U. Nguyen, "The human toll and humanitarian crisis of the Russia-Ukraine war: the first 162 days," BMJ global health, 2022, vol. 7(9), p. e009550.

[7] S.Filippo; S. Petris, and E. Borgogno-Mondino, "A methodological proposal to support estimation of damages from hailstorms based on copernicus sentinel 2 data times series," International Conference on Computational Science and Its Applications, Springer, Cham, 2020. pp. 737-751, doi: 10.1007/978-3-030-58811-3_53.

[8] G. Ghazaryan, O. Dubovyk, V. Graw, and et al., "Local-scale agricultural drought monitoring with satellite-based multi-sensor time-series," GIScience & Remote Sensing, 2020, vol. 57.5, pp. 704-718, doi: 10.1080/15481603.2020.1778332.

[9] N. Kussul, A. Kolotii, A. Shelestov, B. Yailymov, and M. Lavreniuk, "Land degradation estimation from global and national satellite-based datasets within the UN program," 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 383-386, doi: 10.1109/IDAACS.2017.8095109.

[10] P. Defourny and et al., "Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world," Remote sensing of environment, 2019, 221: 551-568, doi: 10.1016/j.rse.2018.11.007.

[11] N. Kussul, S. Drozd, and H. Yailymova, "Detection of war-damaged agricultural fields of Ukraine based on vegetation indices using Sentinel-2," 12th International Conference on Dependable Systems," Services and Technologies (DESSERT'2022), December 9-11, 2022, Greece, Athens, doi: 10.1109/DESSERT58054.2022.10018739.

[12] L. Sosa, A. Justel, and I. Molina, "Detection of Crop hail Damage with a Machine Learning Algorithm Using Time Series of Remote Sensing Data," Agronomy, 2021, vol. 11.10, p. 2078, doi: 10.3390/agronomy11102078.

[13] A. Hazem Ghassan, "Impacts of war in Syria on vegetation dynamics and erosion risks in Safita area, Tartous, Syria," Regional environmental change, 2018, vol. 18.6, pp. 1707-1719, doi: 10.1007/s10113-018-1280-3.

[14] P. Pereira, F. Bašić, I. Bogunovic, and D. Barcelo, "Russian-Ukrainian war impacts the total environment," Science of The Total Environment, 2022, vol. 837, p. 155865.

[15] The Armed Conflict Location & Event Data Project (ACLED). [Online]. Available: https://acleddata.com/about-acled/.

[16] Earth Observation for Ukraine (EO4UA). [Online]. Available: https://cloudferro.com/en/eo4ua/.

Image Segmentation as an Instrument for Setting Attention Regions in Convolutional Neural Networks for Bias Detection Purposes

Bojana Velichkovska, Danijela Efnusheva, Marija Kalendar and Goran Jakimovski
Faculty of Electrical Engineering and Infomation Technologies, "SS. Cyril and Methodius University" in Skopje, Rugjer Boshkovikj Str. 18, Skopje, N. Macedonia
{bojanav, danijela, marijaka, goranj}@feit.ukim.edu.mk

Keywords: Artificial Intelligence, Deep Learning, Medical Image Processing, Convolutional Neural Networks, Attention Regions, Lung Segmentation.

Abstract: Convolutional neural networks (CNNs) are constantly being used for medical image processing with increased application in publicly available datasets and are later being actively applied in medical practice. Therefore, since patient lives are at stake, it is important that the functionality of the neural network is beyond reproach. In this paper, due to dataset availability, we present two lung segmentation approaches using traditional image processing and deep learning methodologies; these approaches can later be used to focus a CNN for image segmentation and classification tasks, with implementations spanning everything from disease diagnosis to demographic and bias analysis. The aim of this paper is to provide a framework for segmentation in medical images of the chest cavity, as a way of applying attention regions and localizing sources of bias in images. Both of the proposed segmentation tools, the traditional image approach using computer tomography scans and the CNN applied to chest X-rays, provide excellent lung segmentation comparable to popular methods in the image processing sphere. This allows for an all-encompassing application of the developed methodology regardless of different image formats, therefore making it widely applicable in setting attention regions for CNNs.

1 INTRODUCTION

Deep learning approaches based on convolutional neural networks (CNNs) have allowed computers to achieve excellent results in the field of computer vision. Namely, CNNs have found application in tasks like image classification, object detection, and semantic classification [1][2]. Furthermore, CNNs have significantly contributed in medical image processing [3][4][5]. Researchers have successfully applied CNNs in many medical applications, such as tumor classification [6], detection of skin lesions [7], heart anomalies [8][9], etc.

It is often necessary to focus the attention of the network, namely to restrict the recognition of the network to a specific region in the image. This region is known as a region of interest (ROI) [1] or attention mask, and it is given as an input to a CNN in order to provide the focus. However, before being applied for that purpose, the ROI must be detected and properly defined. Naturally, it is easier to use already existing methodologies and defined approaches as means to extrapolate the ROI in an image.

One of the standard approaches of introducing a ROI as an input for CNNs is by assuming a fixed, rectangular ROI alike a bounding box. This ROI can then be cropped and used as a separate input to the CNN. However, there are several limitations to this approach. One important constraint to consider is that the ROI assumes a rectangular shape, and as such it is not applicable to problems where the investigated elements have an arbitrary shape, which is often the case when working with medical images. An additional issue is that with this approach certain background information, that might be essential for understanding the context of the features obtained, will be ignored due to the selective cropping.

On the other hand, irregularly-shaped ROI offer the ability to select all sections of interest, background included, and feed them to the network for the analysis. For example, with a specific image in mind, for the purposes of one study the required accent can be on the bones, whereas another might require the lungs.

Therefore, in this paper, we investigate two methodologies – traditional image processing and deep learning – in order to perform lung segmentation on two different types of medical images: chest X-rays and computer tomography (CT). The purpose is to analyze the quality of the obtained segmentation results and understand whether they could be used as attention regions in CNNs in bias analysis of medical images. The topic of bias has been widely investigated in the past few years [10]. Reported cases of bias include medical personnel, medical datasets, and medical AI-based applications [11]. With awareness levels rising, researchers have begun to analyze the presence of bias in medical images [12], with nearly 100% accuracy in gender bias and 90% accuracy in racial bias. Curiosity arises in the case of racial bias and the sources which are indicative of its presence from mere chest X-rays, therefore we wish to develop an instrument which would allow investigation of the elements which the network uses to detect it.

Due to the type of data available, we center our approach around evaluation of lung segmentation as a tool. However, the developed pipeline is applicable to other forms of segmentation (e.g., bone, soft tissue), should the data required become available. In order to evaluate the quality of the results, we compare our proposed CNN to the U-Net network [13] which is one of the standards in the field of image processing.

This paper is organized as follows. Section two provides the dataset, technology, and evaluation metrics used. Section three gives an overview of the results and in section four we conclude the paper.

2 MATERIALS AND METHODS

The pipeline for lung segmentation used in this research consists of three stages: data preprocessing, model training, and model evaluation. For that purpose, in this section we describe the datasets used in the paper, then address the technology used, and finally we provide the evaluation metrics used for evaluating the performance of the network from the results obtained.

2.1 Datasets

For the purposes of this research, we use two different image datasets. The first dataset contains chest X-rays collected by radiologists at two clinics (Montgomery and Shenzhen) [14]. The data encompasses chest X-rays from approximately 600 patients with tuberculosis. Additionally, following anatomical landmarks, the X-rays were also accompanied by corresponding lung segmentations. The second dataset contains CT images of patients suffering from pulmonary fibrosis created by the Radiological Society of North America [15].

The reason for the different data sets lies in the format. Namely the research focuses on X-ray images and CT images, as to offer a comprehensive approach for lung segmentation methods and applications on different medical imaging formats.

Both datasets are represented in the DICOM format. DICOM is a medical image processing information communication and management standard which is used to store, exchange, and transmit medical images. The standard includes protocols for exchange, compression, and 3D visualization of results for multiple medical procedures, like magnetic resonance, radiography, computed tomography, etc.

When it comes to the first dataset, one record in this standard consists of a set of pixels that represent a static position of the chest. On the other hand, when it comes to the second dataset, the representation of CT images is more complex, in that, it includes a set of static positions taken at different depths (or sections) of the body in order to create an overall image of the chest cavity. This format is widely used due to its advantageousness; namely, the DICOM standard offers high quality of the stored information, i.e., the images have higher dimensionality compared to what the human eye can perceive.

2.2 Traditional Image Processing Methodology

For the traditional image processing, we relied on the Hounsfield scale, which is a quantitative scale for describing radiodensity. The scale provides a clear overview of the information embedded in the image, or simply put, by applying a simple range of values one can understand the tissue displayed in the image. A detailed overview of the Hounsfield units is given in Table 1, where, e.g., a unit measure above 1000 means the image contains bone, calcium, or metal, whereas a unit measurement below -1000 signifies air, and so on.

However, the results obtained after applying the Hounsfield units scale can generate masks which contain artefacts (anomalies in the image, e.g., missing pixel values in the center of a mass whether it be bone or an organ). In such cases, we apply a single round of basic morphological operations, in

our case, erosion and dilation, to compensate for the errors. Dilation adds pixels to the boundaries of objects in an image, which fills out the potential anomalies in the center of the lungs. Erosion removes pixels on object boundaries, meaning it decreases the volume expansion of the lungs (originating from the dilation).

Table 1: Hounsfield units.

Unit Measure	Representation of
> 1000	Bone, calcium, metal
100 to 600	Iodinated CT contrast
30 to 500	Punctate calcifications
60 to 100	Intracranial hemorrhage
35	Gray matter
25	White matter
20 to 40	Muscle, soft tissue
0	Water
-30 to -70	Fat
< -1000	Air

2.3 CNNs

CNNs as a segmentation tool, are a structure of one or more convolutional layers, often followed by sampling layers and one or more deconvolutional layers. The input and the output layers of a CNN serve the purpose of defining the functionality of the CNN. Namely, the input layer defines what the network requests as input data, and the output layer defines what the network will provide as a result of the input. On the other hand, the type of hidden layers structured in the CNN and the way these layers are connected define the behavior of the network, or rather what the network will observe, compute and learn from.

In order to segment the lung from the chest X-rays we use a CNN consisting of four convolutional and four deconvolutional layers. The last three of the convolutional layers are followed by an undersample layer, and the first three deconvolutional layers are preceded by oversampling layers. The final layer is followed by a softmax activation function that provides the segmentation result.

2.4 Evaluation Metrics

The choice of metrics when evaluating the method's performance is important when training deep learning models [16]. The reason for this is because the same model can give different results if the analysis is performed using different metrics; namely, one evaluation metric can suggest good results, whereas the model is actually underperforming when another evaluation metrics is considered.

One of the most often used evaluation metrics is accuracy. When it comes to segmentation, accuracy as a metrics is indicative of pixel-wise classification. Now, the negative aspect of this metrics comes to light when working with significantly imbalanced data, which in the case of segmentation problems can often be the case. In cases as this, the value of accuracy can easily reach the actual percentage representation of the class that dominates the data set, while the model itself is still a weak classifier because it knows how to recognize only the dominant class.

As a result, a better metrics for evaluating segmentation models is the confusion matrix and the all-encompassing metrics which can be extrapolated from it. In this paper, we use the dice coefficient, given in (1). The values used to calculate the dice coefficient are as follows: true positives (TP), false positives (FP), and false negatives (FN). A TP is an outcome where the model correctly predicts the positive class. A FP is an outcome where an instance is predicted positive when it is actually negative, whereas a FN is an outcome when an instance is predicted negative when it is actually positive.

$$\text{diceScore} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \quad (1)$$

The dice coefficient calculates the similarity of two samples, and its values can range from 0 to 1, with 0 indicating no overlap between two segmented areas, whereas 1 represents full overlap between the proposed and the true segmentation areas. Therefore, the higher the value of the dice coefficient the better the segmentation.

3 RESULTS

The results obtained are divided into two separate groups: results from the traditional image processing obtained from the CT scans and results from CNNs obtained from the chest X-rays.

3.1 Traditional Image Processing Methodology

The results obtained through the traditional image processing in the CT scans can be observed in Figure 1. In the observed masks artefacts can be seen in several images. This means that single dilation cannot fix the artefact, therefore in the

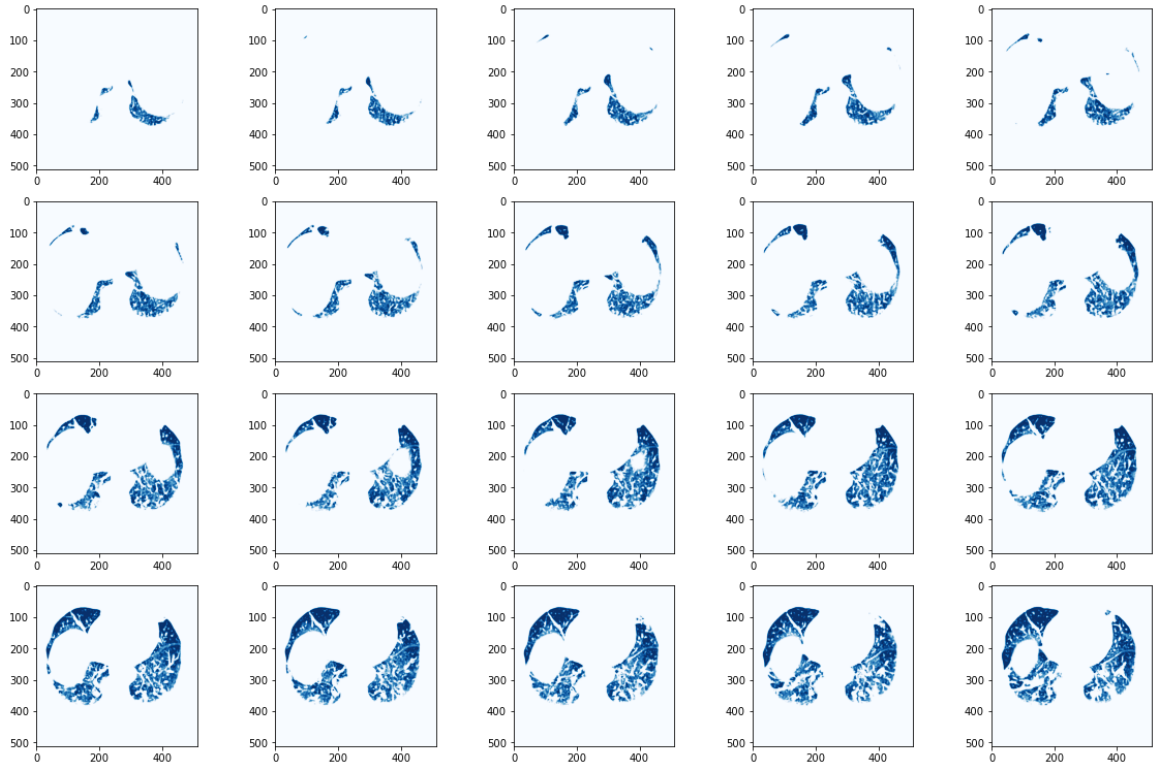


Figure 1: Results obtained from traditional image processing using the Hounsfield scale from all slices of a randomly selected CT scan.

future we can expand these results and test the number of dilation-erosion repetitions required to fix the artefacts from the segmentation errors based on the Hounsfield scale.

The result from the final segmentation can be seen in Figure 2. This is a three-dimensional rendering obtained by merging all the sections (layers) from the CT scans into one. The rendered result shows that using the Hounsfield scale provides excellent lung segmentation results, in spite of the small artefacts and excess selected regions, which can be seen in the upper left and bottom right corners of the rendition.

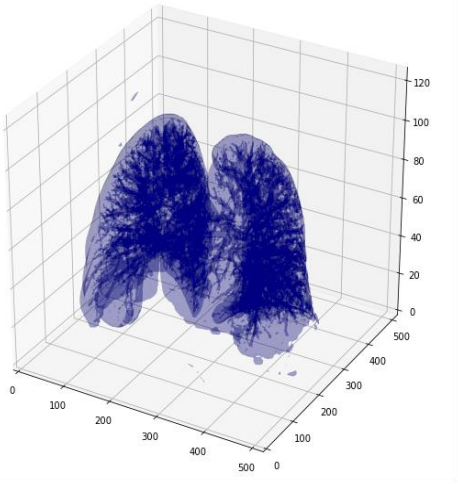


Figure 2: Three-dimensional rendering of the segmented lungs obtained from all slices in a single CT scan.

3.2 CNNs

We investigated two different CNNs. One is our proposed network consisting of four convolutional layers and four deconvolutional layer. The deconvolutional layers are concatenated with the convolutions for the purpose of eliminating noise, whereas the other is U-Net. Both models were trained and evaluated on the chest X-ray dataset.

The difference between the U-Net network and our network is the number of layers which encode the images. Namely, our proposed method uses less layers to analyze the input, whilst it provides comparable results to U-Net. We are using a two-layered concatenation, whereas U-Net concatenates three input layers. This makes the proposed model

faster to train compared with U-Net, but the overall performance is only marginally affected.

Both of the networks were trained on the same portion of the Montgomery and Shenzhen chest X-ray dataset, and both were optimized using Adam [17]. During the training stages, the monitored metrics was the dice coefficient, which was later also used to evaluate the overall performance. The results for the U-Net are given in Figure 4, while the results for our proposed method are given in Figure 3. In both figures, the first column of images is the predicted segmentation, the second column are the actual labels, whereas the third column describes the difference between the two. The difference map contains four separate colors, each of which depicts a certain aspect of the confusion matrix. The light pink indicates lung segments which were correctly identified, whereas the black shows the correctly

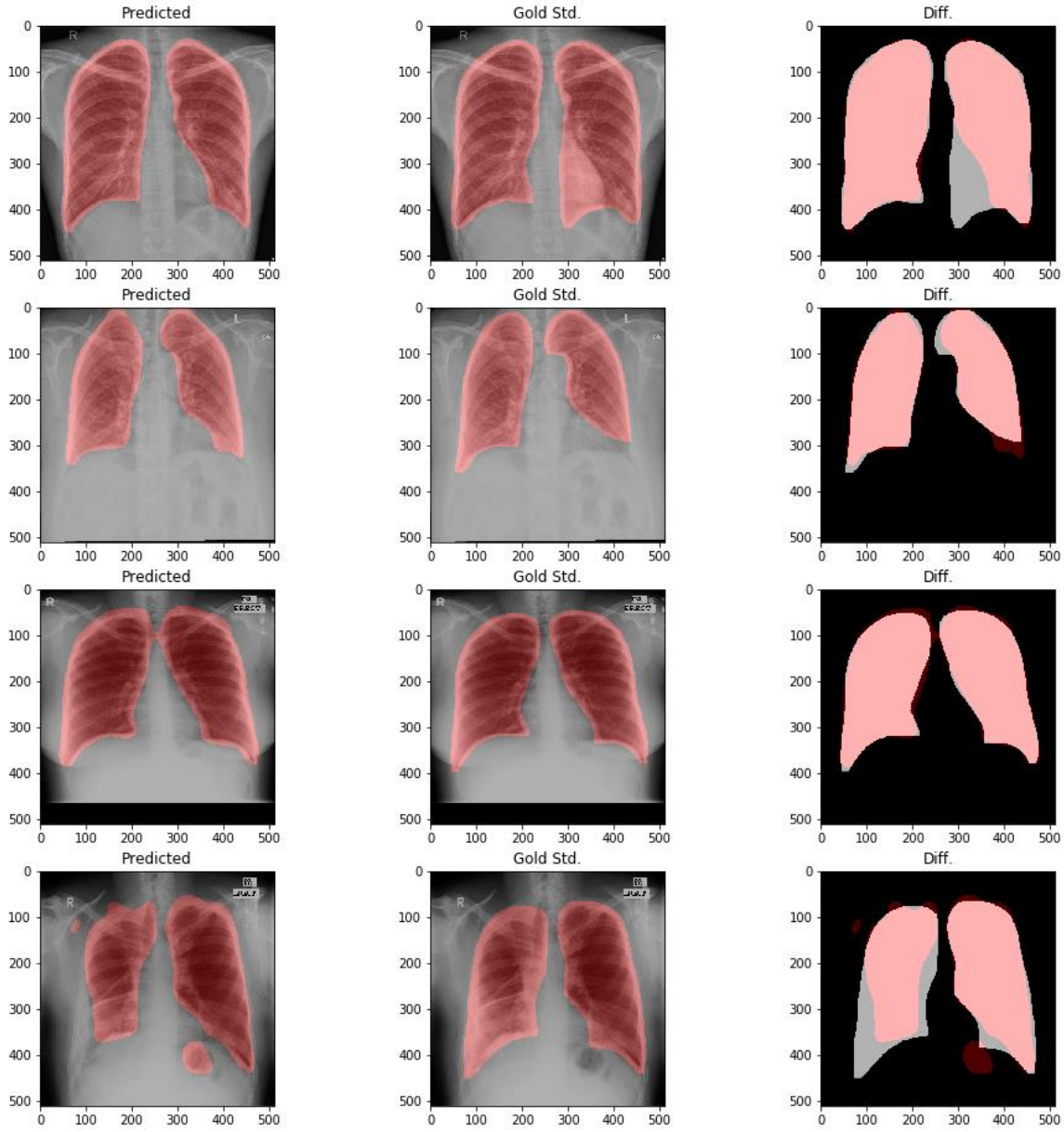


Figure 3: Results obtained from the proposed CNN. Each row represents a different chest X-ray. For the columns: the first column displays the chest X-ray overlapped with the segmentation mask predicted by the CNN, the second column displays the chest X-ray overlapped with the segmentation label (or rather the annotation provided by the dataset creators), and the third column shows the difference between the predicted and the actual masks.

removed background elements. The light gray shows the sections which should have been identified as lungs but were instead classified as background, and the burgundy shows where the model misclassified the background as lung tissue.

Major differences in the results can be noted in the last row of X-ray images in the result figures for both datasets. We can see that the proposed CNN struggles with properly segmenting a section of the left lung, which is not the case with the results obtained from U-Net. Therefore, it can be concluded that one of the limitations of the proposed CNN is the ability to differentiate the lung in instances where clouding of the lower lung is present. However, if we observe the first row of both figures, it can be noted that there are cases where the U-Net also has a difficult time identifying lower sections of the lung.

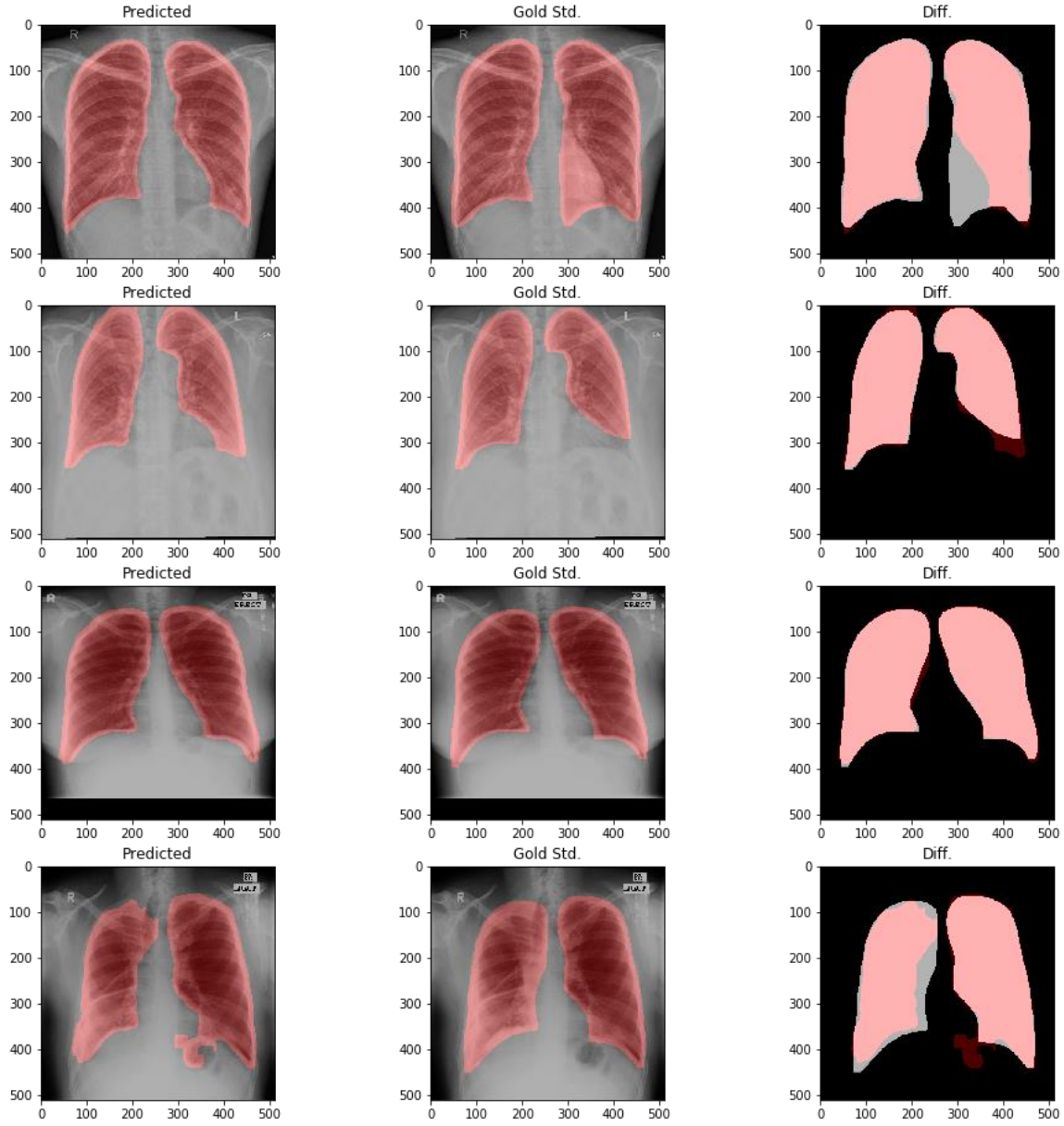


Figure 4: Results obtained from U-Net. Each row represents a different chest X-ray. For the columns: the first column displays the chest X-ray overlapped with the segmentation mask predicted by the CNN, the second column displays the chest X-ray overlapped with the segmentation label (or rather the annotation provided by the dataset creators), and the third column shows the difference between the predicted and the actual masks.

The dice coefficient for the proposed CNN is 0.95, while for U-Net it is 0.97, making the results between the two CNNs comparable, as can be seen in the values of the dice coefficient and the selected samples, with the added advantage of faster training time in the case of our proposed approach.

4 CONCLUSIONS

With this paper we proposed two separate methods for lung segmentation obtained from two different medical imaging technologies represented by a single format. Both of the proposed segmentation tools provide excellent lung segmentation in their corresponding datasets, which provides two different approaches in obtaining proposed attention regions for CNNs in classification and segmentation problems requiring focus on image sections. Since the results are comparable to a widely-used and renowned U-Net, the proposed instrument is likely to be an effective tool for focusing CNNs in future research. Namely, the overall idea is to utilize the obtained results for investigating presence of bias in imaging datasets, as well as understanding where exactly that bias originates from. Therefore, as a future step, this research will be expanded into segmenting different aspects of the chest cavity (e.g., bones, soft tissue, etc.) and applying focus on those areas in order to determine which elements of the image (and therefore which parts of the human body) contribute to presence of bias the most.

REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection Region Proposal Networks," Advances in Neural Information Processing Systems, vol. 28, 2015.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation.," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[3] P. Dutta, P. Upadhyay, M. De, and R. G. Khalkar, "Medical Image Analysis using Deep Convolutional Neural Networks: CNN Architectures and Transfer Learning," 2020 IC on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020.

[4] P. Bir, and V. E. Balas, "A Review on Medical Image Analysis with Convolutional Neural Networks," 2020 IEEE IC on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2020.

[5] P. Kalyani, S. Srivastava, A. Reddyprasad, R. Krishnamoorthy, S. Arun, and S. Padmapriya, "Medical Image Processing from Large Datasets Using Deep Learning," 2021 3rd IC on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021.

[6] J. E. A. Ovalle, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. Á. Guevara-López, "Representation learning for mammography mass lesion classification with convolutional neural networks", Computer Methods and Programs in Biomedicine, vol. 127, pp. 248-257, 2016.

[7] B. Shetty, R. Fernandes, A.P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. Lakshmanna, "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network," Sci Rep, vol. 12, no. 1, pp. 18134, 2022.

[8] S. P. Jillella, C. Rohith, S. Shameem and P. S. S. Babu, "ECG Classification For Arrhythmias using CNN & Heart Disease Prediction using Web application," 2022 First IC on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022.

[9] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, et al., "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation," IEEE Transactions on Medical Imaging, vol. 37, no. 2, pp. 384-395, 2018.

[10] S. Romano, D. Fucci, G. Scanniello, M. Teresa Baldassarre, B. Turhan, and N. Juristo, "Researcher Bias in Software Engineering Experiments: a Qualitative Investigation," 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Portoroz, Slovenia, 2020.

[11] J. A. Sabin, "Tackling Implicit Bias in Health Care", New England Journal of Medicine, vol. 387, no. 2, pp. 105-107, 2022.

[12] J. W. Gichoya, I. Banerjee, A. R. Bhimireddy, and et al., "AI recognition of patient race in medical imaging: a modelling study," in The Lancet Digital Health, vol. 4, no. 6, pp. 406-414, 2022.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Medical Image Computing and Computer-Assisted Intervention, Springer, vol. 9351, pp. 234-241, 2015.

[14] S. Jaeger, S. Candemir, S. Antani, Y. X. Wáng, P. X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," Quant Imaging Med Surg, vol. 4, no. 6, pp. 475-477, 2014.

[15] E. Colak, F. C. Kitamura, S. B. Hobbs, and et al., "The RSNA Pulmonary Embolism CT Dataset," Radiology: Artificial Intelligence, vol. 3, no. 2, p. e200254, 2021.

[16] M. M. Jawaid, R. Rajani, P. Liatsis, C. C. Reyes-Aldasoro, and G. Slabaugh, "Improved CTA Coronary Segmentation with a Volume-Specific Intensity Threshold," Medical Image Understanding and Analysis, pp. 207-218, 2017.

[17] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", 2014.

KNN-Based Algorithm of Hard Case Detection in Datasets for Classification

Anton Okhrimenko[1] and Nataliia Kussul[1,2,3]

[1]*Institute of Physics and Technology, Igor Sikorsky Kyiv Polytechnic Institute, Peremohy Avenue 37, Kyiv, Ukraine*

[2]*Department of Space Information Technologies and System, Space Research Institute National Academy of Science of Ukraine an State Space Agency of Ukraine, Glushkov Avenue 40, Kyiv, Ukraine*

[3]*Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*

ant.okhrimenko@gmail.com, nataliia.kussul@gmail.com

Keywords: KNN, Dataset Quality Assessment, Imbalanced Datasets, Hard Cases.

Abstract: The machine learning models for classification are designed to find the best way to separate two or more classes. In case of class overlapping, there is no possible way to clearly separate such data. Any ML algorithm will fail to correctly classify a certain set of datapoints, which are surrounded by a significant number of another class data points at the feature space. However, being able to find such hardcases in a dataset allows using another set of rules than for normal data samples. In this work, we introduce a KNN-based detection algorithm of data points and subspaces for which the classification decision is ambiguous. The algorithm described in details along with demonstration on artificially generated dataset. Also, the possible usecases are discussed, including dataset quality assessment, custom ensemble strategy and data sampling modifications. The proposed algorithm can be used during full cycle of machine learning model developing, from forming train dataset to real case model inference.

1 INTRODUCTION

In various machine learning tasks a size and an quality of training data have a huge impact on its performance. For general case, the more data, the best model. But the data quality is important, we can't just duplicate our data or continuously sample it from the same source or object. Data must be diverse to cover the biggest possible volume at the feature space. This lead us two the second important requirement for training data: its quality. Speaking about classification task, the dataset must be separable, e.g. there must exist a surfaces which clearly separate data point and which are smooth enough to avoid overfitting. That means that data point forms some kind of clusters, each of which contains same-class data points only.

However, for the class overlapping problem the clear and unambiguous dividing surfaces can not be composited. Similar, the data point can not be distributed to the separate clusters. As result, a part of data point can not be clearly distinguished at the feature space and there are some subareas which contains a mix of different class points without any logic or structure. So typically the researcher faces the tasks to determine if given dataset fits for given classification task, to correct data acquisition process and, in case of classes overlapping problem still exists, to achieve the best possible result using given dataset.

For deeper problem understanding lets answer an question, why does the classes overlapping problem raise? It could be caused by an inaccurate data collection and/or labeling; or by improper data representation (lack of data complexity). In the last case, adding more dimensions to the feature space could greatly improve the data fitness for the given task. As a result, the researchers need an algorithm to determine weak points of the datasets, hard case percentage and subspaces with ambiguous data. Having such an algorithm, it becomes possible to make a decision to modify data gathering process, to append new feature or sample more numerous and accurate data points within and around questionable subspaces.

The most simple dataset investigation method is its visualization on two-dimensional plot. Large-scale datasets with numerous features can not be visualized directly without additional transformations. In this case, algorithms like PCA [1], tSNE [2] can be used for converting a large number of original features into a smaller set of converted features, and plotting it in

2D figure. Those methods have its weak points, such as human subjectivity and inability to effectively visualize and analyze high dimensional data due to high losses during the conversion.

However, in most cases researcher can not influence the dataset collection process and forced to work with partially overlapping data. In this case he also needs an algorithm, which will predict potential classification accuracy on given dataset and define unreliable data samples to correct model learning process.

At last, defined at the previous steps questionable subspaces at the feature space could be used at the inference time, making prediction using another set of rules, than for data which fall in reliable part of feature space.

So, hard case detection in feature space is very important scientific problem for image recognition quality. Reliable hard case extraction would improve quality of image recognition with convolutional neural networks (CNN). Typically CNN feature extractor converts image to embedding, which is a vector at the feature space. Then, the classification using this vector is performed. The problem there is that CNN is subject to change and class separation depends not only on data quality but on feature extractor quality.

It this study we propose a novel algorithm for hard case detection based on KNN classifier and discus its possible usages during machine learning model development.

2 RELATED WORKS

There are numerous researches conducted on dataset overlapping problem. Most of them are focused on the training stage enhancing providing methods to smooth the influence of dataset imperfectness. In general, class imbalance issue can be moderated by completely ignoring the data at the overlapping region, ignoring a majority class at the overlapping region or by making a separate rules for trusted and untrusted data [3]. This demonstrates us the importance of reliable method of overlaping area detection.

There are researches on KNN algorithm performance on overlapping and imbalanced datasets [4]. During the experiments generated datasets were used. The imbalance and overlapping percentages were main hyper-parameters for its generation as well as classes ratios at the overlapping zone.

Work [5] describes a method to deal with both class overlapping and imbalanced dataset problems using data undersampling. A result of two issues combinations is as follow: if class overlapping problem exists in imbalanced dataset, the machine learning model tends to classify all the data point in a controversial zones to the majority class, while ignoring the minority classes. Authors propose to determine data points of majority class which are lying near the data points on minority classes using KNN-based approach. Later those point are removed from the training dataset in order to balance minority and majority classes. This way model is trained to pay almost equal attention to all represented classes.

Authors of [6] also investigate the case of simultaneous class overlapping and class imbalance problem. In a similar way to the previous work they propose to delete majority class at the overlapping zone. The next stage is generating of the new datasets for several ensemble models by random balanced data sampling. The final model makes prediction based on a number of simple classifiers, among which each one has been learned on its own unique dataset.

Another way to deal with a class imbalance problem is to make a separate classification for different areas in the feature space. In [7] researchers propose to train two classification model, first one makes a binary decision about belonging to non-overlapping zones and second one based on SVM makes a classification decision in case of non-overlapping region. Several researches introduce two models, first one for overlapping area and second one for non overlapping.

Similar approach can be also used for dynamic model selection from an ensemble during the inference time [8]. Aside from class overlapping zones there are ”local unfair” zones where most models fail to make a correct prediction and as a result the final model fails also. However there are still models which make correct prediction. Thus, authors of [8] have proposed solution to detect such ones and select a subset of models that suit this zone. This way final model will have better metrics and comparable performance across the all feature space.

3 PROPOSED ALGORITHM OF HARD CASE DETECTION

Two possible problems that can arise during solving a classification task are classes overlapping and outliers (Figure 1). We propose algorithm which addresses with both problem simultaneously by detecting hard cases in dataset: data points, which any classification model will fail to recognize correctly with high probability. Finding such untrusted data points in the dataset will make great help for researcher during whole cycle of machine learning model developing.

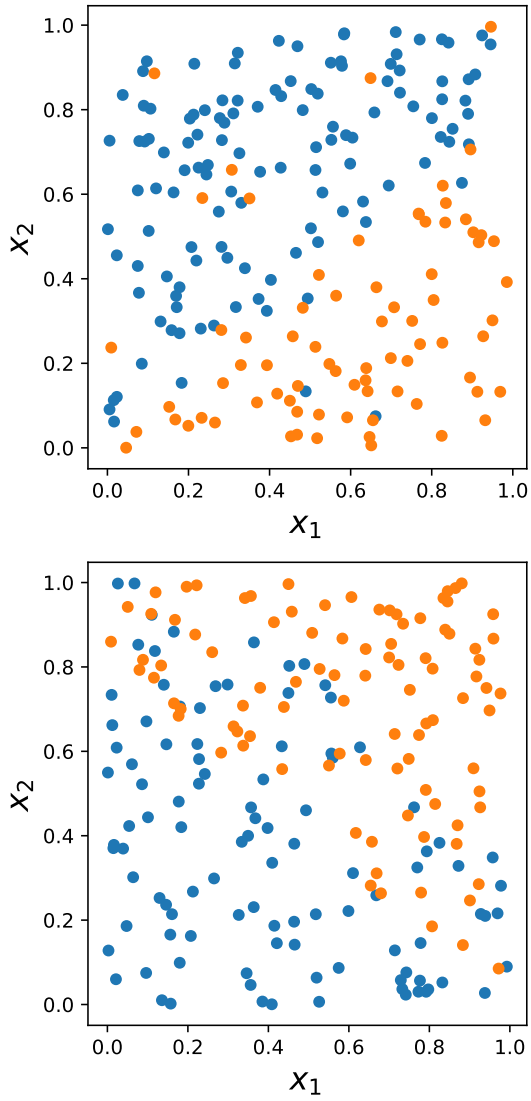


Figure 1: The outliers problem (up) and class overlapping problem (bottom).

3.1 Dataset Generation

In our study an artificially generated dataset is be used (Figure 2). It has two-dimensional features and three classes with a significant overlapping. Also several outliers are present. Three classes were sampled using random normal distribution $N(0.5, 0.2)$, $N(0.3, 0.3)$ and $N(0.2, 0.5)$. Values in range $[0.0, 2.5]$ are ignored.

For demonstration purpose we utilize two-dimensional dataset. Despite the small number of dimensions, the proposed algorithm works with any-dimensional data.

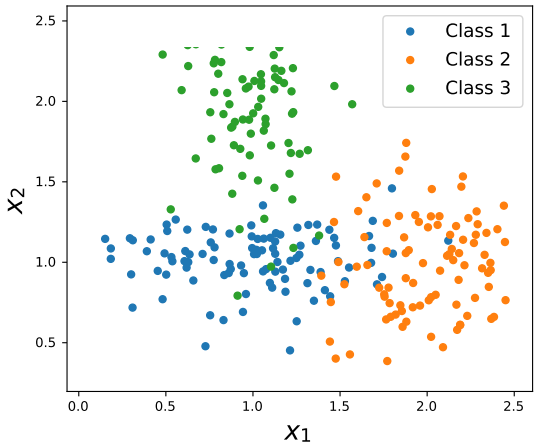


Figure 2: The generated dataset, that consists of three classes with class overlapping.

3.2 Hard Case Detection Algorithm

Let us define $\hat{X}$ as a set of data samples and $\vec{x}_i \in \hat{X}$ as a single data sample from this set. Similarly, lets define $\hat{Y}$ as a set of all possible classes and $y_i \in \hat{y}$ as a ground truth class for data sample $\vec{x}_i$.

We need to answer the question if it is possible for some data sample $\vec{x}_i$ to be correctly classified as class y_i. To do so we will use an ensemble of KNN classifiers with different neighbor number $n = [0, 1, \ldots, N], n \in \mathbb{N}$. For each $\vec{x}_i \in \hat{X}$ we get a vector $\vec{m}_i$ where m_i^j element is the result of classification of $\vec{m}_i$ by KNN classifier with neighbor number j which was trained using dataset $\hat{X} \setminus \vec{x}_i$.

$$\vec{m}_i : m_i^j = KNN(\vec{x}_i, j, \hat{X} \setminus \vec{x}_i) \tag{1}$$

This way each $\vec{x}_i$ will have its corresponding vector $\vec{m}_i$ and it is possible to construct a matrix M from this vectors:

$$M : M_{ij} = \{m_i^j\} \tag{2}$$

Now we can compare each vector $\vec{m}_i$ with groundtruth class y_i.

There are several cases:

- most elements of $\vec{m}_i$ match the true class y_i;
- first elements of $\vec{m}_i$ match the true class y_i, rest don't;
- most elements of $\vec{m}_i$ don't match the true class y_i;
- the predicted class m_i^j constantly changes depending on j, so-called class jumping;

To set a data sample as reliable we regard first two situations as obligatory. In fact the first one always

true when the second one is true, so there is one condition remains.

The data sample can't be reliable if third or fourth situation is present. The first case means that this sample is probably an outlier and the last one mean that data sample is surrounded with other sample with another class labels and probably within an overlapping zone.

Thus, the first condition C_1: from first k elements of vector $\vec{m}_i$ at least r must be equal to the right class y_i. k and r are hyperparameters and quite small. In general case, best values depends on dataset density.

The second condition C_2: data sample is unreliable if the most frequent class from first k elements of vector $\vec{m}_i$ isn't y_i.

The last condition C_3: if frequent class switching is present, data sample is unreliable. To put in algorithmic form lust check it with $1D$ convolution along the vector with kernel $K = [-1, 1]$ If two element are same, the convolution result will be zero. And for each case of class switching convolution result will be non-zero. Sum of the convolution result must not exec some value q which must be low.

The final rule for data sample classification is as follow:

$$C_1 \wedge \bar{C_2} \wedge \bar{C_3} \tag{3}$$

There are hyperparameters: k, r and q. Varying its values we can make some condition more important then another. As a general rule $q < r$ and $q < k$.

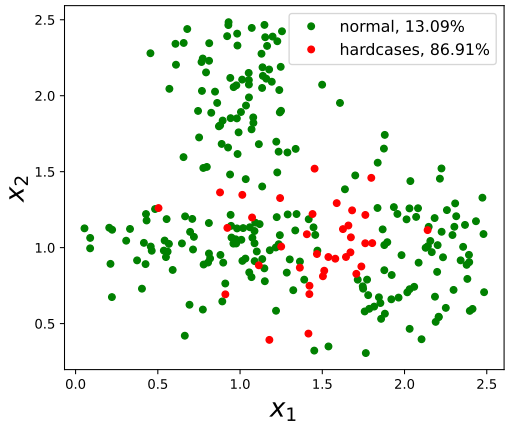


Figure 3: Detected hardcases, marked with a red color. Given a fraction of normal and hardcase data point in relation to a whole dataset size.

The example algorithm output is shown on Figure 3. It correctly detect most of questionable data points, although it has some problems with extreme data points.

4 PRACTICAL APPLICATION POSSIBILITIES

An algorithm of finding unreliable data samples is able to become one more tool at the researcher's toolbox and can be used during whole cycle of machine learning model development, from data collection to model deployment.

At the next subsections possible usecases f the proposed algorithm are discussed in details.

4.1 Dataset Quality Assessment

Introduced algorithm can prove its usefulness during both dataset collection and dataset research processes.

Most of modern machine learning problems require a huge datasets with high feature dimension and tremendous number of samples. Often there is no dataset that fits to current task. So many researchers are forced to create it, for example via outsourcing or crowd-sourcing. At this situation a tool to control the dataset quality is extremely important, because it can detect a problems at an early stages and give a possibility for a researcher to correct data gathering process.

Another situation take place when there are multiple datasets for a given problem. Usually the best ones are chosen or their combinations. Being able to determine a percentage of hardcases at given datasets, researcher can choose the dataset with the best quality. Also it possible to construct a new dataset from several ones with lowest hard case number aiming to minimize a fraction of unreliable data samples at united dataset.

4.2 Training Data Sampling Correction

As soon as it is possible to determine reliable and unreliable data sample at the training dataset, a lot of possible training enchantments are arisen. The first and the most oblivious one is to modify data sampling strategy.

During the model training, one can use reliable data for training input more frequently then unreliable ones. Balancing the frequency allow to feed the model is desired proportion, up to completely removing hard cases data. The resulting model will make decision more (heavily) based on reliable data samples rather than on unreliable ones. This allows it to make correct prediction with a high confidence at the non-overlapping zones and pay less attention to overlapping zones, because it is no way to correctly determine class at such areas anyway.

4.3 Custom Ensemble Strategy

Being able to clearly distinguish different data sample types by its quality gives an opportunity to train several models, each of which runs at its own areas of the feature space. This way the classification rules for zone with reliable data samples will be different from unreliable ones. Moreover, the models can be set to demand less confidence from the point at overlapping area during the training stage. The final model will consist of two subsets: first one for trusted data, and second one for data, which lies nearly the unreliable data points at the feature space.

However, the unreliable data sample can either lie at an overlapping zones, or to be an outlier. It is clear that the last case we can't threat with second set of classifiers and we need to filter such cases, so there is a room for improvements.

4.4 Dataset Modifications

Many methods of dealing with dataset imbalance and overlapping problems propose to delete majority class data sample at the overlapping zones [9]. To the contrary it is possible to make a class label modification without data deletion.

We need change class labels at the overlapping zones. To which class do we need to change labels is a subject for discussion and the answer is highly depends on current goals. For best possible metrics it make sense to change all class labels at the overlapping zone to the most frequent one. To fight an imbalance problem all class labels at questionable zone can be changed to the most underrepresented class.

As in the previous subsection, we need to filter outliers, which class labels can be changed to the labels of surrounding data points.

5 CONCLUSIONS AND FURTHER WORK

This work is dedicated to investigation of a class overlapping problem along with other possible hard cases such as outliers. There are many reasons for such a problem and the most important are inaccurate data sampling and lack of data dimension (e.g. there is a need to increase the dimension of feature space). Regardless of the reasons, there are some points which simply can't be classified correctly, because they have similar features with another data samples with different class label. As an example we can consider thematic segmentation problem to determine land use based on satellite data [10]. Due to similar spectral characteristics of such crops as wheat and barley, it is impossible to separate them at the crop specific map. Such data points interfere the model training process and could become a source of incorrect prediction on real data.

At this study a novel algorithm has been introduced, which allows any researcher to have more clear understanding of datasets quality and it does not depend on feature space dimension. Apart from visual estimation methods based on dataset decomposition, this method is able to produce a clear numeric data quality metrics such as the percentage of trustful data. More over, it allows to distinguish reliable data samples and unreliable ones, opening an opportunity to adjusting.

The proposed algorithm could make an enchantment during the whole cycle of machine learning model development. This include correction and adjustment of the dataset collection process, as well as dataset choosing and mixing. Having hard case detected, one can apply different rules for reliable and unreliable data during training and inference runs. The algorithm also opens room for dataset modification, according to which class labels of untrustworthy data samples are changed to match the current goals, such as high metrics or class balance.

In this study we consider artificially generated datasets, but our further steps will be related to utilization of this algorithm for real world problem of land use classification on satellite data.

REFERENCES

[1] H. Abdi and L. J. Williams, "Principal component analysis. wiley interdisciplinary reviews: computational statistics," Wiley Interdisplinary Reviews: Computational Statistics, 2010.

[2] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," Journal of Machine Learning Research, vol. 9, no. 86, pp. 2579-2605, 2008. [Online]. Available: http://jmlr.org/papers/v9/vandermaaten08a.html

[3] W. A. Almutairi and R. Janicki, "On relationships between imbalance and overlapping of datasets," EPiC Series in Computing, vol. 69, 2020.

[4] V. García, R. A. Mollineda, and J. S. Sánchez, "On the k-nn performance in a challenging scenario of imbalance and overlapping," Pattern Analysis and Applications, vol. 11, 2008.

[5] M. M. Nwe and K. T. Lynn, "Knn-based overlapping samples filter approach for classification of imbalanced data," Studies in Computational Intelligence, vol. 845, 2020.

[6] L. Chen, B. Fang, Z. Shang, and Y. Tang, "Tackling class overlap and imbalance problems in software

defect prediction," Software Quality Journal, vol. 26, no. 1, pp. 97–125, Mar 2018. [Online]. Available: https://doi.org/10.1007/s11219-016-9342-6.

[7] Y. Tang and J. Gao, "Improved classification for problem involving overlapping patterns," IEICE TRANSACTIONS on Information and Systems", vol. 90, no. 11, pp. 1787–1795, Nov 2007. [Online]. Available: https://doi.org/10.1093/ietisy/e90-d.11.1787.

[8] N. Lässig, S. Oppold, and M. Herschel, "Metrics and algorithms for locally fair and accurate classifications using ensembles," Datenbank-Spektrum, vol. 22, 2022.

[9] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," ACM Comput. Surv., vol. 52, no. 4, aug 2019. [Online]. Available: https://doi.org/10.1145/3343440.

[10] N. Kussul, A. Shelestov, M. Lavreniuk, I. Butko, and S. Skakun, "Deep learning approach for large scale land cover mapping based on remote sensing data fusion," International Geoscience and Remote Sensing Symposium (IGARSS), vol. 2016-November, 2016.

Human Activity Recognition with Wearables using Federated Learning

Borche Jovanovski[1], Stefan Kalabakov[1,2], Daniel Denkovski[1], Valentin Rakovic[1], Bjarne Pfitzner[3], Orhan Konak[3], Bert Arnrich[3] and Hristijan Gjoreski[1]

[1]*Faculty of Electrical Engineering and Information Technologies, "SS. Cyril and Methodius University" in Skopje, Rugjer Boshkovikj Str. 18, Skopje, N. Macedonia*

[2]*Jozef Stefan International Postgraduate School, Jamova Cesta Str. 39, Ljublana, Slovenia*

[3]*Digital Health – Connected Healthcare, Hasso Plattner Institute, University of Potsdam, Prof.-Dr.-Helmert-Str. 2-3, Potsdam, Germany*

kalabakovs@gmail.com, {borchej, valentin, danield, hristijang}@feit.ukim.edu.mk, {bjarne.pfitzner, orhan.konak, bert.arnrich}@hpi.de

Keywords: Activity Recognition, Machine Learning, Federated Learning, Deep Learning, Human Activity Recognition, Deep Neural Network.

Abstract: The increasing use of Wearable devices opens up the use of a wide range of applications. Using different models, these devices can be of great use in Human Activity Recognition (HAR), where the main goal is to process information obtained from sensors located in them, especially in eHealth. The high volume of data collected by various smart devices in contemporary ML scenarios, leads to higher processing consumption and in many cases results in compromised privacy. These shortcomings could be overcome by using Federated Learning (FL), a learning paradigm that allows for decentralized training of models such that user's personal data does not need to ever leave their devices, which substantially reduces to possibility of a breach. This paper analyses the behaviour and performances of FL when applied to the context of HAR. The obtained results show that FL can achieve comparable performances to those of centralized Deep Learning, while facilitating improved data privacy and diversity, as well as fostering real-time continuous learning.

1 INTRODUCTION

The ever-increasing trend of ubiquitous Internet of Things devices is paving the road for every day use and application of Wearable Sensor Technology (WST) [1]. WST has many potential benefits, such as: broad spectrum of usability, real-time information sharing, low cost of devices using this technology, custom hardware component of devices themselves, use for medical purposes in treatments and remote monitoring of patient condition, training and fitness, etc. The dominance of this technology is increasingly seen from a medical point of view. Specifically, the devices we carry enabled remote monitoring of the condition and health of patients. The data collected by the device is sent to the appropriate doctor, avoiding the unnecessary visit to the health institution. Continuous monitoring of patients' health helps to identify potential problems through preventive interventions, by improving the quality of care and saving money because the cost of prevention is often less than the cost of treatment [2]. The use of this type of devices, regardless of their purpose, is increasing day by day.Wearable Sensor Technology is a technology with a wide range of usability, which includes smart watches [3], fitness trackers, VR headsets, smart jewelry, webglasses and Bluetooth headsets. The way to use thesedevices is by attaching them to the human body, but there are some that do not have physical contact withthe user. A common factor among these WST devicesis the fact that each one monitors and collects datain real time. Due to the small form factor, these devices do not have much processing power or energy. This problem is exacerbated by the implementation of Machine Learning (ML) models, where the devices themselves are not able to store large amounts of data or apply complex DNNs [4]. Therefore, the information itself needs to be forwarded to a central server so that it can be processed. But there is a problem in terms of security and real-time continuous learning. Because of all this, when it

comes to WST scenarios [5], the application will be most compatible with Fed erated Learning (FL). Analyzing FL with respect to HAR, the comparison between FL and DL scenarios, the influence of the number of users and the number of epochs/rounds in the models where the FL is used as well as its behavior for different situations are described in this paper.

The paper is organized as follows. Section 1 focuses on the implementation of WST using FL. Furthermore, Section 2 presents the related work in the area of FL and HAR, while Section 3 presents the data set background. Section 4 presents the methodology and describes the Feed-Forward Neural Network (FFNN) used in conjunction with the FL paradigm. Our evaluation setup, metrics as well as experimental scenarios are described in Section 5, while Section 6 presents the results and a discussion of them. Finally, we conclude the paper with our interpretation and reflection on the given question and outline the open perspectives and future work in Section 7 end spacing.

2 RELATED WORK

Advanced ML and DL solutions combine a large number of data sets collected by multiple sensor devices and therefore need to be located in one centralized location in order to achieve a single unit with high accuracy. The disadvantages of training centralized models appear in terms of the costs of establishing their own communication, as well as the security of privacy. Federated Learning (FL) can mitigate these disadvantages, by combining the information of locally trained models, obtained from the underlying clients. In [6] the authors evaluated FL as a kind of classifier for Recognizing Human Activities and compared performance with Centralized Learning based on a deep neural network and softmax regression, that are trained on synthetic and real data sets. The emphasis here is on the impact of erroneousclients with corrupted data on FL, as well as communication costs. It is concluded that FL manages to achieve a model with a worse but for now acceptable accuracy of 89% compared to 93% in Centralized Training when it comes to Deep Neural Networks. Inthis paper, the global model trained for FL in relation to faulty datasets shows that the accuracy here is comparable to the Centralized Learning model. A FL algorithm was further proposed that aims to first identify and reject users that contain corrupted data. Complex models that are built on the basis of Deep Neural Networks, and at the same time have implemented the HAR, the best benefits this solution wouldcertainly achieve by using FL. In [7], a new aggregation algorithm called FedDist is proposed here andimplemented to meet the needs of penetrating applications. This algorithm allows you to modify your model architecture by identifying differences between client-specific neurons. This allows the specifics of the customers to be taken into account without disturbing the generalization. In [8] the authors propose a new hybrid HAR method (FedHAR) that combinessemi-supervised learning and FL. The combinationof Active Learning and Spreading Labels to semiautomatically track the local flow of unlabeled sensor data relies entirely on FL to build a global modelof activity in a scalable and privacy-conscious manner. In [9], a new HAR FL system called FedDL has been proposed that captures basic user relationships and applies them to the dynamic learning of personalized models for different users. A dynamic layer sharing scheme is designed that learns the similarity between the weights of the user models to form the sharing structure and merges the iterative layer models from the bottom up. It also merges local models based on the dynamic sharing scheme, significantly accelerating convergence while maintaining high accuracy. In [10], a Meta-HAR federated representation learning framework is proposed in which the embedded signals are meta-learned by FL while the learnedsignal representations are entered into a personalizedclassification network for each user to predict activity. In [11], a new challenge has been explored that is of interest - tackling heterogeneity in users' activitiesusing FL. A framework with two different versions of activity-based federated aggregation is proposed, which uses overlapping gain of information across activities, with one using a model distillation update and the other using a weighted α-update. ClusterFL [12] is proposed as a FL system that is aware of similarities and can provide high model accuracy and low communication costs for HAR applications. In terms of the cluster connection learned, ClusterFL effectivelydiscards nodes that merge more slowly or have little correlation with the other nodes in each cluster, resulting in a significant acceleration of convergence whilemaintaining accuracy. Furthermore, our model will be presented, which consists of a set of data collected bytwo sensors, an accelerometer and a gyroscope placedin two locations on the human body on the wrist and thigh. Accuracy will be presented as a comparison when using Federated Learning versus Deep Learning for different scenarios, for example when using one or the other sensor at both locations,

one or the other sensor separately for both locations and the twosensors together for both locations.

3 DATA

The dataset [13] consists of data from ten subjects, who performed the following activities of daily living: lying, walking, transition, sitting, standing, kneeling, all fours, bending, cycling, running. During the data collection process, subjects had two Inertial Measurement Units (IMUs) attached to their body, one on thewrist of the dominant hand and one on the thigh of thedominant leg. Both IMUs consist of: an accelerometer and a gyroscope. The sampling frequency of the IMUs is 50Hz. The number of samples in the training subset is 30607, while there are 6920 samples in the testing subset. The dataset contains features and logs that are organized into two directories: combined features and features. Both directories have the same data, organized in two slightly different ways. The Features-id and Labels-id files contain the extracted features (statistical + time domain + frequency domain) from the two IMU sensors (accelerometer + gyroscope) from the two IMU locations (wrist + thigh) for the object specified in their name. The data in the JSI-FOS database are approximately equally distributed among the ten subjects that participated in thedata collection, and the same can be said for the distribution of activities, hence we conclude that the distribution of activity and frequency of occurrence are the same in the test subsets and a train.

4 METHODOLOGY

4.1 Artificial Neural Networks

The inspiration from the way neurons in the human brain work is translated into these so-called Artificial Neural Networks [14], which are a kind of discipline when it comes to using Machine Learning. As technology breaks down, much more advanced machines with large capacities that can work with large amounts of data are being developed, rewarding this tin of Artificial Neural Networks (success is seen in the training of complex neural networks in some domains that require a lot of data and computing capacity). There are several types of Artificial Neural Networks, of which we can mention Feed-forward, recurring, etc. We will stick strictly to Feed-forward [15] which are successfully used for clustering, regression, template classification, optimization, control, etc. Characteristic of Feed-forward as an Artificial Neural Network is that its connections between the nodes themselves do not form a loop. On the opposite side of the Neural Networks moving forward, we have the Recurrent Neural Networks in which certain pathways are cyclically represented. The simplest form of Neural network is the retraction model because the information itself is processed in only onedirection, so the data passes through several hidden nodes, always moving in one direction and never going back. The Neural Network with more pronounced power is usually seen in its simplest form and is represented as a single-layer perceptron [16]. The modelcontains a series of inputs that are connected in a layer and multiplied by the corresponding weights. Each value is then added together to obtain the sum of the corresponding weighted input values. If the sum of the values is above a certain threshold, usually set to zero, the output value is often 1, while if the sum falls below the threshold, the output value is -1. This type of perceptron is very important in neural networks, especially when it comes to classification. Furthermore, the single-layer perceptron can incorporate aspects of machine learning, using a property known as the delta rule, the neural network can compare the outputof its nodes with predicted values, allowing the network to adjust its weights through training to produce more accurate output values. This way of training andlearning produces a form of gradient descent. In multilayered perceptrons, the process of updating weightsis represented by an analog, while the process is morespecifically defined as back-propagation. In that situation, the existence of a hidden layer in the network isadjusted according to the output values produced by the last layer. Feed Forward neural networks are characterized as quite simple in that their simplistic architecture takes advantage of certain applications when using machine learning.

4.2 Federated Learning

Federated learning [17] is a Machine Learning methodology in which all models experience is a multitude of situations with different data sets locatedin different locations, without disturbing the training sequences and data (Figure. 1). By sharing the device model, we can collectively train the model asa whole. Looking at this concept, anyone can participate directly or indirectly in the FL of their devices, e.g. edge devices, then smartphones and IoT devices, can benefit from device data without leaving the device, especially for devices that are confined to a computer where communication itself is presented as a bottleneck with smaller devices. The concept of moving data calculations is a powerful concept in

terms of building any intelligent system, while protecting the privacy of all individuals. The FL paradigm downloads the current model and calculates the updated model on the device itself by utilizing its local data. These locally trained models are then sent from the devices back to the central server where they are collected, processed, commonly with an average weighted operation, and then a single consolidated and improved global model is sent back to the target devices. FL allows machine learning algorithms to gain experience from a wide range of data sets, even if they are present in different locations. This ensures that a number of different entities cooperate with each other and build a model that will be used by all stakeholders, and will not share their security information, but only the model they need will be strictly shared.

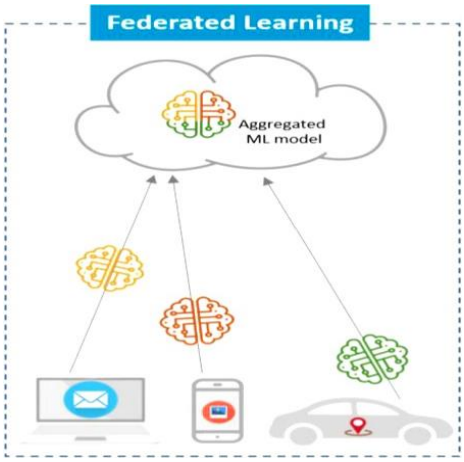


Figure 1: Federated Learning concept.

5 EXPERIMENTAL SETUP

This section presents the various scenarios and the model on which the tests were performed, as well as the behavior of the strategies we use during implementation.

5.1 Evaluation Setup

Behavior between FL and DL is presented in the following way: the collected data from the 10 users is divided into two possibilities, one for testing and the other for training. The test subset contains 20 percent of the original dataset, while the remaining 80 percent is intended for training. No validation set is used, and the goal is to directly report changes in performance made by different settings on the test subset. It was decided to use a more personalized evaluation setup instead of a Leave-One-Item-Out strategy, because FL provides a unique opportunity to train centralized models without explicitly sharing data between users. Splitting the data into two subsets strictly at each user serves to further merge with the corresponding subsets of the other users, with the final test and training subsets containing the data from all users, respectively. The way we use these two features during training and evaluation of DL and FL models is as follows: for DL models, the training feature updates the model in each epoch, while the test feature evaluates after each epoch and at the very end of training. On the other hand, in FL the training facility is used to train each local user model after each round, while testing it for evaluation. Finally, after each round we have an evaluation of the centralized model which contains the merged data from test subsets of all users in the database.

5.2 Metrics

Here, the macro F-score will be used to quantify the performance of each of our models, and this is done due to the unbalanced distribution of activity instances in the JSI-FOS database, in this study. The macro F-score allows avoiding the bias of metrics such as the accuracy of activities with a larger number of examples by calculating the F-score for each activity separately and reporting the average of the results obtained. This F-score metric is defined as the harmonic mean of the precision and recall metrics for a particular label. There is no value interpretation here, for example, precision metric, F-score, and macro F- score values closer to 1.0 represent better classification performance, while values closer to 0.0 represent worse classification performance. Note that the macro F-score and accuracy metrics typically report very close values on datasets with a balanced distribution of activity instances. Equations (1), (2), and (3) show how each of the precision, recall, and F-score metrics are calculated, respectively.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

5.3 DL and FL Models

The parallel between the DL and FL models is presented in terms of using a different number of training epochs/rounds. The main thing about this experiment is that we will look at the behavior of DL with respect to the FL model when we use the

different number of epochs, that is, when the database will be fully used. The maximum number of epochs/rounds used for training is 50. Note that when comparing DLmodels and FL models, we treat one epoch (DL) andone round (FL) as equivalent. Here, we aim to understand the impact of individual FL parameters on the overall FL performance. To this end, we chose to observe the changes in performance produced by varying the number of epochs/rounds and the percentage of clients used for training during each round of the FL model trained when using the full dataset. We will include a metric called F-score, with which we will present the behavior of the model itself in relation to the number of clients per round, i.e. not only does it change the number of users who have access to each page, we also change the percentage of clients that are used for training in each round. The model fitted to this experiment is characteristic of 50 rounds.

6 RESULTS AND DISCUSSION

In this section, we present the results obtained from the various experiments regarding the behavior between the FL and DL models. FL is an online learning strategy and transfers the model weights to the FL centralized server in each round of operation, FL- based macro F-score curves are presented as continuous with respect to the data transfer volume. In contrast, since DL is an offline strategy and the data needs to be fully transferred to the centralized server to perform the model training. DL and FL strategies, multiple runs were conducted to calculate 95% confidence intervals for the macro F-score. We will look at the result for the macro F-score when using the full dataset (Figure 2). It is noteworthy that even in the case where the full dataset is used, the FL model slightly approaches the DL model in the case of 35 epochs, this is a result because the DL model has a fairly relative growth since the beginning compared to the FL model. The performance degradation of FL relative to macro F may come as a result of the small number of epochs used to train the local FL models. A slight degradation in terms of the performance itself is observed with the DL strategy, in the situation of the full feature dataset. We conclude that by increasing the number of local epochs in the FL strategy, it approaches the DL strategy, which shows the benefit of FL implementation.

Figure 3 shows a heatmap that presents the achieved macro F-score of the FL model when varying the percentage of clients used for training in one round and the total number of rounds used for training. More specifically, on the horizontal axis, Figure 2 show the percentage of clients used for training in each epoch and on the vertical axis, it shows the maximum number of rounds. After roughly 20 rounds of training, the performance of the FL model plateaus and the achieved performance varies by at most two percentage points. Furthermore, Figure 3 also suggests that using a larger percentage of clients, i.e., above 50%, for training during one round produces better results. In our case, the best results seem to be produced by an FL model that uses 80% of all clients to update the centralized model in each training round.

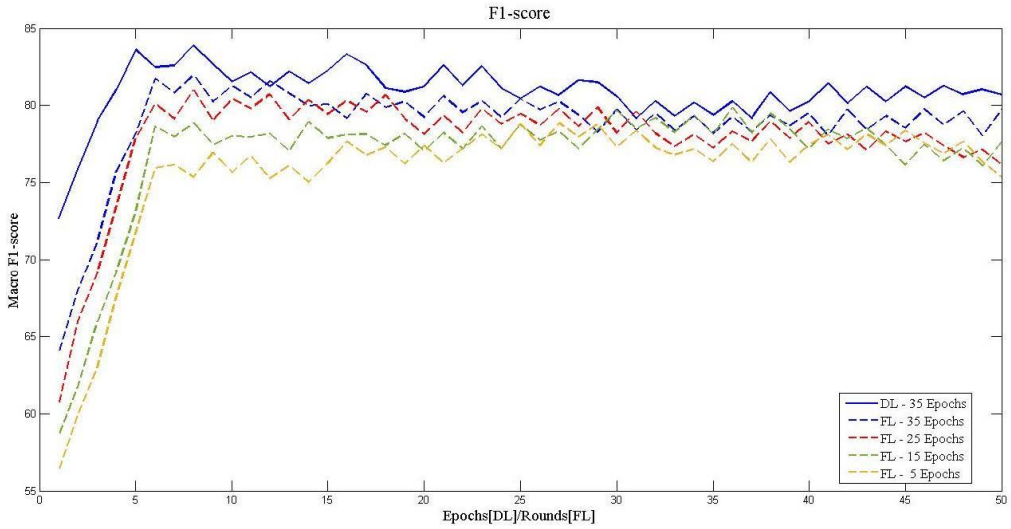


Figure 2: A comparison between FL and DL in terms of Full-featured dataset.

Figure 3: Macro F-scores achieved by an FL model when varying the number of clients used for training in each round and the total number of rounds used for training.

7 CONCLUSIONS

This paper presents the advantage that the FL strategy has in terms of IoT implementation, as well as the advantages that can be seen in terms of eHealth. The FL strategy is presented as a distributed collaborative approach to Artificial Intelligence - AI that is able to offer all of this to distributed IoT devices without the need for data sharing. The emphasis here is on the use of FL in scenarios used for HAR. What can be concluded is that the importance of the number of clients and local epochs for FL model training are key parameters in this experiment. The correct selection of the parameters, i.e. the optimal ones, leads to the fact that the performance of the FL strategy is quite close to that of the already existing DL strategies, which justifies the implementation of FL.

The future direction of research will continue to deepen the benefits and advantages that the FL strategy offers us, in terms of creating different models that will have different demands and testing their behavior in relation to DL.

ACKNOWLEDGMENTS

This work has been supported by the WideHealth project - European Union's Horizon 2020 research and innovation programme under grant agreement No. 952279. Additionally, Stefan Kalabakov would also like to thank the Slovene Human Resources Development and Scholarship Fund (Ad futura) for theirsupport.

REFERENCES

[1] A. Alsiddikya, W. Awwada, K. Bakarmana, H.Fouad, A. S.Hassanein, and A. M. Solimanc, "Priority-based data transmission using selective decision modes in wearable sensor based healthcare applications", Computer Communications, 1 July 2020, pp. 43-51.

[2] Z. Xiao, X. Xu, H. Xing, F. Song, X. Wang, and B. Zhao, "A federated learning system with enhanced feature extraction for human activity recognition", Knowl. Based Syst. 2021, vol. 229, p. 107338.

[3] L. Tu, X. Ouyang, J. Zhou, Y. He, and G. Xing, "FedDL: Federated Learning via Dynamic Layer Sharing for Human Activity Recognition", in Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra Portugal, 15-17 November 2021.

[4] Y. Liu, J. Peng, J. Kang, A. M. Iliyasu, D. Niyato (IEEE Member), and A. A. Abd El-Latif (IEEE Fellow), "A Secure Federated Learning Framework for 5G Networks", IEEE Wireless Communication Magazine, 12 May 2020.

[5] M. Wasilewska, H. Bogucka, and A. Kliks, "Federated Learning for 5G Radio Spectrum Sensing", Sensors 2022.

[6] K. Sozinov, V. Vlassov, and S. Girdzijauskas, "Human activity recognition using federated learning", in Proceedings of the 2018 IEEE Intl Conf on Parallel, Distributed Processing with Applications, Ubiquitous Computing Communications, Big Data Cloud Computing, Social Computing Networking, Sustainable Computing Communications.

[7] S. Ek, F. Portet, P. Lalanda, and G. Vega, "A Federated Learning Aggregation Algorithm for Pervasive Computing: Evaluation and Comparison", 2021 IEEE International Conference on Pervasive Computing and Communications (PerCom), 25 May 2021.

[8] C. Bettini, G. Civitarese, and R. Presotto, "Personalized Semi-Supervised Federated Learning for Human Activity Recognition", arXiv 2021, arXiv:2104.08094.

[9] L. Tu, X. Ouyang, J. Zhou, Y. He, and G. Xing, "FedDL: Federated Learning via Dynamic Layer Sharing for Human Activity Recognition", in Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra Portugal, 15-17 November 2021.

[10] C. Li, D. Niu, B. Jiang, X. Zuo, and J. Yang, "Meta-HAR: Federated Representation Learning for Human Activity Recognition", in Proceedings of theWeb Conference 2021 (WWW'21); Association for Computing Machinery: Ljubljana, Slovenia.

[11] G.K. Gudur and S.K. Perepu, "Resource- constrained federated learning with heterogeneous labels and models for human activity recognition", in Proceedings of the Deep Learning for Human Activity Recognition: Second InternationalWorkshop, DL-HAR 2020, Kyoto, Japan, 8 January 2021, Springer: Berlin/Heidelberg, Germany, 2021.

[12] X. Ouyang, Z. Xie, J. Zhou, J. Huang, and G. Xing, "ClusterFL: A Similarity-Aware Federated Learning System for Human Activity Recognition", MobiSys '21: Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services, June 2021.

[13] S. Kozina, H. Gjoreski, M. Gams, and M. Lustrek, "Three-layer Activity Recognition Combining Domain Knowledge and Meta- classification", Journal of Medical and Biological Engineering, vol. 33(4), pp. 406-414.

[14] R. Dastres and M. Soori, "Artificial Neural Network Systems", International Journal of Imaging and Robotics (IJIR), 2021, vol. 21 (2), pp.13-25.

[15] Al-Z. Malek, S. Almajali, and A. Awajan, "Experimental evaluation of a multi-layer feedforward artificial neural network classifier for network intrusion detection system", International Conference on New Trends in Computing Sciences (ICTCS) 2017.

[16] J. Singh and R. Banerjee, "A Study on Single and Multi-layer Perceptron Neural Network", 2019 3rd International Conference on Computing Methodologies and Communication (IC-CMC), 27-29 March 2019.

[17] Q. Li, Z. Wen, Z. Wu, and et al., "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection", IEEE Transactions on Knowledge and Data Engineering (TKDE), 5 Dec 2021.

Research of Methods to Increase the Efficiency of the Object Detection System on the Raspberry Pi Platform

Daria Koshutina and Svetlana Antoshchuk
Odesa Polytechnic National University, Shevchenko Avenue 1, Odesa, Ukraine
{*d.v.koshutina, asg*}*@op.edu.ua*

Keywords: Robotics, Raspberry Pi, Python, Machine Learning, Deep Learning, Convolutional Neural Networks, Computer Vision, Object Detection, Single-Shot MultiBox Detector, R-CNN.

Abstract: The work is devoted to the consideration and solution of the problems of object detection efficiency. This article analyzes object detection methods. The existing methods and systems of object detection are considered. On the basis of the researched methods, prospects and further directions for the development of object detection programs are defined. This research is relevant in today's world, because smart devices, robots and robotic systems are increasingly being used to improve life. Therefore, the object detection system is an important part of robotics and automation. The development of a real-time object detection algorithm on the Raspberry Pi platform is described. The method of automatic detection and recognition of objects is described. To check the effectiveness of the methods, a system was designed and implemented, which is a camera connected to Raspberry Pi using the algorithm developed during the work. The problem consists in creating algorithms and methods to improve the response time and accuracy of object detection in real time. The system was created on the basis of already existing research results, refinement and implementation of the methods proposed in them In the course of the study, the results of the development of the object detection system based on the developed algorithms were presented and their effectiveness was investigated.

1 INTRODUCTION

Modern life is becoming increasingly automated, accelerating the pace of the economy and society in general. However, some areas are still not fully developed, the desired results that can affect the life of mankind have not been achieved. One of these areas is the development of a real-time object detection system.

Such systems are very popular in space developments and in public safety management bodies, in particular, in places of mass gathering of people, with the aim of preventing terrorism, as well as recognizing persons who have committed crimes (airports, train stations, banks, supermarkets and shopping centers, cultural, entertainment and sports facilities). It is also important to control order on the city streets. Most of such systems have many inaccuracies and shortcomings, which complicates the process of detecting objects in real time.

Object detection is a computer technology related to computer vision[1] and image processing that deals with detecting instances of semantic objects of a certain class (such as people, buildings, or cars) in digital images and videos[2]. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance[3].

The main goal of the article is analyzing existing solutions in the field of object detection in real time, including their classification.

2 RESEARCH OF REAL-TIME OBJECT DETECTION METHODS

Object detection methods are usually based on either machine learning or deep learning[4].

Deep learning is an industry of machine learning, which is based on the set algorithms, which attempt to model high-level abstractions in data by applying a deep graph with several processing layers constructed from several linear or nonlinear transformations[5].

The You Only Look Once (YOLO) method is less accurate than regional convolutional neural networks, but at the same time it is much faster, allowing you to detect objects in real time[6]. The essence of this

method is the initial division of the image into a grid of cells. In addition, each cell defines the class of the object in the area related to this cell[7]. The main idea of this approach is to define an object as a problem of regression to spatially separated bounding blocks and associated class probabilities[8].

The Single-Shot MultiBox Detector method is represented by two components: a neural network for image classification and a convolutional layer for detecting and classifying objects in the image. This method, like YOLO, divides the image into a grid, but also operates with the concept of an anchor region[9]. Multiple anchor regions can be assigned to each grid cell. Each of them is determined and responsible for the size and shape of the object inside the cell.

The SSD method is used to describe architectures that use a single convolutional neural network to directly predict the location of regions and their classes, without applying a second stage of classification. In this entire method, at the output of the neural network, several thousand forecasts are formed for the possible regions of the location of objects of different shapes at different scales, then with the help of suppression of non-maximums, a selection of several most likely areas is made. Such a single structure, simultaneously taking into account different image scales, provided the SSD method with the highest indicators in terms of speed and quality of object detection compared to other modern approaches[10].

The operation of the SSD method is based on a fixed set of rectangles, which check the presence of an object on each of them[11]. Suppose we have some m×n feature map, which is obtained from one of the convolutional layers of a neural network. Let's go through it with a convolution with a 3×3 kernel, which at the output produced 4+Cl channels, where Cl is the number of classes.

That is, we divide our image with a grid, because each feature at the output of the convolutional layer absorbs information about the pixels of a square in the original image and therefore can detect an object located in this square. The earlier layer we use to extract the feature map, the larger its size (i.e., m and n) and the smaller objects we will be able to detect.

Unlike R-CNN, where there is at least a minimal probability of finding an object in candidate regions, there is no region filtering step in SSD[12]. As a result, a much larger number of describing rectangles at different scales is generated compared to R-CNN, and most of them do not contain an object. In order to solve this problem in SSD[13], firstly, the suppression of non-maximums of combining similar rectangles into one is used. Secondly, the hard negative mining technique is used, according to which only a part of negative examples is used in each iteration of training, in SSD the ratio of the number of negative examples to positive is 3 to 1.

For each object marked on the image, there can be several predictors from whom we are ready and want to get a description of the object. Let's introduce an indicator function that is equal to one if the i-th anchor has an IoU greater than 0.5 with the j-th object in the image, and zero if not.

The overall target loss function is the weighted sum of localization loss (loc) and confidence loss (conf) (1):

$$L(x,c,k,g) = \frac{1}{N}(L_{conf}(\overline{x},c) + \alpha L_{loc}(x,l,g)), \quad (1)$$

where N is the number of matching blocks by default. If $N = 0$, wet sets the loss to 0. The localization loss is the smooth $L1$ loss between the predicted box parameters (l) and the ground truth field (g). We regress to offsets for the center (cx; cy) of the default frame (d), as well as for its width (w) and height (h).

L_{class} - is responsible for the correct definition of the class of the object and is summarized by many anchors.

L_{loc} - is the sum of all anchors to which the object is mapped, i.e., $X_{ij} \neq 0$ and fines for errors in defining the rectangle of the object are summed up.

If we have chosen m layers, then evenly spreading from the scales we get (2):

$$S_k = S_{min} + \frac{S_{max} - S_{min}}{m-1}(k-1). \quad (2)$$
$$k = 1, \cdots, m$$

Also, for each scale, in order to choose not only square anchors, we will set a set of aspects (3):

$$A = \{a_r | r = 1, \cdots, T\} \quad (3)$$

Then the dimensions of the anchors will be calculated according (4):

$$w_k^r = S_k\sqrt{(a_r)}, \quad (4)$$
$$h_k^r = \frac{S_k}{\sqrt{(a_r)}}$$

Anchor positions are chosen simply. If we attach squares to a layer with a feature of size m×n, then the centers of the squares will be at the points (5):

$$\left(\frac{i+0.5}{m}\right), \left(\frac{j+0.5}{n}\right). \quad (5)$$
$$i = 0.1, \cdots, m; j = 0.1, \cdots, n$$

The SSD architecture is the most suitable for real-time image processing (especially when using Mobile Net networks)[3][13], but it must be taken into account that high accuracy requirements cannot always be met. This method has the following advantages:

- the speed of determining the object in real time;
- one pass through a convolutional neural network;
- searching for objects on layers, which does not degrade the quality of the input image. That is, more bins by default results in more accurate detection, although it affects speed;
- having MultiBox on multiple layers also results in better detection due to the fact that the detector works with multi-permit features;
- SSD mixes objects with similar categories (e.g., animals). This is probably because the seats are shared by several classes;
- the SSD-500 (the highest resolution variant using 512×512 input images) achieves the best MAP on Pascal VOC2007 at 76.8%, but at the expense of speed, where its frame rate drops to 22 fps. The SSD-300 is thus a better compro-mise, with 74.3 mAh at 59 fps.

3 CONSTRUCTION OF THE OBJECT DETECTION ALGORITHM ON THE RASPBERRY PI PLATFORM

The analysis of existing object detection methods allowed us to highlight the main drawback: the SSD method gives the worst performance for small objects, because they may not be displayed on all object maps. Increasing the resolution of the input image alleviates this problem, but does not completely solve it.

To implement the task, the Raspberry Pi platform was used. To implement object detection, the SSD300 method was used, whose algorithm was described above. Since the recommended input image size for this method is 300×300, a camera with the appropriate image resolution was chosen.

Also, let me show you a diagram of the algorithm for detecting objects on the Raspberry Pi platform. The algorithm of this method is shown in Figure 1.

Step 1. We take the image out, if the object is in the camera's field of view.

Step 2. We pass the original image through a series of convolutional layers and obtain a set of feature maps for different scales.

As a result of the analysis of the existing methods and the analysis of their advantages and disadvantages, we choose the SSD300 method for the implementation of the system. The main advantage of this version of the method is that the input image is reduced to the size of 300×300 pixels. Which in turn improves the processing time of this image[13].

Step 3. We apply a 3×3 convolutional filter to each point of each feature map to obtain rectangles.

Step 4. For each rectangle, the spatial displacement probability of finding the object is simultaneously estimated

Step 5. In the learning process, the real descriptive objects of the rectangles are compared with the predicted ones to exclude late detections.

Step 6. We compare the object class in the image with the object class of the training model.

Step 7. Combine all layers of the image and suppress non-maximums.

Step 8. We display the result.

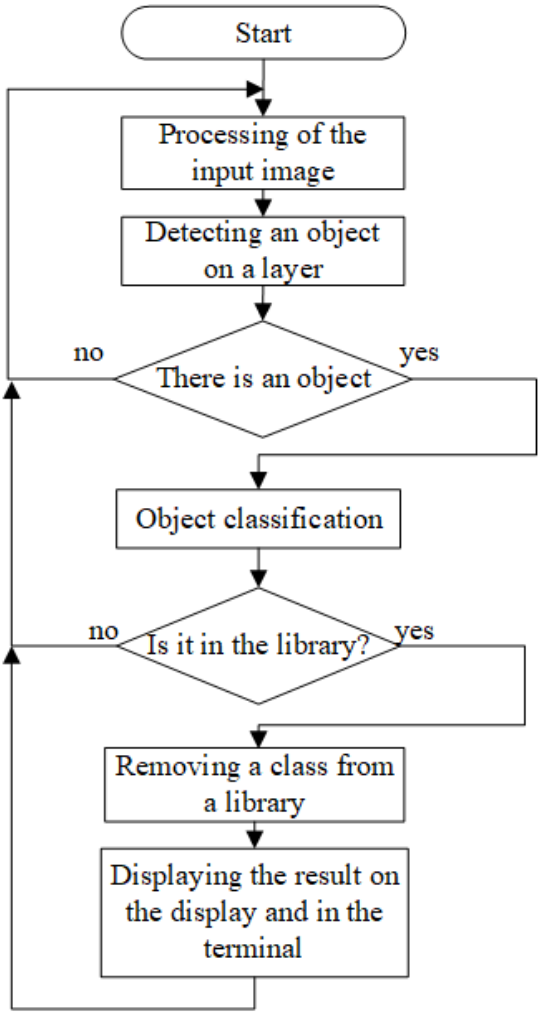


Figure 1: Block diagram of the object detection algorithm.

4 OBJECT DETECTION SYSTEM ON RASPBERRY PI

Based on the algorithm described above, we will develop an object detection system using the Raspberry Pi platform.

In order to test the theories regarding the effectiveness of the methods of detecting objects defined above, it is necessary to conduct a research experiment. For this, a system of detecting objects using the proposed method was developed.

The main functional capabilities of the system are:

- detection of the centers of coordinates of the position of the object;
- object classification;
- detection of the result of system operation.

The SSD method is implemented in the Python 3 programming language using the TensorFlow[14] and Keras deep learning libraries.

This system consists of 3 elements: camera, Raspberry Pi and display. During the execution of the program, the camera constantly reads the image on the video stream and processes it. It transfers the received results to the terminal on Raspberry Pi, it is shown in Figure 2.

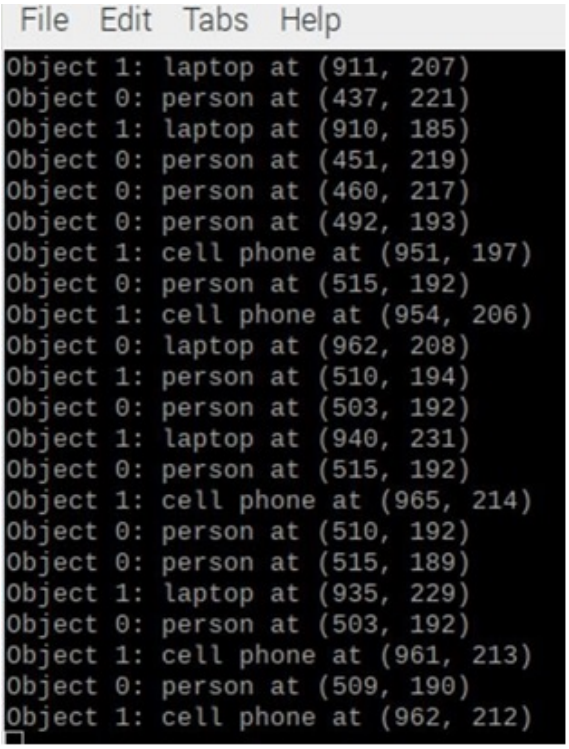


Figure 2: The result of executing the program in the terminal on Raspberry Pi.

At the same time, it is transmitted to the display, which is shown in Figure 3.

Figure 3: The result of program execution on the display.

5 CHECKING THE EFFICIENCY OF THE PROPOSED SOLUTIONS

To evaluate the quality of object detection, such metrics as the measure of the intersection of the found and reference rectangles that contain objects (Intersection, I), completeness (Recall, R) and accuracy (Precision, P) of object detection are used.

The measure of the intersection of the found and reference rectangles I (6) show how clearly the convolutional neural network found rectangles relative to the rectangle of the reference marking.

$$I = \frac{S_i}{S_j + S_{gt} - S_I}, \tag{6}$$

where S_i - the area of intersection of the true and calculated rectangle;

S_j - the area of the rectangle found by the algorithm;

S_{gt} - the area of the reference rectangle (grounded truth).

The completeness of R (7) shows the sensitivity of the algorithm to errors of the 2nd kind, that is, omissions, and is equal to the ratio of the number of correctly found objects to the number of these objects in the reference markup.

$$R = \frac{t_p}{t_p + f_n}, \tag{7}$$

where t_p - true positives are those objects expected to be seen and received at the output;

f_n – false negatives – objects that we expected to see, but the algorithm did not identify them (misses).

Accuracy P (8) shows the sensitivity of the algorithm to errors of the 1st kind, i.e., false positives, and the ratio is equal to the number of correctly found objects to the total number of rectangles found by the algorithm.

$$P = \frac{t_p}{t_p + f_p}, \tag{8}$$

where f_p - false positives are objects that should not be at the output, but the algorithm erroneously returned them at the output (false activations).

To test the effectiveness, we will take 3 methods:

- the YOLO method;
- the SSD300 method;

The above methods have been tested using a camera and a computer.

- the SSD300 method created on the Raspberry Pi platform (SSD300_Raspberry).

This method has been developed and implemented on the Raspberry Pi platform. The same camera as in the previous methods was used to obtain the image.

The parameters for evaluating efficiency are calculated according to formulas 6-8.

During the experiment, 100 images were processed. To construct graphs, the threshold coefficient in the object detection algorithm is changed in the range from 0 to 1 with a step of 0.01. The threshold coefficient is the minimum value of the probability estimate at which the decision to detect the object will be made. Figures 4-6 show the graphs of dependencies of accuracy, completeness, and crossing measure on the given threshold.

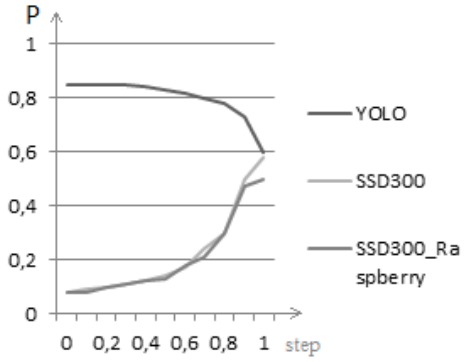


Figure 4: Dependencies of the accuracy (P) of the methods: YOLO, SSD300 and SSD300_Raspberry from the given threshold.

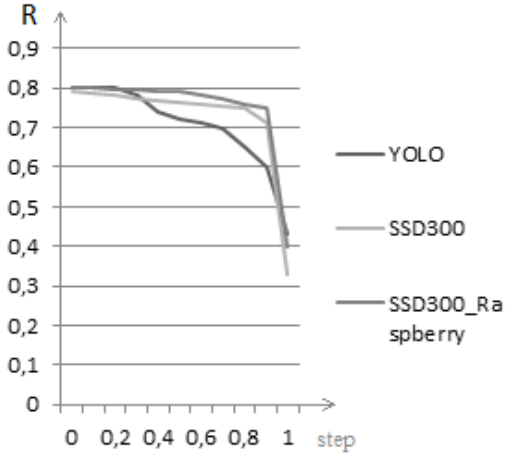


Figure 5: Dependencies of the completeness of the activation (R) of the methods: YOLO, SSD300 and SSD300_Raspberry from the given threshold.

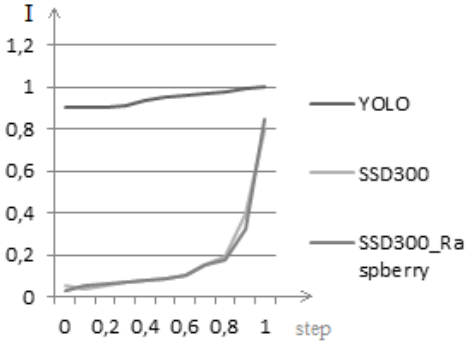


Figure 6: Dependencies of the measure of the intersection of the found and reference rectangles (I) when using the methods: YOLO, SSD300 and SSD300 Raspberry from the given threshold.

To assess the quality of the system, the mAP (mean average precision)[15] metric is calculated, which is the average value of AP across all classes of objects. AP (Average Precision)[15] is the average value of the maximum precision for different completeness values. The area under the graph of dependence of accuracy on completeness (AUC – area under curve) and mAP were used as integral estimates of the quality of the detectors. The named indicators are presented in the Table 1.

As a result of the assessment of the quality of the system, it was found that the SSD300 Raspberry method works faster by 0.2 ms, and detects objects more accurately by 3% than the SSD method on a computer. However, compared to the YOLO method, it is slower by 20 ms, but the accuracy is higher by 8.7%. From this, it can be concluded that the system

on the Raspberry Pi platform is better suited for the implementation of a system that recognizes objects in real time.

Table 1: Performance test results.

Method	Characteristics		
	AUC	mAP, %	Time, ms
YOLO	0.882	66.4	76
SSD300	0.573	72.1	58
SSD300_Raspberry	0.534	75.1	56

6 CONCLUSIONS

We hope you find the information in this template useful in the preparation of your submission.

Detection of objects in real time is a complex task that requires further improvement of existing or creation of new methods for the implementation of this task. Using the SSD method, we have an advantage in the speed of work, but the disadvantage is the quality of object classification.

In the course of the research, already existing methods of detecting objects in the image were analyzed. Based on the results of the analysis, an opinion was formed about the further development of this topic. Based on the developed methods and algorithms, with the help of tools and tools of Python programming technology and the Raspberry Pi platform, a real-time object detection system was designed. Recommendations have been developed to improve the operation of the methods, due to the improvement of technical equipment. To conduct a study of the effectiveness of the methods of detecting objects on the image, a comparative analysis was performed using an integral assessment of the quality of the detectors, using the area under the graph of the dependence of accuracy on completeness and mAP. During the comparison, integral indicators were invented that confirm the effectiveness of the developed accounting methods.

REFERENCES

[1] D. H. Ballard and C. M. Brown, Computer Vision, New Jersey, USA: Prentice Hall, 1982.

[2] S. Dasiopoulou, V. Mezaris, I. Kompatsiaris, V.-K. Papastathis, and M. G. Strintzis,"Knowledge-assisted semantic video object detection," IEEE Trans. Cir. Syst. for Video Techn., vol. 15, no. 10, pp. 1210-1224, 2005.

[3] Y. Jang, H. Gunes, and I. Patras, "Registration-free face-ssd: Single shot analysis of smiles, facial attributes, and affect in the wild," Comp. Vis. Image Und., vol. 182, pp. 17-29, 2019

[4] M. Babichev and V. Lytvynenko, Eds., Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making. Boston, MA: Springer Cham, 2022, pp. XVII, 721.

[5] G. Stockman and L. G. Shapiro, Computer Vision. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001

[6] . Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method based on improved yolov4-tiny," arXiv preprint arXiv:2011.04244, 2020

[7] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in Twenty-second int. joint conf. on AI. Citeseer, 2011.

[8] H. Zhou, F. Jiang, and H. Lu, "Ssda-yolo: Semisupervised domain adaptive yolo for cross-domain object detection," Comp. Vis. and Im. Und., vol. 229, p. 103649, 2023

[9] M. A. Feroz, M. Sultana, M. R. Hasan, A. Sarker, P. Chakraborty, and T. Choudhury, "Object detection and classification from a real-time video using ssd and yolo models," in Comp. Int. in Patt. Rec.: Proc. of CIPR 2021. Springer, 2022, pp. 37–47

[10] S. Kanimozhi, G. Gayathri, and T. Mala, "Multiple real-time object identification using single shot multibox detection," in 2019 ICCIDS. IEEE, 2019, pp. 1-5.

[11] K. Wadhwa and J. K. Behera, "Accurate real-time object detection using SSD," Int. Res. J. of Eng. and Techn., vol. 7, no. 5, 2020.

[12] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented r-cnn for object detection," in Proceedings of the IEEE/CVF Int. Conf. on Comp. Vis., 2021, pp. 3520-3529.

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in Computer Vision–ECCV 2016: 14th Eur. Conf., Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, 2016, pp. 21-37.

[14] O. Ferm, "Real-time object detection on Raspberry Pi 4: Fine-tuning a SSD model using tensorflow and web scraping," 2020.

[15] L. Liu and M. T. Özsu, Eds., Mean Average Precision. Boston, MA: Springer US, 2009, pp. 1703-1703.

Sensitivity Analysis of Enzyme-Substrate-Inhibitor Interaction Based on Nonlinear Dynamic Model

Vasyl Martsenyuk and Aleksandra Klos-Witkowska
University of Bielsko-Biala, Department of Computer Science, Willowa Str. 2, Bielsko-Biala, Poland
{*vmartsenyuk,awitkowska*}*@ath.bielsko.pl*

Keywords: Morris Sensitivity Analysis, Enzyme, Substrate, Inhibitor, Michaelis-Menthen Model.

Abstract: The development of electrochemical biosensors is cutting-edge in current research in medicine, biology, and ecology. The modeling and study of the enzyme-substrate-inhibitor interaction are required for biosensor design. The aim of the paper is to construct a three-stage model for an electrochemical biosensor and to perform its sensitivity analysis. The research is based on the Morris method. The main results are concerning the comparative impact of the model parameters related to biochemical reaction rates on the dynamics of the changes of the concentrations of enzyme, substrate, inhibitor, three complexes, and the reaction product throughout three stages. The results have both theoretical and practical relevance as the model parameters studied come from a real case study of biosensors for alpha-chaconine.

1 INTRODUCTION

Sensitivity analysis meaning the study of the effects of different parameters on outputs plays an important role in system research. Such impacts are studied based on both experimental and simulated data. Mathematically it can be described as the importance of factors in nonlinear relations. The complexity of the sensitivity analysis is essentially growing for the systems incorporating spatial and temporal relations.

Morris sensitivity was formulated in [1] at first. In [2] the method was extended by selecting a subset of trajectories from a large set, to maximize their spread and may the factors to deviate greatly. The work [3] generalizes Morris method for the dependent inputs.

Till now Morris analysis leaves a popular tool for the purpose of qualitative research of the systems. The comparison of Morris with the Sobol' techniques is presented in [4] for environmental models. In [5] Spearman, Sobol, and Morris approaches are compared for the purpose of radiological impact assessment of a nuclear power plant discharge.

Currently, the most promising approach appears to be global sensitivity one which is based on the concept of the active subspace against variance-based methods like Sobol' and Morris ones. In [6] active subspace technique is compared with variance-based methods using different test functions. Being more computationally effective, however, the development of the global sensitivity technique requires expanding methods for approximating the gradient of the model function.

Last time Morris method with modifications has actively been used in a variety of applications, including energy storage [7], building engineering [8], urban water supply [9], [10], environmental modeling [4], [11], safety systems [12] etc.

The given research is devoted to electrochemical biosensor designing which requires analysis of the interaction of multiple substances determined by a series of rate constants. Biosensor outputs are featured by the parameters, the temporal and spatial impacts of which should be investigated.

The reactions which are used in electrochemical biosensing come from the reactions that are catalyzed by an enzyme. They are commonly known as reversible [13] or irreversible [14] reactions. The irreversible one-complex Michaelis-Menthen (IR1CMM) mechanism is a keystone in modeling enzyme kinetics. Its reaction scheme

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

represents two-step process [15, 16, 17], where the enzyme E combines with the substrate S to form a complex C which then breaks down into the product P releasing E in the process.

Michaelis-Menthen model was further developed and applied for describing various reactions lying at the base of the electrochemical biosensor design. In

the work [18] we have developed and studied the generalized model of multi-substrate multi-inhibitor interactions using the law of the delayed mass action. The method for the estimation of the model parameters based on the machine learning technique was offered.

The objective of the given work is to implement the sensitivity analysis of the biosensor model with the respect to the impact of the reaction rates on the change of the concentrations of the substances.

The experimental data was gathered from the alpha-chaconin biosensor. For the purpose of the biosensor design, three-stage experiment was used [18]. To identify the parameters of the mathematical model of the biosensor for the determination of alpha-chaconin, a comparison of simulated and experimental responses was carried out. Potentiometric measurements were carried out after placing the transducers in the measuring cell, which was filled with 5 mM phosphate buffer with a pH of 7.0. The solution was stirred. After stabilization of the output signal (the first stage), the necessary amount of substrate was added to the measuring cell (the second stage), and after stabilization of the response to the substrate, certain volumes of concentrated solutions of alpha-chaconin were introduced to measure the level of inhibition (the third stage). Initially, based on the results of the experiment, the response of the butyrylcholinesterase biosensor to the addition of butyrylcholine chloride to the substrate cell and the subsequent introduction of alpha-chaconine was obtained.

2 MODEL DESCRIPTION

Let $n_e(t)$, $n_s(t)$, $n_i(t)$, $n_p(t)$, $n_{es}(t)$, $n_{ei}(t)$, $n_{esi}(t)$ be concentrations of enzyme, substrate, inhibitor, product, as well as enzyme-substrate, enzyme-inhibitory and enzyme-substrate-inhibitory complexes, which change over time t; k_s, k'_s, k_i, k'_i, and k_p be the corresponding rate constants of the forward and backward reactions of complex formation and the product; a be a constant whose numerical value determines the inhibition or activation of the enzyme. The change in product concentration time is directly proportional to the response of the biosensor.

The model of enzyme-substrate-inhibitor interaction is based on the following biochemical assumptions.

- The rate of the change of n_e is additionally proportional to the rates of forward and backward reactions of the enzyme with substrate and enzyme with inhibitor. Let k_s, k_i be the rates of the corresponding forward reactions, k'_s, k'_i be the rates of the backward ones.
- The product of the reaction of the enzyme with the substrate is formed with the rate k_p, which is also the dissociation rate for the complex enzyme-substrate.
- Let the formation rate of enzyme-inhibitor complexes is proportional to the concentration of free (available) enzymes n_e and available inhibitors n_i, and it leads to a decrease of n_e.
- Let the dissociation of enzyme-substrate and enzyme-inhibitor molecules increase the concentration of enzymes.
- Let a be a constant whose numerical value determines the inhibition or activation of the enzyme.

The equations

$$\begin{aligned}
\frac{dn_e}{dt} &= -k_s n_e n_s - k_i n_e n_i + k'_s n_{es} + k'_i n_{ei} + k_p n_{es} \\
\frac{dn_{es}}{dt} &= k_s n_e n_s - k'_s n_{es} - a k_i n_{es} n_i + a k'_i n_{esi} - k_p n_{es} \\
\frac{dn_{ei}}{dt} &= k_i n_e n_i - k'_i n_{ei} - a k_s n_{ei} n_s + a k'_s n_{esi} \\
\frac{dn_{esi}}{dt} &= a k_i n_{es} n_i - a k'_i n_{esi} + a k_s n_{ei} n_s - a k'_s n_{esi} \\
\frac{dn_s}{dt} &= -k_s n_e n_s - a k_s n_{ei} n_s + k'_s n_{es} + a k'_s n_{esi} \\
\frac{dn_i}{dt} &= -k_i n_e n_i - a k_i n_{es} n_i + k'_i n_{ei} + a k'_i n_{esi} \\
\frac{dn_p}{dt} &= k_p n_{es}
\end{aligned}$$

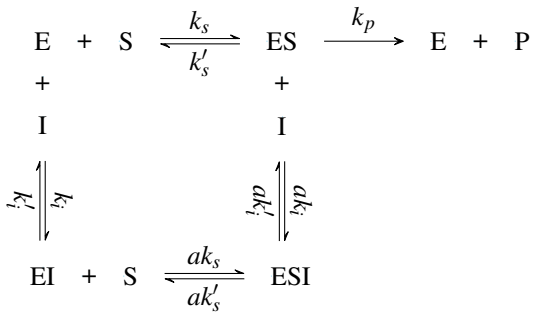
describe the biochemical reactions taking place for concentrations of enzyme, substrate, inhibitor, product, enzyme-substrate, enzyme-inhibitory, and enzyme-substrate-inhibitor.

The first equation is considered for enzyme concentration $n_e(t)$. The first term on the right-hand side, $-k_s n_e n_s$, represents the change in enzyme concentration due to the reaction going with a rate $-k_s n_e n_s$. The rate of this reaction is proportional to the enzyme concentration n_e and to the substrate concentration n_s. The negative sign in this differential equation means that the process of ES formation results in a decrease in the concentration of the enzyme E.

The next term, $-k_i n_e n_i$, similarly to the first term, accounts for the reaction E + I $\rightleftharpoons$ EI. The formation rate of EI complexes is proportional to the concentration of free (available) enzymes $n_e(t)$ and available inhibitors $n_i(t)$, and it leads to a decrease of $n_e(t)$, so it goes in negative.

Dissociation of ES and EI molecules increases the concentration of enzymes. It is taken into account by adding terms $k'_s n_{es}$ and $k'_i n_{ei}$. Formation of the product also releases enzyme molecules as $k_p n_e s$. All the

other equations are composed according to the following reactions



3 METHOD OF MORRIS SENSITIVITY ANALYSIS

Morris proposed his sensitivity analysis method in 1991 [1]. The method is described in details in [19]. This method is also called the elementary effects method [27], which can effectively identify and rank the importance of input parameters of a model by changing the value of only one parameter in an instance and finding its effect on the model output. Therefore, it is possible to calculate the "elementary effect (EE)" of each parameter on the output, one by one, and finally evaluate the influence of all of them on the results. On this basis, sensitivity factors can be compared globally and the nonlinearity of the model can be described qualitatively.

The experimental plan is composed of individually random One-At-a-Time (OAT) experiments. Each model input $X = (k_s, k_i, k'_s, k'_i, k_p, a) \in \mathbb{R}^6$ is assumed to vary across p selected levels in the space of the input factors. Hence, the region of experimentation, Ω, is a 6-dimensional p-level grid. According to the principle of Morris sensitivity analysis, the factors are assumed to be uniformly distributed in the range of $[0,1]$ and are then transformed from the unit hypercube to their actual distribution space.

For a given value of X, the elementary effect of the input factor k_s on the model solution n_e is defined as follows:

$$EE_{k_s}(X,t,\Delta) := \frac{n_e(X + e_1\Delta(k_s^{\max} - k_s^{\min}),t) - n_e(X,t)}{\Delta},$$

where Δ is a perturbation value that is selected from the collection $1/(p-1), ..., 1-1/(p-1)$, p is the number of levels, $X \in \Omega$ is any selected value in Ω, such that the transformed point $X + e_1\Delta(k_s^{\max} - k_s^{\min})$ is still in Ω, and $e_1 \in \mathbb{R}^6$ is a vector of zeros, but with a unit as its first component. The finite distribution of each elementary effect of the input factor k_s on output $n_e(t)$ is obtained by randomly sampling different

parameter	k_s	k_i	k'_s	k'_i	k_p	a
minimal value	50	2000	0.5	0.2	0.001	0.1
maximal value	5000	200000	50	20	0.1	0.9

Table 1: Ranges of values of the parameters analyzed.

X vectors from Ω and is denoted by $F_{k_s}^{n_e}$. The distribution for overall outputs is denoted be F_{k_s}.

In a similar way, we may introduce the elementary effects of arbitrary input factors $k_s, k_i, k'_s, k'_i, k_p, a$ on any output from $n_e, n_{es}, n_{ei}, n_{esi}, n_s, n_i, n_p$ and the corresponding distributions.

In addition, Morris proposed two sensitivity measures for each elementary factor, $\mu^\star$ and σ, which are, respectively, the mean and standard deviation of F_{k_s}. In order to estimate these quantities, Morris suggested performing sampling on r elementary effects from F_{k_s} via an efficient design that constructs r trajectories of $(k+1)$ points in the input space, each providing k elementary effects, one per input factor. The formula for computing $\mu^\star$ and σ is given by the following

$$\mu^\star_{k_s} = \frac{1}{N}\sum_{r=1}^{N} EE_{k_s,r}$$

$$\sigma_{k_s} = \sqrt{\frac{1}{N-1}\sum_{r=1}^{N}(EE_{k_s,r} - \mu^\star_{k_s})^2}$$

where $EE_{k_s,r}$ corresponds to the rth EE of k_s and N is the sample size. The higher the value of $\mu^\star$, the greater the influence of the corresponding parameters on the output value of the model. The higher the value of σ, the greater the interaction of a certain parameter with other parameters. Thus, the effect of that certain parameter has on the model output is nonlinear.

4 RESULTS

The effects of rates $k_s, k_i, k^\natural, k'_i, k_p, a$ on concentrations $n_e, n_{es}, n_{ei}, n_{esi}, n_s, n_i, n_p$ were investigated with the help of the method presented above. The sensitivity was analyzed for the parameters at the ranges presented in Table 1.

The charts on the Figure 1 show us the Morris indicators of sensitivity $\mu^\star$, σ for six parameters, namely, k_s, $k^\natural$, k_i, k'_i k_p, a, and their effect on n_e

The plot on the left side displays the change of mean value $\mu^\star$ of the effect of the parameters. We see that the biggest impact on n_e during the time interval (180-300s) is related to the dissociation rate $k^\natural$ (in green).

With time, the influence of the dissociation rate, $k^\natural$, on n_e decreases. Another analyzed factor is k_p.

As you can see after the 300s, this factor has a decisive influence on n_e. Moreover, it can be seen that the value of $\mu^\star$ for k_p is close to a constant value, which means that the influence of k_p on n_e is linear and additive. The least influence on n_e at this stage makes k_s, which decreases over time.

From the σ diagram it can be seen that the seriality of k'_s, k_p, k_s is preserved for non-linearity and the level of interaction of the influence of parameters on n_e. All other parameters have no effect on n_e, which can also be seen from the equations of the model at this stage.

The charts in the Figure 2 show us the Morris indicators of sensitivity $\mu^\star$, σ for six parameters, namely, k_s, k'_s, k_i, k'_i, k_p, a, and their effect on n_s.

The plot on the left side displays the change of mean value $\mu^\star$ of the effect of the parameters. We see that the biggest impact on n_s during the time interval (180-460s) is related to the rate of formation of the enzyme-substrate complex (k_s) which increases with time similar to the speed of product development (k_p). With time, the influence of the dissociation rate, k'_s on n_s decreases.

In the time frame (460-750s), the influence of factors k'_s, k_i, k_s, k_p, k_i is noticeable and it increases with time. It is worth noticing that influence on n_s comes from the dissociation rate k'_s (in green) is greater than influence comes from k_i, k_s, k_p, k'_i parameters. The effect of the factors on parameter σ (on the right) looks very similar except for the effect of the parameters k_p and k'_i. In the time interval (460-750s) the effect coming from the dissociation of the enzyme-inhibitor complex (k_i) outweighs the effect coming from the formation of the reaction product (k_p).

The charts in the Figure 3 show us the Morris indicators of sensitivity $\mu^\star$, σ for six parameters, namely, k_s, k'_s, k_i, k'_i, k_p, a, and their effect on n_i. The plot on the left side displays the change of mean value $\mu^\star$ of the effect of the parameters. In a time frame (180-460s) significant effect on the inhibitor concentration (n_i) of the dissociation constant of the enzyme-inhibitor complex (k'_i) was noted. This influence increases with time. Other parameters k_i, k_p, k_i, k_s are constant over time. In time intervals (460-750s) value that comes from the formation of the reaction product is close to constant, which means that the effect of k_p on n_i is linear and additive. The nature of the changes seen in the graphs σ is similar to $\mu^\star$.

Figure 4 illustrates sensitivity analysis indicators from n_{es} at different stages of modeling. As time passes (180-460s) on left, the effect on n_{es} of the parameter k'_s (in green) and k_s (red) is noticeable, which decreases with time. In contrast, the parameter k_p, denotes the formation of the reaction product which is almost constant over time with a slight upward trend.

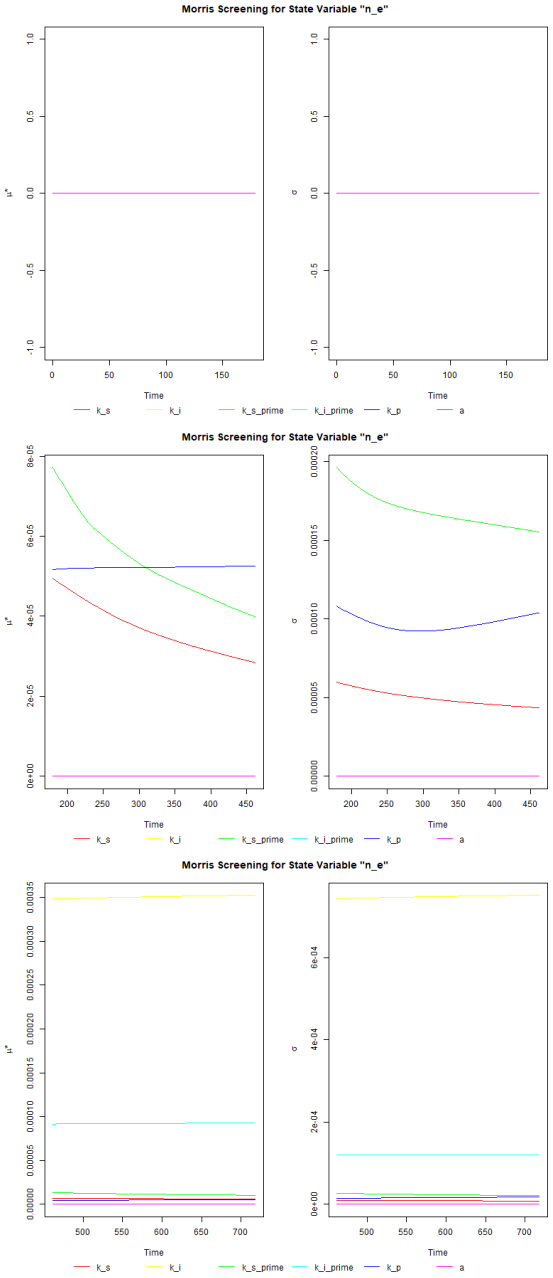


Figure 1: Sensitivity analysis indicators for n_e at the different stages of modeling.

The similar nature of the changes for k'_s and k_s is shown in the graph for σ, on right. The parameter k_p slightly decreases up to 270s, in the interval (271-320s) it remains constant, while from 350-460s) a slight upward trend is noticeable.

Taking into consideration the change of mean value $\mu^\star$ of the effect of the parameters in the time frame (460-750s) the biggest influence on n_{es} comes from k_i (binding constant of complex enzyme-

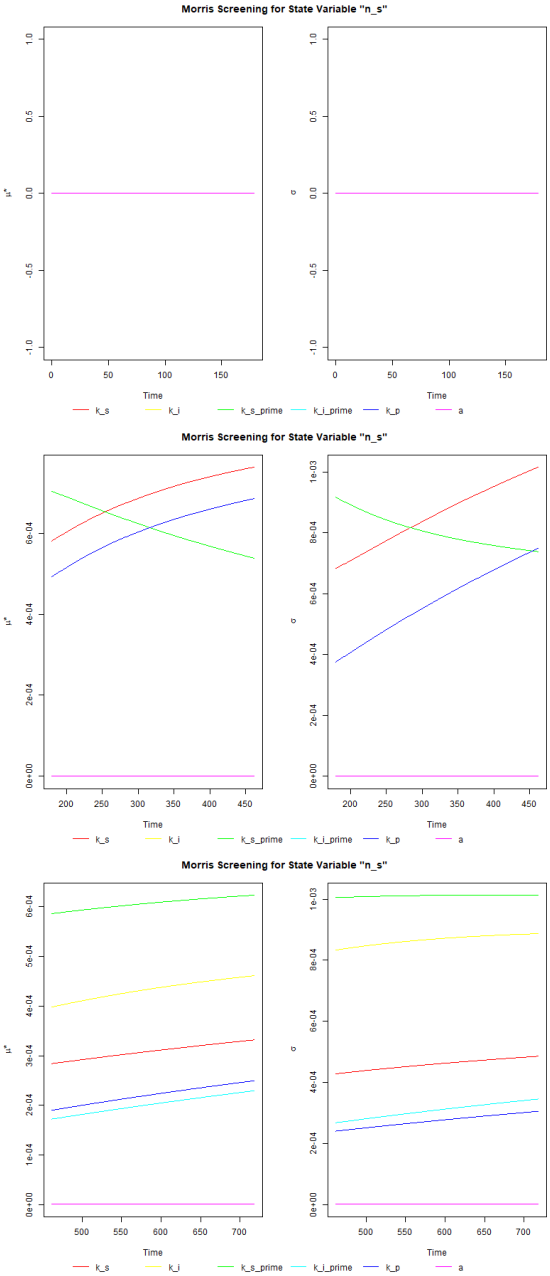


Figure 2: Sensitivity analysis indicators for n_s at the different stages of modeling.

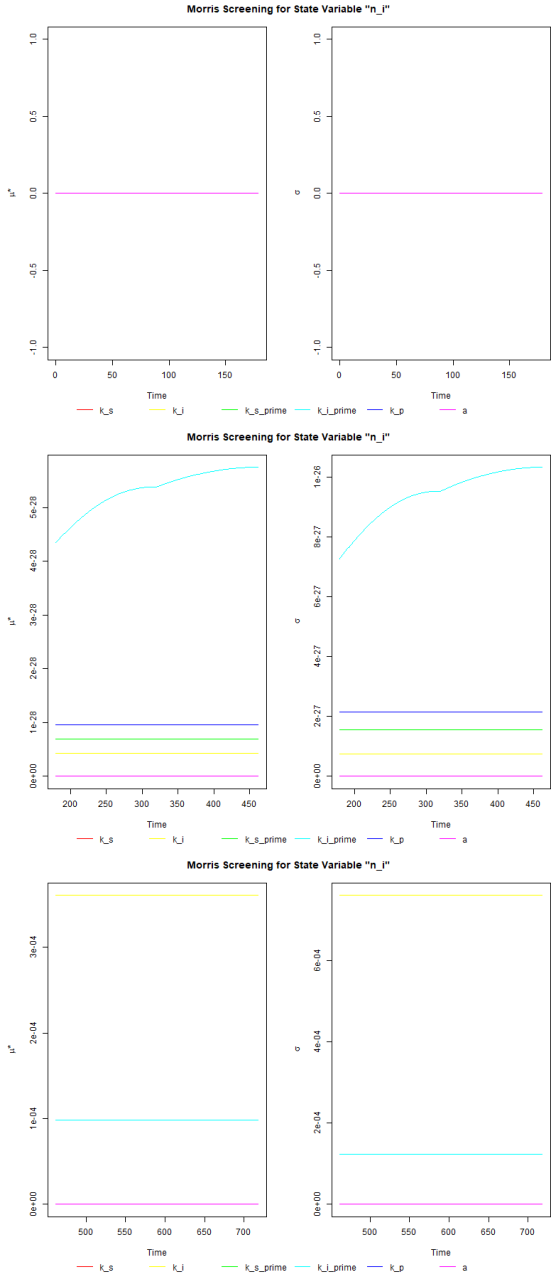


Figure 3: Sensitivity analysis indicators for n_i at the different stages of modeling.

inhibitor) like k_s' (in green), k_i' (in blue) and k_s (in red) the influence of these factors decreases with time. The influence of the parameter (k_p) derived from the formation of the product reaction increases with time. A similar character of changes is shown in the graph for σ.

Figure 5 illustrates sensitivity analysis indicators from n_p at different stages of modeling. As time passes (180-460s) on left, the effect on product concentration of the parameters k_s, k_p, and k_s' is noted. With the passage of time there is a noticeable increase in the influence of the parameters k_p and k_s (for k_s minimally larger) in contrast to the parameter k_s', which decreases over time.

In the time frame (460-750s) There is a noticeable effect on the concentration of the product (n_p) of all parameters, i.e. parameters k_i, k_s', k_p, $k_i\prime$ and k_p. This influence increases with time.

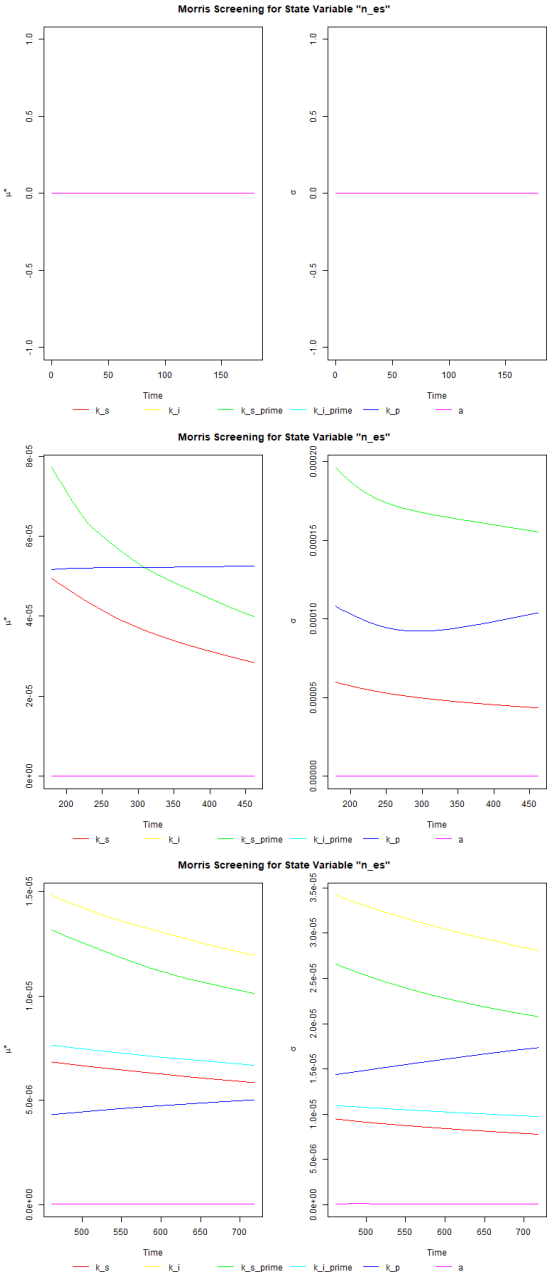


Figure 4: Sensitivity analysis indicators for n_{es} at the different stages of modeling.

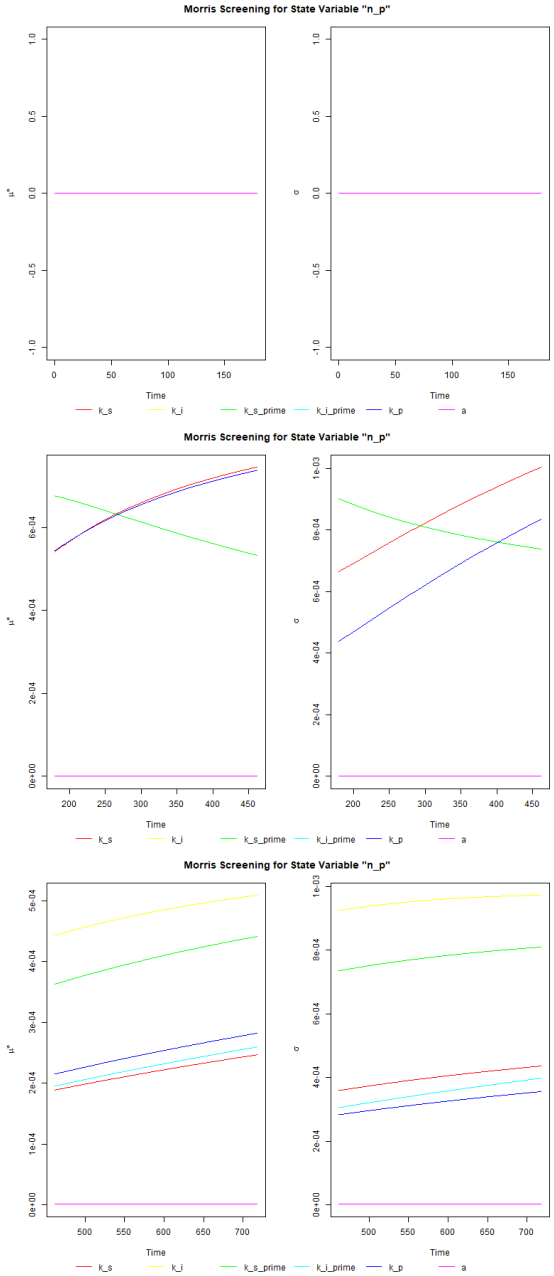


Figure 5: Sensitivity analysis indicators for n_p at the different stages of modeling.

A similar nature of the changes was observed for the σ except for the influence of the parameter k_p, which in the time interval (460-750s) for $\mu^\star$ exceeds the influence from $k^\star$ and k_p in contrast to σ, where its influence is the smallest.

On Figure 6 on the left side, displays the change of mean value $\mu^\star$ of the effect of the parameters on n_{ei}. We see a definite increasing effect over time of the dissociation rate of enzyme-inhibitor complex (k_i') in the time interval (180-460s). Other parameters are constant over time.

A similar character of changes was observed for σ, on the right side. In the time frame (460-750s) character of changes for $\mu^\star$ and σ is also similar. We see the dominant influence of the factor derived from the formation of the enzyme-inhibitor complex (in yellow), which over time is almost constant with a slight upward trend, as is the k_p factor. The k_s' factor de-

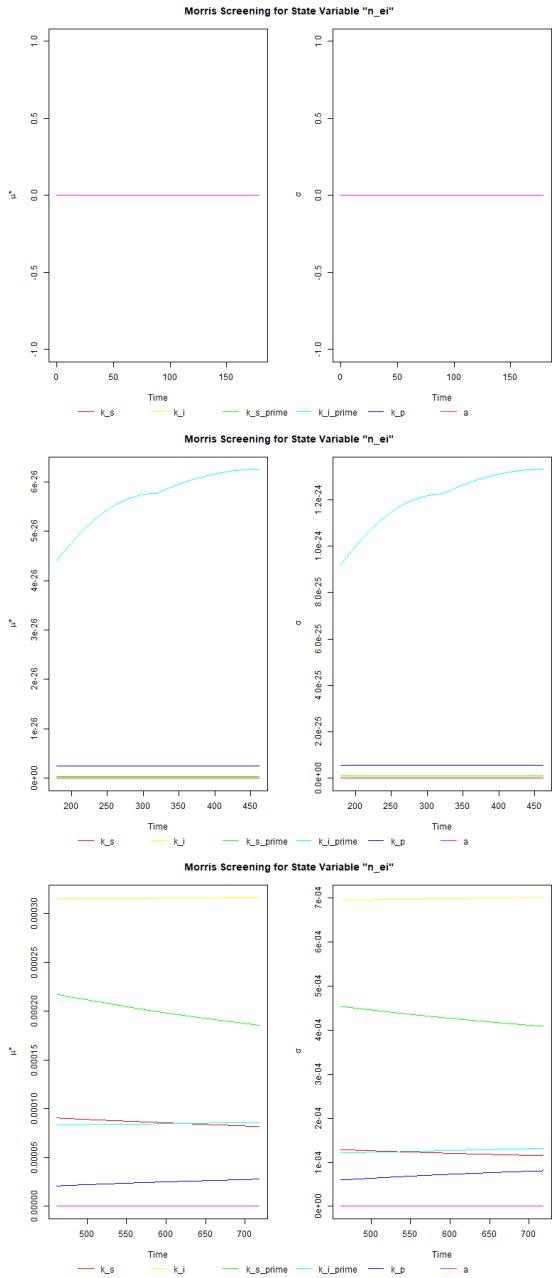


Figure 6: Sensitivity analysis indicators for n_{ei} at the different stages of modeling.

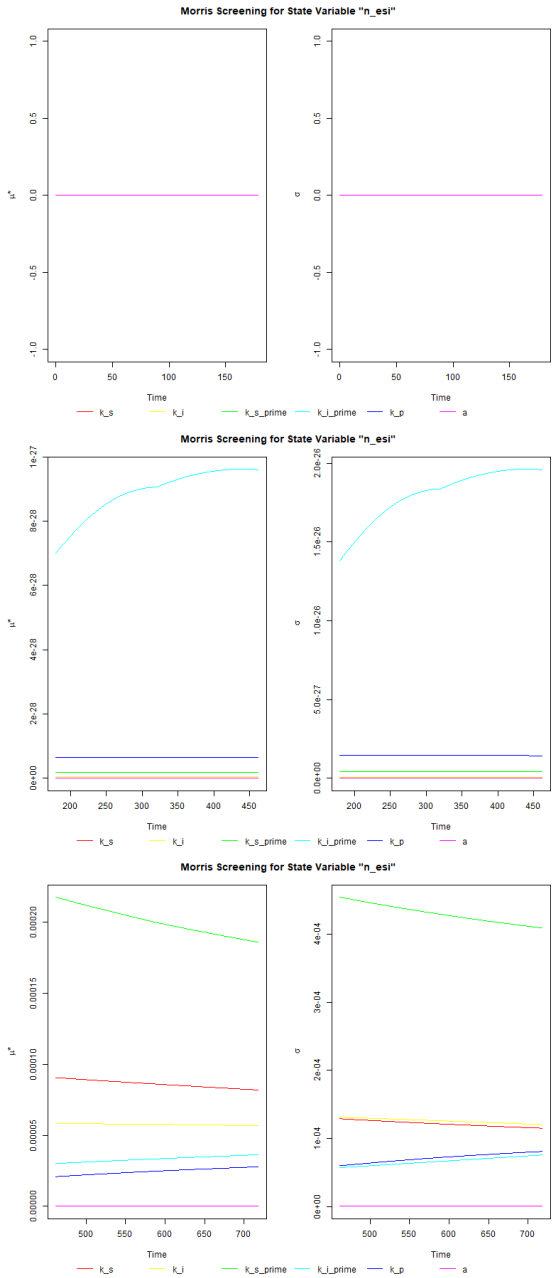


Figure 7: Sensitivity analysis indicators for n_{esi} at the different stages of modeling.

rived from the dissociation rate of complex enzyme-substrate decreases with time, as the k_{si} and k_i' factors for which the nature of the changes is smoother.

As is shown on Figure 7 on left, in time frame (180-460s) significant effect on the inhibitor concentration (n_i) of the dissociation constant of the enzyme-inhibitor complex (k_i') was noted. This influence increases with time. Other parameters k_i, k_p, k_i, k_s are constant over time.

In the time interval (460-750s) the greatest effect on n_{esi} on k_s' (dissociation rate of the enzyme-inhibitor complex), which decreases with time similarly to k_s (binding constant of enzyme-substrate complex). The influence of the parameter k_i on n_{esi} is invariant over time, while the influence of the parameters k_i' and k_p presents a slight upward trend with time. The trend of changes for σ is similar, although it is possible to weigh in the range (460-750s) reduced

influence of binding constant of complex enzyme-substrate(k_s) (in red) and equalization the influence coming from k_i' and k_p.

5 CONCLUSIONS

For the reasons given in the work we have shown the Morris method is powerful tool for sensitivity analysis of the models for biochemical reactions. Here we have applied the method for studying the enzyme-substrate-inhibitor interactions which is used during the design of electrochemical biosensor.

The method can be applied for the comparative investigation of the influence of the parameters (factors) on the model outputs (trajectories). Moreover such study can be conducted for the various stages of the experiment.

In turn it is of importance when constructing the calibration curves basing on the responses of the biosensor.

ACKNOWLEDGMENTS

The authors are grateful to reviewers for valuable comments that allowed to improve the paper significantly. The paper presents applied artificial intelligence good practice related to enzymatic analytics basing on modeling with the help of dynamical systems. It was fulfilled within the framework of the work package 2 of the Erasmus+ project FAAI [20] (project no. 2022-1-PL01-KA220-HED-000088359), which was devoted to the comprehensive investigation of the good practices in applied artificial intelligence, including the biosensor design. In particular, financing of management, dissemination activities, meetings, and intellectual output, which follow from this research, was conducted from the project budget. On the other hand, all activities related to preparing and publishing the paper were financed by the University of Bielsko-Biala within the framework of the discipline "Technical Informatics and Telecommunications".

REFERENCES

[1] M. D. Morris, "Factorial sampling plans for preliminary computational experiments," Technometrics, vol. 33, no. 2, pp. 161-174, May 1991, [Online]. Available: https://doi.org/10.1080/ 00401706.1991.10484804.

[2] J. P. Norton, "Selection of Morris trajectories for initial sensitivity analysis," IFAC Proceedings Volumes, vol. 42, no. 10, pp. 670-674, 2009, 15th IFAC Symposium on System Identification, [Online]. Available: https://doi.org/10.3182/ 20090706-3-FR-2004.00111.

[3] Q. Ge and M. Menendez, "Extending Morris method for qualitative global sensitivity analysis of models with dependent inputs," Reliability Engineering System Safety, vol. 162, pp. 28-39, 2017, [Online]. Available: https://doi.org/10.1016/j.ress.2017.01.010.

[4] M. Jaxa-Rozen and J. Kwakkel, "Tree-based ensemble methods for sensitivity analysis of environmental models: A performance comparison with sobol and morris techniques," Environmental Modelling Software, vol. 107, pp. 245-266, 2018, [Online]. Available: https://doi.org/10.1016/j.envsoft.2018. 06.011.

[5] V. Nicoulaud-Gouin and et al., "Sensitivity analysis in a radiological impact assessment of a nuclear power plant discharge. a comparison of the morris, spearman and sobol' approaches," Journal of Environmental Radioactivity, vol. 242, p. 106770, 2022, [Online]. Available: https://doi.org/10.1016/j.jenvrad.2021. 106770.

[6] X. Sun, B. Croke, A. Jakeman, and S. Roberts, "Benchmarking active subspace methods of global sensitivity analysis against variance-based sobol' and morris methods with established test functions," Environmental Modelling Software, vol. 149, p. 105310, 2022, [Online]. Available: https://doi.org/ 10.1016/j.envsoft.2022.105310.

[7] A. Janse van Rensburg, G. van Schoor, and P. van Vuuren, "Stepwise global sensitivity analysis of a physics-based battery model using the morris method and monte carlo experiments," Journal of Energy Storage, vol. 25, p. 100875, 2019, [Online]. Available: https://doi.org/10.1016/ j.est.2019.100875.

[8] Y. Shen and M. Yarnold, "A novel sensitivity analysis of commercial building hybrid energy-structure performance," Journal of Building Engineering, vol. 43, p. 102808, 2021, [Online]. Available: https://doi.org/10.1016/j.jobe.2021.102808.

[9] D. King and B. Perera, "Morris method of sensitivity analysis applied to assess the importance of input variables on urban water supply yield – a case study," Journal of Hydrology, vol. 477, pp. 17-32, 2013, [Online]. Available: https://doi.org/10.1016/j.jhydrol. 2012.10.017.

[10] E. A. Baker, A. Cappato, S. Todeschini, L. Tamellini, G. Sangalli, A. Reali, and S. Manenti, "Combining the Morris method and multiple error metrics to assess aquifer characteristics and recharge in the lower Ticino Basin, in Italy," Journal of Hydrology, vol. 614, pp. 128536, 2022, doi: 10.1016/j.jhydrol.2022. 128536.

[11] L. Gatel, C. Lauvernet, N. Carluer, S. Weill, J. Tournebize, and C. Paniconi, "Global evaluation and sensitivity analysis of a physically based flow and reactive transport model on a laboratory experiment," Environmental Modelling Software, vol. 113, pp. 73-83, 2019, doi: 10.1016/j.envsoft. 2018.12.006.

[12] C. Wang, M. Peng, and G. Xia, "Sensitivity analysis based on Morris method of passive system performance under ocean conditions," Annals of Nuclear Energy, vol. 137, pp. 107067, 2020, doi: 10.1016/j.anucene.2019.107067.

[13] B. P. English and et al., “Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited,” Nature Chemical Biology, vol. 2, no. 2, pp. 94, Dec. 2005, doi: 10.1038/nchembio759.

[14] B. Palsson, “On the dynamics of the irreversible Michaelis-Menten reaction mechanism,” Chemical Engineering Science, vol. 42, no. 3, pp. 447-458, 1987.

[15] J. Keener and J. Sneyd, Eds., Mathematical Physiology. Springer New York, 2009, doi: 10.1007/978-0-387-75847-3.

[16] S. Schnell, “Validity of the Michaelis-Menten equation - steady-state or reactant stationary assumption: that is the question,” FEBS Journal, vol. 281, no. 2, pp. 464-472, Nov. 2013, doi: 10.1111/febs.12564.

[17] W. D. Ristenpart, J. Wan, and H. A. Stone, “Enzymatic reactions in microfluidic devices: Michaelis-Menten kinetics,” Analytical Chemistry, vol. 80, no. 9, pp. 3270-3276, May 2008, doi: 10.1021/ac702469u.

[18] V. Martsenyuk, A. Klos-Witkowska, S. Dzyadevych, and A. Sverstiuk, “Nonlinear analytics for electrochemical biosensor design using enzyme aggregates and delayed mass action,” Sensors (Basel), vol. 22, no. 3, pp. 980, Jan. 2022.

[19] J. Ren, W. Zhang, and J. Yang, "Morris sensitivity analysis for hydrothermal coupling parameters of embankment dam: A case study," Mathematical Problems in Engineering, vol. 2019, pp. 1-11, Jun. 2019.

[20] “FAAI Job Hub – The Future is in Applied Artificial Intelligence 2022-1-PL01-KA220-HED-000088359,” http://faai.ath.edu.pl, [Accessed on Feb. 24, 2023].

Random Forest Algorithm in Unravelling Biomarkers of Breast Cancer Progression

Nadiia Kasianchuk[1,2], Dmytro Tsvyk[3], Eduard Siemens[4] and Halina Falfushynska[4,5]
[1]*Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznańskiego Str. 6, Poznań, Poland*
[2]*Faculty of Pharmacy, Bogomolets National Medical University, Taras Shevchenko Str. 13, Kyiv, Ukraine*
[3]*Institute of International Relations, Taras Shevchenko National University of Kyiv, Illyenka Str. 36, Kyiv, Ukraine*
[4]*Anhalt University of Applied Sciences, Bernburger Str. 55, Köthen, Germany*
[5]*Institute of Biological Sciences, University of Rostock, Albert Einstein Str. 3, Rostock, Germany*
nadkas2@st.amu.edu.pl, tsvykdima@gmail.com, halina.falfushynska@uni-rostock.de, eduard.siemens@hs-anhalt.de

Keywords: Breast Cancer, K-Means, Random Forest, Genes, Prognostic Factors, Classification, Biomarkers.

Abstract: Breast cancer is the leading cause of cancer death among women. As its development involves a multidimensional network of gene-environment interactions, advanced data analysis tools and bioinformatics are vital to uncover the nature of cancer. The initial database contained the expression values of 19737 genes in 1082 patients. Random Forest algorithm was used to distil the genes with the strongest influence on four substantial prognostic factors (survival period, tumour size, lymph node seizure, and metastasis). The obtained set consists of 230 potential biomarkers that facilitate the critical cancer-related pathways, such as p53, Wnt, VEGF, UPP, thereby influencing cell proliferation, tumouri- and angiogenesis. A considerable contrast in the expression was shown between the patients at different stages of cancer progression. The obtained set will simplify the diagnostics and prediction of tumour progression, enhance treatment outcomes and elaborate better strategies for curing breast cancer.

1 INTRODUCTION

Cancer is reported to be the second most common cause of death globally, accounting for an estimated 10 million deaths, or one in six in 2020[1] with breast cancer taking 2.26 million lives annually. As indicated by the World Health Organization (WHO), the current dramatic rise in the total number of diagnosed breast cancer cases is precipitated by alterations in human lifestyle and increased life expectancy. Though, the growing general awareness of the problem and enhanced screening technologies should not be overlooked as a contributing factor. With that, however, it is yet a long way to go, as people in low-middle-income countries and vulnerable strata in high-income states still face higher mortality rates. The reasons thereof include insufficient availability, affordability and accessibility of cancer care, as well as limited access to clinical innovations [1]. Regardless of the wealth, accumulated in a country, the implementation of cancer screening programmes is of the highest necessity, since it saves lives and also funds to be invested into treatment.

Cancer control is chiefly aimed at mitigating the prevalence and mortality rate of cancer [1, 2]. The disease screening programmes implement systematic evidence-based early diagnosis, the cornerstone whereof being a selection of suitable biomarkers of malignant neoplasm and a study of target pathways that aggravate cancer progress. Thus, the proper and in-time diagnosis and personalised therapy curb mortality and disability rates among persons affected by common cancers.

Such clinical biomarkers for breast cancer, as estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) provide vital material for breast cancer prognosis and insights into the patient's probable reaction to treatment. That notwithstanding, other parameters are also deemed prospective to this end. *BRCA1*, *BRCA2*, *CXCR4*, caveolin, miRNA, and FOXP3, despite additional expenses, may come in handy to describe breast malignant neoplasm and tumour progression, to prevent metastasis and recurrences [3]. Yet, in step with the diversification of the raw data obtained, the analysis thereof becomes increasingly complex and demanding, however, a

[1]https://www.who.int/news-room/fact-sheets/detail/cancer

handsome reward for the work well performed constitutes the remarkably uplifted prognostic precision [4].

Hence, up-to-date risk assessment tools and novel approaches are rendered indispensable as modern science is broadening the range of applicable multi-gene tests and various biomarkers, including genomic, biochemical, and histopathological signatures [5]. Such a comprehensive data processing may be efficiently conducted only by means of transformative AI and machine learning solutions, so that the fruits thereof would provide valuable prognostic signals.

Whereas some pioneering attempts have been taken in the field [4,6,7,8], meaningful and comprehensive options for integrative data-processing are still limited. The key purpose of this work is the development of a proper network of the most promising breast cancer biomarkers. The paper marks our starting point in shedding light on mighty, but feasible data analysis tools for biomedical research. The integration of the IT data-processing solutions into life science and medical institutions will, with regards to breast cancer, contribute to detecting biomarker discrepancies and minimalising misreported biomarker status.

2 MATERIALS AND METHODS

2.1 Data Preparation

The database presented by The CGA through cBioPortal website[2] contains clinical pictures of breast cancer patients supplemented by the expression values of 19737 genes. The latter is described as z-scores relative to normal samples (log RNA Seq V2 RSEM), thus showing the number of standard deviations below or above the population mean.

$$z = \frac{x - \mu}{\sigma}$$

Data cleaning and pre-processing are standardly applied to the obtained information. All indicators with missing values were excluded from further analyses to avoid inaccurate interpretations. Furthermore, genes with statistically outlying z-scores (e.g. *OR2B2*) were also justifiably removed.

Python environment version 3.9.13 (Python Software Foundation) was employed for the purposes of the present analyses, with NumPy, pandas and Matplotlib libraries being used for data manipulations. Sklearn package was of much help for the algorithms of classification, regression and clustering.

2.2 K-Means Data Clustering

All samples were clustered according to their survival rate, so that the workflow of the Decision Tree could be optimised.

K-means represents an unsupervised clustering algorithm, praised in data analysis mainly due to its simplicity, efficacy and high-speed data processing. Firstly, k objects are randomly chosen as centroids across the data, whereto other samples are then being assigned. Such process is based on the smallest distance between data sample and existing clusters, measured in the Euclidean distance metric.

$$d = \sqrt{((x_2 - x_1)^2 - (y_2 - y_1)^2)}$$

The number of clusters was drawn on the Silhouette score, calculated in the following way.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

The coefficient has a range of [-1;1]. S=1 indicates perfect clusterisation (the distance between any elements in the cluster is significantly lower than that between any in neighbouring clusters). S=0 shows that some samples are located right upon or near the line, delimiting two adjacent clusters. S=-1 depicts poor clustering choice. The highest Silhouette coefficient is obtained for k = 2 and k = 3, however values for k = 4 and k = 5 are still acceptable (Figure 1).

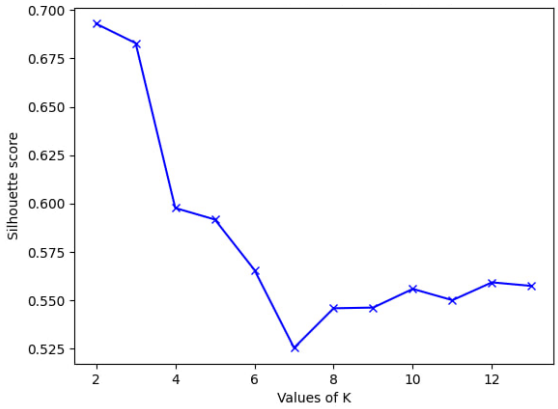


Figure 1: Silhouette analysis for Optimal k.

2.3 Random Forest Classification

For the purposes of classifying genes, Random Forest was trained on the prepared data, with gene expression values shown as z-scores relative to

[2]https://www.cbioportal.org

normal samples and survival expressed in months. The samples were clustered according to their survival. The further classification was performed based on the affiliation with a specific cluster. Categorical variables needed for the Random Forest (e.g. presence of metastasis), were transformed into numerical values (e.g. M0 – 0, M1 – 1).

Such hyperparameters as quantity of decision trees in the random forest and maximum depth of each tree were checked and chosen for the best performance and computing speed.

3 RESULTS AND DISCUSSION

3.1 Main Features of the Data Cohort

After cleaning and preparation, the analysed database contained the information about 1082 breast cancer patients, 1070 (98,89%) of whom were females and 12 were (1,11%) males (Table 1).

Table 1: Basic characteristics of the dataset.

Sex					
Female				98,89%	
Male				1,11%	
Diagnosis Age, years					
Mean	58,40 ± 13,19	Median	58	Mode	62
TNM (tumour–node–metastasis), %					
T1	25,51	N0	47,32	M0	82,62
T2	57,95	N1	32,81	cM0	0,55
T3	12,66	N2	11,00	M1	1,94
T4	3,60	N3	7,02	MX	14,88
TX	0,28	NX	1,85		
Tumour Type, %					
Infiltrating Ductal Carcinoma				71,53	
Infiltrating Lobular Carcinoma				18,58	
Medullary Carcinoma				0,55	
Metaplastic Carcinoma				0,74	
Mixed Histology (NOS)				2,68	
Mucinous Carcinoma				1,57	
Other				4,34	
Cancer subtype, %					
BRCA_LumA				46,12	
BRCA_LumB				18,21	
BRCA_Basal				15,80	
BRCA_Her2				7,21	
BRCA_Normal				3,33	
NA				9,33	

The mean diagnosis age was 58,40 ± 13,19 years. Median value for the same indicator equalled to 58 and the mode was 62 years. The most common type of tumour was Infiltrating Ductal Carcinoma (71,53%), followed by Infiltrating Lobular Carcinoma (18,58%). 2 samples with tumours, labelled as Breast Invasive Carcinoma and Infiltrating Carcinoma (NOS), were moved to group 'Other', for they did not meet the necessary criteria.

The American Joint Committee on Cancer introduced the tumour–node–metastasis (TNM) cancer staging system, which is used to indicate the severity of cancer. In such system, T-value characterises the size and the extent of a tumour, N shows the presence of cancer cells nearby or in the lymph nodes and M indicates whether cancer metastasised (spread to the distant parts of the body). The majority of patients from the database had T2 tumour stage (57,95%) and no metastases (M0 – 82,62%). 47,32% of the patients had no cancer cells near the lymph nodes (N0) and 32,81% had mild affection of lymph nodes (N1). Where TNM indicators mattered for the analyses, patients without TNM scores were not taken into consideration.

3.2 Clusterisation

The patients were clustered regarding their overall survival period. The number of clusters was selected due to the Silhouette coefficient. As $k = 2$ and $k = 3$ had the highest score, both were processed through the Random Forest algorithm. For $k = 3$, the high survival group comprised only 18 patients, medium and poor survival clusters numbered 254 and 810 entries respectively. For $k = 2$, the lower survival group was also considerably bigger (846 compared to 235 of higher survival). The noticeable imbalance in data distribution owes to the data's background.

3.3 Classification

Random Forests possess excellent prediction accuracy, ergo, they are employed particularly in forecasting treatment response in cancer cell lines, in localising tumour and in identifying its stage in patients.

In light of such massive amount of data, reliance solely on the Random Forest results might compromise the outcome of the analysis. The issue was handled by building several Random Forest Decision Trees trained on a set of key indicators of cancer severity (survival clusters for $k = 2$ and $k = 3$, metastasis, tumour size, nodes, etc.). The number of Decision Trees in the Random Forest varied from 50 to 100 and the depth of each tree was set at 5-7 to optimise the computing speed and to avoid overfitting of models.

Eventually, the number of genes was condensed from 19737 to several dozens, on which the further analyses were carried out (Table 2).

Table 2: Genetic prognostic markers classified by the Random Forest.

Group	Number of genes
Survival (2 clusters)	50
Survival (3 clusters)	59
Tumour size	94
ZNF497, C9orf106, GPR89B, AVP, CUEDC2, PTGIR, MCOLN1, LEKR1, GUSBP19, ZNF696, PRELID2, MTA1, SH2D7, DCP1B, CMTM2, C6orf26, SGCG, GNA12, CLIN, C13orf38, PAQR5, ZFPM1, MED1 ,HAPLN3, SLC2A10, C13orf23, B3GALT4, GREM2, THG1L, FAM55B, CYP3A7, C2orf47, FDX1, AKR1D1, ATP1B4, PPIF, NOS1, COQ10B, TARBP1, EIF4A1, OPLAH, TMPRSS15, ZDHHC21, ACBD6, TUBA4A, BTN2A1, FAM181A, ENTPD5, CUZD1, POM121L1P, IL22, FAM71F1, C15orf52, RDH16, LTA, SLC5A10, EBAG9, MED8, IL18BP, C3orf38, ASB17, PE12L, KLHL32, C3orf10, EMILIN1, FAM189B, MLYCD, FRA, C12orf34, C16orf96, BEX5, RFXANK, USP38, WDR36, CDKN1B, FHOD1, CADM2, SLC2A4, DHTKD1, CD84, GPR171, VPS18, SELPLG, C1orf177, TMEM181, CLASP2, C6orf162, NKD1, LY9, DNMB, MFAP3L, PTMA, CCL15CCL14, WASF2	
Cancer in lymph nodes	112
CSDE1, MKI67IP, HIAT1, ZNF711, ADRA2A, GCHFR, ALDOA, PMS2CL, BLVRA, MARS, SMUG1, ZNF542, CLEC1A, EGR4, C14orf4, ZNF629, AGGF1, C12orf65, SLC37A3, LOC100271832, OR1J4, EYA4, C10orf105, PDHX, ECHDC3, RFPL2, CPSF4, ITCH, SEC61B, ITIH3, C12orf47, PMS2L3, IGFBP5, CPNE7, PRKAR2A, FNBP4, MLLT1, JMJD5, SLC22A9, GABRA3, ZNF384, WRAP53, REEP6, RBM39, CLEC12A, PYGL, CTDSP2, NMT1, C1orf130, TMEM127, DYTN, FBXO46, PYGO2, SPCS2P4, BOLA1, ITGB4, ARHGEF5, TPM3P9, TDRKH, AGFG1, ERMN, PPP2R5A, FRYL, ARSG, KIF14, HLAB, RSPH6A, IWS1, LOC90784, CCDC23, SMIM6, PRRC1, FBXO33, TMEFF2, CPAMD8, RBM7, MAPK8IP3, CXorf30, CCPG1, MMR, MRPL41, CDK1, PCDH17, PARVA, TTPA, RUVBL2, ABCA11P, OR4N4, COBL, OGDHL, RPL23P8, ZIM2, BAZ2B, SH2B2, FKBP6, TECPR1, VPS37A, XKR9, MREG, KLHL31, C1orf52, CCDC59, DDX4, SRD5A1, C2orf34, RANBP3, LRR, WFDC5, GOLGA9P, ENPP5, SOX15, MAD2L2	
Metastasis	24
DPY19L3, LONRF1, SETDB1, C2orf3, MRPL9, GFAP, ZNF516, C9orf11, FAM2B, LRRC37A16P, ERP27, RNF121, FAM22A, PM2D2, LGALS4, EIF4E1B, SNORA13, WDR67, SASH3, CAPN7, C6orf81, RASA4CP, NUTM2A-AS1, PM20D2	

3.4 Tumour Stage

Tumour size is a centrepiece in diagnostics and prognosis of cancer, hence, it underlies the widely used TNM staging system. The Random Forest algorithm identified 94 genes which affect UPP, TGFβ1 and TLR4-dependent pathways, as well as the coverage of the progestin and adipoQ receptor family.

As the present findings show, *WASF2* stiffly impacts tumour growth [9]. It is notably to blame for activating the actin-related protein 2/3 complex, which enables migration and invasion of cancer cells. The upregulation of *PTMA* exacerbates histological malignancy and heightens the possibility of cancer recurrence [10]. The properties alike are viewed as potential therapeutic targets for cancer treatment.

CUEDC2 provokes endocrine resistance in patients by inhibiting estrogen and progesterone receptors via the ubiquitin-proteasome pathway. Therefore, it undermines the effect of such first-line treatment as tamoxifen in estrogen receptor-positive breast cancer patients [11].

PAQR5 regulates cyclic adenosine monophosphate synthesis and manipulation of *MAPKs*. The downregulation of *PAQR5* has its hand in the increased methylation of the promoter DNA and in a poor survival outcome in renal cancer patients [12].

Such genes as *PAQR5* and *CUEDC2* still lack rightful scientific attention and thus have not yet been institutionalised as prognostic markers. As our study indicates, the exploratory horizons of the matter in question are much to be broadened.

3.5 Lymph Nodes

Distant metastatic activity of cancer poses a pervasive threat for human health and wanes the chances for survival. Given the onset of the tumour cells spread usually takes place in the lymph nodes, such developments are regarded as a proven prognostic factor [13].

112 genes were chosen by the Random Forest algorithm and subsequently structured by the expression values in N0 group (no cancer in lymph nodes) (Figure 2). Predictably, the highest divergence is shown between the group free of seizures and that with the highest number of the invaded nodes. 64 genes of 112 chosen are downregulated in patients (N0-N2 stages) with *CTDSP2*, *RMP7*, and *C12orf65* being the most altered. The selected genes engage in such pathways as p53, Wnt, PI3K/Akt, VEGF and mtRQC, thereby influencing cell proliferation, tumouri- and angiogenesis.

The prognostic potential of the chosen genes is corroborated by recent studies. *CPSF4* is linked to the unfavourable cancer outcome, since it tends to upregulate the key cancer development genes, such as *MDM4* and *VEGF* [14]. *REEP6* itself showed almost 4-fold increase in the z-score in N3 group in contrast to N0, which implies its prospects as a novel

biomarker, resonating with a freshly published study on the Triple-Negative Breast Cancer prognosis [15].

GCHFR is considerably upregulated in cancer tissues. Furthermore, its expression has been associated with the activation of ferroptosis which renders a tumour resistant to therapy [16].

Pygo2 is the most overexpressed gene in the chosen set. It gives rise to the sensitivity to common chemotherapeutics, as well as promotes dedifferentiation and progression of cancer cells [17]. Such effect may have roots in the gene's ability to target the Wnt/-catenin pathway, which is critical for the proper cell proliferation and development. Although *Pygo2* has already undergone some study as a therapeutic target for metastatic breast cancer, the data on the matter is still insufficient, hence, the supposed correlation is subject to further examination.

Therefore, the further *in silico* and *in vitro* research of these genes is expected to bring about new insights indeed into molecular machinery of cancer progression, thus laying ground for the intricated set of cancer biomarkers to be developed.

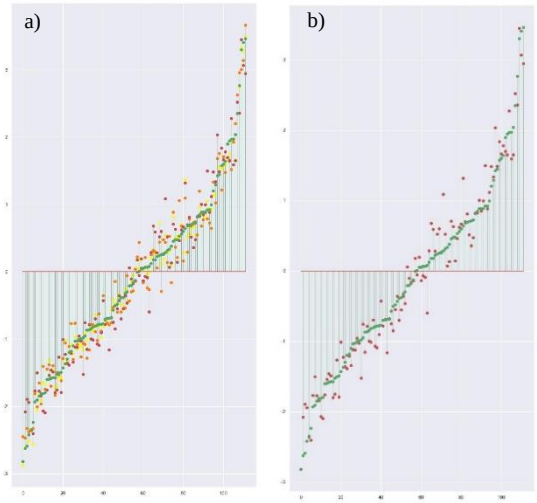


Figure 2: The expression of the classified genes a) in all groups; b) in N0 and N3 groups only. Horizontal and vertical axes represent genes and samples respectively.

3.6 Metastasis

Metastasis is defined as the spread of cancer cells from the initial tumour to the distant parts of the body via blood or lymph. It is a hallmark of cancer progression and is the most common reason for cancer-related deaths.

The present paper marks out 24 genes with divergent expression profiles in groups M0 (no metastasis) and M1 (metastasised cancer). Genes were categorised by z-score values in the metastasis-free group.

eIF4E is expressed almost seven-fold higher in M1 patients than in M0 within our database. There is evidence of *eIF4E* being employed to predict the rapamycin sensitivity in breast cancer cell lines. Furthermore, the cells might enhance resistance by altering the *eIF4E* activity [18].

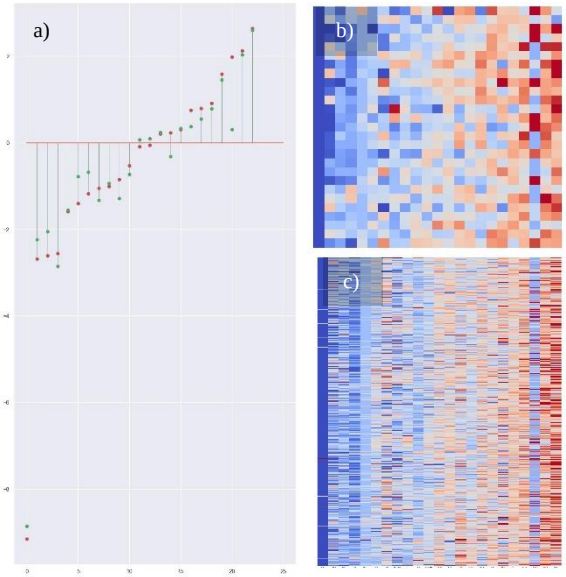


Figure 3: mRNA expression levels after *Z*-score normalization in BRCA tumour tissues (a) with M0 average values shown in green and M1 depicted in red. Heatmaps represents the expression values in all patients in (b) M0 and (c) M1 groups. Horizontal and vertical axes represent genes and samples respectively.

The *ZNF516* expression demonstrates dramatic increase in cancer tissues comparing to the healthy ones. The gene strongly influences the cell growth, proliferation and motility by repressing the transcription of the epidermal growth factor receptor (*EGFR*) [19]. *ZNF516* belongs to the protein transduction cascade which facilitates key processes in cancer development (e.g. angiogenesis, metastasis and immunosuppression). However, the detailed mechanism of such actions in breast cancer are still elusive [20].

The drastically low z-score (-8,8542 and -9,1529 for M0 and M1, respectively) in *SNORA13* gene could be precipitated by the errors to have taken place in performing the clinical analyses or curation of the results.

4 CONCLUSIONS

Machine learning algorithms like Random Forest are brilliant tools for processing vast amounts of data. The set of cancer-related genes was distilled to simplify the establishment of prognostic biomarkers.

The TNM cancer staging system includes 3 principal progression parameters – tumour diameter, lymph nodes invasion and metastasis, which were used for the classification. 230 genes were distinguished as prognostic factors, among which 94 impact tumour growth, 112 affect lymph nodes and 24 influence metastasising.

The present findings reveal the facilitating effects of such genes as *WASF2, CUEDC2, CPSF4, Pygo2, ZNF516* on the extension of cancerous tissues. These genes also engage in such pathways as p53, Wnt, PI3K/Akt, VEGF, UPP and TGFβ1, thereby influencing cell proliferation, tumouri- and angiogenesis.

The theoretical value of the present paper is expressed in employing the IT machinery in the analyses of yet unsystematised biomedical data; and in laying out the genetic canvas which is to be extended in our further works. The applicatory novelties of the research comprise setting up a basis for advanced diagnostic model to be utilised by breast cancer practitioners, as well as molecular biologists and scientists alike.

For the purposes of this study the initial version of the algorithm was employed, thus it has its limitations to be resolved in the offing. The prospective avenues of improvement include, *inter alia*, increasing credibility of the algorithm and implementing more sophisticated multifunctional analyses. With some fine-tuning, this model will be useful in guiding choices on how to treat patients with breast cancer.

ACKNOWLEDGMENTS

This work was partly supported by EMBO IG 4728-2020, Jacek Arct and 'New Technologies for Women' scholarships for NK and Alexander von Humboldt Stiftung (Philipp Schwartz-Initiative) for HF.

REFERENCES

[1] S.C. Shah, V. Kayamba, R.M. Jr. Peek, and D. Heimburger, "Cancer Control in Low- and Middle-Income Countries: Is It Time to Consider Screening?" vol. 5, pp. 1-8, 2019, doi: 10.1200/JGO.18.00200.

[2] J. Li, X. Guan, and et al., "Non-Invasive Biomarkers for Early Detection of Breast Cancer," Cancers (Basel), vol. 12(10), pp. 2767, 2020, doi: 10.3390/cancers12102767.

[3] B.K. Banin Hirata, J.M. Oda, R. Losi Guembarovski, C.B. Ariza, C.E. de Oliveira, and M.A. Watanabe, "Molecular markers for breast cancer: prediction on tumor behavior," Dis. Markers, vol. 2014, pp. 513158, 2014, doi: 10.1155/2014/513158.

[4] A.N. Richter and T.M. Khoshgoftaar, "A review of statistical and machine learning methods for modeling cancer risk using structured clinical data," Artif. Intell. Med., vol. 90, pp. 1-14, 2018, doi: 10.1016/j.artmed. 2018.06.002.

[5] A. Zaremba, P. Zaremba, S. Siry, Y. Shermolovich, and S. Zagorodnya, "In vitro and in silico study of anti-influenza activity of 2-dioxopyrimidin-5-trifluoromethyl-tetrahydrothiophene with subsequent increase in its affinity for the target protein," Proceedings of the 7th International Electronic Conference on Medicinal Chemistry, 1-30 November 2021, MDPI: Basel, Switzerland, doi:10.3390/ ECMC2021-11439.

[6] L. Peng, W. Chen, W. Zhou, F. Li, J. Yang, and J. Zhang, "An immune-inspired semi-supervised algorithm for breast cancer diagnosis," Comput. Methods Programs Biomed., vol. 134, pp. 259-265, 2016, doi: 10.1016/j.cmpb.2016.07.020.

[7] H. Falfushynska, O. Lushchak, and E. Siemens, "The Application of Multivariate Statistical Methods in Ecotoxicology and Environmental Biochemistry," Proceedings of International Conference on Applied Innovation in IT, vol. 10 (1), pp. 99-104, 2022.

[8] P. Rzymski, N. Kasianchuk, D. Sikora, and B. Poniedziałek, "COVID-19 Vaccinations and Rates of Infections, Hospitalizations, ICU Admissions, and Deaths in Europe during SARS-CoV-2 Omicron wave in the first quarter of 2022," Journal of Medical Virology, vol. 95(14), 2022, doi: 10.1002/jmv.28131.

[9] P.S. Rana, A. Alkrekshi, W. Wang, V. Markovic, and K. Sossey-Alaoui, "The Role of WAVE2 Signaling in Cancer", Biomedicines, vol. 9(9), pp. 1217, 2021, doi: 10.3390/biomedicines9091217.

[10] Y.H. Kuo, A.L. Shiau, and et al., "Expression of prothymosin α in lung cancer is associated with squamous cell carcinoma and smoking", Oncol. Lett., vol. 17(6), pp. 5740-5746, 2019, doi: 10.3892/ol.2019.10248.

[11] S. Roy, S. Saha, and et al., "Molecular crosstalk between CUEDC2 and ERα influences the clinical outcome by regulating mitosis in breast cancer", Cancer Gene Ther., vol. 29(11), pp. 1697-1706, 2022, doi: 10.1038/s41417-022-00494-x.

[12] C. Tao, W. Liu, and et al., "PAQR5 Expression Is Suppressed by TGFβ1 and Associated With a Poor Survival Outcome in Renal Clear Cell Carcinoma", Front. Oncol., vol. 11, pp. 827344, 2022, doi: 10.3389/fonc.2021.827344.

[13] X. Chen and H. Ishwaran, "Random forests for genomic data analysis," Genomics, vol. 99(6), pp. 323-9, 2012, doi: 10.1016/j.ygeno.2012.04.003.

[14] N.E. Reticker-Flynn, W. Zhang, and et al., "Lymph node colonization induces tumor-immune tolerance to promote distant metastasis", Cell, vol. 185(11), pp. 1924-1942.e23, 2022, doi: 10.1016/j.cell.2022. 04.019.

[15] Y. Song, K. Sun, and et al., "CPSF4 promotes tumor-initiating phenotype by enhancing VEGF/NRP2/TAZ signaling in lung cancer", Med. Oncol., vol. 40(1), pp. 62, 2022, doi: 10.1007/s12032-022-01919-1.

[16] Y. Zhou, Y. Che, Z. Fu, H. Zhang, and H. Wu, "Triple-Negative Breast Cancer Analysis Based on Metabolic Gene Classification and Immunotherapy", Front. Public Health, vol. 10, pp. 902378, 2022, doi: 10.3389/fpubh.2022.902378.

[17] V.A. Kraft, C.T. Bezjian, and et al., "GTP Cyclohydrolase 1/Tetrahydrobiopterin Counteract Ferroptosis through Lipid Remodeling", ACS Cent Sci. vol. 6(1), pp. 41-53, 2020, doi: 10.1021/acscentsci.9b01063.

[18] M. Saxena, R.K.R. Kalathur, and et al., "2-Histone Interaction Is Critical for Cancer Cell Dedifferentiation and Progression in Malignant Breast Cancer", Cancer Res. vol. 80(17), pp. 3631-3648, 2020, doi: 10.1158/0008-5472.CAN-19-2910.

[19] S. Satheesha, V.J. Cookson, and et al., "Response to mTOR inhibition: activity of eIF4E predicts sensitivity in cell lines and acquired changes in eIF4E regulation in breast cancer", Mol. Cancer, vol. 10, pp. 19, 2011, doi: 10.1186/1476-4598-10-19.

[20] L. Li, X. Liu, L. He, and et al., "ZNF516 suppresses EGFR by targeting the CtBP/LSD1/CoREST complex to chromatin", Nat. Commun., vol. 8(1), pp. 691, 2017, doi: 10.1038/s41467-017-00702-5.

Mathematical and Statistical Methods of Analyzing the Successful Implementation of German-Ukrainian Projects

Cornelia Scott[1] and Oksana Vasylenko[2]
[1]*Department of Economics, Anhalt University of Applied Sciences, Strenzfelder Allee 28, Bernburg (Saale), Germany*
[2]*Departments of Electrical Engineering, Mechanical Engineering and Engineering Management, Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*
cornelia.scott@hs-anhalt.de, oksana.vasylenko@hs-anhalt.de

Keywords: German-Ukrainian Projects, DAAD, Regression Analysis, Factor Analysis, Digin.Net, DigiJED, Idea-East Hub.

Abstract: The extraordinary dynamism, complexity and ever-increasing interdependence of all transformation processes in the modern world in the face of crisis conditions necessitates strengthening partnerships and coordination of actions at the national, regional and international levels to ensure the quality and sustainability of higher education systems around the world. The research focuses on the successful practical implementation of German-Ukrainian projects of the Anhalt University of Applied Sciences, such as Digin.Net, Digin.Net 2, Study Visits, DigiJED, DigiJED 2, GLSs and Idea-East Hub. Based on the mathematical and statistical methods of analysis, a regression model was built to predict the attraction of funding for the development of education in Ukraine. A meta-analysis of existing practical projects and their financial plans was conducted, considering the definition of specific criteria for successful selection among competitors. Using factor analysis of quantitative and qualitative assessment of the indicators of implemented projects, the authors' model of criteria for successful selection of projects in DAAD was created.

1 INTRODUCTION

Education and science are at the centre of a large-scale transformation of German-Ukrainian partnership that will change not only the academic world, but also the environment of cooperation, its conditions, and life in general [1], [2], [3]. During the war and in the post-pandemic period, international academic cooperation proved to be not only very effective and sustainable, but also a lifeline to save and preserve the entire nation, its potential, and intellectuals. The article is devoted to a critical analysis of the existing German financial support for Ukraine in the field of education development based on statistical data and correlation and regression analysis.

The German-Ukrainian cooperation is examined on the example of the practical implementation of German-Ukrainian projects of the Anhalt University of Applied Sciences, such as Digin.Net, Digin.Net 2 [4], Study Visits, DigiJED, DigiJED 2 [5], Idea-East Hub [6], etc.

The purpose of the study is to conduct a meta-analysis of project financial plans in order to identify specific criteria for successful selection among competitors and, based on the factor analysis of quantitative and qualitative assessment of the indicators of implemented projects, to create an authors' model of criteria for successful selection of projects in DAAD.

2 FINANCING GERMAN-UKRAINIAN EDUCATIONAL AND SCIENTIFIC PROJECTS

From 2013 to 2020, Germany committed 405,15 million euros to development finance to Ukraine for Education. Of this amount, 100% was provided as Official Development Assistance (ODA) grants, while 0 euro (0%) was provided in the form of ODA loans [7]. Development finance from Germany to Ukraine for Education was provided to different sub-sectors, as shown in the figure below. The largest commitments were 349,18 million euros to Higher education, 34,53 million euros to Education facilities and training and 12,26 million euros to Teacher training (Figure 1).

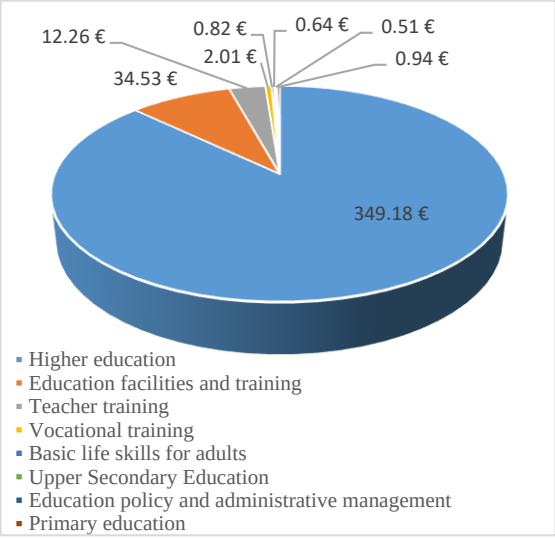


Figure 1: Total finance by sub-sector targeted 2013-2020, in million €.

Analyzing the ten sectors that received the most funds in 2013-2020, we can clearly see that most of the money is allocated to higher education, to support student learning.

Table 1: Top ten supported Sub-Sector in 2013-2020.

	Amount, in mn €	Year	Donor	Recipient	Sub-Sector	Project Title
1.	47,5 mn	2020	Germany	Ukraine	Higher education	Student costs
2.	46,9 mn	2019	Germany	Ukraine	Higher education	Student costs
3.	45 mn	2018	Germany	Ukraine	Higher education	Student costs
4.	43 mn	2017	Germany	Ukraine	Higher education	Student costs
5.	41,4 mn	2016	Germany	Ukraine	Higher education	Student costs
6.	38,3 mn	2015	Germany	Ukraine	Higher education	Student costs
7.	34,6 mn	2014	Germany	Ukraine	Higher education	Student costs
8.	34,2 mn	2013	Germany	Ukraine	Higher education	Student costs
9.	6,02 mn	2017	Germany	Ukraine	Teacher training	Professional integration of internally displaced persons
10.	3,57 mn	2019	Germany	Ukraine	Education facilities and training	Operating expenses, Goethe-Institute (GI)

2.1 Main Financial Sources

The main sources of funding for most educational projects within the framework of cooperation between Germany and Ukraine are DAAD, Erasmus, and BMBF. DAAD is the world's largest funding organization for the promotion of international student and research exchange, with an operating budget of around 634,7 million euros in 2021. Its most important funding providers include the Federal Foreign Office – AA (33,5%), the Federal Ministry of Education and Research – BMBF (32,5%), the Federal Ministry for Economic Cooperation and Development – BMZ (7,5%) and the European Commission and other international organizations European Union – EU (20,7%) [8].

In 2021-2027, the Erasmus+ budget of 21,208 billion euros provides an opportunity for Europeans to study, train and gain experience abroad.

Accordingly, DAAD, Erasmus and BMBF are the biggest investors in education. Funding of educational projects by foreign institutions is very important for Ukraine, especially considering the significant potential in the field of education.

2.2 Forecast Based on Regression Analysis

In 2020, Germany allocated 59,69 million euros to development funding to Ukraine for education, which was 100 % provided in the form of ODA grants. The disbursement rate was 105,2 %.

Consider the schedule of Germany's allocation of funds to Ukraine for the period 2013-2022. Moreover, the data for 2013-2020 were used from the Aid Atlas resource [7], and the calculation of the statistics of funding amounts was made on the basis of a forecast based on preliminary data and on the basis of statistical analysis using regression and trend line analysis (as of January 10, 2023, Figure 2).

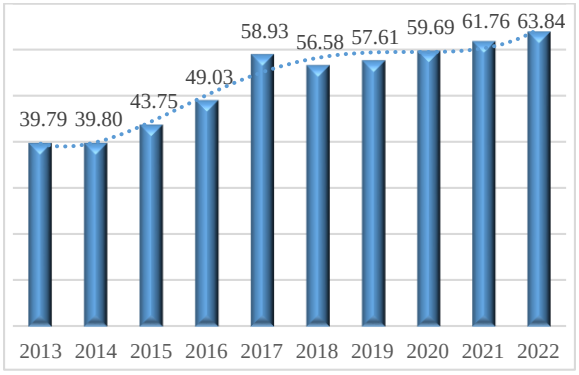


Figure 2: German funds to Ukraine for the development of education in 2013-2022, in million €.

Using data regression analysis and trend line analysis, it was found that the most effective trend line is a polynomial trend line. This curve is characterized by increasing and decreasing variables.

To get the approximation reliability value of R-squared (1):

$$R^2 = 0{,}9669, \quad (1)$$

we had to increase the equation to the 4^{th} degree and use the corresponding one (Figure 2):

$$y = 0{,}0417x^4 - 0{,}9356x^3 + 6{,}7781x^2 - 14{,}069x + 50{,}183 \quad (2)$$

R^2 is close to 1 (0,9669), indicating minimal approximation error. Consequently, this (2) can be used to predict data in future years.

3 JOINT GERMAN-UKRAINIAN RESEARCH: IMPLEMENTED PROJECTS

The Anhalt University of Applied Sciences (HSA) is an active player in the dynamic process of internationalization. It has about 2 500 international students from more than 110 countries and 12 English-language study programs. International students, who now make up around 30 per cent of the total, have been the focus of the activities of the current internationalization strategy. The University is also very strong in research and has 87 300,00 euros of external funds per faculty member, compared to the federal average of 45 400,00 euros in engineering and computer science at HAWs University in the 10 strongest researches in the federal territory. In its research strategy (Senate Resolution of January 2021), the HSA set out to advance research, knowledge transfer and technology not only regionally but also internationally to address important future challenges, such as the energy transition, climate change and digitalization, through innovative applied research. According to the internationalization strategy of the University, it is actively involved in many projects focused on Eastern and South-Eastern Europe, including Ukraine, where the HSA is strategically positioned as an important partner. HSA is constantly expanding existing contacts in Ukraine and bringing them to a qualitatively higher level in order to create more international networking and interdisciplinary study programs, especially with double degrees, as well as to create an international innovative network of doctoral students and faculty based on this international network and innovative Universities. Innovative competitions and forums, innovative training and relevant professional development activities contribute to the sustainable involvement of all target groups of the University in the process of internationalization and increase the competitiveness of the University.

3.1 HSA-Ukraine Cooperation

Anhalt University of Applied Sciences maintains intensive contacts with Ukrainian Universities and regularly conducts student and academic exchanges. This is done in particular through the ERASMUS+ (KA 107), Eastern Partnership and professional international internship programs, as well as through various research funding in the fields of renewable energy, data science and the Internet of Things (IoT). HSA actively cooperates and already has double degree agreements with some Universities in the country. The most powerful technical Universities of the country are involved in cooperation:

- National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” (KPI, Kyiv);
- Odesa Polytechnic National University (OPNU, Odesa);
- Oles Honchar Dnipro National University (DNU, Dnipro);
- State University of Intelligent Technologies and Communications (SUITT, Odesa);
- Tavria State Agrotechnological University (Melitopol);
- Ternopil Volodymyr Hnatiuk National Pedagogical University (TNPU, Ternopil);
- International Humanitarian University;
- Kharkiv National University of Radio Electronics (NURE, Kharkiv);
- Landwirtschaftliche Universität Kiew (Kyiv).

The English-language master's double degree programs: DDP Communication and Embedded Systems, DDP Data Science and a German-language bachelor's degree program in German – DDP Electrical and Information Engineering were created. More than 60 students have graduated from these programs and are working in German and Ukrainian companies, more than 40 professors and associate professors have completed internships at HSA, and more than 380 have completed international online internships in 2022.

In cooperation with Ukrainian Universities, in particular within the framework of the DAAD programs, HSA has already accumulated initial and positive experience. The Center of Excellence for Artificial Intelligence and Cloud Infrastructure Analysis - AI-aided Data Analysis and Data Transport Infrastructures (CoE AIDA-TI) was established in Kyiv under the leadership of HSA. The IDEA-East Hub, Digin.Net, Digin.Net 2, DigiJED,

DigiJED 2, Study Visits, GLS KDI 40, FIT4Ukraine, GLS Computer Systems and Networks/ Internet of Things (IoT), INTEGRA projects expand and consolidate promising cooperation in the academic sphere, they support a highly qualified didactic center for advanced training of all participants in the educational process, which has been focusing on the didactic features of digital learning for many years.

3.2 Funding of German-Ukrainian Projects of the HSA

A critical analysis of the progress in increasing the funding of German-Ukrainian cooperation shows how the interest in creating a clear and strong network between German and Ukrainian educational institutions is rapidly growing, which leads to the conviction that all possible resources and opportunities should be used to intensify this cooperation. This is evidenced by the data on the financing of projects from 2015 to 2025 (approved projects), with the grantee being the HSA (Table 2).

Let's consider Figure 3, which demonstrates the history of German-Ukrainian cooperation (as of March 02, 2023) and the prospects for the development and support of the education and science sector in terms of close cooperation and funding.
Thus, the total amount of funding for projects from 2015 to 2022 (excluding projects managed by the international office HSA) is 755 784,5 euros. Funding for 2023-2025 is planned and agreed in the amount of 797 245,00 euros.

3.3 DAAD Funding: Case Study

All DAAD programs have a standard format for preparing a financial plan for all countries. The financial plan consists of the following three major sections:

1) Personnel expenses;
2) Material expenses;
3) Funded individuals.

Table 2: Funding of German-Ukrainian HSA projects leadership and participation of prof. E. Siemens.

Years **Projects**	2015	2017	2019	2020	2021	2022	2023	2024	2025	**Total**
Study Visits	9 600,0	9 600,0	9 600,0			9 600,0				**38 400,0**
GLS Ukraine							15 000,0	45 000,0		**60 000,0**
AIDA TI						28 384,5	10 665,0			**39 049,5**
Pers Init			10 200,0			26 700,0	64 080,0			**100 980,0**
DigInNet 2			28 014,9	90 600,0	91 985,1	117 000,0	80 000,0			**407 600,0**
Idea East Hub						62 500,0	62 500,0	62 500,0	62 500,0	**250 000,0**
Sachsen Anhalt						30 000,0	245 000,0			**275 000,0**
DigiJED 2						232 000,0	150 000,0			**382 000,0**
Total, in €	9 600,0	9 600,0	47 814,9	90 600,0	91 985,1	506 184,5	627 245,0	107 500,0	62 500,0	**1 553 029,5**

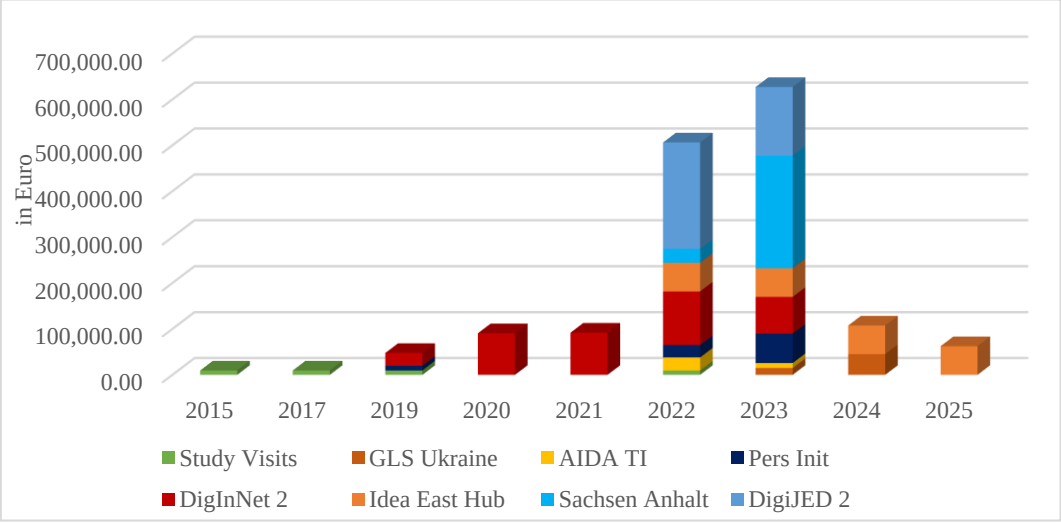


Figure 3: Funding of projects 2015-2025, in €.

Each of these sections is divided into subsections depending on the project program. For example, the first section is related to the financing of the personnel for project implementation and supervision and may contain the following items:

- Academic staff;
- Research assistants;
- Student assistants;
- Other personnel.

The second section describes, for example, the material expense required to implement the project:

- Fees,
- Mobility of project staff;
- Stay of project staff;
- Material resources internal/external to be procured within the project.

The section on supported persons contains information on their financing and justification in a particular project according to subsections:

- Funded individuals/ Supported persons;
- Mobility of funded individuals;
- Accommodation of funded individuals.

In addition, each subsector has a list of eligible funding activities and rules for their use [9]. The factors should be in the focus:

- Targeted - the goal;
- Timely - planning of all activities;
- Geographical - countries, regions;
- Legal - to be guided by the laws of the countries participating in the project, visas, personal data protection, health insurance.
- Duration - dates;
- Relevance - transportation costs, accommodation costs, additional costs;
- Cost-efficiency - purchase of goods.

The financial plans of the projects that received funding were analyzed and it was determined that project funding is closely correlated with the criterion of Relation of the Program Goals with the Goals of the Project. The Spearman rank correlation coefficient has a value of ~ +0,87, which indicates a significant direct linear relationship between the described criterion and the financing of a particular project. This correlation can be considered significant at the significance level α=0,01.

Figure 4 represents the aggregated results of the financial plans of HSA projects supported by DAAD in 2015-2025. A critical analysis shows that all DAAD projects are aimed at supporting and financing Human Resources such as salaries and mobility in education and research – 87,89 % in the categories of Personnel Expenses and Funded Individuals. Limited funding is allocated for the technical infrastructure of the project – 12,11 %.

These results confirm that DAAD programs are focused on supporting the educational and scientific activities of Students, Doctoral candidates, Postdoctoral students, Young scientists, Experienced scientists and Professors in accordance with the activities planned in the projects.

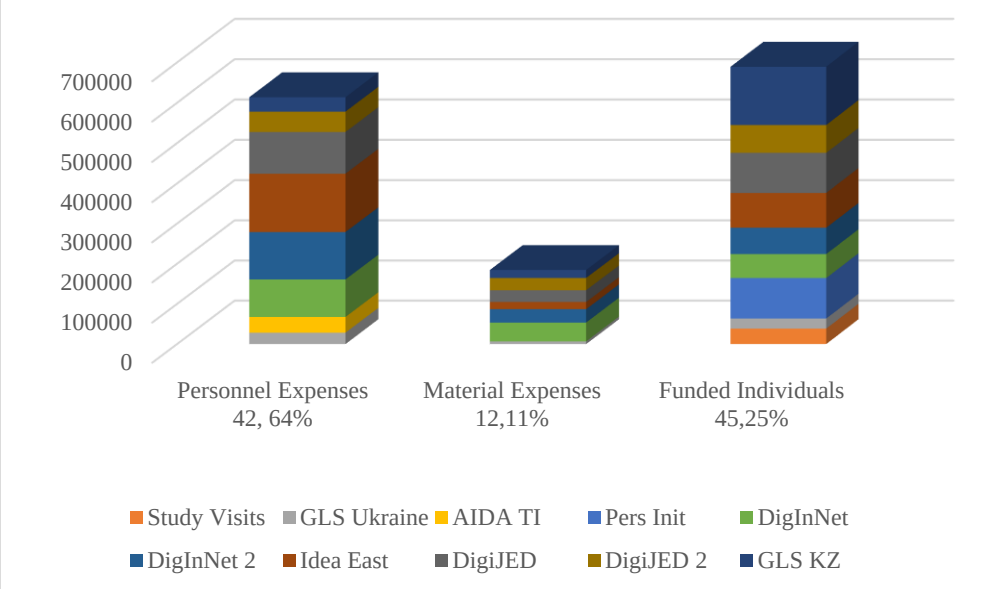


Figure 4: Main sections of HSA-DAAD projects funding 2015-2025, in %.

4 MODEL OF CRITERIA FOR SUCCESSFUL PROJECT SELECTION

The decision on funding is made by DAAD based on the evaluation of applications by the selection committee. Experience has shown that only some of the applications can be approved that meet the program's goals mentioned. These are particularly well-organized activities, scientific, research and teaching work, visits and seminars which guarantee a close cooperation with Universities, scientists and students in Germany.

In order to research the positive dynamics of supporting new projects and obtaining funding in the future from DAAD, a detailed critical analysis of the selection criteria for each program, their weighting coefficient was carried out and a unique author's model of the criteria for successful selection of projects in the selected program was created.

Consider those programs in which HSA has participated and, thanks to a correct prospective assessment of the selection factors, received a positive decision on the application description and, as a consequence, subsequent funding.

4.1 Selection of Criteria

For program Study Visits and Study Seminars for Groups of Foreign Students to Germany 2021/2022 the main criteria for successful selection are summarized in Table 3. As part of this program, the application from the HSA was approved by a selection committee, and the "*Studienpraktikum Communication and Embedded System, KPI, OPSU*" project was funded and successfully implemented.

The main criteria for the successful selection of the program Support for the internationalization of Ukrainian higher education institutions – shaping the digital future together: German-Ukrainian higher education institution collaborations (2019–2021) are summarized in Table 3. The application was approved by a selection committee under this program and the „*DigIn.Net: German-Ukrainian Digital Innovation Network*" project was funded and successfully implemented in 2019-2021. After the successful implementation of this project, DAAD proposed to prepare an application to participate in the program Support for the internationalization of Ukrainian higher education institutions. The application was approved by the selection committee for this program, and the project "*DigIn.Net 2: German-Ukrainian Digital Innovation Network - 2*" (2021-2023) is currently funded and actively implemented.

The criteria for successful selection for the program Internationalization of Universities of Applied Sciences (HAW/FH) "HAW.International" (Module B 2022 - 2025), are given in Table 3. The application was approved by the selection committee of this program, and the project „*IDEA-East-Hub: International Innovation Hub for Data Science and renewable Energy - East*" is now funded and successfully implemented during the period 2022-2025.

Table 3: Main criteria for the successful selection of projects, DAAD.

	Study Visits	DigIn.Net	DigIn.Net 2	Idea East Hub	DigiJED, 2	GLS
Relation of the Project to the Program Goals, the Measures and Objectives	*X*			*X*	*X*	*X*
Plausibility of Project Description and Financing Plan		*X*	*X*	*X*	*X*	*X*
Quality of the Project Program	*X*			*X*		*X*
Contentual Preparation	*X*		*X*	*X*		*X*
Technical Requirements (Digital Existing Infrastructure)		*X*	*X*	*X*		*X*
Experience with Digital Teaching and Administration		*X*	*X*	*X*	*X*	*X*
Involvement of Young Scientists		*X*	*X*		*X*	*X*
Correlation of Subject - Cultural	*X*					
Encounter and Exchange of Experience	*X*			*X*		*X*
Sustainability of the Project		*X*	*X*	*X*	*X*	
Concept for Further Development			*X*	*X*	*X*	*X*
Plausibility of Ukraine Expertise		*X*	*X*		*X*	
Language Skills	*X*					*X*
University Partnerships	*X*			*X*	*X*	*X*

In the program Ukraine digital: Ensuring academic success in times of crisis (2022, 2023) the main criteria for successful selection are given. The application was approved by the selection committee for this program, and the projects "*DigiJED and DigiJED 2: Digital Education with Joined Efforts*" are funded and actively implemented in 2022-2023.
The program "German-language study programs (GLS) in East-Central, Southeast and Eastern Europe as well as the Caucasus and Central Asia - 2023 - 2024" describes the main criteria for a successful

selection. Two applications from Anhalt University of Applied Sciences (one for Ukraine, the other for Kazakhstan) were approved by the selection committee and are funded and implemented from 2023.

4.2 Methodology

Analyzing the above criteria of the listed programs and considering the selection criteria of other relevant DAAD programs, the factors influencing the successful decision of the competition committee were analyzed, and on their basis the weight of each criterion was determined using a multi-criteria evaluation approach. Factor analysis was conducted to reduce the sample size of the criteria, as well as to group them based on correlations. The principal components method was used as a selection method, and Varimax with Kaiser normalization was used as a rotation method. To optimize the calculations, the statistical software package SPSS was used [10], the results of which are presented in the following Tables 4, 5.

4.2.1 Implementation

The main criteria with corresponding weighting coefficients are:

- Relation of the Project to the Program Goals, the Measures and Objectives (X1, weight - 8%);
- Plausibility of Project Description and Financing Plan (X2, weight - 10%);
- Quality of the Project Program (X3, w. - 6%),
- Contentual Preparation (X4, w. - 8%);
- Technical Requirements (Digital Existing Infrastructure) (X5, w. - 8%);
- Experience in Digital Teaching and Administration (X6, w. - 10%);
- Involvement of Young Scientists (X7, w. - 8%);
- Correlation of Subject - Cultural (X8, w. - 2%),
- Encounter and Exchange of Experience (X9, w. - 6%);
- Sustainability of the Project (X10, w. - 8%),
- Concept for Further Development (X11, w. - 8%);
- Language Skills (X12, w. - 4%);
- Plausibility of Ukraine Expertise (X13, w. - 6%);
- University Partnerships (*X*14, w. - 8%).

The factor analysis resulted in the data presented in Table 4 with the values of the total explained variance and Table 5 - the matrix of returned components or factor matrix were obtained. Table 4 shows that the value of three Eigen factors is greater than one. Therefore, only three factors are selected for analysis. The first factor explains 34,308 % of the total variance, the second – 12,97 % and the third – 9,433. Only 14 criteria are included in the analysis, as the rest of them in total give a very small variance, and therefore have little influence on decision-making (this was also considered when finding weighting coefficients).

Table 4: Total explained variance.

	Initial Eigenvalues			Sums of Squares of Loads		
	Total	% of variance	Cumulative % of variance	Total	% of variance	Cumula-tive % of variance
1	5,146	34,308	34,308	3,466	23,105	23,105
2	1,945	12,970	47,278	2,536	16,907	40,013
3	1,415	9,433	56,711	2,505	16,698	56,711
4	0,990	6,601	63,312			
5	0,936	6,238	69,550			
6	0,760	5,068	74,617			
7	0,693	4,622	79,240			
8	0,612	4,083	83,323			
9	0,529	3,529	86,852			
10	0,473	3,151	90,004			
11	0,433	2,889	92,893			
12	0,339	2,262	95,1555			
13	0,301	2,007	97,161			
14	0,245	1,635	98,797			
15	0,181	1,203	100,000			
Method of selection: Principal component analysis						

Further, in each row of the resulting factor, the factor loadings that have the highest absolute value are noted (Table 5). These factor loadings should be understood as correlation coefficients between criteria and factors. Thus, the criterion *X*1 is most strongly correlated with factor 1, the correlation value is 0,866; the criterion *X*2 is also most strongly correlated with factor 1 with a value of 0,641; the criterion *X*5 is most closely related to factor 2 (0,783), etc. In most cases, the inclusion of a variable in one factor, based on the correlation coefficients, is unambiguous.

Table 5: Matrix of returned components.

	Components		
	1	2	3
X1	0,866	0,628	−0,191
X2	0,641	0,157	0,215
X3	0,727	−0,153	0,311
X4	0,533	−0,106	0,394
X5	−0,362	0,783	4,52E-02
X6	−1.2E-02	0,763	−3,8E-02
X7	3.58E-02	0,525	0,543
X8	−0,117	0,719	−0,267
X9	2.56E-02	0,551	−8,8E-02
X10	0,652	−9.5E-02	0,285
X11	0,485	−0,110	0,665
X12	0,199	0,802	0,108
X13	0,125	-0,292	0,392
X14	-0,251	0,592	0,392
Method of selection: Principal component analysis Rotation method: Varimax with Kaiser normalization a. The rotation is performed in 8 iterations			

4.2.2 Results Interpretation

Therefore, based on the above, the criteria can be attributed in the following order to three factors:

Factor 1. Qualitative component of the application, 40%:

- Plausibility of Project Description and Financing Plan (*X*2, weight - 10%);
- Relation of the Project to the Program Goals, the Measures and Objectives (*X*1, weight - 8%);
- Contentual Preparation (*X*4, weight: 8%);
- Sustainability of the Project (*X*10, weight: 8%);
- Quality of the Project Program (*X*3, weight: 6%).

Factor 2. Quantitative component of the application, 38%:

- Experience in Digital Teaching and Administration (*X*6, weight: 10%);
- Hardware and Software (Digital Existing Infrastructure) (*X*5, weight: 8%);
- University Partnerships (*X*14, weight: 8%),
- Encounter and Exchange of Experience (*X*9, weight: 6%);
- Language Skills (*X*12, weight: 4%);
- Correlation of Subject - Cultural (*X*8, weight: 2%).

Factor 3. Prospects and forecasting further development, 22%:

- Involvement of Young Scientists (*X*7, weight: 8%);
- Concept for Further Development (*X*11, weight: 8%);
- Plausibility of Ukraine Expertise (*X*13, weight: 6%).

Thus, it is possible to outline and demonstrate the author's model of the main criteria for successful project selection (Figure 5).

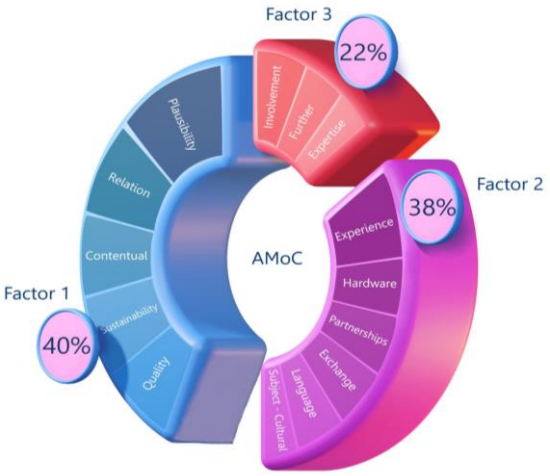


Figure 5: Model of the criteria for successful project selection.

When preparing a new application for funding in any DAAD program, it is worth paying attention to the conservative provision on „Plausibility of Project Description and Financing Plan" (*X*2) and one of the innovatively significant provisions that has emerged as a result of reform changes in modern learning formats - "Experience in Digital Teaching and Administration" (*X*6), which have the greatest weight.

3 CONCLUSIONS

The analysis of the growing financial flows to the higher education sector in Ukraine shows a rapid trend towards the importance and necessity of organizing effective measures to strengthen and expand German-Ukrainian cooperation in the field of digital internationalization.

The successful implementation of the HSA projects contributes to the prospective development of international cooperation, which is constantly evolving and expanding through new partners,

connections, scientific schools, and the creation of new joint double degree programs.

Financial grants from DAAD programs are primarily aimed at supporting human resources in projects (over 87%). Using factor analysis, the author's model of the criteria for successful selection of DAAD projects was developed, tested and presented, which allowed to identify 3 main factors: *Factor 1. Qualitative component of the application*, *Factor 2. Quantitative component of the application*, *Factor 3. Prospects and forecasting further development* and the 2 main criteria: *"Plausibility of Project Description and Financing Plan" and "Experience in Digital Teaching and Administration"* that should be paid the most attention to when preparing documentation for submission to DAAD programs.

ACKNOWLEDGMENTS

This work has been supported by the DAAD projects (DigIn.Net 2 - 57602278, Study Visits - 57618778, DigiJED 2 - 57678267, GLSs – 57663389-57663461, Idea-East Hub - 57603761).

REFERENCES

[1] United Nations, "The Sustainable Development Goals Report," New York, 2022, 68 p.

[2] P. Pistor (Ed.), "The Internationalisation of Higher Education – Perspectives from the THEA Ukraine Project and Beyond," Münster, 2021, 168 p.

[3] S. Chang, and C. Gomes, "Why the Digitalization of International Education Matters," Journal of Studies in International Education 26 (2), 2022, pp. 119-127.

[4] DigIn.Net 2, "German-Ukrainian Digital Innovation Network Project," 2022, [Online]. Available: https://digin-net.de/en/.

[5] DigiJED, "Digital Education with Joined Efforts," 2022, [Online]. Available: https://www.hs-anhalt.de/international/internationale-projekte/digijed-digital-education-with-joined-efforts.html.

[6] IDEA East Hub, "International Innovation Hub for Data Science and renewable Energy - East," 2022, [Online]. Available: https://www.hs-anhalt.de/en/international/intercultural-encounters/idea-east-hub.html.

[7] A. Atteridge, G. Savvidou, S. Sadowski, F. Gortana, L. Meintrup and A. Dzebo, „Aid Atlas. Germany to Ukraine for Education during 2013–2020," SEI, 2019, [Online]. Available: https://aid-atlas.org/profile/germany/ukraine/education/2013-2020?usdType=usd_commitment.

[8] C. Papsch, „DAAD: Budget and funding bodies," 2022. [Online]. Available: https://www.daad.de/en/the-daad/what-we-do/facts-figures/budget-and-funding-bodies/.

[9] DAAD, „Ukraine: Kurze Einführung in das Hochschulsystem und die DAAD-Aktivitäten 2022", Bonn: Impressum, 2022, pp. 23-24.

[10] A. Bühl, P. Zöfel, „SPSS 12: Einführung in die moderne Datenanalyse unter Windows," Pearson Studium, 2005, 744 p.

Multi-Criteria Optimisation for Building it Management Model within a Company

Larysa Globa[1], Stanislav Dovgyi[2], Oleh Kopiika[2] and Oleksii Kozlov[2]
[1]*Igor Sikorsky Kyiv Polytechnic Institute, Peremohy Avenue 37, Kyiv, Ukraine*
[2]*Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine, Chokolivsky Boulevard 13, Kyiv, Ukraine*
lgloba@its.kpi.ua, sdovgii@gmail.com, okopiyka@gmail.com, alexey.ua84@gmail.com

Keywords: Service-Oriented Information Technology, IT Services, Multi-Criteria Optimisation.

Abstract: The paper proposes approaches to multi-criteria optimisation when building an IT management model within a company using modern service-oriented IT solutions. It is proposed to employ a methodology that helps companies achieve appropriate levels of efficiency from IT implementation by maintaining a balance between the realisation of benefits, optimisation of risks and use of resources. The methodology enables administering and managing IT, both in the field of its functional responsibility and business tasks addressed by the company, while considering the IT needs of internal and external stakeholders. At the same time, a specific objective shall be defined, clarified, and specified in line with the balanced scorecard (financial, customer, internal processes, learning and growth). In addition, the task of effective IT implementation related to developmental priorities is fulfilled. In this case, the decision-making problem is characterised by a multi-criteria pattern. Specifically, the article proposes implementing decision-making principles using algorithms or procedures addressing multi-criteria optimisation problems. Almost all known approaches to solving these problems involve their scalarization, where consideration of the benefits is one of the main issues that must be addressed to select the best solution. Under this principle, the methods for solving multi-criteria optimisation problems are classified according to characteristics of information: decision-making under certainty; decision-making when information on the benefit system is unavailable; decision-making with the gradual acquisition of information on the benefit system; decision-making under multi-connectivity, heterogeneity, and multi-criteria information about the benefit system.

1 INTRODUCTION

Information is becoming an increasingly valuable business asset and, in many cases, the main one. The rapid development of information technology (IT) has removed physical barriers to producing and transmitting information and changed our understanding of how information is obtained (produced) and consumed. In the modern world of information technology, the share of services for storing and transmitting data through the "cloud" and software as a service (SaaS) models offered by software vendors, etc., is constantly growing.

The world's leading consulting companies believe that rapid improvement and spread of cloud computing is currently one of the key trends to significantly affect in the coming years the global development of the IT industry and business, finance, public administration, healthcare, education and many other sectors of human life [1-10]. In the context of advanced development of information and communications technology and yet another global downturn, the technology enabling organisations and other entities to eliminate high spending on their IT infrastructure and obtain all required IT resources online is viewed as a promising and cost-effective modernisation choice, and the best investment in the future.

All of the above determines the urgency of changing the company's IT management business model in line with developing a service-oriented information economy.

Mush research on this topic has already been done [11-16]. Notably, the evolution in IT service management has occurred more or less synchronously with fundamental societal changes and therefore was inevitable.

2 STAGES OF IT SERVICE MANAGEMENT GOALS EVOLUTION

In terms of IT development, three stages of IT service management goals and IT outputs evolution can be distinguished:

1) IT infrastructure management;
2) IT service management;
3) information business service management.

These stages are described in more detail in Table 1. To some extent, each stage may be associated with changes in the perceived role of IT within a company - the transition from a technology unit to a full-fledged partnership (Figure 1).

The enhanced IT role and importance should indisputably lead to evolution in understanding IT service management goals, and IT outputs [17].

Equally important is that evolution in the understanding of IT service management goals and IT outputs may be associated with a shift in social consciousness: transition from a product-oriented industrial economy to a service-oriented information economy (with certain reservations as the economic paradigm has been changing quite slowly and gradually).

Table 1. Three Stages of it management evolution.

Stage	Tasks	Priorities	Characteristics
Stage 1	IT infrastructure management	Stability and control over IT infrastructure	IT Department delivers IT systems. The goal of IT management is to mitigate the risk of IT system failure.
Stage 2	IT service management	Quality and efficacy of IT processes	IT Department delivers technological services. The goal of IT management is to maintain the agreed parameters of IT services.
Stage 3	Information business service management (Web-service)	Maximum IT efficiency	IT Department delivers information business services. The goal of IT management is to serve the business needs best.

Apart from general market factors, the evolution in the understanding of actual IT outputs has been driven by the need to lay the groundwork for solutions to complex practical issues in the IT industry.

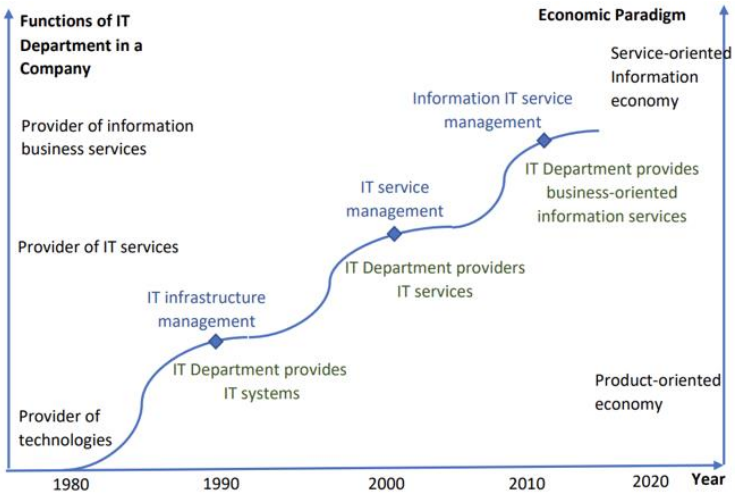


Figure 1: Stages in the evolution of IT service management goals and IT outputs.

IT executives constantly face numerous problems and lack systemic tools to address them. These problems include:

1) justification of continuous IT spending growth;
2) justification of the value generated through investing in IT; issues related to a mismatch between outputs and expectations;
3) management of IT requirements; issues relating to aligning business units' expectations and IT department capabilities;
4) persistent lack of IT resources, IT competencies, and problems associated with delegating complex tasks to external IT providers.

This paper aims to present approaches to multi-criteria optimisation when building a business model of IT management within a company tailored to the needs of business units in the modern service-oriented information economy setting.

3 METHODOLOGY FOR THE GOVERNANCE AND MANAGEMENT OF COMPANY IT

It is apparent that users need not an IT system as such but its operational results [18-24]. Achieving these results generally requires more assets than just an IT system. It is impossible to meet complex needs focusing on the IT system only. To move forward, all stakeholders should arrive at an understanding of common tasks and objectives.

It can therefore be concluded that information is an essential resource. At all stages of its lifecycle, information is critically dependent on specialised technologies. Information and rapidly developing information technology are vital for the development of society.

Today, more than ever, business leaders should strive to:

- maintain high quality of information to support managerial decisions;
- generate value for society from IT-enabled investments, i.e., achieve strategic goals and realise business benefits through effective and innovative use of IT;
- achieve operational excellence through the reliable and efficient use of technology;
- maintain IT-related risks at an acceptable level;
- optimise spending on IT services and technology;
- improve compliance with laws, regulations, contractual terms and policies related to IT use.

Thus, the modern characteristics of IT implementation are, namely:

1) control over information management services;
2) increased IT value;
3) IT department as a provider of information management services;
4) the goal of IT management – the most accurate compliance with the needs of society.

Therefore, to improve the efficiency of IT implementation, we propose a methodology designed to assist in addressing IT governance and management tasks.

This methodology helps in achieving the appropriate levels of efficiency from IT implementation and maintaining a balance between the realisation of benefits and the optimisation of risks and resources. The methodology enables administering and managing IT, both in the field of its functional responsibility and business tasks addressed, while considering the IT needs of internal and external stakeholders. The methodology is universal and of benefit to executives who make decisions.

The methodology is based on the five key principles of company IT governance and management, as shown in Figure 2.

Principle 1: Meeting Stakeholder Needs. The main task of any company is to create value for stakeholders by maintaining a balance between the realisation of benefits and the optimisation of risk and resources. The methodology describes all required processes and other factors supporting business value generation with IT. Since business tasks may vary, the methodology can be customised to suit the context of a particular company. This can be achieved by cascading high-level tasks to manageable and specific IT-related goals and associated processes and practices (Figure 3).

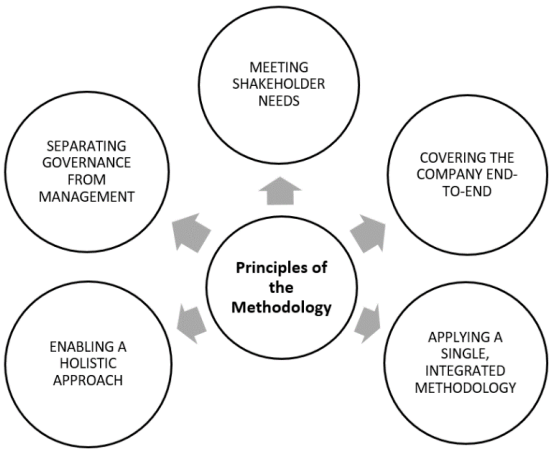


Figure 2: Principles of the methodology.

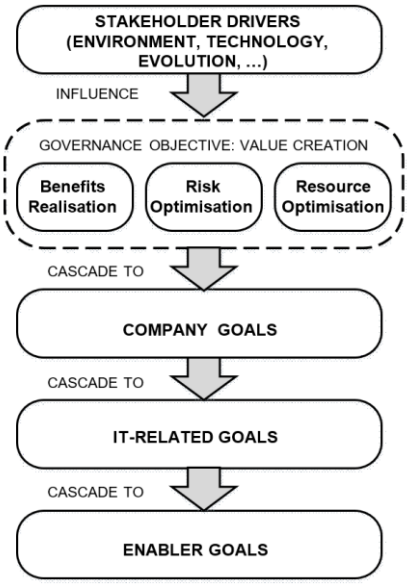


Figure 3: Goals cascade.

Principle 2: Covering the Company End-to-End. The methodology integrates IT management into company management; it covers all functions and processes within a company.

The methodology focuses not only on "IT-function" performance but also on information and related assets of the company that shall be managed like any other assets. It is based on the fact that IT-related governance and management enablers operate company-wide and across the entire value generation chain, including all internal and external aspects and roles relevant to IT governance and management.

Principle 3: Applying a Single, Integrated Methodology. There are many IT-related standards providing guidance on specific IT activities. The methodology ensures compliance with these standards. Thus, it provides an integrated approach to the governance and management of company IT.

Principle 4: Enabling a Holistic Approach. Efficient and effective governance and management of company IT require a structured approach considering numerous interacting components of the goal.

The methodology defines a set of enablers that support the implementation of IT governance and management system. Enablers can be broadly defined as anything that can help achieve specified objectives. The methodology includes seven categories of enablers (Figure 4): Principles, Policies and Approaches; Processes; Organisational Structure; Culture, Ethics and Behavior; Information; Services, Infrastructure and Applications; People, Skills and Competencies.

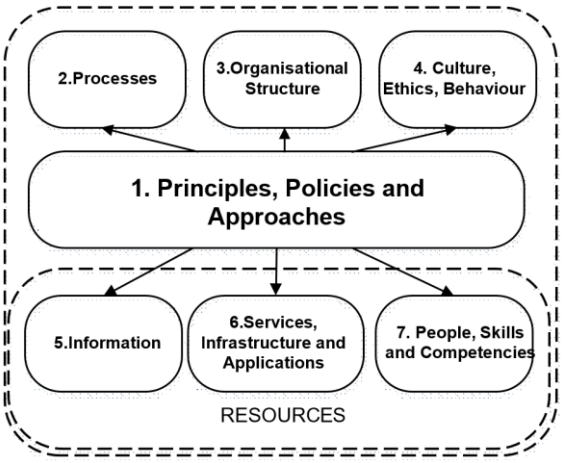


Figure 4: Enablers.

Principle 5: Separating Governance from Management. The methodology makes a clear distinction between governance and management. These two processes encompass different activities, require different organisational structures, and serve different purposes. According to the methodology, the difference between governance and management is as follows:

- in most cases, governance is the responsibility of the Board of Directors chaired by the Chairman;
- some functions may be delegated to special business units at an appropriate level;
- to a large extent, management is the responsibility of company directors who report to the director general (CEO).

Together, these principles help to build an effective governance and management methodology that optimises investments in information technology for the benefit of stakeholders.

The importance of this goal is that it enables setting priorities for the implementation and improvement and ensures IT governance based on understanding of company (strategic) objective and associated risks. In practice, this goal:

1) defines important and reasonable objectives and tasks at different levels of responsibility;
2) allows retrieval from the knowledge base of all data relating to the defined objectives that may be of use for implementation, improvement or quality assessment projects;
3) clearly identifies and demonstrates (in some cases in great detail) how the enablers help achieve company objectives.

The Table of concordance between the company and IT-related goals (Table 2) is given for illustrative purposes only and should be used with reservations.

Table 2. IT-related goals.

BSC Dimension	Information and Related Technology Goal
Financial	1) Alignment of IT and business strategy 2) Foreign laws and regulations IT compliance and support for business compliance 3) The leading role of executive management in making IT-related decisions 4) Managed IT-related business risks 5) Realised profit from IT-enabled investments and services portfolio 6) Transparency of IT costs, benefits and risks
Customer	1) Provision of IT services in response to business requirements 2) Adequate use of applications, information and technology solutions
Internal processes	1) IT flexibility 2) Security of information, processing infrastructure and applications 3) Optimisation of IT assets, resources and capabilities 4) Ensuring the operation and support of business processes by integrating applications and technology into business processes 5) Benefiting from programs and projects implemented within the established time frame and budget while meeting requirements and quality standards 6) Availability of reliable and relevant information for decision-making 7) Internal policies compliance
Learning and growth	1) Competent and motivated IT personnel 2) Knowledge, expertise and proactivity for business innovation

Therefore, it is necessary to define the company goal, clarifying and detailing it in accordance with the balanced scorecard (financial, customer, internal processes, learning and growth). Besides, the problem of effective IT implementation, associated with different priorities, is addressed.

4 MULTI-CRITERIA OPTIMISATION METHODS

In the case under consideration, the decision-making is based on multiple criteria. Therefore, special attention should be paid to identifying the causes for such a multi-criteria pattern and finding ways to use appropriate mathematical methods for the goals cascade.

Consequently, “IT implementation” aims to establish the values for a larger number of qualitative and quantitative parameters, i.e. partial goals. These goals may compete with each other.

One more argument for the multi-criteria decision-making pattern is the time-space factor. The results of activities and costs are distributed, so the “value” of IT implementation today, tomorrow and in a year, development prospects, etc., are to be compared. Known convolution techniques (for example, for time distribution, i.e. discounting) are characterised to a large extent by subjectivism (coming from an author of the method, not from the decision-maker expert (DM)).

Setting up multi-criteria decision-making problems essentially involves two aspects. On the one hand, these problems are very similar to the problems of making decisions under uncertainty because different options must be evaluated regarding their possible benefits and associated risks. On the other hand, multi-criteria decision-making under certainty and considering a large number of criteria is based on the rejection of the traditional assumption that one of the alternatives is always selected using one criterion only. In such situations, a scalar optimisation problem is replaced with a vector optimisation problem.

To take account of these two aspects, a mathematical model that reflects a multidimensional system of targeted functions shall be developed as a scalar criterion of optimality. This approach requires the decision-maker to have special preference functions. The validity of this assumption is not obvious in advance. However, this approach has a well-developed theory for solving such problems [25].

Specifically, the decision-making principles are implemented as algorithms or procedures for solving multi-criteria optimisation (MCO) problems. Almost all known approaches to solving these problems include their scalarization. At the same time, one of the main issues is to consider the DM benefits when choosing the best solution. According to this principle, the methods for solving multi-criteria optimisation problems can be classified based on the characteristics of information about DM benefits:

1) decision-making under certainty;
2) decision-making when information about the DM benefit system is unavailable;
3) decision-making with a gradual acquisition of DM benefit system information;

4) decision-making under multi-connectivity, heterogeneity, and multi-criteria benefit system information.

Classification of multi-criteria decision-making methods is shown in Figure 5.

A) The first group of tasks related to the context of certainty regarding the DM benefit system is discussed below.

1) A particularly large class of decision-making methods is based on the existence of vector estimates of the preference relationship, which depends on the system of objective functions only. In this case, it is possible to construct the corresponding value function. In addition, the value function determines the structure of DM benefits precisely. More detailed information about the theory of the value function is given in [26].
2) One of the decision-making methods for multi-criteria optimisation, which allows narrowing the Pareto set, implies arranging the criteria according to their importance. Questions about the criteria importance addressed to the DM are, on the one hand, quite clear to the DM. On the other hand, DM's advice can be effectively used in optimisation algorithms. However, many methods require the DM to precisely define the importance of the criteria (weight coefficients). An example of the qualitative ordering of criteria (without defining weight coefficient) is lexicographic ordering, and multi-criteria problems with criteria strictly ranked by importance are called lexicographic optimisation problems. The range of multi-criteria problems relating to lexicographic ordering is obviously too narrow. The lexicographic ordering principle can be used when the threshold (minimum allowable) values of the criteria are set in the criterion space. In such a case, the specified ordering of initial prioritized criteria will be defined by a sequence of these criteria maximization up to the corresponding threshold values. As a result, the multi-criteria problem converts into a lexicographic problem with a vector criterion. However, it should be noted that, in general, the lexicographic optimisation problems may be unstable because minor changes in their parameters (source data) can seriously affect the choice of optimal alternatives. Therefore, special methods are used to solve them.
3) Another algorithm for decision-making under certainty includes the methods of targeted programming. The targeted programming method for solving multi-criteria problems is based on ordering criteria (goals) according to their importance. The initial problem is solved by sequentially solving several problems with one targeted function in such a way that solving a problem with a less important goal cannot

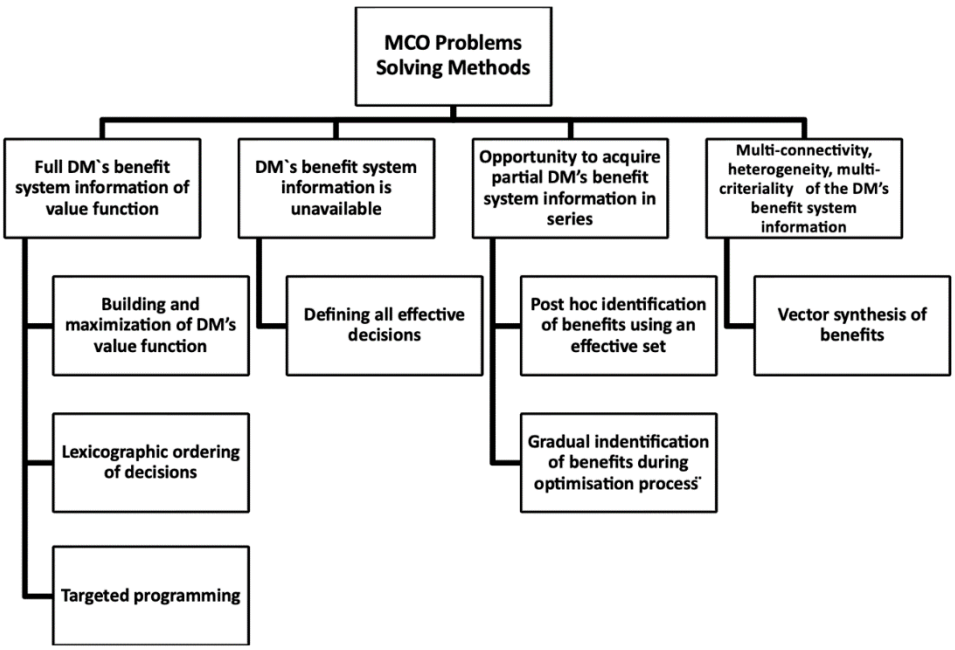


Figure 5: Classification of methods for solving multi-criteria problems.

affect the optimal value of a higher-priority objective function. The result is a satisfactory solution to the problem. The targeted programming is usually applied to linear models. Its main difference from linear programming is that many goals are formalized not as targeted functions but as restrictions in another, more general model. For this purpose, the estimated quantitative values of targeted functions are applied together with the so-called variable deviations characterizing the degree of goals' achievement for a particular decision.

B) The standard procedures for identifying effective solutions are used if no information about the DM benefit system is available. Then, the DM selects the best alternative by explicit evaluation.

C) In the case of gradual acquisition of DM benefit system information, there are two groups of multi-criteria methods for decision-making.

1) 1) Below is an example of a multi-criteria optimisation method, which belongs to the group specified on the classification scheme as a "post hoc identification of benefits using an effective set". In this case, the so-called ELECTRA methods are used, proposed by a well-known DMT (decision-making theory) [27].

   The ELECTRA methods aim to narrow the Pareto set of alternatives, which is accomplished as follows. Based on the DM responses, the weight (value characterising the importance of the specific criterion) is determined for each criterion (it is assumed that the criteria are numerical). All modifications of the ELECTRA methods include obtaining qualitative information from the DM about the relative importance of the criteria (expressions such as "criteria 3 and 4 are equally important and have greater importance than criterion "1") and its transformation into quantitative, numerical one. The problem here is that it can generally be done in various ways.
2) Interactive or human-computer procedures for solving MCO problems (HCPs). With HCPs, the ranges of acceptable values can be explored to find the best solution. The search for a solution is carried out in coordination with the DM and decision-making support system (DMSS). The source data for the HCPs are the decision options and the criteria used to evaluate the decision. HCPs include two phases: the analysis phase performed by the DM and the calculation phase performed by a computer.

The calculation phase. The DMSS:

1) performs calculations based on preliminary information or data obtained from the DM in the previous step;
2) calculates the decision (decisions);
3) produces the supporting information for the DM.

The analysis phase. The DM: (1) evaluates the presented decision (decisions) and determines whether the decision (one of them) is acceptable. If so, the HCP is complete; otherwise, the DM analyses supporting information, (3) provides the DMSS with additional information to calculate a new decision (decisions).

D) The fourth group of tasks characterised by multi-connectivity, heterogeneity and multi-criteria information about the DM's benefit system is discussed below.

Such information includes a balanced scorecard (financial, customer, internal processes, learning and growth) and information about effective IT implementation related to business development priorities. This information is influenced by the time-space factor. Both the results of activity and costs are distributed, so it is necessary to compare the "value" of IT implementation today, tomorrow and a year from now, and the development prospects, etc.

Finding the optimal solution will be called synthesizing information about the DM's benefit system. The synthesis aims to find a system which compromises and optimises indicators with limited input data and a spectrum of defined conditions. Note that the synthesis of a system of this type must be vector, i.e. performed taking into account the values of a set (vectors) of indicators, including technological and economic ones, which are taken into account in advance (predicted) in the criteria of superiority (criteria of optimality of the system).

Synthesis is called a vector, which is performed taking into account several indicators, that is, based on vectors $K(\kappa_1, \kappa_2, ..., \kappa_m)$. This is determined by the properties of a multi-criteria complex system. In contrast to the vector, synthesis, carried out according to one quality indicator, is called a scalar synthesis.

Therefore, when conducting a vector synthesis, it is necessary to determine such values of the control variables $x \in D$ that simultaneously provide a minimum of all the introduced optimality criteria $Q_k(x)$, $\kappa = 1,2, ..., s$. These criteria are certainly contradictory; optimisation according to each of them leads to different values of the control variables x. Therefore, to take into account the entire set of partial criteria, it is necessary to analyse the vector optimality criterion Q(x) = [Q1(x), ..., Qs(x)], which

leads to the solution of the multi-criteria optimisation problem.

Solving the optimal synthesis problem is selecting variables x that belong to the permissible domain D and providing the optimal value of the characteristic $Q(x)$. The characteristic that shows the relative "advantage" of one option compared to others is called an optimality criterion (objective function, efficiency criterion, utility function, etc.).

The extreme value of the optimality criterion Q(x) (quantitative value) characterises one of the essential properties of the system. Depending on the specific task, it is necessary to obtain either the maximum or the minimum of this function.

Thus, for each criterion $Q_1(x)$, $Q_2(x)$,..., $Q_s(x)$, it is necessary to find the vector $x = (x_1\ x_2,\ \ldots,\ x_n)$ that provides the minimum (maximum) value criterion of optimality

$$Q=Q(x_1,x_2,\ldots,x_n), \quad (3)$$

when solving the system of inequalities

$$Q(x_1,x_2,\ldots,x_n)\geq 0, \quad (4)$$

$$x_{j-} < x_j < x_{j+},\ j = 1,2,...,n, \quad (5)$$

where x_{j-}, x_{j+} — values of the j-th controlled variable, characterising the range of its possible changes based on real conditions.

Therefore, the solution of the optimisation problem is reduced to solving the optimisation condition of expressions (3) - (5), that is, to determine the optimal value that satisfies the inequalities (4), (5) and finding the minimum (maximum) value of the optimality criterion (3).

5 CONCLUSIONS

1) The paper discusses the stages of IT service management goal evolution.

2) It is proposed to use a methodology that helps companies achieve optimum efficiency from IT implementation, maintaining a balance between obtaining benefits and optimising risks and resources.

3) The methodology is based on five key principles for business IT governance and management: meeting stakeholder needs; covering the company end-to-end; applying a single, integrated methodology; enabling a holistic approach; separating governance from management.

4) The methodology aims to build a goal that: defines necessary and reasonable objectives and tasks at different levels of responsibility; allows the retrieval from the knowledge base of all data relating to the defined objectives that may be required for implementation, improvement or quality assessment projects; clearly identifies and demonstrates (in some cases in great detail) how the enablers help achieve company objectives.

5) Under this principle, the methods for solving multi-criteria optimisation problems are classified according to characteristics of information: decision-making under certainty; decision-making when information about the benefit system is unavailable; decision-making with the gradual acquisition of information about the benefits system; decision-making under multi-connectivity, heterogeneity, multi-criteria information about the system of benefits.

REFERENCES

[1] R. Partha Pratim, "An Introduction to Dew Computing: Definition, Concept and Implications," IEEE Access, vol. 6, pp. 723-737, 2018.

[2] M. Ahmadreza, Y. Mohammad Hossein, and L. Alberto, "Green Cloud Multimedia Networking: NFV/SDN Based Energy-Efficient Resource Allocation," IEEE Transactions on Green Communications and Networking, vol. 4, no. 3, pp. 873-889, 2020.

[3] V. Luis, R. Luis, C. Juan, and L. Maik, "It's probable that you've misunderstood 'Cloud Computing' until now," Sigcomm Comput. Commun. Rev. TechPluto, vol. 39, no. 1, pp. 50-55, 2009.

[4] S. He, L. Guo, Y. Guo, and M. Ghanem, "Improving Resource Utilisation in the Cloud Environment Using Multivariate Probabilistic Models," in Proceedings of 2012 IEEE 5th International Conference on Cloud Computing (CLOUD), 2012, pp. 574-581.

[5] M. Ming and M. Humphrey, "A Performance Study on the VM Startup Time in the Cloud," in Proceedings of 2012 IEEE 5th International Conference on Cloud Computing (Cloud2012), 2012, pp. 423.

[6] B. Dario, D. Salvatore, L. Francesco, P. Antonio, and S. Marco, "Workload-Based Software Rejuvenation in Cloud Systems," IEEE Transactions on Computers, vol. 62, no. 6, pp. 1072-1085, 2013.

[7] N. Seyed, H. Li, S. Venugopal, W. Guo, M. He, and W. Tian, "Autonomic Decentralized Elasticity Based on a Reinforcement Learning Controller for Cloud Applications," Future Generation Computer Systems, vol. 94, pp. 765-780, 2019.

[8] "Evolving business support systems for future services," TM Forum, 2018. [Online]. Available: https://inform.tmforum.org/research-and-analysis/reports/evolving-business-support-systems-for-future-services/. [Accessed: Mar. 27, 2023].

[9] "Prospects for the development of the cloud computing market in Ukraine: advantages and risks. Analytical note," National Institute of Strategic Education, 2019. [Online]. Available: http://www.niss.gov.ua/articles/1191.

[10] N. Susan, "Standardization and Modularity in Data Center Physical Infrastructure," 2011, Schneider Electric, p. 4.

[11] L. Francesco, P. Antonio, and U. Steven, "Value-Oriented and Ethical Technology Engineering in Industry 5.0:

A Human-Centric Perspective for the Design of the Factory of the Future," Applied Sciences, vol. 10, no. 12, pp. 4182, 2020. [Online]. Available: http://doi:10.3390/app10124182.

[12] C. Baotong, W. Jiafu, S. Lei, L. Peng, M. Mithun, and Y. Boxing, "Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges," IEEE Access, vol. 6, pp. 6505-6519, 2018. [Online]. Available: http://doi:10.1109/ACCESS.2017.2783682. ISSN 2169-3536.

[13] P. Antonio, L. Francesco, N. Letizia, and M. Giovanni, "A Digital Twin based Service Oriented Application for a 4.0 Knowledge Navigation in the Smart Factory," IFAC-PapersOnLine, vol. 51, no. 11, pp. 631-636, 2018. [Online]. Available: http://doi:10.1016/j.ifacol.2018.08.389. ISSN 2405-8963.

[14] Y. Yong, S. Kathryn, and L. Dongni, "The evolution of production systems from Industry 2.0 through Industry 4.0," International Journal of Production Research, vol. 56, no. 1-2, pp. 848-861, 2018. [Online]. Available: http://doi:10.1080/00207543.2017.1403664. ISSN 0020-7543.

[15] G. Lucia, P. Antonio, and U. Steven, "Designing Smart Operator 4.0 for Human Values: A Value Sensitive Design Approach," Procedia Manufacturing, vol. 42, pp. 219-226, 2020. [Online]. Available: http://doi:10.1016/j.promfg. 2020.02.073. ISSN 2351-9789.

[16] P. Mattia and P. Lucia, "A new perspective on technology-driven creativity enhancement in the Fourth Industrial Revolution," Creativity and Innovation Management, vol. 31, no. 1, pp. 109-122, 2022. [Online]. Available: http://doi:10.1111/caim.12468. ISSN 0963-1690.

[17] "Information services in в XXI," [Online]. Available: http://itclub-vologda.ru/sites/default/files/news/attachment/information_services_xxi_information_management_3_2013.pdf.

[18] McD. James, "What Can Digitization Do For Formulated Product Innovation and Development," Polymer International, vol. 70, no. 3, pp. 248-255, Mar. 2021, doi: 10.1002/pi.6056.

[19] A. Szajna, R. Stryjski, W. Wozniak, N. Chamier-Gliszczynski, and M. Kostrzewski, "Assessment of Augmented Reality in Manual Wiring Production Process with Use of Mobile AR Glasses," Sensors, vol. 20, no. 17, p. 4755, Aug. 2020, doi: 10.3390/s20174755.

[20] B. Hendrik Sebastian and H. Evi, "Impact of IoT challenges and risks for SCM," Supply Chain Management, vol. 24, pp. 39-61, Jan. 2019, doi: 10.1108/SCM-03-2018-0142.

[21] G. Tymchik, V. Kolobrodov, M. Kolobrodov, A. Vasyura, P. Komada, and Z. Azeshova, "The output signal of a digital optoelectronic processor," Proc. SPIE 10808, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2018, no. 108080W, Oct. 2018, doi: 10.1117/12.2501595.

[22] I. Chyzh, V. Kolobrodov, A. Molodyk, and et al., "Energy resolution of dual-channel opto-electronic surveillance system," Proc. SPIE 11581, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2020, article no. 115810K, Oct. 2020, doi: 10.1117/12.2580338.

[23] S. Dovgiy, O. Kopiika, and O. Kozlov, "Architectures for the Information Systems, Network Resources, and Network Services," in Proc. of the Workshop CPITS-II-2021, 2022, pp. 293-301.

[24] S. Dovgiy and O. Kopiika, "Standard Model of System Architecture of Enterprise IT Infrastructure," in Progress in Advanced Information and Communication Technology and Systems, 2022, pp. 181-201. [Online]. Available: http://doi: 10.1007/978-3-031-16368-5_9.

[25] I. M. Sable, Selection of optimal parameters in problems with many criteria, M.: Drofa, 2006. ISBN 5-7107-7989-X.

[26] A. Rubinstein, Lecture Notes in Microeconomic Theory, 2nd ed., Princeton University Press, 2013. ISBN 978-0-691-15413-8.

[27] B. Rua, Problems and methods of decision-making in problems with multiple objective functions. Questions of analysis and decision-making procedures, M.: Mir, 1976.

Investigation of Stochastic Fluctuation Spreading through a Longwall Gob

Victor Nazimko[1], Artem Merzlikin[2] and Ludmila Zakharova[1]
[1]*Institute of Geotechnical Mechanics, National Academy of Sciences of Ukraine, Simferopol Str. 2A, Dnipro, Ukraine*
[2]*Department of Mining, Donetsk National Technical University, Shybankova Square 2, Pokrovsk, Ukraine*
victor.nazimko@gmai.com, artem.merzlikin@donntu.edu.ua

Keywords: Random Fluctuation, Risk, Stochastic Simulation, Ground Pressure, Coal Mine, Longwall, Gob.

Abstract: We considered a problem of random fluctuation spreading in complex multilayer rock strata during longwall extraction of a coal seam at a great depth. This issue is urgent because it is not easy to forecast harmful and dangerous events in underground environment having extremely uncertain geologic structure and variation of the rock properties. This paper aimed to investigate the process of the rock property random fluctuation to understand if it is possible to forecast the state of the longwall and ground pressure distribution in the gob in particular. We used a numerical computer simulation of the stress-strain redistribution during the longwall advance. The computer models were verified by the results of the ground pressure experimental monitoring in situ. We investigated the distributions of the fluctuation disturbance and revealed the underlying patterns that follow a simple and complex system. We concluded that it is possible to forecast the harmful situation during longwalling but it is need to select a direct and simple chain of cause-effects. We developed efficient proactive technologies to control the longwall operations.

1 INTRODUCTION

Certain processes are rear in real life. On the contrary, uncertainty evolves in any area of human activity, namely politics, economics, nature, technique, and many others. This paper discusses the mining technology process, particularly, coal mining because of the current energy global crisis [1].

Those countries having coal deposits can mitigate the crisis, intensifying the development of their own reserves. However, the depth of mining steadily increases which worsens development conditions because of the high level of ground pressure, coalbed methane content and temperature. These factors elevate the risks of dangerous dynamic processes, coal and rock bursts [2]. These problems has increased even more due to the high level of geologic uncertainty that expands the range of possible technological outcomes, in particular, risks of underground roadways failure [3], and stability of the longwall operation [4].

The modern long-walling technology of underground coal extraction is very complex employing powerful equipment, which operates intensively in hard and uncertain conditions. The practice has shown that big damage, which causes financial lost, is triggered by a small random fluctuation of geologic environment, the voltage of energy, or simple neglect of the regulations and technical standards [5].

Despite the great amount of research that has been made on the longwall reliability problem [6],[7], the chain from the triggering fluctuation down to failure or breakage of the entire longwall panel has not been investigated properly. This paper aims to trace how this random fluctuation spreads and multiplies on an example of the rock strata surrounding a moving longwall face.

2 METHODOLOGY

We used numerical simulation of the stress state of surrounding rock mass in dynamics employing the longwall movement simulation. First, we simulated certain situation, monitoring the stress state of surrounding strata along the whole panel. Then a small fluctuation was introduced into a spot on the immediate roof of the coal seam and this simulation has been repeated. After this, we found the difference

between the previous and secondary simulation to investigate it.

We employed two numerical models, namely simple and complex one. The simple model simulated only one rock layer that covered the extracted coal seam. Real geological strata was presented in the complex model and the task was repeated for the 3D state of the rock mass, where adjacent rock layers interacted. Such an approach helped to trace the spread of the fluctuation in a single rock layer and the entire 3D rock mass.

2.1 Simple Model

A thin plate presented the strata of the immediate and main seam roof [8]. This plate rested on the elastic base from the coal seam and its floor (Figure 1).

The bending and sag of this plate were found solving the differential equation by the final difference method on a regular grid (1):

$$\frac{\partial^4\omega}{\partial x^4} + 2\frac{\partial^4\omega}{\partial x^2\partial y^2} + \frac{\partial^4\omega}{\partial y^4} = \frac{q}{D} - \frac{K\omega}{D}, \quad (1)$$

where ω is the sag, x and y are coordinates of 2D space, q is overburden pressure of the strata, K is the base rigidity, and D is the cylindrical rigidity of the plate-roof. The roof was caved when

$$\sigma_e = \sqrt{\left(\sigma_{x(y)} + 3\tau_{xy}^2\right)} > [\sigma], \quad (2)$$

where σ_e is equivalent stress, σ_x is the normal stress, τ_{xy} is shear stress, and $[\sigma]$ is a tensile limit of the roof rock.

The roof strata caves not for the entire thickness immediately behind the longwall face [9],[10]. First, the immediate roof caves, then the main roof caves by portions while the face withdraws. The caving process starts when condition (2) is satisfied. However, the height of the caving depends on the difference between σ_e and $[\sigma]$. The more the height the bigger the rigidity of the gob (3):

$$K_g = \frac{c_1 K\sigma_e}{[\sigma]} \le c_2 K, \quad (3)$$

where K is the base rigidity in the gob area, c_1 and c_2 are empirical coefficients.

Such a simple algorithm was available to simulate feasibly the periodic caving of the roof after proper tuning of the empirical coefficients.

We used "Pokrovs'k" coal mine geological conditions (Ukraine) as a prototype to simulate the strata behavior. The depth of mining was 1000 m. A flat 2.0 m coal seam has been extracted by the 256 m longwall with complete caving of the roof. Rate of advance was 300 m per month.

2.2 Complex Model

FLAC3D commercial code was employed to simulate the caving dynamics of strata in the 3D state. This code has essential advantage that allows simulating the real path of the model loading for the condition when the ground moves irreversibly [11].

FLAC3D solves equation of Newton second law (4):

$$\left(\frac{d\sigma_{ij}}{dx_j} + pd_i = \rho\frac{dv_i}{dt}\right), \quad (4)$$

where ρ is density, v is velocity of the ground movement.

Therefore, this code accounts for dynamics of the ground movement explicitly which allows tracing the path of the fluctuation and understanding how it spreads in the gob.

Damping of the unbalanced forces in the model was simulated according to (5), (6):

$$F_i^l + F_i^l = M^l\left(\frac{dv_i}{dt}\right), \quad (5)$$

$$N_i^l = -\alpha\left|N_i^l\right| sign\left(v_i^l\right), \quad (6)$$

where F is an active force, N is a damping force, M is mass. Technically, the damping process has been realized iteratively. To reach the static equilibrium, the share of unbalanced forces should be less than a predetermined limit, for example 10^{-5}. However, the rock mass that surrounds a moving longwall face is not in equilibrium. Thus, we control the rate of the longwall face advance by the number of iterations [12]. The fewer number the more rate of the advance.

The overall view of the model is depicted in Figure 1. We removed the upper left part of the model to show it inside.

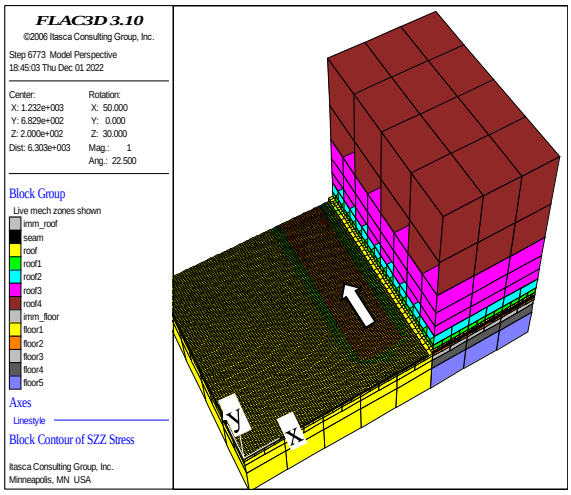


Figure 1: View of the model.

Dimensions of the model were 2048 m by 1280 m by 1500 m along the longwall (X-axis) and along the panel (Y-axis) correspondingly. The coal seam was covered by 1100 m overburden strata. Four hundred meter strata bedded the seam. Such dimensions allowed placement the longwall panel inside the model and provided relevant boundary conditions. The arrow indicates direction of the longwall advance.

Table 1 presents the initial properties of the rock mass.

Table 1. Properties of the rock in the model.

Name of the rock	Bulk modulus, MPa	Shear modulus, MPa	Cohesion, MPa	Angle of friction, degree	Tension limit, MPa	Dilation, degree
rock	24.0	14.4	9.0	35	7.0	0.1
roof2	24.0	14.4	9.0	35	7.0	0.1
roof1	24.0	14.4	7.0	28	5.0	0.5
roof	24.0	14.4	7.0	28	5.0	1.0
main_roof	20.0	12.0	4.0	35	3.0	2.0
imm_roof	19.0	11.0	4.0	35	3.0	2.0
seam	20.0	12.0	4.0	32	2.5	2.0
imm_floor	20.0	12.0	7.0	28	5.0	1.0
floor1	24.0	14.4	9.0	35	7.0	0.1

Fluctuation of the rock properties was inserted in one zone of immediate roof at the range of 840≤Y≤ 850 m and 976≤X≤1024 m. This range is indicated in Table 1 as 'imm roof' line.

Here is fragment of a subroutine in FISH code that has been used to simulate the longwall movement:

```
def advance_dp
    x_vent = 1168 ;positioning the face
    x_conv = 864
    ystart = 170.0
    yfinish = ystart + d_y        ;-----   ;advance
        along Y-axis
    loop while yfinish <= ymax
       command
            mo null range group seam     x x_vent
            x_conv    y ystart yfinish
            ini szz -0.1e0   sxx -0.1e0 syy -0.1e0
            range group seam    x x_vent x_conv    y
            ystart yfinish
            interface 1 face range  x x_vent x_conv
            y ystart yfinish z -0.2 0.2
            interface 1 prop kn 100e9 ks 100e9 fric
            30 coh 15e9 ten 10e9
       end_command
          command
               run01;----cycling
               id_contact
            end_command
         ystart = yfinish
         yfinish = ystart + d_y
     end_loop;--while
     num_panel = 0
end ;--advance_dp
```

The startup room of the longwall was set 170 m from the boundary to prevent its impact. The face was moving by 10 m-runs (d_y = 10 m). The extracted portion of the coal seam was replaced by an empty room (model null). An interface was put down to the immediate floor to prevent intrusion of the roof into the floor. It should be noticed that we input the dilation angles of the rocks to simulate their actual dilation due to caving. This approach helped to simulate the real contact of the roof and floor in place of equivalent rigidity of the gob [13],[14] which increased reliability of the simulation results [15].

The contact moment was determined after every step of the advance. The state of the model was saved before the longwall went into the fluctuation zone to analyze all components of the strata stress-strain state.

3 RESULTS OF SIMULATION

Figure 2 demonstrates the difference between the initial stress state in the gob plane and that disturbed by a fluctuation. Fragment (a) shows the effect of the small fluctuation, the magnitude of which is 1% from the average level of the ground pressure, whereas fragment (b) demonstrates the influence of the big fluctuation. Its magnitude exceeds 23%. These distributions were made with the simple model [8]. This simulation proved that process of any fluctuation spreading is stochastic. This process develops even if the magnitude of the fluctuation is less than 1%. The fluctuation multiplies itself along both the longwall line and the line of the advance direction.

There is a spatial threshold and an incubation period when a local fluctuation starts to develop. For example, the effect of the fluctuation became evident when the face withdrew to 205 m from the spot where the fluctuation occurred and to 550 m for the fluctuation magnitude of 24% and 1% respectively. Therefore, the less the fluctuation magnitude the longer the incubation period and the farther the longwall face withdraws before the effect of the fluctuation becomes obvious. When the total disturbance accumulates to a sufficient amount, the effect from the initial fluctuation bursts up and spreads as the chain reaction [16]. Thus, the big fluctuation caused a quick or prompt effect whereas the influence of the small fluctuations delayed.

Histograms of the accumulated perturbations are symmetrical both in simple and complex models (Figure 3).

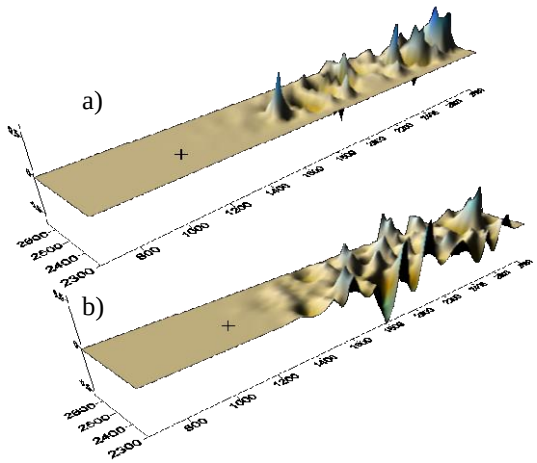


Figure 2: Evolution of the disturbance: a) small fluctuation, b) big fluctuation.

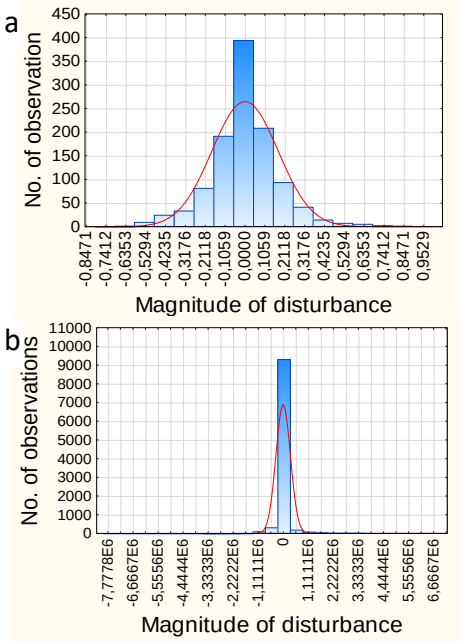


Figure 3: Distribution of the disturbance: a) simple model; b) complex model: STD is 10% and 3% correspondingly.

The histograms correspond to the normal distribution but have abnormal excess which means prevailing the minor agitations.

The analysis of the complex model results has shown that they fairly correspond to well-known pattern of stress state around the longwall gob [17],[18]. Figure 4 illustrates distribution of the damaged zone around the gob far behind the longwall. Shear strain plays the principal role in the process of rock transition over the strength limit [19]. The tensile stress caused damage in the immediate and main roof of the coal seam.

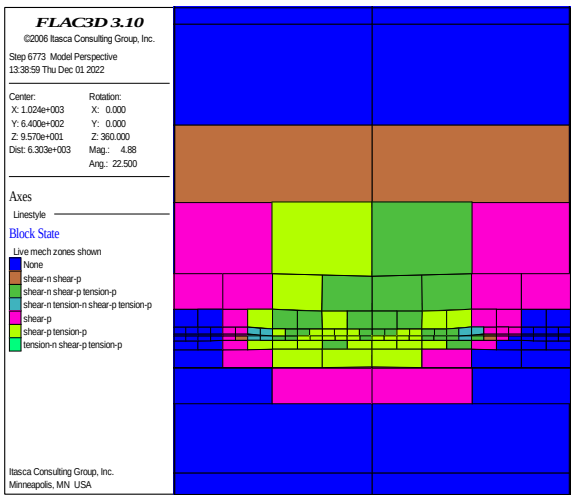
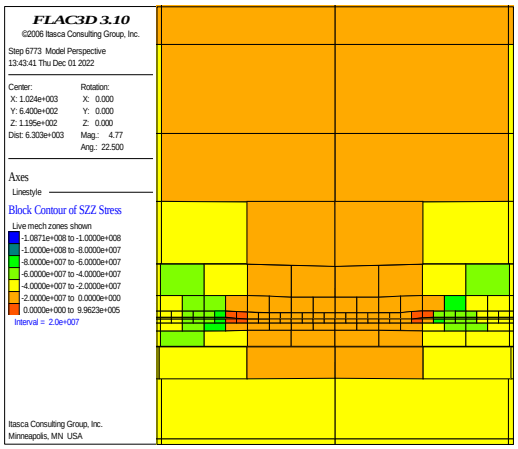


Figure 4: Damage zone distribution around the gob.

Extraction of the coal seam caused evident stress relief around the gob (Figure 5). The minimum stress vertical component occurs at the ribs of the gob where stress is even tensile. The compression stress level is 12-14 MPa in the middle of the gob or 43% from the geostatic background.



Fugure 5: Ground pressure distribution around the gob.

The maximum stress concentration occurs at the abutment zone [20],[21] where the stress' component exceeds 60 MPa.

Figure 6 demonstrates periodic variation of ground pressure in the gob along the longwall advance. The step of stress variation varied in the range of 10-20 m which perfectly matched the experimental data (Figure 7) [22],[23].

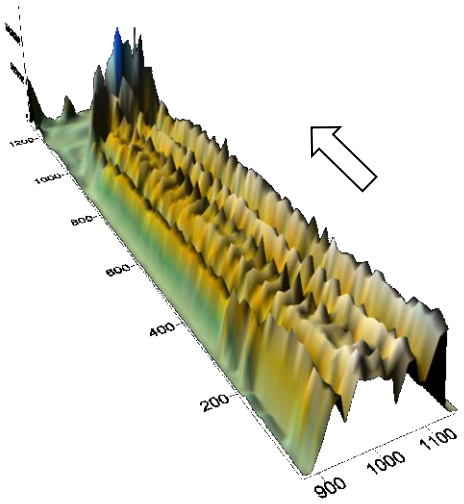


Figure 6: Ground pressure distribution along the gob: the arrow indicates the direction of the longwall advance.

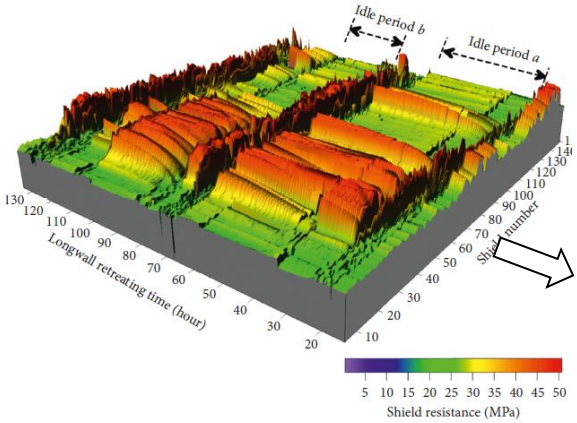


Figure 7: Experimental variation of the ground pressure according to [21].

Figure 8 presents the disturbance of the fluctuation in the complex model. These distributions are similar to those provided by the simple model. The common feature of the both distributions is that magnitude of the perturbations does not depend on the amplitude of the triggering fluctuation. This is a consequence of the fundamental law of nature that governs the complex systems [24].

In addition, even a small fluctuation might find the shortest way to disturb any point of a big complex system through the intermediate numerous chains, events, and reasons [25],[26]. That is why the complex model did not reveal the delay and incubation period.

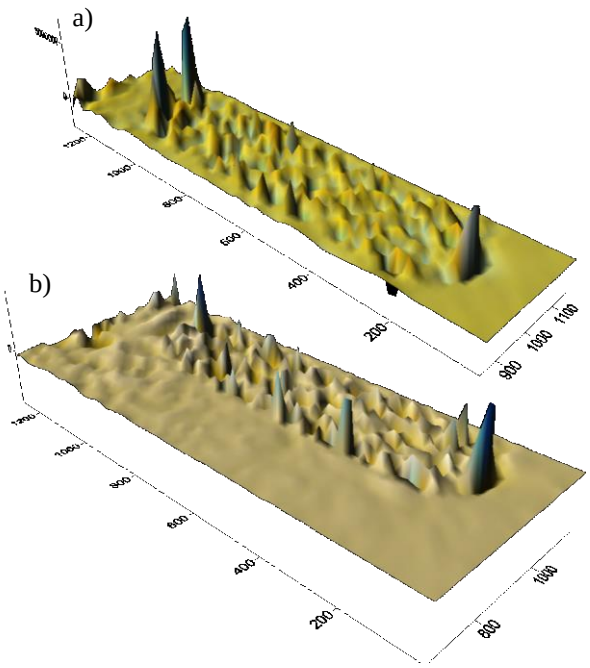


Figure 8: Distributions of the perturbations in the complex model: a) and b) are big and small fluctuation respectively.

4 DISCUSSION

Does it mean that it is impossible to forecast dangerous or harmful situation during longwalling? It is definitely not. The longwall technology involves a set of technological processes, which interact and have direct and indirect links. We found such chains, decomposed them, and revealed typical patterns that could be used to forecast the critical state of the longwall and to prevent it by employing proactive control [27].

We proposed to monitor the current variation of the longwall output using a sliding window having the optimal width varying in the range from 12 to 16 days. The everyday output was reduced according to a coefficient that is in the exponential dependence on the time tardiness from the current day. We used an entropy potential, which is proportional to the sum of derivatives from the output standard deviation. The failure of the longwall operation occurs if the positive value of the entropy potential keeps going a certain period that is proportional to the predetermined risk level. This approach helped to take the proactive measures against the risk.

5 CONCLUSIONS

Stochastic numerical simulation has been used to investigate the spreading and multiplying the random fluctuation in the rock strata during longwall mining a coal seam. It was found that the histogram of the perturbations caused in the rock mass by rock property fluctuation is symmetrical and is similar to the normal distribution having abnormal excess.

The magnitude of the disturbance does not depend on the amplitude of the triggering fluctuation. The simple cause-effect chains can functioned with delay, when there is the incubation period while the fluctuation will spread over the entire volume of the rock mass. The less the fluctuation magnitude the longer the incubation period. However, in the complex system of the multilayer rock strata any fluctuation might find the shortest paths to any point of the strata without evident delay. The fundamental laws of chaos' evolution in nature govern the process of fluctuation spreading. It is possible to forecast and prevent harmful or dangerous situation by employing proactive control of the longwall. We selected simple direct cause-effect chains [28] to realize such a control.

ACKNOWLEDGMENTS

National Academy of Sciences of Ukraine, grant 0120U100081 and Ministry of Education and Science, Ukraine, grant 0121U110527 supported this work.

REFERENCES

[1] T.D. Berrill, "The 2022 Energy Crisis & Market Failure Review of Energy Policy in Queensland View project," 2022, [Online]. Available: https://www.researchgate.net/publication/363289876.

[2] W. Pytel, K. Fuławka, and P. Mertuszka, "Effect of Near-Wave Field Mining Tremors on the Rock Bolt Support Behaviour within Stress Domain," in IVth International Innovative Mining Symposium, 2019, pp. 1-12, [Online]. Available: https://doi.org/10.1051/e3sconf/201910501007.

[3] G. Xu and M. Gutierrez, "Study on the damage evolution in secondary tunnel lining under the combined actions of corrosion degradation of preliminary support and creep deformation of surrounding rock," Transp. Geotech., vol. 27, p. 100501, Mar. 2021, [Online]. Available: https://doi: 10.1016/j.trgeo.2020.100501.

[4] B. Zhao and J. Wang, "Study on interval rupture mechanism and support optimization of layered roof," Eng. Fail. Anal., vol. 141, p. 106690, Nov. 2022, [Online]. Available: https://doi: 10.1016/j.engfailanal. 2022.106690.

[5] S. Li, Z. Jia, and Q. Ye, "Study on Dynamic Response of Damper under Gas Explosion Impact," Sustainability, vol. 15, no. 4, p. 3356, 2023, [Online]. Available: https://doi.org/10.3390/su15043356

[6] S. S. Peng, F. Du, J. Cheng, and Y. Li, "Automation in U.S. longwall coal mining: A state-of-the-art review," Int. J. Min. Sci. Technol., vol. 29, no. 2, pp. 151-159, Mar. 2019, [Online]. Available: https://doi: 10.1016/j.ijmst.2019.01.005.

[7] C. Mark, "An updated empirical model for ground control in U.S. multiseam coal mines," Int. J. Min. Sci. Technol., vol. 31, no. 2, pp. 163-174, Mar. 2021, [Online]. Available: https://doi: 10.1016/j.ijmst. 2020.12.024.

[8] V. Nazimko and K. Babenko, "Spreading of ground pressure fluctuation in the gob," in Proceedings of the 36th International Conference on Ground Control in Mining, Jul. 2017, pp. 131-137.

[9] R.H. Trent and W. Harrison, "Longwall mining introduction," AIME, New York, 1982.

[10] D.C. Oyler, C. Mark, D.R. Dolinar, and R.C. Frith, "A study of the ground control effects of mining longwall faces into open or backfilled entries," Geotechnical and Geological Engineering, vol. 19, no. 2, pp. 137-168, 2001.

[11] Itasca Consulting Group, "FLAC3D Fast Lagrangian Analysis of Continua in 3 Dimensions." 2008.

[12] V. Nazimko and K. Babenko, "Use of FLAC3D for mining induced seismicity prediction," in Proceedings of the 2nd International FLAC/DEM Symposium. February, 14-16, Melbourne, Australia, pp.85-92, 2011

[13] G.S. Esterhuizen, C. Mark, T.M. Klemetti, and R.J. Tuchman, "Proceedings of the International Workshop on Numerical Modeling for Underground Mine Excavation Design Ordering Information," 2009. [Online]. Available: www.cdc.gov/niosh.

[14] Y. Lu and L. Wang, "Numerical simulation of mining-induced fracture evolution and water flow in coal seam floor above a confined aquifer," Computers and Geotechnics, vol. 67, pp. 157-171, 2015.

[15] W. Pan, X. Nie, and X. Li, "Effect of premining on hard roof distress behavior: a case study," Rock Mechanics and Rock Engineering, vol. 52, no. 6, pp. 1871-1885, 2019.

[16] P.A. Blythe, A.K. Kapila, and M. Short, "Shock-induced chain-branched ignition," Proc. Combust. Inst., vol. 32, no. 2, pp. 2371-2377, 2009, [Online]. Available: https://doi: 10.1016/j.proci.2008.05.071.

[17] R. Sňuparek and P. Konecný, "Stability of roadways in coalmines alias rock mechanics in practice," Journal of Rock Mechanics and Geotechnical Engineering, vol. 2, no. 3, pp. 281-288, 2010.

[18] G. Feng and P. Wang, "Stress environment of entry driven along gob-side through numerical simulation incorporating the angle of break," Int. J. Min. Sci. Technol., vol. 30, no. 2, pp. 189-196, Mar. 2020, [Online]. Available: https://doi: 10.1016/j.ijmst.2019. 03.003.

[19] G. S. P. Singh and U. K. Singh, "A numerical modeling approach for assessment of progressive caving of strata and performance of hydraulic

powered support in longwall workings," Computers and Geotechnics, vol. 36, no. 7, pp. 1142-1156, 2009.

[20] M. Salamon, "Mechanism of caving in longwall coal mining," in Proceedings of the 31st US Symposium on Rock Mechanics Contributions and Challenges, pp. 161-168, Golden, CO, USA, June 1990.

[21] S. S. Peng, Longwall Mining, CRC Press, Boca Raton, FL, USA, 3rd edition, 2019.

[22] R. Trueman and I. Hutchinson, "The use of shield monitoring data for predicting in advance roof control problems on longwall faces," Min. Technol. Trans. Inst. Min. Metall., vol. 127, no. 4, pp. 209-218, Oct. 2018, [Online]. Available: https://doi: 10.1080/25726668.2018.1458478.

[23] J. Cheng, Z. Wan, and Y. Ji, "Shield-Roof Interaction in Longwall Panels: Insights from Field Data and Their Application to Ground Control," Adv. Civ. Eng., vol. 2018, pp. 1-18, Dec. 2018, [Online]. Available: https://doi: 10.1155/2018/3031714.

[24] D. Orrell, "Role of the metric in forecast error growth: how chaotic is the weather?," Tellus A Dyn. Meteorol. Oceanogr., vol. 54, no. 4, p. 350, Jan. 2002, [Online]. Available: https://doi: 10.3402/tellusa.v54i4.12159.

[25] R.L. Devaney, An Introduction to Chaotic Dynamical Systems. Boca Raton: Chapman and Hall/CRC, 2021.

[26] Y. Hu, W. Li, Q. Wang, S. Liu, and Z. Wang, "Study on failure depth of coal seam floor in deep mining," Environmental Earth Sciences, vol. 78, no. 24, 2019.

[27] V.V. Nazimko, M.O. Ilyashov, and I. Diedich, "Prospective tendencies of coal mining risk management," Mining of Mineral Deposits, vol. 13, no. 1, 2019, pp. 111-117. [Online]. Available: https://doi.org/10.33271/mining13.01.111.

[28] C. Liu, Z. Yang, P. Gong, and et al., "Accident analysis in relation to main roof structure when longwall face advances toward a roadway: a case study," Advances in Civil Engineering, vol. 2018, Article ID 3810315, 11 p., 2018.

Cultural Heritage Reconstruction using Virtual and Augmented Reality

Nilufar Tukhboeva
Department of Architecture, Ajou University in Tashkent, Asalobod Str. 113, Tashkent City, Uzbekistan
n.tukhboeva@posteo.de

Keywords: Augmented Reality (AR), 3D Reality- Based Modelling, Tourism Industry, Cultural Heritage, Aksaray.

Abstract: Augmented reality and virtual reality (commonly abbreviated as AR and VR respectively) are reality technologies that either enhance or replace a real-life environment with a simulated one. AR enhances your surroundings by adding digital elements to a live view, usually by using the camera on a smartphone. VR is a completely captivating experience that replaces a real-life environment with a simulated one. This article presents the connection between the reconstruction of medieval architecture and the further improvement of ways to use it effectively, the proper use of architectural monuments to improve travel opportunities is an important challenge. Particular attention is paid to the sustainability of old cities, preserving the buffer zone of architectural monuments, development can improve the tourism industry and convey new opportunities. The first was the creation of a highly appealing visualization system that allows a virtual visit to Aksaray (Amir Timur Palace in Shakhrisabz, Uzbekistan). The second is the possibility of using this technology for a virtual exploration of Aksaray, from a technical and architectural point of view.

1 INTRODUCTION

The current great role of increasing the economic potential of a country's tourism industry is great. The tourism industry is always trying to keep track of new technologies, as the new generation of travelers are active smartphone users who use smart devices to access social platforms. New technological improvements such as “virtual reality” (VR) and “augmented reality” (AR) can improve the tourism industry and open up new opportunities. Recently, various fields of tourism, have already introduced augmented reality technology.

AR is an interactive experience and evolving technology that overlays computer-generated enhanced images or virtual objects onto a live real-world environment in real time, thus providing a composite view.

The technology that makes AR possible is now much more powerful than ever before and is compact enough to deliver AR experiences to tourist sites through smart devices, making AR more sociable, communicative and inclusive, combining art, culture, architecture, design with projections and preservation of architectural landmarks with surrounding sound and special effects.

Augmented reality tools that superimpose digital content on a real environment are a key component of Industry 4.0.

All over the world the scientific research on the expedient use of architectural monuments pays much attention to the preservation and restoration of historical monuments, increasing their durability.

Active measures are being taken to solve the problems related to urban planning, design and architecture, further improvement of education and science in construction, innovative development of construction, preservation of architectural monuments and their appropriate use.

In this regard, the preservation of architectural monuments of the Middle Ages and the further improvement of the ways of their effective use, the proper use of architectural monuments to improve the possibilities of movement are important tasks.

The aim of the project is VR technology, to enrich the gaps in the study of cultural heritage sites related to the buildings of Amir Timur and Temurids in the Middle Ages in the historic cities, using a graphic reconstruction and creation of visual animation. The project proposal is implemented on the basis of historical sources and the results of archaeological research (Figure 1) [1].

Figure 1: Amir Timur's Aksaray Entrance Peshtak.

2 VIRTUAL REALITY FOR THE CULTURAL HERITAGE

The integration of the research in the area of computer graphics, computer vision and 3D survey systems found as a common action field the virtual reconstructions of Cultural Heritage.

The current virtual technologies (VR and AR), devices and sensors are able to immerse the user in a virtual environment, offering the possibility of interacting with additional contextual heritage data (reality-based 3D model, pictures, technical documents, historical information, tourist guides, etc.) [2]. The VR applied to Cultural Heritage has already become a potential tool for 'tourist' users to navigate and interact in a virtual scene in total symbiosis with the environment. The system has an attractive impact for dissemination purposes because it adds extra information and allows a direct and intuitive access of digital objects [3].

VR offers more than a replica of the real context; it allows virtual hypothetical reconstructions of the past aspect that can be developed for different epochs and can be related or overlaid with the archaeological surviving evidences [4]. VR creates a direct connection between tangible data (museum collection, archaeological remains etc.) and ancient invisible space, reproducing the scene, the environment and the atmosphere of past context in an immersive and attractive mode [5].

The potentiality of VR applications to promote knowledge Cultural Heritage is proven by the development of a significant number of projects in this research area [6].

3 CASE-STUDY: AMIR TIMUR'S AKSARAY IN SHAKHRISABZ

In 1976 archaeologist Sultonov X.T. began archaeological excavations in Aksaray building for the reason of the outstanding structure and conservation of the architectural building 3-4 meters south of the ruins of the entrance. At a depth of 1.5 meters above the current earth surface it was discovered a wonderful combination of different colored tiles on the surface of the building (Figure. 2).

Figure 2: Archaeological excavations in Aksaray.

During the excavations numerous ceramics have been found which decorated rooms of the palace. These include limestone glazed and poured with gold water fragments of cornices and embossed epigraphic ceramic tiles with patterns, carved marble tiles and molding's out of ganch and ceramics.

In order to build the ancient monument, it was selected a flat topography and the stable ground for the construction. The foundation of the building is ribbon-shaped and choosing its shape results from the distribution of the mass of the building onto the ground (Figure. 3). To increase the bottom surface area of the foundation in historical monuments trapezoid and curved forms were used. The foundation is 7 meters deep, and the material consists of large stones. (Figure. 4)

In order to ensure equal resistance in the structures of the monument it was used same materials or materials with similar physical and mechanical properties. The main building material was burnt brick which was laid using special type of mortar - ganch. The brick used in the construction of the monument was square shaped, measuring 24 to 28 cm on each side, 3 to 5 cm thick.

The restoration of the monument is an endless, continuous work of repairing, cleaning and replacement that has to be documented to help future activities. The modern answer to this necessity is the management of all the works inside a BIM system, as it combines the digital three-dimensional or

multidimensional representation of an object with an information database (spatial position, technical features, properties of the materials, realization phases, maintenance operation, etc.) [7]. Thanks to this working method, it is possible to create a proper geographic and informative system of the heritage, including the management of its life cycle, from the phase of the project, to the phases of the use and maintenance.

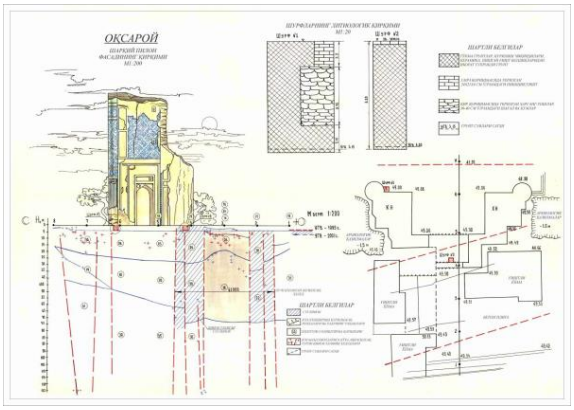


Figure. 3: The foundation.

Figure 4: Stacking of stone rocks.

Therefore, the main effort was spent in the last years to build a dedicated BIM system with this specific goal: supporting to the operations of the restoration yard.

3.1 Virtual Reconstruction

AR focuses on the simplicity and ease of providing an explanation experience, scientific researchers and tourists can gain knowledge and skills through 3D simulations generated by computers and other electronic devices, more so than in traditional ways. AR can explain architectural styles, graphics, or designs of palaces, mausoleums and mosques (Figure 5) [8].

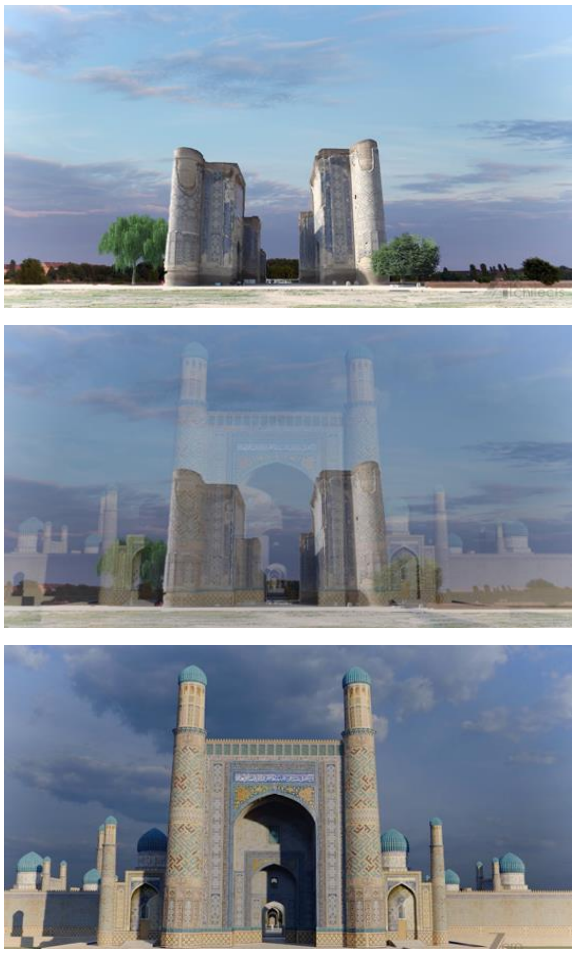


Figure 5: Virtual graphic reconstruction of Aksaray, performed by author.

Improving Information and Understanding: Scientific researchers and tourists know that the process of guided tours should focus on inspiration and interaction. Their goal is to get tourists interested in the subject or theme. Tourists should not be mere listeners and passive observers, some tourists, especially teenagers, need practice and hands-on experience.

The ability to connect reality and digital content is improving, opening up more possibilities for academic researchers and tourists.

For example, AR can improve the amount and quality of information about Amir Timur's palace, making the explanation environment more educational, productive and contextual. To improve the quality of explanation by producing and delivering rich, constructive and useful content. With AR, everything about Amir Timur's palace can be explained more scientifically than ever before through interaction with the computer display, sound, text, and 3D effects (Figure 6) [9].

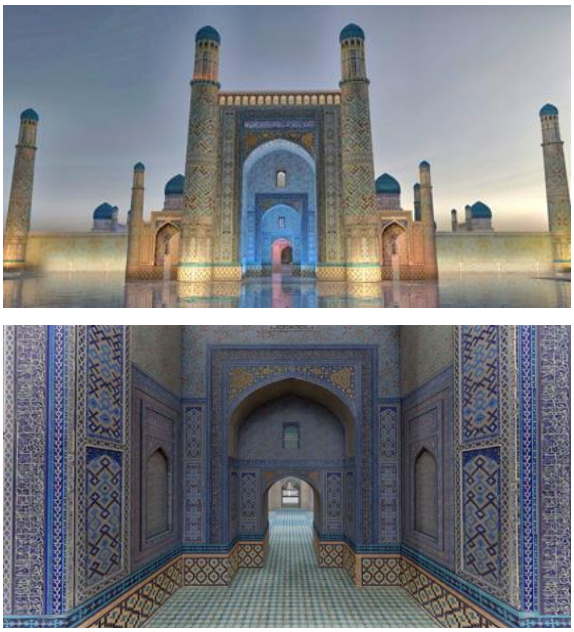


Figure 6: 3D effects.

Figure 7: 3D models (Scheme of schematic reconstruction of Aksaray) by author.

Simplifying abstract concepts: AR has the potential to make tourists more engaged and motivated in learning about restoring history and architectural heritage.

AR technology has the ability to explain abstract and complex concepts by visualizing objects that are difficult to imagine and turning them into 3D models, thereby facilitating the perception of abstract and complex content.

It is especially useful for visual tourists and virtually anyone to translate theoretical material into real-world concepts (Figure 7) [10].

3.2 3D Visual Guide

Visitors to museums and heritage sites can access AR applications via smart devices and discover historical content related to the sites in the form of 3D digital media, videos and images (Figure 8). AR simply provides instant and relevant enriched information about what they see to increase their motivation to learn and their level of understanding.

Moreover, AR offers an enhanced combination of different types of historical sources and visualizations, including 3D models, images, text, video, animation and sound, along with traditional methods to enhance the tourist environment.

Figure 8: Aksaray 3D animation and sound.

4 CONCLUSIONS

The research presented here aimed at applying new VR immersive techniques to the field of architecture and facility management. The main goal was to evaluate how VR could be used in the sector of Cultural Heritage as a support for conservation and maintenance activities. The work represented the follow-up of the previous research activities of survey and modelling of the Amir Timur's Aksaray in Shakhrisabz. The research work led to the creation of a complete high-resolution 3D model, currently used as a support of the maintenance activities of the Amir Timur's Aksaray in Shakhrisabz. Furthermore, in order to promote a wider use of the 3D virtual model to a larger group of users also including non-experts of 3D, an ad hoc building information modelling (BIM) system has been developed. This system is able to connect the modelling software with an external navigation system that encourages a simple and shared use of the 3D model by different stakeholders of the conservation process. It works like a catalyst that collects and connects data and information. The presented work succeeded in adding a further step in terms of virtualization by creating a VR immersive system that not only allows the classical virtual visit of the object, but also provides the technical information stored in the BIM system.

Based on the above historical data it is possible to schematically create restoration drawings of the layout of Amir Timur's Aksaray in Shakhrisabz. AR technology is becoming more and more accessible and therefore more widespread. Despite the growing use of VR in many businesses these days, AR in tourist orienteering is still new and unexplored. However, the potential of AR in tourist orientation is great, providing new ways to guide as guides gain attention and motivate tourists better, and tourists gain new tools to visualize complex concepts as well as gain practical skills.

Overall, the future of AR as an imaging technology looks positive; more attempts have recently been made to expand the use of AR in tourism guidance, and new AR tools for tourism guidance will continue to be developed as the technology becomes more highly developed and advanced than ever before. In addition, a great deal of research is being done to improve the compatibility and applicability of AR in tourist orienteering.

REFERENCES

[1] G. Pugachenkova and L. Rempel, “Outstanding Monuments of Architecture of Uzbekistan,” Tashkent: Gosizhudlit, 1958.

[2] L. Calori, C. Camporesi, S. Pescarin, and A. Guidazzoli, “Open heritage: an integrated approach to web 3D publication of virtual landscapes,” in 3D-ARCH 2005: Virtual Reconstruction and Visualization of Complex Architectures, 2005, pp. 1-5.

[3] M. Forte, S. Pescarin, and L. Pujol Tost, “VR applications, new devices and museums: visitor’s feedback and learning: a preliminary report,” in 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST, 2006.

[4] M. Forte and N. Danelon, “Regium@Lepidi 2200 Project,” Archeomatica, vol. 6(1), pp. 42-48, 2015.

[5] M. Navvab, F. Bisegna, and F. Gugliermetti, “Experiencing the tangible past through virtual reconstruction: cultural heritage of buildings and their environmental boundaries,” Archeomatica, vol. 4(3), pp. 36-41, 2013.

[6] S. Pescarin, A. Pagano, M. Wallergard, W. Hupperetz, and C. Ray, “Evaluating virtual museums: archeovirtual case study,” in 40th Conference on Computer Applications and Quantitative Methods in Archaeology, 2012.

[7] C. Achille, F. Fassi, and L. Fregonese, “4 year history: from 2D to BIM for CH,” in 18th International Conference on Virtual Systems and Multimedia, 2012, pp. 377-382.

[8] G. Clavijo, “Diary of a trip to Samarkand to the court of Timur (1403-1406),” 1970.

[9] H. Sultonov, G. Dresvyanskaya, S. Lunin, and Z. Usmanova, “Shakhrisabz Part II,” Tashkent University, 1993, and K. Abdurashidov, F. Kabulov, B. Rakhmanov, “Engineering problems of architectural monuments,” Tashkent Science, 2011.

[10] T. Jung, N. Chung, and M. Leue, “The determinants of recommendations to use augmented reality technologies: the case of a Korean theme park,” Tourism Management, vol. 49, pp. 75-86, 2015.

On Fractal-Based Estimates of Subsidence Volumes for Various Types of Soils

Tetyana Mokrytska[1] and Anatolii Tushev[2]
[1]*Department of Chimestry, Dnipro National University, Gagarin Avenue 72, Dnipro, Ukraine*
[2]*Department of Mathematics, Dnipro National University, Gagarin Avenue 72, Dnipro, Ukraine*
mokritska@i.ua, anavlatus@gmail.com

Keywords: Particle Size Distribution, Fractal Dimension, Loess, Subsidence.

Abstract: In this paper we study the protracted degradation of the soil structure during technogenesis in loess and various types of silty and clay soils. We develop a new mathematical model of the degradation which is based on applying fractal theory to soils, and our modeling is mainly dealing with particle size distribution by volume. The values of the fractal dimension of the particle size distribution by volume were also calculated for various types of soils. The predicted values of the porosity coefficient in the state of complete decomposition of micro-aggregates were calculated according to the new model. The difference of the porosity coefficients in the initial (natural) and predicted states characterizes the value of the volume deformation of the soil. The particle size distribution of the samples was determined before and after several different sample preparation methods. The obtained results on volumetric deformation of loess and silty-clay soils do not contradict the known data on the deformation features of the deposits of the region formed in the epochs of interglacial and terrestrial glaciation.

1 INTRODUCTION

The forecast of soil deformation under external anthropogenic impact is a topical scientific problem. The amount of deformation depends on the properties and type of soils [1, 2]. During operation of the structure, the state, degree of aggregation and soil structure change [3]. The forecast of final deformations can be obtained using the methods of research of complex systems: neural networks, the method of group accounting of arguments and others [4, 5]. The theory of fractals has been successfully applied to describe structure and properties of soils [6, 7, 8] and to predict the mechanical behavior of the soil environment [9, 10]. In the presented paper, the forecast of final deformations is based on the application of fractal theory.

In modern publications devoted to the problems of deformation and destruction of various geomaterials, considerable attention is paid to the description of the geometric properties of the structures formed by particles and pores. Russell [11] described soil compaction using an example of a medium with particle and pore sizes that change during crushing. Using elements of fractal analysis Zhang and Zhang [12] showed that the ratio between volumetric deformation and the index of crushing is constant during crushing. Seblany et al. [13] performed an analysis of the distribution of pores in the granular material. The influence of the geometric properties of pores of shale on adsorption phenomena was performed in [14] using the Frenkel-Halsey-Hill model. Also the investigation of the influence of the phase composition of geomaterials on their strength and deformability are continued. Xue and others [15] investigated the effect of temperature and stress on electroosmosis in clays. Espitia et al. [16] showed that the Poisson's ratio in clays changes nonlinearly with increasing pressure and this has an effect on water-retaining properties. The effect of dispersion and cohesion on the behavior of soils is also very essential. In [17], internal and external factors of soil aggregates stability were studied, the author identified the types and explained the mechanism of soil aggregates stability.

The above brief review of publications shows that the search for new models of geomaterials and soils is carried out in various aspects, a common tendency is to study materials as complex dynamic systems.
The study of microaggregate composition allows to solve several problems of different types. It is well known that the number of aggregates, particles

composing aggregates and free particles does not remain constant. In [18, 19] we obtained some estimations of subsidence volume for loess soils which are dealing with the fractal dimension of the particle size distribution. In the presented paper we develop a new fractal-based model which describes the process of the emergence of a new structure in a dispersed soil after the disintegration of microaggregates. This model is based on the fractal dimension of the particle size distribution by volume of particles, and it is more appropriate for estimations of subsidence volume. The theoretical conclusions are supported by results of experiments. The techniques of microaggregate analysis are described in [20].

The final deformations were studied as the result of the disintegration of microaggregates in dusty-clay soils of various types (loess-like loams and clays). Samples of natural density and moisture of loess-like loams and clays were examined. We studied Quaternary loess-like loams of eolian-deluvial and deluvial origin and Neogene-Quaternary clays of continental origin. The total number of experiments to determine the microaggregate composition was 200 for loess and 80 for clay. Samples selected from natural outcrops and from bores on the Dnipro raised platform and Donetsk folded structure of the Ukrainian array. The conditions of occurrence of the loess horizons of the Dnipro Upland in the area of the city of Dnipro are characterized with a map (Figure 1).

2 FRACTAL-BASED MODELING OF SOIL SUBSIDENCE

In [21] the particle size distribution $N_s(L > d_s)$ was defined as the number of particles being of any size L larger than d_s, where d_s runs over the real numbers. In the same way we can define the particle size distribution by volume $V_s(L > d_s)$ (and by mass $M_s(L > d_s)$) as the volume (mass) of particles being of any size L larger than d_s, where d_s runs over the real numbers. Certainly, $N_s(L > d_s)$, $V_s(L > d_s)$ and $M_s(L > d_s)$ are real functions.

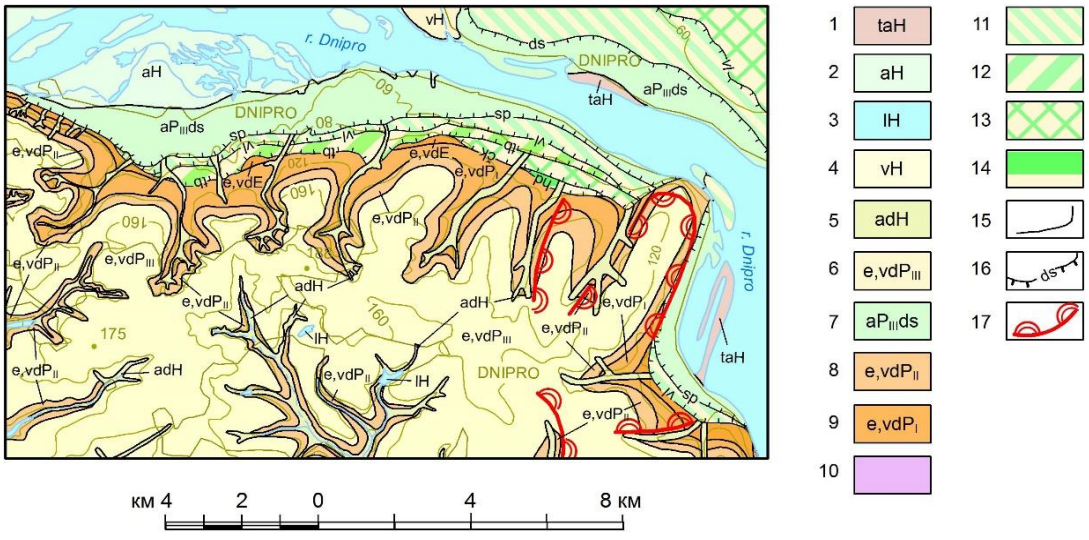


Figure 1: Geological map of Quaternary. Holocene: 1 technogenic sediments; washed; 2 alluvial; 3 limnetic; 4 aeolian; 5 alluvial-deluvial. Upper Quaternary: 6 eluvial, aeolian-deluvial sediments; 7 alluvium of the first terrace (Desnianska terrace); 8 eluvial; aeolian-deluvial sediments of Middle Pleistocene; 9 eluvial, aeolian-deluvial sediments of Lower Pleistocene; 10 pre-quaternary sediments. Terrace alluvium buried under subaerial sediments; 11 vilshanska; 12 trubizka; 13 cherkasska; 14 hadjibeiska; 15 geologic boundaries; 16 terrace inner margin and name; 17 landslide-prone slopes.

The particle size distribution by volume has fractal dimension DV_s if

$$V_s(L > d_s) \approx \gamma {d_s}^{-DV_s}, \qquad (1)$$

where γ is a coefficient and the sign $\approx$ means "approximately". In our further equations we use the sign $=$ taking in mind (1) and realizing that the obtained formulas give us only some estimations of the real situation.

In fact (1) gives us an estimation of the particle size distribution by volume $V_s(L > d_s)$ which is based on a fractal topological invariant DV_s. As it was noted in [11], modeling the particle size distribution using fractals provides a good fit to experimental data (e.g. [20, 22, 23, 24, 25]).

Since in the presented paper we study volumetric characteristics of soil subsidence, the particle size distribution by volume is more convenient for us. Using the fractal dimension of the particle size distribution by volume essentially simplifies obtained formulas and facilitates calculations. The particle size distribution by volume can be easily obtained from the particle size distribution by mass, which appears as the result of the standard procedure of sieve analysis.

The particles forming the ground may have only a finite set of sizes. We denote these sizes

$$d_1, d_2, ..., d_{n-1}, d_n, \qquad (2)$$

which are running from the biggest d_1 to the smallest d_n. Everywhere below we assume that the relation $\alpha = \alpha_j = d_{j+1} / d_j$ does not depend on j, where $1 \le j \le n$. This assumption corresponds to the idea of the self-similarity of fractal structures. In addition, all known mathematical fractals are constructed on this principle. Now, we define an "imaginary" particle size $d_{n+1} = \alpha d_n$ so that the relation $\alpha = d_{j+1} / d_j$ holds for all $1 \le j \le n$.

We will use the following denotations:

- A_j is the structure formed only by particles of one fixed size d_j;
- $V_s(A_j)$ is the volume of particles of size d_j which form the structure A_j;
- $V_p(A_j)$ is the volume of pores of the structure A_j;
- $V(A_j) = V_p(A_j) + V_s(A_j)$ is the volume of the whole structure A_j.

We will assume that the porous structure A_J, formed only by particles of size d_j, is similar to the porous structure A_i, formed only by particles of size d_i, for all $1 \le i, j \le n$. This property corresponds to the nature of self-similar fractal structures and is preserved in the process of subsidence and after it. Then, as the structure A_J is similar to A_i for all $1 \le i, j \le n$, we can assume that $V_p(A_j) / V_s(A_j) = V_p(A_i) / V_s(A_i) = \mathrm{k}_p$ for all $1 \le i, j \le n$, i.e. the coefficient of porosity of the structure k_p of A_J does not depend on j.

As it was mentioned in [12], one of the most common soil models is the granular medium model. We also assume that particles are spherical. It is well known that $k_p = (1-s)/s$, where s denotes the shape factor, which for spheres becomes $s = \pi / 6$, as it was mentioned in Appendix A of [26]. Then below we assume that $k_p = (6-\pi)/\pi$.

Equation (1) allows us to estimate the volume $V_s(A_s)$ of all particles of size d_s. We should note that an increment of the function $V_s(L > d_s)$ is only due to the particles of sizes from (2) (i.e. in points $d_1, d_2, ..., d_{n-1}, d_n$). Thus, $V_s(A_j) = V_s(L > d_{j+1}) - V_s(L > d_j)$ and using (1) we can calculate the volume of particles of size d_j for each d_j from (2) as the following

$$V_s(A_j) = \gamma {d_j}^{-DV_s} \beta, \qquad (3)$$

where $\beta = (d_{j+1} / d_j)^{-DV_s} - 1$ and $1 \le j \le n$.

Then it follows from the definition of the coefficient of porosity k_p that

$$V_p(A_j) = \gamma \, \mathrm{k}_p \, {d_j}^{-DV_s} \beta, \qquad (4)$$

where $1 \le j \le n$. Thus, (3) and (4) give the following

$$V(A_j) = \gamma (1 + \mathrm{k}_p) {d_j}^{-DV_s} \beta \qquad (5)$$

where $1 \le j \le n$. It follows from (4), (5) that $V(A_{j+1}) > V_p(A_j)$.

We assume that particles of different sizes are distributed in the soil evenly and this property is

preserved during the process of subsidence and after it. In the process of subsidence, the largest particles of the size d_1 will form a porous structure A_1 with the largest pores. The particles of the next size d_2 will be in these pores and they will form a porous structure A_2, whose pores will contain particles of the farther smaller size d_3, etc. to the particles of the smallest size d_n. We recall that

$$k' = V_p / V_s = (V - V_s) / V_s = (V / V_s) - 1$$

is the coefficient of porosity of the soil after subsidence. We have proved that

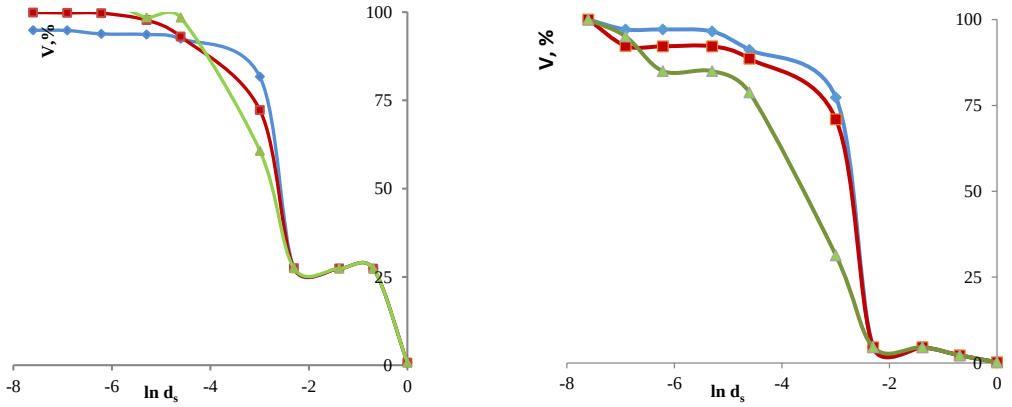
$$k' = \frac{k_p(1-\alpha^{DV_s})}{1-(\alpha^{DV_s})^n} . \quad (6)$$

3 THE RESULTS OF EXPERIMENTAL STUDIES

The values of the particle size distribution by volume for soils of various origin and various preparation methods of samples were calculated. Typical curves of the granulometric composition of the loess of the Kryvyi Rih region and clays of the Western Donbass are presented in Figure 2.

The curves show the changes in the total volume of particles under different conditions of sample preparation for pipetting analysis. It follows from the graphs that the change in the microaggregate composition, depending on the preparation method of the sample for analysis, is manifested better in the clay of the Western Donbass than in the loess of the Kryvyi Rih region. Three different methods of sample preparation are used, therefore, in this case, different degrees of decomposition of microaggregates are achieved. In the first method, technogenic mechanical impact is simulated during mechanical shaking. The second and third methods characterize mainly the chemical impact that occurs in the conditions of large natural-technogenic systems, for example, in cities. Thus, one can get an idea of soil degradation under the conditions of mechanical and chemical technogenic impact.

According to the results of particle size analysis, the values of the fractal dimension of the particle size distribution by volume were calculated for various preparation methods of samples for analysis. Examples are presented in Figure 3.



a) b)

Figure 2: The graph of the particle size distribution by volume $V_s(L > d_s)$ (in %) with logarithmic X-axis for various preparation methods of samples: a) Kryvyi Rih, b) Western Donbass. The volume percentage $V, \%$ of particles whose size is bigger than the specified size d_s (in mm): with the aggregate method of preparation; with the semi-dispersed method of preparation; with the dispersed method of sample preparation for pipetting analysis.

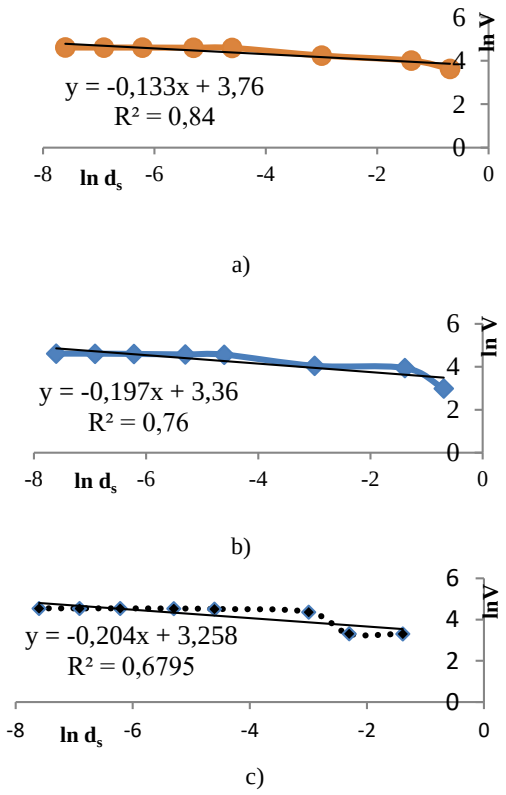


Figure 3: The trend equation calculated by the results of particle size analysis of clay samples with the dispersed preparation: a) red-brown clay, b) green-gray clay, c) Kryvyi Rih loess. Denotations: $ln\,d_s$ - the logarithm of particle size d_s; lnV - logarithm of the volume percentage V of particles whose size is bigger than d_s; R^2 - the coefficient of determination.

The trend equation has the following form: $y = DV_s\,x + b$, where DV_s is the fractal dimension of the particle size distribution by volume. The value of DV_s is then define from the trend equation: a) Red-brown clay $DV_s = 0{,}133$, b) Green-gray clay $DV_s = 0{,}197$, c) Kryvyi Rih loess $DV_s = 0{,}204$. The porosity coefficient of red-brown clays will take values from 0.197 to 0.2. The porosity coefficient of green-gray clays will be in the range of 0.234 – 0.239. In loess, these values will vary from 0.372 to 0.380.

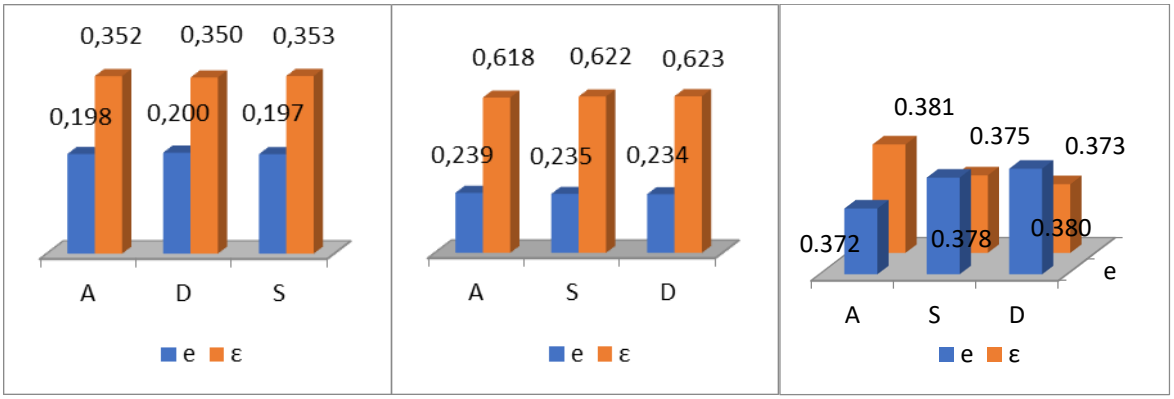
The tendencies of formation of the "new" structure in the studied soils are different. The disintegration of microaggregates will lead to some denser packing of particles of red-brown clay, the packing of particles of green-gray clay of Western Donbass will be relatively dense.

The loosest packaging is expected for particles of loess of the Kryvyi Rih structure. In previous studies, prognosed deformation values for the loess of the Middle Dnipro were calculated (Table 1). According to our calculations we have discovered that for all studied soil samples while disintegration of microaggregates the "new" porous structure is completely formed by particles of the smallest size, the particles of bigger sizes are presented in the "new" structure as separated inclusions. So, for each soil sample the porosity coefficient was calculated by (6) and then we calculated the average predicted porosity coefficient e for each type of soil and each method of sample preparation (Table 1).

The next Figure 4 illustrates gathered in Table 1 average predicted values of the porosity coefficients and the volumetric deformation of the final deformation with the complete disintegration of microaggregates. The deformations associated with the formation of a new structure during the destruction of microaggregates in loess are larger than in some clays, despite the large particle sizes.

Table 1: Average predicted values of the porosity coefficient and volumetric deformation of silty-clay soils of different types and genesis.

Soil Name Indicator	Values				
	The porosity coefficient of soil in natural state e_0.	Predicted values of the porosity coefficient e and the volumetric deformation $\varepsilon = e_0 - e$.	Aggregate method of sample preparation, A	Dispersed method of sample preparation, D	Semi-dispersed method of sample preparation for analysis, S
Red-brown clay	0,55	e	0,198	0,200	0,197
		ε	0,352	0,350	0,353
Green-gray clay	0,857	e	0,239	0,234	0,235
		ε	0,618	0,623	0,622
Kryvyi Rih loess	0,805	e	0,372	0,380	0,378
		ε	0,381	0,373	0,375



a) b) c)

Figure 4: Predicted values of the porosity coefficient and volumetric deformation during the decomposition of rock microaggregates of various ages and origins: a) red-brown clay, b) gray-green clay, c) Kryvyi Rih loess. Denotations: A - aggregate preparation; S - semi-dispersed preparation; D - dispersed preparation; e - the coefficient of porosity; ε - the volumetric deformation.

4 CONCLUSIONS

In this article, we present the results of theoretical and experimental studies of the degradation process of clay soils in Ukraine. Rocks of different time of formation and origin were studied. We consider the degradation of clay soils as a consequence of technogenic influence, leading to irreversible destruction of the structure. The final state of the soil is considered, in which the decay of all aggregates occurred. As a result of the performed experiments and the application of new theoretical models, the following conclusions can be drawn:

- the difference in the predicted values of the porosity coefficient of the soils, calculated by the new method, confirms the influence of the type, genesis and age of the dispersed soil on the density of compaction during the formation of the “new” structure;
- the developed techniques can be considered as the basis for creating a new classification of silty-dispersed soils by the values of the fractal dimension of their particle size distribution;
- for silty loess and clay soils, after the disintegration of microaggregates, the "new" porous structure is completely formed by particles of the smallest size, the particles of bigger sizes are presented in the “new” structure as separated inclusions.

REFERENCES

[1] Z. Luo, D. Ning, Z. Li, and X. Tian, “Three-dimensional fully coupled study of groundwater seepage and soil deformation under the action of cyclic pumping and recharge,” Proc. IAHS, vol. 382, pp. 515-520, 2020, [Online]. Available: https://doi.org/10.5194/piahs-382-515-2020.

[2] H. N. Ngugi, S. M. Shitote, and N. Ambassah “Effect of Variation in Moisture Content on Soil Deformation and Differential Settlement of Frame Structures in Nairobi Area and its Environs,” The Open Construction & Building Technology Journal, vol. 15, pp.106-128, 2021, doi: 10.2174/1874836802115010106.

[3] J. Novotná and B. Badalíková, “The Soil Structure Changes under Varying Compost Dosage,” Agriculture, vol. 64, no. 4, pp. 143-148, 2018, doi: 10.2478/agri-2018-0015.

[4] Y. Saku, M. Aizawa, T. Ooi, and G. Ishigami, “Spatio-temporal prediction of soil deformation in bucket excavation using machine learning,” Advanced Robotics, vol. 35, no. 23, pp. 1404-1417, 2021.

[5] K.H.E. Meurer, C. Chenu, and et al., “Modelling dynamic interactions between soil structure and the storage and turnover of soil organic matter,” Biogeosciences, vol.17, pp. 5025-5042, 2020, doi: 10.5194/bg-17-5025-2020.

[6] Y. Zhang, X. Zhong, J. Lin, and et al., “Effects of fractal dimension and water content on the shear strength of red soil in the hilly granitic region of southern China,” Geomorphology, vol. 351, article ID 106956, 2020, doi: 10.1007/s12303-021-0037-z.

[7] D. Zhao, Y. Guo, G. Wang, X. Guan, X. Zhou, and J. Liu, “Fractal analysis and classification of pore structures of high-rank coal in Qinshui Basin, China,” Energies, vol. 15, article ID 6766, 2022, doi: 10.3390/en15186766.

[8] Y. Gao, M. Wang, Y. Li, L. Li, and Y. Chen, "Volume fractal and surface fractal analysis of the pore structure of natural and heat-treated spruce wood using the mercury intrusion porosimetry test," Holzforschung, vol. 76, no. 10, pp. 907-915, 2022, .

[9] H. OuYang, G. Dai, W. Qin, Ch. Zhang, W. Zhu, and W. Gong, “Experimental study on the mechanical behaviors and particle breakage characteristics of calcareous sand from South China Sea under repeated one-dimensional impacts,” Acta Geotechnica, vol. 17, pp. 3927-3946, 2022, doi: 10.3390/ma15134690.

[10] X. Li, W. Wei, L. Wang, and J. Cai, "Fractal dimension of digital 3D rock models with different pore structures," Energies, vol. 15, no. 20, article ID 7461, 2022, [Online]. Available: https://doi.org/10.3390/en15207461.

[11] A.R. Russell, “A compression line for soils with evolving particle and pore size distributions due to particle crushing,” Géotechnique Letters, vol. 1, pp. 5-9, 2011, [Online]. Available: http://dx.doi.org/10.1680/geolett.10.00003.

[12] J. Zhang and B. Zhang, “Fractal pattern of particle crushing of granular. advances in civil engineering”, vol. 2018, article ID 2153971, 2018, doi: 10.1155/2018/2153971.

[13] F. Seblany, U. Homberg, E. Vincens, P. Winkle, and K.J. Witt, “Merging criteria for defining pores and constrictions in numerical packing of spheres,” Granular Matter, vol. 20, article no. 37, 2018, doi: 10.1007/s10035-018-0808-z.

[14] D.A. Wood and B. Hazra, “Characterization of organic-rich shales for petroleum exploration & exploitation: a review-part 1: bulk properties, multi-scale geometry and gas adsorption,” J. Earth Sci., vol. 28, pp. 739-757, 2017, doi: 10.1007/s12583-017-0732-x.

[15] Z. Xue, X. Tang, and Q. Yang, “Influence of voltage and temperature on electro-osmosis experiments applied on marine clay,” Applied Clay Science, vol. 141, pp.13-22, 2017, doi: 10.1016/j.clay.2017.01.033.

[16] J.M. Espitia, B. Caicedo, and L. Vallejo, “Effect of suction and stress on Poisson's ratio of argillaceous rocks,” Géotechnique Letters, vol. 7, no. 1, pp. 53-59, 2017.

[17] E. Amezketa, “Soil aggregate stability: a review,” J. Sustainable Agricalture, vol. 14, no. 2-3, pp: 83-151, 1999.

[18] T.P. Mokritskaya, A.V. Tushev, E.V. Nikulchev, and K.A. Samoylich, “On the Fractal Characteristics of Loess Subsidence,” Contemporary Engineering Sciences, vol. 9, no. 2, pp. 799-807, 2019, [Online]. Available: http://dx.doi.org/10.12988/ces.2016.6687.

[19] T.P. Mokritskaya and A.V. Tushev, “On fractal-based estimations of soil subsidence,” in Basics of Computational Geophysics. Amsterdam: Elsevier, 2021, pp. 319-333.

[20] T.P. Mokritskaya, A.V. Tushev, K.A. Samoylich, and P.N. Baranov, “Deformations of loess soils caused by changes in the microaggregate structure,” Bulletin of Engineering Geology and the Environment, vol. 78, pp. 3729-3739, 2019, doi: 10.1007/s10064-018-1361-z.

[21] A.R. Russell, “Water retention characteristics of soils with double porosity,” Eur. J. Soil Sci., vol. 61, no. 3, pp. 412-424, 2010.

[22] G.R. McDowell, M.D. Bolton, and D. Robertson, “The fractal crushing of granular materials,” J. Mech. Phys. Solids, vol. 44, no. 12, pp. 2079-2101, 1996, [Online]. Available: https://doi.org/10.1016/S0022-5096(96)00058-0.

[23] M.R. Coop, K.K. Sorensen, K. K. Bodas Freitas, and G. Georgoutsos, “Particle breakage during shearing of a carbonate sand,” Géotechnique, vol. 54, no. 3, pp. 157-163, 2004, doi: 10.1680/GEOT.2004.54.3.157.

[24] I. Einav, “Breakage mechanics – part I: Theory,” J. Mech. Phys. Solids, vol. 55, no. 6, pp. 1274-1297, 2007, doi: 10.1016/j.jmps.2006.11.003.

[25] I. Einav, “Breakage mechanics – part II: Modelling,” J. Mech. Phys. Solids, vol. 55, no. 6, pp. 1298-1320, 2007, [Online]. Available: https://doi.org/10.1016/j.jmps.2006.11.004.

[26] R. Kong, F. Zhang, G. Wang, and J. Peng, “Stabilization of Loess Using Nano-SiO_2,” Materials, vol. 11, no. 6, article ID 1014, 2018, doi: 10.3390/ma11061014.

[27] W.T. Oh and S. Vanapalli, “Undrained shear strength of unsaturated soils under zero or low confining pressures in the vadose zone,” Vadose Zone J., vol. 17, no.1, pp. 1-13, 2018, doi: 10.2136/vzj2018.01.0024.

Modeling of Vegetable Oil Miscellaneous Drive Process in Final Distiller Spray Zone

Abduhoshim Qadirov[1], Ulugbek Mannanov[2], Anvar Khamdamov[1], Absalom Xudayberdiyev[1] and Nematilla Kurbanov[1]

[1]*Namangan Institute of Engineering and Technology, Kosonsoy Str., 7, Namangan, Uzbekistan*

[2]*Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan*

abdulhoshimqadirov@gmail.com, umannanov@gmail.com, kurbanovnm@gmail.com, anvarkhamdamov@rambler.ru, jarayon@rambler.ru

Keywords: Steam, Temperature, Distiller, Missella, Extraction , Oil, Nozzle, Drop, Liquid, Pressure, Time.

Abstract: This work discusses the kinetics of the separation process of a solution consisting of vegetable oil and a low-boiling hydrocarbon solvent in a distillation process. The efficiency of the distiller is affected by capital costs, energy costs, and quality characteristics of the resulting oil. The mass transfer rate depends on the nature of the phase equilibrium, mass transfer coefficient, and contact area between the phases. The issues of mathematical modeling of heat and substance exchange processes going in the spraying zone of the final distillation device of vegetable oil are presented. The influence of various factors such as initial temperature and concentration of the feed liquid, vacuum pressure, and particle diameter on the distillation process is also discussed. The process of scattering liquid in a channel sprayer is given to two zones A and B, that is, mathematical and computer models of the processes of heat and substance exchange, which are carried out by turning the liquid into a drop, and the results of the computer model are analysed.

1 INTRODUCTION

The kinetics of the separation process of a solution consisting of vegetable oil and a low-boiling hydrocarbon solvent depends on the mass transfer between the solution and the mixture of water and heated vapors of the solvent Distillers of various designs are used in oil extraction enterprises. Important indicators of distiller efficiency are capital costs, energy costs, especially costs associated with providing superheated steam to the final distiller for contact processing of mistella, and the quality characteristics of the resulting oil [1-3,17].

Capital costs in distillation equipment are roughly proportional to size and inversely proportional to throughput. The mass transfer rate depends on the nature of the phase equilibrium, is proportional to the mass transfer coefficient and the contact area between the phases. The increase in mass transfer occurs in conditions of high turbulence of one or both phases and at high speeds of their movement relative to each other [4-6, 10,11].

When a steam nozzle is used to spray the mistella in the final still, the phase contact surface is large and very high droplet flow velocities are obtained by co-flowing the superheated water vapor near the nozzle mouth. The intensity of interphase transfer can be adjusted by changing the temperature and flow rate of the water vapor supplied to the nozzle, while the flow rate of the spray liquid remains constant. Heating the oil in the still for a long time has a negative effect on its quality, but to evaporate the solvent from the surface of the droplet, it must be heated to some degree relative to the temperature corresponding to the conditions of phase equilibrium. It depends on the pressure in the evaporation zone and the concentration of the solution on the droplet surface [7,12].

The driving force of the transfer of substances from the droplet to the external vapor environment is the difference in the partial vapor pressures of the volatile component on the droplet surface and the one expelled from it. The larger the partial vapor pressure of the volatile component in the outer stream of vapor mixtures, the smaller the diffusion flux of the mass of the vaporizing component on the droplet surface. The higher the temperature and concentration of the evaporating component on the surface of the droplet, the greater its pressure near this surface. It is necessary to choose a distiller taking into account the

above factors for the technological flow of steam mixtures [8-9,16].

2 METHODS AND MATERIALS

We will consider the process of substance exchange under study conditionally divided into A and B zones.

Modeling the heat substance exchange process in zone A. Under the conditions under consideration, we accept the following system of equations to calculate the concentration process in the drop formation zone based on equation [13] and the mathematical expression characterizing the balance (1):

$$\begin{cases} G_{\text{р}} = k_a F(a_{\text{р}} - a) = G_0\left(1 - \frac{a_0}{a}\right) \\ \frac{G\frac{a_0}{a}\left(1-\frac{a_0}{a}\right)}{G_{\text{г}}} = \frac{M_{\text{р}}}{M_{\text{г}}}\frac{P_{\text{р}}}{P_{\text{ум}}-P_{\text{р}}} \\ P_{\text{р}} = f(t,a) \\ cadt + (i-ct)da - \frac{\alpha_V a}{G}(t_{\text{г}} - t) = 0 \end{cases} \quad (1)$$

where a - solvent concentration at a given time, %; $a_{\text{р}}$ - solvent equilibrium concentration, %; k_a - volume coefficient of heat transfer.

Results ($Q_k < \varepsilon$) of the analysis, $\alpha_V \approx 0$ it can be considered. The value of the volume coefficient of heat transfer $\left(k_{a,}\text{c}^{-1}\right)$ determines the efficiency of this zone.

According to the information obtained by Beloborodov V.V., the volume coefficient of heat transfer during the final distillation by spraying liquid for vegetable oil-extractive gasoline solution is 0.02 c^{-1}. Based on this, it is convenient to express the mass transfer equation through the volumes of the light volatile component that passes into the liquid and vapor phase. In this case, the drop formation zone is considered as an object with accumulated indicators (2):

$$V_{\text{р}} = k_V V \frac{(a_{\text{р}} - a)\tau_{\text{к}}}{100}, \quad (2)$$

where $V_{\text{р}}$ - the volume of the solvent passing into the V - vapor phase, is the total volume of the liquid phase, and $\tau_{\text{к}}$ - the time of droplet formation.

We can characterize volumes by mass costs and write:

$$\frac{G_{\text{р}}}{\rho_{\text{э}}} = k_V \frac{G_0\frac{a_0}{a}}{\rho}\frac{(a_{\text{р}} - a)\tau_{\text{к}}}{100}, \quad (3)$$

$\rho, \rho_{\text{э}}$ - light volatile component and solution density, respectively.

According to (3) for light volatile component and mistletoe costs

$$\frac{G\left(1-\frac{a_0}{a}\right)}{\rho_{\text{р}}} = k_V \frac{G_0\frac{a_0}{a}}{\rho}\frac{(a_{\text{р}} - a)\tau_{\text{к}}}{100}. \quad (4)$$

After some mathematical transformations, we write the (5) that characterizes the concentration of the mistletoe at the exit of the droplet formation zone:

$$a^* = \frac{100a_0 + k\frac{\rho_{\text{р}}}{\rho}a_{\text{р}}a_{\text{к}}\tau_{\text{к}}}{100 + k\frac{\rho_{\text{р}}}{\rho}a_0\tau_{\text{к}}}. \quad (5)$$

When calculating the currents in this zone, the values of the following indicators were obtained: according to the data of [14], the time of droplet formation is a few microseconds, so we take it to be equal to 0.03 s;

The ratio of extraction gasoline to 95% mistella density is about 7:9; multiplying it by the average time of droplet formation, we get a value of 0.023; we accept the volumetric coefficient of mass transfer in the range of 1-10. Here, only the liquid phase is assumed, not the full size of the droplet formation zone.

The studies carried out in the K model show that the intensity of the mistella concentration process by spraying in the A zone is significantly affected by the volumetric coefficient of substance delivery. In our example, the coefficient of mass transfer is 4 times larger than in [15]. The influence of temperature on the studied process is not very significant, but it affects the coefficient of mass transfer, and as such, it should also affect the studied process.

Zone B. The droplets formed in zone A fall down and concentrate by evaporating the solvent. In accordance with its hierarchical structure, the following events affect the mistella concentration process: evaporation of the solvent from the droplet; cooling of the drop due to evaporation; heating of the droplet with the vapor phase; molecular-diffusion transfer of the solvent in the droplet; enrichment of the vapor phase with solvent vapors; vapor phase temperature change.

The mathematical model of the process in this zone can be written in the form of the following system of (6):

$$\begin{cases} a_j = a_{j-1} + \left(a_{\text{р}j-1} - a_{j-1}\right)\frac{\pi D}{4R^2}\Delta\tau \\ \frac{G_0}{G_{\text{г}}}\left(\frac{a_0}{a_j} - \frac{a_0}{a_{\text{к}}}\right) = \frac{M_{\text{р}}}{M_{\text{г}}}\frac{\text{P}_{\text{р}}}{\text{P}_{\text{ум}}-\text{P}_{\text{р}}} \\ P = (b_{11}t_{\text{м}} - b_{10})(100 - a_{\text{р}}) \\ cadt + (i-ct)da = 0 \end{cases} \quad (6)$$

a_{j-1}, a_j - mistella $j$concentration at the entrance and exit of the site, respectively; b_{10}, b_{11} - constant coefficients for the seed-gasoline mist (at a temperature of 80-100°C and a concentration of 90-100% $b_{10} = 3{,}470\ kPa/°C, b_{11} = 0{,}064\ kPa/°C$,%); R - drop radius, m; D - liquid diffusion coefficient, m^2/s; τ-concentration time, s.

Calculations are carried out in the following sequence:

1) Finding distribution of mistletoe concentration in different diameters of the drop depending on the time of the drop;
2) Determination of spray mistletoe concentration versus time at different initial liquid temperatures
3) Calculation of mistletoe change over time at different initial liquid concentrations
4) Determination of the effect on the ratio of open water vapor and mistletoe consumption during spray concentration for different initial liquid concentrations
5) Show the change in mistletoe concentration over time at different values of pressure (40 and 100 kPa).

The results of the research conducted in the mathematical model for different diameters of the drop are presented in the form of graphs (1-2), which indicate that the concentration efficiency decreases when the diameter of the drop increases. So when the particle diameter is 10 µm, distillation actually takes place in one tenth of a hundredth of a second. In production, distillation takes place in a few seconds when the particle diameter obtained by spraying is 80 µm in the final distillation equipment.

3 RESULTS AND DISCUSSION

The average residence time of a particle in the spray zone of the distiller is approximately 1-2.5 s. The distillation process suddenly deteriorates when the particle diameter is 0.1 mm. Curve 4 in Figure 1 corresponds to the concentration of a droplet composed of a mixture of particles larger than the indicated diameter in a uniform mass distribution. It can be seen that at the beginning of the distillation process, the relationship is characterized by a relatively large curvature, which deteriorates significantly in the later stages of the process. For the situation under consideration, it can be assumed that there are no conditions for obtaining oil of standard requirements in a typical device. This indicates that it is necessary to pay serious attention to the problem of obtaining a finely dispersed phase with a maximum diameter of the sprayed particle not exceeding 100 µm . That's why we channel the spray nozzle and we have proposed a new effective construction.

The calculation results of the influence of the initial temperature of the liquid sprayed on the distillation process are presented in Figure 2. Calculations show that with a decrease in temperature, the intensity of the distillation process also decreases, but to bring the mistella concentration to the standard requirements, the initial temperature of the feed liquid is considered to be 90 °C.

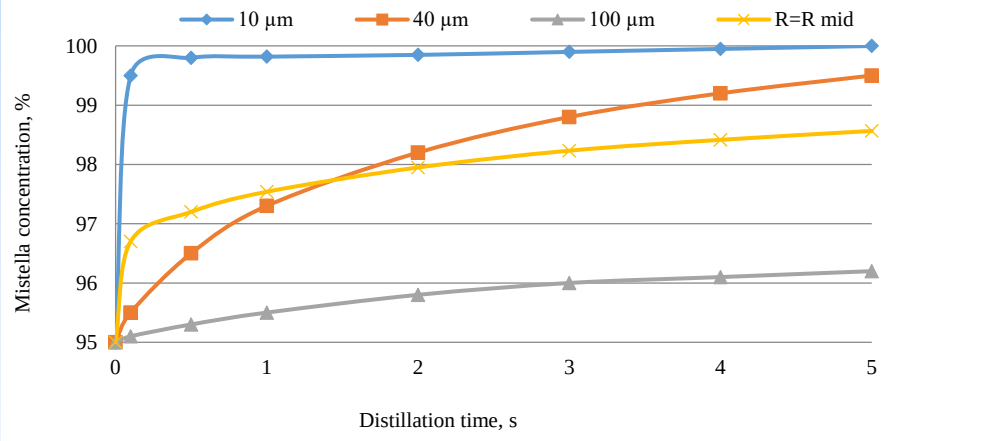


Figure 1: Variation of the mistletoe concentration depending on the drop time at different diameters of the droplet.

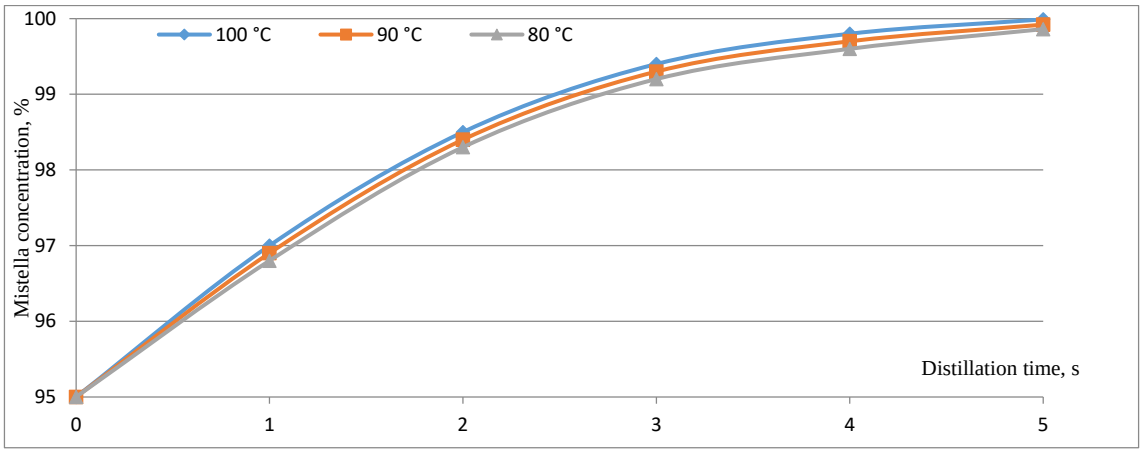


Figure 2: Changes in the concentration of micelles depending on the time of descent at different initial temperatures of the liquid.

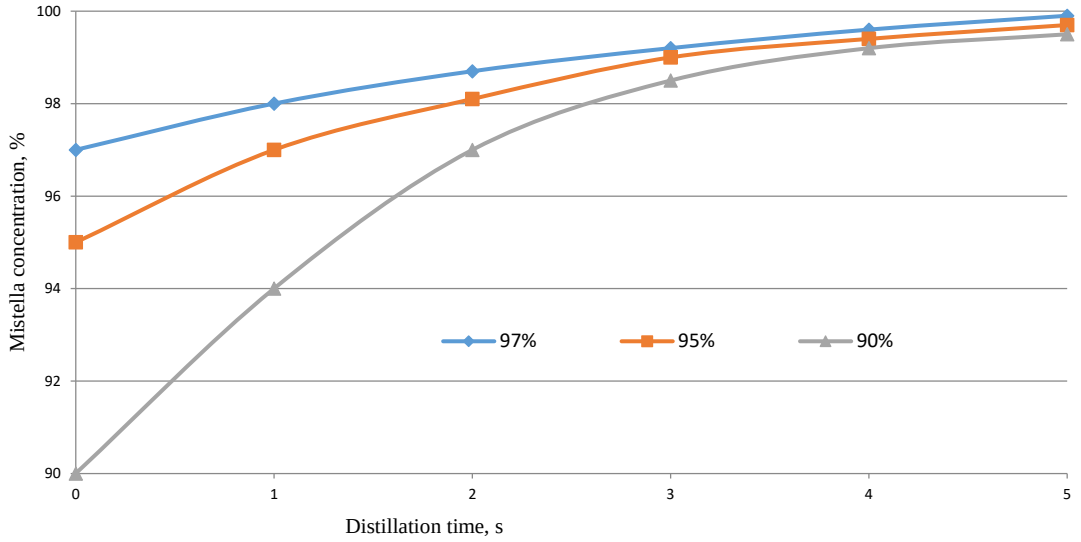


Figure 3: Changes in the concentration of mistella as a function of the time of deposition at different initial concentrations of the liquid.

Depending on the ratio of water vapor and mistletoe costs, the theoretical distribution of concentration will have its critical cost ratio for a certain value of concentration. It will be impossible to get standard oil . In practice, the consumption of water vapor should be much higher due to the uneven distribution of phases in the conditions of a complex hydrodynamic structure.

The influence of the initial concentration of the feed liquid on the distillation process is illustrated in Figure 3. It can be seen that at the end of the process, the curves become much closer to each other, but do not merge. The difference is very noticeable, due to the slight increase in concentration at the back of the curve due to more time in the vapor phase than the liquid. This fact confirms that as the concentration of mistella in the oil increases, it becomes more difficult to drive the solvent.

When look at the Figure 4 using a vacuum of 40 kPa in the device, compared to atmospheric pressure, the process is significantly intensified.

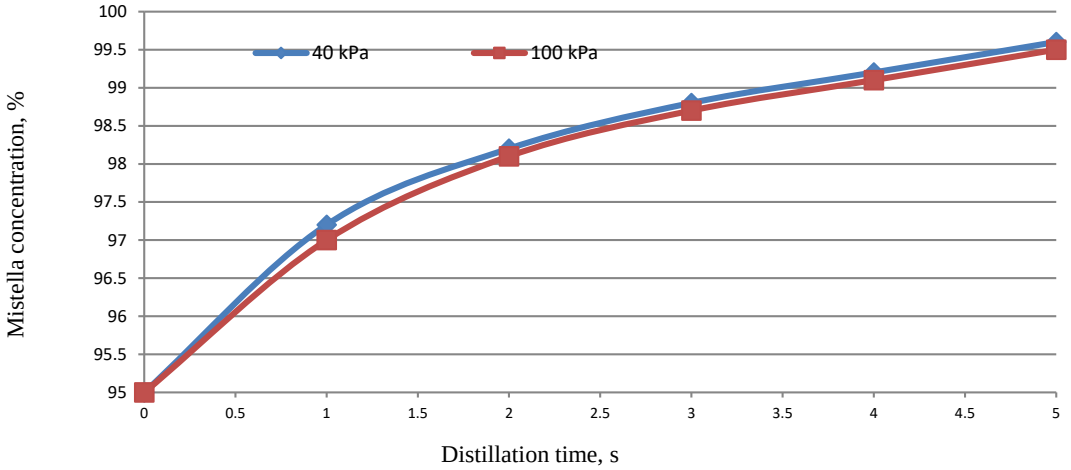


Figure 4: Using a vacuum of 40 kPa in the device.

4 CONCLUSION

A complete mathematical model of the heat substance exchange process in the spray zone of the still. Shown in the image above. The upper block represents the solution concentration zone during droplet formation. Here, a mistletoe with initial parameters is G_0, a_0, t_0concentrated to a state with intermediate parameters . G_0^*, a_0^*, t_0^*As mentioned above, the appropriateness of the mathematical model of this zone depends on the initial temperature of the liquid. This zone is considered as an object of an ideal mixing model from the point of view of the hydrodynamic structure of a concentric flow.

REFERENCES

[1] R. W. Faidley and R. L. Panton, “Measurement of liquid jet instability induced by surface tension variations,” Exp. Therm. Fluid Sci., vol. 3, pp. 383-387, 1990.

[2] E. B. Saitov and T. B. Sodiqov, “Modeling an Autonomous Photovoltaic System in the Matlab Simulink Software Environment,” in AIP Conference Proceedings, 2022, vol. 2432, p. 020022.

[3] E. B. Saitov, Sh. Kodirov, B. M. Kamanov, N. Imomkulov, and I. Kudenov, “Increasing the Efficiency of Autonomous Solar Photovoltaic Installations for Power Supply of Agricultural Consumers,” in AIP Conference Proceedings, 2022, vol. 2432, p. 040036.

[4] F. Zikrillayev, E. B. Saitov, J. B. Toshov, B. K. Ilyasov, and M. B. Zubaydullayev, “A Software Package for Determining the Optimal Composition and Parameters of a Combined Autonomous Power Supply System Based on Renewable Energy Sources,” in AIP Conference Proceedings, 2022, vol. 2432, p. 020021.

[5] N. F. Zikrillayev, K. C. Ayupov, S. A. Valiyev, S. B. Shergoziyev, and G. X. Mavlonov, “Functional possibilities of strong compensated leggier Si with the different deep energetic levels impurities in electronics,” J. Adv. Res. Dyn. Control Syst., vol. 12, no. 7, Special Issue, pp. 2751–2758, 2020.

[6] A. M. Khamdamov, S. H. Sultonov, and S. A. Bozorov, “The main results of the study of the processes of deodorization of vegetable oils using wooden nozzles,” J. Pharm. Negat. Results, vol. 13, no. S06, pp. 3844-3851, Dec. 2022, doi: 10.47750/pnr.2022.13.S06.512.

[7] N. R. Yusupbekov, H. S. Nurmuhamedov, and S. G. Zokirov, “Kimyoviy texnologiya asosiy jarayon va qurilmalari,” T.: Fan va texnologiya, 2015, pp. 1-848.

[8] W. A. Sirignano and C. Mehring, “Review of theory of distortion and disintegration of liquid streams,” Prog. Energy Combust. Sci., vol. 26, pp. 609-655, 2000.

[9] A. G. Laptiev, M. I. Farakhov, and N. G. Mineev, “Osnovy rascheta i modernizaciya teplomassoobmennyh ustanovok v neftehimii,” Kazan, Russia: Kazan. gos. energet. un-t, 2010, 574 p.

[10] Y. Qodirov and M. Raximov, “Yoglarni qayta ishlash texnologiyasi,” Darslik: “Iqtisod-Moliya”, 2013, pp. 300.

[11] S.V. Butova, I.A. Sorokina, N.V. Korolkova, and M.N. Shaxova, “Rascheti oborudovaniya maslojirovoy promыshlennosti: uchebnoye posobiye”, Voronej: FGBOU VO Voronejskiy GAU, 2017, pp. 151.

[12] G.F. Vasileva, V191 Dezodoratsiya masel i jirov (Predisloviye d.t.n. N. S .Arutyunyana), SPb: GIORD, 2000, pp. 192.

[13] A.M. Hamdamov and D.A. Igamberdieva, “Matematicheskoe modelirovanie ravnovesnogo sostoyaniya ekstrakcionnogo benzina i zhirnyh kislot,” Science Time, no. 4 (40), 2017, pp. 209-213.

[14] A.M. Hamdamov and D.A. Igamberdieva, “Mathematical modeling of the process of final distillation of a multi-stage rotary-disk apparatus,” Scientific Bulletin of Namangan State University, vol. 2, no. 3, pp. 128-135, 2020.

[15] A.M. Khamdamov, O.K. Askarova, O.A. Mansurov, and S.H. Sultonov, “Simulation of a Multistage Distillation Process in a Rotary Disc Device,” Annals of the Romanian Society for Cell Biology, Apr. 2021, pp. 5939-5948.

[16] D.R. Chen, D.Y. H. Pui, and S.L. Kaufman, "Electrospraying of conduction liquids for monodisperse aerosol generation in 4 nm to 1,8 μm diameter range," J. Aerosol Sci., vol. 26, pp. 963-977, 1995.

[17] V. Chernyak, “The kinetic theory of droplet evaporation,” Journal of Aerosol Science, vol. 26, no. 6, pp. 873-885, Sept. 1995.

Effects of Temperature on the Efficiency of Photovoltaic Modules

Islom Jurayev, Isroil Yuldoshev and Zukhra Jurayeva
Tashkent State Technical University named after I. Karimov, Universitet Str. 2, Tashkent, Uzbekistan
nauka-jir@mail.ru, yuldashev.i2004@gmail.com, zuxra_199@mail.ru

Keywords: Photovoltaic Module, Heating Temperature, Solar Radiation, Power, Efficiency.

Abstract: This article presents calculations of the efficiency and power of modules of various technologies using mathematical expressions and equations. In the calculations, we used experimentally measured values of solar radiation, module temperature and passport values of some parameters obtained by module manufacturers under Standard Testing Conditions (STC). The results of analytical calculations and comparisons are presented in graphical and tabular forms. The object of the study, were the following types of photovoltaic modules: silicon monocrystalline (mc-Si), silicon polycrystalline (pc-Si), thin-film based on copper, indium, gallium and selenide (CIGS), thin-film based on cadmium telluride (CdTe). In this work is determined the influence of the temperature factor, external parameters on the energy performance and efficiency of photovoltaic modules. At the maximum module temperature of 85°C, the power losses from nominal power value of the modules, respectively, had the following percentages: pc-Si, mc-Si - 33%, CdTe – 20%, CIGS– 29%. According to the calculations the decrease in the efficiency of modules under the influence of heating temperature was 2-2.4% in silicon modules, 0.6-0.7% in thin-film modules.

1 INTRODUCTION

At present, renewable energy is more widely implemented in the energy supply system of consumers. In particular, to compensate for the energy deficit in the energy system during peak hours of energy consumption, to provide a backup power supply system, to provide individual consumers located far from the centralized power supply system and for a number of other consumers. One of the promising technologies of renewable energy, in particular solar energy, is photovoltaic modules (PVM) for converting solar energy into electrical energy. The performance and efficiency of the PVM is directly related to environmental factors, the movement of the sun, the geographical latitude of the area, the angle of inclination and orientation of the modules, the design parameters and manufacturing technology of the modules and other factors [1], [2].

During the operation of the PVM, in the order of 80-85% of solar radiation is absorbed by the module, of which only 10-15% of solar energy at the point of maximum power is converted into electrical energy, the rest is converted into heat [3]. In hot operating conditions, the modules are heated above the normal operating temperature under the influence of solar radiation. As a result of overheating, the efficiency of the modules decreases [4]. A decrease in efficiency leads to an increase in the average price of energy production. In commercial use, photovoltaic modules based on crystalline silicon are mainly used, the cost of production of which is high. In recent years have been developed, modules based on thin-film technology, such as amorphous silicon, cadmium telluride, copper-indium-gallium-selenide and a number of others. Due to the lower consumption of materials, thin-film modules are cheaper than modules based on crystalline silicon [5]. The influence of environmental parameters on the parameters of the modules was investigated in many researches [6], [7], [8], [9], [10]. And the influence of the heating temperature on the characteristics of the modules has been studied by follow works [11], [12], [13], [14]. In these works are determined, the correlations of the energy and electrical parameters of the modules with the heating temperature and external parameters.

In the geographical conditions of Tashkent of the Republic of Uzbekistan, we have studied the PVM of various manufacturing technologies in the summer season. In this paper, the following goals were set for the studied types of PVM, depending on the temperature factor:

- calculation of efficiency and power values by mathematical expressions and equations;
- calculation of power losses and efficiency;
- presentation of the results of these calculations in graphical forms;
- comparison of calculation results in tabular forms.

For achieve these goals, we used the passport values of some parameters of PVM measured by manufacturers under STC testing conditions.

2 PVM TESTING CONDITIONS

As a rule, PVM manufacturers in the module passport give the characteristics of the modules obtained by testing in laboratory conditions, according to the relevant testing standards. The regulations for testing PVM in various climatic and geographical conditions were published by the International Electrotechnical Commission in the IEC 61853 standard "Photovoltaic Module. Power Rating", which provides HTC, LIC, HTC, NOCT and STC tests. There is also the PV-USA Test Condition (PTC) standard, which is not included in the IEC standard. Table 1 shows the types and parameters of testing conducted by PVM manufacturers [15].

Table 1: Types and parameters of testing of modules.

Para-meter	STC	NOCT	PTC	HTC	LIC	LTC
Illumi-nation W/m^2	1000	800	1000	1000	200	500
The tempe-rature of the solar cell,ºC	25	43-50	-	75	25	15
Ambi-ent tempe-rature, ºC	-	20	20	-	-	-
Wind speed, m/s	-	1	1	0	0	0
Height above groud level,m	-	-	10	-	-	-
AM	1.5					

As can be seen from Table 1, environmental parameters have different values in all PVM testing conditions. But, in real operating conditions of the PVM, the values of these parameters vary depending on the seasons of the year.

3 MATERIALS AND METHODS

In this paper, in order to achieve these goals, we used a method for calculating the efficiency and energy parameters of the studied types of PVM using mathematical expressions and equations. For the calculations were used measured values of solar radiation and PVM temperature in accordance with the experimental studies previously conducted by us. Experimental studies of PVM were conducted in the conditions of the summer season on the heliopolygon of the Department of "AES" of the Tashkent STU named after I.Karimov. To carry out measurements of parameters, these types of PVM were installed at an angle of inclination of 45° to the horizon (Figure 1).

Figure 1: Photos of various types of PVM.

For the study was used the following types of the photovoltaic modules:

- Monocrystalline silicon PVM (mc-Si), Sky(AR) 290W;
- Polycrystalline silicon PVM (pc-Si), ODA50-18-P 50W;
- Thin-film PVM based on copper, indium, gallium and selenide (CIGS), SC-50MDF 48W;
- Thin-film PVM based on cadmium telluride (CdTe), ASP-S1-90.

In the passport data of the studied types of photovoltaic modules, according to Table 2, are given the parameter values obtained by the module manufacturers according to Standard Test Conditions (STC), at a solar radiation flux density of 1000 W/m2, ambient temperature T = 25°C, AM 1.5).

Table 2: Passport data of photovoltaic modules.

PVM Characteris-tics	Types of PVM			
	CdTe	pc-Si	mc-Si	CIGS
Module sizes, mm	1200x 600	670x 530	960x 1620	1075x 610
Area of modules, S_m, m^2	0.72	0.35	1.56	0.65
Electrical characteristics				
Nominal power, P_n, W	90	50	290	48
Open-circuit voltage, U_{oc},V	122	22,14	39,8	24,5
Short-circuit current, I_{sc}, A	1.06	2.89	9.6	2.7
Voltage at max power, U_{max}, V	96	18.5	32.2	19.8
Current at max power, I_{max}, A	0.94	2.7	9.1	2.52
Temperature coefficients:				
I_{sc}, α, %/^{0}C	0.060	0.065	0.033	0.008
U_{oc}, β, %/^{0}C	-0.321	-0.380	-0.360	-0.280
P_n, γ, %/^{0}C	-0.214	-0.450	-0.440	-0.380

4 RESULTS AND DISCUSSIONS

Experimental measurements were previously carried out by us in the month of July of the summer season of 2021. In work [16], were compared only the values of environmental parameters, temperature, short-circuit current and open circuit voltage of various types of PVM measured in July and August of the summer season. The difference between this work and the previous study in the new work is the carrying out of computational calculations of the values of efficiency, power, their reduction and loss of values under the influence of the temperature factor for the studied types of PVM. These calculations were carried out using mathematical expressions and equations depending on the temperature and the amount of solar radiation. To calculate the set of efficiency and power values, we used the measured values of solar radiation and module temperature from the experiments conducted in July. And also, in this work, during the hours of the maximum arrival of solar radiation, we compare the low, high and maximum values of the surface temperature of the modules. A comparison of the measured values is given in Table 3.

Table 3: Comparison of temperature values of different types of PVM.

Parameters		Measurement time 13-00			
		Types of PVM			
		pcSi	CdTe	mcSi	CIGS
W, w/m^2	Solar radiation	916	940	940	915
T_h◦C	The high value of tempera-ture on the surface of the module	70,3	71,0	67,6	77,9
T_{hts}◦C	Tempera-ture at module hotspots	77,8	76,6	79,6	81,1
T_l◦C	The low value of tempera-ture of the module	64,1	71,1	65	69,1
Ta◦C	Air tempera-ture	44,8	44,8	44,8	44,8
T_{hs}/Ta, times	Specific tempera-ture rise of the module	1,74	1,71	1,78	1,81

From the comparisons, it can be seen that the temperature increase of the module relative to the air temperature at 13-00 o'clock in the afternoon is 71-81%, and the average daily temperature increase is 50-70%.

The dynamics of the heating temperature of the modules of four types according to the measured values during the day is shown as a graph in Figure 2.

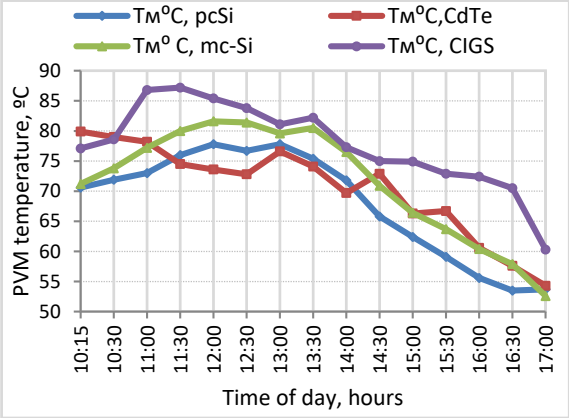


Figure 2: Dynamics of module heating temperature during the day.

As you can see from the graph, the modules have different temperatures. This is due to the different properties of the module materials and their convective heat exchange with the environment.
The efficiency of PVM, as given in [17], may be calculated using (1)

$$\eta_{stc} = \frac{P_n}{W_{stc} \cdot S_m}, \qquad (1)$$

where, P_n-nominal power of PVM, W_{stc}-solar radiation in STC, S_m-area of PVM. By using (1) will calculate the efficiency of the modules according to the values of parameters as Table 2.

$$\text{CdTe, } \eta_{stc} = \frac{90}{1000 \cdot 0.72} \cdot 100\% = 12{,}5\%$$

$$\text{pc-Si, } \eta_{stc} = \frac{50}{1000 \cdot 0.35} \cdot 100\% = 14{,}1\%$$

$$\text{mc-Si, } \eta_{stc} = \frac{290}{1000 \cdot 1.56} \cdot 100\% = 18{,}6\%$$

$$\text{CIGS, } \eta_{stc} = \frac{48}{1000 \cdot 0.65} \cdot 100\% = 7{,}4\%,$$

As mentioned above, the efficiency of the module depends on the temperature and temperature coefficient of the module. The efficiency of the module depending on the temperature factor as given in [18] may be calculated using (2),

$$\eta_{\text{м}} = \eta_{stc} \cdot [1 - \gamma(\text{T}_{\text{м}} - \text{T}_{stc})] + \varphi \cdot Log_{10} W \qquad (2)$$

here, φ is the coefficient of solar radiation. In most sources, the radiation coefficient is assumed to be zero. Then (2) takes a simplified form,

$$\eta_{\text{м}} = \eta_{stc} \cdot [1 - \gamma(\text{T}_{\text{м}} - \text{T}_{stc})], \qquad (3)$$

where, η_{stc} is the nominal efficiency (efficiency) of the module according to STC; T_M at the temperature of the module 25 °C and at W of solar radiation equal to 1000 W/m^2; γ - is the temperature coefficient of the module according to power, T_{stc} is the temperature of the module according to STC.

The values of the listed parameters for calculations are taken from Table 2. T_M is the temperature of the module, ^{0}C, experimentally measured data were used in the calculations. According to the (3), the values of the η_M efficiency of modules of various types tested in environmental conditions were calculated according to the experimentally measured values of the T_M temperature of the modules during the daytime from 10-15 to 17-00 hours. Experimental measurements were carried out on July 26, 2021. When calculating the value of η_M according to the (3), γ is the temperature coefficient for power, while it was assumed that during the day the value of this coefficient remains unchanged, which was taken according to the manufacturer's passport of each module according to Table 2. The values of η_M were calculated according to the (3), in depending on the heating temperature of the modules. In Figure 3 and Figure 4 constructed efficiency graphs.

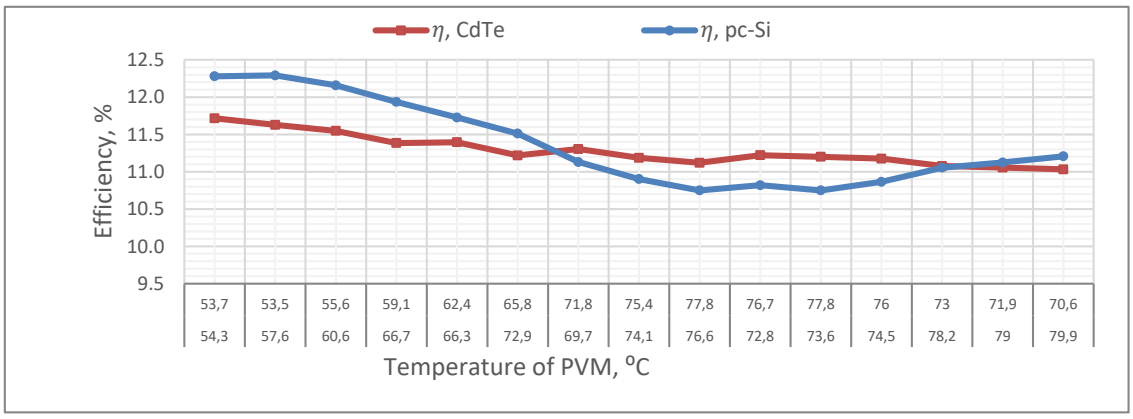


Figure 3: Graph of the dependence of the efficiency of CdTe and pc-Si types of PVM on its heating temperature.

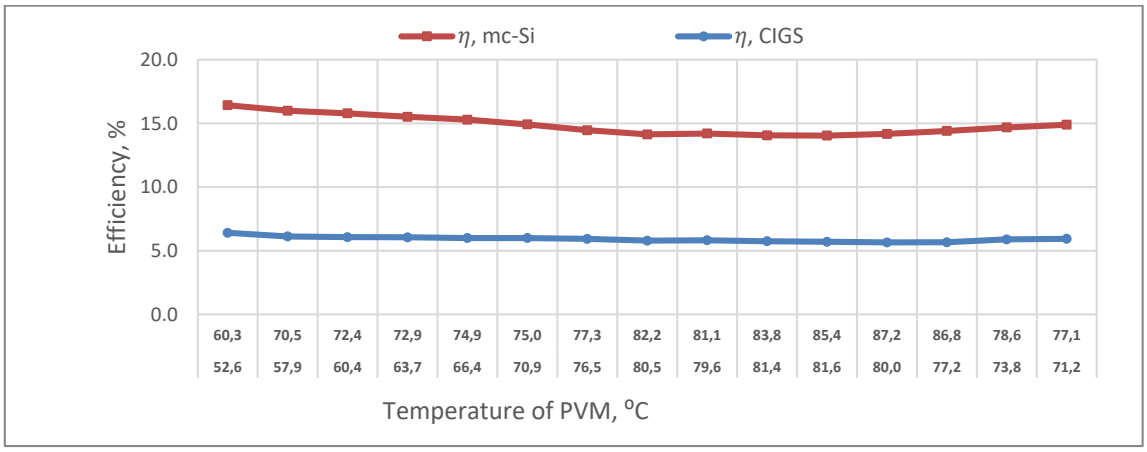


Figure 4: Graph of the dependence of the efficiency of mc-Si and CIGS types of PVM on its heating temperature.

As can be seen from Figure 3 and Figure 4, with an increase in the temperature of the module, the efficiency decreases. The analysis of graphical values shows that the decrease in the efficiency of modules during the daytime of measurement is the following indicators.

pc-Si–13 %, CdTe–6 %, mc-Si –14 %, CIGS –11 %,

The relatively low reduction turned out to be for a CdTe type photovoltaic module and was almost half as much as the other modules. A comparison of the change in efficiency with an increase in temperature is given in Table 4.

Calculations of efficiency indicators and temperature changes based on experimental measurements show that in a hot climate, the efficiency loss or temperature coefficient of efficiency in silicon crystal modules such as pc-Si, mc-Si is on average 36.5% higher relative to thin-film technology modules such as CdTe, CIGS. According to the analytical data, the effect of increasing the temperature of the module was relatively stable in the PVM of thin-film technology of the CdTe type.

The power P_m of the module, as given in [17], [18] may be calculate using (4),

$$P_m = \eta_m \cdot W \cdot S_m \qquad (4)$$

or substituting (3) instead of η_m in (4) get the following dependence,

$$P_м = \eta_{stc} \cdot [1 - \gamma(T_м - T_{stc})] \cdot W \cdot S_м . \qquad (5)$$

Here, η_m is the efficiency of the module calculated by (3). W is the solar radiation flux density, w/m^2; S_m is the module area, m^2, the values according to Table 2 were taken for calculation. To calculate the power of the modules according to the (5), the values of solar radiation measured in July were used. A graph of the change in the value of solar radiation falling on the surface of various types of PVM during the day from 10-15 to 17-00 is shown in Figure 5.

Table 4: Comparison of the effect of temperature on the efficiency of PVM.

Parameters	Types of PVM			
	pc-Si	CdTe	mc-Si	CIGS
Min temperature of PVM, $T_{min,}$ ^{0}C	53,7	54,3	52,6	60,3
Max temperature of PVM, $T_{max,}$ ^{0}C	77,8	79,9	81,6	87,2
Increasing of temperature of PVM, ΔT_0,%	24,1	25,6	29	26,9
Changing pf temperature of PVM relative to T_{stc}, 25^{0}C,(T_m-T_{stc})	52,8	54,9	56,6	62,2
Decreasing of efficiency,$\Delta\eta_m$,%	-2,0	-0,6	-2,4	-0,7
Decreasing of efficiency relative to η_{stc}	-3,4	-1,4	-4,6	-1,7
$\Delta\eta_m/\Delta T_{stc}$, temperature coefficient relative to T_{stc}, %/^{0}C	-0,064	-0,026	-0,081	-0,027

According to the (5), we calculated the power values for the four tested module types. The changes in the power values are given in the form of graphs in Figure 6 and Figure 7. The graph of the power changes during the daytime of the measurements change according to the polynomial law and in proportion to the density of the solar radiation flux.

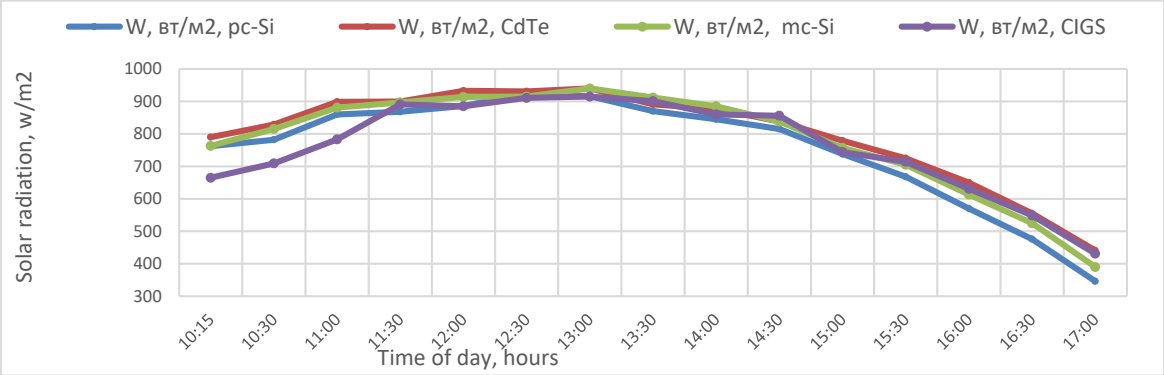


Figure 5: Graph of changes in solar radiation falling on various types of PVM during the day.

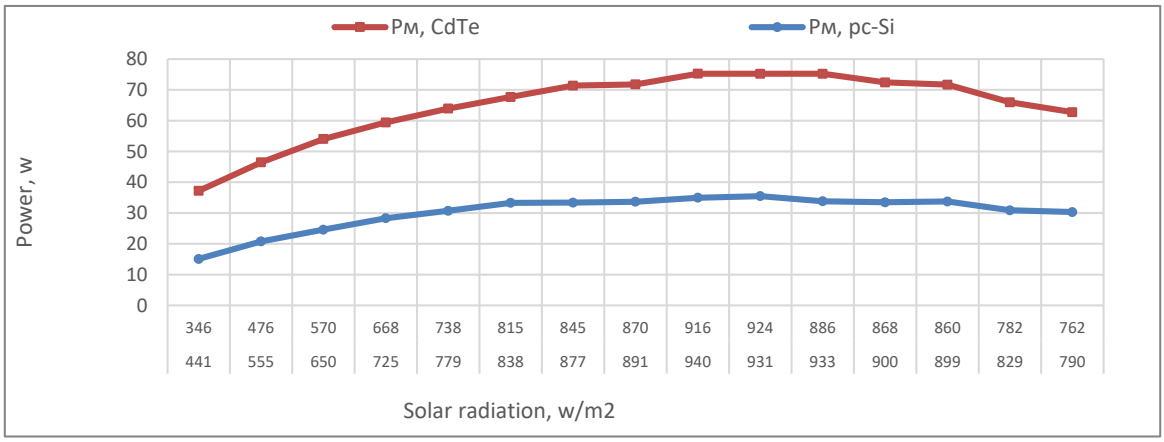


Figure 6: Graph of the dependence of CdTe and pc-Si types PVM power on solar radiation measured during the day.

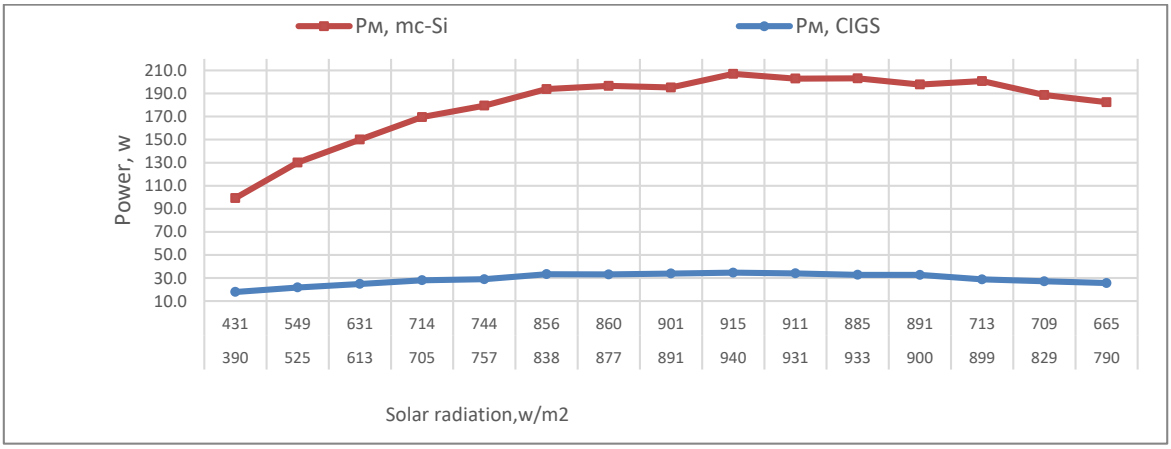


Figure 7: Graph of the dependence of the power of the mc-Si and CIGS type PVM on solar radiation measured during the day.

From the analysis of the graphs of power changes depending on the density of the solar radiation flux, the following values can be calculated.

The increase in the *W* - density of the solar radiation flux on the frontal surface of each module from 10-15 hours to 13-00 hours was, respectively, the following values, pc-Si – 1.7 times, CdTe – 2.1 times mc-Si – 2.4 times, CIGS – 2.1 times.

Accordingly, the increase in power proportional to solar radiation by modules is the following values, pc-Si – 2.2 times, CdTe – 2.1 times, mc-Si – 2.1 times, CIGS – 1.9 times.

The data show small deviations between the modules in the reception of solar radiation to the frontal surface and a corresponding increase in their power.

The effect of temperature increase on the output power of photovoltaic modules for various materials and manufacturing technology was given in [19]. Based on the results of experimental study of various types of PVM, it is possible to construct graphs of power loss depending on the temperature of the module (Figure 8). In this case, the power values were calculated by (4), depending on the heating temperature of the modules.

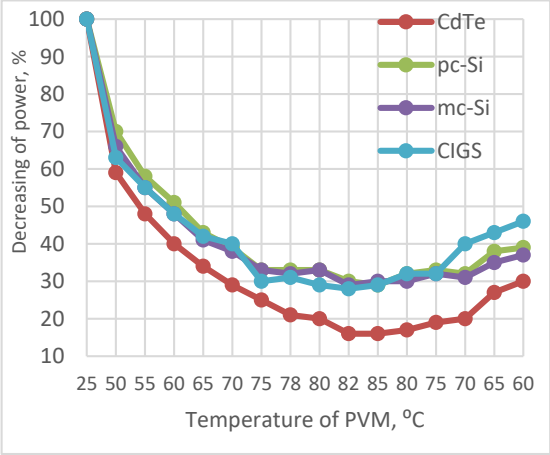


Figure 8: Graph of the dependence of the power of various types of PVM on their heating temperature.

Graphs of changes in power losses depending on the temperature of the modules vary according to the polynomial law. From the analysis of the data in Figure 8, based on the measured temperature values and calculations, the power losses of the PVM relative to the nominal power value of the module amounted to the following values.

At the maximum temperature of the module 85^0C, power loss, respectively made up of the following values, pc-Si-33%, mc-Si - 33%, CdTe – 20%, CIGS – 29%. These data show that the relatively low power sensitivity to temperature rise was for a CdTe type module. This can be explained with the optimal design and good convective heat exchange of this thin-film module.

5 CONCLUSIONS

In accordance with the goals of this work, the values of efficiency and power of various types of PVM were calculated using mathematical expressions and equations. Analysis of the calculated values, presented graphs and tabular comparisons of parameters shows that the influence of the temperature factor on the values of changes in efficiency, power, as well as the loss of these parameters of photovoltaic modules of various types in natural environmental conditions is significant. In hot climates, photovoltaic modules based on thin-film technologies have relative stability to heating compared to crystalline silicon modules. The magnitude of the decrease in the performance of photovoltaic modules primarily depends on the geographical conditions of the area. In our research of various types of modules in the environment conditions of Tashkent, photovoltaic modules based on thin-film technologies had relative stability to heating compared to crystalline silicon modules. When studying these types of modules in the environment of other geographical areas, the efficiency and energy indicators may change in a different ratio in contrast to those presented in this study.

ACKNOWLEDGMENTS

We express our gratitude to the head of the department of “AES” TashSTU named after Islam Karimov, Doctor of Technical Sciences I.A. Yuldoshev for assistance in organizing experiments.

The work was carried out with the financial support of the Ministry of Innovative Development of the Republic of Uzbekistan within the framework of the project F-OT-2021-497 “Development of scientific foundations for the creation of solar cogeneration plants based on photovoltaic thermal batteries”.

REFERENCES

[1] R. Schmalensee and V. Bulovic. The future of Solar Energy. An interdisciplinary MIT study. Massachusetts Institute of Technology (MIT), 2015, ISBN 978-0-928008-9-8.

[2] G. N. Tiwari and S. Dubey. Fundamentals of Photovoltaic Modules and Their Applications. Centre for Energy Studies, Indian Institute of Technology (IIT) Delhi, New Delhi, India, 2010, 423 p., ISBN: 978 1 84973 020 4.

[3] U. Sajjad, M. Amer, H.M. Ali, A. Dahiya, and N. Abbas,"Cost Effective Cooling of Photovoltaic Modules to Improve Efficiency," Case Studies in Thermal Engineering, vol. 14, Article ID: 100420. [Online]. Available: https://doi.org/10.1016/j.csite. 2019.100420.

[4] S. Chander, A. Purohit, A. Sharma, Arvind, S.P. Nehra, and M.S. Dhaka, "A Study on Photovoltaic Parameters of Mono-Crystalline Silicon Solar Cell with Cell Temperature," Energy Reports, vol. 1, pp. 104-109. [Online]. Available: https://doi.org/10.1016/j.egyr. 2015.03.00.

[5] S. Amar, M. Bahich, Y. Bentahar, E. Afifi, and E. Barj, "A Study of the Temperature Influence on Different Parameters of Mono-Crystalline Silicon Photovoltaic Module," Journal of Power and Energy Engineering, vol. 9, pp. 29-42. [Online]. Available: https://doi.org/10.4236/jpee.2021.96003.

[6] M. Mehdi, N. Ammari, A. Merrouni, H.E.Gallassi, M. Dahmani, and A. Ghennioui, "An experimental comparative analysis of different PV technologies performance including the influence of hot-arid climatic parameters: Toward a realistic yield assessment for desert locations," Renewable Energy, vol. 205, 2023, pp. 695-716. [Online]. Available: https://doi.org/10.1016/j.renene.2023. 01.082.

[7] M. Yaghoubirad, N. Azizi, A. Ahmadi, Z. Zarei, and S.F. Moosavian, "Performance assessment of a solar PV module for different climate classifications based on energy, exergy, economic and environmental parameters," Energy Reports, vol. 8, 2022, pp. 15712-15728. [Online]. Available: https://doi.org/ 10.1016/j.egyr.2022.12.070.

[8] R. Korab, M. Połomski, T. Naczyński, and T. Kandzia, "A dynamic thermal model for a photovoltaic module under varying atmospheric conditions," Energy Conversion and Management, vol. 280, 2023, 116773, [Online]. Available: https://doi.org/10.1016/j.enconman.2023.116773.

[9] M.M. Fouad, L.A. Shihata, and E.I. Morgan, "An integrated review of factors influencing the perfomance of photovoltaic panels," Renewable and Sustainable Energy Reviews, vol. 80, 2017, pp. 1499-1511. [Online]. Available: https://doi.org/ 10.1016/j.rser.2017.05.141.

[10] N. Hussain, N. Shahzad, T. Yousaf, and et al., "Study of soiling on PV module performance under different environmental parameters using an indoor soiling station," Sustainable Energy Technologies and Assessments, vol. 50, 2022, [Online]. Available: https://doi.org/10.1016/j.seta.2022.10260.

[11] E. Skoplaki and J. Palyvos, "On the temperature dependence of photovoltaic module electrical performance," A review of efficiency power correlations, Solar Energy, no. 83, pp. 614-624, 2009. [Online]. Available: https://doi.org/10.1016/j.solener. 2008.10.008.

[12] P. Khaledi, M. Behboodnia, and M. Karimi, "Simulation and Optimization of Temperature Effect in Solar Cells CdTe with Back Connection Cu2O," International Journal of Optics, vol.2022, Article ID 1207082, 10 p. [Online]. Available: https://doi.org/10.1155/2022/1207082.

[13] G Notton, C Cristofari, M. Mattei, and P. Poggi, "Modelling of a double-glass photovoltaic module using finite differences," Appl.Therm.Eng. 2005,25: 2854-77. [Online]. Available: https://doi 10.1016/ j.applthermaleng.2005. 02.008.

[14] R. Ajdid, M. Ouassaid, and M. Maaroufi, "Power Output Evaluation of Polycrystalline PV Panel under Various Irradiances and Temperatures in Moroccan Regions," 1st International Conference on Electrical and Information Technologies ICEIT-2015.

[15] "Comparing major solar panel testing conditions," [Online]. Available: https://sinovoltaics.com/solar basics/comparing-major-solar-panel-testing conditions.

[16] I.R Juraev, I.A. Yuldoshev, and Z.I. Juraeva, "Experimental study of the characteristics of photovoltaic modules under field conditions," Scientific and technical journal "Alternative energy", Karshi, no. 2 (02), 2021.

[17] P. Singh and N. Ravindra, "Temperature dependence of solar cell performance-an analysis," Solar Energy Materials & Solar Cells, vol. 101, 2012, pp. 36-45. [Online]. Available: https:// doi:10.1016/j.solmat. 2012.02.019.

[18] H. Zhu, W. Lian, L. Lu, P. Kamunyu, C. Yu, S. Dai, and Y. Hu, "Online Modelling and Calculation for Operating Temperature of Silicon-Based PV Modules Based on BP-ANN," International Journal of Photoenergy, vol.2017, Article ID 6759295, 13 p. [Online]. Available: http://dx.doi.org/10.1155/ 2017/6759295.

[19] M. Hasanuzzaman, A. Malek, M. Islam, A. Pandey, and N. Rahim, "Global advancement of cooling technologies for PV systems," A review. Solar Energy 137, 2016, pp. 25-45. [Online]. Available: https://doi.org/10.1016/ j.solener.2016.07.010.

Autonomous Solar Power Plant for Individual use Simulation in LTspice Software Package Booster Voltage Converter

Nurullo Zikrillayev[1], Elyor Saitov[2], Javokhir Toshov[1], Usmonjon Akhmedov[2], Batir Muradov[1] and Doston Muxtorov[1]
[1]*Tashkent State Technical University, Islam Karimov, University Str. 2, Tashkent, Uzbekistan*
[2]*Kokan branch of Tashkent State Technical University, Fergana region, Usmon Nosir Str. 4, Kokand, Uzbekistan*
zikrillayev.nurullo@mail.ru, elyor.saitov@mail.ru, javokhir.toshov@mail.ru, usmonjon@gmail.com, muxtorov.doston@gmail.com

Keywords: Autonomous, Converter, Wind Characteristics, Voltage Circuits, Program, DC Voltage, LTspice.

Abstract: This work discusses the development of portable and inexpensive solar stations with low power, which are of great interest to farmers, summer residents, geologists, and travelers. The main focus is to ensure the reliability of the installations during transportation in various conditions, convenience and ease of operation, as well as ensuring the minimum size, weight, and reasonable price. The widespread use of alternative energy sources, particularly solar power plants, can save natural resources, improve the ecology of the habitat, reduce carbon dioxide emissions, and stimulate the training of specialists in the field of alternative energy sources. The purpose of this work is the development of portable solar stations that provide lighting for at least 8 hours, charging cell phones, and which are convenient for transportation. The work discusses the use of a DC voltage converter as an intermediate link between the solar battery and the autonomous voltage inverter, which provides high efficiency with a minimum of power elements and minimized weight and size.

1 INTRODUCTION

Our modern society is constantly moving somewhere, and this movement, otherwise called "permanent hypodynamia" - makes some sections of the population retire from society and its modern achievements. But in the modern world, probably, no one succeeds in completely retiring from all its features and achievements, and therefore, although some minimal connection with it is necessary for everyone, even the most inveterate modern “natives” [1].

In this article, we will consider mobile solar power plants based on solar batteries - through which, any of us, being away from home, is able to establish normal communication with relatives and friends, as well asprovide our other, minimal energy needs.

The sun is a source of inexhaustible energy reserves, andthe widespread use of this energy is one of the most urgent tasks for the Republic of Uzbekistan, which has significant solar energy resources. Other alternative energy sources (AES) are also significant. The main supplier of solar electrical energy is solar cells, the principle of which is based on the direct conversion of solar radiation energy directly into electrical energy. Solar cells generate energy at low operating costs and do not pollute the environment. In connection with the increase in the cost of petroleum products, a sharp increase in demand for alternative energy sources is predicted. It should also take into account the environmental and social impact of the use of solar power plants in remote and hard-to-reach areas. Widespread use of AES will save natural resources, improve the ecology of the habitat, reduce carbon dioxide emissions, improve the living conditions of the population and stimulate the training of specialists in the field of AES [2].

Calculations show that the use of RES in remote areas of the republic gives not only a social and environmental effect, but also a direct economic effect.

The purpose of this work was the development of portable solar stations that provide lighting for at least 8 hours, charging cell phones. At the same time, these stations should be convenient for transportation. At the same time, it was necessary to accurately calculate the power of the solar flux, which would not only

have to provide the necessary energy to the system for lightingat night with a duration of 8-10 hours, but also recharge the battery. At the same time, the device had to have optimal dimensions for ease of transportation andplacement [3].

In accordance with the requirements of [4], we havedeveloped a universal portable autonomous solar power plant for individual use, designed for the electrification of farms and summer cottages, as well as for use in remote and hard-to-reach places where there is no traditional power supply.

To expand the operating range of the solar inverter, a DC voltage converter is used as an intermediate link between the solar battery (SB) and the autonomous voltage inverter (AVI).

Such a solar energy converter (PESB) can be a step-up converter or an inverting converter. Such schemes provide high efficiency due toa minimum of power elements, while the weight and size of the filter are minimized by high conversion frequencies [5, 7].

2 MATERIALS AND METHODS

Since consumers require a standard voltage of 220 V with a frequency of 50 Hz, a step-up stabilized voltage converter will be used between the battery and the consumer. At night, the converter (inverter) will be powered from the battery, which is charged during the day.

The choice of voltage of the SB and AB panels is made taking into account the following considerations:

1) Ensure the safety of the solar battery and the battery, which decreases with increasing voltage;
2) Achieving reliability battery and solar battery. In high voltage circuits with a voltage of 220 V, reliability is reduced;
3) The high voltage battery has a large voltage spreadbetween the battery cells and requires a complex balancing system to prevent failure;
4) Reducing the input voltage ripple of the converter when operating from a solar battery in the current section of its current-voltage characteristic.

2.1 Methods

The paper investigates the effectiveness of the application of the controller control of the software using the UC3845A microcircuit [6].

The controller contains a high-frequency transistor SW, allowing for an efficiency of up to 86%. The circuit for switching on the microcircuit, modeled in the LTspice program, is shown in Figure 1.

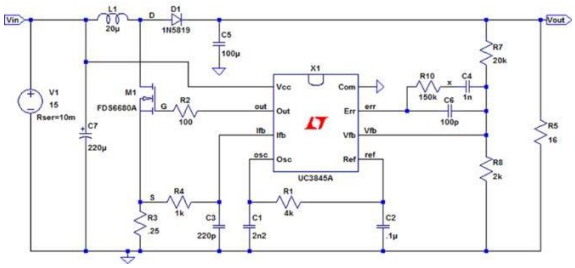


Figure 1: Schematic diagram of a boost converter [7].

The COMP pin is the output of the error amplifier. V_{fb} isthe inverting input of the error amplifier. It is usually connected to the output of the switching power supply through a resistor divider. I_{fb}, a voltage proportional to the current of the inductor is connected to this input.

The PWM uses this information to stop the output switch from working. O_{sc} adjusts the oscillator frequency and maximum duty cycle, programmed by connecting a resistor to V_{ref}. Operation up to 500 kHz is possible. The out pin directly drives the gate of the power MOSFET. V_{cc} - This pin is the positive power supply. V_{ref} provides charging current for capacitor C_2.

The switching circuit is intended for use in the converter as an inverting type, (manufacturer's recommendation) [7]. The output characteristics of voltage (U_{out}) and current (I_{out}) of the boost converter are shown in Figure 2.

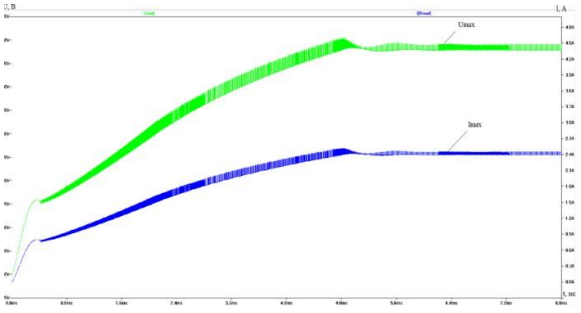


Figure 2: Boost converter voltage (U_{out}) and current (I_{out}) output characteristics.

On Figure 3 shows the transient process of the output voltage U_{out} and current I_{out}.

Based on the diagram of the transient process, we seethat the transient process lasts 5 m×s. And it

takes 5.2 m×s to reach the first maximum, and the voltage value at this point is 51 V.

On Figure 4 shows the transient current on the inductive element IL.

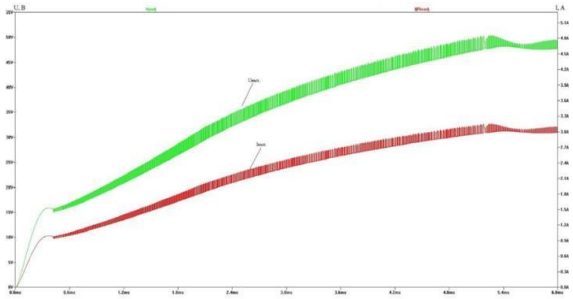


Figure 3: Transient process of output voltage U_{out} and current I_{out}.

Figure 4 shows the transient process of the output voltage (U_{out}) and current (I_{out}), which was obtained using the LTspice program.

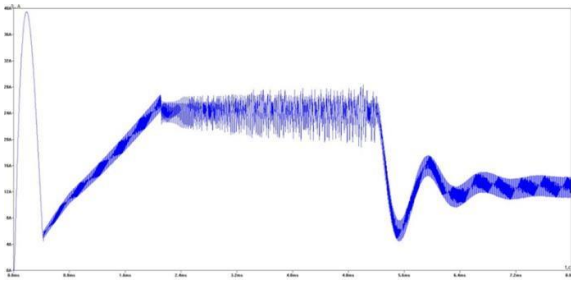


Figure 4: Transient current inductance in the program.

Study of the dependence of the output voltage on theload.

This dependence U_{out} ($R_н$) was studied on the model (Figure 4) and the results are shown in Table 1.

Table 1: Values of the dependence of U_{out} on $R_н$.

U_{out},V	15,5	36	44	48,5	48,5	48,5	48,5
$R_н$, Om	1	5	10	15	20	25	30

A graph of this dependence was also built (Figure 5).

As seen in Figure 5, the output voltage increases as the load increases to 15 Om. With an increase in load above 15 Om, voltage stabilization is realized, this indicates that the operating range of this 48 V circuit is achieved at 15-16 Om [8].

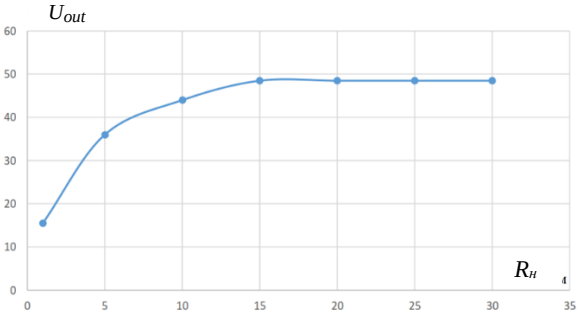


Figure 5: Graph of dependence of output voltage U_{out} on $R_н$.

2.2 Investigation of the Dependence of the Output Voltage on the Supply Voltage

This dependence U_{out} (E_{pit}) was studied on the model (Figure 5) and the results are shown in Table 2.

Table 2: Values of the dependence of U_{out} on E_{pit}.

U_{out},V	0,5	3,5	7,5	48	48,5	48,5	48,5
E_{pit},V	1	4	8	12	16	20	24

A graph of this dependence was also built (Figure 6).

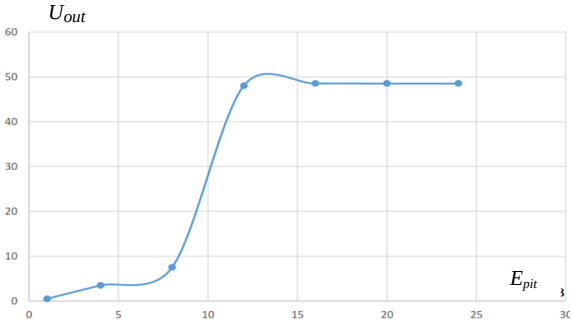


Figure 6: Graph of dependence of output voltage U_{out} on E_{pit}.

As seen in Figure 6, the change in the output voltage U_{out} is not linear and depends on the change in the input voltage U_{out}. Namely, the required 48 V is achieved at a supply voltage of 12 V, which indicates the correctness of the calculations of the power circuit and the possibility of alternative power supply from the battery [9].

2.3 Study of the Dependence of the Ripple Factor on the Load and Output Voltage

Coefficient of voltage (current) ripple according to the average value - A value equal to the ratio of the average value of the variable component of the pulsating voltage (current) to its constant component [10].

When calculating this indicator, information was collected on the minimum, average and maximum valuesof the output voltage, at various values of the supply voltage and load. The received data is substituted into a special formula. The ripple factor is calculated as the difference between the maximum and minimum values, divided by the average, after which the resulting amount is multiplied by 100%:

$$k_n = U_{\text{out}(max)} - U_{\text{out}}(min) \times 100\%.$$

The dependence k_n (E_{pit}) was studied on the model (Figure 6) and the results are shown in Table 3.

Table 3: Dependence values of k_n on $R_н$.

k_n, %	14,5	6,8	4	3	2	2	2	1,2
$R_н$,Om	1	5	10	15	20	25	30	40

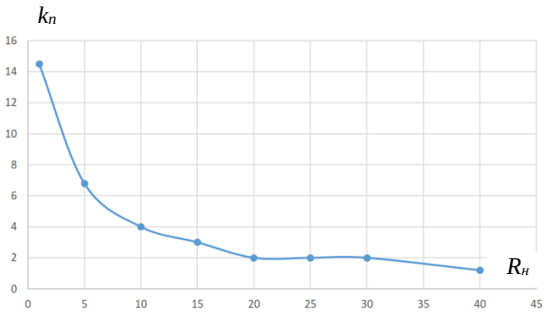


Figure 7: A graph of this dependence was also built.

A graph of this dependence was also built (Figure 7) k_n,%. From this graph it can be seen that with increasing load, the ripple coefficient decreases significantly in the area from 0 Om to 16 Om. With further increase in load, the degree of fall decreases. The system is designed for loads above 15 Om.

The dependence k_n (E_{pit}) was studied on the model (Figure 7) and the results are shown in Table 4.

Table 4: Dependence values of k_n on E_{pit}.

k_n, %	3	2	1,8	1,6
E_{pit}, V	12	16	20	24

A graph of this dependence was also built (Figure 8).

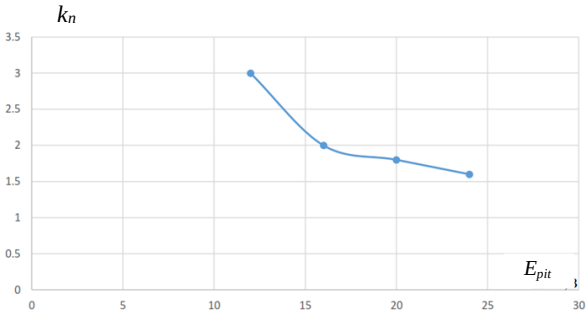


Figure 8: Graph of dependence of output voltage k_n on $R_н$.

Also, with an increase in the input voltage, it can be seenthat the ripple coefficient decreases. In the used range of supply voltages, the ripple factor does not exceed 3%.

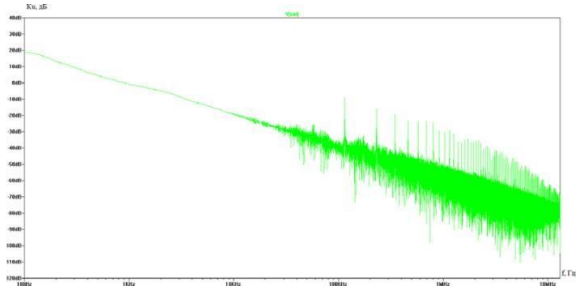


Figure 9: Frequency response of the boost converter.

The spectrum characterizes the type of the PPN harmonic (Figure 9). A sloping view indicates that this characteristic is rigid.

Such a solar station provides continuous operation of one TV and room illumination for 4 hours, and without a TV it illuminates the room and provides cell phone charging for 20 hours. To ensure the operation of such stations, it is necessary to place the solar panel in the sun from 6 to 10 hours, depending on weather conditions.

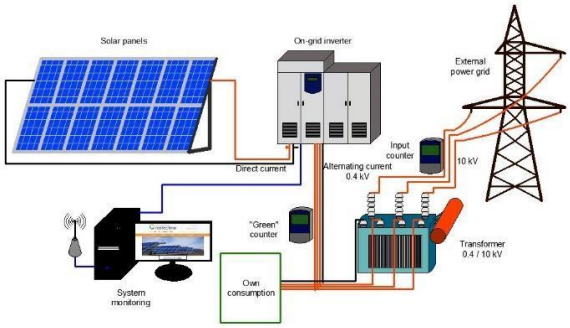


Figure 10: General and appearance of the solar panel device.

Our experiments have shown that the service life of such solar stations is at least 20 years, from time to time there is a need only to replace the batteries. In Figure 10 is the appearance and arrangement of such solar stations.

3 CONCLUSIONS

The study analyzed the coefficient of voltage (current) ripple of a solar-powered station designed for individual use and in remote locations where there is no traditional power supply. The ripple factor was calculated using data collected on the minimum, average, and maximum values of the output voltage at various supply voltage and load values. The results showed that with an increase in load and input voltage, the ripple coefficient decreases significantly. The system is designed for loads above 15 Ohms, and the ripple factor does not exceed 3% in the range of supply voltages used.

The solar station was found to provide continuous operation of one TV and room illumination for 4 hours and room illumination with cell phone charging for 20 hours. The station's service life was found to be at least 20 years, with only occasional battery replacements. This solar station is designed for individual use, as well as for use in remote and hard-to-reach places where there is no traditional power supply. Possible consumers of products can be budgetary organizations or individuals whose specific activity is carrying out work in remote and hard-to-reach places where there is no traditional power supply.

ACKNOWLEDGMENTS

In this article, we used the experience of Sergei Gennadyevich Mikhalchenko, Honorary Professor of the National Research Tomsk Polytechnic University.

REFERENCES

[1] S. V. Mitrofanov, D. K. Baykasenov, and M. A. Suleev, “Simulation Model of Autonomous Solar Power Plant with Dual-Axis Solar Tracker,” in 2018 International Ural Conference on Green Energy (UralCon), October 2018, doi: 10.1109/URALCON. 2018.8544275.

[2] B. E. Umirzakov, D. A. Tashmukhamedova, M. A. Tursunov, Y. S. Ergashov, and G. K. Allayarova, “Escape Depth of Secondary and Photoelectrons from CdTe Films with a Ba Film,” Technical Physics, vol. 64, no. 7, pp. 1051-1054, Jul. 2019, doi: 10.1134/S1063784219070260.

[3] Y. S. Ergashov, D. A. Tashmukhamedova, and B. E. Umirzakov, “On the Synthesis of Nanoscale Phases of Metal Silicides in the Near-Surface Region of Silicon and the Study of Their Electronic Structures by Passing Light,” Journal of Surface Investigation, vol. 11, no. 2, pp. 480-484, Mar. 2017, doi: 10.1134/S1027451017020252.

[4] S. V. Mitrofanov, D. K. Baykasenov, and A. U. Nemaltsev, “Operational Experience of a Solar Power Plant with a Dual-Axis Solar Tracking System in the Conditions of the Southern Urals,” E3S Web of Conferences, vol. 114, pp. 01003, Jan. 2019, doi: 10.1051/e3sconf/201911401003.

[5] Y. Ergashov, B. Donaev, S. Khudainazarov, and J. Normuminov, “Formation of Photoelectron Spectra of Alloys Niobium-Molybdenum-Zirconium,” E3S Web of Conferences, vol. 264, pp. 05036, April 2021, doi: 10.1051/e3sconf/202126405036.

[6] R. Suwarno, “Optimization Circuit Based Buck-Boost Converter for Charging the Solar Power Plant,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 6, no. 2, pp. 254-258, May 2017, doi: 10.11591/ijeecs.v6.i2.pp254-258.

[7] E. B. Saitov, Sh. Kodirov, B. M. Kamanov, N. Imomkulov, and I. Kudenov, “Increasing the Efficiency of Autonomous Solar Photovoltaic Installations for Power Supply of Agricultural Consumers,” in AIP Conference Proceedings, vol. 2432, no. 1, 040036, 2022.

[8] E. B. Saitov, “Renewable Energy Development in Uzbekistan: Current Status, Problems and Solutions,” in E3S Web of Conferences, vol. 216, 01134,2020.

[9] S. Sharma and N. Sengar, “Review of Solar PV Training Manuals and Development of Survey-Based Solar PV System Training Formats for Beginners,” in Solar Energy, vol. 241, pp. 72-84, Jul. 2022. doi: 10.1016/j.solener.2022.05.055.

[10] T. G. Zoryna, S. A. Aliaksandrovich, Y. S. Valeeva, M. V. Kalinina, L. E. Ilikova, and E. Y. Suvonovich, “Measures to stimulate the development of electric transport as a tool for the development of the territory,” in 2022 8th International Conference on Energy Efficiency and Agricultural Engineering (EE and AE) 2022 - Proceedings, 2022, doi: 10.1109/EEAE53789. 2022.9831360.

Application of the Time Method for Studying Linear Reproducing Systems for Assessing the Dynamic Accuracy of Devices Based on Magnetoelectric Systems

Oleksandr Sytnik[1], Miraziz Sagatov[2], Konstantin Klyuchka[1] and Sergey Protasov[1]
[1]*Cherkasy State Technological University, Shevchenka Boulevard Str. 460, Cherkasy, Ukraine*
[2] *Tashkent State Technical University named after Islam Karimov, University Str. 2, Tashkent, Uzbekistan*
sytnyk.ets@gmail.com, informtgtu@mail.ru, ux0cx@ukr.net, protasov.serhey@gmail.com

Keywords: Magnetoelectric Systems, Time Method, Differential Equation, Transfer Function, Hurwitz Stability Conditions, EMF Thermocouples.

Abstract: For the successful use of magnetoelectric measuring systems, an important point is the possibility of assessing their accuracy in a dynamic mode of operation. To obtain such estimates, methods based on frequency transformations and, accordingly, analysis of the composition of the spectral components of the currents of the interacting circuit and the magnetic field are traditionally used. At the same time, frequency methods have a number of limitations, in particular, due to the finiteness of the number of terms of the Fourier series in the analysis of periodic functions of time, as well as some other limitations when using the Fourier integral for non-periodic functions. In addition, there are certain limitations when using the well-known complex-spectral method and the method of typical effects. In the presented article, in addition to the indicated methods, it is proposed to consider the possibility of using a temporary research method to assess the dynamic properties of magnetoelectric systems. Also, as an example, the article presents an analysis of the dynamic properties of a magnetoelectric measuring system with electromagnetic damping, which can be extended to more complex measuring systems of this type.

1 INTRODUCTION

When assessing the dynamic accuracy of magnetoelectric systems (i.e., those whose operation is based on the interaction of a circuit with a current and a magnetic field), the frequency method for studying linear reproducing systems is usually used [1, 2]. This research method is based on the assumption that electric currents (initializing currents) flowing through the circuit can be represented as a set of constantly acting sinusoidal components.

It is easy to conclude that the validity of using the frequency method is limited to cases where the currents in the circuit are accurately described by periodic functions of time, as well as by functions of time expressed by the Fourier integral [3, 4].

At the same time, experience shows that in specific physical examples, the above time functions are not enough to accurately describe the currents in the circuit, and, consequently, to assess the accuracy of reproducing the effects of a specific device. The time functions used in such methods of studying magnetoelectric systems as the complex spectral method and the method of typical effects [5, 6] also turn out to be insufficient.

In this regard, of particular interest is the use of the time method for studying linear reproducing systems to evaluate the dynamic properties of magnetoelectric systems [7].

With this research method, to describe the impacts, time functions $\alpha_{in}(t)$ are used that satisfy the condition:

$$|\alpha_{in}^{\kappa}(t)|\begin{cases}=0; & at\ -\infty\leq k\leq+\infty,-\infty\leq k\leq+\infty\\ \leq M\Omega^{\kappa}; & at\ +0\leq t\leq\infty,-\infty\leq k\leq+\infty\end{cases} \quad (1)$$

where M and Ω are some given real positive numbers.

The set of time functions satisfying condition (1) will be referred to below as the class of functions $S(M,\Omega)$.

From a practical point of view, the determination of the parameters M and required for the time method for the class of impacts to be reproduced is associated, as a rule, with less difficulties than the determination of the spectral composition of these impacts, required in the frequency and complex-spectral methods, and the determination of the typical impact, required in

the method of typical influences. It is characteristic that in the study of the reproducing properties of magnetoelectric systems, in most practical cases, the relation can be used to determine the parameter Ω:

$$\Omega \approx \frac{|i^{(1)}(t)|_{\max max}}{|i(t)|_{\max max}}.$$

where $|i(t)|_{max\,max}$ – the limiting value of the module of the currents to be investigated in the circuit at $-\infty \le t \le +\infty$;
$\left|i^{(1)}(t)\right|_{\max max}$ – is the limiting value of the modulus of the rate of change in time of the currents to be investigated in the circuit at $-\infty \le t \le +\infty$.

2 METHODS

The main purpose of magnetoelectric systems is, as is known, the conversion of an electric current flowing through a certain circuit, $i_к(t)$, into a deviation of a pointer mechanically connected to this circuit with a current by the value $L(t)$**.** The required nature of this transformation is described either by the:

$$L_0(t) = K(0)i_k(t) \quad (2)$$

expressing the ideal tracking process, or by the

$$L_\tau(t) = K(0)i_k(t+\tau) \quad (3)$$

expressing the ideal registration process.

In (2) and (3): $L_0(t)$ and $L_\tau(t)$ are the functions of time describing the required reproduction of the impact $i_k(t)$: $K(0)$ is the sensitivity of the device with the magnetoelectric unit for direct current; τ – admissible time of displacement of registration of the investigated currents in the circuit.

Obviously, the actual reproduction of the impact $i_k(t)$, described by the time $L(t)$, function differs both from $L_0(t)$, and from $L_\tau(t)$. The mutual deviation of the time functions $L_0(t)$ and $L(t)$ is the instantaneous error of the tracking device (4):

$$\Delta_0(t) = L(t) - L_0(t) \quad (4)$$

and the mutual deviation of the time functions $L_\tau(t)$ and $L(t)$ is the instantaneous registration error (5)

$$\Delta_\tau(t) = L(t) - L_\tau(t)\,. \quad (5)$$

The functions $\Delta_0(t)$ and $\Delta_\tau(t)$ quite fully characterize the dynamic properties of the magnetoelectric system with respect to the current $i_k(t)$. When determining these functions, we will proceed from the assumption that the relationship between $L(t)$ and $i_к(t)$ is expressed by a linear differential equation with constant coefficients.

In the classical theory of magnetoelectric systems [8], the relationship between $L(t)$ and $i_к(t)$ is expressed by a linear inhomogeneous second-order differential (6):

$$L(t) + \frac{2s}{\omega_0}L^{(1)}(t) + \frac{1}{\omega_0^2} - L^{(2)}(t) = K(0)\, i_к(t), \quad (6)$$

where s - the degree of calming of the moving system of the circuit with the investigated currents; ω_0 – natural frequency of the mobile system.

At the same time, a thorough study of the dynamic properties of magnetoelectric systems, carried out by R. R. Kharchenko and N. N. Evtikhiev, showed that to describe the relationship between $L(t)$ and $i_к(t)$ it is necessary to use linear differential equations with orders higher than the second. Therefore, in this paper, to describe the relationship between $L(t)$ and $i_к(t)$, we use the linear inhomogeneous differential:

$$L(t) + \sum_{k=1}^{n} b_k L^{(к)}(t) = K(0)\, i_к(t), \quad (7)$$

whose order (the number n) is not bounded from above.

The coefficients b_k in this case are real positive numbers determined by the design features of the used magnetoelectric device.

Consider the structure of the solution of the differential (7) for the case when the initializing currents $i_k(t) \in S(M, \Omega)$ and the transfer function of the device (8)

$$K(p) = K(0)\frac{1}{1 + \sum_{k=1}^{n} b_к p^{к}}, \quad (8)$$

is an analytic function of a complex variable inside $|p| < R_{min}$ and outside the circle $|p| \le R_{max}$. The relationship between coefficients b_k and numbers R_{min} and R_{max} is established by the well-known theorem from higher algebra on the limits of zeros of polynomials with real coefficients [9].

A) Initializing currents $i_k(t) \in S(M, \Omega)$, $\Omega < R_{min}$.

It can be proved that when solving the $\Omega < R_{min}$ differential (6), the function $L(t)$ can be represented as the sum of two components (9)

$$L(t) = L(t)_{forc} + L(t)_{fr}. \quad (9)$$

The first of these components is forced - $L(t)_{forc}$, due to the reaction of the moving system of the circuit with current to smooth changes in time of the current $i_k(t)$ and its derivatives $\{i_k{}^{(k)}(t)\}$ is determined by the (10):

$$L(t)_{forc} = \sum_{k=0}^{\infty} \frac{1}{\kappa!} K_{(0)}^{(k)} i_k^{(k)}(t), \tag{10}$$

where

$$K^{(\kappa)}(0) = \lim_{p \to 0} \frac{d^{\kappa}}{dp^{\kappa}} K(p).$$

The second component is free - $L(t)_{fr}$, due to the reaction of the circuit system with current to abrupt changes in time of the current $i_k(t)$ and its derivatives $i_k{}^{(k)}(t)$, at $t = 0$, can be determined by the (11):

$$L(t)_{fr} = \sum_{k=1}^{N} \frac{1}{(s_{\kappa}-1)!} \lim_{p \to -\beta_{\kappa}} \frac{d^{\kappa}}{dp^{\kappa}} [K(p) I_{\kappa}(p) (p+\beta_{\kappa})^{sk} e^{pt}] \tag{11}$$

where N – the number of poles of the transfer function of the magnetoelectric system $K(\mathrm{p})$; β_k – k^{th} pole $s_k{}^{th}$ order $K(p)$; $I_k(p)$ – image (according to Laplace) of the initializing current $i_k(t)$.

From the possibility of division into components $L(t)$ follows the possibility of division into components $\Delta_{\tau}(t)$ and $\Delta_0(t)$. Obviously (12) – (14),

$$\Delta_0(t)_{forc} = \sum_{k=1}^{\infty} \frac{1}{\kappa!} K^{(\kappa)}(0)\, i_{\kappa}^{(\kappa)}(t), \tag{12}$$

$$\Delta_{\tau}(t)_{forc} = \sum_{k=2}^{\infty} \frac{1}{\kappa!} [K^{(\kappa)} - {}_{(0)} K(1)_{(0)}{}^{\kappa} K^{-k+1}{}_{(0)}] i_{\kappa}^{(\kappa)}(t), \tag{13}$$

$$\Delta_0(t)_{fr} = \Delta_{\tau}(t)_{fr} = L(t)_{fr}. \tag{14}$$

The above formulas allow us to estimate the limiting values $|\Delta_0(t)_{forc}|_{\max max}$, $|\Delta_{\tau}(t)_{forc}|_{\max max}$ and $|L(t)_{fr}|_{max\, max}$ for currents $i_c(t) \in S(M,\Omega), \Omega < R_{min}$ at $-\infty < t \le +\infty$.

It should be noted that for magnetoelectric systems, which are known to be stable systems, of greatest interest from a practical point of view are estimates of the limiting values $|\Delta_0(t)_{forc}|$ and $|\Delta_{\tau}(t)_{forc}|$.

To estimate the limiting values of $|\Delta_0(t)_{forc}|$ and $|\Delta_{\tau}(t)_{forc}|$, we use the inequalities

$$|i_c^{(k)}(t)| \le M, \Omega^k, 0 \le k \le \infty,$$

following from (1). The above inequalities allow us to write that for $-\infty < t \le +\infty$

$$|\Delta_0(t)_{forc}| \le K(0) M \sum_{k=1}^{\infty} \frac{1}{\kappa!} \left| \frac{K^{(\kappa)}(0)}{K(0)} \right| \Omega^{\kappa}, \tag{15}$$

$$|\Delta_{\tau}(t)_{forc}| \le K(0) M \sum_{k=1}^{\infty} \frac{1}{\kappa!} \left| \frac{K^{(\kappa)}(0)}{K(0)} - \left(\frac{K^{(1)}(0)}{K(0)} \right)^{\kappa} \right| \Omega^{\kappa}. \tag{16}$$

Analysis of estimate (15) shows that the minimum values $|\Delta_0(t)_{forc}|_{\max max}$ for currents $i_c(t) \in S(M,\Omega), \Omega < R_{min}$ occur at minimum values of the coefficients $\left\{ \left| \frac{K^{(k)}(0)}{K(0)} \right| \right\}$ or, the same thing, at the minimum values of the coefficients b_k and ratio $\frac{\Omega}{R_{min}}$.

In turn, it follows from estimate (16) that for currents $i_c(t) \in S(M,\Omega), \Omega < R_{min}$, the minimum values $|\Delta_{\tau}(t)_{forc}|_{\max max}$ occur at the minimum values of the coefficients,

$$\left\{ \left| \frac{K^{(\kappa)}(0)}{K(0)} - \left(\frac{K^{(1)}(0)}{K(0)} \right)^{\kappa} \right| \right\},$$

which, in particular, is ensured when the following relations are performed

$$\begin{cases} b_2 = \frac{1}{2!} b_1^2; \\ \ldots\ldots\ldots\ldots; \\ b_n = \frac{1}{n!} b_1^n, \end{cases} \tag{17}$$

when for $2 \le k \le n$:

$$\left| \frac{K^{(\kappa)}(0)}{K(0)} - \left(\frac{K^{(1)}(0)}{K(0)} \right)^{\kappa} \right| = 0$$

For the differential (6) used in the classical theory of devices using the magnetoelectric principle, the minimization $|\Delta_{\tau}(t)_{forc}|_{\max max}$ conditions established by relations (17) are reduced to the equality

$$\frac{1}{\omega_0^2} = \frac{1}{2!} \left(\frac{2s}{\omega_0} \right)^2,$$

which matches $s = (\sqrt{2})^{-1}$ or $\frac{1}{\sqrt{2}}$.

It should be noted that at $n \ge 5$, magnetoelectric systems that satisfy conditions (17) are physically unfeasible, since for $n \ge 5$, the polynomial

$$b(p) = 1 + \sum_{k=1}^{n} \frac{1}{\kappa!} b^{\kappa}{}_1 p^{\kappa}$$

does not satisfy the Hurwitz stability conditions [10 - 12].

From a practical point of view, the use of inequalities (15) and (16) to assess the dynamic accuracy of a magnetoelectric system is inappropriate. This is explained, first of all, by the well-known difficulties of summing infinite series

$$\sum_{k=1}^{\infty}\frac{1}{\kappa!}\left|\frac{K^{(\kappa)}(0)}{K(0)}\right|\Omega^{\kappa}$$

and

$$\sum_{k=2}^{\infty}\frac{1}{\kappa!}\left|\frac{K^{(\kappa)}(0)}{K(0)}-\left(\frac{K^{(1)}(0)}{K(0)}\right)^{\kappa}\right|\Omega^{\kappa},$$

representing upper bounds for the modules of functions

$$\varepsilon_0(p)=\frac{K(p)}{K(0)}-1 \quad \text{and} \quad \varepsilon_\tau(p)=\frac{K(p)}{K(0)}-e^{\frac{K^{(1)}(0)}{K(0)}p}$$

respectively in a circle $|p|\le\Omega<R_{min}$.

In this regard, it is advisable to estimate the limiting values $|\Delta_0(t)_{forc}|$ and $|\Delta_\tau(t)_{forc}|$ using the inequalities

$$|\Delta_0(t)_{forc}|\le K(0)M\eta_i(\Omega), \tag{18}$$

$$|\Delta_\tau(t)_{forc}|\le K(0)M\varphi_i(\Omega), \tag{19}$$

where $\eta_i(\Omega)$ and $\varphi_i(\Omega)$ are rather simply calculated majorants of the series

$$\sum_{k=1}^{\infty}\frac{1}{\kappa!}\left|\frac{K^{(\kappa)}(0)}{K(0)}\right|\Omega^{\kappa}$$

and

$$\sum_{k=2}^{\infty}\frac{1}{\kappa!}\left|\frac{K^{(\kappa)}(0)}{K(0)}-\left(\frac{K^{(1)}(0)}{K(0)}\right)^{\kappa}\right|\Omega^{\kappa}.$$

It can be proved [7] that the function $\eta_i(\Omega)$ can be used as a majorant

$$\eta_1(\Omega)=\frac{\sum_{k=1}^{n}b_k\Omega^{\kappa}}{1-\sum_{k=1}^{n}b_k\Omega^{\kappa}},$$

and as a majorant $\varphi_i(\Omega)$ – the function

$$\varphi_1(\Omega)=\eta_1(\Omega)+e^{b_1\Omega}-1-2b_1\Omega.$$

It follows from inequalities (18) and (19) that for a magnetoelectric device characterized by differential (7), a sufficient condition for registering an initializing current $i_k(t)\in S(M,\Omega),\Omega<R_{min}$ with a reproduction scale $K(0)$, a time offset $\tau=\frac{K^{(1)}(0)}{K(0)}=-b_1$, and limit values $|\Delta_0(t)_{forc}|$ and $|\Delta_\tau(t)_{forc}|$, not exceeding for $-\infty<t\le+\infty$ some numbers $\bar{\Delta}_0$ and $\bar{\Delta}_\tau$, is the fulfillment of the relations

$$K(0)M\eta_i(\Omega)\le\bar{\Delta}_0 \tag{20}$$

and

$$K(0)M\varphi_i(\Omega)\le\bar{\Delta}_\tau, \tag{21}$$

at $0\le\Omega<R_{min}$.

An analysis of the functions $\eta_1(\Omega)$ and $\varphi_1(\Omega)$ shows that, at sufficiently small values ΩR_{min}^{-1} the influence of the coefficients $b_k, 2<k\le n$ on the nature of these functions is extremely insignificant.

This allows, when registering initializing currents $i_c(t)\in S(M,\Omega),\Omega<R_{min}$, to estimate the limit values $|\Delta_0(t)_{forc}|$ and $|\Delta_\tau(t)_{forc}|$, to use the relations (22), (23)

$$|\Delta_0(t)_{forc}|\le K(0)M\frac{2s\frac{\Omega}{\omega_0}+\frac{\Omega^2}{\omega_0^2}}{1-2s\frac{\Omega}{\omega_0}-\frac{\Omega^2}{\omega_0^2}}=K(0)M\varepsilon_0\left(\frac{\Omega}{\omega_0}\right) \tag{22}$$

and

$$|\Delta_\tau(t)_{forc}|\le K(0)M\left[\varepsilon_0\left(\frac{\Omega}{\omega_0}\right)+e^{2s\frac{\Omega}{\omega_0}}-1-4s\frac{\Omega}{\omega_0}\right]= \tag{23}$$

$$=K(0)M\varepsilon_\tau\left(\frac{\Omega}{\omega_0}\right),$$

derived from differential (6).

To estimate the attenuation intensity $L(t)_{fr}$ in the case under consideration, following relation can be used

$$|L(t)_{fr}|\le K(0)M\frac{e^{-s\omega_0 t}}{\left(1-\frac{\Omega}{\omega_0}\right)\sqrt{|1-s^2|}}=K(0)M\varepsilon\left(\frac{\Omega}{\omega_0},t\right). \tag{24}$$

Graphs of the functions

$$\varepsilon_0\left(\frac{\Omega}{\omega_0}\right);\ \varepsilon_\tau\left(\frac{\Omega}{\omega_0}\right) \quad \text{and} \quad \varepsilon\left(\frac{\Omega}{\omega_0},t\right)$$

for different values of s are shown in Figures 1-3.

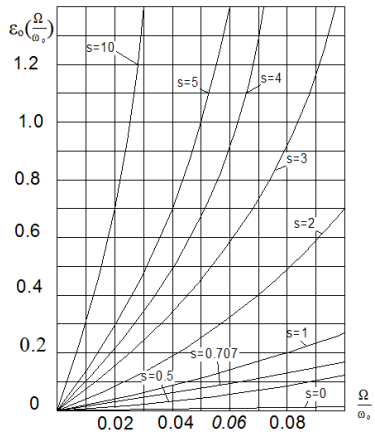


Figure 1: Graph of the function $\varepsilon_0\left(\frac{\Omega}{\omega_0}\right)$.

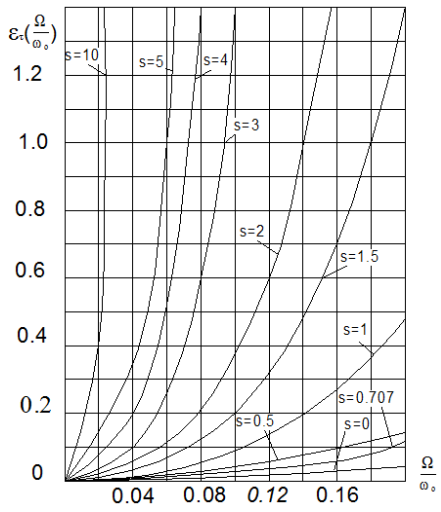


Figure 2: Graph of the function $\varepsilon_\tau\left(\frac{\Omega}{\omega_0}\right)$.

It follows from these graphs that the circuit with a current, characterized by the differential (6) at $K(0) = 2$ mm/mA, $\omega_0 = 500$ Hz and $s = 0{,}5$ will make it possible to display the initializing current $i_k(t) \in S(M_1, \Omega)$, $M = 100$mA, $\Omega = 25\ s^{-1}$ with a limit value $|\Delta_0(t)_{forc}|$, not exceeding $2 \times 100 \times 0{,}05 = 10\ mm$, and limit value $|\Delta_\tau(t)_{forc}|$ not exceeding $2 \times 100 \times 0{,}01 = 2\ mm$, (at $\tau = -0{,}002\ s$). The limiting majorant in this case is the function

$$K(0)M\varepsilon\left(\frac{\Omega}{\omega_0}, t\right) = 275e^{-25t}.$$

Note that currents $i_k(t) \in S(M, \Omega)$, $M = 100$ mA, $\Omega = 25\ s^{-1}$ include, in particular, currents $i_{1k}(t) = 100 \sin 25t$; $i_{2k}(t) = 100e^{-25t}$, etc.

In cases where the amplitude-frequency characteristic of the magnetoelectric system $M^{(\omega)}$ and its phase-frequency characteristic $\varphi(\omega)$ are known, the following relations can be used to estimate the limit values $|\Delta_0(t)_{forc}|$ and $|\Delta_\tau(t)_{forc}|$, when displaying the initializing currents $i_c(t) \in S(M, \Omega), \Omega < R_{min} \leq \omega_{kp}$

$$\Delta_0|(t)_{forc}| \cong M(0)M\left|\varphi^{(1)}(0)\right|\Omega, \qquad (25)$$

$$|\Delta_\tau(t)_{forc}| \cong M(0)M\left|\frac{M(\Omega)}{M(0)} - 1\right|. \qquad (26)$$

obtained by substituting into inequalities (20) and (17) the relations

$$\sum_{k=1}^{\infty}\frac{1}{\kappa!}\left|\frac{K^{(\kappa)}(0)}{K(0)}\right| \Omega^\kappa > \frac{1}{1!}\left|\frac{K^{(1)}(0)}{K(0)}\right|\Omega,$$

$$\sum_{k=2}^{\infty}\frac{1}{\kappa!}\left|\frac{K^{(\kappa)}(0)}{K(0)} - \left(\frac{K^{(1)}(0)}{K(0)}\right)^\kappa\right| \Omega^\kappa > \frac{1}{2!}\left|\frac{K^{(2)}(0)}{K(0)} - \left(\frac{K^{(1)}(0)}{K(0)}\right)^2\right|\Omega^2,$$

as well as the relations proved in [8]

$$\frac{K^{(1)}(0)}{K(0)} = \varphi^{(1)}(0),$$

$$M(\Omega) \cong M(0) + \frac{1}{2!}M^{(2)}(0)\Omega^2,$$

$$\frac{1}{2!}M^{(2)}(0) = \frac{1}{2!}\left|K^{(2)}(0) - \frac{K^{(1)2}(0)}{K(0)}\right|.$$

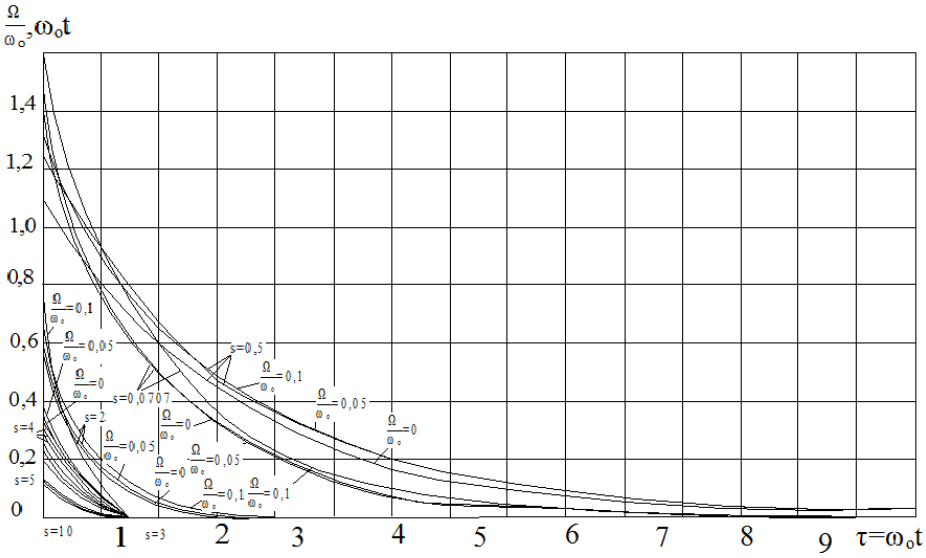


Figure 3: Graph of the function $\varepsilon\left(\frac{\Omega}{\omega_0}, t\right)$.

B) Initializing currents $i_k(t) \in S(M,\Omega)$, $\Omega > R_{max}$.

It can be proved that for $\Omega > R_{max}$, the representation of $L(t)$ as a sum of two components

$$L(t) = L(t)_{forc} + L(t)_{fr}.$$

is preserved. The possibility of determining $L(t)_{fr}$ by formula (11) also remains. At the same time, formula (10) is not suitable for determination $L(t)_{forc}$.

For $\Omega < R_{max}$, $L(t)_{forc}$ is defined this way:

$$L(t)_{forc} = \sum_{k=0}^{\infty} \frac{1}{K!} Q^{(K)}(0) i_c^{(-K)}(t),$$

where

$$Q^{(\kappa)}(0) = \lim_{z\to 0} \frac{d^{\kappa}}{dz^{\kappa}} K(z^{-1}),$$

and

$$i_{\kappa}^{(-\kappa)}(t) = \frac{1}{(k-1)!} \int_0^t i_{\kappa}(\tau)(t-\tau)^{\kappa-1} d_{\tau}$$

in case if

$$K(p) = K(0) \frac{1}{1 + \sum_{k=1}^{n} b_{\kappa} p^{\kappa}},$$

for $k \le n, Q^{(k)}(0) = 0$ accordingly,

$$L(t)_{forc} = \sum_{k=n+1}^{\infty} \frac{1}{\kappa!} Q^{(\kappa)}(0) i_{\kappa}^{(-\kappa)}(t). \qquad (27)$$

Analysis of formula (27) shows that in relation to the currents in the circuit $i_k(t) \in S(M,\Omega), \Omega > R_{min}$, the magnetoelectric system, characterized by the differential (7), manifests itself as an "opaque" system with a forced component of the "suppression" error of influences $\delta(t)_{forc}$, determined by the

$$\delta(t)_{forc} = L(t)_{forc}.$$

Formula (27) also implies the possibility of using the studied magnetoelectric system for (n + 1 + q) - a short integration of currents $i_k(t) \in S(M,\Omega), \Omega > R_{max}$ with a scale

$$\frac{1}{(n+1+q)!} Q^{(n+1+q)}(0)$$

and a forced component of the integration error $j(t)_{forc}$, determined by the

$$j(t)_{forc} = L(t)_{forc} - \frac{1}{(n+1+q)!} Q^{(n+1+q)}(0) i_c^{(n+1+q)}(t).$$

The evaluation of the limit values $|\delta(t)_{forc}|$ and $|j(t)_{forc}|$ for the currents in the circuit $i_k(t) \in S(M,\Omega), \Omega > R_{max}$ can be made by similar methods for evaluating the limit values $|\Delta_0(t)_{forc}|$ and $|\Delta_\tau(t)_{forc}|$ for the currents $i_k(t) \in S(M,\Omega)$, $\Omega < R_{min}$.

C) Initializing currents $i_k(t) \in S(M,\Omega)$, $R_{min} \le \Omega \le R_{max}$.

At $R_{min} \le \Omega \le R_{max}$, the division of $L(t)$ into forced and free components is generally impossible. This is explained by the fact that among the currents in the circuit $i_k(t) \in S(M,\Omega), R_{min} \le \Omega \le R_{max}$ there are those that are able to bring the moving system of the circuit of the magnetoelectric system into a state of generalized resonance, a phenomenon first described by S.P. Strelkov. A special case of this state is the well-known resonance that occurs when sinusoidal currents flow through the circuit of the magnetoelectric system, the frequency of which coincides with the natural frequencies of the moving system of the circuit with current.

3 RESULTS AND DISCUSSION

In the above analysis of the properties of the magnetoelectric system, it was assumed that the nature of the transfer function of the device is determined only by its design parameters and is completely independent from the parameters of the electrical circuit of the device.

However, this assumption is valid only in cases where the influence of EMF generated in a current-carrying circuit on the properties of the magnetoelectric system is practically imperceptible. It is justified, in particular, when analyzing the properties of magnetoelectric systems with single-turn circuits with oil and magneto-inductive damping. At the same time, for multi-turn circuits with electromagnetic damping, the above assumption is generally non-uniform.

As an example, let us analyze the properties of a magnetoelectric system with electromagnetic damping when using it to study the EMF of thermocouples.

Let $e_T(t)$ be the EMF of thermocouple to be studied; r_T is the ohmic resistance of the thermocouple; $e_k(t)$ – EMF generated by the circuit; r_k – ohmic resistance of the loop; r_∂ – damping loop resistance (shunt).

For the case under consideration, the relationship between $i_k(t)$, $e_T(t)$ and $e_k(t)$is expressed by the (28)

$$i_{\kappa}(t)=\frac{r_{\partial}e_T(t)+(r_T+r_{\partial})e_{\kappa}(t)}{r_{\kappa}r_T+r_{\kappa}r_{\partial}+r_Tr_{\partial}}. \quad (28)$$

Let's pretend that

$$e_c(t)=-k_{cur}L^{(1)}(t), \quad (29)$$

where k_{cur} – constructive constant of the loop with current, and

$$i_c(t)=\frac{1}{K(0)}\left\{L(t)+\sum_{k=1}^{n}b_{\kappa}L^{(\kappa)}(t)\right\} \quad (30)$$

in accordance with the differential (7) without taking into account the effect of electromagnetic damping of the current loop due to the EMF generated by it. Substituting the expanded expressions for $e_k(t)$ (29) and $i_k(t)$ (30) into (28), we can write that

$$\frac{K(0)r_{\partial}e_T(t)}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}= \quad (31)$$

$$=L(t)+\left(b_1+\frac{k_{cur}\,K(0)(r_T+r_{\partial})}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}\right)L^{(1)}(t)+\sum_{k=1}^{n}b_{\kappa}L^{(\kappa)}(t).$$

This means that the transfer function of the magnetoelectric system with respect to EMF of thermocouple is a fractional rational function

$$N(p)=\frac{K(0)r_{\partial}}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}\times$$

$$\times\frac{1}{1+\left(b_1+\frac{k_{cur}\,K(0)(r_T+r_{\partial})}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}\right)p+\sum_{k=2}^{n}b_{\kappa}p^{\kappa}}.$$

Let the studying EMF of thermocouple $i_k(t)\,\epsilon\,S(M,\Omega),\Omega>R'_{min}$, where R'_{min} is the convergence radius of the series

$$\sum_{k=0}^{\infty}\frac{1}{\kappa!}N^{(\kappa)}(0)p^{\kappa}=N(p).$$

By analogy with the previous one, it can be argued that in the case under consideration $i_k(t)\,\epsilon\,S(M,\Omega),\Omega<R'_{min}$, the effects will be reproduced by the magnetoelectric system with a reproduction scale

$$N(0)=\frac{K(0)r_{\partial}}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}},$$

displacement time

$$\tau=-\left(b_1+\frac{k_{cur}K(0)(r_T+r_{\partial})}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}\right)$$

and limit values $\left|\Delta_0(t)_{forc}\right|$ and $\left|\Delta_{\tau}(t)_{forc}\right|$, estimated with inequalities

$$|\Delta_0(t)_{forc}|\le\frac{K(0)r_{\partial}M}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}\eta'_1(\Omega) \quad (32)$$

and

$$|\Delta_{\tau}(t)_{forc}|\le\frac{K(0)r_{\partial}M}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}\varphi'_1(\Omega), \quad (33)$$

where

$$\eta'_1(\Omega)=\frac{\left|b_1+\frac{k_{cur}\,K(0)(r_T+r_{\partial})}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}\right|\Omega+\sum_{k=1}^{n}b_{\kappa}\Omega^{\kappa}}{1-\left|b_1+\frac{k_{cur}\,K(0)(r_T+r_{\partial})}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}\right|\Omega-\sum_{k=1}^{n}b_{\kappa}\Omega^{\kappa}},$$

$$\varphi'_1(\Omega)=\eta'_1(\Omega)+e^{\eta_1^{I(1)}(0)\Omega}-1-2\eta'^{(1)}_1(0).$$

When registering EMF of thermocouple $e_T(t)\,\epsilon\,S(M,\Omega),\Omega\ll R'_{min}$, the condition for minimizing the limit value $\left|\Delta_0(t)_{forc}\right|$ is the fulfillment of the relation

$$b_1+\frac{k_{cur}\,K(0)(r_T+r_{\partial})}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}=0 \quad (34)$$

and relations

$$b_k=0 \text{ at } 2\le k\le n. \quad (35)$$

The condition for minimizing the limiting value $\left|\Delta_{\tau}(t)_{forc}\right|$ in the case under consideration is the fulfillment of the relations

$$b_{\kappa}=\frac{1}{\kappa!}\left[b_1+\frac{k_{cur}\,K(0)(r_T+r_{\partial})}{r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial}}\right]^{\kappa}, \quad (36)$$

at $2\le k\le n$.

An analysis of the differential (31), inequalities (32) and (33), as well as the conditions for minimizing the limit values $\left|\Delta_0(t)_{forc}\right|$ and $\left|\Delta_{\tau}(t)_{forc}\right|$, shows that the use of the effect of electromagnetic damping of the current circuit, increasing in

$$\left[1+\frac{k_{cur}\,K(0)(r_T+r_{\partial})}{(r_{\kappa}r_{\partial}+r_{\kappa}r_T+r_Tr_{\partial})b_1}\right]$$

times the damping intensity $L(t)_{fr}$ at the same time leads to an increase in the displacement time τ and the limit value $\left|\Delta_0(t)_{forc}\right|$. From conditions (36) it follows, in particular, that for circuits with current, characterized by differential (6), the dynamic correction of the magnetoelectric system in order to reduce the limit value $\left|\Delta_{\tau}(t)_{forc}\right|$ takes place at

$$\frac{r_Tr_{\partial}}{r_T+r_{\partial}}=\frac{k_{cur}K(0)\omega_0}{\sqrt{2}(1-\sqrt{2}s)}-r_{\kappa}, \quad (37)$$

in cases where $r_\tau \gg r_\partial$ condition (37) reduces to the relation

$$r_\partial \approx \frac{k_{cur} K(0)\omega_0}{\sqrt{2}(1-\sqrt{2}s)} - r_\kappa. \qquad (38)$$

With the resistance of the thermocouple $r_T = \frac{k_{cur} K(0)\omega_0}{\sqrt{2}(1-\sqrt{2}s)} - r_c$, the dynamic correction of the magnetoelectric system in order to reduce the limit value $|a_\tau(t)_{forc}|$is provided in the absence of r_∂ (i.e., at $r_\partial = \infty$ 1).

Similarly, the properties of more complex measuring devices, in which magnetoelectric systems are used as recording organs, can be investigated.

CONCLUSION

The possibility of using the time method for solving problems of dynamic accuracy associated with the design and operation of magnetoelectric systems, as well as various types of measuring devices, in which current-carrying circuits as part of the magnetoelectric system are used as recording organs, is shown.

REFERENCES

[1] A. Verlan and S. Moskalyuk, “Matematicheskoye modelirovaniye nepreryvnykh dinamicheskikh sistem,” AN USSR, In-t problem modelirovaniya v energetike, Kiev, Nauk. dumka, 1988, pp. 288.

[2] A. Verlan and M. Sagatov, “Inverse problems of the dynamics of observation interpretation systems,” J. Phys.: Conf. Ser., vol. 2131, p. 032109, 2021, [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/2131/3/032109.

[3] A. Sytnik, K. Klyuchka, and S. Protasov, “Development of the Method for Creating Explicit Integral Dynamic Models of Measuring Transducers,” Eastern-European Journal of Enterprise Technologies, vol. 5, no. 4, pp. 40-48, 2017, [Online]. Available: doi:10.15587/1729-4061.2017.111444.

[4] A. Verlan, V. Fedorchuk, and J. Sterten, “Using Non-linear Integral Models in Automatic Control and Measurement Systems for Sensors’ Input Signals’ Recovery,” in 11th World Conference “Intelligent System for Industrial Automation” (WCIS-2020), Advances in Intelligent Systems and Computing, vol. 1323, Springer, Cham, [Online]. Available: https://doi.org/10.1007/978-3-030-68004-6_3.

[5] Z. Bubnicki, “Modern Control Theory,” Springer, 2005, pp. 424, [Online]. Available: https://doi.org/10.1007/3-540-28087-1.

[6] I. Nagratah, “Control Systems Engineering,” 2005, pp. 858.

[7] W. Levine, “The Control Handbook: Control System Fundamentals, Second Edition,” 2017, pp. 1-786, [Online]. Available: https://doi.org/10.1201/b10383.

[8] A. Shchepetov, “Osnovy proyektirovaniya priborov i sistem,” Akademiya, 2019, pp. 458.

[9] A. Kurosh, “Kurs vysshey algebry. Uchebnik dlya vuzov,” Lan', 2020, pp. 432.

[10] I. Kuz'mitskiy and G. Kulakov, “Teoriya avtomaticheskogo upravleniya: uchebnik,” BGTU, Minsk, 2010, pp. 572.

[11] G. Kulakov and K. Artsiomenka, “Compare of Transient Quality in Automatic Control Systems with Classic PID Algorithm and Optimal Regulator,” ENERGETIKA. Proceedings of CIS higher education institutions and power engineering associations, vol. 62, no. 2, pp. 192-200, 2019, [Online]. Available: https://doi.org/10.21122/1029-7448-2019-62-2-192-200.

[12] J. Sevinov, O. Zaripov, and Sh. Zaripova, “The algorithm of adaptive estimation in the synthesis of the dynamic objects control systems,” International Journal of Advanced Science and Technology, vol. 29, no. 5s, pp. 1096-1100, 2020, [Online]. Available: http://sersc.org/journals/index.php/IJAST/article/view/7887.

A Review Paper on IoT Solutions in Health Sector

Vishnu Vardhana Reddy Chintha[1] and Mallikarjuna Reddy Ayaluri[2]
[1]*Department Of CSE, Anurag University, Venkatapur,Ghatkesar Rd, Hyderabad, Telangana, India*
[2]*Department of Artificial Intelligence, Anurag University, Venkatapur,Ghatkesar Rd, Hyderabad, Telangana, India*
ch.vishnuvardhana@gmail.com, mallikarjunreddycse@cvsr.ac.in

Keywords: Healthcare, Internet of Things, Sensors.

Abstract: The past decade has been marked by extensive research in the health services sector and their technological modernization. In particular, the Internet of Things (IoT) has shown potential applications in connecting medical devices, sensors and disparate healthcare professionals to provide quality medical services in faraway places. This has improved patient safety, reduced healthcare costs, improved access to healthcare services, and increased operational efficiency in the healthcare sector. An organization of human body wearable sensors, with each having a very precise identifier, could acquire healthiness statistics that are more detailed than what has previously been available for the regular observed data in scientific/clinical contexts. MIoT statistics when stored, analyzed, and evaluated by comparing with the records from extraordinary human beings, permits the customization and innovation of medical services with significant improvements in reduced costs and results. The current study updates the potential medical services of IoT technologies (MIoT). Here, the progression of the adoption of MIoT was reported in terms of latest technologies, medical care and services to solve various health problems. In summary, the present study serves as a comprehensive source of data on various application areas of MIoT with the aim of helping future researchers interested in this field.

1 INTRODUCTION

IoT is being pushed to play a big role in all facets of healthcare administration because to the rapid development of IoT devices and the desire to make healthcare more affordable, individualized, and watchful. This has raised the expense of healthcare and put a pressure on rural and remote medical facilities. Advances in Internet of Things (IoT) technology and rising needs as well as best practices in smart healthcare have a lot in common.

The healthcare industry has expanded quickly in recent years and is a significant source of wealth and jobs. A few years ago, a physical examination in the hospital was the only way to diagnose illnesses and bodily anomalies. The majority of patients seek their treatment in hospitals. This has increased healthcare costs and has also put a strain on health facilities in rural and remote areas. There are strong synergies between Internet of Things (IoT) technological advances and increasing demands as well as guidelines throughout smart healthcare. The growing development of IoT devices, as well as the intention to make healthcare more cost-effective, customized, and vigilant, are attempting to push IoT to play a significant role in all areas of healthcare administration. It is poised to enable the IoT segment, the Medical Internet of Things (MIoT). MIoT is divided as two subcategories: clinical and individualized/personal. Individualized/personal MIoT incorporates devices consumers use for self-monitoring, such as activity/heart rate trackers, smart clothing, and smartwatches (such as Fitbit [1] and Apple Watch [2]). These general-purpose gadgets are not subject to stringent regulation and are designed for consumer usage without the assistance or advice of a doctor. Clinical MIoT devices are specifically designed for health monitoring with physician guidance and engagement. Examples comprise connected inhalers [4] and smart continuous glucose meters [3]. Strict regulations apply to these devices, which are only permitted for use after being clinically validated. An overview of the developing field of clinical MIoT is given in this study.

A convergence of both technological and social factors is driving the development of Internet of Things technology in the medical field (Figure 1).

Figure 1: Advancement of Internet of Things technology in medical field.

The elderly citizens in Western countries are burdened to have physical facilities, and treatment costs are rising far faster than the general inflation rate [5]. Since many healthcare facilities are already online, they may benefit from the expanded access to high-speed connectivity, low-cost cloud computing, and in-depth data analysis. In this new scenario, MIoT technology is attractive. Not only will it enable the personalization of clinical care delivery, enabling significant cost savings, however it will also make it possible for better results by enhancing flexibility, personalization, and the efficient use of data obtained. MIoT reduces the duration it takes to detect medical conditions [6], provides efficient and high-quality care that reduces hospital costs [7]. By connecting to MIoT, patients can give doctors continuous feedback and track their own improvement to increase the involvement of patient and their happiness. The extensive collection of historical data collection from numerous resources enabled with the introduction of MIoT also creates new opportunities towards extending the common screening approaches used by practitioners (Figure 2).

3 RELATED WORK

Several recent studies have addressed components of MIoT devices. MIoT systems focused on individualized medical centers [8] and infant/kid apps [9] were explored. Technologies in remote health surveillance systems which includes high fever monitoring and aged care [10], as well as technologies for Ambient Assisted Living (AAL), smartphones, and wearable devices have all been thoroughly studied [11]. Focusing with a goal of improving healthcare in underdeveloped nations, MIoT applications in rural healthcare are also reviewed [12]. Other reviews include research papers concentrating on connectivity [13], safety and confidentiality facts of MIoT devices [14,15]. MIoT in home automation [16], elevated surroundings and security sector [17], and MIoT-services for mental illness [18] have also been the focus of several reviews. Unlike the previous studies, this paper specifically focused on clinical aspects of MIoT and covered them comprehensively.

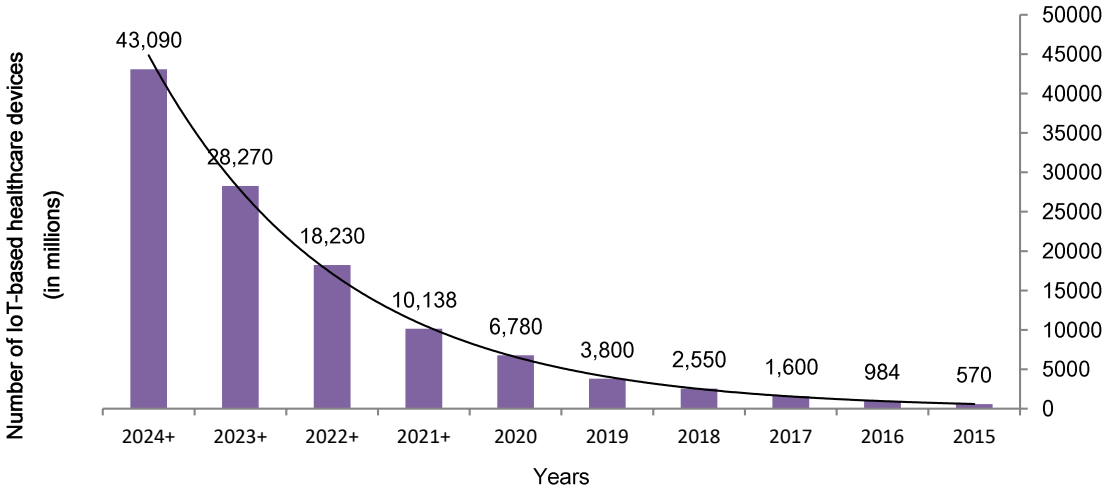


Figure 2: Estimation of IoT devices in future.

Authors in [19], created a process known as We-Care. The system aims to collect data and monitor the overall health of the aged using MIoT technology and provide health alerts to medical teams via sensors worn as watches. The vital signs of the elderly are gathered by We-Care and sent to the medical staff. The system is created to be user-friendly, economical, and energy-efficient. Both medical teams and older people have found this technique to be effective. The technology also supplied a better substitute for conventional healthcare and elderly tracking techniques as well as a dependable way to alert in an emergency. Most significantly, fewer trips to medical facilities are now necessary.

In [20] HEARTFAID aims to develop an innovative service platform for early detection and effective treatment of heart disease in the elderly. Its creative computerized systems can help improve processes such as diagnosis, prognosis, and treatment delivery, and can be applied to EHR, remote monitoring, signal and image processing, and pattern recognition of previously collected data. It offers services such as targeted clinical decision support for heart patients, analysis and decision of interventions for them.

The creation of an early detection and warning system for elderly people who experience a stroke abruptly while exercising was discussed by researchers in [21]. The described solution connects the senior to his smart devices in order to accomplish this purpose. This intelligent device is Self-learning and wirelessly connected to your computer, The system is made up of a number of fundamental parts, including a knowledge base, a real-time patient monitoring interface, self-learning devices, and a number of parts that guard the network against outside threats. Despite the fact that this study is sound in theory, it only looks at older athletes.

Authors in [22] proposed the development and deployment of an MIoT healthcare tracking device. This is quite helpful for many rescue services. This article described the development of 2nd generation Intel GALILEO to support intensive care units, Critical care units, and as well as other services.

The usage of digital handheld electronic components has suggested to improve the standard of patient care (Table 1). Recent research has focused on integrating navigation and location systems into these wheelchairs in [23]. The idea is to connect a real-time obstacle avoidance system with a MIoT steering system. When actual data is captured and examined using image processing techniques, GPS devices can identify barriers.

Table 1: Related work of various healthcare techniques.

No.	Authors	Sensors Used	Diseases	Description
1	Varathar ajan, R et al.[27]	Human Wearable sensor devices	Fever, Diabetes, Blood Pressure and Heart rate	ECG, EMG, pulse oximeter, mica2 motes and SpO2 sensor were used to continuously monitor individual health conditions.
2	Gia, T.N et al.[28]	Arduino Due, Bluetooth, Wi-Fi, ZigBee or 6LoWPAN.	Heart Issues	A health monitoring system that employs secondary network infrastructure based on Internet of Things and contains multiple sensing devices including Wi-Fi and Bluetooth.
3	Rahulamathavan, Y et al.[29]	Body sensors	Health criticality of any patient	Criticality Aware data transmission (CARE) is a scheme proposed in CPS-based medical systems for boosting the computational frequency of patients sensed physiological attribute values.
4	Mukherjee, A et al. [30]	Body sensors	Intensive care patients	The proposed "DROPS" scheme recognizes radio protocols dynamically in a resource smart Wearable medical system.

Wheelchair monitoring has become more involved and simple for patients because of mobile computing [24]. Through the integration of numerous sensors, mobile technology, and cloud computing, the described intelligent wheelchair was created. A mobile app that is part of the system enables patients to communicate with their wheelchairs and carers. The software also enables caretakers to keep an eye on wheelchairs from a distance. In [25] Insulin is a state in which glucose levels in the body rise for an extended period of time. This is one of the most common human ailments. In general, there are three main types of insulin: Type 1, type 2 , gestational. Diseases and their types can be identified using three tests: A randomized plasma glucose test, a fasting

plasma glucose test, and an oral glucose tolerance test. However, the most common diagnostic method for detecting diabetes is "fingerpicking" followed by measurement of blood glucose levels. Recent developments in MIoT technology have been leveraged to create a variety of non-invasive, friendly , affordable, and reliable wearable glucose meters. An M-IoT-based non-invasive blood glucose meter has been proposed for real-time blood glucose monitoring. In order to detect cardiac problems in real time, researchers in [26] presented an Internet of Things (IoT)-based ECG monitoring system made up of a wireless data gathering system and a receiving processor. ECG, EMG, pulse oximeters, mica2 motes, and SpO2 sensors among other sensors in [27] used to continuously monitor patient health parameters. The authors in [28],developed a health monitoring system that makes use of IoT secondary networks and has numerous sensors, including Wi-Fi and Bluetooth. Researchers in [29] ,developed Criticality Aware Data Transmission (CARE) as a method for accelerating the processing of a patient's sensed physiological parameter values in CPS-based healthcare systems. Mukherjee et al. [30] recommended "DROPS," as a method that dynamically chooses radio protocols in a wearable IoT healthcare system with limited energy.

3 CONCLUSIONS

Internet of Things technology is a technological innovation which provides improvements and easy alternatives in areas of healthcare such as hospital record keeping, sample collection and device integration. The use of IoT technology offers an excellent opportunity to lower surgical risks in complex cases The MIoT system was analysed from a variety of perspectives in the recent review. Information on existing health services where MIoT-based solutions have been studied is included in this study. Keeping these ideas in mind, MIoT technology has aided medical practitioners in the monitoring and diagnosis of several health issues, the measurement of numerous health indices, and the provision of diagnostic resources in remote areas. Due to this, the healthcare industry has changed from being controlled by hospitals to become more patient-centered. MIoT addresses the right strategy for information systems to manage global results and promotes hospital system transformations, this information-based service offers up new potential in healthcare. When used properly, MIoT can assist in precisely resolving a various medical challenges such as speed, price, and complexity.

REFERENCES

[1] Fitbit Inc. Fitbit technology. [Online]. Available: https://www.fitbit.com/technology.

[2] Apple Inc. Apple watch. [Online]. Available: https://www.apple.com/watch.

[3] Abbott Diabetes Care Inc. Continuous glucose monitoring–FreeStyle libre system, [Online]. Available: https://www.freestylelibre.us.

[4] Teva Pharmaceutical Industries Ltd. ProAir Digihaler with connected mobile app., [Online]. Available: https://www.proairdigihaler.com.

[5] V. Murthy and A. Okunade, "Is the health care price inflation in US urban areas stationary evidence from panel unit root tests," Journal of Economics, Finance and Administrative Science, vol. 23, no. 44, pp. 77-94, 2018.

[6] S. Reddy, "Can tech speed up emergency room care," 2017, [Online]. Available: https://www.wsj.com/articles/can-tech-speed-upemergency-room-care-490629118.

[7] D. D. Maeng, A. E. Starr, J. F. Tomcavage, J. Sciandra, D. Salek, and D. Griffith, "Can telemonitoring reduce hospitalization and cost of care: a health plan's experience in managing patients with heart failure," Population Health Management, vol. 17, no. 6, pp. 340-344, 2014.

[8] J. Qi, P. Yang, G. Min, O. Amft, F. Dong, and L. Xu, "Advanced internet of things for personalised healthcare systems: A survey," Pervasive and Mobile Computing, vol. 41, pp. 132-149, 2017.

[9] A. S. Yeole and D. R. Kalbande, "Use of internet of things in healthcare: A survey," in Proc. ACM Symp. Women in Research, New York, NY, USA, 2016, pp. 71-76.

[10] A. T. Thakar and S. Pandya, "Survey of IoT enables healthcare devices," in International Conference of Computing Methodologies and Communication (ICCMC), 2017, pp. 1087-1090.

[11] J. J. Rodrigues et al., "Enabling technologies for the internet of health things," IEEE Access, vol. 6, pp. 129-13141, 2018.

[12] E. Anandhavijayalaxmy, V. Dhanalakshmi, and S. Sivasankaran, "Trends and technologies in IoT based healthcare system: A survey," Biometrics and Bioinformatics, vol. 10, no. 3, pp. 41-47, 2018.

[13] M. M. Alam, H. Malik, M. I. Khan, T. Pardy, A. Kuusik, and Y. Le Moullec, "A survey on the roles of communication technologies in IoT-based personalized healthcare applications," IEEE Access, vol. 6, pp. 36611-36631, 2018.

[14] M. A. Al-Garadi, A. Mohamed, A. Al-Ali, X. Du, and M. Guizani, "A survey of machine and deep learning methods for internet of things security," arXiv preprint 1807.11023, 2018.

[15] S. Hameed, F.I. Khan, and B. Hameed, "Understanding security requirements and challenges in internet of things (IoT): A review," Journal of Computer Networks and Communications, 2019.

[16] M. Talal, A. Zaidan, B. Zaidan, A. Albahri, A. Alamoodi, O. Albahri, and et al., "Smart home-based IoT for real-time and secure remote health monitoring of triage and priority system using body

sensors: Multi-driven systematic review," Journal of Medical Systems, vol. 43, no. 3, p. 42, 2019.

[17] M. Thibaud, H. Chi, W. Zhou, and S. Piramuthu, "Internet of things (IoT) in high-risk environment, health and safety (EHS) industries: A comprehensive review," Decision Support Systems, vol. 108, pp. 79-95, 2018.

[18] I. Torre Díez, S.G. Alonso, S. Hamrioui, E.M. Cruz, L.M. Nozaleda, and M.A. Franco, "IoT-based services and applications for mental health in the literature," Journal of Medical Systems, vol. 43, no. 1, p. 11, 2019.

[19] S. Pinto, J. Cabral, and T. Gomes, "We-care: An IoT-based health care system for elderly people," in IEEE International Conference on Industrial Technology (ICIT), pp. 1378-1383, 2017, doi: 10.1109/ICIT.2017.7915565.

[20] M. Maksimovic, V. Vujovic, "Internet of Things Based E-health Systems: Ideas, Expectations and Concerns," in Handbook of Large-Scale Distributed Computing in Smart Healthcare. Scalable Computing and Communications, Springer, 2017.

[21] S. Hong, H. Park, D. Kim, Y. Seo, I. Hussain, and S. J. Park, "Service based healthcare monitoring system for the elderly - physical activity and exercise," in Advances in Human Factors and Ergonomics in Healthcare and Medical, pp. 337-342, 2018.

[22] P. Gupta, D. Agrawal, J. Chhabra, and P.K. Dhir, "IoT based smart healthcare kit," in International Conference on Computational Techniques in Information and Communication Technologies, 2016.

[23] Y. K. Lee, "Real time image processing based obstacle avoidance and navigation system for autonomous wheelchair application," in Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 380-385, Kuala Lumpur, Malaysia, December 2017.

[24] A. Ghorbel, "Cloud based mobile application for remote control of intelligent wheelchair," in Proceedings of the 14th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 1249-1254, Limassol, Cyprus, June 2018.

[25] R. S. Istepanian, "The potential of Internet of m-health Things 'm-IoT' for non-invasive glucose level sensing," in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5264-5266, Boston, MA, USA, March 2011.

[26] M. L. Liu, "Internet of Things-based electrocardiogram monitoring system," Chinese Patent, vol. 102, p. 118, 2012.

[27] R. Varatharajan, G. Manogaran, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system," Future Gener. Comput. Syst., 2018.

[28] T. N. Gia, A. M. Rahmani, B. Negash, and et al., "Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach," Future Gener. Comput. Syst., pp. 641-658, 2018.

[29] Y. Rahulamathavan, A. Roy, C. Roy, S. Misra, and M. Rajarajan, "CARE: Criticality-aware data transmission in CPS-based healthcare systems," in Proc. 2018 IEEE Int. Conf. Commun. Workshops (ICC Workshops), Kansas City, MO, USA, May 20-24, 2018, pp. 1-6.

[30] A. Mukherjee, S. Misra, A. Roy, and C. Roy, "DROPS: Dynamic radio protocol selection for energy-constrained wearable IoT healthcare," IEEE J. Sel. Areas Commun., pp. 338-345, 2020.

Conversion and use of Solar Energy Calculation Methodology for Photovoltaic Systems

Elyor Saitov, Obid Jurayev, Sevara Axrorova, Jushqinbek Ismailov and Bakhtiyor Baymirzaev
Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan
elyor.saitov@mail.ru, obid.jurayev@mail.ru, sevara.axrorova@mail.ru, juruenergy@gmail.com, b.baymirzayev76@gmail.com

Keywords: Calculation, Temperature, Photovoltaic (PV) Module, Solar Cell, Solar Radiation Intensity, Short Circuit Current, Open Circuit Voltage, Electric Power.

Abstract: The needs of the population and industry for electrical energy are limited by oil and gas reserves, which leads to the need to use renewable energy sources. Myanmar is one of the developing countries in Asia. Its specific power consumption is low compared to neighboring countries. Currently, one of the most important tasks of the electricpower industry is to ensure reliable, uninterrupted power supply to all industrial and domestic facilities. At the same time, the development of small solar photovoltaic installations (PMT), operating both in parallel with the grid and in autonomous mode, can improve the power supply of closely located consumers more efficiently and faster than the development of a large power system. Therefore, the work devoted to the improvement of the equipment of a small solar photovoltaic installation is relevant and of great practical importance.

1 INTRODUCTION

RES should transform the energy sectors of the countries of Central Asia, but, most importantly, they should help vulnerable categories of the population who are disconnected from the main power grid or face regular and prolonged power outages. Therefore, the importance of off-grid solar and wind power plants, rooftop solar panels or small and mini-hydro power plants should not be underestimated. Unfortunately, these aspects of the transformation of the energy sector in the countries of Central Asia are often neglected [1,2]. The transition to a diversified energy mix in Central Asia, where the predominant share of energy consumption will be covered by renewable energy, is a long-term initiative and requires comprehensive measures [3]. The purpose of this policy brief, however, is to draw attention to the priority tasks that already today require an urgent response from the authorities in order to stimulate the development of renewable energy in eachof the Central Asian countries. This policy brief examines the development of renewable energy in Central Asia from three parallel perspectives:

- restoration of intra-Central Asian electricity trade in order to effectively use the hydropower potential of the region;
- construction of large solar and wind power plants that will supply electricity directly to the central power grid;
- creation of autonomous objects of renewable energy sources in order to supply electricity to settlements that are disconnected from the central energy system, or those in which there is a regular and prolonged shortage of electricity [4,5].

All photovoltaic (PV) systems can be divided into two types: autonomous and connected to the electrical network. Stations of the second type transfer excess energy to the network, which serves as a reserve in the event of an internal energy shortage.

An autonomous system (Figure 1) generally consists of a set of solar modules placed on the roof, a storage battery (AB), a discharge controller - battery charge, and connecting cables. To obtain alternating voltage, an inverter-converter of direct voltage to alternating voltage is added to the kit.

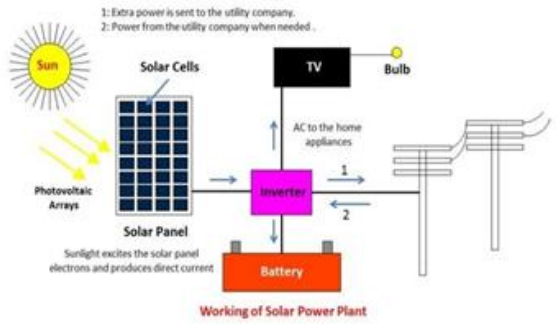


Figure 1: Structural diagram of an autonomous photovoltaic station.

2 METHODS AND MATERIALS

The calculation of FES means the determination of the nominal power of modules, their number, connection scheme; choice of type, operating conditions and capacity of AB; inverter and charge-discharge controller capacities; determination of parameters of connecting cables [6,7].

It is necessary to first determine the total (calculated) power of all consumers connected at the same time. This power is indicated in the product data sheets.

The amount of energy consumed W (kWh) during the time T is determined (1):

$$W = P_{cal} \times T, \quad (1)$$

where $P_{cal.}$ is the calculated load power, kW.

It is believed that solar power supply is economically feasible with a daily energy consumption of up to 4 kWh [8,9].

The calculated load power $P_{cal.}$ is determined by the statistical method [10,11].

According to this method, the calculated load of a group of receivers is determined by two integral indicators: the general average load P_h (kW) and the general standard deviation σ from the (2):

$$P_{cal} = P_h + \beta \cdot \sigma , \quad (2)$$

where β is a static coefficient depending on the distribution law and the accepted probability of exceeding the load P according to the load schedule from the level P_h; σ - standard deviation for the accepted averaging interval.

The standard deviation for the group graph is determined by the (3):

$$\sigma = \sqrt{D_P} = \sqrt{P^2_{nor.\,kvt} - P_h^2} , \quad (3)$$

where $P_{nor.\,kvt}$ is the active rms power, kW.

The root-mean-square (rms) value of the active power of a separate power receiver (EP) for the considered period oftime is determined by the (4):

$$P_{nor.\,kvt} = \sqrt{\frac{\sum p_i^2 t_i}{\sum t_i}} , \quad (4)$$

where $P_{nor.\,kvt}$ is the root mean square value of the active power of the electrical receiver, kW; p_i is the active power consumed by the ED for the considered time interval t_i (determined from the load graph for active power), kW; t_i - time interval for which p_i is determined, min, h.

The statistical method allows you to determine the design load with any accepted probability of its occurrence. In practical calculations, it is quite sufficient to take the probability of exceeding the calculated load from the average by % 5, 0, which corresponds to β=2.5 then (5):

$$P_{cal} = P_h + 2{,}5\ \sigma . \quad (5)$$

Based on the known load power, the inverter power is selected, which, taking into account the losses in the inverter, should be at least 25% more than the calculated power, i.e. (6)

$$P_{ind} \geq 1{,}25\ P_{cal} . \quad (6)$$

The nominal power range of inverters (Figure 2) is 150, 300, 500,800, 1500, 2500, 5000 W. For powerful stations (more than 1 kW), the station voltage is selected at least 48 V, because At higher powers, inverters work better with higher input voltages [12,13].

Figure 2: Appearance of the inverter.

The capacity of the storage battery AB is selected froma standard range of capacities rounded to the side greater than the calculated one. Estimated battery capacity (Ah) is determined by the (7):

$$C_{cal} = \frac{W}{U_{AB} \times \delta} \quad (7)$$

where W is the amount of energy consumed (Wh); U_{AB} – battery voltage; δ is the admissible depth of AB discharge.

When calculating the battery capacity in a fully autonomous mode, it is necessary to take into account the presence of cloudy days, during which the battery must ensure the operation of consumers [14, 15]. For the maximum number of consecutive “days without sun”,

you can take the set number of days during which the battery will feed the load on its own without recharging. The service life of the battery depends on the depth of discharge of the battery δ. The larger δ, the faster the AB will fail. The recommended value of the depth of discharge is 20% (no more than 30%). This means that it is possible to use 20% of the value of the nominal battery capacity. Under no circumstances should the battery discharge exceed 80%. The time of complete discharge of a battery with a capacity of SAB under the influence of a load of power P can be determined (8):

$$T_{cal}=\frac{C_{AB}\cdot 8.5}{P}. \quad (8)$$

The battery capacity also depends on the ambient temperature. The decrease in AB capacity with decreasing temperature takes into account the temperature coefficient K_c, the values of which are taken from the graph (Figure 3). Total required battery capacity (9):

$$C_{AB}=C_{cal.}\cdot K_c \quad (9)$$

useful in the preparation of your submission. If any, the appendix should appear directly after the references without numbering, and not on a new page.

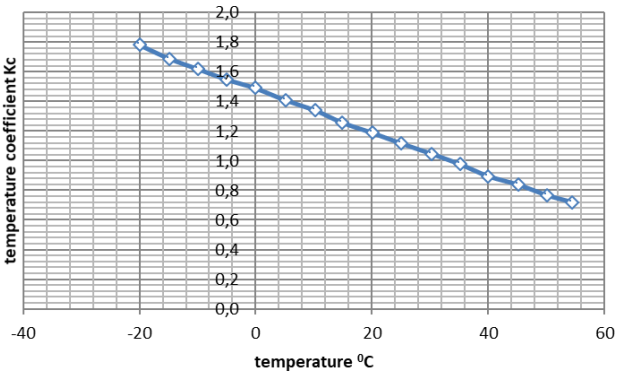


Figure 3: Dependence $K_c = f(T)$.

The cables connecting the inverter and batteries carry a very large current. Therefore, it is necessary to correctly select the cable section based on the maximum currents that the inverter can consume [16]. In order to minimize the voltage drop in the wires between the battery and the inverter, and thereby increase the efficiency of the inverter, the cable should be thick and short enough [17].

When choosing a cable section with a length of 2 m, depending on the power of the inverter and the voltage of the battery, it is recommended to use [18].

In order to calculate the required wire size for a particular installation, you need to know the power of the inverter or charger, or the maximum current flowing through these wires. You also need to know the distance from the battery to the inverter and the DC voltage in the system [19].

Typically, most 12V systems operate in the 11 to 12V range. not more than 0.25 V.

For 24V or 48V systems, the cable length can be 2 or 4 times longer. For a voltage of 220 V, the length can be 20 times longer. To calculate the cross section of the wire, you can also use the (10):

$$r=\frac{\Delta U}{I\cdot L}, \quad (10)$$

where r is the resistivity of the wire, Om/m; ΔU is the maximum allowable voltage drop in the wire, V; I is the passed current, A; L is the total cable length in the system in meters (multiply by 2 for the positive and negative wires) [20].

The resistivity must not be less than 0.0024 Om/m. According to [21], the minimum wire cross-section is determined - 6 mm. The thicker the wire, the less will be the loss in the transfer of energy from the battery to the load [22].

3 RESULTS AND DISCUSSION

The solar module is made in the form of a panel enclosed in an aluminum profile frame (Figure 4). Thepanel is a photovoltaic generator consisting of a glass plate, on the back side of which solar cells are placed between two layers of a sealing (laminating) film, electrically interconnected by metal tires. The modules are made from pseudo-square monosilicon photovoltaic cells (PVCs) coated with an anti-reflection coating. The operating voltage of photovoltaic modules is usually 12V or 24 V. Specifications of solar modules are given in Table 1.

Factory-made photovoltaic cells have a specific power rating, expressed in watts of peak power (Wh) [23]. This is anindicator of their maximum power under standard test conditions, when solar radiation is close to its maximum value of 1000 W/m^2, and the surface temperature of the photocell is 25 °C. In practice, photocells rarely have to work in such conditions. One solar module, depending on the number of plates, has a power of 100 to 3000 watts. During a selected period of time, a module with a power of P_m generates an amount of energy equal to (11):

$$W_M=\frac{P_M\cdot E\cdot \eta}{1000} \quad (11)$$

where W_M is the generated amount of electricity, kWh; 1000 is the maximum value of solar radiation, W/m^2; η – system performance coefficient (0.5 in summer and 0.7 in winter); E is the value of insolation for theselected period for a given latitude, kWh/m^2.

a)

b)

Figure 4: Installed view of solar power plant for utility house: a) Installation of a solar power plant on the ground; b) Installing the solar power plant on the roof of the house.

The value of E for latitude 52.5° is given in Table 1. By dividing the energy consumed by the load during the selected period by the energy generated by one module during the same period, we determine the required number of modules of the photovoltaic system [24] (12):

$$\eta_M = \frac{W}{W_M}. \quad (12)$$

The values of solar radiation power (W/m^2) and incoming solar energy (kWh/m^2) in the settlement are given in Table 1 and Table 2 (according to the weather station).

Table1: Solar radiation power, W/m^2.

No.	month, decade	Month							
		April	May	June	July	August	September	October	November
1.	1	170,72	195,59	121,96	249,99	159,84	139,08	49,02	28,52
2.	2	188,31	139,40	196,20	166,50	165,44	100,44	37,89	23,64
3.	3	180,14	167,48	194,38	243,93	73,40	77,16	52,29	15,34
4.	Average	179,72	167,49	170,85	220,14	132,89	105,56	46,40	22,5
5.	For a month	539,17	502,47	512,54	660,42	398,68	316,68	139,2	67,5

The total capacity for the period from April to November in the region is 3136.66 kW/m^2, solar energy input – 75.28 kWh/m^2

The cost of a solar power plant for an autonomous powersupply system consists of the cost of solar modules, a storage battery, an inverter, an AB charge controller and connecting fittings (wires, switches, fuses, etc.)

Table 2: Solar energy input, kWh/m^2 day.

No.	month, decade	Month							
		April	May	June	July	August	September	October	November
1.	1	170,72	195,59	121,96	249,99	159,84	139,08	49,02	28,52
2.	2	188,31	139,40	196,20	166,50	165,44	100,44	37,89	23,64
3.	3	180,14	167,48	194,38	243,93	73,40	77,16	52,29	15,34
4.	Average	179,72	167,49	170,85	220,14	132,89	105,56	46,40	22,5
5.	For a month	539,17	502,47	512,54	660,42	398,68	316,68	139,2	67,5

The cost of a solar battery is equal to the product of the number of modules and the cost of one module. The costof the inverter depends on its power and type. The costof connecting fittings can be taken approximately equal to 0.1 - 1% of the system cost.

The payback period of FES is determined (13):

$$T = \frac{K}{C \times 8760 \times \xi(Z)}, \quad (13)$$

where K - capital unit costs; C is the electricity tariff; Z - annual operating costs; ξ - rated power utilization factor (14):

$$\xi = \frac{W}{P_H \times 8760} = \frac{\sum_1^{365} E_i \times S \times k_{Ti} \times \eta}{P_{max} \times S \times k_{Tmax} \times \eta \times 8760} = \frac{\sum_1^{365} E_i \times k_{Ti}}{P_{max} \times k_{Tmax} \times 8760}, \quad (14)$$

where E_i is the daily value of daily irradiance in clearweather during the year, kWh/m^2 day; S – working surface of the receiver, m^2; η – Efficiency of converting solar energy into electricalenergy; k_{Ti} –daily value of the Clarity Index; k_{Tmax} – the same - the maximum value P_{max} – maximum value of solar radiation fluxdensity, kW/m^2.

Figure 5: Measuring solar energy at a height of 2 metersand weather conditions at a height of 3 meters.

Tashkent State Technical University named after Islam Karimov, together with the Department of Alternative Energy Sources and Juru Energy, conducted experimentsaccording to the above Figure 5 by definition of albedo in vox-zones.

Performed in accordance with ASTM E1918-16 using two pyranometers in the following versions:

1) 1.5 m height measurement from descending pyranometer to surface;
2) measurement at a height of 2 meters as the worst case scenario.

The weather sensor AWS-600 was installed at a heightof 3 m, in order to obtain the necessary data, it was decided to choose a height of 3 m from the ground, which was also recommended manufacturer's representative and instruction manual.

To measure solar radiation and weather conditions, 3 sets of the following instruments were used:

1) 2x SMP-11V/CMP-10 pyronometers (Kipp&Zonen);
2) 1x Tripod (for mounting sensors at appropriate heights);
3) 1x Logbox SE-Logbox (Kipp&Zonen);
4) 1x Weather station WS-600 (Lufft);
5) 1x Power supply kit (12V battery).

Temperature dependence of sensitivity show in Figure 6.

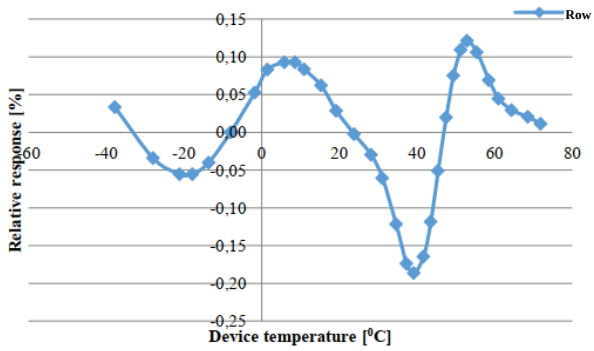


Figure 6: Temperature dependence of sensitivity.

Conventional temperature dependence measurements during final inspection. The pyranometer is installed inside the climate chamber and is illuminated by a white light source at normal incidence. A CMP22 pyranometer outside the chamber is used to monitor lamp stability.

The pyranometer is tested in the temperature range from 70°C to 40°C in steps of 10°C. The relative temperature dependence is shown below. The measurement error of this characteristic is ±0.1% (k=2).

3 CONCLUSIONS

A methodology has been developed for calculating the power of autonomous solar power plants and its elements, which allows taking into account the change in load during the day. Taking into account the change in the load schedule at the interval of the night time of the day allows you to accurately determine the required capacity of the battery. The expression for determining the required battery capacity of an autonomous solar power plant can be easily generalized for any number of changes in the steps of the load curve. The developed method for calculating the power of autonomous solar power plants makes it possible to exclude unjustified overestimation of the power of the elements of the power plant and the rise in the cost of an autonomous solar power plant.

ACKNOWLEDGMENTS

The work was financially supported by the Ministry of Innovative Development of the Republic of Uzbekistan within the framework of the project F-OT-2021-497 - “Development of the scientific foundations for the creation of solar cogeneration plants based on photovoltaic thermal batteries”. The authors express their gratitude to the professor of Tashkent State Technical University N.F. Zikrillayev for his scientific and practical help in writing this article.

REFERENCES

[1] B. E. Umirzakov, D. A. Tashmukhamedova, M. A. Tursunov, Y. S. Ergashov, and G. K. Allayarova, “Escape Depth of ‘Secondary and Photoelectrons from CdTe Films with a Ba Film’,” Technical Physics, vol. 64, no. 7, pp. 1051-1054, 2019, doi: 10.1134/ S1063784219070260.

[2] Y. S. Ergashov, D. A. Tashmukhamedova, and B. E. Umirzakov, “On the synthesis of nanoscale phases of metal silicides in the near-surface region of silicon and the study of their electronic structures by passing light,”

Journal of Surface Investigation, vol. 11, no. 2, pp. 480-484, 2017, doi: 10.1134/S1027451017020252.

[3] S. V. Mitrofanov, D. K. Baykasenov, and A. U. Nemaltsev, "Operational experience of a Solar Power Plant with a Dual-Axis Solar Tracking System in the conditions of the Southern Urals," E3S Web of Conferences, Jan. 2019, doi: 10.1051/e3sconf/20196901015.

[4] R. Suwarno, "Optimization Circuit Based Buck-Boost Converter for Charging the Solar Power Plant," Indonesian Journal of Electrical Engineering and Computer Science, vol. 6, no. 2, pp. 254-258, May 2017, doi: 10.11591/ijeecs.v6.i2.pp254-258.

[5] E. B. Saitov, Sh. Kodirov, B. M. Kamanov, N. Imomkulov, and I. Kudenov, "Increasing the Efficiency of Autonomous Solar Photovoltaic Installations for Power Supply of Agricultural Consumers," AIP Conference Proceedings, vol. 2432, 040036, 2022, doi: 10.1063/5.0058767.

[6] F. Zikrillayev, E. B. Saitov, J. B. Toshov, B. K. Ilyasov, and M. B. Zubaydullayev, "A Software Package for Determining the Optimal Composition and Parameters of a Combined Autonomous Power Supply System Based on Renewable Energy Sources," AIP Conference Proceedings, vol. 2432, 020021, 2022, doi: 10.1063/5.0058098.

[7] E. B. Saitov, J. B. Toshov, A. O. Pulatov, B. M. Botirov, and Y. M. Kurbanov, "Networked interactive solar panels over the roof photovoltaic system (PVS) and its cost analysis at Tashkent state technical University," E3S Web of Conferences, vol. 216, 01133, 2020, doi: 10.1051/e3sconf/202021601133.

[8] S. Sharma and N. Sengar, "Review of solar PV training manuals and development of survey based solar PV system training formats for beginners," Solar Energy, vol. 241, no. 2, pp. 72-84, Jul. 2022, doi: 10.1016/j.solener. 2022.05.055.

[9] T.G. Zoryna, S.A. Aliaksandrovich, Y.S. Valeeva, M.V. Kalinina, L.E. Ilikova, and E.Y. Suvonovich, "Measures to stimulate the development of electric transport as a tool for the development of the territory," in 2022 8th International Conference on Energy Efficiency and Agricultural Engineering (EE and AE), Proceedings, 2022, pp. 1-6, doi: 10.1109/EEAE53789.2022.9831360.

[10] M. Abdel-Basset, D. El-Shahat, R.K. Chakrabortty, and M. Ryan, "Parameter estimation of photovoltaic models using an improved marine predators algorithm," Energy Conversion and Management, vol. 227, pp. 113491, 2021.

[11] R. Araneo, U. Grasselli, and S. Celozzi, "Assessment of a practical model to estimate the cell temperature of a photovoltaic module," International Journal of Energy and Environmental Engineering, vol. 5, no. 1, pp. 1-16, 2014.

[12] V.Z. Antonopoulos, D.M. Papamichail, V.G. Aschonitis, and A.V. Antonopoulos, "Solar radiation estimation methods using ANN and empirical models," Computers and Electronics in Agriculture, vol. 160, pp. 160-167, 2019.

[13] M.S. El Nozahy, M.M.A. Salama, and R. Seethapathy, "A probabilistic load modelling approach using clustering algorithms," 2013 IEEE Power and Energy Society General Meeting, pp. 1-5, 2013.

[14] K. Kant, A. Shukla, A. Sharma, and P.H. Biwole, "Thermal response of poly-crystalline silicon photovoltaic panels: numerical simulation and experimental study," Solar Energy, vol. 134, pp. 147-155, 2016.

[15] A. Nouar Aoun, "Methodology for predicting the PV module temperature based on actual and estimated weather data," Energy Conversion and Management: X, vol. 14, pp. 100182, 2022.

[16] P. Midya, P.T. Krein, R.J. Turnbull, R. Reppa, and J. Kimball, "Dynamic maximum power point tracker for photovoltaic applications," 27th Annual IEEE Power Electronics Specialists Conference Record, pp. 1710-1716, 1996.

[17] 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," Renewable Energy, vol. 132, pp. 284-295, 2019.

[18] International Electrotechnical Commission, "Ampacity charts," [Online]. Available: https://www.cerrowire.com/products/resources/tables-calculators/ampacity-charts/.

[19] Y. Chen, Y. Liu, Z. Tian, Y. Dong, and et al., "Experimental study on the effect of dust deposition on photovoltaic panels," Energy Procedia, vol. 158, pp. 483-489, 2019.

[20] M.M. Fouad, L.A. Shihata, and E.S.I. Morgan, "An integrated review of factors influencing the performance of photovoltaic panels," Renewable and Sustainable Energy Reviews, vol. 80, pp. 1499-1511, 2017.

[21] National Electric Code, "Litz wire," [Online]. Available: https://www.elektrisola.com/en-us/Litz-Wire/Info.

[22] E. Skoplaki and J.A. Palyvos "On the temperature dependence of photovoltaic module electrical performance: a review of efficiency/power correlations". Sol. Ener., vol. 83, pp. 614-24, 2009.

[23] M.A. Mamun , M.M. Islam, M. Hasanuzzaman, and J. Selvaraj, "Effect of tilt angle on the performance and electrical parameters of a PV module: Comparative indoor and outdoor experimental investigation". Ener. and Buil. Env., vol. 3, pp. 278-290, 2022.

[24] H. Nisar, A. K. Janjua, H. Hafeez, S. Shakir, N. Shahzad, and A. Waqas. "Thermal and electrical performance of solar floating PV system compared to on-ground PV system-an experimental investigation", Sol. Ener., vol. 241, pp. 231-147, 2022.

3D Scene Reconstruction with Neural Radiance Fields (NeRF) Considering Dynamic Illumination Conditions

Olena Kolodiazhna[1], Volodymyr Savin[1], Mykhailo Uss[2] and Nataliia Kussul[1,3,4]
[1]*Institute of Physics and Technology, Igor Sikorsky Kyiv Polytechnic Institute, Peremohy Avenue 37, Kyiv, Ukraine*
[2]*Department of Information-Communication Technologies, National Aerospace University, Chkalova Str. 17, Kharkiv, Ukraine*
[3]*Department of Space Information Technologies and System, Space Research Institute National Academy of Science of Ukraine an State Space Agency of Ukraine, Glushkov Avenue 40, Kyiv, Ukraine*
[4]*Anhalt University of Applied Sciences, Bernburger Str. 57, Köthen, Germany*
kolodyazhna.lena@gmail.com, vladimyr.savin@gmail.com, mykhail.uss@gmail.com, nataliia.kussul@gmail.com

Keywords: Computer Vision, Neural Radiance Fields, Dynamic Illumination, View Synthesis, 3D Scene Reconstruction.

Abstract: This paper addresses the problem of novel view synthesis using Neural Radiance Fields (NeRF) for scenes with dynamic illumination. NeRF training utilizes photometric consistency loss that is pixel-wise consistency between a set of scene images and intensity values rendered by NeRF. For reflective surfaces, image intensity depends on viewing angle and this effect is taken into account by using ray direction as NeRF input. For scenes with dynamic illumination, image intensity depends not only on position and viewing direction but also on time. We show that this factor affects NeRF training with standard photometric loss function effectively decreasing quality of both image and depth rendering. To cope with this problem, we propose to add time as additional NeRF input. Experiments on ScanNet dataset demonstrate that NeRF with modified input outperforms original model version and renders more consistent 3D structures. Results of this study could be used to improve quality of training data augmentation for depth prediction models (e.g. depth-from-stereo models) for scenes with non-static illumination.

1 INTRODUCTION

3D scene reconstruction is a long-standing problem in computer vision consisting in understanding 3D structure of a scene given its 2D images. It is applied to diverse domains, including Augmented Reality (AR) and Virtual Reality (VR). For example, 3D scene reconstruction enables occlusions and collision processing between augmented content and the physical world for natural, seamless, and realistic interactions in AR. Multiple methods and tools, including Markov Random Fields [1], local stereo matching algorithms [2, 3] and deep neural networks [4], are applied to solve 3D scene reconstruction task. The complexity of this task is due to the need for simultaneous consistent reconstruction of global scene structures and their local details that require massive computations and a large amount of data. Availability of accurate and reliable data is crucial in deep neural networks training.

Manual data collection with further annotation and synthetic data generation are common approaches for training datasets acquisition. The process of manual collection is very time-consuming and expensive: one need to collect GT depth data, image data along with accurate camera poses. Consequently, this process also requires additional specific equipment, for example, depth cameras. Fully synthetic data cannot completely replace real data in neural networks training. In order to guarantee reasonable performance in real operation conditions, models obtained after fitting on such fully synthetic data should be finetuned on real data from the target domain [5].

One of the recent advances in view synthesis with reliable results is Neural Radiance Fields (NeRF) [6]. It is an MLP (multilayer perceptron) network that can generate novel views of the scene given a limited amount of pictures of the scene with corresponding camera poses during the training process. NeRF optimizes underlying continuous volumetric function using a sparse set of input views [6]. This method may be used for new data generation and augmenting training datasets for depth prediction neural networks. In, for example, [7] authors uses synthetic im-

ages generated by NeRF to solve localization tasks. They demonstrate that additional synthetic data improves the accuracy of regression of the camera's position. One of the advantages of this MLP network is that it can synthesize not only RGB images but also depth maps that are important for depth estimation networks. And as mentioned before, 3D scene reconstruction should be close to the original environmental structures, so it requires reliable depth maps.

During training NeRF optimizes photometric loss function that is equal to pixel-wise difference between intensities of original and generated images. However, this loss function has its limitations in dynamic scenes and scenes with illumination changes: such data violate the brightness consistency assumption important for such photometric loss functions due to its dependence on image intensity. Discussed loss function is also a common loss function that is minimized in depth estimation networks that are trained in unsupervised mode (e.g. in [8, 9]). Authors of [10] analyze the problem of using photometric consistency loss functions for datasets with bad or dynamic illumination. They demonstrate that standard photometric loss function fails for such data.

The main contribution of this work is using a time variable as a sequential image index added to the NeRF model input parameters for static scenes but with dynamic illumination. This modification considers some illumination changes across all datasets and allows their compensation. To evaluate its influence on scene reconstruction, the depth generation quality of NeRF was measured by calculating relative depth errors on two ScanNet data [11] scenes with some changes of illumination and with the presence of reflections.

The rest of this paper is organized as follows. Section 2 reviews related works in generating image and depth data and scene reconstruction with illumination changes. Section 3 discusses the problem of NeRF training for scenes with dynamic illumination and proposes modifications of NeRF architecture and loss function. Section 4 discusses experiments that demonstrate the effect of the proposed modifications. Finally, Section 5 summarizes main article's contributions and future research directions.

2 RELATED WORK

Having a certain number of images of the scene captured from different positions, NeRF reconstructs its entire 3D structure and enables the synthesis of novel views. This method is not the first to address this problem. Other variants of generating novel views include, for example, 3D grid-based optimization of the representation of the scene [12] or methods optimized by neural networks that map XYZ coordinates into sign distance functions (SDF) [13, 14]. However, due to their computational complexity, these methods require high-quality ground truth 3D data and have low-scaled capability focused on generating high-resolution images. On the contrary, NeRF optimizes the representation of the scene in the form of a continuous differentiable function that allows training model in an end-to-end manner.

However, the original NeRF model [6] also has drawbacks due to, for example, time consuming and requirement of accurate camera poses. Authors of BARF model [15] mitigate exact camera poses requirement by adding them to the optimization process. Both NeRF and BARF models have common limitation of supporting only a single scene. They both fail to generalize other environments and allow image synthesis only for the specific scene used in the training process. There are some NeRF models such as [16, 17] that overcome the mentioned problem by using MVS-based (multi-view stereo) approaches.

In general, models that are based on NeRF assume that scenes are static, but it is not always true. Neural Scene Flow Fields (NSFF) [18] is one of the NeRF representatives that enables the optimization for dynamic scenes. The authors of this model modified the basic representation of the scene as NeRF, considering the dynamic conditions of the environment. As a result, obtained new representation of the scene, NSFF, simulates the dynamic scene as a time variable continuous function of the environment representation, geometry, and movement of the 3D scene. Such an approach enables the interpolation of changes both in space and time. In contrast to this work, in this paper we consider time dependence NeRF for static scenes with dynamic illumination.

3 METHODOLOGY

3.1 NeRF Training Loss

NeRF has recently gained success in generating novel views for complex scenes [6]. It represents a scene using the fully-connected deep network. The input to the model is a 5D vector-valued function which arguments are spatial location (x, y, z) and viewing direction (θ, ϕ). The output is a volume density σ and RGB color $\vec{c}$. The model can be written as [6]:

$$F_w : (\vec{x}, \vec{d}) \rightarrow (\vec{c}, \sigma) \quad (1)$$

To train NeRF, we need a dataset with RGB images of the scene, camera poses, and camera intrinsic parameters. The training process involves rendering corresponding views of the scene and minimization of the photometric loss between observed and synthesized images.

The rendering process is illustrated in Figure 1. First, we march camera rays through each pixel of the image and sample some points. Then, these points are fed to the MLP network that predicts color and density for each of them. At the last stage, classical volumetric rendering [19] is used to aggregate all colors and densities for each sampled point and get the final result for the pixel.

During training, NeRF minimizes photometric loss function. Generally, it uses two networks: coarse and fine [6]. But for simplicity in our work, we only consider the first coarse subnetwork. Given M images $(I_1,...,I_M)$, the goal of NeRF training is optimize the following synthesis-based objective:

$$L_F = \sum_{i=1}^{M} \sum_{u} ||\hat{I}_i(u;w) - I_i(u)||_2^2, \tag{2}$$

where w is the network parameters that also depend on the view directions, u denotes pixels coordinates, $\hat{I}(u;w)$ is the synthesized RGB value at pixel u.

3.2 Photometric Loss Limitations

One of the characteristics of real data is dynamic illumination. Usually, existing datasets have static scenes without changes of lighting. Illumination changes may be caused by some reasons e.g., from external sources such as the sun or from car lights. Also, these changes may be due to camera exposures. NeRF models take into account positions and input ray directions that can compensate effects caused by reflective surfaces but they do not consider time variable. Time dependence is important for scenes with dynamic illumination. Consequently, standard NeRF models may fail in rendering a correct color for images from such datasets. Also, as it shown in [10], depth estimation networks that use photometric-consistency loss functions fail in recovering 3D scene structure for datasets with bad or dynamic illumination. Given this, NeRF-based models may have the same problems as they also use photometric loss function.

3.3 Model Modifications

3.3.1 Depth Loss

The original NeRF model and all its modifications predict color along with the density σ that can be interpreted as an opacity of the objects. Using the obtained density, we can calculate distances to objects. During NeRF training, the predicted depth maps are not further optimized. This leads to the incorrectly predicted distances to the objects and, thus, low quality of the 3D scene reconstruction. There are some datasets that have (incomplete) depth maps that can be used to improve NeRF quality. To do this, we propose to add an additional loss function, which is defined as the MSE error between the GT (ground truth) values of the distance maps and the predicted ones:

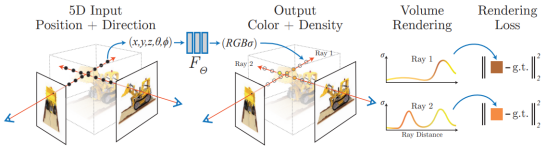


Figure 1: NeRF rendering process [6].

$$L_D = ||D - \hat{D}||_2^2, \tag{3}$$

where D - GT depth maps, $\hat{D}$ - predicted depth maps. The overall loss function takes the following form:

$$L = \sum_{i=1}^{M} \sum_{u} ||\hat{I}_i(u;w) - I_i(u)||_2^2 + ||D - \hat{D}||_2^2, \tag{4}$$

where M denotes number of images, u is pixel coordinates, $\hat{I}_i(u;w)$ is synthesized RGB color in pixel u for image i, $I_i(u)$ is GT RGB color in pixel u for image i, D is GT depth maps, $\hat{D}$ is predicted depth maps by NeRF.

3.3.2 Time as Additional Variable of NeRF Model

In order to model dynamic illumination, we propose to use time t as supplemental input variable to the network. Each of the elements of t corresponds to the sequential index of the input image for NeRF training. Time and images indices have linear dependence. This is due to the fact that the processed dataset is a video sequence with a fixed value of frames per second (fps). Variable t is additionally normalized to the $[0,1]$ interval, and positional encoding is applied. This variable can be added in two variants to the model. The first one corresponds to time t added with 3D coordinates denoted as $(\vec{x},t)$ and the second one is time t added with input rays directions denoted as $(\vec{d},t)$.

The modified NeRF model is defined as:

$$F_w : (\vec{x},\vec{d},t) \rightarrow (\vec{c},\sigma) \tag{5}$$

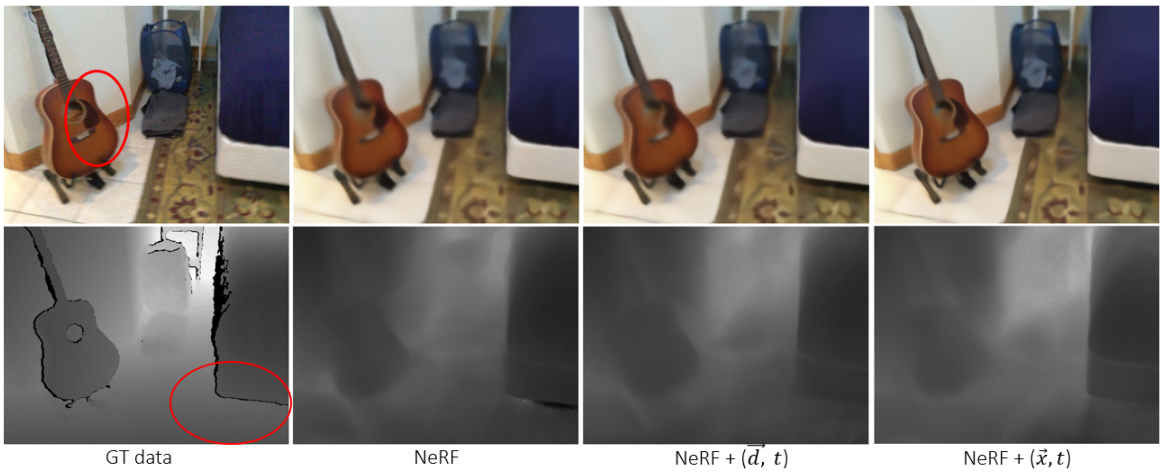


Figure 2: Qualitative comparisons for scene0000_00. First line corresponds to RGB images and second line to depth maps. Adding time to the NeRF model improve both quality of RGB images and depth maps as shown in highlighted regions.

4 EXPERIMENTS

4.1 Dataset Description

ScanNet [11] is an RGB-D dataset that consists of 2.5 million images collected in more than 1500 different indoor locations. Camera intrinsics and extrinsics (camera poses) are provided for each scene. In this paper, we use two ScanNet scenes: scene0000_00 and scene0005_01. For these scenes illumination changes with viewing angle because of reflective surfaces and also with time. Time dynamics is most probably caused by the camera auto exposure. Scene0000_00 consists of 5577 images and scene0005_01 consists of 1449 images. For experiments we use only a part of the scene0000_00 that contains 500 images. We select only a part that contains illumination changes.

4.2 Experiment Description. Metrics

To demonstrate the effect of our proposed modifications we conducted the following experiments:

1) Train original NeRF model.
2) Train original NeRF model with added loss function.
3) Train original NeRF model with both added depth loss function and time variable.

Training configuration including train/validation split, number of sampled points, number of random rays is the same as in [15].

We evaluate our models using two metrics: MAE (mean absolute error) and MARE (mean absolute relative error). We measure errors between GT depth and predicted one.

MAE can be calculated as:

$$MAE = \frac{1}{N}\frac{1}{n}\frac{1}{m}\sum_{k=1}^{N}\sum_{i,j}^{n,m}|y_{k,i,j}) - \hat{y}_{k,i,j}|, \quad (6)$$

where $y_{k,i,j}$ - GT depth value for (i, j) pixel, $\hat{y}_{k,i,j}$ - predicted depth value for (i, j) pixel, N - number of images in the dataset, n - image height, m - image width.

And MARE can be calculated as:

$$MARE = \frac{1}{N}\frac{1}{n}\frac{1}{m}\sum_{k=1}^{N}\sum_{i,j}^{n,m}\frac{|y_{k,i,j} - \hat{y}_{k,i,j}|}{y_{k,i,j}} * 100\%, \quad (7)$$

where $y_{k,i,j}$ - GT depth value for (i, j) pixel, $\hat{y}_{k,i,j}$ - predicted depth value for (i, j) pixel, N - number of images in the dataset, n - image height, m - image width.

4.3 Experiment Results

The performance metrics of trained models in different modes are shown in Table 1. For the scene0000_00 NeRF models were trained without adding depth loss function. Base model results in low performance, achieving 28.1% relative depth error. Adding time variable allow us to improve the model quality by 8-11%. Qualitative results for this dataset can be found in Figure 2. For dataset scene0005_01 we analyze both modifications: depth loss and additional time t variable. Modification with depth loss reduces the relative depth error from 30% to 1.88%. Adding time variable further improves quality of synthesized color images as well as depth maps. It is especially shown for case of adding time variable with the input 3D coordinates $(\vec{x}, t)$. Relative depth error in this mode is equal to 0.93%. Qualitative results for dataset

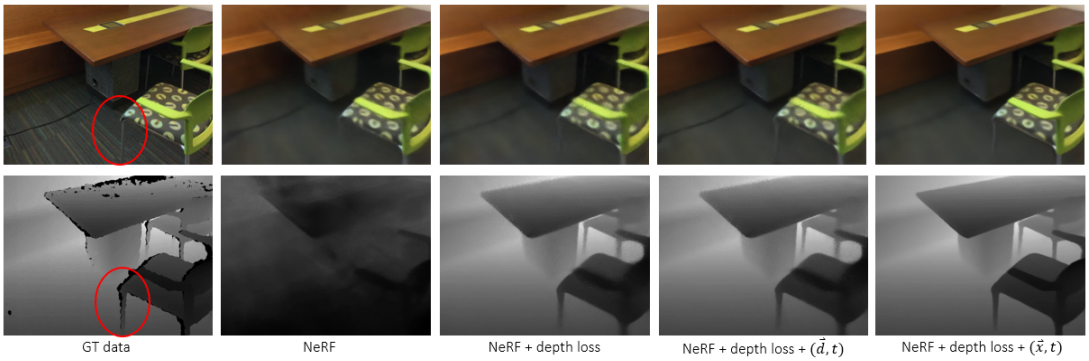


Figure 3: Qualitative comparisons for scene0005_01. First line corresponds to RGB images and second line to depth maps. Adding depth loss function as well as time variable to NeRF model improve synthesis quality and allow model to better generate thin structures of the scene (highlighted region).

scene0005_01 can be found in Figure 3. Visual comparison of generated image and depth data shows us quality improvement with our mode modifications.

Table 1: Metrics for ScanNet dataset.

Dataset	Train. mode	MAE	MARE
scene0000_00	base model	0.686m	28.1%
scene0000_00	$(\vec{d})$	0.498m	20%
scene0000_00	$(\vec{x},t)$	0.429m	17.3%
scene0005_01	base model	0.577m	30%
scene0005_01	depth loss	0.03m	1.881%
scene0005_01	depth loss + $(\vec{d})$	0.03m	1.875%
scene0005_01	depth loss + $(\vec{x},t)$	0.015m	0.93%

5 CONCLUSIONS

In this paper, we study how dynamic illumination affect quality of a scene representation by NeRF model. Dynamic illumination can be caused by illumination sources with power changing in time (sunlight in cloudy weather), light sources switched on or off during scene acquisition, or by camera automatic exposure. We argue that such changes cannot be modelled by standard NeRF using position and viewing angle direction as inputs and lead to rendering quality degradation. To cope with this problem, we propose to extend NeRF input with additional time variable. This idea was previously used for scenes with dynamic objects, we demonstrate that same approach is useful for static scenes with dynamic illumination. Experiments on ScanNet dataset show that extending NeRF input with time variable leads to improvement of quality of synthesized images (e.g. for small structures) and to relative depth error decreasing by 10-28%. From the practical point of view, results of this work can be used to improve quality of data augmentation for training depth prediction models where quality of both image and depth rendering is highly important.

REFERENCES

[1] L. Tardon, I. Barbancho, and C. Alberola-López, Markov Random Fields in the Context of Stereo Vision, 01 2011.

[2] S. Zhu and L. Yan, "Local stereo matching algorithm with efficient matching cost and adaptive guided image filter," vol. 33, no. 9, 2017. [Online]. Available: https://doi.org/10.1007/s00371-016-1264-6

[3] M. Bleyer, C. Rhemann, and C. Rother, "Patchmatch stereo - stereo matching with slanted support windows," in BMVC, January 2011. [Online]. Available: https://www.microsoft.com/en-us/research/publication/patchmatch-stereo-stereo-matching-with-slanted-support-windows/

[4] H. Laga, L. V. Jospin, F. Boussaid, and M. Bennamoun, "A survey on deep learning techniques for stereo-based depth estimation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 4, pp. 1738–1764, apr 2022.

[5] J. Watson, O. M. Aodha, D. Turmukhambetov, G. J. Brostow, and M. Firman, "Learning stereo from single images." Berlin, Heidelberg: Springer-Verlag, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-58452-8_42

[6] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing

scenes as neural radiance fields for view synthesis," in ECCV, 2020.

[7] A. Moreau, N. Piasco, D. Tsishkou, B. Stanciulescu, and A. de La Fortelle, "Lens: Localization enhanced by nerf synthesis," 2021. [Online]. Available: https://arxiv.org/abs/2110.06558

[8] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," 2016. [Online]. Available: https://arxiv.org/abs/1609.03677

[9] H. Li, A. Gordon, H. Zhao, V. Casser, and A. Angelova, "Unsupervised monocular depth learning in dynamic scenes," in Proceedings of the 2020 Conference on Robot Learning, ser. Proceedings of Machine Learning Research, J. Kober, F. Ramos, and C. Tomlin, Eds., vol. 155. PMLR, 16–18 Nov 2021, pp. 1908–1917. [Online]. Available: https://proceedings.mlr.press/v155/li21a.html

[10] K. Wang, Z. Zhang, Z. Yan, X. Li, B. Xu, J. Li, and J. Yang, "Regularizing nighttime weirdness: Efficient self-supervised monocular depth estimation in the dark," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2021, pp. 16 055–16 064.

[11] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "Scannet: Richly-annotated 3d reconstructions of indoor scenes," in Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.

[12] M. Waechter, N. Moehrle, and M. Goesele, "Let there be color! large-scale texturing of 3d reconstructions," in Computer Vision – ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 836–850.

[13] C. M. Jiang, A. Sud, A. Makadia, J. Huang, M. Nießner, and T. Funkhouser, "Local implicit grid representations for 3d scenes," in Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2020.

[14] E. Penner and L. Zhang, "Soft 3d reconstruction for view synthesis," vol. 36, no. 6, 2017.

[15] C.-H. Lin, W.-C. Ma, A. Torralba, and S. Lucey, "Barf: Bundle-adjusting neural radiance fields," in IEEE International Conference on Computer Vision (ICCV), 2021.

[16] A. Chen, Z. Xu, F. Zhao, X. Zhang, F. Xiang, J. Yu, and H. Su, "Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 14 124–14 133.

[17] Y. Wei, S. Liu, Y. Rao, W. Zhao, J. Lu, and J. Zhou, "Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo," in ICCV, 2021.

[18] Z. Li, S. Niklaus, N. Snavely, and O. Wang, "Neural scene flow fields for space-time view synthesis of dynamic scenes," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[19] J. T. Kajiya and B. V. Herzen, "Ray tracing volume densities," Proceedings of the 11th annual conference on Computer graphics and interactive techniques, 1984.

Prospects for the Development of the use of Pumped Storage Power Plants in the Energy System of the Republic of Uzbekistan

Murodilla Mukhammadiev, Kurbon Dzhuraev and Faxriddin Nosirov
Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan
djqs1983@outlook.com, nosirov@mail.ru

Keywords: Pumped Storage Power Plants, Electric Power System, Renewable Energy Sources, Hydropower Complex, Agility, Energy Storage, Capacity, Electricity, Energy-Economic Efficiency, Capital Investment, Annual Costs, Saving Fuel Resources, Environmental Factor.

Abstract: The article discusses the need to use pumped storage power plants (PSPP) to increase the reliability, stability, maneuverability and energy-economic efficiency of the electric power system (EPS) with power plants based on renewable energy sources (RES), which make it possible to cover the minimum and maximum daily load schedules, and also regulating the capacities of solar-wind power plants, for which the tasks of equalizing power fluctuations and energy storage are especially important. For a technical and economic assessment and determination of the optimal parameters of a PSPP, a mathematical model is proposed under the conditions of its construction with existing structures, taking into account its water-use purpose - low-pressure, daily and seasonal operating modes, with the criterion of minimum investment, payback period, annual costs for PSPP and environmental factors with the maximum generated peak power and fuel savings. On the basis of the proposed mathematical model, a methodology and program have been developed to determine the optimal energy and economic indicators of a PSPP.The schemes of using pumped storage power plants at four energy and water facilities, that is, the Tuyamuyun hydroelectric complex, Arnasai, Talimarjan and Khodjikent reservoirs, were considered, and for these facilities, on the basis of the developed methodology and program, their energy and economic parameters of the PSPP were determined.

1 INTRODUCTION

In the global energy sector, when diversifying the energy balance through the use of traditional power plants and alternative energy sources, including renewable ones, the most important aspects are increasing the reliability, stability, maneuverability and energy-economic efficiency of the electric power system (EPS) and power plants based on renewable energy sources (RES) - solar, wind, hydraulic, etc.

Large-scale work is being carried out in Uzbekistan to develop the energy industry, taking into account the best international experience and current trends in the development of the world electric power industry [1,2]

To date, the total installed capacity of Uzbekistan's power plants is about 14,140.6 MW. About 85.7 % of electricity generation is accounted for by thermal power plants (12,129 MW), mainly powered by natural gas, 13.3 % by HPP (1,878 MW) and 1 % by the station unit (132.9 MW) [1]. The operated power plants of Uzbekistan generate more than 62.9 billion tons per year. kWh of electricity, and the consumption is more than 62.0 billion rubles. kWh [1-4].

At the same time, certain work is being carried out to develop the use of renewable energy, primarily on the use of hydropower potential, solar and wind energy.

The tasks of RES development are specified in Government Resolutions and Decree No. PP-2947 "On the program of measures for the further development of hydropower for 2017-2021" 2.05.2017 [5], No. PP-3012 "On the program of measures for the further development of renewable energy, improving energy efficiency in economic and social sectors for 2017-2021" 26.05.2017 [6] and UP No. 5044 of May 18, 2017 "On the formation of the joint-stock company "Uzbekhydroenergo"" [7].

Thus, the need for a consistent increase in the use of renewable energy sources, the creation of new environmentally friendly generating capacities on their basis, the provision of technical and technological re-equipment of existing hydroelectric power plants based on the use of modern technologies was noted [5-6].

These resolutions are aimed at improving the efficiency of water resources management, taking into account international best practices, improving the balance of energy resources and ensuring, on this basis, the needs of enterprises and the population for electricity [5-7].

Among the priority issues of the development of the electric power industry of Uzbekistan in such areas as thermal power plants (TPP), nuclear power and RES [2, 3, 8]. Major investment projects are being worked out to increase the generating capacity of the energy system in the period up to 2030, including [2, 3, 8]:

1) construction of energy-efficient thermal generating capacities with a total volume of about 15 GW;
2) construction of modern CCGT units of combined cycle of electricity production with a total capacity of about 9.5 GW (at the Syrdarya, Navoi, Talimarjan, Takhiatash, Turakurgan TPP, etc.);
3) construction of regulating capacities on the basis of maneuverable gas-piston installations, aviation GTI, modern energy storage systems with a total capacity of about 3.2 GW (at the Syrdarya, Tashkent TPP, Mubarak TPP, etc.);
4) construction of new objects RES (solar, wind and hydroelectric power plants) with a total capacity of more than 8,400 MW, in order to increase the specific share of RES in the total generation volume to 25%.
5) construction of a NPP with a capacity of 2,400 MW.

However, along with the positive aspects of the development and saturation of the energy system of Uzbekistan with large power plants of huge capacity, the difficulties with covering the minimum and maximum daily loads of the EPS are compounded (Figure 1). In addition, the demand for electric energy was not fully met, the deficit was about 9.4 percent of the demand [9]. The limited range of power control of large-block units and the lack of the possibility of frequent starts and stops without a sharp decrease in the reliability and efficiency of the power equipment of power plants makes it difficult to cover the uneven part of the electrical load schedules and leads to the need to increase maneuverable capacities.

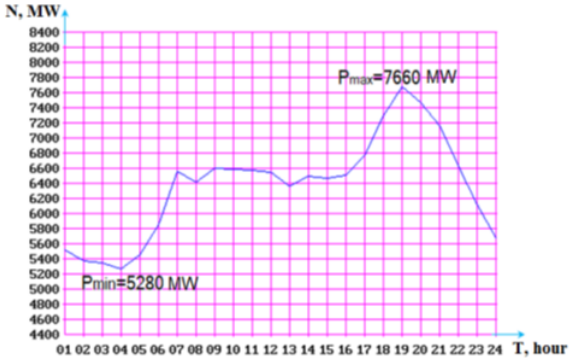


Figure 1: Schedule of daily loads of EPS of the Republic of Uzbekistan (09.10.2019).

The insufficient share of highly maneuverable power plants in the structure of generating capacities of the EPS of the Republic of Uzbekistan forces the regulation of capacities by TPP. This leads to a deterioration in energy supply indicators, in particular, voltage and frequency surges occur in the power system, since thermal power plants cannot quickly change the power. Frequent starts and stops of turbine units lead to fuel overspending, a decrease in the operational resource of heat and power equipment, an increase in repair costs, and a deterioration of the environmental situation in the areas where the TPP is located [9]. In general, according to the Ministry of Energy of the Republic of Uzbekistan, in 2018, the average specific fuel consumption was 354.6 g. t./kWh, and the EPS of countries where proper capacity regulation is carried out using PSPP, this figure does not exceed 330.0...340.0 g.t./kWh [10].

As is known, for the normal operation of the EPS, maneuverable capacities should be about 25% of its total capacity [11]. The most promising maneuverable capacities are HPP, but their share in the EPS of Uzbekistan is only about 13.3%.

The most optimal solution to this problem is the creation and use of pumped storage power plants (PSPP). The peculiarities of the creation and use of PSPP in our region are that the available hydro resources mainly have a water management purpose, and other priorities have so far been secondary, especially with regard to the tasks of improving the energy efficiency of existing reservoirs.

The most important issue in the development of water resources has always been meeting the needs of water management and drinking water supply [12]. Therefore, at the current stage of hydropower development, the development of

hydropower resources should be carried out on seasonal inland watercourses, in reservoir systems, on hydraulic structures without compromising the interests of irrigation and water supply [13]. This, as calculations show, significantly reduces the capital investment for the development of hydropower resources and makes it possible to more fully use the potential of the accumulated volume of water.

In general, we can say that the hydropower complex (HPC) is the parallel operation of electric power facilities (HPP, TPP, WPP, SPP, etc.) and water management systems (PS, geoTPP pumps, HS etc.) [14].

Currently, the construction of new PSPP requires the construction of new reservoirs, which requires large capital investments [15]. It can also be noted that the issues of detailed design of reservoirs for PSPP hydroelectric power plants and their operating modes are only beginning to be developed. When designing reservoirs, it is necessary to take into account a large number of factors: terrain, soil properties, water filtration, groundwater, rocks located in this area, the possibility of using local materials for the construction of a dam, the location of settlements and farmland in the area, the availability and development of road transport capabilities and the environmental situation in this area. For the PSPP, it is necessary to additionally take into account the water management purpose of water resources and their use regimes.

To reduce capital investment in the construction of a PSPP, existing reservoirs can be used [16], but this issue has not been studied.

Another economical option for the construction of a PSPP is the construction of only one reservoir, and instead of the second one, use a large river or canal, as well as create a special pool near large PS and TPP.

As is known, agriculture plays an important role in the economy of Uzbekistan. The most important aspect of the development of agriculture in our region is the water resources necessary for irrigation of land. For this reason, most of the reservoirs in our country, including “Charvak”, serve for the needs of agriculture. During the construction of any new hydropower facility, or the reconstruction of the old one, it is necessary to take into account the needs of agriculture. It is advisable, in conditions of shortage, to use water resources simultaneously for the needs of agriculture and for the needs of the country's energy system.

In this regard, improving the energy, economic and environmental efficiency of hydraulic storage systems, directly related to the development and improvement of methods and methods for determining the optimal regime and technical and economic parameters of the PSPP, and assessing their significance in ensuring the reliability, stability and maneuverability of the EPS and power plants based on RES, is relevant.

2 METHODS AND MATERIALS

For this purpose, a mathematical model is proposed for the technical and economic assessment and determination of the optimal parameters of the PSPP under the conditions of construction of additional facilities with existing structures.

Let there be a set of options with fixed values of capital investments for the construction of the HPC K_{HPC} and the annual costs of the I_{HPC}.

At the same time (1), (2)

$$K_{HPC} = \sum_{i=1}^{n} K_{E,i}\ ; \tag{1}$$

$$I_{HPC} = I_{WMEc} + \sum_{i=1}^{n} I_{E.i}\ . \tag{2}$$

where, I_{WMEc} – annual costs for water management facilities to environmental factors; K_{HPC}, I_{HPC} – capital investments and annual costs for hydropower facilities.

The task will be to distribute limited capital investments between objects $i = 1...n$ and choose the appropriate solution option $j = 1...m$.

As is known [14, 16], the criterion for such calculations is (3):

$$C_{HPC\,ij} = I_{HPC,\,ij} + \lambda K_{HPC,\,ij} + \sum U_{i\,j} \rightarrow min; \tag{3}$$

where, $\sum U$ is the total damage during the commissioning of hydropower facilities (for example, an increase in water loss).

Given the values of the multiplier λ for known values of I_{HPC}, K_{HPC}, and $\sum U$, after a certain number of steps, we can find a solution that satisfies the above criterion and is acceptable for the balance of investments. In this case, the multiplier λ will be the standard of efficiency for the considered set of options.

The set goal, i.e., the increase in additional capacity of ΔN_{HPC} should be provided with a minimum of annual costs with the condition (4)

$$\left(\sum_{i=1}^{n} K_i\right)_j \leq K_{HPC}\,. \tag{4}$$

With the known given economic indicators of the $\gamma_{K,i}$ and $\gamma_{I,i}$, the mathematical model of such a system can be described as follows (5):

$$\left.\begin{aligned} I_i &= \left(\sum_{i=1}^{n} \gamma_{I,i} \cdot N_i\right)_j \rightarrow min \\ \sum_{i=1}^{n} N_i &\geq \Delta N_{HPC} \\ \left(\sum_{i=1}^{n} \gamma_{K,i} \cdot N_i\right)_j &\leq K_{HPC} \end{aligned}\right\} \quad (5)$$

The efficiency of the functioning of any WMC (water management complex) is associated with the balance of the volume of water in reservoirs. When it comes to the accumulation of hydraulic energy, the water balance is compiled separately for the upper and lower streams.

The structural scheme of the PSPP with two reservoirs is shown in Figure 1.

For the upstream, the following condition must be met (6):

$$V_{U.B}(t) = V_{U.B}(t_0) + \int_{t_0}^{t_l}[Q_{flow}(t) - Q_{idle}(t) + \\ +Q_{sed}(t) - Q_{loss}(t) - Q_{draw}(t) \pm Q_{PSPP}(t)]\, dt; \quad (6)$$

where, $V_{u.b}$ (t_0) is the initial volume of the reservoir; Q_{flow} (t), Q_{idle} (t), Q_{loss} (t), $Q_{loss}(t)$, $Q_{draw}(t)$ are the flow rates, idle discharges, sediment, losses and water withdrawal for various needs at time t; Q_{PSPP} (t) is the performance of a hydropower plant, where the sign (+) corresponds to the pumping, and the sign (–) corresponds to the turbine mode.

The volume of the lower reservoir depends on the following balance ratio (7):

$$V_{D.B}(t) = V_{D.B}(t_0) + \int_{t_0}^{t_l}[Q_{idle}(t)+Q_{sed}(t) - \\ -Q_{WMS}(t) - Q_{loss}(t) - Q_{draw}(t) \pm Q_{PSPP}(t)]\, dt; \quad (7)$$

where, Q_{WMC} (t) is the water consumption for the participants of the *WMC* at time *t*. For the lower reservoir, the sign (+) before the Q_{HP} value (t) corresponds to the turbine mode, and the sign (–) corresponds to the pumping mode. If it is impossible to catch filtration waters coming from the upper stream in the lower reservoir, then it is necessary to take this factor into account when calculating the value of the $V_{D.B}$.

This mathematical model determines the relationship between economic costs and the potential of reservoirs in the PSPP. In particular, it follows that the local conditions of the lower stream and the near-dam area can become determining factors for the value of $V_{D.B}$.

Based on the mathematical model for the technical and economic assessment and determination of the optimal parameters of the PSPP, a methodology and a program for determining the optimal energy and economic indicators of the PSPP were developed, which allows us to solve the above problem and makes it possible to determine capital investments in the PSPP (K_{PSPP}), as well as electricity generation in the turbine mode (TR) (E_{TR}), electricity consumption in the pumping

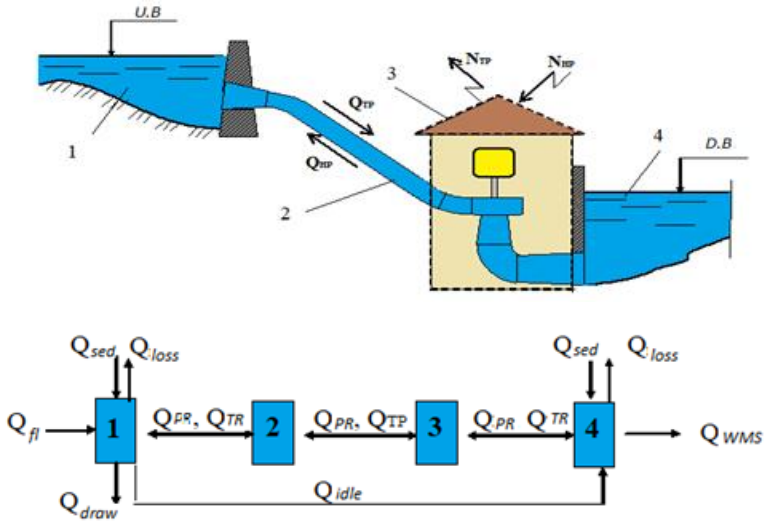


Figure 1: Structural diagram of the PSPP: 1 - upper reservoir; 2 - pressure pipe; 3 - the building of the PSPP; 4 - lower reservoir.

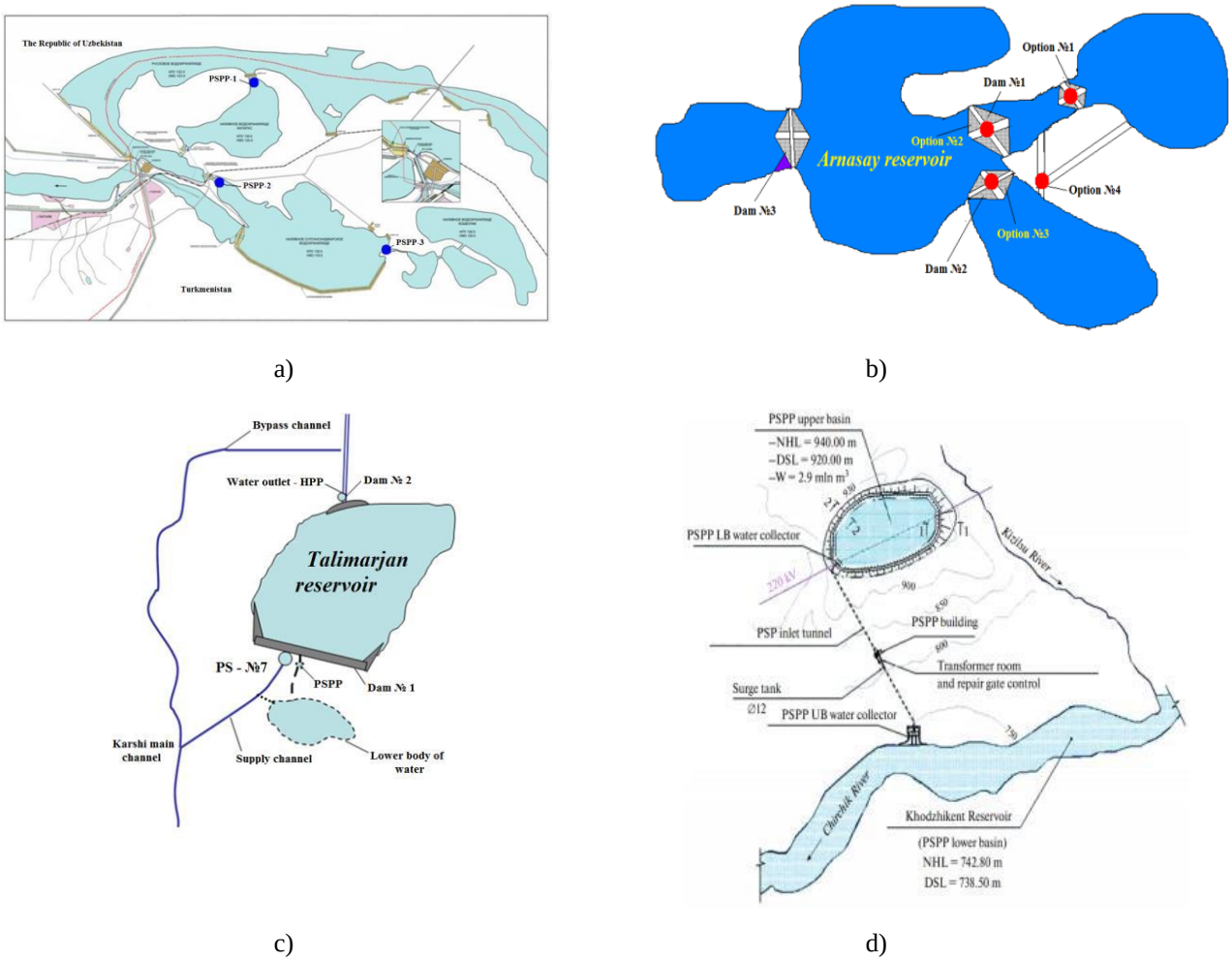


Figure 2: Schemes of energy and water facilities using the PSPP: a) the scheme of the Tuyamuyun hydroelectric complex; b) the scheme of the Arnasay reservoir; c) the scheme of the Talimarjan reservoir; d) the scheme of the Khodjikent reservoir.

mode (PS) (E_{PS}), annual fuel economy (D_{fuel}), fuel economy (E_{fuel}), annual costs of the PSPP (I_{PSPP}), economic efficiency per year (E_{PSPP}), the payback period of capital investments ($T_{payback}$) (5-6 years) and the profitability of capital investments R [17].

3 RESULTS AND DISCUSSION

Based on the developed methodology and program, the energy and economic indicators of the PSPP were studied when they were used by energy and water management facilities [18] at the Tuyamuyun hydroelectric complex, Arnasay, Talimarjan and Khodjikent reservoirs (Figure 2) and the following results were obtained:

- The construction of a PSPP with a total capacity of 39.4 MW in the Tuyamuyun hydroelectric complex will make it possible to receive at least 86.5 million kWh of electricity per year, with an annual economic efficiency of 24.3 billion. sum, and also save 12,950. 0 tons of organic fuels.
- The construction of a PSPP with a total capacity of 159.8 MW at the Arnasay reservoir will make it possible to receive at least 350.0 million kWh of electricity per year, with an annual economic efficiency of 109.5 billion. sum, as well as save 52,508. 3 tons of organic fuels [24,25].
- The use of a PSPP with a capacity of 18.473 MW at the Talimarjan reservoir makes

it possible to obtain 40.456 million kWh of electricity per year, with an annual economic efficiency of 5.440 billion. sum, as well as savings of 6 068.45 t. u. t. of organic fuels.

- Construction of the Khodjikent PSPP with a capacity of 200 MW with an annual economic efficiency of 564.55 billion rubles . sum, makes it possible to generate 452.600 million kWh of electricity per year, save 67.90 thousand tons of organic fuels.

4 CONCLUSION

The use of the PSPP increases the reliability, stability, maneuverability and energy-economic efficiency of the EPS and RES-based power plants; makes it possible to cover the minimum and maximum daily loads of the EPS, regulate the capacities of solar and wind power stations, equalize power fluctuations by accumulating energy.
Currently, the existing reservoirs of the republic are mainly used for irrigation purposes, partly for drinking water supply. For a more complete use of the potential of water resources in the operated reservoirs, it is necessary to create hydropower complexes with PSPPs that can solve issues of electricity production during the mutual operation of reservoirs.

A mathematical model is proposed for the technical and economic assessment and the choice of determining the parameters, a methodology and a program for determining the energy and economic indicators of the PSPP for the energy and water facilities of Uzbekistan are developed. They can be used in the design, development, feasibility study and determination of optimal options for complexes.

The potential and technical possibilities of creating a PSPP at four energy and water management facilities-Tuyamuyun, Arnasay, Talimarjan and Khodjikent facilities allows to obtain a total peak capacity of about 418 MW and additionally generate more than 930.0 million kWh/year of electricity, as well as save fuel resources up to 139 thousand tons of u.t./year. At the same time, it becomes possible to reduce the annual emission of CO_2 emissions to 213 thousand tons / year and achieve environmental efficiency of 32.11 billion rubles. sum/year (or $ 3,195 million / year) and as a result, the economic efficiency of the PSPP will be about 700 billion sums/year.

REFERENCES

[1] Electric power industry of Uzbekistan, [Online]. Available: http://energo-cis.ru/wyswyg/file/news/Power_System_uzbekistan.pdf.

[2] The concept of providing Uzbekistan with electric energy for 2020-2030, [Online]. Available: http://minenergy.uz/uploads/1a28427c-cf47-415e-da5c-47d2c7564095_media_.pdf.

[3] The Ministry of Energy: goals, objectives, plans and achievements, [Online]. Available: https://minenergy.uz/ru/lists/view/10.

[4] “Enerdata Statistical Yearbook of World Energy 2019,” [Online]. Available: https://yearbook.enerdata.ru/renewables/wind-solar-share-electricity-production.html.

[5] Resolution of the President of the Republic of Uzbekistan No. PP-2947, “On the program of measures for the further development of hydropower for 2017-2021,” May 2, 2017, [Online]. Available: http://lex.uz/pages/getpage.aspx?lact_id=3219734.

[6] E.B. Saitov and T.B. Sodiqov, "Modeling an Autonomous Photovoltaic System in the Matlab Simulink Software Environment," AIP Conference Proceedings, vol. 2432, 2022, 020022.

[7] E.B. Saitov, Sh. Kodirov, B.M. Kamanov, N. Imomkulov, and I. Kudenov, “Increasing the Efficiency of Autonomous Solar Photovoltaic Installations for Power Supply of Agricultural Consumers,” AIP Conference Proceedings, vol. 2432, 2022, 040036.

[8] F. Zikrillayev, E.B. Saitov, J.B. Toshov, B.K. Ilyasov, and M.B. Zubaydullayev, “A Software Package for Determining the Optimal Composition and Parameters of a Combined Autonomous Power Supply System Based on Renewable Energy Sources,” AIP Conference Proceedings, vol. 2432, 2022, 020021.

[9] E.B. Saitov, J.B. Toshov, A.O. Pulatov, B.M. Botirov, and Y.M. Kurbanov, “Networked interactive solar panels over the roof photovoltaic system (PVS) and its cost analysis at Tashkent state technical University,” E3S Web of Conferences, vol. 216, 2020, 01133.

[10] Decree of the President of the Republic of Uzbekistan, “On the formation of the joint-stock company Uzbekhydroenergo,” no. UP-5044, May 18, 2017, [Online]. Available: https://lex.uz/docs/3210104.

[11] Resolution of the President of the Republic of Uzbekistan, “On the program of measures for the further development of renewable energy, improving energy efficiency in economic and social sectors for 2017-2021,” no. PP-3012, May 26, 2017, [Online]. Available: “http://www.lex.uz/pages/getpage.aspx?lact_id=3221897.

[12] “The development of generating capacities - thermal power plants, nuclear power plants, RES-are the main ones in the development of the electric power industry of Uzbekistan,” [Online]. Available: https://e-cis.info/news/567/82429/.

[13] Infopedia, “Accumulating electric stations,” https://infopedia.su/18x172e0.html

[14] P. Denholm, E. Ela, B. Kirby, and M. Milligan, “The role of energy storage in the production of electricity

from renewable sources," Technical Report NREL/TP-6A2-47187, National Renewable Energy Laboratory, January 2010.

[15] V. Yu. Sinyugin, V. I. Magruk, and V. G. Rodionov, "Hydraulic accumulating power plants in modern electric power industry," ENAS, Moscow, 2008.

[16] I. Sh. Klychev, M. M. Mukhamadiev, O.Kh. Nizamov, E. K. Mammadyarov, and K. S. Dzhuraev, "Method for calculating the capacity of combined autonomous power plants," Applied Solar Energy, vol. 50, no. 3, pp. 196-201, 2014.

[17] M. M. Mukhammadiev, K. S. Dzhuraev, S. R. Juraev, Zh. B. Kulanov, and D. A. Mamatkulov, "Program for determining the energy and economic parameters of pumped storage power plants," Certificate of official registration of the computer program No. DGU 07363, December 19, 2019.

[18] M. Mukhammadiev, K. Dzhuraev, S. Juraev, A. A. Uulu, and A. Makhmudov, "International Scientific Conference 'Construction Mechanics, Hydraulics and Water Resources Engineering' (CONMECHYDRO - 2021)," E3S Web of Conferences, vol. 264, pp. 1-14, 2021.

System of Circadian Lighting Based on Determination of Human Thermal Comfort Level

Yehor Zheliazkov[1,2]
[1]Igor Sikorsky Kyiv Polytechnic Institute, Peremohy Avenue 37, Kyiv, Ukraine
[2]Technical University of Dresden, Mommsenstrasse Str. 11, Dresden, Germany
yzheliazkov2@gmail.com

Keywords: Circadian Lighting, Visual Comfort, Thermal Comfort Lighting, Adaptive Lighting.

Abstract: Comfortable level is an important element that has considerable implications on circadian energy and the perception of lit environments. Light has been also defined in terms of the circadian system, but not only for human vision. Light source and systems development should consider the needs of both the visual and non-visual systems. Many complaints about indoor conditions are related to unsatisfactory thermal environments. Most research on Thermal Comfort (TC) considered physical parameters for settings and users yet marginalized the influence of user's psychological aspects in the process of thermal sensation. This research examines the effect of mood states on human evaluation of the thermal environment in virtual settings. There are ten university participators from Technical university of Dresden have taken part in experiments. The experiment followed the "Lab 2.0" using climate controlled chamber and TC evaluation using psychological responses. Linear model was used to analyse the data. The results revealed a relationship between TC, mood state and quality of the indoor environment. Humans' judgment on TC is a variable mental reaction. The research presents differences between the evaluation people to their thermal environments. This study expands research on the indoor environment quality and develops TC evaluation strategies.

1 INTRODUCTION

We spend most of our time indoor in artificial thermal environments which affect human beings. Properly designed indoor environments are essential for pleasant living and optimized task performance in the work place. Amongst other parameters, Thermal Comfort (TC) level is one of the scales that measure human`s solace in physical settings. It indicates the thermal balance of the body as determined by the operative temperature and moderated by environmental and personal parameters.

Lighting systems allow to adjust the individual characteristics and human healthy demands. Particularly, during adjusting the lighting parameters, it is possible to create a positive effect on a human circadian rhythms. The violation of these rhythms is an important problem during the durable operation of computers or laptops [1].

In this article lighting will be describe through the connection between thermal and circadian lighting. Circadian lighting means lighting that coincides with our circadian system for the entire day. Outdoor lighting changes color throughout the day from a warm lighting to a cool daylight color and again to a warm color [2]. It is important to pay attention that circadian lighting isn't limited to light bulbs and fixtures, it includes electronic devices as well.

Studies on TC are classified within the domain of the physical and technical sciences. Three major trends of TC studies were identified: positivistic and reductionist approach, multivariate and systems-oriented approach and applied and conservation oriented approach [3]. The former is a physical and laboratory-based trend, which exported set of standards, indexes and bioclimatic charts in indoor and outdoor thermal conditions.

This article reports a multi-domain (thermal and visual) comfort investigation, aiming to identify whether the colour of light affects the human thermal perception. For this, it was planned an experiment associating different thermal settings and colours of light.

It is also important to receive measurements in order to realize electronic system of lighting.

2 THEORETICAL PART OF CIRCADIAN AND COMFORT THERMOLIGHTING

2.1 Radiation Lighting in Case of Temperature Field

Lighting is a fundamental requirement of our daily life, especially for human circadian rhythm. The effect of light, which is emitted by electronic devices, has either positive or negative effect on a person, even if it does not get in contact with your eyes [4].

Decades of research demonstrate that radiation responses vary across an organism's circadian period. The emerging field of chronoradiobiology examines the biological relationships between the complex mechanisms of circadian regulation and cellular radiation responses with the goal of improving the therapeutic index of radiation treatments.

Understanding circadian regulation, disruptions, and downstream effects that can impact radiation therapy could lead to potential improvements for patients. Data on circadian disruption and clock gene regulation may lead to new approaches to personalize care [5].

2.1.1 Wien's Law and Connection Between Room and Color Temperature

Radiation laws of the black body: The black body represents an ideal radiation body. This means that the entire incident radiation is absorbed by the black body and also emitted again. Since this assumption assumes thermal equilibrium, the temperature of the black body does not change either [6, 7].

Radiation temperature must be relatively low because of the anti-proportional relationship between wavelength λ and temperature T due to Wien's displacement law (1) (DIN 5031-7 1984).

$$\lambda_{\max} = \frac{2897.8\mu m \cdot K}{T}. \qquad (1)$$

It is important in our work to find connection between room and lighting temperature in order to make some basically explanation of lighting impact on thermal comfort.

The position of the radiation maxima can be determined using Wien's displacement law

$$E = \frac{2\pi h c_0^2}{\lambda^5 \left[\exp\left(\frac{hc_0}{k_B T_{room} \lambda} \right) - 1 \right]}, \qquad (2)$$

where coefficients $h = 6.63 \cdot 10^{-34}\ Joule/s$ – Planck constant, $c_0 = 3 \cdot 10^8\ m/s$ – speed of the lighting, $k_B = 1.38 \cdot 10^{-34}\ Joule/K$ – constant of Boltzmann, T_{room} – air temperature inside of the room or chamber ($T_{room} = 19...25°C$), λ – wavelength of radiation $[nm]$, S – area of falling light $[m^2]$, r – distance between eyes and lighting source $[m]$.

After the series of transformation (2) were transformed into (3)

$$I = \frac{E}{S} = \frac{2\pi h c_0^2}{S\lambda^5 \left[\exp\left(\frac{hc_0}{k_B T_{room} \lambda} \right) - 1 \right]} \Rightarrow \frac{I}{r^2} = \frac{2\pi h c_0^2}{r^2 S\lambda^5 \left[\exp\left(\frac{hc_0}{k_B T_{room} \lambda} \right) - 1 \right]} \qquad (3)$$

where I – intensity of lighting $[W/m^2]$, I/r^2 – relation for illuminance $[lx]$,

Illuminance describes the quantity of luminous flux falling on a surface. It decreases by the square of the distance (inverse square law). Relevant standards specify the required illuminance (e.g. EN 12464 "Lighting of indoor workplaces") [8]. Equation (3) were transformed into (4):

$$illu\min ance = \frac{2\pi h c_0^2}{r^2 S\lambda^5 \left[\exp\left(\frac{hc_0}{k_B T_{room} \lambda} \right) - 1 \right]} = \frac{2\pi E c_0}{r^2 S\lambda^4 \left[\exp\left(\frac{E}{k_B T_{room}} \right) - 1 \right]} \qquad (4)$$

Equations (1) and (4) are basically in order to calculate and receive connections between lighting parameters and climate parameters, especially room temperature.

2.1.2 Representation of the Different Kinds of Electric Light Sources.

Light, the basis for all vision, is an element of our lives that we take for granted. We are so familiar with brightness, darkness and the spectrum of visible colours that another form of perception in a different frequency range and with different colour sensitivity is difficult for us to imagine. Visible light is in fact just a small part of an essentially broader spectrum of electromagnetic waves, which range from cosmic rays to radio waves [9].

Human eye can experience a specific portion of the electromagnetic spectrum, which referred as 'Visible light' in Figure 1 [10]. Radiant energy from light is transmitted, absorbed or reflected, when it strikes a surface. From observer's point of view, color of surface is only visible if the frequency of the spectrum reflected back to the eye of observer. The brighter the surface, the occurrence rate of reflection

increase, likewise, darker surfaces absorbs more radiant energy with minimal reflection.

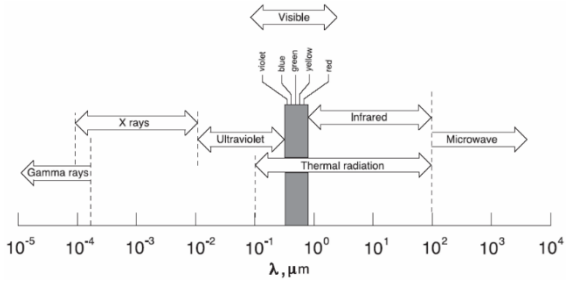


Figure 1: Electromagnetic spectrum.

Electric light sources can be divided into two main groups, which differ according to the processes applied to convert electrical energy into light. First group comprises the discharge lamps; they include a wide range of light sources, e.g. all forms of fluorescent lamps, mercury or sodium discharge lamps and metal halide lamps. Second group comprises the thermal radiators, they include incandescent lamps and halogen lamps.

Low-pressure and high-pressure lamps belong to the category of "Discharge lamps". This means, that they do not leave a lot of heat during the working process.

The electric discharge lamp is family of artificial light sources that generates light by sending an electric discharge through an ionized gas. In all types of electric discharge lamps, an electric current is passed through a gas or vapor which renders it luminous. The most commonly used elements in these types of lamps are neon, mercury and sodium vapors as in Figure 2.

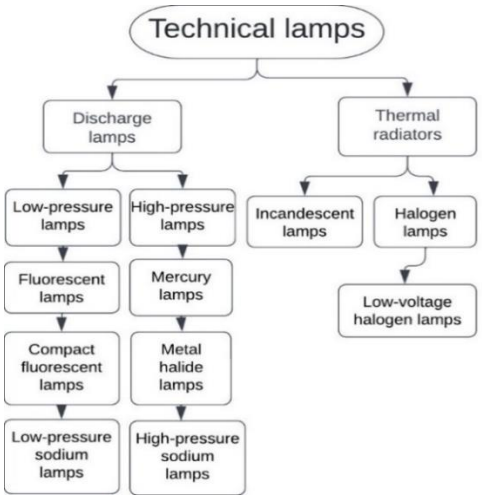


Figure 2: Kinds of electric light sources according to the means of their light production.

The advantages of electric discharge lamps are: longer life, operating cost is low, discharge lamps can be designed for different colors, and, of course, produce less heat [11]. Last aspect was taken into account for experiment.

2.2 Circadian Rhythm and Circadian Lighting

This "Circadian rhythms" [12] mean changes in physiological, biochemical and behavioral functions of the body. These rhythms are synchronized with the natural light dynamics of the surrounding world ("day and night") by means of internal molecular "hours", surround temperature, humidity and other conditions. Circadian rhythms are also 24-hour cycles that are part of the body's internal clock, running in the background to carry out essential functions and processes. One of the most important and well-known circadian rhythms is the sleep-wake cycle.

Human organism can adopt to the certain time of the day and the different systems of the body follow circadian rhythms that are synchronized with a master clock in the brain. This master clock is directly influenced by environmental cues, especially light, therefore circadian rhythms are tied to the cycle of day and night.

As properly aligned, a circadian rhythm can promote consistent and restorative sleep. But if this circadian rhythm is thrown off, it can create significant sleeping problems for the organism, including insomnia. Earlier researches are also revealing that circadian rhythms play an integral role in diverse aspects of physical and mental health.

Certain list of circadian rhythm with a certain time and characteristic of human organism was represented in Figure 3. This shows certain period of time, when people have highest or lowest either blood pressure, and body temperature; time of deepest sleep and so on.

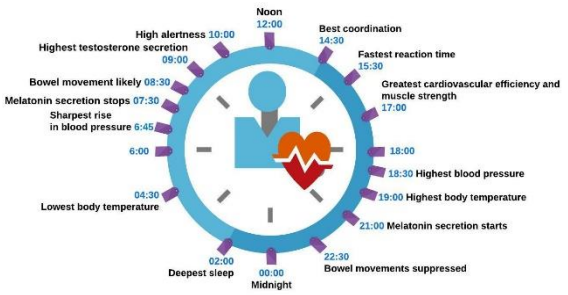


Figure 3: Circadian rhythm during the different period of day.

The terms circadian and integrative lighting are now often used interchangeably [13]. Throughout this report, we have chosen to use the term circadian lighting, as the project was funded for studying circadian lighting, lighting that is intended to support the natural sleep and wake cycle of humans that is normally entrained by the presence and absence of naturally varying daylight.

Research has indicated that light affects on the human healthy both of non-visual and our visual systems and that electric light can make impact on circadian rhythm [14]. Circadian lighting is the concept that electric light can be used to support human healthy by minimizing the effect of electric light on the human circadian rhythm. Research workers have discovered that long-term exposure to certain wavelengths of blue light at a specific intensity can have either positive or negative impact on melatonin production. The conception of using lighting to influence human circadian rhythm is a relatively new one in the lighting industry: all aspects of this exciting new topic are still being studied for a long time, and the current data is still in its infancy. Nowadays, there are three main electric light approaches to implementing a circadian lighting system: intensity tuning and stimulus tuning.

Intensity tuning is the most familiar and cost-effective solution to circadian lighting. Fixtures maintain a fixed correlated color temperature (CCT) while the intensity (brightness) of the fixture is adjusted, through a controlled dimming system, to correlate with time of day. Light fixtures are set to a lower intensity in the early morning, transition to a higher intensity as the day progresses, and reduce to a lower intensity in the evening.

Color tuning involves changing the light intensity and CCT to mimic the daytime/nighttime cycle. We experience cooler color temperatures (ranging from 4000 Kelvin up to about 10,000 Kelvin) when the sun is highest in the sky and people are typically most alert during the day. Therefore, cooler CCTs are used in spaces and during times when it's appropriate to promote alertness and attention. Warmer color temperatures (ranging from < 2700K to 3500K) represent daylight hours when the sun is rising and setting when people are falling asleep or waking up. Circadian lighting systems are set to adjust based on the CCT we typically observe at any given time of the day.

Stimulus tuning is lighting technology that replaces the "blue light hazzard" with "good blue" light wavelengths. This circadian lighting approach more closely mimics the daylight natural spectrum. Stimulus tuning light fixtures can be programmed for reducing blue light wavelengths during the evening/nighttime hours to limit melatonin suppression without changing the CCT. This lighting approach is similar to color tuning and is the most effective when paired with intensity tuning [15]

The consequences of circadian rhythm disturbances are reduced sleep quality, drowsiness, low work productivity and bad mood. The effect of radiation on the human body includes visual perception, influence on the psychophysiological and emotional state. At the same time, there is damage to the elements of the eye and skin, as well as a number of other negative factors for physical health (photobiological danger).

2.3 Main Aspects for Visual Comfort Circadian Lighting

Most of all visual comfort can be usually defined through a set of criteria based on the level of light in the room, the balance of contrasts, CCT and the absence or presence of glare [16]

There are actually many different definitions of visual comfort level and what people should measure as part of it.

The most widely used and accepted metrics tend to focus on:

- The measure or the quantity of natural light over the year: the potential of the certain building and its location to provide enough daylight to occupants. This is typically explored through daylight autonomy or through the useful daylight illuminance;;
- Lighting distribution as perception: whether a space will be too highly contrasting or too bright;
- Another aspects: there are another conceptions that don't have a metric but on which academics widely agree. The quality (through its spectral) of a view is also a consideration but is hard to quantify; however, the first methodologies are starting to appear.

During our research, office conditions will be accept. Actually, in cellular offices, good conditions have been observed in general and giving the users full responsibility for controlling their own indoor climate during occupied hours works very well. The occupants have demonstrated a sparing use of supplementary mechanical cooling/heating equipment. They appear to have a preference for using windows and clothing to modify conditions in mild mid-season weather. However, most occupants appear to have upper and lower "tolerance" limits,

beyond which active intervention will be applied if the opportunity is available [17].

The control strategy should determine both time and rate control. It should also determine different control modes in relation to different weather conditions. The actual control strategy should reflect the demands of the building owner, the needs of the users and the requirements in standards and regulations [17, 18].

Therefore, actual task of the work is defining of comfort temperature and dates, which are suitable and comfort for humans, who are staying in the office for a very long time.

3 METHODS OF RESEARCHING

3.1 Climate Chamber at the Technical University of Dresden

For experiment was used climate chamber at the Technical university of Dresden.

The physical behavior of the climate room is extremely flexible. The climate room at the TU Dresden (Figure 4) has an inner dimension of [19].

The interior surfaces are divided into 73 individually tempered surface areas and the connected ventilation system can be controlled by temperature, humidity and volume flow. In order not to exert any significant influence on the measurement results, the ventilation system only serves to supply the subjects with fresh air during the test.

Activities are currently underway at various scientific institutions that focus on refining the local comfort models and on checking the global comfort models.

Figure 4: Circadian rhythm during the different period of day.

The surface temperatures of the walls, ceiling and floor can be regulated in a range of T = 10 … 50 °C. Air volume flow maximal, actually maximal volumetric flow per unit time 600 m³/h.

In order to define important for measuring parameters of lighting there were used optical devices. There are digital color temperature sensor and illuminance device FLAD23CCT with ALMEMO D6-connector [20], which were used during the experiment, are in Figure 5.

Figure 5: Color temperature sensor with device ALMEMO 2590-2.

Certain technical dates for measuring device are represented in Table 1.

Color temperature and illuminance are determined as a means to plot and evaluate lighting systems.

Digital color temperature sensor with “TrueColorSensorchip” and integrated signal processor “The TrueColorSensorchip” (3 sensors on 1 chip) detects - separately - each of the three colors - red, green, blue (RGB). The respective sensitivities of these 3 color sensors are adapted to the standard spectral curves as per CIE and DIN. On the basis of these RGB values the computer calculates the color point within the RGB range in terms of coordinates X and Y and determines the correlated color temperature (CCT) in Kelvin.

Compact sensor, particularly suitable for mobile applications. The display shows simultaneously both this color temperature and the illuminance in a range of lux (lx) or kilolux (klx).

Device as in Figure 5 is useful both for measuring of lighting, and different climate parameters such as for example air temperature, humidity level and list of others.

Table 1: Technical dates for measuring devices.

Spectral sensitivity	380 nm - 720 nm
Correlated Color Temperature (CCT)	10...65,000 lux (factory settings) or 0.05...170.00 kLx
Illuminance V-Lambda:	10...65,000 lux (factory settings) or 0.05...170.00 kLx
Room temperature conditions	-10…+40 C°
Refresh rate:	1.5 seconds for all channels
Settling time:	3 s (when operating the data logger in sleep mode, a sleep delay of 3 s must be programmed)
Supply voltage	6...13 V DC

During the experiment test people are sitting in the middle of the room and accepting the light as represented in Figure 5 There is represented front view a), side view b) and plan view c) with an area dimension and lamps, each of the lamps numbered, has a certain dimension and Guth position index (p_1, p_2, p_3, p_4) [21]. More detailed information about Guth position index and another parameters can be found in [22]. Lighting in Figure 6 is represented in case of arrows.

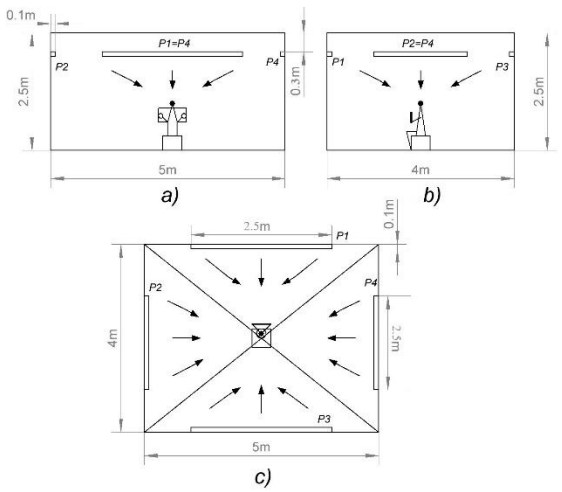


Figure 6: Schematic representation of elevation and person, which is sitting in the middle of the room: a) front view; b) side view; c) plan view.

Person in Figure 6 can be represented with a big scale as in Figure 7. There is explanation of siting person position.

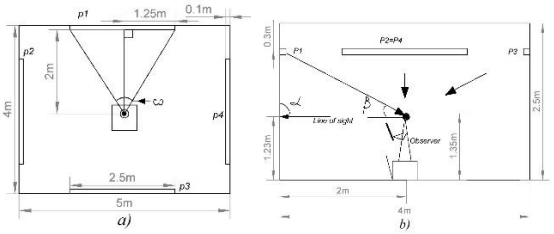


Figure 7: Detailed schematic representation of the test person a) plan view; b) side view.

3.2 Experiment Task and Participants

One of the tasks for measuring is defining of human comfortable level. There were prepared special conditions in order to define impact on human comfortable level through the lighting and to confirm some theoretical moments between lighting parameters and comfort parameter.

Participants. There are five persons were called for an experiment, especially boys and girls. Each of them has been interviewed before experiment about different biological and personal parameters: age, sex, how did people get to the labor (bicycle, walking, car or something else), height, weight, first imagination of temperature feeling (hot, warm, cold) in the climate chamber.

3.3 Realization

During the experiments, both Correlated Color Temperature (CCT) and illuminance are only variable parameters. Another parameters such as room and surface temperature, airflow are constants.

The influence of brightness on human sensation has been extensively studied in the past and is also established in various standards and regulations. In order to be able to assess the influence of color temperature on the thermal sensation of humans, tests are currently being carried out on test persons in the Combined Energy Lab of the TU Dresden [10]. A uniform operative room temperature of 22°C is set in the indoor climate room. One result could be that at the selected operative room temperature of 22 °C, the thermal sensation of the test persons is influenced by the choice of color temperature - "warm" light could provide a higher thermal sensation than "cold" light. If this correlation can be shown, this obviously has a very great potential for the operation of a heating system or its energy consumption as well as CO2 emissions.

With the help of LED lighting, the color temperature can be varied in the range of T = 2600 ... 4400 K and brightness can be varied in the range

of 79…250 lx. This results in a three-stage experiment with color temperatures T1 = 2600 K, T2 = 3500 K and T3 = 4400 K with constant ambient conditions (room perimeter surfaces and supply air have 22 °C).

The color temperature is thus the only variable influencing thermal comfort. The test person answers an online questionnaire based on DIN EN ISO 28802 every 5 minutes. Experiment passes during the certain period of day-time: either in the morning, or in the afternoon.

Figure 8 shows one of the example of colour temperature changing in climate room.

Figure 8: Experimental setup, left T1= 2600 K, right T3 = 4400 K.

There are some more conditions:

- test person must not use any other devices such as computers, mobile phones, tablet or any other lighting devices;
- each test person has a right to take along either book or newspaper;
- each test person has a right to take along a bottle of water and drink it.

4 METHODS OF RESEARCHING

4.1 Results of Lighting Calculations

There were different participants, who have already taken part during the experiment. Certain information about biological and personal parameters, which were mentioned in Section 3.2, are shown in Table 2. During the each position there people must sitting on the chair in the middle of the room.

Dates in Table 2 are important and it is possibility to compare lighting impacts on the human, who has certain individual characteristics. Unfortunately, there are only 5 test persons, but the acute accent was taken not on statistical research, but searching for common zone, which would be comfortable for each test person.

Table 2: Personal parameters of test persons.

Number of person	Age	Sex	Person gets to university	Height, cm	Weight, kg
1	25	male	On feet	181	76
2	39	female	On feet	171	76
3	29	male	On feet	180	71
4	33	male	On feet	170	80
5	23	male	Bicycle	172	

As it was mentioned above, each test person had to answer online questionnaire every 5 minutes. Here are some results of experiment in Figure 9 a) and 9 b).

Table 3 shows lighting conditions, which were used during the experiments.

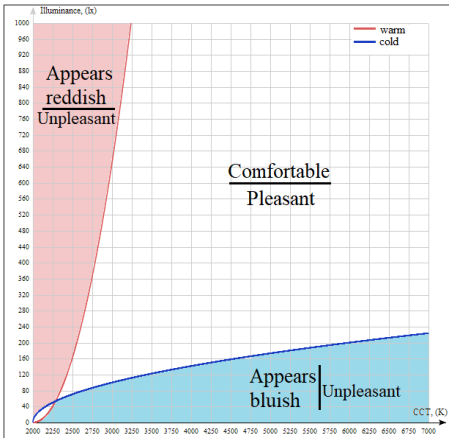
Table 3: Experimental middle measuring results based on questionnaire of one of the statistical person.

Questionnaire of one of the statistical person							
Time	Period	Phase	CCT, K	Brightness, lx	Feeling of room temp.	Did you feel any changes	Do you agree with temp.
9:23		Clear all date	4400	222			
9:46		Test realization	4172	81	Something warm	Cooling	Yes
9:55	0:00	Phase 1	4133	85	Neutral	Cooling	Yes
10:15	0:20	Start of Pause	4138	84	Neutral	Cooling	Yes
10:20	0:25	Phase 2	3346	69	Something cold	Cooling	Yes
10:40	0:45	Start of Pause	2600	78	Neutral	No changes	Yes
10:45	0:50	Phase 3	2600	85	Neutral	No changes	Yes
11:05	1:10	Finish	2600	88	Neutral	No changes	Yes

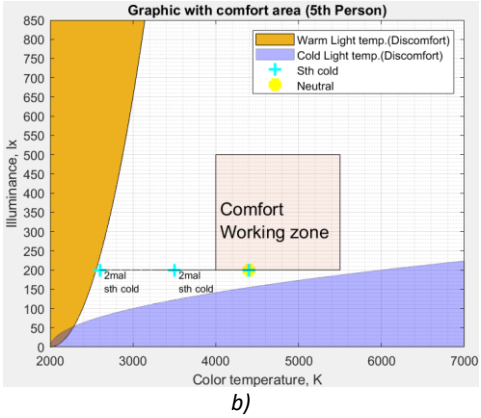
All the results obtained from the tested persons were then processed and brought to Kruithof coordinate plane that contains two curves:

- hyperbolic function for bluish or cold colour temperature $y = \sqrt{10(x-2000)}$;
- parabolic function for reddish or warm colour temperature $y = (x-2000)^2 / 1538.46$.

Final curves are shown on Figure 9 where a) presents the approximation of Kruithof curves with a square of cold and warm colour temperature. At the same time there was received an experimental measuring results based on questionnaire of the statistical persons b). Square in the center of graphic is «comfort zone», which is limited by four direct lines ($CCT_1 = 4000K$, $CCT_2 = 5500K$, $E_{V1} = 200lx$, $E_{V2} = 500lx$) and shows the most pleasant zone for a work.



a)



b)

Figure 9: Practical realization of Kruithof curves with an: a) warm and cold color temperatures; b) results based on experiments with one of the test persons.

Comfortable thermal lighting has background connection with adaptive lighting.

Adaptive lighting means lighting that could automatically adjust to the certain color temperature of your smart lights throughout the day to better match human circadian rhythm and the light outside.

An adaptive lighting system automatically adjusts its light output and operation to provide targeted light levels based on environmental conditions, user schedules, or other application specific criteria. An adaptive system can also often be manually tuned, over time, in terms of light level, and in some cases, color, to provide optimal lighting conditions as designated by system operators, building owners or occupants. This feature set is accomplished by combining controllable luminaires with lighting controls and communication hardware that are able to interpret changes in the environment and adjust the luminaires accordingly [23].

According to equations (2) and (5) for lighting parameters, there was received certain connections between lighting parameters. First of all parameters of color temperatures $CCT_1 = 4000K$ and $CCT_2 = 5500K$ were taken into account, the next point is illuminance for a working process ($E_{V1} = 200lx$ and $E_{V2} = 500lx$), and the last one is range of visible lighting between $\lambda = 380nm$ and $\lambda = 700nm$.

As the result, there were built graphics, which can in the best way show dependence between lighting parameters (room temperature and color temperature). Certain dependence is represented in Figure 10.

Table 4: Connections between lighting parameters.

λ, nm	T_{farb}, K	E^*_{Joule} 10^{-27}	$illuminance,$ lux
724.5	4000	2.745	132.0219
526.9	5500	3.775	649.023
594.16	4877.4	3.3476	500
700	4140	2.841	200
380	7626	5.234	3326.2
700	4140	2.841	200.219

According to (4), there was received graphic as in Figure 10 a).

It is also important to take into account restriction in Figure 9 b), especially range of illuminance $E_V = 200...500lx$ for office work. As the result, it was received graphic in Figure 10 b).

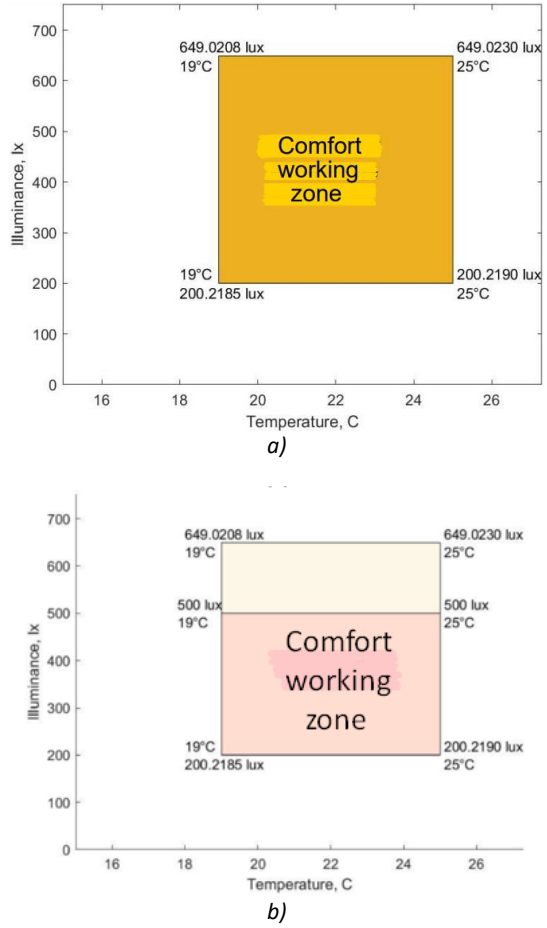


Figure 10: Connection between illuminance and room temperature: a) common situation; b) situation according to Kruithof curve.

As the result of Table 4, there were received 3-dimensional zone for comfort zone in Figure 11, which could theoretically passed according to results of calculations.

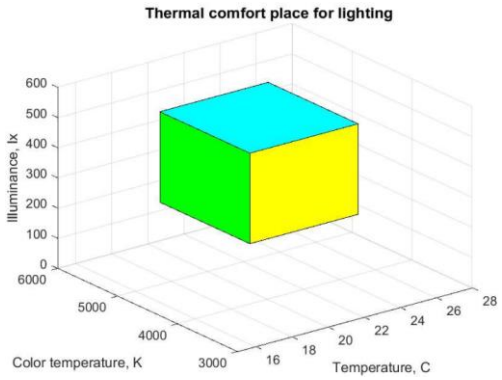


Figure 11: 3D-model of thermal comfort with connection between illuminance and room temperature.

4.2 Results of Human Questionnaire

According to answers during the experiment, there were received the results, which are represented in Table 5 and Table 6 and appropriate graphics in Figure 11 and Figure 12.

Some questions during the questionnaire were connected with acceptations and opinion about level of room temperature. In Table 5 there are characteristics in range of “freeze-cold-something cold-neutral-something warm-warm-hot” and answers, which prefer test person.

Certain dates for each color temperature in Table 5 were also represented in Figure 13.

Table 5: What does test person think about room temperature, especially feeling of room temperature.

Question:	What do you think about room temperature?						
Options:	freeze	cold	something cold	neutral	something warm	warm	hot
Meaning:	-3	-2	-1	0	1	2	3
Color temperature	Number of responses by subjects during the last two surveys						
2600		2	4	4	0		
3500			4	5	1		
4400		2	3	5	0		

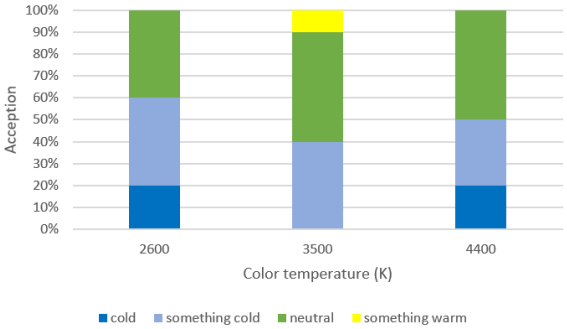


Figure 12: How does test person accept actual room temperature.

Another question during the questionnaire was connected with agreement with room temperature.

Table 6: What does test person think about room temperature, especially feeling of room temperature.

Question:	Do you agree with room temperature?	
Options:	No	Yes
Meaning:	0	1
Color temperature	Number of responses by subjects during the last two surveys	
2600	2	8
3500	2	8
4400	3	7

Certain dates for each color temperature in Table 6 were represented in Figure 13.

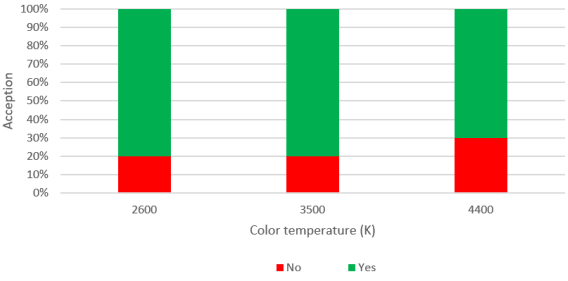


Figure 13: Does test person accept actual room temperature.

There are many different particular characteristics of human organism and certain results need more processing. But people can image which of color temperature can be acceptable, when both air parameters, and room temperature are constant. It is a background for creating a graphic for Sample Standard Deviation. As the result, it will be receive zone of agreement combining both common zone of actual room temperature, and zone of air acceptations.

By the way, principle of fuzzy logic will be used in order to define correct point in a certain period of time as in [22]. It will be done through the system IoT.

5 CONCLUSIONS

The optimal CCT is more beneficial than increased illuminance in moderate ambient indoor lighting, as it provides better lighting comfort.

This work is dedicated to the development of "smart lighting in the building" systems that create a positive impact on the physiological properties of human. By applying IoT and sensor technology we designed and implemented a smart lighting system, which can dynamically control lighting environment inside a room.

During the projecting of lighting systems, the choice of color temperature and spectral composition of sources should be carefully considered in order to avoid disruption of natural circadian rhythms or, if necessary, to increase the efficiency of human physiological parameters (for example, daytime or nighttime working period). Particularly, the beginning of the color temperature is in the numerical range of 4000-5500 K.

There was defined original and practical zone from “Kruithof curve”, which forms square “Comfort working zone”, which is limited by four direct lines ($CCT_1 = 4000K$, $CCT_2 = 5500K$, $E_{V1} = 200lx$, $E_{V2} = 500lx$) and shows the most pleasant zone for a work.

ACKNOWLEDGMENTS

Author is thankful to the Technical university of Dresden, first of all to Faculty of mechanical and science engineering, Chair of building energy systems and heat supply for possibility to conduct the experiment in climate chamber, which is a part of the “Combined Energy Lab 2.0”.

Author is also thankful personal to Joachim Seifert, Maximilian Beyer and Lars Schinke for common work in this direction and possibility to conduct the experiment in climate chamber, which is a part of the “Combined Energy Lab 2.0”, at the Technical university of Dresden.

Moreover, author is thankful to supervisor Julia Yamnenko for common working in direction, connected with a system of circadian lighting.

REFERENCES

[1] O. K. Suzer, "The effects of correlated colour temperature on wayfinding: A study in a virtual airport environment," 6th International Conference on Spatial Cognition (ICSC 2015), Italy, Rome, vol. 16, no. 1, pp. S72, 2015.

[2] SMUD Energy Education and Technology Center "Circadian Lighting and You," pp. 1-23, 2017.

[3] W. Heijs and P. Stringer, "Research on residential thermal comfort: some contributions from environmental psychology," Journal of environmental psychology, vol. 8, no. 3, pp. 235-247, 1988.

[4] J. Zhang, K. Lv, X. Zhang, M. Ma, and J. Zhang, "Study of Human Visual Comfort Based on Sudden Vertical Illuminance Changes," vol. 12, no. 8, 16 p., July 2022, doi: 10.3390/buildings12081127.

[5] N. Nelson, J. Lombardo, L. Matlack, and et al., "Chronoradiobiology of Breast Cancer: The Time Is Now to Link Circadian Rhythm and Radiation Biology," International Journal of Molecular Sciences, vol.23, no. 3, 26 p., 25 January 2022, doi: 10.3390/ijms23031331.

[6] C. Ferk, "CFD Modellierung eines klimaadaptiven Fassadenelementes mit Fokus auf aktiver und passiver Solarenergienutzung," Masterwork for obtaining the academic degree Graduate engineer, 198 p., August 2016.

[7] C. Zürcher and T. Frank, "Bauphysik – Bau und Energie," ETH Zürich, Switzerland, 4th Edition, 340 p., 2014. ISBN 978-3-7281-3608-4

[8] Zumtobel Lighting GmbH, "The Lighting Handbook," 6th Edition, Dornbirn, Austria, 244 p., 2018. ISBN 978-3-902940-72-8.

[9] K. Alam, "Case studies on an integral, simulation-based energy-efficient open loop control for daylight and artificial lighting," Submitted as a project report for obtaining the academic degree Master of science in Engineering, Wels, Austria, 126 p., December 2016.

[10] M. Karlen, C. Spangler, and J. R. Benya, "Lighting Design Basics," 3rd Edition, 272 p., September 2017. ISBN: 978-1-119-31227-7.

[11] Electric Discharge Lamp – Definition, Types, Advantages, Disadvantages & Applications, [Online]. Available: https://www.tutorialspoint.com/electric-discharge-lamp-definition-types-advantages-disadvantages-and-applications.

[12] Y. Zheliazkov, J. Seifert, J. Yamnenko, M. Beyer, and L. Schinke, "System of adaptive lighting in the building based on determination of human circadian rhythms," 2022 IEEE. Proceedings of 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine, 886 p., February 22-26, 2022, doi: 10.1109/TCSET55632.2022.9766921.

[13] W. Osterhaus, I. Erhardtsen, M. Gkaintatzi-Masouti, K. Gert Nielsen, and F. Dobos, "Case Studies Circadian Lighting," Aarhus University, AFRY and IVE, 100 p., 2020.

[14] The lighting practice. What is Circadian Lighting? [Online]. Available: https://www.thelightingpractice.com/what-is-circadian-lighting/.

[15] Y. Chaopu, W. Fang, T. Jiancheng, Y. Fan, L. Yanfeng, and L. Chun, "Change of blue light hazard and circadian effect of LED backlight displayer with color temperature and age," Optics Express 26, pp. 27021-27032, October 2018, doi: 10.1364/OE.26.027021.

[16] A. Liu, A. Tuzikas, Z. Arturas, P. Vitta, R. Vaicekauskas, and M. Shur, "Cultural Preferences to Color Quality of Illumination of Different Objects," ArXiv, Physics and Society, San Diego, California, United States, vol. 8835, 8 p., May 2013, doi: 10.48550/arXiv.1305.6308.

[17] M. Kolokotroni, P. Heiselberg, and W. Miller, "Ventilative Cooling and Thermal Comfort," INIVE EEIG, pp. 39-43, Brunel University of London, 2022. ISBN: 978-2-930471-62-4.

[18] A. Tabadkani, A. Roetzel, H. Xian Li, A. Tsangrassoulis, and S. Attia, "Analysis of the impact of automatic shading control scenarios on occupant's comfort and energy load," Applied Energy, vol. 294, p. 21, July 2021, doi: 10.1016/j.apenergy.2021.116904.

[19] J. Seifert, L. Schinke, A. Buchheim, and M. Beyer, "The new climate room for transient investigations of thermal comfort," The REHVA european HVCA Journal, vol. 54, no. 1, pp. 20-22, 2017.

[20] Optical radiation "Digital sensor for color temperature and illuminance FLAD23CCT with ALMEMO® D6 plug," October 2021, [Online]. Available: https://www.mrclab.com/digital-sensor-for-color-temperature-and-illuminance.

[21] J. Shuster, "Addressing Glare in Solid-State Lighting," White Paper WP528001EN, Eaton, January 2017.

[22] Y. Zheliazkov and J. Yamnenko, "Impact of lighting parameters on human comfortable feeling", 2022 IEEE. 63th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), pp. 1-10, 2022, doi: 10.1109/RTUCON56726.2022.9978745.

[23] S. Havassy, C. Jackson, and M. Siminovitch, "Saving Energy in Buildings with Adaptive Lighting Systems: Solutions for the Retail Sector", California Lighting Technology Center, UC Davis, pp. 1-2, June 2015, doi: 10.3390/buildings12081127.

Optimization of the Composition of Operating Units in Power Plants by Genetic Algorithm

Tulkin Gayibov, Sherkhon Latipov and Bekhzod Pulatov
Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan
tulgayibov@gmail.com, sherkhonlatipov@gmail.com, b.pulatov27@gmail.com

Keywords: Optimization, Objective Function, Energy Characteristic, Constraints, Penalty Function, Genetic Algorithm, Composition of Operating Units.

Abstract: One of the main tasks to be solved during planning the short-term modes of power systems is the optimization of compositions of operating units in power plants. In general case, it is a complex problem of nonlinear mathematical programming. Its solution, in essence, comes down to determining for each time interval of the planning period the composition of units to be put into operation or to be stopped. Currently, there are many methods and algorithms for solving of this problem. On the powers obtained at this stage and generalized energy characteristics, the optimal compositions of operating units in power plants are determined. The effectiveness of proposed algorithm is researched on the examples of power systems mode optimization with determination the composition of operating units in power plants. The high accuracy of results of optimization and the reliability of convergence of iterative process is ensured due to direct use in calculations the real, obtained in tabular form, energy characteristics of power plants with effective consideration of functional constraints by penalty functions, as well as due to the ability of genetic algorithm to solve multi-extremal problems without any simplifications.

1 INTRODUCTION

In problems of optimal planning of short-term modes of power system, the unevenness of consumers load schedules sets the conditions for changing the composition of operating units in power plants. In order to reliably provide consumers with high-quality electricity at minimal economic costs, it is advisable to use the optimal composition of operating units in power plants in each time interval of the planning period. Thus, here the task is to determine for each time interval of planning period the composition of units to be put into operation or to be stopped in power plants participating in optimization in accordance with the changing of load schedule,.

Currently, in the existing literature there are many methods and algorithms for solving the problem under consideration [1-13].

Due to the simplicity and ease of using, the most widely used algorithm is based on the preliminary construction of energy characteristics for all the possible combinations of operating units [1, 2]. According to this algorithm, stopping of the unit when the load decreases begins from that one, turning off which provides the greatest savings in fuel costs. By analogy with this in the intervals of increasing the consumer loads, the start-up of the next units begins from that one, in which the maximum savings of fuel costs is ensured.

Thus, the use of such algorithm provides for the multiple solution of the optimization problem to build a series of dependencies for each interval of the power system load change. In addition, the calculation of fuel costs using a simplified method, as well as considering the possibility of successively turning off (or turning on) the units when the load changes, introduces additional errors that reduce the effect of optimization.

The algorithms proposed in [3, 4, 9], which are based on the use of mixed integer linear programming, have also become quite widespread. However, they allow us to obtain the approximate solutions of the problem only.

In [5, 6], the use of classical Lagrange relaxation methods and the selection of options by priority for solving the problem under consideration are proposed. Despite the simplicity of the algorithm, these methods are characterized by problems associated with obtaining a solution with sufficient

accuracy and a rather long duration of the calculation process. In [10], an algorithm for choosing the composition of units in power plants and determining their optimal capacities based on the differential assessment method is described. However, this algorithm has not found wide application for solving practical problems due to some computational difficulties.

In [7, 8], particle swarm optimization algorithms are proposed for solving the problem under consideration. They are based on the use of quadratic energy characteristics of plants, which are obtained by approximation the real characteristics of power plants, usually given in tabular form. These algorithms have good computational qualities and, in most cases, allow us to obtain the optimal solution of the problem with sufficient accuracy. Along with this, in some cases, they encounter difficulties associated with taking into account functional constraints in the form of inequalities, as well as decreasing the accuracy of the results due to the use of quadratic approximation functions of energy characteristics of power plants.

In connection with the indicated shortcomings, which have place for existing methods and algorithms for solving the problem under consideration, the problem of their improvement to overcome the above mentioned difficulties remains an urgent task. The urgency of the problem is also associated with the need to frequently solving of this problem for modern power systems, which include solar and wind power plants with significant capacity [12-14].

Over the past few years, a number of works on the use of artificial intelligence methods, in particular, a genetic algorithm, in solving the problems of optimization of modes of power systems, such as [7, 8, 15-17] have appeared. The positive features of the genetic algorithm, determined by the ability to work with objective functions that have discontinuities, determine the global extremum of multiextremal problems without simplifying them, creates a good conditions for their effective use in solving the problem under consideration.

This paper proposes an effective genetic algorithm for optimization of composition of operating units in power plants, taking into account regime and technological constraints, where many of the difficulties typical for existing algorithms are successfully overcome.

2 METHODS AND MATERIALS

To present the essence of the proposed algorithm, consider a power system containing only thermal power plants (TPPs) which are involved in optimization. Since if there are hydro power plants (HPPs) participating in optimization, they are taken into account using indefinite Lagrange multipliers, which physically represent the equivalents of the fuel costs of water consumption in them. Accordingly, at known values of these multipliers, multiplying them by the energy characteristics of the corresponding HPPs, we obtain equivalent fictitious TPPs (in calculation sense). Further optimization is carried out as for a power system containing only TPPs involved in optimization. The values of these multipliers can be determined as in [18,19]. The proposed algorithm provides for accounting for other types of power plants using existing methods that do not affect its efficiency.

Thus, the objective function, which is a function of the total fuel costs in the TPP of power system for the planning period *T*, has the following form (1):

$$F = \sum_{t=1}^{n_T} \sum_{i=1}^{n} \left[B_{it}(P_{it}, c_{it}) + B_{it}^{S}(\tau_{it}, c_{it}) \right], \qquad (1)$$

where n, n_T are the number of TPPs involved in optimization and time intervals in considered period of planning the power system mode *T*, respectively; P_{it}, B_{it} – active power and fuel costs of *i*-th TPP in *t*-th time interval of the planning period; c_{it} is a combination of operating units in *i*-th TPP in *t*-th interval; τ_{it} is idle time of the *i*-th TPP unit after shuting down and start-up it in the *t*-th time interval; B_{it}^{S} - fuel costs associated with the start-up of the unit of *i*-th TPP (start-up consumption), left in the off state for a while τ_{it}. It also takes into account the costs associated with a reduction the service life of the unit as a result of the next start.

The dependence of the starting fuel consumption on the idle time of the unit is non-linear. However, it has been proven that when the idle time of the unit does not exceed 20 hours, the dependence is approximately linear and it passes through the origins of the coordinate axes. In this case, this function can be represented as $\tau \cdot B_{oi}^{S}(c_i)$ and, in accordance with this, the optimization problem for the planning period can be reduced to the problem of interval optimization, when it is solved separately

for each time interval. In this case the objective function for any interval is represented as follows:

$$F = \sum_{i=1}^{n} \left[B_i(P_i, c_i) + B_{oi}^{S}(c_i) \right], \qquad (2)$$

where $B_{0i}^{S} = \frac{B_i^{S}}{\tau_i}$.

Minimization of (2) by optimization the composition of operating units in TPPs and their respective capacities is carried out taking into account constraints on: the number of simultaneously started (or stopped) units, the balance of active power in power system (3)

$$W = \sum_{i=1}^{n} P_i - P_D - \pi = 0 \qquad (3)$$

the minimum and maximum allowable powers of power plants (4)

$$P_i^{\min} \le P_i \le P_i^{\max}, \qquad (4)$$

the minimum and maximum power flows in controlled power transmission lines (PTL) (5)

$$P_l^{\min} \le P_l \le P_l^{\max}, \quad l \in L, \qquad (5)$$

where W is the function of unbalance of active power in power system; P_D is the total active load in power system; π is the total losses of active power in electrical networks of power system; P_i, P_i^{min}, P_i^{max} are calculated and maximum allowable powers of TPP; L is the set of PTL in which the active power flows are controlled; P_l, $P_l^{\min}$, $P_l^{\max}$ are calculated and specified limit values of the active power flow along the l-th controlled PTL.

In the proposed algorithm, the considered problem of integer programming is solved by methods of continuous mathematical programming. The calculations are carried out in two stages. At the first stage, for each TPP, according to the given energy characteristics of the units and all possible combinations of jointly operating units, a generalized energy characteristic is built. At the second stage, the optimization of the power system mode is carried out according to the generalized energy characteristics of TPP obtained at the first stage, taking into account all regime and technological constraints.

If the relative increments of fuel costs for units in TPP are given, then such characteristic for any particular combination of them is constructed by summing the abscissas of the units included in it by the equality of the relative increments in them.

Let is explain the methodology for construction the generalized energy characteristics of power plants using the example of TPP that has two possible combinations of operating units with the corresponding characteristics of relative increments (CRI) (Figure 1 a) and consumption characteristics (Figure 1 b) of fuel costs. Analyzing these CRIs, we can draw the following conclusions: at plant capacities in P_1^{min}–P_2^{min} range, the single possible composition of operating units is the composition corresponding to the characteristic 1, and in the range of P_1^{max}–P_2^{max} is the composition corresponding to the characteristic 2. To determine the power of the plant, at which it is necessary to switch from one composition of units to another one in range of P_2^{min}–P_1^{max} we should compare the fuel costs B_1 and B_2, determined for the operating conditions of combinations the units with characteristics 1 and 2, respectively. Let's assume that combination of units 2 differs from combination 1 by one included unit. In this case, B_2 should include in its composition the starting consumption of the newly switched on unit B^S_{o2}. For example, if at power P_2^{min} $B_1 > B_2$, then at the same point the transition to the composition with combination 2 is carried out. Otherwise, the transition point is searched in specified range P_2^{min}–P_1^{max}, at which the condition $B_1 = B_2$ is provided. At the transition point from composition of unit 1 to 2 the following equality is carried out (6):

$$\int_{P_1^{min}}^{P_{12}} b_1(P_1)dP_1 = \int_{P_2^{min}}^{P_{12}} b_2(P_2)\cdot dP_2 + B_{02}^{S}, \qquad (6)$$

where P_{12} is the power of the plant at the transition point; B_{02}^{S} is the starting consumption of fuel costs for the next unit to composition 1.

If $B_1 < B_2$ remains within the range $P_2^{\min}$–$P_1^{\max}$, then the power at which the forced transition to the composition with combination 2 is carried out is P_1^{max}.

Figure 1 shows the generalized CRI obtained on the basis of the calculation according to the described algorithm (Figure 1 a) and the corresponding consumption characteristics of fuel costs (Figure 1 b) for the example under consideration.

In general case, in modern power plants there can be many combinations of compositions of operating units. Accordingly, obtaining a generalized energy characteristic for the plant by the described algorithm becomes a time-consuming, but solvable task. On the other hand, in many plants the simultaneous start-up of more than one unit is not allowed because technical reasons. In such cases, the construction of a generalized characteristic based on

the determination of transition points to the next compositions of units is simplified by reducing the number of compared variants of combinations.

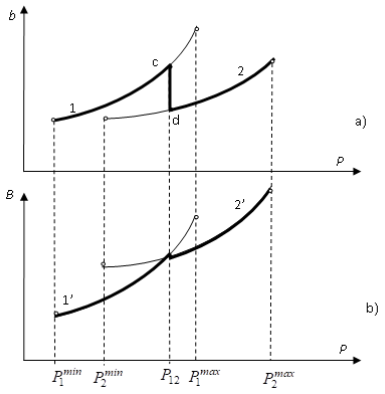


Figure 1: Construction of generalized characteristics of relative increments - a) and consumption characteristics of fuel costs - b) of TPP.

Thus, for the entire control range of each power plant, all transition points from one composition to another are determined. As a result, usually, sawtooth similar CRIs for plants are obtained (Figure 1 a). Accordingly, the resulting problem of optimization of mode of power system with such CRI of plants becomes multi-extremal.

After obtaining the generalized energy characteristics of power plants, the second stage of calculation is performed. Where the objective function which is the sum of the generalized consumption functions of TPPs involved in optimization is minimized (7):

$$F = \sum_{i-1}^{n} B_i(P_i) \to \min \qquad (7)$$

subject to all the above presented constraints.

The proposed algorithm provides for solving the last optimization problem by genetic algorithm. At the same time, to improve the convergence of the iteration process, the constraint on the balance of active power in power system (2) is taken into account by the allocation of a balancing plant. In this case, the power of this plant goes into the number of dependent variables. Therefore, the constraints on its maximum and minimum allowable values are taken into account using penalty functions in exponential form (8), (9):

$$P_1 = P_н + \pi - \sum_{i=2}^{n} P_i\,, \qquad (8)$$

$$G_{\max} = \alpha e^{\beta\left(P_1 - P_1^{\max}\right)},\ G_{\min} = \alpha e^{\beta\left(-P_1 + P_1^{\min}\right)}, \qquad (9)$$

where α, β are weight (penalty) coefficients.

Accounting of functional constraints in the form of inequalities is also carried out by exponential penalty functions (10):

$$G_{l.\max} = \alpha_l e^{\beta_l\left(P_l - P_l^{\max}\right)}\ G_{l.\min} = \alpha_{l.} e^{\beta_{l.}\left(-P_l + P_l^{\min}\right)} \qquad (10)$$

The generalized objective function, which is minimized by the genetic algorithm, taking into account constraints (4), has the following form (11):

$$F = \sum_{i=1}^{n} B_i(P_i) + G_{\max} + G_{\min} + \sum_{l=1}^{L_p} G_{l.\max} + \sum_{l=1}^{L_p} G_{l\min} \to \min \qquad (11)$$

According to the optimal capacities of powers and generalized energy characteristics obtained as a result of solving the last problem, the optimal compositions of the operating units in them are determined.

Due to the performance of optimization using the real, usually given in tabular form, energy characteristics of plants, their powers obtained at the second stage of calculation are automatically optimal for certain already optimal compositions of operating units in them. Unlike many existing algorithms for solving of this problem, it is not required to perform additional optimization of power system mode using energy characteristics for the selected optimal compositions of units in plants.

Thus, the calculations according to the proposed algorithm are performed in the following order:

1) construction of the generalized energy characteristics of power plants based on the determination of transition points in the CRI of individual combinations of units;
2) optimization of power system mode using the generalized energy characteristics of plants and taking into account all the constraints by genetic algorithm;
3) determination of the optimal compositions of operating units in power plants according to the corresponding ones obtained as a result of implementation of Section 2, plants capacities.

3 RESULTS AND DISCUSSION

The efficiency of the proposed algorithm was researched on the example of optimization the composition of operating units in the problem of optimal distribution of power system load of 800 MW between three TPPs with two units of the same type. The CRI of fuel equivalent consumption of one unit and reference fuel equivalent consumption at its minimum load are shown in Table 1.

Table 1: The characteristics of relative increments of fuel equivalent consumption of TPP units.

TPP – 1 B_0=35,2 t.f.e./h.	P_1, MW	100	108	115	130	159	188	210
	b_1, $\frac{t.f.e.}{MW \cdot h.}$	0.301	0.303	0.305	0.307	0.309	0.330	0.332
TPP – 2 B_0=28,5 t.f.e./h.	P_2, MW	80	96	109	110	160		
	b_2, $\frac{t.f.e.}{MW \cdot h.}$	0.316	0.317	0.319	0.334	0.3341		
TPP – 3 B_0=50,0 t.f.e./h.	P_3, MW	150	240	270	271	300		
	b_3, $\frac{t.f.e.}{MW \cdot h.}$	0.295	0.296	0.297	0.300	0.3001		

To compare the calculation results, Table 2 shows the results of optimal distribution of power system load at all the possible compositions of operating units in TPPs.

Table 2: The results of the optimal distribution of power system load between TPPs at various possible compositions of operating units.

TPP-1		TPP-2		TPP-3		Total fuel consumption, t.f.e./h.
Number of operating units	P_1, MW	Number of operating units	P_2, MW	Number of operating units	P_3, MW	
1	120.0	1	80.0	2	600.0	258.75
1	210.0	2	290.0	1	300.0	263.73
1	100.0	2	160.0	2	540.0	263.18
2	390.22	1	109.78	1	300.0	262.41
2	200.0	1	80.0	2	520.0	263.94
2	337.54	2	162.46	1	300.0	264.91
2	200.0	2	160.0	2	440.0	268.75

Comparing the results given in Table 2, we determine the mode in which the total consumption of fuel equivalent in TPPs is minimum 258.75 t.f.e./ h. and the optimal compositions of operating units in them are 1, 1, 2, respectively. Table 3 shows the results of solving of the problem under consideration by the algorithm proposed here.

Comparing the results obtained in Table 3 with the reference one given in Table 2, we can verify the high accuracy of the proposed algorithm.

The effectiveness of the proposed algorithm was also studied in more complex example, where it is required to determine the optimal compositions of operating units in three TPPs with 5, 12 and 7 units of the same type, respectively, during the day with optimal coverage of the daily load schedule of power system shown in Table 4.

Table 3: The results of optimizing the composition of operating units by the proposed algorithm.

TPP-1		TPP-2		TPP-3		Total fuel consumption, t.f.e./h.
Number of operating units	P_1, MW	Number of operating units	P_2, MW	Number of operating units	P_3, MW	
1	120.0	1	80.0	2	600.0	258.75

Table 4: Daily load schedule of power system.

t, h.	1	2	3	4	5	6
P_{Dt}. *MW*	3650	3500	3450	3550	3650	3800
t, h.	7	8	9	10	11	12
P_{Dt}. *MW*	3900	4000	4100	4200	4150	4050
t, h.	13	14	15	16	17	18
P_{Dt}. *MW*	3950	3900	3960	4000	4100	4200
t, h.	19	20	21	22	23	24
P_{Dt}. *MW*	4400	4550	4590	4500	4200	3900

Table 5 shows the results of optimization the composition of operating units in TPPs, by the proposed algorithm for several hours of the day with total loads of 3500 MW, 4000 MW, 4500 MW and 4200 MW.

Table 5: The results of optimization the composition of operating units in TPPs.

t, h.	TPP -1		TPP-2		TPP-3		Total fuel consumption for the *t*-th interval, t.f.e./h.
	Number of operating units	*P*, MW	Number of operating units	*P*, MW	Number of operating units	*P*, MW	
1	5	1050	3	350	7	2100	1131.3
2	5	1050	6	850	7	2100	1302.1
3	5	1050	9	1350	7	2100	1473.0
4	5	1050	7	1050	7	2100	1370.2

In order to evaluate the effectiveness of the proposed algorithm, reference results are also obtained based on a simple selection and comparison of them for all possible combinations of operating units in power plants. Comparing them with the results shown in Table 5 showed their complete agreement.

Thus, the proposed algorithm for optimization the composition of operating units in power plants based on genetic algorithm allows us to reliably obtain the solution of the problem with high accuracy. A characteristic feature of genetic algorithm makes it possible to directly use the generalized energy characteristics of plants without their correction in order to reduce them to a convex programming problem. This ensures an increase the accuracy of optimization and, accordingly, the effect of optimization.

4 CONCLUSION

The paper has accomplished the following:

1) A new algorithm for optimization the composition of operating units in power plants of power system based on genetic algorithm, which allows us to obtain the optimal solution of the problem, taking into account regime and technological constraints in the form of equalities and inequalities with sufficient reliability and accuracy is proposed.
2) An increase the accuracy of the results when using the proposed algorithm is ensured by optimization with the direct use the real energy characteristics of power plants, which are usually specified in tabular form.
3) To implement the described algorithm for optimization the composition of operating units in power plants with direct use of their energy characteristics specified in tabular form, it is advisable to apply a genetic algorithm with real coding of variables.

REFERENCES

[1] A. Bhardwaj, V.K. Kamboj, V.K. Shukla, B. Singh, and P. Khurana, “Unit commitment in electrical power system - a literature review,” in IEEE International Power Engineering and Optimization Conference (PEOCO), Melaka, Malaysia, 2012, pp. 275-280, doi: 10.1109/PEOCO.2012.6230874.

[2] E.B. Saitov and T.B. Sodiqov, “Modeling an Autonomous Photovoltaic System in the Matlab Simulink Software Environment,” in AIP Conference Proceedings, vol. 2432, 2022, 020022, doi: 10.1063/5.0057797.

[3] E.B. Saitov, Sh. Kodirov, B.M. Kamanov, N. Imomkulov, and I. Kudenov, “Increasing the Efficiency of Autonomous Solar Photovoltaic Installations for Power Supply of Agricultural Consumers,” in AIP Conference Proceedings, vol. 2432, 2022, 040036, doi: 10.1063/5.0057798.

[4] F. Zikrillayev, E.B. Saitov, J.B. Toshov, B.K. Ilyasov, and M.B. Zubaydullayev, “A Software Package for Determining the Optimal Composition and Parameters of a Combined Autonomous Power Supply System Based on Renewable Energy Sources,” in AIP Conference Proceedings, vol. 2432, 2022, 020021, doi: 10.1063/5.0057802.

[5] A. Bhardwaj, “Unit Commitment in Power System: A review,” International Journal of Electrical and Power Engineering, vol. 6, no. 1, pp. 51-57, 2012, doi: 10.11591/telkomnika.v11i1.238.

[6] N.P. Padhy, “Unit commitment – a bibliographical survey,” IEEE Transactions on Power Systems, vol. 19, no. 2, pp. 1196-1205, 2004, doi: 10.1109/TPWRS.2004.825866.

[7] H.Y. Yamin, “Review on methods of generation scheduling in electric power systems,” Electric Power Systems Research, vol. 69, no. 2-3, pp. 227-248, 2004, doi: 10.1016/j.epsr.2003.12.007.

[8] S. Virmani, E.C. Adrian, K. Imhof, and S. Mukherjee, “Implementation of a Lagrangian relaxation based unit commitment problem,” IEEE Transactions on Power Systems, vol. 4, no. 4, pp. 1373-1380, 1989, doi: 10.1109/59.32645.

[9] K. Senjyu, T. Shimabukuro, K. Uezato, and T. Funabashi, “A fast technique for unit commitment problem by extended priority list,” IEEE Transactions on Power Systems, vol. 18, no. 2, pp. 882-888, 2003, doi: 10.1109/TPWRS.2003. 811140.

[10] M. M. Morato, J. D. Vergara-Dietrich, P. R. C. Mendes, J. E. Normey-Rico, and C. Bordons, “A Two-Layer EMS for Cooperative Sugarcane-based Microgrids,” Int. J. Electr. Power Energy Syst., vol. 118, p. 105752, 2020.

[11] B. O. Anyaka, J. F. Manirakiza, K. C. Chike, and P. A. Okoro, “Optimal unit commitment of a power plant using particle swarm optimization approach,” Int. J. Electr. Comput. Eng., vol. 10, no. 2, pp. 1135-1141, Apr. 2020.

[12] Y. Wang, J. Yan, J. Li, Z. Li, and W. Zhang, “A new model of economic dispatch considering energy conservation and environmental protection in electricity market,” Energy Procedia, vol. 17, pp. 1769-1777, 2012.

[13] R. Storn and K. Price, “Differential evolution – a simple and efficient adaptive scheme for global optimization over continuous spaces,” Tech. Rep. TR-95-012, Int. Comput. Sci. Inst., Berkeley, CA, USA, 1997.

[14] M. F. Anjos, “Recent Progress in Modeling Unit Commitment Problems,” in Model. Optim. Theory Appl., New York, NY, USA: Springer, 2013, pp. 1-29.

[15] N. Troy, E. Denny, and M. O’Malley, “Base-load cycling on a system with significant wind penetration,” IEEE Trans. Power Syst., vol. 25, no. 2, pp. 1088-1097, May 2010.

[16] T. Gayibov and B. Pulatov, “Optimization of Short-term Modes of Hydrothermal Power System,” E3S Web Conf., vol. 209, p. 07014, 2020, doi: 10.1051/e3sconf/202020907014.

[17] T. Gayibov and E. Abdullaev, “Optimization of daily operation mode of photovoltaic systems of enterprises,” E3S Web Conf., vol. 264, p. 04063, 2021, doi: 10.1051/e3sconf/202126404063.

[18] T. M. Mohan and T. Nireekshana, “A Genetic Algorithm for Solving Optimal Power Flow Problem,” in 2019 3rd Int. Conf. Electron. Commun. Aerospace Technol., Coimbatore, India, 2019, pp. 1438-1440, doi: 10.1109/ICECA.2019.8822090.

[19] C.-L. Chiang, “Improved genetic algorithm for power economic dispatch of units with valve-point effects and multiple fuels,” IEEE Trans. Power Syst., vol. 20, no. 4, pp. 1690-1699, Nov. 2005, doi: 10.1109/TPWRS.2005.857924.

Hydraulic Energy Storage of Wind Power Plants

Boborakhim Urishev[1] and Fakhriddin Nosirov[2]
[1]*Karshi Engineering-Economics Institute, Mustakillik Str. 225, Karshi, Uzbekistan*
[2]*Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan*
bob_urishev@mail.ru, nosirov@mail.ru

Keywords: Wind Power Station, Pumped Storage Power Plant, Energy Storage, Upper Reservoir, Renewable Energy, Upper Reservoir.

Abstract: The article discusses information on the need to accumulate energy from renewable sources to improve their efficiency, as well as some examples of the integration of systems for hydraulic energy storage and renewable sources, which ensure an increase in the reliability and volume of energy generation. The method for determining the parameters of the hydraulic energy storage system of a wind power plant, which is based on the balance of the daily load produced and spent on energy storage, is presented. This technique, with changing daily loads, makes it possible to determine the main parameters of the complex, including the volume of accumulated water, the coefficient of energy use of the wind power station. A functional diagram of the programmed control of the parameters of a pumped storage and wind power plant for the optimal use of the wind potential in hydraulic energy storage presented. Based on the results of calculations using the proposed method, the main parameters of the system based on a pumped storage and wind power plant with a capacity of 100 MW were determined, the efficiency of hydraulic energy storage was determined in comparison with lithium-ion batteries.

1 INTRODUCTION

Traditional centralized power systems have significant drawbacks, such as significant energy losses due to the remoteness of some consumers, insufficient flexibility of the production process caused by its low maneuverability, lack of proper regulation of consumption and energy tariffication. All these disadvantages ultimately lead to an increase in fuel consumption, the degree of CO_2 emissions and the cost of energy produced. In this regard, recently, localized power systems such as micro grid, smart grides, distributed generation clusters and virtual power plants are increasingly being used. Such systems can help solve very important problems, such as optimization, stabilization, flexibility of the power system, integration of renewable energy sources and “smart” control centers in the process of energy production and distribution [1,2,3]. Accumulation of an excess part of energy for use in peak hours is a necessary procedure for local power systems, for which electrochemical, regenerative - fuel and other energy storage devices are most often used.

Currently, the method of hydraulic energy storage has the best performance among storage devices and is considered as an effective direction for expanding the possibilities of using renewable energy sources, despite the fact that the decisive factors are the availability of sufficient natural resources, conditions for the construction of reservoirs and obtaining the necessary pressure [4,5].

The development of integration of systems for the use of wind power plants and hydraulic storage can lead to an increase in the share of wind energy on average up to 20% in the total energy consumption [6]. Investigating this issue depending on the power consumption, the author of [7] asserts that up to 80% of wind energy can be realized in power supply systems, provided that pumped storage is used. An increase in the utilization rate of wind energy is also observed with an increase in the duration of energy storage, so, for example, up to 90% of wind energy can be used to accumulate water in the upper reservoir for 1 ... 3 days [8].

In [9], an overview of the prospects for the development of hydraulic energy storage in the light of sustainable development was presented. According to this review, the most promising direction is the hybrid use of pumped storage power plants (PSP) of low power in a complex with wind and solar power

plants. This is confirmed by the results of the study, indicating an increase in the reliability and volume of energy generation by hybrid solar and wind power plants compared to their individual work [10].

The authors of [11, 12, 13] came to the conclusion that an integrated system for the use of wind energy and PSP under certain conditions is the most economically and technically competitive technology in different geographic latitudes.

In [14], the issues of large-scale integration of renewable energy sources with the use of pumped storage power plants into the Irish power system are considered on the example of the use of wind energy. The main key issues were considered: capital costs, optimal parameters and aspects of operation. The results of the study showed that the PSP allows increasing the efficiency of the use of renewable energy sources and reduces operating costs.

The integration of a PSP into a wind power plant system in Spain increases the profitability of this integrated system and minimizes wind energy losses [15].

The work [16,17,18] studied the possibility of maximum use of wind energy in combination with existing autonomous thermal power plants by using pumped storage power plants in the isolated energy system of the islands of Karpathos and Kasos, located in the Southeast Aegean Sea. The results of this study proved the economic feasibility of using low-power wind turbines to power pumping units of pumped storage power plants, even without taking into account the undoubted environmental benefits.

With the optimal integration of wind energy with a pumped storage power plant in lake Turkana (Kenya), the hydraulic energy storage of the wind turbine made it possible to reduce the energy deficit to 46%, which is equivalent to receiving an income of more than 10 thousand dollars per day [19].

2 METHODS AND MATERIALS

A schematic of a low-power PSP with a wind turbine is shown in Figure 1.

According to this scheme, the main energy producer is a wind power station (WPS), consisting of several wind turbine, which, in hours of minimum energy consumption, having excess power, supplies power to pumping units (PU) of a pumped storage power plant for accumulating water in the upper reservoir. During peak hours of energy consumption, when the power of the wind turbine is not sufficient, water from the upper reservoir is supplied to the hydro turbines to generate energy, and then to the lower reservoir.

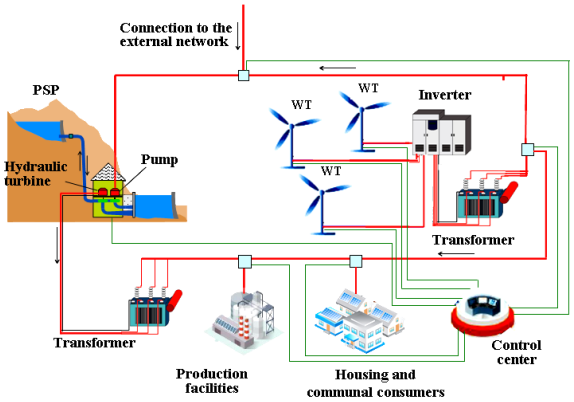


Figure 1: Scheme of low power pumped storage power plant with wind farm.

With known values of the power of the wind flow, the installed power of the wind farm must be determined based on the values of the daily load schedule. Assuming that the energy generated by the wind turbines is used for two purposes: the main part of the energy is directed to the needs of the consumer, and the excess is used for storage, i.e. to power the PU. The average graphs of the daily load and energy of WPS with a maximum power of 100 MW are shown using the example shown in Figure 2.

For this example, the following expression is valid, which follows from the (1) for balancing the daily load and the generated energy.

$$E_{WPS} \cdot \eta_E = \int_0^T P(t)dt, \quad (1)$$

where E_{WPS} – generated energy by WPS over time T; η_E – efficiency in the process of transferring energy to the consumer; $P(t)$ – load consumption values at points in time t. From here we can determine the average power of the WPS over time T (2):

$$N_{WWS\ av} = \frac{\int_0^T P(t)dt}{T \cdot \eta_E}. \quad (2)$$

The installed capacity of a wind farm can be determined by the following (3):

$$N_{WPS} = N_{WT} \cdot n, \quad (3)$$

where n – number of wind turbines.

Determination of the type, brand and number of installed wind turbines must be carried out on the basis of a feasibility study, taking into account local conditions on the basis of considering several options.

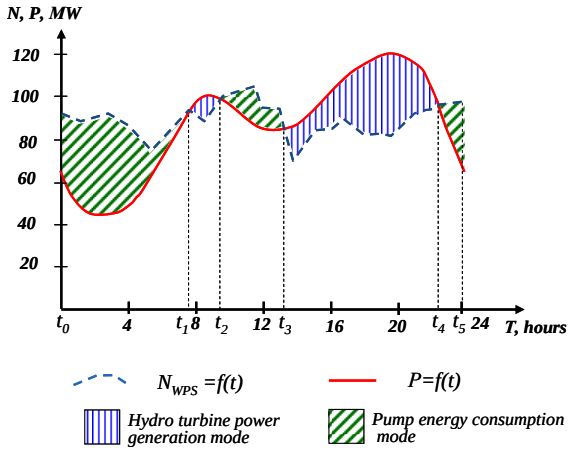


Figure 2: Schedule of the operating mode of a low-power PSP with a WPS.

All options must meet the condition specified in (1).

The obtained values of the power of the WPS N_{WPS}, depending on the time t, can be used to determine the mode of operation of the PSP with the WPS (see Figure 2 graph $N_{WPS}=f(T)$).

The condition for the energy balance of a PSP with a WPS in its pumping mode, i.e. when there is an excess of the generated energy over consumption, it is described by the following balance (4)

$$\sum_{i=1}^{n} N_{\text{WPSi}} t_i - \sum_{i=1}^{n} P_i \cdot T = \sum_{i=1}^{n} N_{\text{Pi}} t_i = E_P \,, \quad (4)$$

N_P, E_P - power and energy consumption in pumping mode.

From here we determine the maximum volume of the upper reservoir according to the well-known (5)

$$V_{\max} = \frac{367 \cdot \eta_{\text{pum}} \left(\sum_{i=1}^{n} N_{\text{WPSi}} t_i - \sum_{i=1}^{n} P_i \cdot t_i \right)}{H_{Pi}} . \quad (5)$$

From (5) it can be seen that the volume of the reservoir V is determined in direct dependence on the amount of energy that must be accumulated, and in inverse dependence on the pressure H_P.

The volume of water in the upper and lower reservoirs during the day will be variable, and depends on the operating time of the PSP in turbine and pumping modes. To assess changes in water volumes, graphs $W_{lev}=f(V)$ are used, characterizing the dependence of the water level W_{lev} in reservoirs on their volumes V.

The excess capacity of the PSP $N_P = N_{WPS} - P$ must be used to power the PSP pumping the volume of water ΔV_H into the upper reservoir, for example, according to the graph in Figure 2 during the time from t_0 to t_1 , t_2 to t_3 and from t_4 to t_5.

When there is not enough WPP power to cover the daily load schedule, the lack of power $N_H = P - N_{WPS}$ is compensated by transferring the PSP to the mode of power generation with water supply to the hydro turbines, which will generate the missing electricity. For example, according to the schedule in Figure 2 during the time from t_1 to t_2 and t_3 to t_4. It is convenient to present the calculation results in tabular form (Table 1).

The amount of electricity is determined by discrete summation of the products $N \cdot \Delta t$ in fixed time intervals, for example, for the period $t_{0\text{-}1}$

$$E_{P\ (0-1)} = \sum_{i=1}^{n} N_{Pi} \cdot \Delta t_i \,,$$

and for the period $t_{1\text{-}2}$

$$E_{H\ (1-2)} = \sum_{i=1}^{n} N_{Hi} \cdot \Delta t_i$$

The volume of water supplied by the PU to the upper reservoir in the first period of operation an be determined as follows (6)

$$\Delta V_{P(0-1)} = \frac{367 \cdot E_{P\ (0\text{-}1)} \cdot \eta_{pum}}{H_{P(0-1)}} , \quad (6)$$

where $H_{P(0\text{-}1)}$, η_{pum} – head and efficiency of pumped storage power plant in pumping mode.

With the alternating supply and withdrawal of water from reservoirs, the water level in them changes and, therefore, the pressure values of the PSP will be different and they are determined by the following relationship (7) [20],

$$H_{Hi,Pi} = H^{G}_{\text{Hi,Pi}} \pm \Delta H_i \,, \quad (7)$$

where ΔH_i is the pressure loss in the water conduits, which depend on the flow rate of the pump or turbine Q and can be calculated using well-known hydraulic calculation methods. In the above formula, the (+) sign corresponds to the pumping mode, and the (-) sign to the turbine mode, H^G – geometric head.

When the time of energy insufficiency of the WPP $t_{1\text{-}2}$ comes, the HP will be stopped and the hydro power plants (HPP) will be put into operation with the supply of a volume of water $\Delta V_{H(1\text{-}2)}$ (8) (Figure 2 and Table 1).

$$\Delta V_{H\,(1-2)} = \frac{367 \cdot E_{H\ (1-2)}}{H_{H\,(1-2)} \cdot \eta_{tur}} \,, \quad (8)$$

$H_{H(1\text{-}2)}$, η_{tur} – head and efficiency of pumped storage power plant in turbine mode.

Table 1: Determination of the capacity of the PSP.

Time of day	Daily load of electric power P, kW	WPS power N_{WPS}, kW	Pumping power regime, N_P, kW	Turbine power, N_H, kW	The amount of electricity in pumping mode, E_P, kW/h	The amount of electricity in turbine mode, E_H, kWh	Volume water supply upper reservoir, V_P , m^3	Volume water supply lower reservoir, V_H, m^3
t_{0-1}	$P_{i\,(0-1)}$	$N_{WPS\,i\,(0-1)}$	$N_{P\,i\,(0-1)}$	-	$E_{P(0-1)}$	-	$\Delta V_{P(0-1)}$	-
t_{1-2}	$P_{i\,(1-2)}$	$N_{WPS\,i(1-2)}$	-	$N_{Hi(1-2)}$	-	$E_{H(1-2)}$	-	$\Delta V_{H(1-2)}$
t_{2-3}	$P_{i\,(2-3)}$	$N_{WPS\,i(2-3)}$	$N_{P\,i\,(2-3)}$	-	$E_{P(2-3)}$	-	$\Delta V_{P(2-3)}$	-
t_{3-4}	$P_{i\,(3-4)}$	$N_{WPS\,i(3-4)}$	-	$N_{H\,i(3-4)}$	-	$E_{H(3-4)}$	-	$\Delta V_{H(3-4)}$
t_{4-5}	$P_{i\,(4-5)}$	$N_{WPS\,i(4-T)}$	$N_{P\,i(4-T)}$	-	$E_{P(4-5)}$	-	$\Delta V_{P(4-5)}$	-

The volume of water in reservoirs corresponding to the operating mode of the PSP and WPP, shown in Figure 2 can be determined by the maximum value of ΔV_H or ΔV_P according to Table 1.

The water flow rate of the PSP in turbine and pumping modes Q_H and Q_P at fixed times t_i can be determined based on the values of powers and heads according to the following formulas.

$$Q_{\text{Hi}} = \frac{N_{Hi}}{9{,}81 \cdot \eta_{tur} \cdot H_{Hi}},$$

$$Q_{\text{Pi}} = \frac{N_{Pi} \cdot \eta_{\text{pum}}}{9{,}81 \cdot H_{Pi}} \qquad (9)$$

One of the specific energy indicators is the utilization rate of the generated and consumed energy, calculated according to the following relationship.

$$K_t = (E_{WPS} - E_P + E_H) / E_{WPS} \qquad (10)$$

Thus, we can summarize that the above method allows you to determine the parameters of the pumped storage power plant, the maximum corresponding to the capacity of the wind farm and the daily electrical load.

For this complex, a functional diagram of the execution of operations for the programmed control of the parameters of PSP and WPP has been developed (Figure 3).

At the first stage of the complex functioning, in order to resolve the issue of "accumulate or generate energy", the power value of the N_{WPS} at time t compared with the corresponding power of the daily load schedule P, because of which one of three possible operations performed.

a) $N_{WPS} = P$, at the same time, the PSP does not work, the loads are covered by the energy of the WPS.
b) in the case of $N_{WPS} > P$, an excess of energy is observed, which must be directed to feed the PU in order to accumulate water in the upper reservoir.
c) the state of the $N_{WPS} < P$ shows the lack of generated energy, which precedes the start of the power plant with water supply from the upper reservoir.

Thus, the main task of the first stage is to determine the time and conditions for the start-ups of the HPP and PU according to the parameters of the N_{WPS} and P.

The second stage begins with the start-up of the PU with a full or almost full volume of water in the lower reservoir. During the operation of the pumps, it is necessary to control the water level in the lower reservoir and the power of the pump in order to perform the following operations:

- when the water level drops to the level of the dead volume, the PU is switched off;
- if N_P does not correspond to the P_H value, which is characterized as the difference in the capacities of the $P_H = N_{WPS} - P$ (excess power of the WPP, directed to power the PU), an operation is performed to increase or decrease the flow rate of water supplied to the PU in order to achieve the correspondence $N_P = P_P$.

At the third stage, when $N_{WPS} < P$, in order to replenish the missing part of the load, the hydro power plant is launched. During the period of operation of the HPP, N_H is regularly measured and it is compared with the value of P_H, which characterizes the difference $P_H = P - N_{WPS}$. If these values are equal, this means that the power of the power plant is sufficient to cover the load curve and the position of the water level in the upper reservoir is checked, i.e.

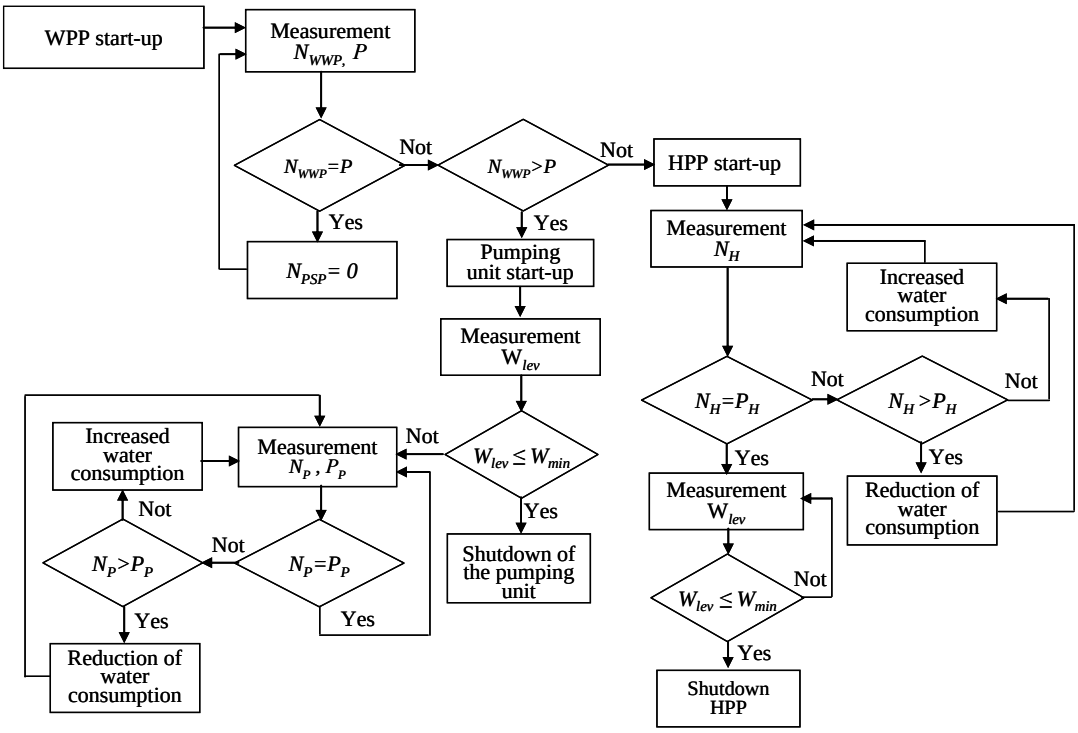


Figure: 3. Functional diagram of PSP with WPS.

$W_{lev} \le W_{min}$. If the water level drops below W_{min}, the power plant is turned off and the pumped storage power plant switches to pumping mode. In case of non-compliance of $N_H \neq P_H$, a command is given to increase or decrease the flow rate of water supplied to the power plant to achieve the balance $N_H = P_H$.

3 RESULTS AND DISCUSSION

Consider a certain local power system with a daily demand from 800 to 1050 MW·h based on a wind farm, generating 850÷1100 MW×h of electricity per day, the schedules of $N_{WPS} = f(T)$ and $P = f(T)$ of which are shown in Figure 2. Based on the results of calculations performed using the characteristic graphs $N_{WPS} = f(T)$ and $P = f(T)$, the following indicators were obtained:

1) The amount of accumulated energy is E_P = 110÷150 MW·h, and the generated energy E_H = 76÷104 MW×h.
2) Reservoir volume V=8 000 000÷10 230 000 m3
3) Installed capacity of WPP N_{WPS} =100 MW
4) Installed capacity of PU and HPP N_P = 55 MW, N_H = 35 MW
5) The average operating time of the PU is 12.5 hours, and the HPP – 11,5 hours.
6) Average head H_P =40,0 m, H_H = 38,5 m.

The calculation results show that the amount of accumulated energy is 13,0 ÷13,6 % of the generated, for which it is necessary to build a 55 MW PSP with two reservoirs with volumes of 8,0 ÷10,23 million m^3. The energy utilization factor of the WPS is 0,91 ÷0,96, which is one of the highest indicators for this type of power plant [7].

In order to answer the question of how economically beneficial it is, we compare the costs of hydraulic energy storage with the costs of its storage by lithium-ion batteries, which are one of the most common storage devices in 2021. According to the financial consulting company Lazard, given in [21], the LCOS of lithium-ion batteries of the LIP type with a capacity of 100 MW is 146... MWh. These costs take into account the cost of storing power for 4 hours for lithium-ion batteries, and for 10 hours for a pumped storage power plant. In addition, it should be noted that the average lifetime of lithium-ion batteries is 10 years, and PSP is 40 years, and in this regard, these plants have advantages despite high capital costs (capital costs of PSP 2623 $ / kW, Li-ion batteries - 1541 $/kW for the same power [22]). If we take into account the difference in the LCOS values of both installations (for example, with their difference of 65 $/MWh), the economic effect from the use of PSP in the considered energy system compared to lithium-ion batteries is $7150÷9750 per day.

4 CONCLUSIONS

The paper has accomplished the following:

1) A method has been developed for determining the parameters of the PSP and WPP complex, based on ensuring the balance of generated, consumed and accumulated energy in the power system.
2) A functional diagram of the programmed control of the parameters of pumped storage power plants and wind power plants has been developed, which makes it possible to achieve optimal use of the wind potential by hydraulic energy storage.
3) The results of comparative calculations of the costs of energy storage have shown the economic efficiency of PSPP in comparison with lithium-ion batteries, estimated at $7150÷9750 per day.

ACKNOWLEDGMENTS

The work was financially supported by the ministry of higher education, science and innovation of the Republic of Uzbekistan within the framework of the F3-ot-2021-235 project “Theoretical foundations for the development of hydropower using hydropower complexes”.

REFERENCES

[1] P. Alstone, D. Gershenson, and D. M. Kammen, “Decentralized energy systems for clean electricity access,” Nature Climate Change, vol. 5, pp. 305-314, 2015, [Online]. Available: https://rael.berkeley.edu/wp-content/uploads/2016/04/Alstone-Gershenson-Kammen-NatureClimateChange-2015-EnergyAccess.pdf.

[2] United States Environmental Protection Agency, “Distributed Generation of Electricity and its Environmental Impacts,” Washington, DC 20460, 2019, [Online]. Available: https://www.epa.gov/energy/distributed-generation-electricity-and-its-environmental-impacts#ref1.

[3] B. Urishev, “Decentralized Energy Systems, Based on Renewable Energy Sources,” Applied Solar Energy, vol. 55, no. 3, pp. 207-212, 2019, doi: 10.3103/S0003701X19030101.

[4] S. Rehman, L. M. Al-Hadhrami, and M. M. Alam, “Pumped hydro energy storage system: A technological review,” Renewable and Sustainable Energy Reviews, no. 44, pp. 586-598, 2015, [Online]. Available: http://www.inme.sjtu.edu.cn/Upload/Files/2017070314353216681 28.pdf.

[5] J. K. Kaldellis and K. A. Kavadias, “Optimal wind-hydro solution for Aegean Sea islands' electricity-demand fulfillment,” Applied Energy, vol. 70, no. 4, pp. 333-354, 2001.

[6] D. Weisser and R. S. Garcia, “Instantaneous wind energy penetration in isolated electricity grids: concepts and review,” Renewable Energy, vol. 30, no. 8, pp. 1299-1308, 2005.

[7] T. Jacob, “Pumped storage in Switzerland an outlook beyond 2000,” Stucky Consulting Engineers, The Economist, pp. 345-363, 2014.

[8] G. C. Bakos, “Feasibility study of a hybrid wind/hydro power-system for low-cost electricity production,” Applied Energy, vol. 72, no. 3-4, pp. 599-608, 2002.

[9] G. Ardizzon, G. Cavazzini, and G. Pavesi, “A new generation of small hydro and pumped-hydro power plants: advances and future challenges,” Renewable Sustainable Energy Rev., vol. 31, pp. 746-761, 2014.

[10] C. E. Hoicka and I. H. Roewlands, “Solar and wind resource complementarity: advancing options for renewable electricity integration in Ontario, Canada,” Renewable Energy, vol. 36, pp. 97-107, 2011.

[11] K. G. Rados, A. Zervos, and G. K. Stamtsis, “On the market of wind with hydro-pumped storage systems in autonomous Greek islands,” Renewable Sustainable Energy Rev., vol. 14, pp. 2221-2226, 2010.

[12] B. Dursun and A. Bora, “The contribution of wind-hydro pumped storage systems in meeting Turkey's electric energy demand,” Renewable Sustainable Energy Rev., vol. 14, no. 7, pp. 1979-1988, 2010.

[13] A. Tuohy and M. O'Malley, “Pumped storage in systems with very high wind penetration,” Energy Policy, vol. 39, no. 4, pp. 1965-1974, 2011.

[14] D. Connolly, H. Lund, B. V. Mathiesen, E. Pican, and M. Leahy, “The technical and economic implications of integrating fluctuating renewable energy using energy storage,” Renewable Energy, vol. 43, pp. 47-60, 2012.

[15] A. K. Varkani, A. Daraeepour, and H. Monsef, “A new self-scheduling strategy for integrated operation of wind and pumped-storage power plants in power markets,” Appl Energy, vol. 88, pp. 5002-5012, 2011.

[16] D. A. Katsaprakakis, D. G. Christakis, K. Pavlopoylos, et al., "Introduction of a wind pumped storage system in the isolated insular power system of Karpathos – Kasos," Appl Energy, vol. 97, pp. 38-48, 2012.

[17] D. A. Katsaprakakis, D. G. Christakis, I. Stefanakis, P. Spanos, and N. Stefanakis, “Technical details regarding the design, the construction and the operation of seawater pumped storage systems,” Energy, vol. 55, pp. 619-630, 2013.

[18] M. W. Murage and C. L. Anderson, “Contribution of pumped hydro storage to integration of wind power in Kenya: an optimal control approach,” Renewable Energy, vol. 63, pp. 698-707, 2014.

[19] B. Urishev, “Selection of Parameters of Pumped Storage Power Plants at Large Pumping Stations for Water Use,” Applied Solar Energy, vol. 54, no. 6, pp. 477-480, 2019, doi: 10.3103/S0003701X18060166.

Research of Ferroresonance in 6-35 kV Electrical Networks Taking Into Account the Dynamic Model of Non-Linear Inductivity of Power Transformer

Shavkat Begmatov
Tashkent State Technical University Islam Karimov, University Str. 2, Tashkent, Uzbekistan
shavkatbegmatov60@gmail.com

Keywords: Ferroresonance, Electrical Distribution Networks of 6-35 kV, Voltage Transformer, Generalised Dynamic Model, Non-Linear Inductance.

Abstract: Considering that distribution networks of 6-35 kV are the longest among electrical networks, one of the specialaspects of improving the reliability of power supply is the study of the effect of ferroresonance on the performance of voltage transformers (VT). Since the ferroresonant mode is quasi-stationary and occurs both at the fundamental frequency and at the subharmonic, the key role in the study is given to the creation of a dynamic model of the nonlinear inductance of the VT. The mathematical models and characteristics of the non-linear inductance of the VT and the resulting mathematical expressions proposed in well-known scientific papers are approximate and do not have sufficient accuracy for the analysis and qualitative assessment of ferroresonance in electrical networks of 6-35 kV. Since ferroresonance is characterized by non-linear abrupt modes of saturation of the VT magnetic circuit, the paper proposes a generalized dynamic model of the nonlinear VT inductance and more accurate analytical equations for the effective analysis of ferroresonance in 6-35 kV electrical networks.

1 INTRODUCTION

In any substation there are VTs containing ferromagnetic cores and mains capacitances, so under certain conditions ferroresonance can occur. For occurrence of ferroresonance processes two conditions must be present: currents in VTs must be sufficient for transition of magnetization curves into saturation area and input resistance of network connected to the winding must have capacitive character [1,2,3].

Considering that medium voltage distribution networks of 6-35 kV are the longest among high voltage networks, one particular aspect of improving the reliability of electrical networks is to investigate the effects of ferroresonance on the performance of VTs [4,5,6].

According to statistics over the last 30-40 years, in 6-35 kV networks, about 80% of the equipment damaged due to ferroresonance is VTs. Up to 10% of the installed power VTs are damaged annually in earth faults and ferroresonance. Practice has confirmed that the most frequent cause of damage is ferroresonance between the capacitance of the network and the inductance of the power VTs. Generally, ferroresonance leads to overvoltage's on the bus bars and inadmissible currents will flow through the high voltage winding of the VTs, causing them to be damaged and causing a power failure. The consequence of such an overvoltage is shown in Figure 1 [5,6].

Figure 1: Damage to the voltage transformer.

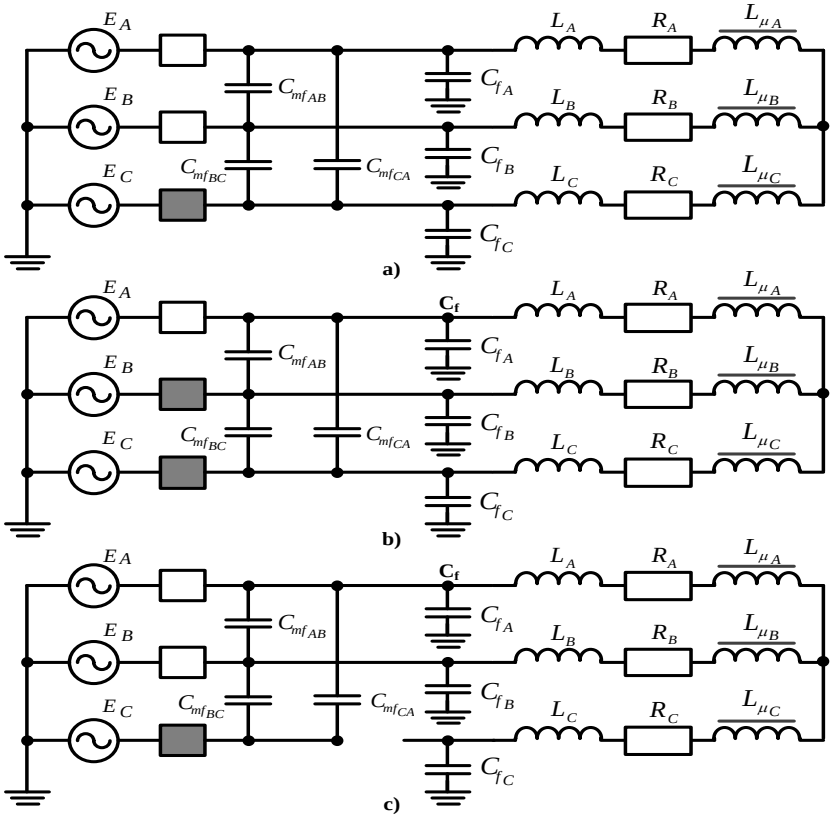


Figure 2: Diagrams of the possible occurrence of ferroresonance.

2 METHODS AND MATERIALS

The task of ferroresonance research is complicated by the fact that in complex three-phase high-voltage electrical networks, with limited experimental approaches due to the high cost of transformers, modeling and calculation of ferroresonance comes to the fore [6,7].

Despite the high level of development of existing mathematical models, which allow for the consideration of a variety of influencing factors, the main problems in the study of ferroresonance on VTs are the lack of:

- a reliable dynamic Weber-Ampere characteristic (WAC) of the VTs;
- accurate methods of determining this characteristic in VTs;
- determination of the range of possible changes in the nonlinear inductance (NI) of VTs;

A number of papers [8] have proposed investigations of ferroresonance in high voltage electrical networks by creating mathematical models of ferroresonance taking into account the influence of nonlinear parameters of VTs windings. However, the obtained mathematical expressions are complex and do not have sufficient accuracy to investigate ferroresonance in electrical networks.

Considering that ferroresonance is characterized by nonlinear jump-like modes of VTs magnetocarbon saturation, it is relevant to create:

- a generalised dynamic model of NI VTs;
- dynamic WAC of NI reflecting real dynamic hysteresis loop of VTs;
- more accurate analytical equations for determining equivalent NI VTs parameters.

In three-phase electrical networks, three most frequent cases of voltage ferroresonance are practically possible: single-phase connection of a line section with an idle transformer with insulated neutral point, Figure 2 a); two-phase connection of the same line, Figure 2 b); break of one phase with fall of a broken wire to ground from power supply side, Figure 2 c) [7,8].

Here, the inductance L takes into account the inductances of the mains and the line; the capacities C_f and C_{mf} correspond to the capacities of the network phases relative to ground and between phases; the active resistances R *is* entered to account for all types of active losses i.e. losses in the ground, in the line

conductors, in the transformer steel; the inductance *L* corresponds to the magnetization of the transformer.

The generalized differential (1) of the phase voltages for the circuits (Figure 2) of possible ferroresonance are as follows:

$$\begin{aligned} U_0 + U_A &= i_A R_A + L_{\mu A}\frac{di_A}{dt} + w_A\frac{dF_A}{dt} + i_{A0}R_A \\ &+ L_{\mu A}\frac{di_{A0}}{dt} + w_A\frac{dF_{A0}}{dt} + \frac{1}{C}\int i_N dt; \\ U_0 + U_B &= i_B R_B + L_{\mu B}\frac{di_B}{dt} + w_B\frac{dF_B}{dt} + i_{B0}R_B \\ &+ L_{\mu B}\frac{di_{B0}}{dt} + w_B\frac{dF_{B0}}{dt} + \frac{1}{C}\int i_N dt; \\ U_0 + U_C &= i_C R_C + L_{\mu C}\frac{di_C}{dt} + w_C\frac{dF_C}{dt} + i_{C0}R_C \\ &+ L_{\mu C}\frac{di_{C0}}{dt} + w_C\frac{dF_{C0}}{dt} + \frac{1}{C}\int i_N dt; \end{aligned} \quad (1)$$

According to the method of symmetrical components, write down the currents on phases A, B, C and on the neutral conductor as follows;

$$\begin{aligned} i_a + i_{A0} = i_A;\ \ i_в + i_{B0} = i_B; \\ i_c + i_{C0} = i_C;\quad i_A + i_B + i_C = i_N; \end{aligned} \quad (2)$$

Equations according to Kirchhoff's second law for the magnetic circuit of the phases

$$H_A l_A = w_A i_A;\quad H_B l_B = w_B i_B;\quad H_C l_C = w_C i_C. \quad (3)$$

By paralleling two phases which are in the same conditions with respect to the point of asymmetry (phases B and C in Figure 2 b; phases B and A in Figure 2 a), c) all three circuits can be reduced to a simplified form (Figure 3). Here, *Re*, *Le*, and *Ce* (C_1 and C_2) are the equivalent line and VTs parameters, respectively.

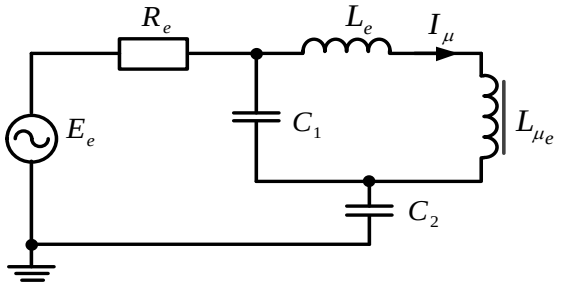


Figure 3: Simplified ferroresonance substitution diagram.

3 RESEARCH RELEVANCE

In [9-12] the NI substitution diagram (Figure 4) in general form and (4) which is its generalized dynamic model were developed

$$i = F_2\left(\Psi,\ \Psi^n;\ \frac{d\Psi}{dt},\ \frac{d^2\Psi}{dt^2}\ldots\right). \quad (4)$$

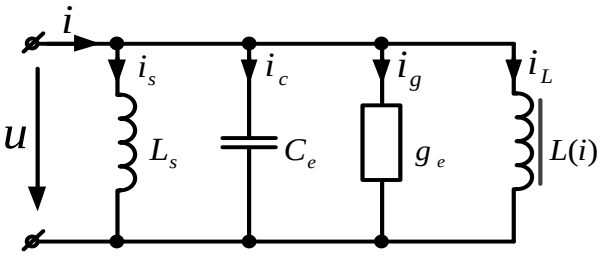


Figure 4: Diagram of the generalized dynamic NI model substitution.

In the schematic (Figure 4) L_s - leakage inductance, C_e - equivalent electromagnetic capacitance, $g_e=1\ /\ R_e$ - equivalent active conductance, are equivalent parameters of NI and if we take them as constant, we obtain in view (5).

$$i = C_e\frac{d^2\Psi}{dt^2} + g_e\frac{d\Psi}{dt} + a\Psi + b\Psi^n + \frac{\Psi}{L_S} \quad (5)$$

where $i_L = a\Psi + b\Psi^n$ is approximation of WAC of NI, obtained on the basis of magnetization curve $B = f(H)$.

The dynamic hysteresis loop approximation assumes that the dependence of the core demagnetization rate on the dynamic strength is linear, i.e. the faster the core demagnetization, the wider the dynamic hysteresis loop. The generalized equation that describes the magnetization and demagnetization processes of a NI core is as follows (6):

$$\frac{dB}{d\tau} = \frac{\mu_e}{\pi}(H \pm H_C) \quad (6)$$

It follows from (5) that for a particular conversion frequency, the dependence of the rate of change in the induction of NI on the dynamic field strength is linear. If we take the mains voltage $u = U_m cos\omega t$ and taking into account the accepted approximation of WAC of NI in the form of $i_L = a\Psi + b\Psi^n$ we obtain:

$$i = i_C + i_g + i_L = \left(a - \frac{I_{cm}}{\Psi_m}\right)\Psi + b\Psi^n \pm \frac{I_{gm}}{\Psi_m}\sqrt{\Psi_m^2 - \Psi^2} \quad (7)$$

Based on (7), the dynamic WAC of the generalised dynamic NI model is constructed, which describes its dynamic hysteresis loop (Figure 5) [13,14,15].

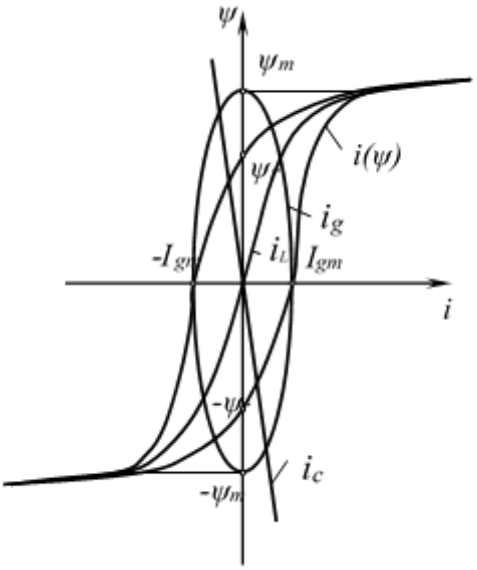


Figure 5: Dynamic characteristics of NI.

The equivalent parameters g_e, C_e, L_S of the dynamic NI model (Figure 4) are respectively determined from the following:

$$g_e = \frac{l\left(H_c + 0{,}125\omega\sigma d^2 B_S\sqrt{2\varepsilon - 1}\right)}{\omega w^2 SB}, \quad (8)$$

$$C_e = \frac{a\Psi_r + b\Psi_r^n - \frac{\omega}{Re}\sqrt{\Psi_m^2 - \Psi_r^2}}{\omega^2\Psi_r}, \quad (9)$$

$$L_S = \frac{\Psi_r}{\frac{1}{\Psi_m}\left(I_{cm}\Psi_r + I_{gm}\sqrt{\Psi_m^2 - \Psi_r^2}\right) - b\Psi_r^n}. \quad (10)$$

It should be noted that (8), (9) and (10) can be successfully applied to determine equivalent parameters R_e, C_e (C_1 and C_2), L_e and in a simplified ferroresonance substitution scheme (Figure 3) since they are equivalent parameters of NI VTs [16-19].

4 CONCLUSION

Therefore, the following conclusions are drawn from the generalized dynamic NI model and the analytical expressions for determining its equivalent parameters:

1) The generalized dynamic NI model describes analytically the non-linear WAC of NI VTs instantaneous basis and allows their equivalent parameters to be determined with high accuracy compared to known models.
2) Analytical expressions (8), (9) and (10) defining equivalent parameters Ge, Ce and Ls NI are simple and precise enough in comparison with known calculation equations. These equations adequately determine the equivalent parameters Re, Ce (C1 and C2), Le of NI VTs.
3) The application of a generalized dynamic model and analytical equations of equivalent NI parameters is an effective way to analyze and calculate ferroresonance in 6-35 kV electrical networks.

REFERENCES

[1] K. S. Demirchyan, L. R. Neuman, and et al., “Theoretical foundations of electrical engineering,” St. Peter, 2006, vol. 1, 450 p., ch. 28.

[2] K. S. Demirchyan, L. R. Neiman, and et al., “Theoretical foundations of electrical engineering,” St. Peter, 2006, vol. 2, 575 p.

[3] E. B. Saitov, Sh. Kodirov, Z. F. Beknazarova, A. Nortojiyev, and N. Siddikov, “Developing Renewable Sources of Energy in Uzbekistan Renewable Energy Short Overview: Programs and Prospects,” AIP Conference Proceedings, vol. 2432, 2022, pp. 020015.

[4] E. B. Saitov and T. B. Sodiqov, “Modeling an Autonomous Photovoltaic System in the Matlab Simulink Software Environment,” AIP Conference Proceedings, vol. 2432, 2022, pp. 020022.

[5] E. B. Saitov, Sh. Kodirov, B. M. Kamanov, N. Imomkulov, and I. Kudenov, “Increasing the Efficiency of Autonomous Solar Photovoltaic Installations for Power Supply of Agricultural Consumers,” AIP Conference Proceedings, vol. 2432, 2022, pp. 040036.

[6] J. Bird, “Electrical and Electronic Principles and Technology,” London and New York, 2014, 455 p.

[7] K. Charles, A. Matthew, and N. O. Sadiku, “Fundamentals of Electric Circuits,” New York, 2014, 458 p.

[8] N. A. Antonov, “Analysis of ferroresonant circuits of 110-500 kV electrical networks by the methods of the mathematical modeling,” Dissertation, Ivanovo, 1988, 200 p.

[9] L. A. Vergara Valdes, “Development of a technique for detecting and compensating nonlinear dynamic processes in the medium voltage networks of the electrical complexes,” Dissertation, Ph.D., Moscow, 2016, 161 p.

[10] A. E. A. Araujo, A. C. Soudack, and J. R. Marti, “Ferroresonance in Power Systems: Chaotic Behaviour”, Proceedings-C, vol. 140, no. 3, 1993, pp. 237-240.

[11] B. A. Abdullaev, A. A. Alimov, and D. A. Khalmanov, “To the problem of the calculation capacity of the nonlinear inductance,” in Seventh World Conference

on Intelligent Systems for Industrial Autamation, Tashkent, 2012, pp. 112-115.

[12] B. Abdullaev, "Generalized models of non-linear passive elements of electrical circuits and systems," Tashkent State Technical University, 2015, 180 p.

[13] A. A. Alimov, D. A. Nosirova, F. A. Akbarov, and Kh. A. Muminov, "To the problem of the calculation capacity of the nonlinear inductance," Journal of Critical Reviews, vol. 7, no. 15, 2020, doi: 10.31838/jcr.07.15.232.

[14] Sh. E. Begmatov, "An innovative way to create secondary power sources with high energy performance," Technical Science and Innovation, vol. 2, pp. 161-166, 2020.

[15] Sh. E. Begmatov, Kh. I. Khalbutaeva, and S. A. Dusmukhamedova, "Study of ferroresonance in electric networks taking into account the generalized model of nonlinear inductance," in Science and Life: International Scientific Conference, Sofia, Bulgaria, 20 October 2020, pp. 41-43.

[16] Sh. E. Begmatov, S. A. Dusmukhamedova, and Kh. Holbutaeva, "Study of ferroresonance using generalized models of passive nonlinear elements," in Rudenko International Conference: Methodological Problems in Reliability Study of Large Energy Systems (RSES 2020), 14 December 2020, doi: 10.1051/e3sconf/202021601115.

[17] B. A. Abdullaev, Sh. E. Begmatov, D. Khalmanov, and S. A. Dusmukhamedova, "Generalized mathematical models of nonlinear passive elements for the research of ferroresonance in electrical networks," Technical Science and Innovation, vol. 2, pp. 115-123, 2021.

[18] Sh. E. Begmatov, "Method for determining the boundary conditions of ferroresonant over voltages in electrical networks," in VII International Scientific and Practical Conference: Energy and Energy Saving - Theory and Practice, Kemerovo, Russia, 7-9 December 2022, in press.

[19] Shavkat Begmatov, Dilshod Khalmanov, and Saidakhon Dusmukhamedova, "Analysis of Ferroresonance in 6-35 kV Electric Networks Including Dynamic Model of Non-Linear Inductivity of Power Transformers and Reactors," Journal of Electrical Engineering & Technology, vol. 17, no. 2, pp. 772-782, 2022, doi: 10.1007/s42835-021-00777-1.

Application of Cycle-Flow Technology in Coal Mines

Javokhir Toshov[1], Buri Toshov[2], Umid Bainazov[1] and Muxriddin Elemonov[1]
[1]*Tashkent State Technical University named after Islam Karimov, University Str. 2, Tashkent, Uzbekistan*
[2]*Navoi State University of Mining and Technology, Tarobiy Str., Navoi, Uzbekistan*
javokhir.toshov@yandex.ru, elyor.saitov@ yandex.ru, bobomuratovsardor1@gmail.com, sirojiddin6870@gmail.com

Keywords: Open Mining, Rock, Quarry, Cyclical-Flow Technology (SFT), Unloading and Loading Hopper, Ansys Workbench, Conveyor, Belt.

Abstract: This article discusses, describes and analyzes the main problems of open mining in overburden work with the transportation of rock mass at the Angren coal open pit. The analysis of the main problems in the design of cyclical-flow technology with belt conveyors for hard rocks at the unloading and loading section of the bunker has been carried out. Calculations were carried out using the Ansys Workbench program. This article discusses, describes and analyzes the main problems of open mining in overburden work with the transportation of rock mass at the Angren coal open pit. In order to study them, a review was made of the stages of using conveyors for the mining industry both in different countries and at the Angren coal mine. In addition, as a result of the analysis, the author found that the main problem in the design of cyclic-flow technology with belt conveyors for hard rocks is due to the operation of the design of the unloading and loading hopper. In order to solve the problem identified by the author, optimization of the "CPT" with belt conveyors is proposed based on the improvement of the design of the bunker using the "Ansys Workbench» program, the prospects for using this calculation and analytical program are emphasized.

1 INTRODUCTION

In view of the recent political and economic events in the world, the problems of providing countries with energy resources have acquired particular urgency. Of course, these problems are relevant for our state. The fuel and energy complex of Uzbekistan occupies a special place in the economy of the republic, is a life support system for the population, and contributes to the political and economic independence of the state [1, 2]. A special place in solving these problems is occupied by the mining industry of Uzbekistan, which today is characterized by the further development of the open pit mining method. Open-pit mining has already acquired a predominant role in the extraction of ores of ferrous and non-ferrous metals, mining and chemical raw materials and building materials. Open pit mining is also being further developed in the coal industry, in particular, at the Angren coal open pit. To substantiate the origins of the problem considered in this paper, we briefly characterize this coal mine [3, 4]. The experience of recent years, as well as the achievements of modern mining technology, dictate new conditions for the organization of technology and the procedure for mining the Angren coal deposit. The existing railway transport within the section limits the capacity of internal dumps, requires large material costs for the operation of transport, does not provide a modern change in the configuration of the side, which reduces its stability, and the presence of a large length of railway tracks requires significant labor and material resources. So, for example, according to the results of 2000 and 2001, railway transport accounts for 36% of the cost of 1 ton of coal. In addition, the irregular work of railway transport reduced the productivity of excavators by 1.5-2 times. The use of railway transport requires the presence of railway tracks on each ledge, which, with a length of one ledge of 6 km, requires the presence of 126 km of tracks inside the section, not counting the railway stations. With the maximum achieved capacity of 5.6 million tons/year in 1992, the Angrensky open pit had 540 km of railway lines.As of January 1, 2002 y. 350 km remained in stock, while the materials for the maintenance of the tracks are forced to buy outside the Republic of Uzbekistan, which requires constant foreign exchange. One of the options for getting out of this situation was the rebuilding of double and triple ledges, however, this experience turned out to be deplorable. The Angrensky section received a lot of local landslides precisely in places where the height of the ledges increased [5].

2 METHODS AND MATERIALS

Structurally, the area of the Angren section is confined to the northeastern flank of the main Angren syncline, which plunges in a southwestern direction and is composed of additional small folds, accompanied in some areas by disjunctive disturbances. The displacement amplitude of the coal deposit is 10-40 m. The angles of dip of the layers vary from 5o to 18o. "Upper" and "Powerful". The "upper" complex is represented by frequent interbedding of coal packs with a thickness of 0.2 to 2.5 m, separated by rock interlayers. The total thickness of the complex ranges from 15.0 to 30.0 m. The average thickness of the complex is 21.65 m, the coal mass accounts for 12.45 m. The "powerful" complex occupies part of the coal deposit [6, 7]. In the "Powerful" complex, the coal mass occupies 89-95% and only 5-15% is occupied by rock. Rock layers are usually thin 0.10-0.15 m. In the central part there is a zone of impoverishment of the "Powerful" complex. In this zone the content of coal packs is reduced to 25-30%.

Working with the world's leading companies producing mining equipment and studying their positive and negative experience of introducing new equipment at various coal mines in Germany, Kazakhstan, Russia, India, France, etc., the company's specialists came to the conclusion that the most appropriate technology for production mining operations at the Angren open pit, is a cycle-flow and in-line technology. The chain of cyclic-flow technology involves: an EKG-type excavator + a mobile crushing plant + a self-propelled loader + a conveyor complex + a spreader. One local conveyor takes rock from 3 ledges along three independent chains, consisting of an excavator + mobile crushing plant + loader.

To ensure a stable and uninterrupted supply of objects of economic sectors and the population of the republic with solid fuel, Uzbekugol JSC carried out active preparatory work to attract investment in the coal industry. The set of planned measures made it possible to increase the share of coal in the country's fuel and energy balance to 15%. At present, the explored coal reserves in the republic amount to 1.9 billion tons, and the predicted resources exceed 5.7 billion tons [8, 9].

Modernization of the process of open pit coal mining has shown that the belt conveyor plays a particularly indispensable role in modern production. In this regard, 23 conveyor belts were designed and manufactured by the China North Heavy Industry Corporation at the Angren mine. However, the experience of using the cyclic-flow technology (SFT) at the Angren coal open pit made it possible to identify problem areas in the process of transporting overburden rocks using a belt conveyor. We came to these conclusions as a result of studying the mechanism of overburden transportation [10, 11].

One of the most important problems is the process of reloading the rock from the bunker to the main belt conveyor, which causes frequent breaks in the main conveyor belt due to pieces of overburden falling from a considerable height. The problem is aggravated by the fact that the conveyor lines move at a significant speed: the speed of the face conveyor belt is 5.0 m/s, the main one is 5.6 m/s.Since the use of CCT reduces the distance of transportation of rock mass by 50-60%, reduces the use of electricity by 5-6 times and reduces the number of employees by almost 6 times.

The study of the operation of the belt conveyor gives us reason to assume that one of the main problems is the bins that receive the load and transfer it to another conveyor line. This problem is already chronic, and interruptions in the operation of conveyors due to a stop for repairs directly affect productivity, which is unfavorably reflected in the operation of the transportation complex as a whole. I studied in detail the scientific research and experience of the specialists of the Angren Division on the problem under consideration, so that international experience in solving this pressing issue, we identified possible factors:

1) In all countries of Europe and the CIS, the use of conveyor transport, in comparison with other modes of transport in open-pit coal mining, allows a number of positive factors (economy, high productivity, ensuring the continuity of cargo transportation, the possibility of complex automation of production, labor safety);
2) Problems with groundwater, which increase the moisture content of rocks as a result of soil sticking, which significantly makes the extracted rock heavier, that is, hydrogeological conditions lead to difficulties in the operation of conveyor transport in coal mines;
3) At present, there are conveyors in foreign practice, the productivity of which reaches 25,000 m3 / hour. The belt speed reaches 10 m/s, the installed motor power is up to 10,000 kW.

The productivity of the CCT at the Angren coal mine reaches 4,000 m3/hour, the speed of the face conveyor belt is 5.0 m/s, the main conveyor belt is 5.6 m/s.Practice shows that it is necessary to carry out work on further reconstruction of the SFT. Based on the results of studies on the efficiency of conveyor

transport at the Angren coal mine, the following shortcomings in the operation of the CLT mechanism were obtained:

- the height of the fall of the product must be minimal. This can be achieved by using a curved chute in which the bulk material moves slowly downward, causing some of the dynamic load of the falling product stream to be absorbed before it reaches the belt;
- a belt with deep grooves has less lateral spillage of the product and holding the load in the center. When transporting fine-grained material, aprons are installed around the loading tray, which prevent it from falling onto the idle branch of the belt;
- bulk material should be loaded at the lowest possible conveyor inclination of no more than 10°, it is preferable to provide a horizontal section for loading;
- the conveyor belt must have additional supports under the loading area, which absorb the inertial force of the falling cargo flow. To do this, roller bearings with a rubber coating or damping devices that absorb shock are used [12].

3 RESULTS AND DISCUSSION

A detailed study of the identified factors at the Angren coal mine, an analysis of the operation of the SFT, made it possible to reveal the following: one of the main reasons for stopping the main conveyor as a result of its rush lies in the features of the mechanism of operation of the unloading and loading hopper of the SFT [13].

For a visual representation of the design of the unloading and loading hopper of the SFT and during its operation in production, this study proposes the use of the Ansys program. In our opinion, this software package allows to carry out calculations of metal structures using the finite element method. The Ansys program also carries out design optimization, which is a computer technology, consisting in the selection of the optimal project from several using finite element analysis [14]. The designer selects the criteria and constraints of the project and creates the same parametric model as in parametric design. As a result of the analysis of the process of operation of the SFT, it was revealed that one of the main reasons for stopping the main conveyor as a result of its rush lies in the features of the mechanism of operation of the unloading and loading hopper of the SFT.

The optimization procedure controls the execution of the analysis based on the decision on the values of the parameters used in the trial calculations. The results of our analyzes based on the Ansys program will be presented in our next papers [15].

Based on this, we have studied in detail the process of operation of the unloading and loading hopper of the SFT using the Ansys program, which is shown in Figure 1.

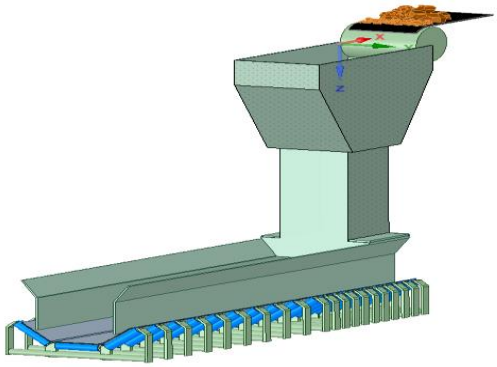


Figure 1: General view of the model for calculation.

The ANSYS software package allows you to calculate metal structures using the finite element method. The ANSYS program also performs design optimization, which is a computer technology that consists in choosing the optimal design from several using finite element analysis. The designer selects the criteria and constraints of the project and creates the same parametric model as in parametric design. The optimization procedure controls the execution of the analysis based on the decision on the values of the parameters used in the trial calculations.

The Ansys program was used as the main method of analysis due to the fact that this program made it possible to carry out the most accurate mathematical calculations of the geometric parameters of this design. The Ansys program, along with other Solidworks programs, Nx and Compas, is successfully used in solving engineering problems. With the help of the AnsysWorkbench program, based on precise geometric parameters, we designed 3D models of the unloading and loading hopper of the SFT located between the conveyor lines. Analysis of the constructed 3D models of the unloading and loading hopper made it possible to determine the reasons for the breakage of the main SFT belt. So, in particular, with the help of graphical and mathematical calculations, it was established that such indicators as: the height of the location of the unloading and loading hopper of the SFT and directly the geometric indicators of the structure itself play an important role.

To identify the most optimal height of the structure under study, its actual indicators of a height of 3.5 m were considered, taking into account the necessary data of materials and the dimensions of the bunker and the main belt. We have obtained graphical and mathematical calculations of the indicated parameters, which are given below (Figure 2).

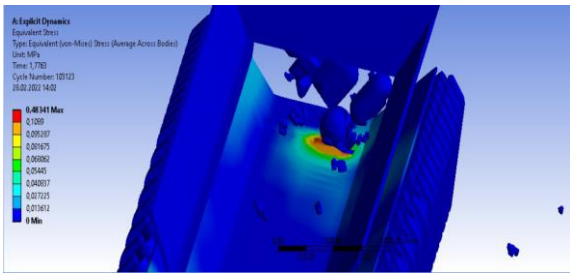


Figure 2: Graph of voltage of the bunker 3.5 m.

We have obtained equation (1) criteria for the rock to touch the tape, which must have *y_A=0* for the rock to hit the tape

$$\begin{cases} x = \left(x_0 + \frac{\vartheta_0 m}{k}\right) - \frac{\vartheta_0 m}{k} e^{\frac{k}{m}t} \\ y = h - \frac{g x_0^2}{2\vartheta_0^2} + \frac{m}{k}\left(\frac{mg}{\text{Á}} - \frac{g x_0}{\vartheta_0}\right) + \frac{m}{k}\left(\frac{g x_0}{\vartheta_0} - \frac{mg}{k}\right) e^{-\frac{k}{m}t} \end{cases} \quad (1)$$

On the graph, the maximum voltage value is indicated in green, the average voltage value is blue, and the minimum value is red. According to this graph, the maximum voltage is reached at 2 seconds (Table 1, Figure 3).

Table 1: Numerical value of rock stress.

Time (s)	Min. (Mpa)	Max. (Mpa)	Average(Mpa)
1,1755e-038			
0,12502	8,2027e-005	2,3605e-004	1,4049e-004
0,25002	8,3377e-005	2,2662e-004	1,8545e-004
0,37503	0	1,6845e-004	1,1457e-004
0,5		1,0315e-004	1,0571e-004
0,62501	8,7678e-005	2,2694e-004	1,7351e-004
0,75001	8,6954e-005	2,2516e-004	1,5231e-004
0,87502	0	1,6332e-004	8,9977e-005
1	2,2611e-005	1,6885e-003	1,3835e-004
1,125	2,9019e-005	1,5561e-003	2,4364e-004
1,25	6,0799e-006	1,7772e-003	1,7693e-004
1,375	1,7992e-006	5,9421e-003	2,4124e-004
1,5	6,0799e-006	3,3168e-003	2,7127e-004
1,625	3,0844e-005	0,45364	1,0513e-002
1,75	2,1866e-004	0,39978	1,1952e-002
1,875	4,3258e-004	0,42374	1,385e-002
2	1,3518e-004	0,48341	1,0494e-002
2,125	1,0315e-004	0,37993	1,0449e-002
2,25	8,2195e-005	0,44386	8,3001e-003

The graph shows the distance traveled by rocks from the bunker to the belt per unit time. It was shown that from the 3.5 m bunker to the belt, the maximum impact rock covered a distance of 5269,5 mm (Table 2, Figure 4).

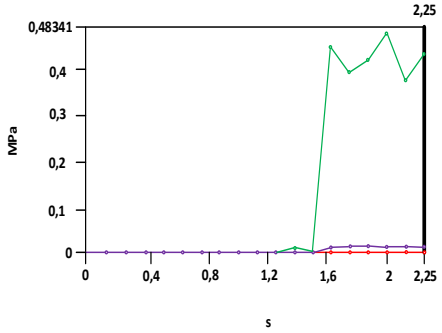


Figure 3: Graph of stress and displacement of rock on the belt from a bunker 3.5 m high.

Table 2: Numerical value of rock movement in the bunker.

Time (s)	Min.(Mpa)	Max. (Mpa)	Average(Mpa)
1,1755e-038			
0,12502	548,32	617,9	583,75
0,25002	1082,3	1235,4	1158,8
0,37503	1611,6	1254,9	1735,7
0,5	2158,7	2492,1	2330,9
0,62501	2753,2	3174,3	2971,9
0,75001	3422,7	3927,8	3684,4
0,87502	3392,6	4393,7	3788,4
1	2790,4	4699,6	3518,1
1,125	2746,1	4719,5	3551,1
1,25	3181,3	4902,2	3847,6
1,375	3573,3	5097,8	4260,8
1,5	4431,8	5170,4	4775,9
1,625	4456,1	5259,8	4840,9
1,75	4447,3	5269,5	4844
1,875	4440,3	5260,6	4832,3
2	4425,5	5253,5	4821,1
2,125	4388	5236	4797,7
2,25	4392,8	5238,1	4798,1

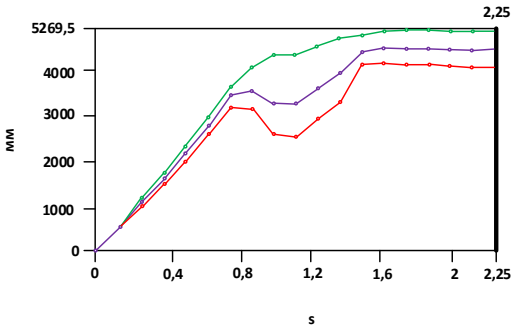


Figure 4: Graph of the movement of rock from the bunker to the conveyor belt.

This graph depicts the thrust and tension of the rock falling from the hopper onto the belt per unit time. The graph shows that the rock that had the maximum impact on the belt from the 3.5-meter bunker covered a distance of 5238 mm. (Table 3, Figure 5).

Table 3: Stresses and displacements of the hopper height.

Time (s)	(B)Equivalent Stress (Max) (MPa)	Total Deformation (Max) (mm)
1,1755E-38	0	0
0,12502	2,3605e-004	617,9
0,25002	2,2662e-004	1235,4
0,37503	1,6845e-004	1854,9
0,5	1,6863e-004	2492,1
0,62501	2,2694e-004	3174,3
0,75001	2,2516e-004	3927,8
0,87502	1,6332e-004	4393,7
1	1,6885e-003	4699,6
1,125	1,5561e-003	4719,5
1,25	1,7772e-003	4902,2
1,375	5,9421e-003	5097,8
1,5	3,3168e-003	5170,4
1,625	0,45364	5259,8
1,75	0,39978	5269,5
1,875	0,42374	5260,6
2	0,48341	5253,5
2,125	0,37993	5236
2,25	0,44386	5238,1

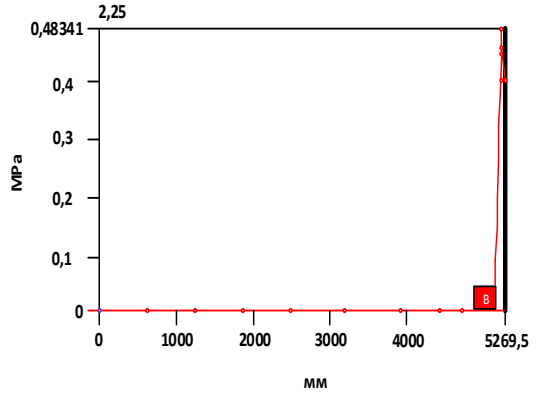


Figure 5: Graph voltage and displacement of the hopper 3.5 m.

4 CONCLUSION

We have studied and analyzed in detail the problems of the CLT operation process during overburden operations at the Angren coal mine. As a result of the analysis, it was revealed that the main problems of overburden transportation are due to various reasons for the rupture of the main line 3.8. In this regard, a scientific study of this problem was carried out, which made it possible to identify one of the main problems of frequent ruptures of the main belt conveyor during operation, due to both the structural and technical features of the main belt itself, and the impact force of the mined rock from the loading and unloading bunker of the SFT. In order to identify the causes of the identified problems, an analysis was made of the operation of the SFT, taking into account its technical parameters (belt speed, degree of hardness and angle of fall of the rock onto the tape, as well as the height of the bunker, from which the rock is unloaded onto a tape 3.5 m high, according to the Ansys program. The analysis showed that one of the main problems is due to the height of the production bunker. Determining the optimal height of the operational hopper allows you to avoid frequent breaks in the main belt conveyor No. 3.8. An important role is played by the peculiarity of the design of the loading and unloading bunker itself.

Our analysis showed that from a bunker height of 3.5 meters, the impact stress of the falling rock on the main line is 0.48341 MPa.

The results of the analysis show that the determination of the optimal height of the loading and unloading hopper makes it possible to extend its service life and reduce the frequency of repair work of this design

REFERENCES

[1] W. Gang, X. Jun, X. Sheng, and W. Haiyang, “Mining a coal seam below a heating goaf with a force auxiliary ventilation system at Longhua underground coal mine, China,” International Journal of Mining Science and Technology, vol. 25, pp. 67-72, Jan. 2015.

[2] A.I. Leleko, “Coal industry of Uzbekistan: stages of development,” Mining Bulletin of Uzbekistan, no. 1, pp. 3-7, 1998.

[3] G. Bulatov and T. Annakulov, “Investigation of the width of the entry of an excavator when loading a mobile crushing plant in the conditions of the Angren coal mine of Uzbekistan,” IOP Conference Series: Earth and Environmental Science, vol. 937, no. 4, article 042088, 2021.

[4] T.J. Annakulov, S.B. Gaibnazarov, O.A. Kuvandikov, F.O. Otajonov, and B.O.Otajonov, “Development of a Methodology for Determining the Economic Efficiency of Cyclic-Flow Technology Schemes for Rock Mining Using Mobile Crushing and Reloading-Conveyor Complexes,” AIP Conference Proceedings, vol. 2432, article 030115, 2022.

[5] M.U. Musirov and N.Yu. Kholmanov, “Investigation of String Vibrations of a Transporting Device,” in Proceedings of the 7th International Conference on Industrial Engineering (ICIE 2021), pp. 126-133, Jan. 2022, doi: 10.1007/978-3-030-85233-7_15.

[6] N.S. Usmanov, I.V. Tsoi, U.U. Irkabaev, I.Kh. Nasriddinov, and T.Zh. Annakulov, "Experience in the implementation of cyclic-flow technology at the overburden complex of the Angrensky open-pit mine," Mining Bulletin of Uzbekistan, no. 1 (60), pp. 82-86, 2015.

[7] B.N. Baratov, F.Y. Umarov, and Z.H. Toshov, "Tricone drill bit performance evaluation," Gornyi Zhurnal, no. 12, pp. 60-63, 2021, doi: 10.17580/gzh.2021.12.11.

[8] Z. Bubnicki, "Modern Control Theory," Springer, 2005, [Online]. Available: https://doi.org/10.1007/3-540-28087-1.

[9] J.B. Toshov, B.R. Toshov, B.N. Baratov, and A.L. Haqberdiyev, "Designing new generation drill bits with optimal axial eccentricity," Mining Informational and Analytical Bulletin, no. 9, pp. 133–142, 2022, doi: 10.25018/0236_1493_2022_9_0_133.

[10] E.E. Sheshko and L.N. Atakulov, "Justification of radius for transition loading section of sandwich belt high-angle conveyor," Gornyi Zhurnal, no. 12, pp. 61-64, 2019.

[11] G. Akanova, U. Kayumov, L. Sagatova, M. Istamov, and L. Atakulov, "Choosing the flow part geometric shape of the dredge pumps for viscous fluids," Mining of Mineral Deposits, vol. 15, no. 4, pp. 75-83, 2021.

[12] K. Forrest and L. Loate, "Power and Accumulation Coal Mining, Water and Regulatory Failure," The Extractive Industries and Society, vol. 5, no. 1, pp. 154-164, Jan. 2018, doi: 10.1016/j.exis.2017.12.007.

[13] V.I. Efimov, N.I. Abramkin, V.V. Vernigor, and B.Kh. Khakimov, "Underground mining equipment kits coal in shallow beds," Sustainable Development of Mountain Territories, vol. 12, no. 4 (46), pp. 510-516, 2020, doi: 10.21177/1998-4502-2020-12-4-510-515.

[14] E.E. Sheshko and L.N. Atakulov, "Elasticity modulus steel-cable rubber conveyor belts," Gornyi Zhurnal, no. 4, pp. 73-76, 2019.

[15] V.A. Bruyaka, V.G. Fokin, E.A. Soldusova, N.A. Glazunova, and I.E. Adeyanov, "Bruyaka Engineering Analysis in ANSYS Workbench: Proc. Allowance," Samara, 2010, 271 p.

Analysis of Auto Parametric Oscillations at the Subharmonic Frequency in Two-Phase Ferro Resonance Circuits

Mukhtorhon Ibadullayev
Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan
mukhtorhon@gmail.com

Keywords: Ferroresonance, Self-Oscillations, Subharmonics, Approximation, Lower Harmonic, Small Parameters, Ferromagnetic Element, Self-Magnetization, Power Balance.

Abstract: Oscillatory processes in nonlinear multiphase electro ferromagnetic circuits (EFMC) have an exceptional variety and complexity, and therefore their research is also associated with solving complex physical problems. Theoretical analysis, in order to identify the main patterns of excitation of subharmonic oscillations of various orders, and the development of engineering calculation methods are of particular importance in designing and creating various switching-type converter devices. On the other hand, two-phase self-oscillating circuits are physical models of power transmission lines (power lines). Consequently, the study of the excitation of the existence of subharmonic oscillations in two-phase systems allows us to establish some patterns of overvoltage with power lines with capacitive compensation caused by harmonic oscillations, and, if possible, take measures to prevent these abnormal modes or mitigate their negative consequences. This article discusses the process of excitation of subharmonic oscillations ω/2 in two-phase EFMC. Using the method of energy relations, the areas of existence and critical values of the circuit parameters are determined

1 INTRODUCTION

It is known that self-oscillations can occur in circuits with electro-ferromagnetic oscillatory circuits. These fluctuations are caused by periodic changes in the nonlinear inductance. Since the change in the nonlinear inductance occurs under the influence of the power supply, the circuit is called auto-parametric [1-7].

In general, auto-parametric oscillations (APO) can be excited in circuits with a nonlinear resistor, inductance, and capacitance due to the ability of circuits to accumulate a certain amount of energy in their elements, which is necessary to excite and maintain the APC at a particular frequency [8-15]. However, the most interesting from a practical point of view are auto-parametric (AP) circuits with an electro ferromagnetic oscillatory circuit capable of developing high power during energy conversion. Ferromagnetic elements in combination with linear capacitances can form circuits in which a significant amount of stored energy during the excitation of the APO can be effectively used to create devices and devices of converter technology.

Of particular interest in this regard are the issues of the development of the theory and methods of calculating auto-parametric (AP) energy converters, which could be the basis for engineering calculations of specific devices.

Of even greater practical interest is the creation of devices based on multiphase AP circuits with an electro ferromagnetic oscillatory circuit (in particular, two-phase and three-phase). Unlike existing energy converters, these devices are made mainly on the basis of single-phase nonlinear circuits, have phase-specific features, and are reliable and easy to operate.

2 METHODS AND MATERIALS

Consider the analysis of the two–phase frequency divider in Figure 1, which is two identical ferroresonance circuits consisting of counter-coupled nonlinear inductors of series-connected capacitors. The secondary windings of nonlinear inductors are connected in series and shunted by a

diode to create a constant magnetic field necessary to obtain even harmonics.

When the device is connected to the power source of the generating voltage U_1 and U_2 and at certain ratios between the input voltages and circuit parameters, auto-parametric oscillations are excited in them at the frequency of the second-order subharmonics. At the same time, the phases of excited oscillations are shifted by 180^0 (Figure 1 b), since the magnetic fluxes created by the currents of the primary and secondary windings of nonlinear inductances are added in one core, while in the other core they are subtracted due to the specified connection of the secondary windings.

Shunting of secondary windings by diode D provides self-magnetization of cores by rectified current.

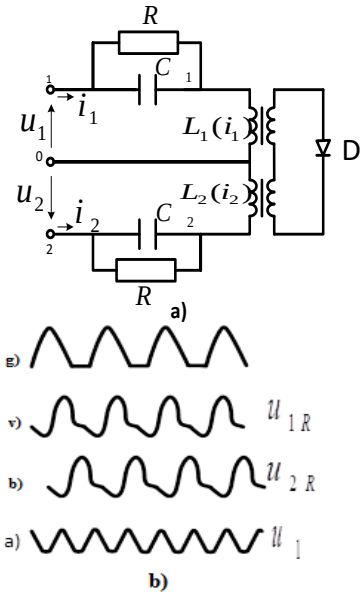


Figure 1: Two-phase frequency divider in two times: a) schematic diagram of the frequency divider; b) voltage waveforms in the tank.

Thus, the device allows for self-excitation of the APO at the second-order subharmonic frequency due to parametric changes in the inductance of a nonlinear reactive element.

3 RESULTS AND DISCUSSION

Subharmonic oscillations can be excited in the system, the amplitudes of which are equal but differ in phase by 180^0 (Figure 1 b) [16].

Depending on this, there are two modes of excitation of the SGC:

$$\Psi_{am}=\Psi_{bm} \qquad 0^0 \text{ and } 180^0$$

$$\Psi_{am}=\Psi_{bm} \qquad 180^0 \text{ and } 0^0$$

To analyze the steady-state regime of changes in magnetic fluxes in ferromagnetic elements (FE), we take as:

$$\Psi_a = \Psi_{0a}+\Psi_{1m}cos(\omega t+\varphi_1)+\Psi_{2m}cos(\omega t+\varphi_2)$$
$$\Psi_b = \Psi_{0b}+\Psi_{1m}cos(\omega t+\varphi_1-180^0)+ \Psi_{2m}cos(\omega t+\varphi_2-180^0)=\Psi_b-\Psi_{1m}cos(\omega t+\varphi_1)- \Psi_{2m}cos(\omega t+\varphi_2) \quad (1)$$

The magnetization characteristic of nonlinear elements is also approximated by an incomplete polynomial of the third degree:

$$i_a= a\Psi_a+b\Psi_a^3 , \qquad i_b=a\Psi_b+b\Psi_b^3 . \quad (2)$$

Substituting (1) into (2) and neglecting terms other than the frequencies ω and 2ω, we obtain expressions for currents (3):

$$i_a=A_{0a}+A_{1a}cos(\omega t+\varphi_1)+A_{2a}sin(\omega t+\varphi_1)+ B_{1a}cos(2\omega t+\varphi_2)+B_{2a}sin(2\omega t+\varphi_2) \quad (3)$$
$$i_b=A_{0b}+A_{1b}cos(\omega t+\varphi_1)+A_{2b}sin(\omega t+\varphi_1)+ B_{1b}cos(2\omega t+\varphi_2)+B_{2b}sin(2\omega t+\varphi_2) ,$$

where:

$$A_{0a}=a\Psi_0+b\Psi_0^3+\frac{3b}{2}\Psi_0\Psi_1^2+\frac{3b}{2}\Psi_0\Psi_1^2- - \frac{3b}{4}\Psi_1^2\Psi_2cos(2\varphi_1-\varphi_2)$$

$$A_{1a}=a\Psi_1+\frac{3b}{4}\Psi_1^3+3b\Psi_0^2\Psi_1- \frac{3b}{2}\Psi_1\Psi_2^2+3b\Psi_0\Psi_1\Psi_2cos(2\varphi_1-\varphi_2)$$

$$A_{2a}=3b\Psi_0\Psi_1\Psi_2sin(2\varphi_1-\varphi_2)$$

$$B_{1a}=a\Psi_2+\frac{3b}{4}\Psi_2^3+3b\Psi_0^2\Psi_2+\frac{3b}{2}\Psi_1^2\Psi_2+\frac{3b}{2}\Psi_0 \Psi_1^2cos(2\varphi_1-\varphi_2)$$

$$B_{2a}= -[\frac{3b}{2}\Psi_0\Psi_1^2sin(2\varphi_1-\varphi_2)]$$

$$A_{0b}=a\Psi_0+b\Psi_0^3+\frac{3b}{2}\Psi_0\Psi_1^2+\frac{3b}{2}\Psi_0\Psi_2^2- - \frac{3b}{4}\Psi_1^2\Psi_2cos(2\varphi_1-\varphi_2)$$

$$A_{1b}= - [a\Psi_1+\frac{3b}{4}\Psi_1^3+\frac{3b}{2}\Psi_0^2\Psi_1+\frac{3b}{2}\Psi_1\Psi_2^2- 3b\Psi_0\Psi_1\Psi_2cos(2\varphi_1-\varphi_2)]$$
$$A_{2b}=3b\Psi_0\Psi_1\Psi_2sin(2\varphi_1-\varphi_2)$$

$$B_{1b}= - [a\Psi_2+\frac{3B}{4}\Psi_2^3+3b\Psi_0^2\Psi_2+\frac{3B}{2}\Psi_1^2\Psi_2+\frac{3B}{2} \Psi_0\Psi_1^2cos(2\varphi_1-\varphi_2)]$$

$$B_{2b} = - [\frac{3b}{2}\Psi_0\Psi_1^2 sin(2\varphi_1-\varphi_2)].$$

The voltage on nonlinear inductors (4)

$$u_a = \frac{d\Psi a}{dt} = -\omega\Psi_{1m}sin(\omega t+\varphi_1)-2\omega\Psi_{2m}sin(2\omega t+\varphi_2)$$
$$u_b = \frac{d\Psi b}{dt} = \omega\Psi_{1m}sin(\omega t+\varphi_1) +2\omega\Psi_{2m}sin(2\omega+\varphi_2)\,. \quad (4)$$

Let 's express currents and voltage in a complex form (5):

$$\dot{I}_{a1m}=(jA_{1a}+A_{2a})\,e^{j\varphi_1}$$
$$\dot{I}_{a2}=(jB_{1a}+B_{2a})\,e^{j\varphi_2}$$
$$\dot{I}_{b1}=(jA_{1b}+A_{2b})\,e^{j\varphi_1}$$
$$\dot{I}_{b2}=(jB_{1b}+B_{2b})\,e^{j\varphi_2} \quad (5)$$
$$U_{a1}=-\omega\Psi_{1m}\,e^{j\varphi_1}$$
$$U_{a2}=-2\omega\Psi_{2m}e^{j\varphi_2}$$
$$U_{b1}=\omega\Psi_{1m}\,e^{j\varphi_1}$$
$$U_{b2}=2\omega\Psi_{2m}e^{j\varphi_2}.$$

Then the complexes of full capacities are expressed accordingly (6):

$$\hat{S}_{a1}=\frac{1}{2}\grave{U}_{a1}\cdot\dot{I}_{a1}$$
$$\hat{S}_{a2}=\frac{1}{2}\grave{U}_{a2}\cdot\dot{I}_{a2} \quad (6)$$
$$\hat{S}_{b1}=\frac{1}{2}\grave{U}_{b1}\cdot\dot{I}_{b1}$$
$$\hat{S}_{b2}=\frac{1}{2}\grave{U}_{b2}\cdot\dot{I}_{b2}$$

or power for subharmonics (7):

$$\hat{S}_{a1}=\frac{1}{2}(-\omega\Psi_{1m}e^{j\phi 1})\cdot(A_{1a}+jA_{1a})\,e^{j\phi 1}$$
$$\hat{S}_{b1}=(\omega\Psi_{1m}e^{j\phi 1})\cdot(A_{2b}+jA_{2b})\,e^{-j\phi 1} \quad (7)$$

or (8)

$$P_{a1}=-\frac{3b}{2}\omega\cdot\Psi_0\Psi_1^2\Psi_2 sin\,(2\varphi_1-\varphi_2)$$
$$Q_{a1}=\frac{\omega}{2}(a\Psi_1^2+\frac{3b}{4}\Psi_1^4+3b\Psi_0^2\Psi_1^2+$$
$$\frac{3b}{2}\Psi_1^2\Psi_2^2+3b\Psi_0\Psi_1^2\Psi_2 cos\,(2\varphi_1-\varphi_2)$$
$$P_{b1}=\frac{1}{2}\omega\Psi_{1m}\,3b\Psi_0\Psi_1^2\Psi_2 sin\,(2\varphi_1-\varphi_2) \quad (8)$$
$$Q_{b2}=-\frac{\omega}{2}[a\Psi_1^2+\frac{3b}{4}\Psi_1^4+3b\Psi_0^2\Psi_1^2+$$
$$\frac{3b}{2}\Psi_1^2\Psi_2^2-3b\Psi_0\Psi_1^2\Psi_2 cos\,(2\varphi_1-\varphi_2)]$$
$$P_{a2}=\frac{1}{2}\omega\Psi_{1m}^2\cdot 3b\omega\cdot\Psi_0\Psi_2 sin\,(2\varphi_1-\varphi_2)$$
$$P_{b2}=-\frac{\omega}{2}\Psi_{1m}^2 3b\Psi_0\Psi_2 sin(2\varphi_2-\varphi_1).$$

According to $\sum_{k=1}^{2}P_k(t)=0$ for two-phase EFMC, the necessary condition for converting the energy of the frequency 2ω into the energy of the frequency ω will be (9):

$$[1\,,\,2,\,3]$$
$$P_{a1}=-P_{a2}$$
$$P_{b1}=-P_{b2} \quad (9)$$

or

$$0°<2\varphi_{1a}-\varphi_{2a}<180°,$$
$$180°<2\varphi_{1b}-\varphi_{2b}<360°.$$

By entering the notation (10):

$$\alpha = b\Psi_0\Psi_2 sin\,(2\varphi_1-\varphi_2);$$
$$\beta = a+\frac{3b}{4}+3b\Psi_0^2+3b\Psi_2^2\,;$$
$$\gamma = 2\varphi_1-\varphi_2$$
$$K_{P1}=\frac{2P}{\omega\Psi_1^2}=-\alpha\cdot sin\gamma \quad (10)$$
$$K_{q1}=\frac{2Q}{\omega\Psi_1^2}=\beta+\alpha\cdot cos\,\gamma$$
$$K_{p2}=\alpha\cdot sin\,\gamma$$
$$K_{q2}=-\beta-\alpha\cdot cos\,\gamma.$$

Squaring and adding them, and using simple pre-formations, we obtain second-order equations with respect to the squares of the amplitudes of the magnetic fluxes $\Psi 1\text{m}^2$ and $\Psi 2\text{m}^2$, where $\Psi 1\text{m}^2$ is the magnetic flux of the subharmonics ($\omega/2$) and $\Psi 2\text{m}^2$ is the component of the magnetic flux of the fundamental harmonic (ω):

$$4\Psi_2^4+\Psi_1^4+16\Psi_0^4+4\Psi_2^2\Psi_1^2+$$
$$8\Psi_1^2\Psi_0-\frac{16}{3b}(K_q-a)\Psi_2^2-\frac{8}{3b}(K_q-a)\Psi_1^2-$$
$$\frac{32}{b}(K_q-a)\Psi_0^2+\frac{16}{9b^2}[K_P^2+(K_q-a)^2]=0, \quad (11)$$

where $K_P=\frac{\omega R}{z^2}$; $K_q=\frac{1}{\omega cz^2}$.

Since in the scheme under consideration Figure 1a subharmonic oscillations has phase shifts $\varphi = 180^0$ with symmetric amplitude, then further analysis of (11) is sufficient to study the process of excitation of SGC in one phase.

To determine the quantitative ratios of the subharmonic oscillations mode in (11), the square of the flow of the magnetizing effect, (Ψ_0^2) which is the control parameter of the frequency division mode, is considered to be set. In this case, (11) will take the form and will also represent a second-order curve:

$$4\Psi_{2m}^4+\Psi_{1m}^2+4\Psi_{2m}^2\Psi_{1m}^2-\frac{16}{3b}(k_q-a-3b\Psi_0^2)$$
$$\Psi_{2m}^2-\frac{8}{3b}(k_q-a)\,\Psi_{2m}^2+ \quad (12)$$
$$\frac{16}{9b^2}(k_q-a-3b\Psi_0^2)\,[(K')^2+1]=0$$

where: $K'=\frac{K_p}{Kq-a-3b\Psi_0^2}$.

Invariants that are equal to (13)

$$\delta=\begin{vmatrix}4 & 2\\ 2 & 1\end{vmatrix}\;S=4+1=5>0$$
$$\Delta=-\frac{64}{3b}\left[\left(\frac{Kp}{Kq-a-3b\Psi_0}\right)^2+1\right]\cdot\Psi_0^2 \quad (13)$$

Since Ψ_{2m}^2 and Ψ_{1m}^2 are the squares of the amplitudes of the main and subharmonic components of the flow couplings, for the existence of the subharmonic oscillations in the system, it is necessary that the curve describing the mode in question (Figure 2) be located on the first quadrant of the plane $\Psi_{2m}^2 0 \Psi_{1m}^2$.This is possible only under the condition that [19-21] (14).

$$\Psi_0^2 > 0, \qquad (14)$$

that is, with a positive magnitude of the magnetizing effect on the ferromagnetic element.

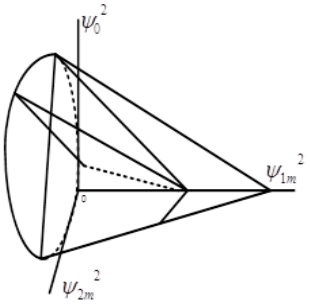


Figure 2: Graphical representation of (12).

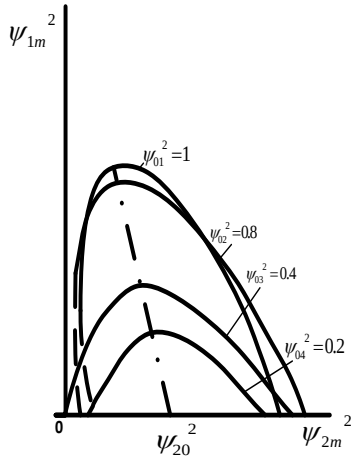


Figure 3: Graphical solution of (12).

If conditions (14) are met, (12) describes real parabolas in the plane $\Psi_{2m}^2 0 \Psi_{1m}^2$, the coordinates of the vertices of which are determined from the expressions (15)

$$\Psi_{20}^2 = \frac{4}{3b}(K_q - a); \quad \Psi_{10}^2 = \frac{2}{3b}(K_q - a). \qquad (15)$$

The angle of rotation of the axes of the parabola relative to the coordinate axes is equal to

$$\alpha = arctg\left(\frac{-4}{2}\right) \approx -63^o30^{\prime}.$$

The parameter of the parabola is determined from the expression (16)

$$P = \frac{-24(Kq-a)[-1-3b\Psi_0^2]}{3a \cdot b \cdot \sqrt{5}}. \qquad (16)$$

From where it can be seen that it is proportional to the magnetizing effect. The construction of the dependence according to (12) is shown in Figure 3.

For the existence of the SGC, it is necessary that the coordinates of the vertices are in the first quadrant of the plane $\Psi_{2m}^2 0 \Psi_{1m}^2$.

Therefore, the conditions for the existence of SGC are inequalities (17):

$$\Psi_{20}^2 > 0 \qquad \Psi_{10}^2 > 0\,. \qquad (17)$$

Since $\Psi_0^2 > 0$ and $\Psi_{20}^2 > 0$, then from (17) the necessary condition for the existence of the SGC will be $\Psi_{10}^2 > 0$. Taking into account (15) and (16), (12) acquires the canonical form of writing in the new coordinate system (Ψ_{2m}^2) and (Ψ_{1m}^2) (18)

$$(\Psi_{1m}^2)' = P(\Psi_{2m}^2). \qquad (18)$$

Since the second-order subharmonic oscillations are excited at positive values of the input effect from (11), it is possible to determine the region of existence by the input effect. Equating $\Psi 1m^2=0$ we have (19) and (20):

$$4\Psi_{2m}^4 - M\Psi_{2m}^2 + F = 0, \qquad (19)$$

here $$M = \frac{16}{3b}(Kq - a).$$

$$G = \frac{16}{9}\left(\frac{Kq-a-3b\Psi_0^2}{b}\right)^2 [\,(K^{\prime})^2 + 1)], \qquad (20)$$

then (21) $$\Psi_2^2 = \frac{-\sqrt{M^2-16F}}{4} \qquad (21)$$

or (22)

$$\Delta\Psi_2^2 = -\sqrt{\frac{256}{9b^2}(Kq-a)^2 - \frac{64}{9b^2}(Kq-a-3b\Psi_0^2)^2 (K)^2}\,. \qquad (22)$$

The critical value of the parameters of the soft excitation circuit of the subharmonic oscillations can be determined from the condition $\Delta\Psi_2^2 = 0$, while (23)

$$4(K_q-a)^2 - (K_q - a - 3b\Psi_0^2)^2 (K^{\prime})^2\,. \qquad (23)$$

It is important to determine in the region of the existence of a second-order subharmonic oscillations bounded by an ellipse according to (12), the region of soft and hard excitation, i.e. determining the limit of variation of the magnetizing effect leading to soft excitation of a soft oscillation.

The condition for soft excitation will be the positivity of the derivative $a(\Psi_{1m}^2) / a(\Psi_{2m}^2)$ at the intersection point of the left branch of the parabola with the axis of the abscissa. The intersection point

is determined from (12). To do this, we will solve these equations with respect to the input effect

$$(\Psi_{2m}^2) = \frac{16}{3b}(K_q - a) \pm \pm\sqrt{\frac{16}{3b}(Kq - a) - 16\left[\frac{16}{9b^2}\left(Kq - a - 3b\Psi_{02}\right)[(K^{|})2 + 1\right]} . \quad (24)$$

Solving the inequality $d\Psi_{1m}^2 / d\Psi_{2m}^2 \geq 0$ $(\Psi_{2m}^2)1$ at the point 0 determined from (24), we obtain an expression for the magnetizing effect in the plane $K_p 0\ \Psi_0^2$, in which a soft excitation mode is possible:

$$4\Psi_0^4 - \frac{32}{3b}(K_q - a)\ \Psi_0^2 + \frac{64}{9b^2}(K_q - a)^2 \left[\frac{Kp^2}{(Kq-a)^2} + 1\right] \geq 0 . \quad (25)$$

Expressions (25), when the right side is equal to zero, describes real ellipses in the $K_p 0 \Psi_0^2$ plane. Coordinates of ellipse centers (26)

$$K_p = 0,\ \ \Psi_0^2 = \frac{4}{3b}(K_q - a). \quad (26)$$

Values of the semi-axes (27):

$$aK_q = \frac{16}{3b}(K_q - a), \quad b_{(\Psi 0^2)} = \frac{8}{3b}(K_q - a). \quad (27)$$

To determine the boundary of the soft excitation, we solve (25) with respect to Ψ_0^2, as a result we obtain (28):

$$(\Psi_0^2)12 \geq \frac{8}{3b}(K_q - a) \pm \pm\sqrt{\frac{64}{9b^2}(Kq - a)^2 - \frac{16}{9b^2}(Kq - a)^2 - [(K1)2 + 1} \quad (28)$$

From here it can be seen that the area of "soft" excitation is limited: from below and from above by the values (Ψ_0^2). The graph of the separation of the areas of "soft" and "hard" excitation is shown in Figure 4.

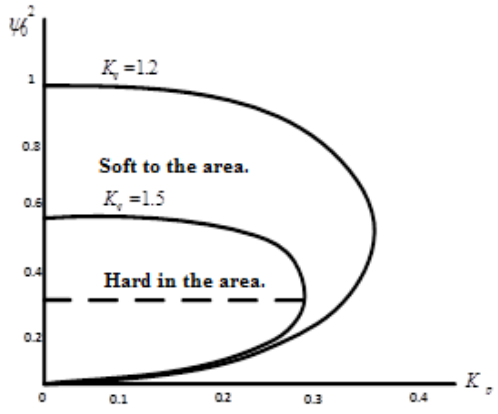


Figure 4: Separation of the areas of "soft" and "hard" excitation is shown.

4 CONCLUSION

The paper has accomplished the following:

1) Using frequency–energy relations, an equation is obtained that characterizes the steady-state mode of existence of the second-order SGC.
2) The results of the analysis of the obtained equations show that the second-order subharmonic oscillations are excited "gently" at certain ratios of the circuit parameters, input voltage and magnetization current.
3) The dependence $Kp = f(Kq)$ is obtained, characterizing the value of the converted power by a reactive nonlinear element to the frequency of the subharmonic oscillations from the power consumption. The analysis of the dependence $K_p = f(K_q)$ also makes it possible to determine the critical values of the circuit parameters characterizing the region of existence of the subharmonic oscillations of a particular frequency close to the engineering calculation method.

REFERENCES

[1] E.B. Saitov, Sh. Kodirov, Z.F. Beknazarova, A. Nortojiyev, and N. Siddikov, "Developing Renewable Sources of Energy in Uzbekistan Renewable Energy Short Overview: Programs and Prospects," AIP Conference Proceedings, vol. 2432, p. 020015, 2022.

[2] Y. Ergashov, B. Donaev, S. Khudainazarov, and J. Normuminov, "Formation of photoelectron spectra of alloys niobium-molybdenum-zirconium," E3S Web of Conferences, vol. 264, no. 05036, 2021, doi: 10.1051/e3sconf/202126405036.

[3] R. Suwarno, "Optimization Circuit Based Buck-Boost Converter for Charging the Solar Power Plant," Indonesian Journal of Electrical Engineering and Computer Science, vol. 6, no. 2, pp. 254-258, May 2017, doi: 10.11591/ijeecs.v6.i2.pp254-258.

[4] E.B. Saitov, T.B. Sodiqov, "Modeling an Autonomous Photovoltaic System in the Matlab Simulink Software Environment," AIP Conference Proceedings, vol. 2432, p. 020022, 2022.

[5] E.B. Saitov, Sh. Kodirov, B.M. Kamanov, N. Imomkulov, and I. Kudenov, "Increasing the Efficiency of Autonomous Solar Photovoltaic Installations for Power Supply of Agricultural Consumers," AIP Conference Proceedings, vol. 2432, p. 040036, 2022.

[6] F. Zikrillayev, E. B. Saitov, J. B. Toshov, B. K. Ilyasov, and M. B. Zubaydullayev, "A Software Package for Determining the Optimal Composition and Parameters of a Combined Autonomous Power Supply System Based on Renewable Energy Sources," in AIP Conference Proceedings, vol. 2432, 020021, 2022.

[7] N.N. Bogolyubov and Yu.A. Mitropolsky, Asymptotic Methods in the Theory of Nonlinear Oscillations. Moscow: Nauka, 1974.

[8] G. Korn and T. Korn, Handbook of Mathematics for Scientists and Engineers. Moscow: Nauka, 1984, pp. 64-74.

[9] L.A. Bessonov, Nonlinear Electrical Circuits. Moscow: Higher School, 1977.

[10] V.I. Ivashev, Oscillations in Nonlinear Electrical Systems. Tashkent: Fan, 1967.

[11] V.V. Bazutkin et al., Overvoltage in Electrical Systems and Protection Against Them. St. Petersburg: Energoatomizdat, 2010.

[12] M. Ibadullaev, A.N. Tovbaev, and A.K. Esenbekov, "Self-oscillations at the frequency of subharmonics in nonlinear electric chains and systems," in International Scientific Conference "Construction and Architecture: Theory and Practice for the Innovation Development" (CATPID-2019), Scopus: E3S Web of Conferences, vol. 138, 2019.

[13] A.S. Karimov, M. Ibadullaev, B. Abdullaev, and A.G. Abdurakhmanov, "Ferromagnetic frequency divider by two," A.S. USSR no. 1467703, 1989, no. II.

[14] M. Ibadullaev, A.N. Tovbaev, and A.K. Esenbekov, "On the general theory of the analysis of subharmonic oscillations in three-phase ferroresonant circuits and systems," Theoretical and Scientific-Practical Journal "Electricity," Moscow, Russia, 2021, no. 12, pp. 35-44.

[15] F.P. Zharkov and G.A. Loginov, "Static ferromagnetic converter of single-phase voltage and multiphase voltage of reduced frequency," A.S. No.280563 BYUL. "Discovery invention," 1970, no. 28.

[16] M. Ibadullaev, A.N. Tovbaev, and A.K. Esenbekov, "Self-oscillations at the frequency of subharmonics in nonlinear electric chains and systems," in International Scientific Conference "Construction and Architecture: Theory and Practice for the Innovation Development" (CATPID-2019), vol. 138, E3S Web of Conferences, 2019.

[17] G.H. Cirtina and V. Alexsandrescu, "Ferorezonanta subarmonica in liniile de transport de energie elektrica," "Bul. inst. poletehn IASI," 1974, SEC 3, vol. 20, pp. 69-76.

[18] M. Ibadullaev and A.N. Tovbaev, "Research of ferroresonance oscillations at the frequency of subharmonics in three-phase non-linear electric circuits and systems," in Rudenko International Conference on Methodological Problems in Reliability Study of Large Energy Systems, RSES 2020, vol. 216, E3S Web of Conferences, Dec. 2020.

[19] A.N. Tovbaev and M. Ibadullaev, "Frequency-energy ratios in the analysis of autoparametric oscillations," Mountain Bulletin, Uzbekistan, 2017, No. 2, pp. 165-170.

[20] J.A. Wright, "Subharmonic oscillations in power systems," IEEE Trans. Power App. Syst., vol. PAS-89, no. 8, pp. 61-65, Aug. 1970.

[21] K. Okumura, S. Nakamura, and A. Kishima, "Experimental and analytical investigation on 1/3-subharmonic oscillations in three-phase circuits," IEEJ Trans. Power Energy, vol. 107, no. 1, 1987, doi: 10.1541/ieejpes1972.107.1.

Effect of Temperature on Electrical Parameters of Photovoltaic Module

Boysori Yuldoshov[1], Elyor Saitov[2], Jasur Khaliyarov[1], Sardor Bobomuratov[1], Sirojiddin Toshpulatov[1] and Fotima Kholmurzayeva[1]
[1]*Termez State University, Barkamol avlod Str. 43, Termez, Uzbekistan*
[2]*Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan*
b.yuldoshov10@mail.ru, elyor.saitov@mail.ru, xjxjasur@mail.ru, bobomuratovsardor1@gmail.com, sirojiddin6870@gmail.com, fotimaxolmirzayeva42@gmail.com

Keywords: Photovoltaic (PV) Module, Solar Cell, Solar Radiation Intensity, Short Circuit Current, Open Circuit Voltage, Electric Power.

Abstract: In this study, the temperatures of the front glass, back sheet, and solar cell (SC) in the rear electrical contact of the PV module during heating, and the open circuit voltage and short circuit current corresponding to these temperatures were experimentally investigated. Measurements were conducted in the city of Termez in conditions of an air temperature of 30ºC and solar radiation intensity of 850-950W/m^2. Two mono c-Si PV modules with the same electric power were selected for the experiment. Three K-type chrome and alumn combined thermocouples were used to measure the temperatures in the layers of the first PV module. During the observations, it was found that the temperature of the PV module glass differs from the back sheet temperature by ~20ºC, and from the temperature of the back electrical contact of the SC by ~25ºC. Also, according to the temperature difference, the open circuit voltage of the PV module decreased from 21V to 19.3V, and the short circuit current increased from 0.65A to 0.75A. During the experiment, the values of electric power changed by ~11-11.6W.

1 INTRODUCTION

Solar energy is going to play an important role in the main supply of electricity in the future. The use of PV modules in the usage of solar energy is the most promising type [1]. The temperature of the PV module is an important parameter that changes the electricity production of PV systems. In addition, it also depends on several parameters such as the thermal properties of the materials used in the manufacture of PV modules, the types of SC, the installation angle of the PV module, and the climatic conditions [2]. However, there are no global monitoring stations to measure such climate conditions and other meteorological parameters, including solar radiation intensity [3]. Despite the advantages of using PV modules, there are disadvantages due to their changing characteristics from time to time. This may affect the performance of the consumer's system. Therefore, an accurate PV module power forecast is required [4].

It is known that an increase in the intensity of solar radiation causes a rise in the SC temperature. Several studies have found that high temperatures can reduce PV module efficiency by up to 12% [5]. Nouar Aoun researched several methods to evaluate the effect of temperature on a PV module [6]. Due to the effect of temperature, only 15-20% of the solar energy in the PV module is converted into electrical energy [7]. Including that part of the solar radiation energy is returned from the surface of the PV module [8]. In addition, dust falling on the surface has a negative effect on the PV module's working potential [9]. A large amount of dust on the surface of the PV module causes it to heat even more [10]. However, the main factor affecting the reduction of PV module output power is high temperature [11]. This is due to the properties of the crystalline silicon semiconductor, a temperature increase of 1ºC leads to a decrease in SC efficiency by about 0.5%. Figure 1 shows the relationship between the efficiency and temperature of a silicon-based PV module [12].

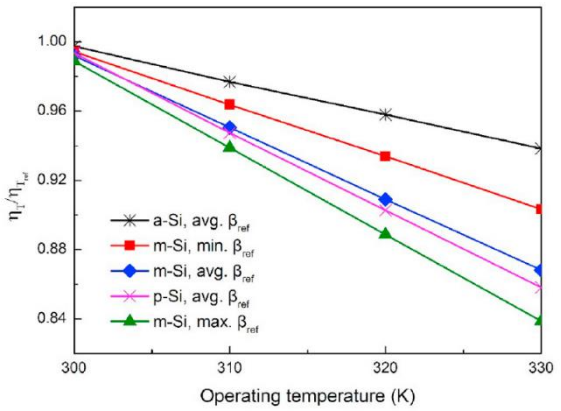


Figure 1: Temperature dependence of PV module efficiency [13].

Malaysian scientists studied the effect of temperature on the PV module by conducting experiments using K-type thermocouples in indoor and outdoor environments [13]. The authors of the study [14] measured the temperature of floating and ground-mounted PV modules using thermocouples and investigated the change in PV module efficiency. Also, Ahmed Amine Hachicha et al. compared the temperatures and electrical parameters of a PV module and a photovoltaic-thermal system using a PT100 thermocouple [15]. A paper by Yogeswara Rao Golive et al., measured the temperature of the SC, ethylene-vinyl acetate (EVA), front glass, and back cover of a c-Si PV module during heating and cooling using T-type thermocouples. In this case, the difference between SC and the back cover was 2.35ºC [16].

It is important to measure the temperature distribution in the parts that make up the PV module due to the decrease in the efficiency of the PV module under the influence of high temperature. This article describes the results of experimental measurements of temperatures and electrical parameters in the PV module structure in hot climate conditions.

2 METHODS AND MATERIALS

This research work differs from the works in the literature mentioned above. Two identical mono c-Si PV modules were used in the experiment. The electric power of PV modules is 15W and consists of 36 SC. The PV modules are mounted on a support device that has a two-axis twist at the same angle. The base device also has a place for measuring instruments (Figure 2).

In the study, the temperatures of the front glass, back sheet, and rear electrical contact of the SC of the first PV module were measured using K-type thermocouples (Model: PeakTech 3340 DMM, measurement range: -20°C to +760 °C). The open circuit voltage and short circuit current of the second PV module was measured using a Multimeter (Model: CHY VC-9205AL, measurement range: 200m V to 1000V, 2mA to 10A). The radiation intensity was recorded using a digital Solar Power Meter (Model: CEM DT-1307, measuring range: up to 1999W/m^2).

Figure 2: Experimental devices for the experiment.

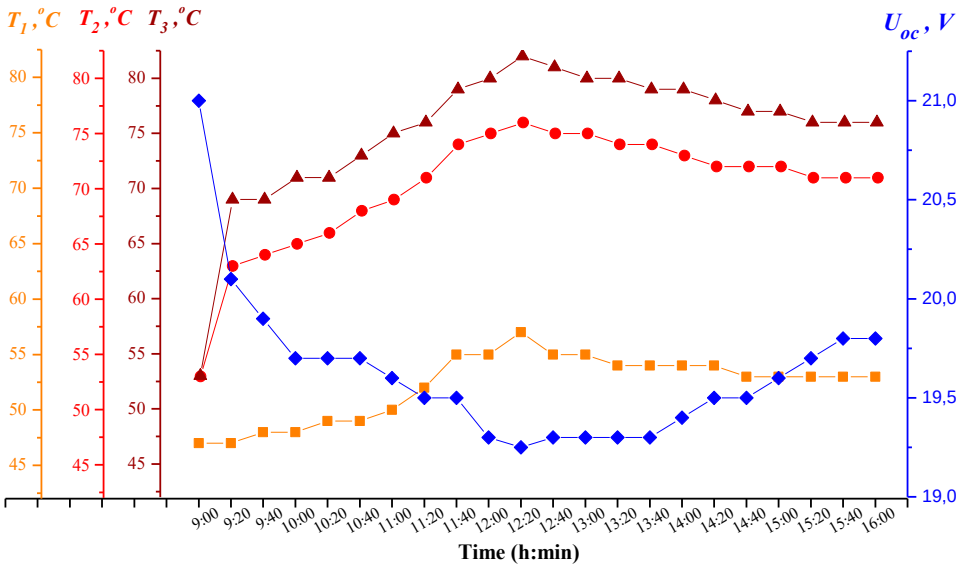


Figure 4: Dependence of PV module open circuit voltage on temperatures of SC parts.

The experiments were conducted in the open air on the territory of Termez State University, at 37°13′27″ north latitude, 67°16′42″ east longitude. The data were measured on October 4, 2022, with an air temperature of 30ºC and solar radiation intensity of 850-950W/m^2, between 9:00 a.m. and 4:00 p.m. During the measurement, the position of the PV modules in the device was changed from 34º to 68º (in tracker mode) with a south orientation.

Three K-type chrome and aluminum combined thermocouples were placed as shown in Figure 3 to determine the temperatures in the components of the SC.

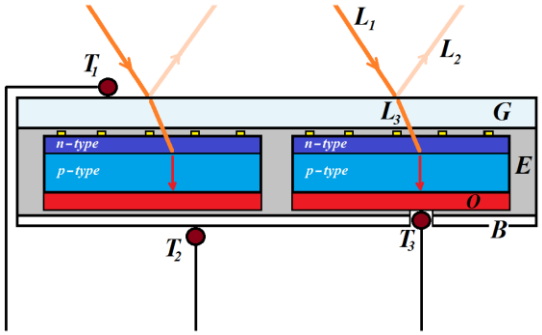


Figure 3: Installations of thermocouples in the PV module. T_1, T_2, T_3- thermocouples, L_1 - falling radiation, L_2 - returned light, L_3 - absorbed light, G - front glass old, E - EVA, O - rear electrical contact, B - back sheet.

In the experiment, the temperature of PV module parts and corresponding electrical parameters were measured at 20-minute intervals. In the results, the dependence of the PV module open circuit voltage on the temperatures in the SC parts, and the dependence of the short circuit current on the intensity of solar radiation were studied. Another goal of the outdoor experiment is to determine the optimal installation angle of the PV module and the zenith point of the sun for the conditions of Termez.

3 RESULTS AND DISCUSSION

The temperature of PV modules depends on the value of the incident solar radiation intensity. When the radiation passes through the front glass of the PV module, some of it is reflected, but most of it is absorbed by the semiconductor material. As a result, the temperature of the SC increases and the heat accumulates in the rear electrical contact of the SC. Therefore, the temperature of the glass of the PV module differs by ~20-25ºC from the temperature of the back sheet and the back electrical contact of the SC (Figure 4).

As you can see, the temperature of the PV module increased between 9:00 AM and 12:20 PM and then decreased again after the solar zenith point. Measurements were made at equal times before the zenith point of the sun (3.5 hours) and after the zenith point (3.5 hours). Nevertheless, SC temperature did not approach the initial value. This process can be explained by the retention of heat in SC for a certain period. It can be seen that the PV module open circuit voltage initially decreased by ~1.75V as the temperature increased, and then increased by ~0.55V

after the zenith point of the sun. However, it is not difficult to understand that the PV module does not reach its initial value due to the heat accumulated in the module.

The increase of short circuit current in a PV module under the influence of temperature is a less useful process, and the change of short circuit current mainly depends on the intensity of solar radiation (Figure 5).

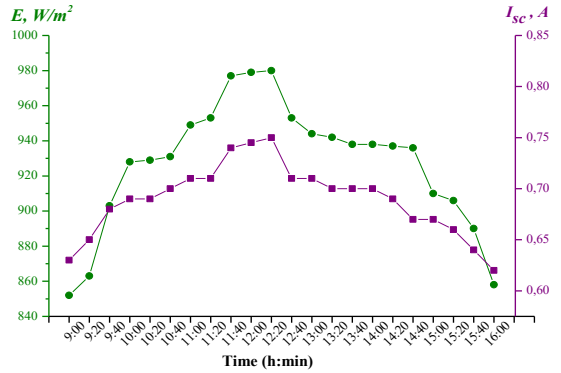


Figure 5: Dependence of short circuit current on the intensity of solar radiation.

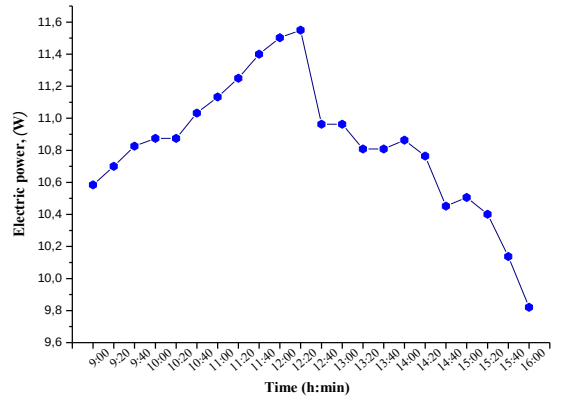


Figure 6: Dependence of electric power on time of day.

It can be seen in Figure 5, the intensity of solar radiation starts from 852W/m^2 at 9:00 and the maximum value reaching the solar zenith point is 980W/m^2. It can be seen that the value of the PV module short circuit current also increased from 0.63A to 0.75A according to the intensity of solar radiation.

Based on the above information, it was observed that the electric power increased from 10.5W to 11.5W (Figure 6). At the end of the experiment, the electric power decreased to 9.8W. This is a very bad indicator for a PV module with a maximum power of 15W. Such a decrease in electric power can be explained by the depletion in the open circuit voltage due to the heat accumulated in the PV module and the fading of the intensity of solar radiation.

4 CONCLUSION

It is a necessary task to study the causes of electrical losses in PV modules. The main factor that reduces the efficiency of PV systems is high temperature. In this study, the influence of high temperature on the performance of PV modules in the conditions of the city of Termez, which is located in the southernmost point of Uzbekistan, was studied. In the study, temperatures and electrical parameters of two PV modules attached to a two-axis support device were measured in the tracker mode. Based on the results obtained from the experiment, the following conclusions were made:

1) The increase in the temperature of PV modules depends on the intensity of solar radiation falling on its surface, and the heat generated by the infrared field of radiation is collected in the back electric contact SC. Therefore, the temperature of the SC can be higher than the temperature of the front glass;

2) The increase in short circuit current in the PV module under the influence of high temperature is not enough to significantly contribute to its power value;

3) Due to the high temperature mainly reducing the open circuit voltage of the PV module, its electrical energy loss reaches ~35%.

REFERENCES

[1] M. Abdel-Basset, D. El-Shahat, R. K. Chakrabortty, and M. Ryan, "Parameter estimation of photovoltaic models using an improved marine predators algorithm," Energy Conversion and Management, vol. 227, pp. 113491, 2021. [Online]. Available: https://doi.org/10.1016/j.enconman.2020.113491.

[2] A. Rodolfo, U. Grasselli, and S. Celozzi, "Assessment of a Practical Model to Estimate the Cell Temperature of a Photovoltaic Module," International Journal of Energy and Environmental Engineering, vol. 5, no. 1, pp. 1-16, 2014. [Online]. Available: https://link.springer.com/article/10.1007/s40095-014-0072-x.

[3] V. Z. Antonopoulos, D. M. Papamichail, V. G. Aschonitis, and A. V. Antonopoulos, "Solar radiation estimation methods using ANN and empirical models," Computers and Electronics in Agriculture, vol. 160, pp. 160-167, 2019. [Online]. Available: https://doi.org/10.1016/j.compag.2019.03.022.

[4] M. S. El Nozahy, M. M. A. Salama, and R. Seethapathy, "A probabilistic load modelling

approach using clustering algorithms," 2013 IEEE Power and Energy Society General Meeting, pp. 1-5, 2013. [Online]. Available: https://ieeexplore.ieee.org/document/6672073.

[5] K. Kant, A. Shukla, A. Sharma, and P. H. Biwole, "Thermal response of poly-crystalline silicon photovoltaic panels: numerical simulation and experimental study," Solar Energy, vol. 134, pp. 147-155, 2016. [Online]. Available: https://doi.org/10.1016/j.solener. 2016.05.002.

[6] N. Aoun, "Methodology for predicting the PV module temperature based on actual and estimated weather data," Energy Conversion and Management: X, vol. 14, pp. 100182, 2022. [Online]. Available: https://doi.org/10.1016/j.ecmx.2022.100182.

[7] 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," Renewable Energy, vol. 132, pp. 284-295, 2019. [Online]. Available: https://doi.org/10.1016/j.renene.2018.08.008.

[8] M. Hasanuzzaman, A. B. M. A. Malek, M. M. Islam, A. K. Pandey, and N. A. Rahim, "Global advancement of cooling technologies for PV systems: a review," Solar Energy, vol. 137, pp. 25-45, 2016. [Online]. Available: https://doi.org/10.1016/j.solener.2016.07.010.

[9] N. J. Middleton, "Desert dust hazards: A global review," Aeol. Res., vol. 24, pp. 53-63, 2017. [Online]. Available: https://doi.org/10.1016/j.aeolia.2016.12.001.

[10] Y. Chen, Y. Liu, Zh. Tian, Y. Dong, Y. Zhou, X. Wang, and D. Wang, "Experimental Study on the Effect of Dust Deposition on Photovoltaic Panels," Ener. Proc., vol. 158, pp. 483-489, 2019. [Online]. Available: https://doi.org/10.1016/j.egypro.2019.01.139.

[11] M. M. Fouad, L. A. Shihata, and E. S. I. Morgan, "An integrated review of factors influencing the performance of photovoltaic panels," Ren. and Sust. Ener. Rev., vol. 80, pp. 1499-1511, 2017. [Online]. Available: https://doi.org/10.1016/j.rser.2017.05.141.

[12] E. Skoplaki and J. A. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: a review of efficiency/power correlations," Sol. Ener., vol. 83, pp. 614-24, 2009. [Online]. Available: https://doi.org/10.1016/j.solener. 2008.10.008.

[13] M. A. A. Mamun, M. M. Islam, M. Hasanuzzaman, and J. Selvaraj, "Effect of tilt angle on the performance and electrical parameters of a PV module: Comparative indoor and outdoor experimental investigation," Ener. and Buil. Env., vol. 3, pp. 278-290, 2022. [Online]. Available: https://doi.org/10.1016/j.enbenv.2021.02.001.

[14] H. Nisar, A. K. Janjua, H. Hafeez, S. Shakir, N. Shahzad, and A. Waqas, "Thermal and electrical performance of solar floating PV system compared to on-ground PV system-an experimental investigation," Sol. Ener., vol. 241, pp. 231-147, 2022. [Online]. Available: https://doi.org/10.1016/j.solener.2022.05.062.

[15] A. A. Hachicha, E. M. Abo-Zahhad, Z. Said, and S. M. A. Rahman, "Numerical and experimental investigations of the electrical and thermal performances of a novel PV thermal system," Ren. Ener., vol. 195, pp. 950-1000, 2022. [Online]. Available: https://doi.org/10.1016/j.renene.2022.06.083.

[16] Y. R. Golive, A. Kottantharayil, and N. Shiradkar, "Improving the accuracy of temperature coefficient measurement of a PV module by accounting for the transient temperature difference between cell and backsheet," Sol. Ener., vol. 237, pp. 203-212, 2022. [Online]. Available: https://doi.org/10.1016/j.solener. 2022.03.049.

Calculation and Analysis of the Thermal State of the Frequency-Controlled Induction Motor Pump Unit

Ravshan Dusmatov
Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan
iis_ravshan@mail.ru

Keywords: Calculation Method Of Thermal State, Thermal Body, Frequency-Controlled Electric Drive, High-Voltage Induction Motor, Power, Pump, Graph Model.

Abstract: The article presents the results of the application of a method for calculating the thermal state of a frequency-controlled induction motor of closed design in steady-state heating mode. The study analyzes the excess temperature of thermal bodies, particularly the stator winding of the frequency-controlled induction motor installed in the pumping unit. Previous studies on regulated electric drives based on induction motors have focused on low-voltage induction motor drives for general industrial purposes. However, high-voltage induction motors have distinct specific features in terms of design and operation. The study of thermal processes in such motor drives is mainly based on three research methods: the method of heating thermal parameters, the method of the temperature field, and the method of equivalent thermal circuits. The analysis of these methods of thermal calculation showed that the most acceptable method of thermal calculation of regulated high-voltage induction motors is the method of equivalent thermal circuits (ETC). Based on the analysis of the excess temperature of thermal bodies, in particular the stator winding of the frequency-controlled induction motor installed in the pumping unit, it is shown that the optimal frequency control range lies within 30-50 Hz, that is, the frequency reduction causes an increase in temperature in all thermal bodies of the engine.

1 INTRODUCTION

As is known, most of the studies of electrical, energy and thermal characteristics of the regulated electric drive on the basis of induction motors at different values of mechanical loads on their shaft in order to obtain technical solutions that improve their parameters related to low-voltage induction motor drives for general industrial purposes [1-8]. Despite the fact that high-voltage induction motors have much incommon with low-voltage induction motors in terms of design, they have their own distinct specific features, both in terms of design and operation. The study of thermal processes in such motor drives is mainly based on three research methods: the method of heating thermal parameters, the method of the temperature field and the method of equivalent thermal circuits [9]. The analysis of these methods of thermal calculationshowed that the most acceptable method of thermal calculation of regulated high-voltage induction motors is the method of equivalent thermal circuits (ETC). The essence of this method is to replace the solution of the one-dimensional stationary Laplace equation to such a transformation of the element size, which would contribute to the achievement of the desired numerical result, as well as to obtain two-dimensional solutions by adding the thermal conductivities of one- dimensional, based on the equivalence of thermal circuits with linear electric circuits.

2 METHODS AND MATERIALS

The article presents the results of studies of the main thermal parameters of high-voltage induction motors obtained by the proposed method of calculating the thermal state by drawing up graph models for each thermal body – the active parts of the motor [10] for the steady-state heating of the frequency-controlled induction motor (IM) of closed design. In this case, the resulting power loss between the source and the temperature of the thermal bodies equivalent transmission have a dimension of thermal resistance [11] and therefore they will be called the total thermal resistance of the thermal bodies R_{ii} ,

as well as thermal resistance between the thermal bodies R_{ij} (where i and j – the order numbers of the thermal bodies), then the final form of the solution of the steady-state heating will take the following form (1)

$$\left.\begin{aligned}\Theta_{1f} &= R_{11f}P_{1f} + R_{12f}P_{2f} + R_{13f}P_{3f};\\ \Theta_{2f} &= R_{21f}P_{1f} + R_{22f}P_{2f} + R_{23f}P_{3f};\\ \Theta_{3f} &= R_{31f}P_{1f} + R_{32f}P_{2f} + R_{33f}P_{3f};\end{aligned}\right\} \quad (1)$$

where $R_{11f} = \dfrac{\Lambda_{11f}\Lambda_{22f}\Lambda_{33f}}{(\Lambda_{11f}\Lambda_{33f} - \Lambda_{13f}\Lambda_{31f})(\Lambda_{11f}\Lambda_{22f} - \Lambda_{12f}\Lambda_{21f})}$;

$$R_{12f} = \frac{\Lambda_{13f}\Lambda_{12f}\Lambda_{22f}}{(\Lambda_{11f}\Lambda_{33f} - \Lambda_{13f}\Lambda_{31f})(\Lambda_{11f}\Lambda_{22f} - \Lambda_{12f}\Lambda_{21f})};$$

$$R_{13f} = \frac{\Lambda_{22f}(\Lambda_{11f}\Lambda_{33f} + \Lambda_{12f}\Lambda_{21f})}{(\Lambda_{22f}\Lambda_{33f} - \Lambda_{23f}\Lambda_{32f})(\Lambda_{11f}\Lambda_{22f} - \Lambda_{12f}\Lambda_{21f})};$$

$$R_{21f} = \frac{\Lambda_{21f}\Lambda_{33f}\Lambda_{11f}}{(\Lambda_{11f}\Lambda_{33f} - \Lambda_{13f}\Lambda_{31f})(\Lambda_{11f}\Lambda_{22f} - \Lambda_{12f}\Lambda_{21f})};$$

$$R_{22f} = \frac{\Lambda_{33f}\Lambda_{11f}\Lambda_{22f}}{(\Lambda_{11f}\Lambda_{33f} - \Lambda_{13f}\Lambda_{31f})(\Lambda_{11f}\Lambda_{22f} - \Lambda_{12f}\Lambda_{21f})};$$

$$R_{23f} = \frac{\Lambda_{12f}\Lambda_{11f}\Lambda_{33f}}{(\Lambda_{22f}\Lambda_{33f} - \Lambda_{23f}\Lambda_{32f})(\Lambda_{11f}\Lambda_{22f} - \Lambda_{12f}\Lambda_{21f})};$$

$$R_{31f} = \frac{\Lambda_{12f}\Lambda_{33f}\Lambda_{11f}}{(\Lambda_{22f}\Lambda_{11f} - \Lambda_{12f}\Lambda_{21f})(\Lambda_{11f}\Lambda_{22f} - \Lambda_{12f}\Lambda_{21f})};$$

$$R_{32f} = \frac{\Lambda_{13f}\Lambda_{22f}\Lambda_{11f}}{(\Lambda_{11f}\Lambda_{22f} - \Lambda_{12f}\Lambda_{21f})(\Lambda_{22f}\Lambda_{33f} - \Lambda_{23f}\Lambda_{32f})};$$

Further, in Figure 1 the changes of total thermal resistances for IM of A4-457-UH-8UZ type at frequency control in the range of 30-50 Hz are presented. As can be seen, the total thermal resistance of thermal bodies of IM with a decrease in the frequency have the character of increasing. The total thermal resistance of the stator winding and the stator steel (R_{11f} and R_{22f}) with a decrease in the frequency value increases slightly, but the total resistance of the rotor R_{33f} increases at a sufficiently high rate, which is associated with a deterioration in the intensity of the airflow.

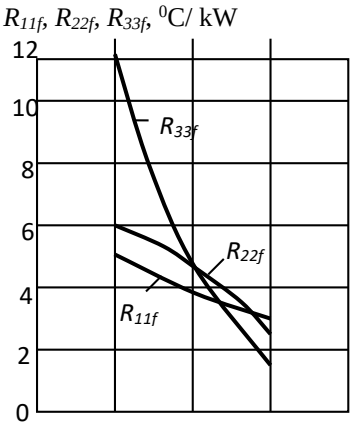


Figure. 1: The change of the total thermal resistances of the thermal bodies of frequency-controlled IM of A4-457-UH-8UZ type in frequency function.

In Figure 2, the changes of thermal resistances between thermal bodies of IM of A4-457-UH-8UZ type at frequency control in the range of 30-50 Hz are presented. As can be seen from the Figure 2, the thermal resistance between the thermal bodies of IM also has an increasing character, but their increases are insignificant. In (Figure 3), the changes of power losses of thermal bodies of IM of A4-457-UH-8UZ type at frequency control are presented, when the law of frequency control in the range of 30-50 Hz for the fan type of load is realized.

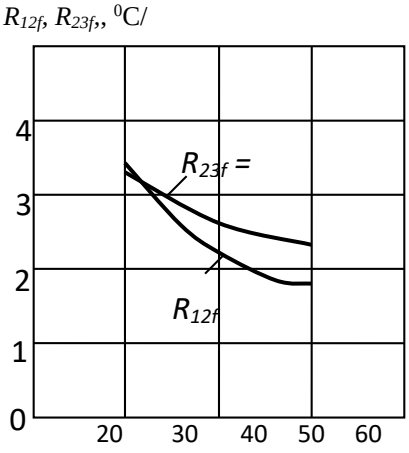


Figure 2: Changes of the thermal resistances between the thermal bodies of frequency-controlled IM of A4-457-UH-8UZ type in frequency function.

All the power losses of thermal bodies of this IM with a decrease in the frequency (respectively the decrease of mechanical power on the motor shaft in proportion ω^2) are also reduced according to the nonlinear law [11, 12, 13].

3 RESULTS AND DISCUSSION

Patterns reduce the loss of heat capacities of all thermal bodies are almost identical. Since in the frequency range 30-50 Hz, the power losses of the rotor IM P3f vary, due to the fact that they consist of electrical losses of the rotor winding (with a decrease in the load, respectively, the current decreases, and electrical losses decrease in proportion to the square of the relative change in the nominal value) as well as losses in the bearings (mechanical losses are associated with a quadratic rate of change) and therefore both losses decrease with a decrease in frequency [14,15].

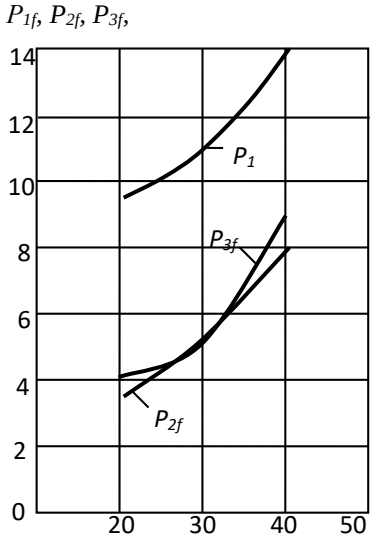


Figure 3: Changes of the power losses of thermal bodies of frequency-controlled IM of A4-457-UH-8UZ type in frequency function.

With the help of the obtained thermal model for each thermal body and solving a system of equations (1), let us determine the excess temperature Θ_{1f}, Θ_{2f}, Θ_{3f} in some parts of the frequency-controlled IM of A4-457-UH-8UZ type in the implementation of the fan nature of the load (Figure 4) [16, 17, 18, 19, 20]. Exceeding the temperature of all thermal bodies of IM at the beginning of the frequency change at a frequency of 40 Hz is reduced, which is associated with a decrease in the power loss of thermal bodies with a decrease in the mechanical load on the motor shaft. At this frequency, there are minor increases in all types of thermal resistances, which especially do not show their effects on the values of excess temperature of thermal bodies, that is, the change in the speed of rotation of the IM has little effect on them. In the frequency range from 30 to 50 Hz due to the decrease in the intensity of the airflow of thermal bodies, a sharp decrease in the power loss occurs, and, despite these excess temperatures of thermal bodies, they reach values corresponding to their nominal values [21, 22. 23, 24, 25, 26].

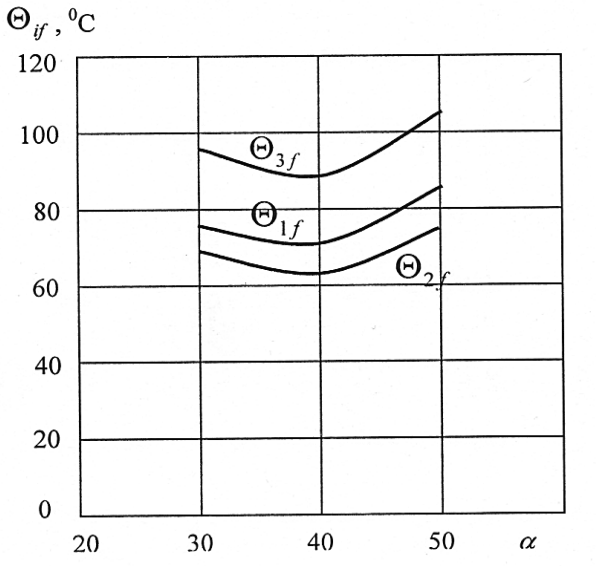


Figure 4: Changes, exceeding the temperature of thermal bodies of frequency-controlled IM of A4-457-UH-8UZ type in the frequency function.

Analysis of the excess temperature of thermal bodies, in particular the stator winding Θ_{1f} of frequency-controlled IM of A4-457-UH-8UZ type (installed in the pumping unit) shows that the optimal frequency control range is 30-50 Hz, that is, a decrease in the frequency cause an increase in temperature in all thermal bodies of the motor. This is due to the design of this type of IM, since this motor having a protected design is blown by a fan located on the shaft of its rotor. Since the performance of the fan is directly related to the speed of rotation of the IM rotor, at low speeds of its rotation, the latter does not contribute to improving the airflow of the motor components.

4 CONCLUSIONS

In order for the IM to work consistently throughout the frequency control range, it is necessary to use an autonomous ventilation system that would allow even at low speed values (rotation frequencies) to improve the degree of air flow of thermal bodies, which in turn will reduce the values exceeding the temperatures of thermal bodies of the frequency-controlled IM pump unit.

Summarizing the above, it can be noted that the developed method for studying the steady-state heating mode of a frequency-controlled induction motor on the basis of a graph model of equivalent thermal circuits allows:

1) To determine the main thermal parameters and the excess temperature of the thermal bodies of the frequency-controlled induction motor;
2) To identify the optimal range of speed control (frequency), taking into account the thermal state of the stator winding.

REFERENCES

[1] I.Ya. Braslavskiy, Z.Sh. Ishmatov, and V.N. Polyakov, "Energy-saving induction motor drive," Moscow: ACADEMA, 2004, pp. 280.

[2] M.M. Khamudkhanov, I.A. Abdullabekov, R.K. Dusmatov, N.B. Khamudkhanova, and B.KH. Fayzullayev, "Controls of the modes of operation of the pumping station with application of frequency-controlled electric drive," in Web of Conferences, II International scientific conference, IOP Conf. Series: Materials Science and Engineering, vol. 862, 2020, pp. 062048IOP. Publishing, [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/862/6/062048.

[3] Islombek Abdullabekov and Khushnud Sapaev, "An Energy Efficient Control System for Water Lifting Units of the Ramadan Pumping Station Based on Frequency Controlled Electric Drives," AIP Conference Proceedings, vol. 2552, no. 1, 2023, article no. 040023, [Online]. Available: https://doi.org/10.1063/5.0130676.

[4] Khushnud Sapaev, Shukhrat Umarov, Islombek Abdullabekov, Nargiza Khamudkhanova, and Maxamanazar Nazarov, "Scheme of effective regulation of pumping station productivity," AIP Conference Proceedings, vol. 2402, no. 1, 2021, article no. 060016, [Online]. Available: https://doi.org/10.1063/5.0071557.

[5] Khushnud Sapaev, Islombek Abdullabekov, and Shukhrat Umarov, "Research energy and resource saving operating modes of the pump unit," E3S Web of Conferences, vol. 216, 2020, article no. 01150, The Authors, published by EDP Sciences, [Online]. Available: https://doi.org/10.1051/e3sconf/202021601150.

[6] Khushnud Sapaev, Islombek Abdullabekov, and Shukhrat Umarov, "Critical frequency of autonomous current inverter when operating on active-inductive load," E3S Web of Conferences, vol. 216, 2020, article no. 01153, The Authors, published by EDP Sciences, [Online]. Available: https://doi.org/10.1051/e3sconf/202021601153.

[7] E.B. Saitov and T.B. Sodiqov, "Modeling an Autonomous Photovoltaic System in the Matlab Simulink Software Environment," AIP Conference Proceedings, vol. 2432, no. 1, 2022, article no. 020022.

[8] E.B. Saitov, Sh. Kodirov, B.M. Kamanov, N. Imomkulov, and I. Kudenov, "Increasing the Efficiency of Autonomous Solar Photovoltaic Installations for Power Supply of Agricultural Consumers," AIP Conference Proceedings, vol. 2432, no. 1, 2022, article no. 040036.

[9] F. Zikrillayev, E.B. Saitov, J.B. Toshov, B.K. Ilyasov, and M.B. Zubaydullayev, "A Software Package for Determining the Optimal Composition and Parameters of a Combined Autonomous Power Supply System Based on Renewable Energy Sources," in AIP Conference Proceedings, 2022, vol. 2432, p. 020021.

[10] A.A. Khashimov, A.U. Mirisaev, and L.T. Kan, "Energy-saving induction motor drive," Tashkent: "Fan va texnologiya," 2011, p. 132.

[11] A.A. Khashimov, A.E. Serov, and A.T. Imamnazarov, "Electromagnetic and thermal processes in frequency-controlled induction motors. In the book: Automatic machines, electric drives, power semiconductor devices and converter equipment," Moscow, Energoatomizdat, 1983, pp. 223-231.

[12] A.A. Khashimov, "Electromechanical and thermal processes of frequency-controlled induction motor drive," Tashkent: Fan, 1997, p. 104.

[13] A.A. Khasimov and R.K. Dusmatov, "The method of drawing up a graph model of the thermal state of an induction motor based on its equivalent thermal circuit," Bulletin of the Turin Polytechnic University in Tashkent, vol. 4, pp. 79-84, 2018.

[14] A.T. Imamnazarov, "Calculation of heating of induction motors by means of the transformed equivalent thermal circuits," Journal of Energy and Resource Conservation Problems, Tashkent, vol. 1, pp. 20-25, 2005.

[15] Kh. Sapaev, Sh. Umarov, and I. Abdullabekov, "Critical frequency of autonomous current inverter when operating on active-inductive load," IOP Conf. Ser.: Mater. Sci. Eng., vol. 216, p. 01153, 2020.

[16] N. Pirmatov and A. Panoev, "Frequency control of asynchronous motors of looms of textile enterprises," E3S Web of Conferences, vol. 216, p. 01120, 2020.

[17] Kh. Sapaev, Sh. Umarov, and I. Abdullabekov, "Research energy and resource saving operating modes of the pump unit," IOP Conf. Ser.: Mater. Sci. Eng., vol. 216, p. 01150, 2020.

[18] O. Toirov and S. Khalikov, "Analysis of the safety of pumping units of pumping stations of machine water lifting in the function of reliability indicators," E3S Web of Conferences, vol. 365, p. 04010, 2023, doi: 10.1051/e3sconf/202336504010.

[19] O. Toirov and S. Khalikov, "Diagnostics of pumping units of pumping station of machine water lifting," E3S Web of Conferences, vol. 365, p. 04013, 2023, doi: 10.1051/e3sconf/202336504013.

[20] H. Afrisal, A.D. Setiyadi, M.A. Riyadi, R. Ismail, O. Toirov, and I. Setiawan, "Performance Analysis of 4-DOF RPRR Robot Manipulator Actuation Strategy for Pick and Place Application in Healthcare Environment," International Journal on Advanced Science, Engineering and Information Technology, vol. 12, no. 6, pp. 2258-2265, 2022, doi: 10.18517/ijaseit.12.6.16591.

[21] O. Toirov, D. Jumaeva, U. Mirkhonov, S. Urokov, and S. Ergashev, "Frequency-controlled asynchronous electric drives and their energy parameters," AIP Conference Proceedings, vol. 2552, article no. 040021, 2023, doi: 10.1063/5.0111331.

[22] N. Pirmatov, O. Toirov, A. Bekishev, N. Kurbanov, O. Zaynieva, and U. Norkulov, "Calculation of U-shaped characteristics and reactive power of synchronous compensator with longitudinal-transverse excitation," AIP Conference Proceedings, vol. 2552, article no. 040020, 2023, doi: 10.1063/5.0115727.

[23] D. Jumaeva, O. Toirov, Z. Okhunjanov, U. Raximov, and R. Akhrorova, "Investigation of the adsorption of nonpolar adsorbate molecules on the illite surface," Journal of Chemical Technology and Metallurgy, vol. 58, no. 2, pp. 353-359, 2023.

[24] O. Toirov, S. Urokov, U. Mirkhonov, H. Afrisal, and D. Jumaeva, "Experimental study of the control of operating modes of a plate feeder based on a frequency-controlled electric drive," E3S Web of Conferences, vol. 288, article no. 01086, 2021, doi: 10.1051/e3sconf/202128801086.

[25] A. Aljwary, Z. Yusupov, O. Toirov, and R. Shokirov, "Mitigation of load side harmonic distortion in standalone photovoltaic based microgrid," E3S Web of Conferences, vol. 304, article no. 01010, 2021, doi: 10.1051/e3sconf/202130401010.

[26] D. Jumaeva, O. Toirov, N. Rakhmatullaeva, R. Akhmedov, and I. Eshmetov, "Energy of adsorption of polar molecules on NaLSX zeolite," E3S Web of Conferences, vol. 288, article no. 01041, 2021, doi: 10.1051/e3sconf/202128801041.

[27] D. Jumaeva, A. Abdurakhimov, K. Abdurakhimov, N. Rakhmatullaeva, and O. Toirov, "Energy of adsorption of an adsorbent in solving environmental problems," E3S Web of Conferences, vol. 288, article no. 01082, 2021, doi: 10.1051/e3sconf/ 202128801082.

[28] A. Shaislamov, N. Abdullaev, and A. Akxrorkhodjaev, "Thermal and Economic Performance of a Solar Hot Water System in Tashkent," in Networked Control Systems for Connected and Automated Vehicles, 2022, pp. 781-789.