



Determining nutrients, dry matter, and pH of liquid organic manures using visual and near-infrared spectrometry

Michael Horf^{a,*}, Robin Gebbers^b, Hans-Werner Olf^c, Sebastian Vogel^a

^a Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB), Department of Agromechatronics, Max-Eyth-Allee 100, 14469 Potsdam, Germany

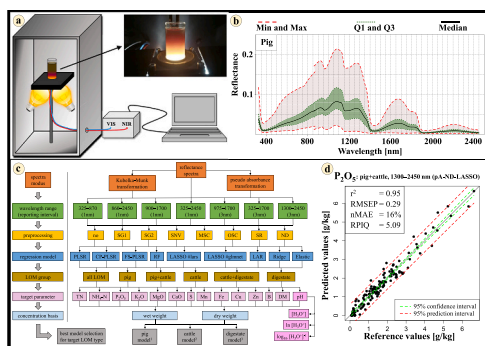
^b Martin Luther University Halle-Wittenberg, Chair of Agricultural Business Operations, Karl-Freiherr-von-Fritsch-Straße 4, 06120 Halle, Germany

^c University of Applied Sciences, Plant Nutrition and Crop Production, Am Krümpel 31, 49090 Osnabrück, Germany

HIGHLIGHTS

- Accurate predictions for macro- and micronutrients in liquid organic manures
- Introduction of two innovative chemometric preprocessing methods (ND, SR)
- Emphasis on important wavelengths with high impact on the prediction accuracy
- Using concentrations on dry weight basis improved predictions of K, Mn, and pH
- Good predictions also obtained with simulated low-cost spectrometers

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Fang Wang

Keywords:

Slurry
Biogas digestates
Optical spectroscopy
Chemometrics
Preprocessing
Low-cost spectrometers

ABSTRACT

In agriculture, overfertilization with liquid organic manures (LOM) is causing environmental issues including eutrophication of non-agricultural ecosystems and nitrate pollution of groundwater. To avoid such problems, a precise and demand-oriented fertilization with LOM is needed. This can only be achieved if the nutrient composition of the LOM is known. However, traditional chemical analysis is cost- and time-intensive and furthermore dependent on a representative sample. Optical spectrometry in the visible and near-infrared range could provide an efficient alternative, if a chemometric calibration assures sufficient accuracy. To improve chemometric calibration, this study investigated several spectral preprocessing and regression algorithms, and compared predictions based either on dry or wet weight concentration. In addition, the capability of low-cost spectrometers was evaluated by simulating low-resolution spectra with smaller wavelength ranges. The reflectance spectra of 391 pig manure, 155 cattle manure, and 89 biogas digestate samples were used to predict plant macronutrients (N, P, K, Mg, Ca, S), micronutrients (Mn, Fe, Cu, Zn, B), dry matter (DM) and pH. The experiments demonstrate the general aptness of optical spectrometry to accurately predict DM, pH and all nutrients except boron in pig, cattle, and digestate LOM, even with simulated low-cost spectrometers. Best results show r^2 between 0.80 and 0.97, ratios of performance to interquartile distance (RPIQ) between 2.1 and 7.8, and mean absolute errors normalized by the median (nMAE) between 5 and 36 %. The regression methods PLSR, LASSO, and least angle regression predominantly performed best. The innovative preprocessing methods named simple

* Corresponding author.

E-mail address: mhorf@atb-potsdam.de (M. Horf).

<https://doi.org/10.1016/j.scitotenv.2023.168045>

Received 2 July 2023; Received in revised form 21 September 2023; Accepted 20 October 2023

Available online 2 November 2023

0048-9697/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

ratios (SR) and normalized differences (ND) proved to be very useful algorithms, especially for N and P predictions, outperforming the accuracy of classical techniques in several cases. Concentrations on dry weight basis improved predictions of K, Mn, and pH.

1. Introduction

Liquid organic manures (LOM), i.e. livestock manures and digestates from biogas production, have become a focal issue in the context of renewable energies, limited mineral fertilizer resources, and precision fertilizer applications by finding a balance between optimal crop supply on the one hand and a prevention of environmental problems caused by overfertilization on the other hand. As excess nutrients will be leached into surface and ground waters (Saeys et al., 2019b) overfertilization results in eutrophication of non-agricultural ecosystems and, in long-term, in the contamination of drinking water with nitrate (Zhang et al., 2020). Furthermore, the relevance of organic fertilizers has become increasingly important due to rising prices for mineral fertilizers on the world market in recent years, strongly influenced by the Russian-Ukraine conflict (Ben Hassen and El Bilali, 2022). LOM contain all essential macro- and micronutrients required for a healthy plant growth (Waldrip et al., 2020) and can be considered a valuable, abundant and low-cost resource with an annually production of more than one billion tons in both the European Union (Foged et al., 2011) and the United States of America (Zhang and Schroder, 2014). The use of LOM instead of mineral fertilizers contributes to a lesser consumption of fossil fuels, preserves mineral resources (e.g. rock phosphates) and enhances soil fertility (Ammaan et al., 2019). As LOM fertilization increases the soil humus content via carbon sequestration, it even contributes to a reduction of CO₂ in the atmosphere (Kitamura et al., 2021).

However, despite these advantages of LOM, it is challenging to exactly quantify the nutrient concentrations, due to its inherent heterogeneity (Aguirre-Villegas et al., 2018). Querying average values from nutrient tables cannot capture the real nutrient diversity (Sanford et al., 2020). The alternative, to carry out classical chemical lab procedures for LOM analyses, are costly, time-consuming and generate hazardous waste. Furthermore, chemical analysis strongly depends on a representative sample, which requires thorough homogenization before sampling LOM (Piepel and Olf, 2023). However, farmers tend to avoid this additional homogenization step because stirring a slurry tank involves extra labour and cost and increases environmentally harmful ammonia (NH₃) and hydrogen sulphide (H₂S) emissions (Park et al., 2020). In fact, farmers often refer to average values of empiric nutrient tables and, moreover, tend to overfertilization in order to ensure a maximum crop supply, while a long-term intact environment is only secondary.

For these reasons, there is a high demand for an alternative measuring technology that quantifies the chemical composition of LOM rapidly, reliably, and at low costs. Optical spectrometry is a fast, clean and non-destructive measuring technology that uses electromagnetic radiation in the ultraviolet (UV), visual (Vis), near-infrared (NIR) or mid-infrared region (MIR). Due to continual improvements in sensor technologies, optical spectrometry is robust enough for on-farm and on-line analysis even though the sensor is exposed to harsh conditions concerning vibrations, temperature, humidity, and dust (Piepel et al., 2022). Furthermore, the rapid development of chemometrics during the last decades (Huang, 2022) has been playing a major role for a continuing progress of optical spectrometry in analytical chemistry (Brereton et al., 2017) and other scientific research areas such as agronomy including LOM analysis. Thus, optical spectrometry may replace lab analysis of LOM if this technique assures sufficient measurement accuracy for LOM analysis on-site, whether in the storage tank, while filling a slurry tanker, or during application in the field.

In regions with high LOM production, surplus LOM has been treated as a cost generating waste since official regulations limit the application as a fertilizer, e.g. EU limit for LOM application: 170 kg nitrogen per ha

and year (EU regulation 2092/91), EU nitrate limit: 50 mg/L in ground-, surface- and drinking water (European Commission directive 91/676/EEC) or US nitrate-N limit: 10 mg/L (equal to 44.3 mg/L nitrate) in ground- and drinking water (US-EPA). Due to recent developments in fertilizer prices and a better awareness of its benefits regarding soil fertility, LOM will be more transferred to other regions as a valuable and profit generating fertilizer, especially if the application of optical spectrometry assures accurate nutrient analysis (Flynn et al., 2023).

First research in the context of manure analysis was carried out by Asai et al. (1993) who used NIR spectrometry (NIRS) to predict total carbon (TC) and total nitrogen (TN) in dried dairy manure. Since then, about 40 studies followed determining nutrient compositions of liquid or dried manures and digestates using Vis, NIR or MIR spectrometry at laboratory scale, all verifying optical spectrometry a general aptness to measure several LOM properties to a certain accuracy (Horf et al., 2022b and references therein). In addition, studies analysing process parameters in biogas plants (e.g. volatile fatty acids), underline this progressive technique for similar substrates (Xu et al., 2023). Within the last few years, additional on-line measurements of LOM at farm and field scale have been successfully tested for TN, ammonium-N (NH₄-N), phosphorous (P), potassium (K), and dry matter (DM) content (Derikx et al., 2021; Saeys et al., 2019a; Williams et al., 2020). In spite of these promising results, there is still a high interest to improve the prediction accuracies, to expand accurate analysis to micronutrients and pH values, and to reduce costs, e.g. by using low cost-spectrometers with smaller wavelength ranges or less resolution.

For these reasons, the main objectives of the present Vis-NIR study were (i) to improve the accuracy of chemometric models to predict the macronutrients TN, NH₄-N, P, K, Mg, calcium (Ca), and sulphur (S), the micronutrients manganese (Mn), iron (Fe), copper (Cu), zinc (Zn), and boron (B), as well as DM and pH in pig, cattle, and digestate LOM, by comparing a large number of preprocessing methods and regression techniques including some new and innovative algorithms, (ii) for reasons of cost reductions, to simulate measurements with low-cost spectrometers using smaller wavelength ranges and lower resolutions, including identifications of sensitive wavelengths for each LOM property, (iii) to compare model performances using either concentrations on wet or on dry weight basis, and (iv) to improve predictions of pH values by using different presentations of the hydronium ion activities. Furthermore, (v) the necessity to develop individual models for different LOM types was evaluated.

2. Materials and methods

2.1. Sample set

A set of 638 samples from liquid livestock manures and biogas digestates were collected from a large number of farms in Northwest Germany between 2018 and 2020. The set included 391 pig manures (227 hog manures, 106 sow manures, 50 piglet manures, and 8 mixed pig manures), 155 cattle manures (70 dairy manures and 85 further cattle manures including bull and young cattle manures), 89 digestates and three additional samples with mixed pig and cattle manures. Each sample was homogenized with a stainless-steel mixer (Blender CB15VXE, Waring Commercial, Torrington, CT, USA) before subsampling for reference analyses and spectral measurements. Samples were stored at -18 °C.

To evaluate the necessity to develop individual models for different LOM types, samples were stratified into six LOM groups: all, pig, cattle, and digestate LOM, a group with exclusively animal manures

(pig+cattle), and a group containing plant residuals (cattle+digestate). It should be noted that the LOM group pig-digestate was not considered due to an imbalanced sample size and less similarities in reflectance spectra.

2.2. Reference analysis

The reference analyses were conducted at LUFA Nord-West (Hameln, Germany), a certified laboratory for slurry analysis using official standard laboratory methods of Germany. The analyses for the main plant nutrients P (commonly expressed as P_2O_5), K (as K_2O), Mg (as MgO), Ca (as CaO), S, as well as the micronutrients B, Mn, Fe, Cu, and Zn are based on acid digestion in a microwave system followed by inductively coupled plasma optical emission spectrometry (ICP-OES; DIN EN ISO DIN e.V., 2009). Nitrogen concentrations (TN and NH_4-N) were determined using the Kjeldahl method (DIN EN DIN e.V., 2012a) and the extraction method based on $CaCl_2$ solution (DIN EN ISO DIN e.V., 2005), respectively. Dry matter content (DM) was determined via drying process (DIN EN DIN e.V., 2001). The laboratory measurement uncertainty concerning the repeatability was between 1.7 and 3.6 % for all analysis parameters. All element concentrations were expressed on a wet and a dry weight basis and both used in chemometric analysis to examine possible effects on prediction accuracy. All pH values were measured with a glass electrode (DIN EN DIN e.V., 2012b).

2.3. Spectral measurements

Spectral measurements were conducted with an ultra-broadband UV-Vis-NIR spectrometer ARCSpectro UV-Vis-NIR FIB (ARCOptix S. A., Neuchatel, Switzerland). It consists of two individual spectrometers: a dispersive spectrometer with a silicon array detector for the UV and Vis range from 200 to 1000 nm, and a Fourier-Transform spectrometer with an InGaAs photo diode for the NIR region from 850 to 2500 nm. The spectral resolution (expressed as full width at half maximum; FWHM) is <5 nm and the reporting interval (RI) is 1 nm, whereby RI means interpolated values of original sampling intervals (SI), which truly differ from 1 nm, dependent from the wavelength range.

About 25 mL of each sample material were filled in small glass vials (2.8 mm diameter, thin borosilicate glass with a planar bottom, Deutsche Metrohm GmbH & Co. KG, Filderstadt, Germany) before spectral

analysis. Immediately before measurements, all liquid samples were intensively stirred. Samples were irradiated by two 50 W halogen lamps at an angle of 45° in a black box to avoid the influence of ambient light (Fig. 1). Diffuse reflected light from the bottom of the sample vials was transmitted to the spectrometer via two glass fibres: one fibre with a high number of OH-bindings for the Vis and one low OH-fibre optimized for the NIR region. Measurements were taken at four different positions by rotating the sample by 90° . At each position, a spectrum was derived by averaging 50 replicates in the Vis region with an integration time of 60 ms and 16 replicates in the NIR region using a medium gain factor. Afterwards, the four obtained spectra were averaged.

As a durable, heat-resistant in-house reflection standard, a dark mat ceramic plate was manufactured and placed in a borosilicate glass vial. This standard was measured every 30 min. The reflectance spectrum of the in-house standard was referenced to a certified 10 % diffuse reflection standard made of PTFE within 250 and 2450 nm (Lake Photonics GmbH, Uhlhingen-Mühlhofen, Germany; traceable to the diffuse reflection standard of the PTB, Braunschweig, Germany). Dark current measurement was carried out on a daily basis.

2.4. Chemometric analysis

2.4.1. Preprocessing methods

As a first step, raw reflection spectra of the slurry samples (I) were converted into reflectance spectra (R) by offsetting with the spectra from the in-house reflection standard (I_0), the conversion factors for each wavelength (z) from the certified standard, and the dark current spectra (I_d) using the following equation:

$$R = \frac{I - I_d \cdot z}{I_0 - I_d} \quad (1)$$

All data from wavelengths within 250 to 324 nm at the lower edge the detector range were removed due to a high noise level. The wavelength data from 2450 to 2500 nm were excluded from the analysis due to a limited range of the certified reflection standard. Next, the spectra were smoothed in three different regions, namely 325–500 nm, 500–1150 nm, and 1150–2450 nm by a Savitzky-Golay filter with a polynomial degree of 1st order using 21 windows, 3rd order using 21 windows, and 1st order using 11 windows, respectively. Subsequently, two preprocessing methods were applied including (i) pseudo-

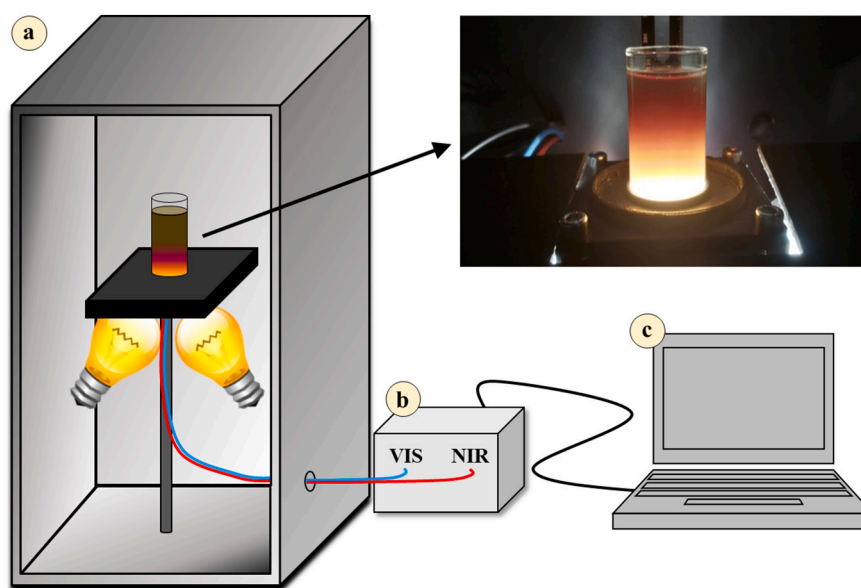


Fig. 1. Setup for spectroscopic measurements of liquid organic manures (LOM). a) Inside a black box, two halogen lamps irradiate an LOM sample filled in a glass vial with planar bottom. Two optic fibres for the visual and near infrared range transmit the reflected light of the slurry sample to b) Vis-NIR spectrometer connected to c) a computer.

absorbance transformation (pA), which is analogous to Lambert-Beer's law for transmitted light, expressed by $\log(1/R)$, and (ii) Kubelka-Munk transformation (KM), expressed by $(1 - R)^2 / 2R$. In both equations, R is the reflectance spectrum.

Reflectance spectra, as well as the transformed pA spectra, and KM spectra were combined with each of the following eight preprocessing methods resulting in 24 different preprocessed spectra for each sample: (i) no preprocessing, (ii) Savitzky-Golay first derivative (SG1; Savitzky and Golay, 1964), (iii) Savitzky-Golay second derivative (SG2; Savitzky and Golay, 1964), (iv) standard normal variate transformation (SNV; Barnes et al., 1989), (v) multiple scatter correction (MSC; Geladi et al., 1985; Martens et al., 1983), (vi) orthogonal signal correction (OSC; Wold et al., 1998), (vii) simple ratios (SR), and (viii) normalized differences (ND; Schirrmann et al., 2013). The last two are both innovative "quotient techniques" inspired from calculating vegetation indices like RVI or NDVI and the quotient mathematics used by Karl Harris, the "father of NIRS", in his first studies with NIRS (Williams, 2019). Within the quotient technique of SR, all possible combinations of ratios between all spectral signals are calculated, resulting in a huge matrix of $(n - 1) * n / 2$ predictor variables (Eq. (2)). Within the second quotient technique ND, all possible differences of two wavelength signals are normalized by their sum, analogue to NDVI (Eq. (3)), resulting in the same number of predictor variables as in SR.

$$SR_{ij} = \frac{y_i}{y_j}; i < j \quad (2)$$

$$ND_{ij} = \frac{y_i - y_j}{y_i + y_j}; i < j \quad (3)$$

Here y_i and y_j are reflectance values at the i^{th} and j^{th} wavelength of a spectrum.

2.4.2. Regression models and validation process

To find the best correlation between the spectral signatures and the 14 dependent variables, nine regression methods were applied to each type of preprocessed spectra and each LOM group: (i) partial least squares regression (PLSR; Wold, 1975; R package "pls"; Mevik and Wherens, 2022), (ii) canonical powered PLSR (CP-PLSR; Indahl et al., 2009; R package "pls"), (iii) a newly developed algorithm called forward stagewise subset selection combined with PLSR (FS-PLSR; Horf et al., 2022a), (iv) random forest (RF; Ho, 1995; Breiman, 2001; R package "randomForest"), least absolute shrinkage and selection operator regression (LASSO; Tibshirani, 1996) calculated with two different R packages (v) "lars" and (vi) "Glmnet" to compare potential deviations, (vii) least angle regression (LAR; Efron et al., 2004), a forward stepwise algorithm with a connection to LASSO (R package "lars"), (viii) ridge regression (Hoerl and Kennard, 1970; R package "Glmnet"), and (ix) elastic net regression (Zou and Hastie, 2005 (R package "Glmnet").

FS-PLSR (Horf et al., 2022a) selects wavelengths by an algorithm that belongs to the family of forward stagewise regressions (Sen and Srivastava, 1990) before using PLSR. The algorithm is inspired by sure independence screening (Fan and Lv, 2014) and involves the following data processing steps: (i) From all predictor variables (wavelengths), the best predictor is selected by the absolute maximum of Kendall's τ correlation between the predictors and the target variable (Fan and Lv, 2014). (ii) A robust linear model is fitted to the data from the selected predictor variable and the target variable. (iii) The residues of this model are used as the new target variable and the best predictor is removed out of the predictor variables. (iv) Steps (i) to (iii) are repeated n times. In this study, n was set to 30 stages. In addition, n could be defined by Kendall's τ of step 1, when surpassing a certain threshold. (v) The set of selected predictors is used to build a PLSR model.

The algorithms of the three regression methods LASSO, ridge and elastic net are very similar and only differ in their penalty term. In contrast to ridge regression, LASSO can shrink coefficients to zero causing an exclusion of predictor variables. Elastic net combines the L1

regulation term of LASSO and the L2 regulation term of ridge regression with a factor α , which was set to 0.5 in this study to weight both terms equally. For predictions using the R package "glmnet", a 10-fold CV was conducted once within the training set. The hyperparameter λ that leads to the minimum RMSEP within the CV was selected for the test set validation. In the R package "lars", the algorithm itself chose the optimal λ . For more detailed information on the shrinkage methods LASSO, LAR, ridge and elastic net algorithms, the reader is referred to Hastie et al., 2009.

For fitting and evaluating the prediction model of each target variable, each LOM group was randomly separated into a training set (75 %) for calibration and an independent test set (25 %) for validation. For splitting, the function *partition()* of the R package *splitTools* was used to balance the partitions as good as possible regarding the distribution of the considered target variable. For PLS modelling, PLS1 was used with classical orthogonal scores, also known as non-iterative partial least squares (NIPALS). PLS1 was preferred to PLS2, as the prediction results are expected to get worse when considering all response variables simultaneously (Bretron et al., 2017; p. 296), which was confirmed in a simple test run. For PLSR, CP-PLSR and FS-PLSR, a 10-fold cross-validation (CV) was conducted within the calibration set. The number of PLSR components having the lowest RMSE in the CV was used for predictions in the test set. Due to the randomness of a 10-fold CV, its iteration may result in slight deviations in the proposed number of PLSR components. Thus, the calibration procedure was repeated 20 times to choose the optimal number of components.

2.4.3. Outliers

Two types of data outliers were removed. "Reference outliers" were identified with a relative conservative criterion. A sample was removed when the reference analysis of a target variable exceeded three times the interquartile range (IQR) from the first or third quartile. "Prediction outliers" were identified as done in Reeves (2001) by exceeding three times the RMSEP of the independent test set validation. After excluding the prediction outliers, a new regression line was calculated; this procedure was repeated once. "Spectral outliers", meaning conspicuous spectra of a dataset, were not found, thus all measured spectra were used for analysis.

2.4.4. Important wavelength ranges and the simulation of low-cost spectrometers

The involved regression methods make use of different variable selection methods to remove adverse effects like noise, and to improve interpretability, modelling speed, and prediction accuracy (Cassotti and Grisoni, 2012). Within these algorithms, spectral wavelengths are either weighted differently or completely removed to get an enhanced prediction result. In PLSR, variable importance in projection (VIP) scores provide insight into the importance or influence of each wavelength on the prediction result, whereby the mean VIP-score of all involved wavelengths is defined to be 1.0. To identify important wavelengths by the shrinkage methods (LASSO, ridge and elastic net regression), regression coefficients calculated for each wavelength were interpreted as importance factor. Coefficients of LAR were also interpreted as an importance factor. RF indicates variable importance by the IncNodePurity index (i.e., the total decrease in node impurities from splitting, averaged over all trees) which is measured in this case by the residual sum of squares.

To assess the effect of excluding potentially unimportant wavelength ranges and to simulate the performance of low-cost spectrometers, the broadband wavelength range of 325–2450 nm and the reporting interval (RI) of 1 nm were reduced to the following six different wavelength ranges and (in three cases) to simulated sampling intervals (SI) of 3 nm: (i) 325–870 nm (RI: 1 nm; exclusively the Vis region), (ii) 860–2450 nm (RI: 1 nm; exclusively the NIR region), (iii) 900–1700 nm (RI: 1 nm; the region of commonly used NIR spectrometers in agriculture), (iv) 975–1700 nm (SI: 3 nm; reduced resolution of low-cost spectrometers;

Rosero-Vlasova et al., 2016; Wolfrum et al., 2020), (v) 325–1700 nm (SI: 3 nm; same as before including the Vis region), and (vi) 1300–2450 nm (SI: 3 nm; the region of strong water absorption bands).

The reduction of the reporting interval to a 3-nm sampling interval (spacing or pixel distance in grating spectrometers) was achieved by calculating the mean of the involved wavelength signals. Whereby, a 3-nm SI often corresponds to a spectral resolution of about 10 nm FWHM (Rosero-Vlasova et al., 2016; Wolfrum et al., 2020). The spectral resolution is always higher than the SI because it needs at least two to three data points to identify a peak and to distinguish between closely related ones; in praxis often three to four points are used.

Each wavelength range was chemometrically analysed using concentrations on wet wt. basis. Additionally, the wavelength ranges of 325–2450 nm (RI: 1 nm) and 975–1700 nm (SI: 3 nm) were analysed using concentrations on dry weight basis to examine possible effects on prediction accuracy and to enable a comparison with past studies, that published results in concentrations on dry weight basis.

2.4.5. Presentations of the hydronium ion activity

In 1924, the pH value was defined as the negative decade logarithm of the hydronium ion activity $a(\text{H}_3\text{O}^+)$, due to very practical reasons. However, this convention of expressing the H_3O^+ activity does not necessarily have to lead to the best results in a chemometric analysis. Thus, it was examined if the hydronium activity itself ($a(\text{H}_3\text{O}^+) = 10^{-\text{pH}}$) would improve prediction results. Additionally, the expression of the pH value using a logarithm with a basis higher or a lower than the standard value of 10, namely 10^6 or e ($2.71 \rightarrow$ natural logarithm) was used for chemometric analysis. Furthermore, the hydronium ion activity was divided by DM (which is equal to nutrient concentrations on dry weight basis).

2.5. Model evaluation

2.5.1. Performance metrics for choosing the best prediction models

The permutation of 24 preprocessing methods with nine regression

algorithms resulted in 216 different predictions for each combination of the 14 target parameters, six LOM groups, seven wavelength ranges and two concentration basis (Fig. 2). Out of these 216 predictions for each case, the best model was selected by evaluating the following performance metrics: squared Pearson coefficient (r^2 ; Eq. (4)), root mean square error of prediction (RMSEP; Eq. (5)), ratio of performance to interquartile range (RPIQ; Bellon-Maurel et al., 2010; Eq. (6)), the mean absolute error (MAE), the MAE normalized by the median of the target variable in the prediction set (nMAE; Eq. (7)), the number of used components as well as outliers, slope, intercept, and bias. The ratio of performance to deviation (RPD), which is often used for elevating regression models in Vis-NIR spectroscopy, was not included since its equivalency to r^2 in case of a normally distributed target variable (Minasny and McBratney, 2013) and its inferiority to RPIQ (in the case of a non-normally distributed target variable; Bellon-Maurel et al., 2010).

$$r^2 = \frac{\left[\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\left[\sum_{i=1}^N (x_i - \bar{x})^2 \right] \left[\sum_{i=1}^N (y_i - \bar{y})^2 \right]} = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (5)$$

$$RPIQ = \frac{Q3 - Q1}{RMSEP} \quad (6)$$

$$nMAE = \frac{MAE}{x_{med}} = \frac{\sum_i |x_i - y_i|}{N * x_{med}} \quad (7)$$

Here N is the number of samples, x_i is the i^{th} reference value, \bar{x} is the mean of the reference values, x_{med} is the median of the reference values, y_i is the i^{th} predicted value, \bar{y} is the mean of the predicted values, \hat{y}_i is the

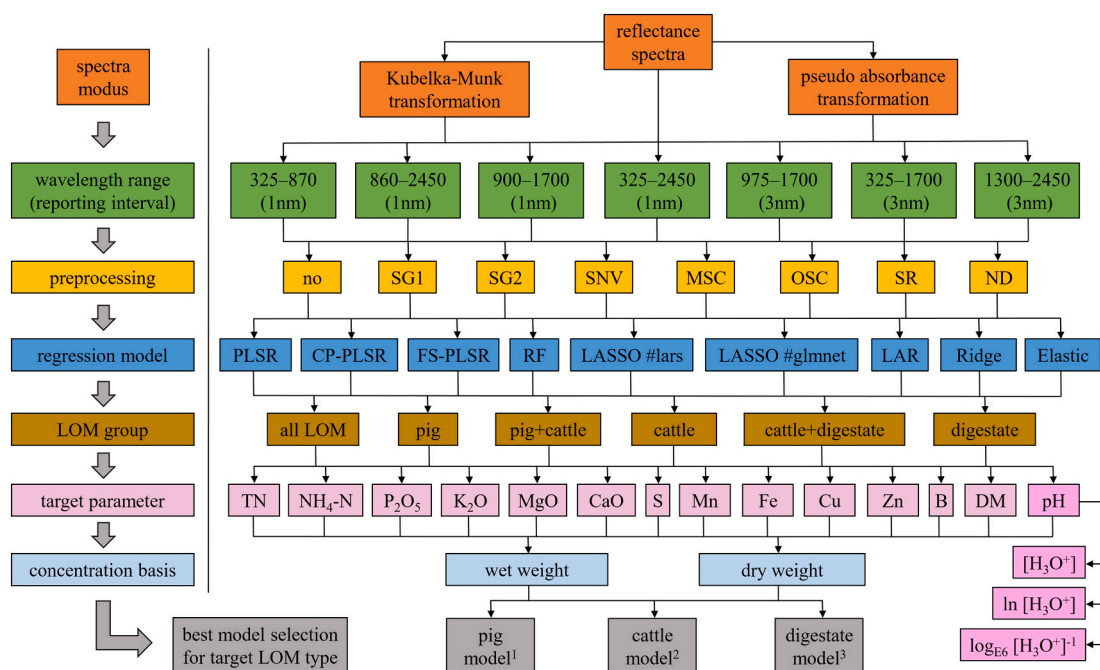


Fig. 2. Flow chart of the study. LOM represents liquid organic manures. For abbreviations of preprocessing methods and regression models see Sections 2.4.1 and 2.4.2, respectively.

¹Selected from the models of the LOM groups pig, pig+cattle, and all LOM.

²Selected from the models of the LOM groups cattle, pig+cattle, cattle+digestate, and all LOM.

³Selected from the models of the LOM groups cattle, cattle+digestate, and all LOM.

i^{th} y-value estimated by the regression equation, Q1 is the first quartile of the reference values, and Q3 is the third quartile of the reference values.

The RPIQ was calculated due to its robustness against non-normal data distributions as it describes the relationship between spread of data and error of prediction. Thus, the higher the RPIQ the better the prediction (Bellon-Maurel et al., 2010).

For the selection of the best models, the precision (random error) of the model was primarily considered. For this reason, r^2 was primarily used. It is independent from the accuracy of the model and thus independent from the slope (systematic scale error) and the intercept (systematic shift error) of the predicted regression line. Furthermore, it is dimensionless and thus enables comparisons between predictions with different units. The further performance metrics RMSEP, RPIQ, and nMAE were bias corrected by slope and intercept to explicitly compare random errors. The reason for focussing on the random error was the assumption that systematic errors can more easily be solved by improving the technical settings and the measurement protocol, whereas random errors can often only be tackled by repeated measurements.

When r^2 was similar among the models, the performance metrics relevant for random errors (bias corrected RMSEP, RPIQ, nMAE) and those relevant for systematic errors (slope, intercept and bias) were considered for choosing the best model. This was also considered when choosing the most appropriate model for each target manure type (pig, cattle, digestate) and target parameter out of the best selected models for each wavelength range, LOM group and concentration basis (see Fig. 2 and Supplementary materials).

2.5.2. Prediction interval

To state the model accuracy for potential usage in practical agricultural, it is proposed to specify a predicted concentration with a 95 % prediction interval, which can be calculated in the two following ways: With the assumption that the residuals of the regression line are normally distributed, the RMSEP can be multiplied by a z-score of 1.96, which represents a two-sided t-test with a large number of samples at the 95 % level. Alternatively, it is proposed to use MAE instead of RMSEP for prediction intervals. As the MAE, in case of normal distributions and zero bias, is mathematically connected to the RMSEP by the Eq. (8), it should lead to similar results.

$$MAE = RMSEP * \sqrt{\frac{2}{\pi}} \quad (8)$$

3. Results and discussion

3.1. Descriptive statistics and correlations among LOM parameters

Table 1 displays several descriptive statistics for the physical and chemical characteristics of all 638 LOM samples utilized in this study. Despite some outliers (e.g. a pH value of 4.0 or a manganese concentration of over 3000 mg/kg wet wt.), the parameters show representative values and concentrations for different animal manure species and

biogas digestates (Waldrip et al., 2020). They cover a large range including a high internal variability with relative standard deviations between 39 and 77 % for all macronutrients, Zn, B, and DM. The micronutrients Mn, Fe and Cu show an extreme dispersion between 114 and 225 %. In contrast, pH values are very constant between 7.3 and 8.3, thus showing a very small dispersion of 5 %. It is noticeable that the concentrations for many nutrients, especially micronutrients, show a right-skewed, log-normal distribution with a higher sample number of lower concentrations, which has to be kept in mind when evaluating performance metrics of regression models. Box-Cox transformations were tested and refused as they deteriorated the prediction results.

In the following figures, the sample set is separated into the three LOM types pig, cattle, and digestate and a distinction is made between concentrations on wet (Fig. 3A) and dry weight basis (Fig. 3B). In general, the concentrations of pig LOM cover a higher value range than cattle or digestate LOM. Furthermore, all dry weight concentrations except Mn in pig LOM are, on average, higher than in cattle or digestate LOM. In contrast, many wet weight concentrations of pig LOM tend to be lower than those of cattle or digestate LOM due to a generally higher water content.

Table 2 lists all correlations between the LOM parameters derived by standard laboratory methods. They differ by LOM type and the concentration based on either wet or dry weight. Internal correlations play an important role for the prediction accuracy of optical less active or inactive parameters, meaning the higher the correlation to optical active parameters, the better the predictions. In this point, it has to be emphasized that target parameters do not necessarily have to be optically active, i.e. that they physically interact with the electromagnetic radiation of the UV, Vis or NIR region. It is also possible to quantify parameters, which do not show optical interactions. This indirect approach is based on their high correlation to at least one optically interactive constituent. In the case of the alkaline metal potassium, the element has no molecule bindings to atoms of non-metals like hydrogen, carbon or oxygen. In LOM, it mainly occurs as K^+ ion in the liquid fraction surrounded by a hydration shell. Only a small part is bound to the organic material as it is easily displaced by ions of higher valence like Ca^{2+} or Mg^{2+} (Harland, 1994). Thus, the prerequisite for a precise prediction of potassium is a correlation to optically active compounds like organic matter or water content. Nevertheless, it is assumed that solved ions like potassium can be indirectly optically active by forming several hydration shells. Such changes in the dynamic of the hydrogen bonding structure may lead to small water band shifts and thus significant changes in the reflectance spectrum (Fornaro et al., 2015).

From Table 2, it can be concluded, that parameters of pig LOM show generally higher correlations than cattle and digestate LOM. Parameters of digestate LOM correlate generally worst, because digestates derive from multiple sources leading to a more varying composition in comparison to animal LOM. The reasons for generally higher correlations in pig than in cattle LOM remain mainly unknown. Only, the strong correlations between P_2O_5 and MgO ($r = 0.97$) or CaO ($r = 0.95$) concentrations (wet wt.) in pig LOM are probably the result of feed additives with magnesium and calcium phosphates as pigs do not produce the enzyme phytase to utilize phosphorous of organic materials mainly

Table 1

Descriptive statistics of laboratory data for DM, pH, and 12 nutrients of all 638 LOM samples; nutrient concentrations were determined by standard chemical methods on wet weight basis. Descriptive statistics grouped by pig, cattle, and digestate LOM are listed in Appendix A.

Parameter ^a	TN	NH ₄ -N	P ₂ O ₅	K ₂ O	MgO	CaO	S	Mn	Fe	Cu	Zn	B	DM	pH
Unit	g/kg							mg/kg					%	-
Min	0.49	0.22	0.05	0.35	0.07	0.18	0.04	0.74	0.32	0.21	0.38	0.21	0.30	4.00
Max	11.10	7.89	6.63	9.95	3.88	7.38	1.52	3156.00	758.00	74.50	141.00	27.00	13.09	8.70
Average	3.94	2.26	1.65	3.57	0.90	1.83	0.39	83.73	72.10	8.42	28.16	2.35	4.95	7.77
Standard deviation (SD)	1.55	0.92	1.17	1.55	0.59	1.18	0.21	188.37	93.84	9.57	21.51	1.82	3.04	0.39
Relative SD (RSD; %)	39	41	71	43	65	65	53	225	130	114	76	77	61	5

^a The laboratory repeatability was between 1.7 and 3.6 % for the selected analysis parameters.

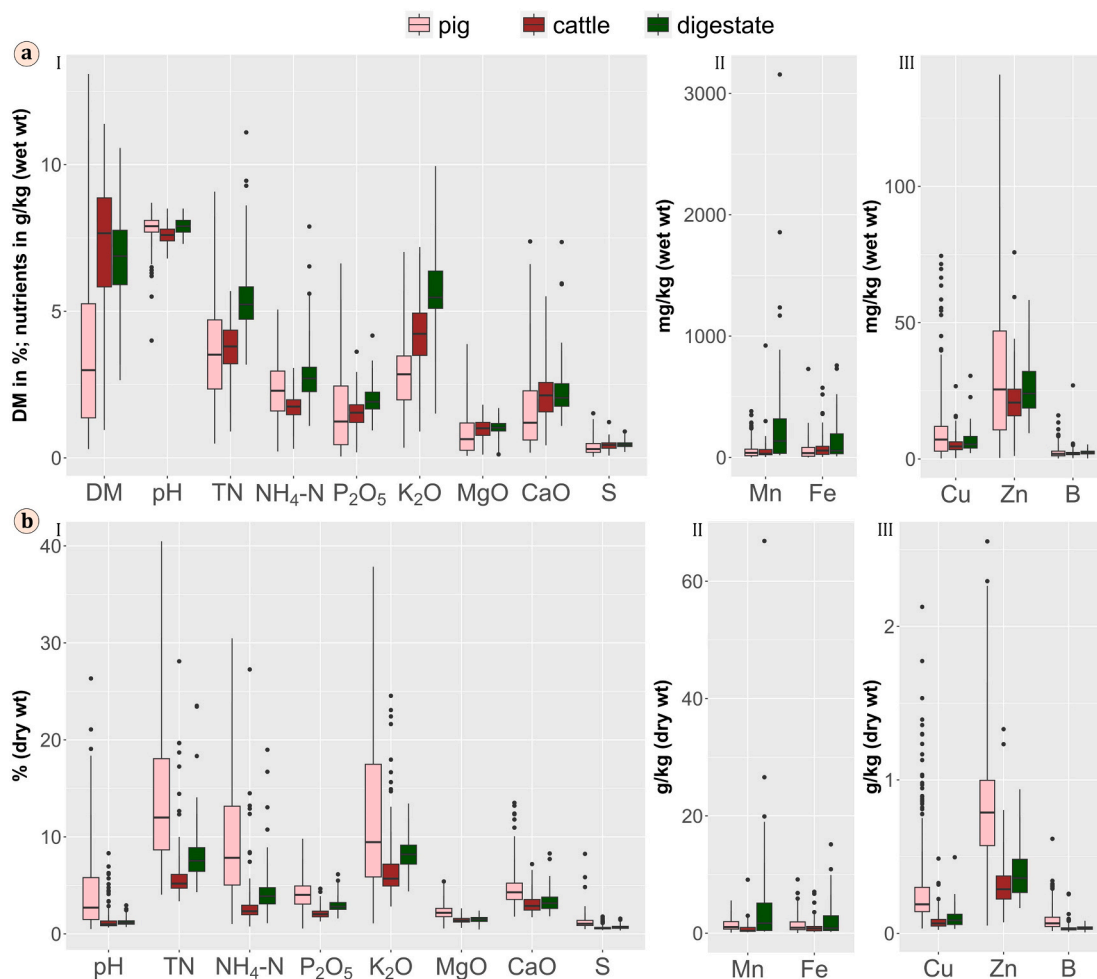


Fig. 3. Box-and-whisker plots of laboratory data for DM, pH, and concentrations of 12 nutrients in a) wet weight and b) dry weight grouped by the LOM types. Plots include interquartile range (IQR) from first and third quartile, median, minimum and maximum, and whiskers with a maximum length of 3*IQR for identifying outliers; pH values in b) were divided by dry matter content, although there is no scientific practice for such a ratio. The reason for nevertheless displaying them on a dry weight basis was improvements in prediction accuracy. Three subfigures A, B, and C are shown due to scaling reasons.

bound in phytate (Malley et al., 2002; Saeys et al., 2005).

Concerning the target parameters, DM correlates most with the other parameters. E.g. in pig and cattle LOM, DM is highly correlated with the macronutrients TN, P₂O₅, K₂O, MgO, CaO, and S with Pearson coefficients (*r*) from 0.65 to 0.93. In pig LOM, DM is additionally correlated with NH₄-N, Mn, Fe, and Zn with *r* from 0.65 to 0.85. The highest correlations can be found between the nitrogen parameters TN and NH₄-N in all LOM types for wet and even stronger for dry wt. concentrations with *r* between 0.82–0.92 and 0.98–0.99, respectively.

3.2. Spectra

Fig. 4 gives an overview about the shape and distribution of all measured reflectance spectra regarding the three sample sets of pig, cattle and digestate LOM. In general, they do not differ much, especially in the region of high water absorption above 1400 nm with its two strong water absorption bands around 1450 nm and 1910 nm resulting in <10 % reflectance. In contrast, in the region of two less pronounced water absorption bands around 975 nm and 1175 nm, the intensities of digestate LOM spectra vary more than those of cattle and pigs, resulting in a larger green area between the Q1 and Q3 quartiles (Fig. 4C). Furthermore, digestate spectra may absorb radiation in that region almost completely. The average intensity of pig spectra in that region is lower due to lower concentrations of solid particles. Thus, the maximum reflectance for pig spectra was only 22 %, whereas for cattle and

digestates, it was much higher with 35 % and 33 %, respectively. However, the main characteristic to distinguish pig from cattle and digestate spectra is the missing absorption band in the visible region around 675 nm, which can be traced back to chlorophyll residues, as green forages are rarely used in pig feed (Saeys et al., 2005). For this reason, it is possible to a certain extent to classify the LOM type by a statistic model analysis (Friedrich et al., 2022).

3.3. Chemometric results

3.3.1. Model evaluations

Tables 3 and 4 give a detailed overview over the 42 best prediction models for all 14 target parameters of pig, cattle, and digestate LOM, validated on an independent test set. Additionally, one regression for each target parameter is exemplarily visualized in Figs. 5 and 8 (Section 3.4). Every regression model is described by the number of samples, the wavelength range, the preprocessing and regression method, and the performance metrics describing the random error (*r*² and the bias corrected RPIQ, RMSEP, MAE, and nMAE), the linear regression parameters describing the systematic error (slope, intercept and bias), as well as the number of components and outliers.

The performance metrics generally correlated with *r*². Regarding nMAE, it was found to be a good alternative to the RMSEP, as nMAE weights every sample equally, whereas RMSEP strengthens higher and weakens smaller values <1.0. Furthermore, nMAE is more robust to non-

Table 2

Pearson coefficients (r) of correlations between DM, pH and concentrations of 12 in a) pig, b) cattle and c) digestate LOM samples, determined by standard lab procedures. Correlations on wet weight basis are displayed in the lower triangle highlighted *italic* and *blue*, on dry weight basis in the upper triangle. Correlation coefficients ≥ 0.65 are underlined.¹ Correlations that do not distinguish between the LOM types, thus using all 638 LOM samples, are displayed in [Appendix A](#).

a - pig															
wet \ dry wt.	DM	pH	TN	NH ₄ -N	P ₂ O ₅	K ₂ O	MgO	CaO	S	Mn	Fe	Cu	Zn	B	
DM		<i>-0.70</i>	<i>-0.74</i>	<i>-0.72</i>	0.39	<i>-0.73</i>	0.11	-0.22	-0.43	0.10	0.23	-0.05	0.15	-0.45	
pH	<i>-0.29</i>		<u>0.79</u>	<u>0.77</u>	-0.43	<u>0.74</u>	0.00	0.41	0.45	-0.17	-0.33	-0.10	-0.38	0.50	
TN	<i>0.85</i>	<i>-0.18</i>		<u>0.99</u>	-0.54	<u>0.90</u>	-0.29	0.12	0.58	-0.27	-0.38	-0.19	-0.47	0.42	
NH ₄ -N	<i>0.64</i>	<i>-0.04</i>	<i>0.92</i>		-0.55	<u>0.90</u>	-0.30	0.10	0.57	-0.27	-0.41	-0.22	-0.50	0.39	
P ₂ O ₅	<i>0.94</i>	<i>-0.16</i>	<u>0.78</u>	<i>0.57</i>		<i>-0.65</i>	0.61	0.45	-0.31	0.21	0.59	0.13	<u>0.69</u>	-0.22	
K ₂ O	<i>0.57</i>	<i>-0.04</i>	<u>0.80</u>	<u>0.86</u>	<i>0.50</i>		-0.39	-0.01	0.55	-0.25	-0.46	-0.20	-0.53	0.39	
MgO	<i>0.92</i>	<i>-0.15</i>	<u>0.76</u>	<i>0.56</i>	<u>0.97</u>	<i>0.48</i>		0.56	-0.24	0.24	0.33	0.18	0.46	0.09	
CaO	<i>0.89</i>	<i>-0.17</i>	<u>0.71</u>	<i>0.49</i>	<u>0.95</u>	<i>0.41</i>	<u>0.92</u>		0.07	0.09	0.33	0.19	0.36	0.16	
S	<i>0.86</i>	<i>-0.20</i>	<u>0.81</u>	<u>0.66</u>	<u>0.84</u>	<u>0.65</u>	<u>0.81</u>	<u>0.82</u>		-0.16	-0.16	-0.06	-0.26	0.16	
Mn	<i>0.73</i>	<i>-0.21</i>	<i>0.51</i>	<i>0.31</i>	<u>0.76</u>	<i>0.27</i>	<u>0.79</u>	<u>0.73</u>	<i>0.62</i>		-0.16	0.22	0.41	0.01	
Fe	<i>0.66</i>	<i>-0.06</i>	<i>0.57</i>	<i>0.41</i>	<u>0.69</u>	<i>0.32</i>	<u>0.66</u>	<u>0.68</u>	<i>0.60</i>	<i>0.28</i>		0.27	0.56	-0.24	
Cu	<i>0.48</i>	<i>-0.27</i>	<i>0.35</i>	<i>0.15</i>	<i>0.47</i>	<i>0.15</i>	<i>0.48</i>	<i>0.51</i>	<i>0.46</i>	<i>0.51</i>	<i>0.41</i>		0.57	0.15	
Zn	<i>0.86</i>	<i>-0.18</i>	<u>0.68</u>	<i>0.44</i>	<u>0.92</u>	<i>0.36</i>	<u>0.91</u>	<u>0.89</u>	<u>0.74</u>	<u>0.81</u>	<u>0.65</u>	<i>0.64</i>		-0.11	
B	<i>0.50</i>	<i>-0.12</i>	<i>0.45</i>	<i>0.31</i>	<i>0.53</i>	<i>0.28</i>	<i>0.52</i>	<i>0.52</i>	<i>0.46</i>	<i>0.48</i>	<i>0.34</i>	<i>0.53</i>	<i>0.55</i>		
b - cattle															
wet \ dry wt.	DM	pH	TN	NH ₄ -N	P ₂ O ₅	K ₂ O	MgO	CaO	S	Mn	Fe	Cu	Zn	B	
DM		<i>-0.81</i>	<i>-0.69</i>	-0.6	-0.18	<i>-0.71</i>	-0.44	-0.33	-0.58	0.08	0.11	-0.46	-0.49	-0.3	
pH	<i>-0.42</i>		<u>0.88</u>	<u>0.83</u>	0.15	<u>0.83</u>	0.38	0.32	<u>0.71</u>	0.04	-0.12	0.42	0.51	0.48	
TN	<i>0.77</i>	<i>-0.16</i>		<u>0.98</u>	0.32	<u>0.81</u>	0.4	0.28	<u>0.79</u>	0.04	-0.09	0.49	0.61	0.57	
NH ₄ -N	<i>0.35</i>	<i>0.09</i>	<u>0.82</u>		0.32	<u>0.77</u>	0.33	0.25	<u>0.75</u>	0.06	-0.11	0.42	0.53	0.61	
P ₂ O ₅	<i>0.78</i>	<i>-0.23</i>	<u>0.82</u>	<i>0.53</i>		0.08	0.29	0.23	0.3	-0.04	0.06	0.19	0.32	0.33	
K ₂ O	<i>0.63</i>	<i>0.05</i>	<u>0.70</u>	<i>0.54</i>	<i>0.53</i>		0.34	0.18	<u>0.65</u>	0.03	-0.13	0.37	0.42	0.38	
MgO	<i>0.83</i>	<i>-0.31</i>	<u>0.78</u>	<i>0.47</i>	<u>0.78</u>	<i>0.58</i>		0.46	0.4	0.17	0.07	0.52	0.63	0.11	
CaO	<i>0.66</i>	<i>-0.34</i>	<i>0.52</i>	<i>0.26</i>	<i>0.56</i>	<i>0.32</i>	<u>0.71</u>		0.39	0.11	-0.05	0.33	0.32	0.23	
S	<i>0.76</i>	<i>-0.34</i>	<u>0.76</u>	<i>0.46</i>	<u>0.76</u>	<i>0.52</i>	<u>0.74</u>	<i>0.64</i>		0.04	-0.06	0.4	0.48	0.48	
Mn	<i>0.24</i>	<i>-0.14</i>	<i>0.10</i>	<i>0.05</i>	<i>0.13</i>	<i>0.11</i>	<i>0.37</i>	<i>0.25</i>	<i>0.17</i>		-0.18	0.06	0.07	0.11	
Fe	<i>0.38</i>	<i>-0.14</i>	<i>0.30</i>	<i>0.11</i>	<i>0.34</i>	<i>0.23</i>	<i>0.33</i>	<i>0.19</i>	<i>0.26</i>	<i>-0.09</i>		-0.01	0.03	-0.19	
Cu	<i>0.24</i>	<i>-0.13</i>	<i>0.27</i>	<i>0.19</i>	<i>0.29</i>	<i>0.13</i>	<i>0.41</i>	<i>0.36</i>	<i>0.23</i>	<i>0.06</i>	<i>0.15</i>		<u>0.76</u>	0.23	
Zn	<i>0.47</i>	<i>-0.05</i>	<i>0.54</i>	<i>0.37</i>	<i>0.61</i>	<i>0.31</i>	<i>0.63</i>	<i>0.47</i>	<i>0.44</i>	<i>0.09</i>	<i>0.29</i>	<i>0.61</i>		0.26	
B	<i>0.29</i>	<i>-0.14</i>	<i>0.30</i>	<i>0.16</i>	<i>0.35</i>	<i>0.15</i>	<i>0.27</i>	<i>0.29</i>	<i>0.35</i>	<i>0.13</i>	<i>-0.06</i>	<i>0.07</i>	<i>0.14</i>		
c - digestate															
wet \ dry wt.	DM	pH	TN	NH ₄ -N	P ₂ O ₅	K ₂ O	MgO	CaO	S	Mn	Fe	Cu	Zn	B	
DM		<i>-0.91</i>	-0.60	-0.60	-0.39	-0.49	0.02	-0.09	-0.59	-0.25	-0.20	-0.23	-0.39	-0.15	
pH	<i>-0.49</i>		<u>0.73</u>	<u>0.75</u>	0.48	0.38	-0.25	0.15	<u>0.67</u>	0.31	0.22	0.22	0.39	0.07	
TN	<i>0.15</i>	<i>0.54</i>		<u>0.98</u>	<u>0.76</u>	0.31	-0.31	0.17	<u>0.81</u>	0.48	0.43	0.30	0.57	0.13	
NH ₄ -N	<i>-0.19</i>	<i>0.63</i>	<u>0.89</u>		<u>0.71</u>	0.25	-0.34	0.20	<u>0.80</u>	0.50	0.38	0.28	0.52	0.07	
P ₂ O ₅	<i>0.57</i>	<i>0.12</i>	<i>0.63</i>	<i>0.37</i>		0.09	-0.02	0.25	0.58	0.30	0.46	0.33	0.62	0.29	
K ₂ O	<i>0.48</i>	<i>0.03</i>	<i>0.33</i>	<i>0.10</i>	<i>0.25</i>		0.00	0.03	0.41	0.09	0.00	0.22	0.38	0.37	
MgO	<i>0.73</i>	<i>-0.32</i>	<i>0.04</i>	<i>-0.23</i>	<i>0.52</i>	<i>0.34</i>		0.26	-0.17	-0.33	0.05	0.27	0.17	0.40	
CaO	<i>0.52</i>	<i>-0.13</i>	<i>0.22</i>	<i>0.09</i>	<i>0.51</i>	<i>0.22</i>	<i>0.59</i>		0.44	0.08	0.08	0.27	0.37	0.44	
S	<i>0.30</i>	<i>0.24</i>	<u>0.66</u>	<i>0.54</i>	<i>0.52</i>	<i>0.44</i>	<i>0.30</i>	<i>0.55</i>		0.54	0.24	0.34	0.58	0.31	
Mn	<i>-0.11</i>	<i>0.13</i>	<i>0.27</i>	<i>0.37</i>	<i>0.00</i>	<i>-0.02</i>	<i>-0.21</i>	<i>0.00</i>	<i>0.36</i>		-0.26	0.00	-0.01	0.00	
Fe	<i>-0.02</i>	<i>0.36</i>	<i>0.41</i>	<i>0.34</i>	<i>0.34</i>	<i>-0.01</i>	<i>0.04</i>	<i>0.10</i>	<i>0.20</i>	<i>-0.33</i>		0.27	0.48	0.07	
Cu	<i>0.15</i>	<i>0.21</i>	<i>0.29</i>	<i>0.22</i>	<i>0.35</i>	<i>0.16</i>	<i>0.32</i>	<i>0.29</i>	<i>0.35</i>	<i>-0.02</i>	<i>0.23</i>		<u>0.72</u>	0.57	
Zn	<i>0.19</i>	<i>0.38</i>	<i>0.57</i>	<i>0.44</i>	<i>0.62</i>	<i>0.28</i>	<i>0.31</i>	<i>0.43</i>	<i>0.55</i>	<i>-0.13</i>	<i>0.46</i>	<u>0.70</u>		0.64	
B	<i>0.45</i>	<i>0.10</i>	<i>0.35</i>	<i>0.14</i>	<i>0.60</i>	<i>0.41</i>	<i>0.56</i>	<i>0.60</i>	<i>0.55</i>	<i>-0.04</i>	<i>0.14</i>	<i>0.58</i>	<u>0.70</u>		

¹p-Values ≤ 0.01 for underlined values, indicating highly significance linear relationships.

normal distributions (similar to RPIQ) as it uses the median as normalizing factor. Additionally, it allows comparing models with different units. Its expression as percentage deviation gives a simple descriptive overview. In contrast to the RMSEP, it is not required to know the unit and distribution of the target variable for evaluating the model performance. For this reason, it was decided to use nMAE as performance metric, too.

In general, good prediction results could be achieved for all target parameters (except boron) of pig, cattle, and digestate LOM. In more detail, these 39 predictions resulted in r^2 values of 0.80–0.97 (mean 0.91), RPIQ values of 2.1–7.8 (mean 4.2), and nMAE values of 5–36 % (mean 15 %), thus outperforming several results published in a comparable context (listed in [Horf et al., 2022b](#)).

The low prediction accuracy for boron ($r^2 = 0.55$ –0.69) confirmed the worse results published in [Malley et al. \(2001\)](#) for boron in hog LOM

($r^2 = 0.76$). At least, these models can be used to distinguish between high and low boron concentrations. The predictions for the remaining target parameters of pig LOM in the present study ($r^2 = 0.86$ –0.97) were comparable to the outstanding results of [Malley et al. \(2001\)](#) ($r^2 = 0.92$ –0.99). Although the transfection mode used by [Malley et al. \(2001\)](#) is slightly more accurate than the reflectance mode, it is less suitable for on-line measurements, e.g. when filling a slurry tanker, as not all slurry particles will pass the required slit of 2 mm ([Saeys et al., 2005](#)). In this context of on-line reflectance spectrometry, the results obtained by different authors for DM and the main macronutrients TN, NH₄-N, P₂O₅ and K₂O in hog LOM ($r^2 = 0.90$ –0.96; [Williams et al., 2020](#)) and further LOM types ($r^2 = 0.86$ –0.96; [Zimmermann and Hartung, 2009](#)) are promising for such applications. Especially the results of the latter publication can be considered as very reliable, as their models were validated with >300 on-line measurements.

The performance metrics describing the systematic error (**slope**, **intercept**, and **bias**) were evaluated as follows: The slope was good in most of the cases lying between 0.83 and 1.08 for about 80 % of all 42 models. However, in some cases, the systematic scale shift was high (e.g. with a slope of 0.51 for MgO in digestate). The reason for such a high systematic error is not clear, as both the reference analysis and the optical measurements of the independent test set were done in the same way as in the training set and all samples for the test set were randomly selected. The intercept was in the range of $\pm 1 \cdot \text{RMSEP}$ for about 70 % of the models and the regression bias was on average 22 % of the RMSEP. However, it has to be kept in mind, that the bias is the sum of differences between all predicted and observed values, meaning that it cannot distinguish between random errors (constant and proportional) and systematic errors (shift bias and scale bias). Furthermore, a positive bias for small values and a negative bias for large values would lead to an unnoticed bias of the whole regression, as an equal positive and negative bias would in total be neutralized to zero.

In this study, two kinds of **outliers** were determined. The number of reference outliers in the reference datasets of the 42 best prediction models was between 0 and 9 % with a mean of 1 %, and the number of prediction outliers in the test set between 0 and 9 % with a mean of 3 %, which was evaluated as an acceptable percentage.

3.3.2. Preprocessing methods

The choice of **spectra mode** (R, pA, KM) did not influence the prediction results significantly. However, converting reflectance spectra (13 \times) into pseudo absorbance (14 \times) or Kubelka-Munk spectra (15 \times) did slightly improve prediction results in a similar number of cases x. However, in [Saeys et al., 2019a](#) a KM or pA transformation of apple reflectance spectra did not improve any prediction performance.

The innovative **preprocessing methods** simple ratios SR (13 \times) and normalized differences ND (2 \times) turned both out to be very beneficial, especially for TN, NH₄-N, Mg, Cu, Zn, and DM predictions, but also for P, S, and Mn. They were outperforming classical methods in more than one third of the 42 best prediction models. It was also observed that, in most cases, SR was superior to ND. This was also noticed by [Tavakoli et al., 2023](#) when calibrating soil spectra. As ND and SR take into account every possible combination of two selected wavelengths, they increase the number of n predictor variables by a factor of $n^2/2$. However, both preprocessing methods require considerable computing time. For this reason, the binning of every 10 wavelengths was observed to be a useful compromise to accelerate the computing speed without losing significant information. Other useful preprocessing methods were SG1 (8 \times), SNV (7 \times), or OSC (5 \times), sometimes SG2 (3 \times) or MSC (2 \times). In only one case, it was the best option to apply no preprocessing method. [Chen et al. \(2010\)](#) examined several preprocessing methods on spectra of fresh poultry manure and combined SNV, MSC, SG1, OSC, or no preprocessing with PLSR. The observed superiority of the OSC method to predict TN, NH₄-N, P₂O₅ and K₂O is not confirmed by the present study, neither for pig, cattle, nor for digestate LOM.

3.3.3. Regression methods

As **regression methods**, LAR (14 \times), LASSO (13 \times), and PLSR (9 \times) predominantly performed best among the 42 best prediction models. Random Forest (4 \times) and FS-PLSR (2 \times) sometimes occurred to be superior. No advantage was achieved with canonical powered (CP)-PLSR as it led either to identical results as ordinary PLSR or, most of the time, to poorer results while choosing a different number of components as PLSR. In addition, the performances of ridge regression, elastic net, and LASSO regression from “glmnet” R – package were not convincing. When comparing these results with similar literature, [Cobbinah et al. \(2022\)](#) is the one of interest, as they tested several regression methods for NIRS analysis of a mixed set of fresh cattle and poultry manure. They obtained best results with stacked regression leading to r^2 values between 0.74 and 0.97 for DM and several macronutrients. Similar to the present study, ridge regression was always inferior to PLSR. However, in

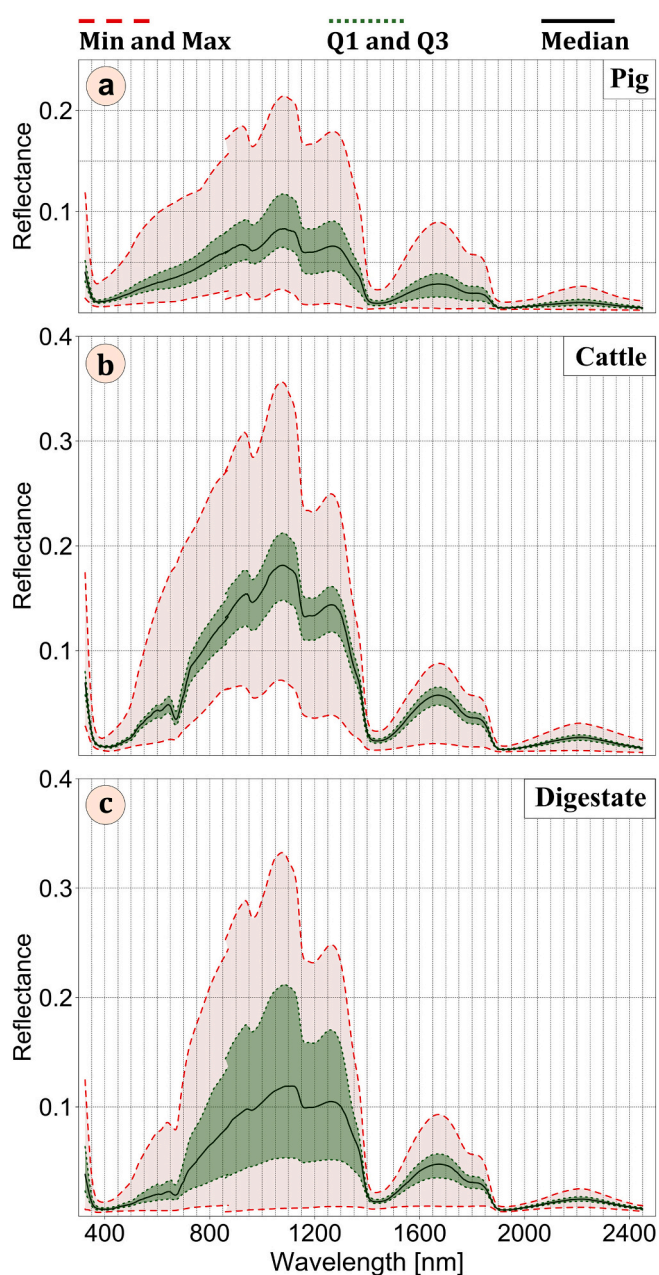


Fig. 4. Summary of reflectance spectra including median, Q1, and Q3 quartiles as well as the minimum and maximum reflectance values of a) pig, b) cattle and c) digestate LOM sample sets. The small shifts around 860 nm are due to spectrometer transitions.

contrast to the present study, elastic net and LASSO regression performed always better and RF often better than PLSR. The reasons for these different results remain unknown; maybe poultry manure has another influence than pig, cattle and digestate.

3.3.4. Important wavelength ranges and the simulation of low-cost spectrometers

To evaluate the exclusion of potentially irrelevant wavelength ranges and to simulate the performance of low-cost spectrometers, the original spectra obtained within a **wavelength range** of 325–2450 nm and a wavelength reporting interval of 1 nm were degraded to 325–870 nm (RI: 1 nm), 860–2450 nm (RI: 1 nm), 900–1700 nm (RI: 1 nm), 975–1700 nm (SI: 3 nm), 325–1700 nm (SI: 3 nm), and 1300–2450 nm (SI: 3 nm). As it needs at least two to three signals to identify a peak, the

Table 3

Best models in the independent test set (25 % of the samples) for **macronutrients** in pig, cattle, and digestate LOM; lowest model performances are highlighted in red, conc. on dry wt. basis in orange letters.

Target variable	Target LOM type	r ²	LOM group	Nr. of samples	Range of wl [nm]	Spectra modus	Prepro- cessing	Regression method	Unit	RMSEP [unit]	MAE [unit]	nMAE [%]	RPIQ	Nr. of comp	Slope	Intercept [unit]	Bias [unit]	Outlier
Total N	pig	0.94	1	638	325–2450	R	ND	LASSO	g/kg wet wt.	0.38	0.30	7.8	5.2	324	0.92	0.36	0.05	5p + 0r
	cattle	0.93	3	155	860–2450	pA	SG1	LAR	g/kg wet wt.	0.27	0.21	5.6	4.1	92	0.93	0.25	-0.03	1p + 0r
	digestate	0.96	6	244	325–2450	pA	SR	LAR	g/kg wet wt.	0.32	0.25	5.9	5.3	90	1.03	-0.17	-0.05	3p + 1r
NH₄-N	pig	0.94	5	546	1300–2450	KM	SR	LASSO	g/kg wet wt.	0.23	0.18	8.9	5.2	509	0.96	0.10	0.01	5p + 0r
	cattle	0.92	3	155	1300–2450	pA	SG1	LASSO	g/kg wet wt.	0.12	0.09	5.3	3.9	225	0.93	0.11	-0.01	2p + 0r
	digestate	0.96	6	244	1300–2450	R	SR	LAR	g/kg wet wt.	0.17	0.14	7.1	6.0	76	0.90	0.17	-0.04	2p + 5r
P₂O₅	pig	0.95	5	546	1300–2450	pA	ND	LASSO	g/kg wet wt.	0.29	0.23	16.3	5.1	762	0.99	0.02	-0.01	3p + 0r
	cattle	0.89	3	155	975–1700	pA	OSC	LAR	g/kg wet wt.	0.19	0.17	10.8	3.2	13	0.71	0.44	0.02	1p + 1r
	digestate	0.89	6	244	325–2450	pA	SG1	LASSO	g/kg wet wt.	0.21	0.17	9.7	3.3	304	0.89	0.10	-0.09	2p + 0r
K₂O	pig	0.95	5	546	325–2450	R	SG1	RF	g/kg dry wt.	15.64	11.78	15.5	5.0	100	0.96	6.07	2.35	7p + 1r
	cattle	0.85	6	244	325–2450	KM	OSC	PLSR	g/kg dry wt.	8.46	6.66	10.0	3.7	14	0.85	13.09	2.66	1p + 5r
	digestate	0.85	6	244	325–2450	KM	OSC	PLSR	g/kg dry wt.	8.46	6.66	10.0	3.7	14	0.85	13.09	2.66	1p + 5r
MgO	pig	0.93	2	391	325–2450	KM	SG2	RF	g/kg wet wt.	0.18	0.14	22.8	4.9	100	0.96	0.10	0.07	4p + 0r
	cattle	0.90	3	155	325–2450	KM	SR	PLSR	g/kg wet wt.	0.11	0.09	8.6	3.5	11	0.83	0.18	0.02	2p + 0r
	digestate	0.85	4	89	975–1700	pA	SR	FS-PLSR	g/kg wet wt.	0.12	0.10	9.2	2.2	5	0.51	0.56	0.06	2p + 1r
CaO	pig	0.93	2	391	325–2450	KM	SG2	RF	g/kg wet wt.	0.27	0.21	18.1	5.4	100	0.92	0.14	0.02	3p + 1r
	cattle	0.91	5	546	325–2450	R	SG1	PLSR	g/kg wet wt.	0.36	0.27	19.5	4.4	21	0.95	0.12	0.02	4p + 1r
	digestate	0.82	1	638	325–2450	R	no	PLSR	g/kg wet wt.	0.55	0.43	27.1	2.6	27	0.76	0.33	-0.10	6p + 2r
S	pig	0.86	2	391	325–1700	KM	SR	LASSO	g/kg wet wt.	0.09	0.07	24.0	3.2	264	0.91	0.03	0.00	2p + 1r
	cattle	0.80	3	155	975–1700	pA	SNV	LAR	g/kg wet wt.	0.07	0.05	12.2	2.9	24	0.66	0.16	0.02	0p + 1r
	digestate	0.92	4	89	325–1700	pA	SG1	LAR	g/kg wet wt.	0.04	0.03	7.4	3.2	33	0.68	0.12	-0.03	1p + 1r

RMSEP: root mean square error of prediction; MAE: mean absolute error; nMAE: mean absolute error normalized by the median; RPIQ: ratio of performance to interquartile distance; comp: components (latent variables); outlier p: prediction outliers in the test set; outlier r: reference outliers in the reference data set; LOM group: 1 = all LOM, 2 = pig, 3 = cattle, 4 = digestate, 5 = pig+cattle, 6 = cattle+digestate.

Table 4

Best prediction models in the independent test set (25 % of the samples) for **micronutrients, DM, and pH** in pig, cattle, and digestate LOM; lowest model performances are highlighted in red, concentrations on dry wt. basis in orange letters.

Target variable	Target LOM type	r ²	LOM group	Nr. of samples	Range of wl [nm]	Spectra modus	Prepro- cessing	Regression method	Unit	RMSEP [unit]	MAE [unit]	nMAE [%]	RPIQ	Nr. of comp	Slope	Intercept [unit]	Bias [unit]	Outlier
Mn	pig	0.90	2	391	325–2450	pA	SNV	LASSO	mg/kg dry wt.	268.75	210.7	19.6	4.3	587	0.90	129.90	-12.1	0p + 0r
	cattle	0.94	3	155	975–1700	KM	OSC	PLSR	mg/kg dry wt.	145.56	113.9	26.1	4.5	12	1.05	-17.36	17.7	0p + 2r
	digestate	0.93	4	89	975–1700	R	SNV	LASSO	mg/kg wet wt.	58.76	44.93	7.2	3.5	385	1.16	-4.18	23.1	1p + 4r
Fe	pig	0.94	2	391	325–2450	R	SNV	PLSR	mg/kg wet wt.	12.15	9.26	29.8	5.2	23	0.95	6.96	4.71	4p + 1r
	cattle	0.92	3	155	325–1700	R	SG2	RF	mg/kg wet wt.	10.33	7.91	13.4	5.7	100	1.06	5.32	8.98	0p + 4r
	digestate	0.85	1	244	975–1700	R	MSC	LAR	mg/kg wet wt.	23.56	18.23	36.0	2.7	32	0.92	7.97	2.03	2p + 15r
Cu	pig	0.87	1	638	325–2450	R	SR	PLSR	mg/kg wet wt.	1.98	1.51	28.1	3.0	35	0.89	0.88	0.15	8p + 23r
	cattle	0.88	3	155	325–2450	pA	SNV	LASSO	mg/kg wet wt.	0.91	0.76	16.2	2.7	395	0.58	1.99	-0.16	1p + 3r
	digestate	0.87	1	638	325–2450	R	SR	PLSR	mg/kg wet wt.	1.98	1.51	28.1	3.0	35	0.89	0.88	0.15	8p + 23r
Zn	pig	0.90	1	638	325–2450	KM	SR	LAR	mg/kg wet wt.	6.04	4.83	21.4	3.0	76	0.84	3.20	-0.86	6p + 8r
	cattle	0.90	1	638	325–2450	KM	SR	LAR	mg/kg wet wt.	6.04	4.83	21.4	3.0	76	0.84	3.20	-0.86	6p + 8r
	digestate	0.80	6	244	325–1700	KM	SNV	LASSO	mg/kg wet wt.	5.61	4.75	21.0	2.1	491	0.75	5.69	-0.30	0p + 1r
B	pig	0.62	1	638	325–1700	pA	SNV	LAR	mg/kg wet wt.	0.78	0.63	31.2	1.7	181	0.75	0.55	0.02	2p + 14r
	cattle	0.55	3	155	325–2450	R	OSC	FS-PLSR	mg/kg wet wt.	0.55	0.42	22.1	1.2	5	0.59	1.06	0.27	1p + 5r
	digestate	0.69	6	244	325–2450	R	SR	LASSO	mg/kg wet wt.	0.58	0.46	23.0	1.7	751	1.08	-0.07	0.11	3p + 2r
DM	pig	0.97	1	391	1300–2450	KM	SR	LAR	%	0.47	0.39	13.4	7.8	72	0.90	0.35	-0.02	5p + 0r
	cattle	0.96	3	155	325–1700	KM	no	LAR	%	0.48	0.37	4.8	6.6	36	0.92	0.58	0.03	2p + 0r
	digestate	0.93	6	244	975–1700	pA	SR	LAR	%	0.60	0.48	6.6	5.0	33	1.02	-0.43	-0.30	2p + 0r
pH	pig	0.96	2	391	325–2450	pA	MSC	LASSO	ln(a(H ₃ O ⁺))-1 dry wt.	0.00	0.00	19.2	6.3	335	0.93	0.00	0.00	6p + 4r
	cattle	0.95	3	155	325–2450	KM	SG1	LAR	dry wt.	0.08	0.06	6.4	4.7	44	0.93	0.05	-0.03	0p + 14r
	digestate	0.96	6	244	975–1700	KM	SG1	LASSO	dry wt.	0.08	0.06	6.0	4.6	29	0.86	0.19	0.03	4p + 16r

RMSEP: root mean square error of prediction; MAE: mean absolute error; nMAE: mean absolute error normalized by the median; RPIQ: ratio of performance to interquartile distance; comp: components (latent variables); outlier p: prediction outliers in the test set; outlier r: reference outliers in the reference data set; LOM group: 1 = all LOM, 2 = pig, 3 = cattle, 4 = digestate, 5 = pig+cattle, 6 = cattle+digestate.

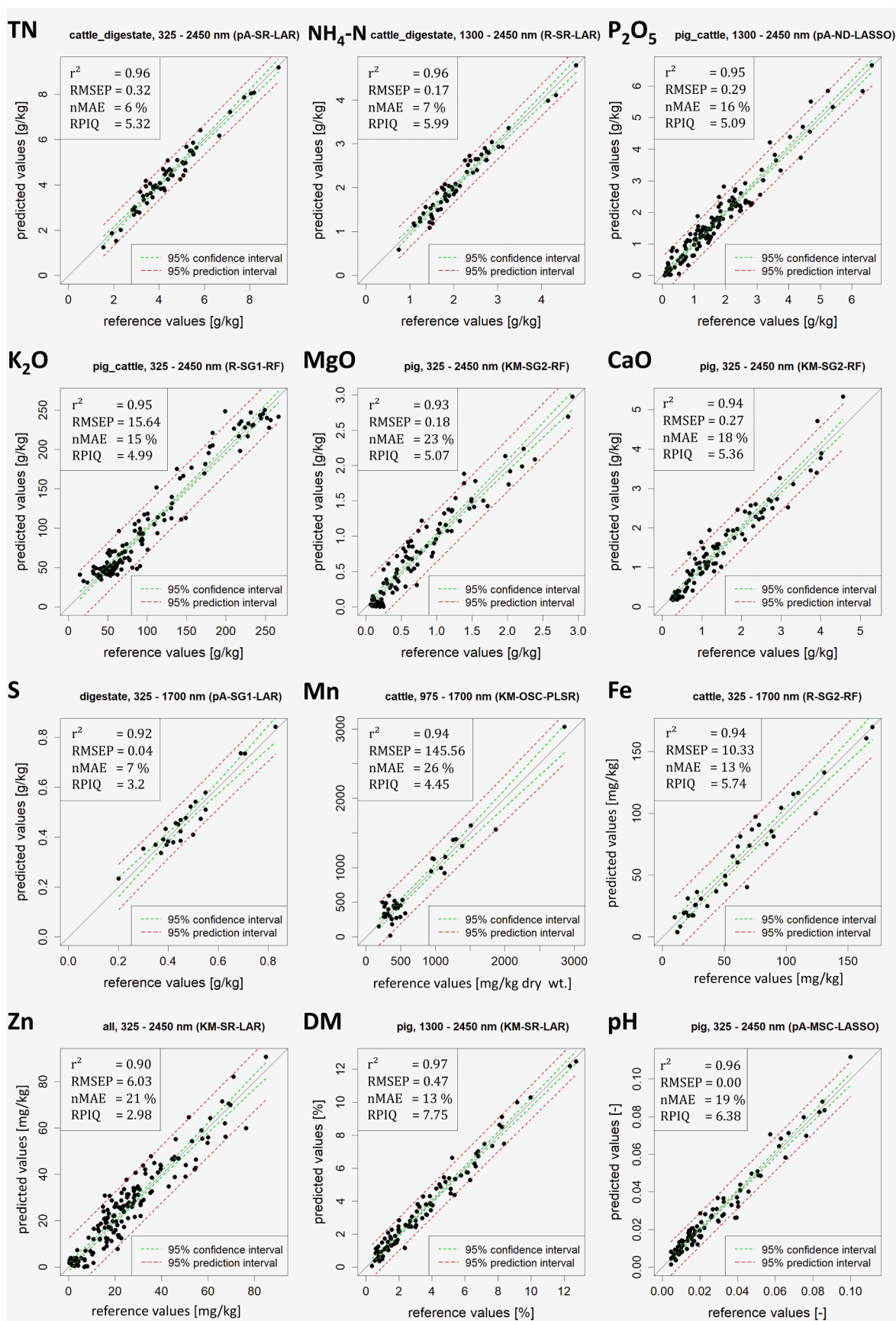


Fig. 5. Plots of random errors of the best prediction models for 12 target parameters (systematic errors were removed by correcting the slope and intercept to the 1:1 line); r^2 = squared Pearson coefficient; for Copper and Boron see Fig. 8.

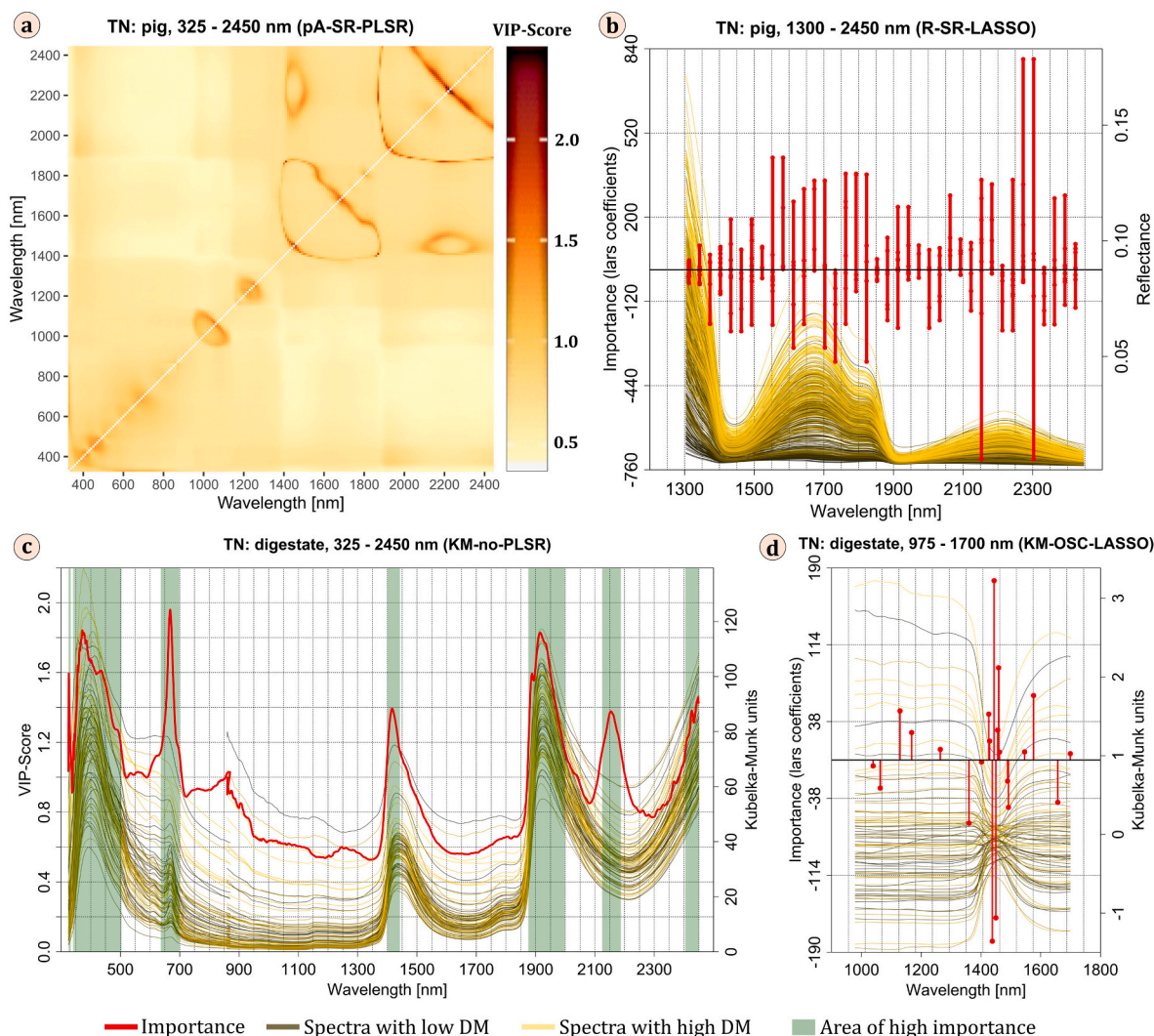


Fig. 6. Visualizations of variables importance for predicting TN. a) VIP-scores of wavelengths ratios using SR as preprocessing and PLSR as regression method. b) “lars”-coefficients of wavelengths ratios using SR as preprocessing and LASSO as regression method. c) VIP-scores of single wavelengths using no preprocessing and PLSR as regression method. d) “lars”-coefficients of single wavelengths using OSC as preprocessing and LASSO as regression method.

spectral resolution expressed in FWHM is always higher than the SI. In praxis, often three to four data points describe a peak at the level of FWHM. Thus, the 3-nm sampling interval can correspond to a simulated spectral resolution of about 10 nm FWHM (Rosero-Vlasova et al., 2016; Wolfrum et al., 2020). However, it is discussable if an SI of 4–5 nm would have been more appropriate to simulate low-cost spectrometers with worse spectral resolutions of about 15 nm FWHM. Further research would be needed regarding this topic.

The prediction results show that, when using only the visual range (325–870 nm), they were considerably deteriorated. Thus, the visual wavelengths do not contain much relevant information. Nevertheless, using the visual range in addition to the NIR range can slightly improve prediction results (see Supplementary materials and compare all predictions of 325–2450 nm vs. 860–2450 nm as well as 325–1700 nm vs. 975–1700 nm).

No considerable changes were recognized, when comparing the results for 900–1700 nm (RI: 1 nm) with 975–1700 nm (SI: 3 nm). Thus, the omission of the range between 900 and 975 nm and the reduction of the reporting interval to a 3-nm sampling interval did not affect prediction accuracy. This can be explained by the fact that numerous interactions with electromagnetic radiation produce broad and strongly overlapping peaks that may not require spectra data in a very high resolution.

It may be assumed, that using a wider range of wavelengths (325–2450 nm) would always lead to the best prediction results and that a reduction to shorter wavelength ranges would always decrease the prediction performance. However, in some cases, the exclusion of unimportant or noisy wavelengths did not affect or even slightly improved prediction results. In example, a reduction of the spectra to 1300–2450 nm increased prediction results for $\text{NH}_4\text{-N}$ in cattle LOM (r^2 from 0.86 to 0.92; RMSEP from 0.16 to 0.12 g/kg; see Supplementary materials). Especially, in respect to the envisaged usage of low-cost spectrometers with smaller wavelength ranges, these results are important for the development of affordable LOM sensor systems.

In other cases, a reduction of the wavelength range led to significant deteriorations of the prediction results. For example, when reducing the wavelength range from 325 to 2450 nm to 975–1700 nm, the prediction performance for TN in pig LOM decreased significantly (r^2 from 0.91 to 0.83; RMSEP from 0.50 to 0.69 g/kg). In contrast, a reduction to 1300–2450 nm resulted in a slightly increased performance ($r^2 = 0.92$; RMSEP = 0.46 g/kg; see Supplementary materials). Thus, there seem to be more wavelengths in the range of 1300–2450 nm than between 975 and 1700 nm, relevant for the prediction of TN and $\text{NH}_4\text{-N}$ in LOM. This assumption is matching with the corresponding VIP-scores and “lars”-coefficients, showing higher values in the range of 1300–2450 nm with the most intensive ones between 1850 and 2450 nm or, more specific,

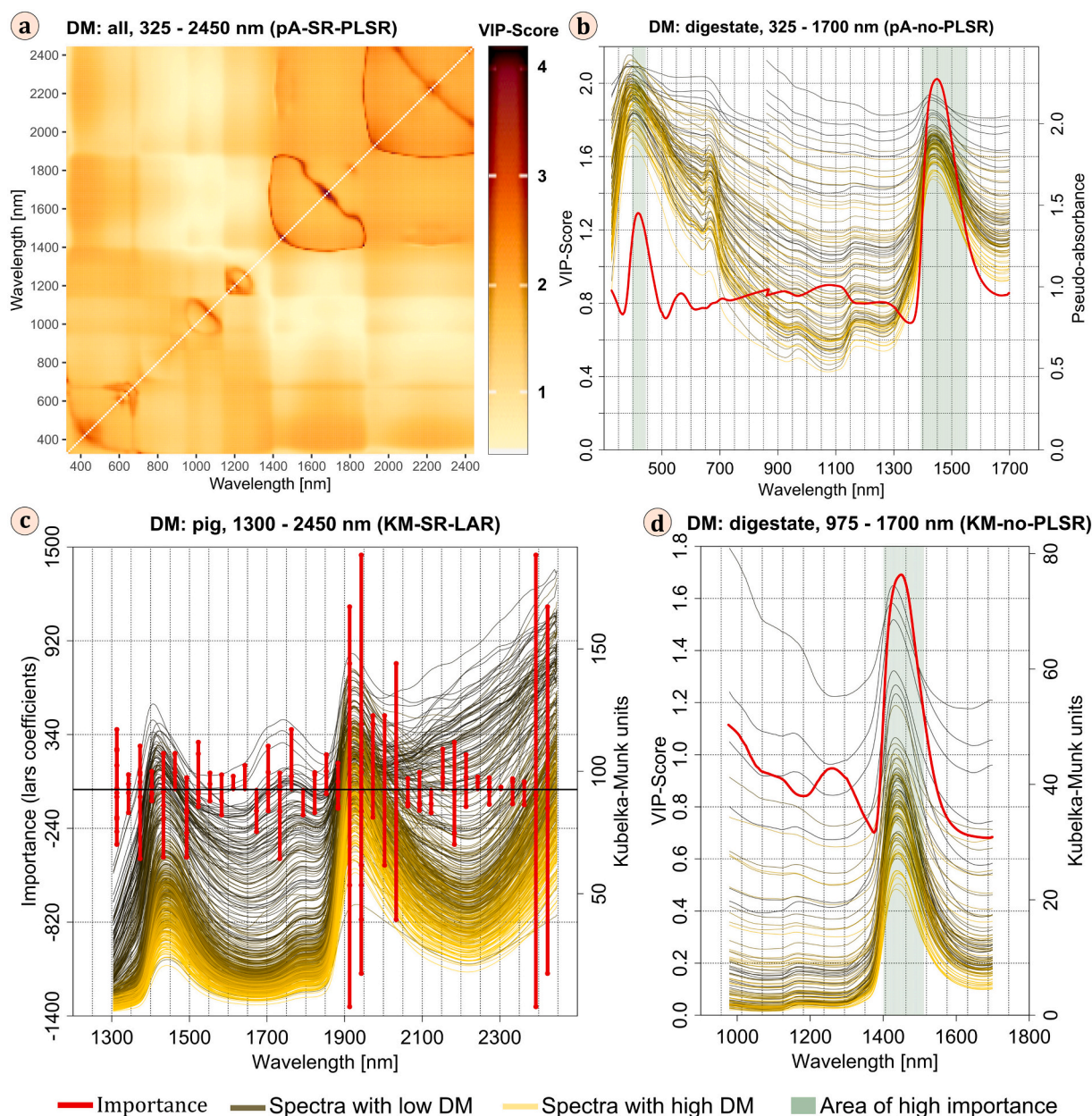


Fig. 7. Visualizations of variables importance for predicting DM. a) VIP-scores of wavelength ratios using SR as preprocessing method. b) VIP-scores of single wavelengths using no preprocessing and PLSR as regression method. c) “lars”-coefficients of single wavelengths using SR as pre-processing and LAR as regression method. d) VIP-scores of single wavelengths using no preprocessing and PLSR as regression method. For further important wavelengths of other target parameters, see Supplementary materials.

between 2100 and 2350 nm (e.g. for TN in pig LOM; Fig. 6A and C). When predicting TN, the models using the range of 975–1700 nm may only benefit from relevant wavelengths around 1500 nm (e.g. for TN in digestate LOM; Fig. 6B and D). Nevertheless, VIP-scores and “lars”-coefficients may change significantly for a target variable depending on the preprocessing and regression method.

In contrast to TN and $\text{NH}_4\text{-N}$, prediction performance for P_2O_5 was almost independent from the tested wavelength ranges. However, there were significant differences between LOM types: The best results for P_2O_5 were achieved for pig LOM ($r^2 = 0.94\text{--}0.96$). They were lower for cattle LOM ($r^2 = 0.86\text{--}0.89$) and much worse for digestate LOM (r^2 around 0.57; see Supplementary materials). Probably, because P_2O_5 in pig manure shows a higher correlation with other constituents as compared to the other LOM (Table 2).

The number of publications comparing the effect of different

wavelength ranges on prediction accuracies for LOM is very low. Cabassi et al., 2015 tested four spectrometers (including portable devices) with different wavelength ranges and resolutions. However, their rather poor results for some macronutrients in cattle LOM using two portable spectrometers as well as for two benchtop spectrometers with an artificially reduced spectral resolution of 10 nm (which may correspond to a sampling interval of about 3 nm) in the range of 1100–1690 nm were contradicting the results of the present study. Sørensen et al., 2007 excluded all wavelengths between 400 and 1200 nm and 2400–2500 nm for their chemometric analysis of pig and cattle LOM spectra, as these wavelengths did not improve prediction results for several macro- and micronutrients. This is partly conflicting with the present study as the variable importance indicates a high relevance for wavelengths from 2400 to 2450 nm for several target variables (Figs. 6AB, 7AB, and 8). These results additionally implicate not to use a

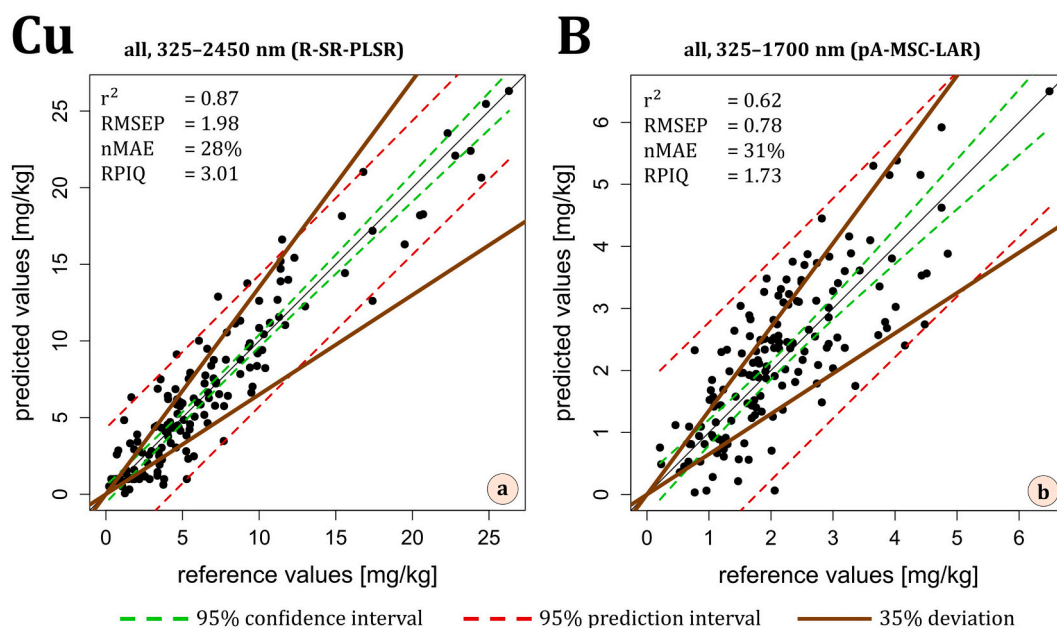


Fig. 8. Regression models for predicting Cu in pig and digestate LOM and B in pig LOM. The figures additionally include the upper and lower limits of 35 % deviation, indicating that predictions of lower concentrations often exceed 35 % deviation.

certified reflection standard that is based on the German PTB reflection standard as it has limited range up to 2450 nm. To profit from the relevant information between 2450 and 2500 nm, strongly influenced by water absorption, it is instead recommended to use the American NIST standard providing reflectance values up to 2500 nm due to lower restrictions in precision.

As the Vis-NIR spectra of LOM are mainly influenced by water, DM was one of the best predicted target parameters with r^2 values above 0.93 and a RMSEP up to 0.47 %. As expected, the two strong absorption bands around 1450 and 1910 nm were found to be most relevant for an accurate prediction of DM (Fig. 7). Nevertheless, also other tested wavelength regions led to similarly good results (see Supplementary materials), possibly caused by smaller water absorption bands near 975 nm, 1175 nm and the beginning of a stronger water absorption band around 2450 nm (Fig. 4). As DM is the parameter that correlates most with other target parameters (r often between 0.65 and 0.93; see Table 2), these relevant wavelength regions for DM often improved the prediction accuracy of several other target parameters (compare Figs. 6, 7, and Supplementary materials).

3.3.5. Wet and dry weight concentration basis

The predictions for K_2O were significantly worse in comparison to other nutrients, when utilizing wet weight concentrations. This confirms the results of Cobbinah et al., 2022 and Chen et al., 2010, which obtained similar results for K_2O . However, when using concentrations based on **dry weight concentrations**, predictions improved significantly. In example, for K_2O in pig LOM using the range of 975–1700 nm, r^2 increased significantly from 0.75 to 0.92 (RPIQ from 2.1 to 4.3; see Supplementary materials). In addition to K_2O , Mn predictions also profited from dry wt. concentrations. For Mn in cattle LOM using the total range of 325–2450 nm, r^2 increased from 0.81 to 0.94 (RPIQ from 2.0 to 4.5; see Supplementary materials). A predicted concentration based on dry wt. is afterwards easily transformed to wet wt. concentration by multiplying with the corresponding DM content.

A similar effect was noticed for pH predictions, although there is no scientific practice for dividing pH values by DM content. Nevertheless, this pH ratio led to much better predictions for pig and cattle and to a lesser extent for digestate LOM. In example, for pig LOM, r^2 increased from 0.62 to 0.95 (RPIQ from 1.6 to 5.5) and for cattle LOM, r^2 increased

from 0.55 to 0.95 (RPIQ from 1.5 to 4.7) using the range of 325–2450 nm (see Supplementary materials).

Although predictions for K_2O , Mn, and pH were superior, when modelling with concentrations on dry wt. basis, it must be kept in mind that when determining the necessary component DM with optical spectrometry instead of a standard laboratory method, the prediction errors would sum up and thus maybe deteriorate the positive effect of dry wt. concentrations. However, the accurate predictions for DM with optical spectroscopy still indicate that the models are superior to those using wet wt. concentrations. Nevertheless, further research is needed regarding this approach.

3.3.6. Presentations of the hydronium activity

It was ascertained that instead of using the decade logarithm of the hydronium activity (as pH is conventionally defined), the inverse of the natural logarithm $\ln(a(H_3O^+))^{-1}$ may slightly improve prediction performance. In example, for pig LOM, r^2 increased from 0.95 to 0.96 (RPIQ from 5.5 to 6.3) when using the 325–2450 nm range and dry wt. concentrations. For digestate LOM, r^2 increased from 0.74 to 0.84 (RPIQ from 2.3 to 3.2) when using $\ln(a(H_3O^+))^{-1}$, the 325–2450 nm range and wet wt. concentrations (see Supplementary materials). In a few cases, using the logarithm with base 10^6 also increased prediction performances, e.g. for digestate LOM, r^2 increased from 0.81 to 0.87 (RPIQ from 2.4 to 3.1) using 975–1700 nm range and dry wt. concentrations. On the other hand, using only the H_3O^+ activity itself ($a(H_3O^+) = 10^{-pH}$) for modelling, the prediction quality significantly deteriorated. Nevertheless, predictions of transformed pH values need to be back transformed afterwards using the corresponding inverse calculation, which was omitted in this study.

3.3.7. LOM groups

By stratifying the samples into six **LOM groups** (all, pig, cattle, digestate, pig+cattle, cattle+digestate), it was tested if it is necessary to use separate models for each LOM type. This was the case in about 50 % of the 42 best prediction models, especially for Mg, S and Mn. In the other cases, it was beneficial to build models based on mixed LOM types. In example, the prediction performance for K_2O in the range of 325–2450 nm improved significantly when using both cattle and digestate samples ($r^2 = 0.85$, RMSEP = 8.5 g/kg dry wt., RPIQ = 3.7,

Table 5
Comparison of 95 % prediction intervals calculated with either RMSEP or MAE. Concentrations on dry wt. basis are highlighted in orange letters.

Target variable	LOM type	Unit	95% prediction interval using RMSEP	95% prediction interval using MAE	Ratio
Total N	pig	g/kg wet wt.	0.75	0.74	1.01
	cattle	g/kg wet wt.	0.52	0.52	1.00
	digestate	g/kg wet wt.	0.63	0.62	1.02
NH₄-N	pig	g/kg wet wt.	0.46	0.44	1.04
	cattle	g/kg wet wt.	0.23	0.22	1.05
	digestate	g/kg wet wt.	0.33	0.34	0.95
P₂O₅	pig	g/kg wet wt.	0.57	0.55	1.04
	cattle	g/kg wet wt.	0.38	0.41	0.93
	digestate	g/kg wet wt.	0.41	0.42	0.97
K₂O	pig	g/kg dry wt.	31.75	30.48	1.04
	cattle	g/kg dry wt.	16.57	16.35	1.01
	digestate	g/kg dry wt.	16.57	16.35	1.01
MgO	pig	g/kg wet wt.	0.35	0.34	1.03
	cattle	g/kg wet wt.	0.21	0.22	0.97
	digestate	g/kg wet wt.	0.23	0.23	0.99
CaO	pig	g/kg wet wt.	0.59	0.56	1.04
	cattle	g/kg wet wt.	0.70	0.67	1.04
	digestate	g/kg wet wt.	1.08	1.04	1.03
S	pig	g/kg wet wt.	0.17	0.17	0.98
	cattle	g/kg wet wt.	0.13	0.13	0.98
	digestate	g/kg wet wt.	0.07	0.08	0.92
B	pig	mg/kg wet wt.	1.53	1.54	0.99
	cattle	mg/kg wet wt.	1.07	1.04	1.03
	digestate	mg/kg wet wt.	1.14	1.13	1.00
Mn	pig	mg/kg dry wt.	526.75	517.66	1.02
	cattle	mg/kg dry wt.	288.18	281.27	1.02
	digestate	mg/kg wet wt.	115.17	110.37	1.04
Fe	pig	mg/kg wet wt.	23.82	22.74	1.05
	cattle	mg/kg wet wt.	24.16	26.34	0.92
	digestate	mg/kg wet wt.	46.17	44.79	1.03
Cu	pig	mg/kg wet wt.	3.87	3.70	1.05
	cattle	mg/kg wet wt.	1.78	1.88	0.95
	digestate	mg/kg wet wt.	3.87	3.70	1.05
Zn	pig	mg/kg wet wt.	11.83	11.87	1.00
	cattle	mg/kg wet wt.	11.83	11.87	1.00
	digestate	mg/kg wet wt.	11.00	11.66	0.94
DM	pig	%	0.92	0.95	0.96
	cattle	%	0.94	0.92	1.02
	digestate	%	1.17	1.18	0.99
pH	pig	ln(a(H ₃ O ⁺)) ⁻¹ dry wt.	0.01	0.01	1.06
	cattle	dry wt.	0.15	0.15	0.99
	digestate	dry wt.	0.15	0.15	0.99

nMAE = 10 %), instead of fitting a separate cattle ($r^2 = 0.76$, RMSEP = 10.4 g/kg dry wt., RPIQ = 1.9, nMAE = 15 %) and digestate model ($r^2 = 0.75$, RMSEP = 11.3 g/kg dry wt., RPIQ = 2.0, nMAE = 12 %; see Table 3 and Supplementary materials). It is assumed that using several LOM groups may better prevent from model overfitting.

3.3.8. Results of further macro- (MgO, CaO, S) and micronutrients (Fe, Cu, Zn)

The prediction performance for MgO and CaO was very similar to P₂O₅ and they were almost independent from the tested wavelength ranges, too. Furthermore, the prediction performance of MgO and CaO in pig LOM outperformed predictions in cattle and digestate LOM. Thus, their performance probably depends on the correlation to P₂O₅, which is high for Mg and Ca in pig LOM ($r = 0.97$ and 0.95), lower for Mg in cattle LOM ($r = 0.78$) and lowest for Ca in cattle LOM as well as for Mg and Ca in digestate LOM ($r = 0.52$ and 0.51 ; see also Table 2 and Supplementary materials).

S was the parameter with the most constant level in predicted r^2 (0.76–0.86 for 28 out of 30 results), independent from the tested wavelength ranges and LOM types, although correlations to other target parameters significantly change with LOM type. Thus, the prediction of

S is probably influenced by several optical interactions over the whole Vis-NIR range (probably due to overtone and combination vibrations of molecule bindings with oxygen and hydrogen as solved ion or bound in organic matter), making S independent from an indirect prediction by other parameters. However, there was one outstanding prediction for S in digestate LOM using the wavelengths from 325 to 1700 nm, with an r^2 of 0.92. The reason for this good result is not clear, as the most intensive “lars”-coefficients for this model were in the range of 975–1700 nm, whereas the best model using only this wavelength range only achieved an r^2 of 0.80 (see Supplementary materials).

Good prediction results were also achieved for the micronutrients Fe, Cu, and Zn with best r^2 values between 0.80 and 0.94 (Table 4). Besides some correlations to other elements (see Table 2), this good prediction performance can probably be traced back to optical active compounds. According to Shen et al., 2014, Cu tends to form complex bonds with organic ligands having functional CONH-R groups such as proteins, urea, amino acids. As Zn and Fe have similar physical and chemical properties, they are expected to form similar complex bonds. However, the r^2 for the predictions of Cu and Zn using exclusively digestate samples, as well as of Zn using exclusively cattle samples, did not exceed 0.69 (see Supplementary materials). The reason for these low prediction

performances is not clear.

3.4. Prediction intervals

To state the model accuracy for potential usage in practical agricultural, it is proposed to specify a predicted concentration with a 95 % prediction interval. It can be calculated either by $1.96 \cdot \text{RMSEP}$ or alternatively by $1.96 \cdot \text{MAE} / (2/\pi)^{0.5}$ (see Eq. (8), Section 2.5.2). The assumption that both methods lead to similar results is confirmed in Table 5, where the ratio of these prediction intervals is always close to 1.0.

Since 2017, the German Agricultural Society (DLG) has been certifying commercially available NIR spectrometers measuring TS, TN, $\text{NH}_4\text{-N}$, P_2O_5 or K_2O in pig, cattle, or digestate LOM. The certificate assures that no sample differs >35 % and at least 60 % of the five tested samples <25 % from their reference lab values (Drücker, 2020). A DLG certificate for the new NIRS technology is an important step forward to establish NIRS for on-line measurements at farm and field scale. However, the coefficient of variation (standard deviation relative to the mean) of a NIRS prediction strongly depends on its concentration level: the lower the concentration, the higher the coefficient of variation (Fig. 8), which matches with the Horwitz function (Souza and Nogueira, 2021). Furthermore, the laboratory measurement uncertainty has to be taken into account, too. Especially lower concentrations of micronutrients are worse in repeatability (higher than the mentioned 1.7 to 3.6 %, which corresponds to the mean deviation of a whole concentration range) and especially in reproducibility (up to approximately 50 %).

For this reason, it is proposed to evaluate NIRS accuracy by a constant error ratio like a 95 % prediction interval, which is independent from the concentration level. In the case of TN in pig LOM, the best model in this study obtained a predicted concentration of ± 0.74 g/kg as 95 % prediction interval (Table 5) for every nitrogen analysis independent from the level of the predicted nitrogen concentration. Thus, the prediction of a high concentration, e.g. 10 g/kg TN, would result in a maximum 7.4 % deviation and the prediction of a low concentration, e.g. 1 g/kg TN, would result in a maximum 74 % deviation from its true value. Although a percentage deviation of 74 % is much higher than 7.4 %, it is still the same amount of 0.74 g TN per kg LOM that is over- or underestimated. Finally, the absolute deviating amount is the value of interest in context of precision farming, not a relative deviation. Nevertheless, it could be possible to calibrate two separate models for lower and higher concentrations to improve the percentage deviations of smaller values. Regarding this aspect, further research is needed.

Furthermore, it has to be kept in mind that lower concentrations of micronutrients determined by standard chemical methods are worse in laboratory repeatability and especially reproducibility (up to approximately 50 %), thus complicating the real deviations of such spectrometric predictions. Although a traditional lab based analysis as such is still more accurate for each nutrient, it has to be emphasized that this method cannot compensate for errors of non-representative samplings. In this context, using optical spectrometry to analyse the complete slurry of a storage tank, while filling a slurry tanker or during field application, may be equal to or even outperform the accuracy of chemical analyses.

4. Conclusions

Good results ($r^2 = 0.80\text{--}0.97$) for all target parameters except boron indicate a sufficient accuracy as alternative technology to nutrient tables and chemical analyses. Regarding chemometrics, the regression methods PLSR, LASSO, and LAR predominantly performed best. Furthermore, it needs to be highlighted that two new preprocessing algorithms, named simple ratios (SR) and normalized differences (ND), were successfully introduced to the topic of multivariate analysis of LOM

spectra. Extended research on combinations of SR or ND with other preprocessing techniques may further increase model performance.

With the aim to cut expenses in practical applications, the performance of low-cost spectrometers with lower resolutions and smaller wavelength ranges was simulated. It turned out that low-cost spectrometers could provide significant information and would not deteriorate prediction performance. Comparing the wavelength ranges of 975–1700 nm and 1300–2450 nm, it is proposed to use the latter, due to a significant improvement of TN and $\text{NH}_4\text{-N}$ predictions and similar ones for other macronutrients, DM, and pH, although performances for micronutrients were slightly superior the range of 975–1700 nm.

As a further aspect, an appropriate evaluation of prediction accuracy was discussed. Due to the normal distribution of the regression residuals, independent from the concentration level, it is proposed to evaluate the prediction accuracy of a spectrometric analysis by a constant value like a 95 % prediction interval rather than by a percentage deviation from a true value. Nevertheless, further research is needed to examine the effect of two separate calibration models for lower and higher concentrations to avoid high percentage deviations for smaller values.

Finally, as the average composition of LOM may change considerably over time (e.g. due to changes of forage crops adapted to climate change), periodical calibrations with new samples are needed to guarantee a constant accuracy over time.

CRedit authorship contribution statement

Michael Horf: Writing – original draft, Conceptualization, Methodology, Software, Validation, Visualization, Formal analysis, Investigation, Resources, Data curation. **Robin Gebbers:** Writing – review & editing, Supervision, Conceptualization, Funding acquisition, Project administration. **Hans-Werner Olf:** Writing – review & editing, Conceptualization, Resources, Funding acquisition, Project administration. **Sebastian Vogel:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was funded by the German Federal Environmental Foundation (Deutsche Bundesstiftung Umwelt—DBU), grant no. AZ 33702.

It was additionally supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany. The Federal Office for Agriculture and Food (BLE) provides coordinating support for future farms and future regions as funding organisation, grant number: 28DE202A21.

Many thanks go to Max-Frederik Piepel for organizing the extensive and diverse LOM sample set. Furthermore, we gratefully thank Uwe Frank for several transports of frozen LOM samples halfway across Germany, Marc Zimme and Benjamin Mahns for their help with constructing the optical measurement setup, Axel Anlauff for his help to produce a reflection standard, Andreas Hoffmann from LUFÄ NW for his valuable advice, and Redmond Shamshiri for all his efforts to improve to final manuscript, especially in the visualization.

Appendix A

Table 1

Descriptive statistics for DM, pH and 12 nutrients of A) pig, B) cattle, and C) digestate LOM samples; nutrient concentrations determined by classical chemical methods on wet weight basis.

A – pig	DM	TN	NH ₄ -N	P ₂ O ₅	K ₂ O	MgO	CaO	S	Mn	Fe	Cu	Zn	B	pH
Unit	%	g/kg	mg/kg											
Min	0.30	0.92	0.66	0.05	0.76	0.07	0.18	0.05	0.74	0.32	0.21	0.38	0.40	5.50
Max	12.75	7.45	4.72	6.52	5.86	3.47	6.61	1.10	381	117	74.50	141	8.94	8.70
Average	3.01	3.13	2.03	1.30	2.58	0.70	1.32	0.31	65.50	23.58	8.51	26.68	2.35	7.71
Standard deviation (SD)	2.46	1.46	0.82	1.32	1.25	0.64	1.14	0.22	74.01	23.64	12.02	27.59	1.60	0.50
Relative SD (RSD; %)	81.61	46.66	40.21	101.8	48.36	92.11	86.45	72.03	113.0	100.3	141.3	103.4	67.99	6.46

B – cattle	DM	TN	NH ₄ -N	P ₂ O ₅	K ₂ O	MgO	CaO	S	Mn	Fe	Cu	Zn	B	pH
Unit	%	g/kg	mg/kg											
Min	0.95	0.90	0.31	0.19	0.90	0.11	0.43	0.08	3.47	3.94	0.35	1.11	0.32	6.80
Max	11.39	5.69	3.07	3.62	7.19	1.81	5.51	1.22	922	574	26.70	75.80	27.00	8.50
Average	7.11	3.76	1.76	1.50	4.21	0.98	2.17	0.42	52.36	74.22	5.15	21.43	2.27	7.60
Standard deviation (SD)	2.67	0.95	0.50	0.66	1.15	0.37	1.02	0.17	104.73	66.81	2.51	9.52	2.88	0.30
Relative SD (RSD; %)	37.55	25.32	28.41	43.71	27.36	37.83	46.96	39.06	200.0	90.02	48.70	44.40	127.12	4.00

C - digestate	DM	TN	NH ₄ -N	P ₂ O ₅	K ₂ O	MgO	CaO	S	Mn	Fe	Cu	Zn	B	pH
Unit	%	g/kg	mg/kg											
Min	2.65	3.18	1.09	0.94	1.51	0.12	1.09	0.20	18.10	9.16	2.16	9.51	0.23	7.30
Max	10.57	11.10	7.89	4.17	9.95	1.70	7.36	0.90	3156	758	30.50	58.30	5.34	8.50
Average	6.97	5.50	2.85	2.02	5.60	1.03	2.33	0.47	269	140	6.64	26.40	2.51	7.91
Standard deviation (SD)	1.65	1.38	1.07	0.55	1.27	0.29	0.99	0.12	433	163	4.20	10.50	0.93	0.27
Relative SD (RSD; %)	23.75	25.07	37.64	27.07	22.59	27.92	42.41	24.76	160.8	117.5	63.30	39.79	37.08	3.45

Table 2

Pearson coefficients (r) for dry matter content (DM), pH and 12 nutrient concentrations in all LOM samples, determined by standard lab procedures. Correlations on wet weight basis are displayed in the lower triangle highlighted *italic* and *blue*, on dry weight basis in the upper triangle. Correlation coefficients ≥ 0.65 are underlined.¹

all LOM														
wet \ dry wt.	DM	pH	TN	NH ₄ -N	P ₂ O ₅	K ₂ O	MgO	CaO	S	Mn	Fe	Cu	Zn	B
DM		<i><u>-0.72</u></i>	<i><u>-0.73</u></i>	-0.08	<i><u>-0.69</u></i>	-0.22	-0.35	-0.52	0.02	0.11	-0.29	-0.24	-0.51	-0.61
pH	<i><u>0.70</u></i>		<i><u>0.99</u></i>	0.23	<i><u>0.90</u></i>	0.34	0.55	<i><u>0.75</u></i>	0.06	0.00	0.23	0.22	0.60	0.51
TN	<i><u>0.32</u></i>	<i><u>0.87</u></i>		0.16	<i><u>0.90</u></i>	0.28	0.49	<i><u>0.73</u></i>	0.04	-0.05	0.19	0.16	0.58	0.57
NH ₄ -N	<i><u>0.74</u></i>	<i><u>0.74</u></i>	<i><u>0.53</u></i>		0.04	<i><u>0.75</u></i>	0.62	0.28	0.19	0.36	0.37	<i><u>0.78</u></i>	0.22	-0.32
P ₂ O ₅	<i><u>0.68</u></i>	<i><u>0.72</u></i>	<i><u>0.50</u></i>	<i><u>0.39</u></i>		0.21	0.42	<i><u>0.69</u></i>	-0.02	-0.10	0.14	0.07	0.55	0.50
K ₂ O	<i><u>0.80</u></i>	<i><u>0.67</u></i>	<i><u>0.42</u></i>	<i><u>0.93</u></i>	<i><u>0.43</u></i>		<i><u>0.74</u></i>	0.32	0.05	0.21	0.42	<i><u>0.65</u></i>	0.40	-0.07
MgO	<i><u>0.80</u></i>	<i><u>0.62</u></i>	<i><u>0.34</u></i>	<i><u>0.83</u></i>	<i><u>0.43</u></i>	<i><u>0.85</u></i>		0.48	0.11	0.17	0.37	0.54	0.44	0.17
CaO	<i><u>0.76</u></i>	<i><u>0.76</u></i>	<i><u>0.56</u></i>	<i><u>0.81</u></i>	<i><u>0.56</u></i>	<i><u>0.79</u></i>	<i><u>0.77</u></i>		0.05	0.05	0.27	0.27	0.42	0.29
S	<i><u>0.23</u></i>	<i><u>0.35</u></i>	<i><u>0.30</u></i>	<i><u>0.24</u></i>	<i><u>0.25</u></i>	<i><u>0.20</u></i>	<i><u>0.24</u></i>	<i><u>0.28</u></i>		-0.08	0.07	0.12	0.00	-0.12
Mn	<i><u>0.44</u></i>	<i><u>0.50</u></i>	<i><u>0.34</u></i>	<i><u>0.46</u></i>	<i><u>0.33</u></i>	<i><u>0.42</u></i>	<i><u>0.43</u></i>	<i><u>0.41</u></i>	<i><u>-0.04</u></i>		0.14	0.29	-0.09	-0.29
Fe	<i><u>0.22</u></i>	<i><u>0.26</u></i>	<i><u>0.17</u></i>	<i><u>0.44</u></i>	<i><u>-0.02</u></i>	<i><u>0.42</u></i>	<i><u>0.37</u></i>	<i><u>0.38</u></i>	<i><u>0.09</u></i>	<i><u>0.21</u></i>		<i><u>0.69</u></i>	0.37	0.01
Cu	<i><u>0.52</u></i>	<i><u>0.56</u></i>	<i><u>0.42</u></i>	<i><u>0.87</u></i>	<i><u>0.15</u></i>	<i><u>0.82</u></i>	<i><u>0.70</u></i>	<i><u>0.65</u></i>	<i><u>0.16</u></i>	<i><u>0.39</u></i>	<i><u>0.66</u></i>		0.26	-0.21
Zn	<i><u>0.36</u></i>	<i><u>0.37</u></i>	<i><u>0.24</u></i>	<i><u>0.46</u></i>	<i><u>0.20</u></i>	<i><u>0.44</u></i>	<i><u>0.43</u></i>	<i><u>0.42</u></i>	<i><u>0.13</u></i>	<i><u>0.16</u></i>	<i><u>0.42</u></i>	<i><u>0.45</u></i>		0.40
B	<i><u>-0.32</u></i>	<i><u>-0.04</u></i>	<i><u>0.14</u></i>	<i><u>-0.12</u></i>	<i><u>-0.01</u></i>	<i><u>-0.17</u></i>	<i><u>-0.20</u></i>	<i><u>-0.18</u></i>	<i><u>0.04</u></i>	<i><u>0.03</u></i>	<i><u>-0.18</u></i>	<i><u>-0.09</u></i>	<i><u>-0.10</u></i>	

¹p-Values ≤ 0.01 for underlined values, indicating highly significance linear relationships.

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2023.168045>.

References

- Aguiar-Villegas, H.A., Sharara, M.A., Larson, R.A., 2018. Nutrient variability following dairy manure storage agitation. *Appl. Eng. Agric.* 34 (6), 908–917. <https://doi.org/10.13031/aea.12796>.
- Ammaan, M., Akila, A., Muthukrishnan, M., Nivedhaa, M.A., Rahul, A., Rishi, B., et al., 2019. Effect of organic manures and biofertilizers on soil microbial population in amaranth (*Amaranthus blitum*). *Int. J. Curr. Microbiol. App. Sci.* 8 (06), 700–704. <https://doi.org/10.20546/ijcmas.2019.806.081>.
- Asai, T., Shimizu, S., Koga, T., Sato, M., 1993. Quick determination of total nitrogen, total carbon and crude ash in cattle manure using near infrared reflectance spectroscopy. *Jpn. J. Soil Sci. Plant Nutr.* 64 (6), 669–675. <https://doi.org/10.20710/dojo.64.6.669>.
- Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43 (5), 772–777. <https://doi.org/10.1366/0003702894202201>.
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the

- quality of the prediction of soil attributes by NIR spectroscopy. *Trends Anal. Chem.* 29 (9), 1073–1081. <https://doi.org/10.1016/j.trac.2010.05.006>.
- Ben Hassen, T., El Bilali, H., 2022. Impacts of the Russia-Ukraine war on global food security: towards more sustainable and resilient food systems? *Foods* 11 (15). <https://doi.org/10.3390/foods11152301>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brereton, R.G., Jansen, J., Lopes, J., Marini, F., Pomerantsev, A., Rodionova, O., et al., 2017. Chemometrics in analytical chemistry-part I: history, experimental design and data analysis tools. *Anal. Bioanal. Chem.* 409 (25), 5891–5899. <https://doi.org/10.1007/s00216-017-0517-1>.
- Cabassi, G., Cavalli, D., Fuccella, R., Marino Gallina, P., 2015. Evaluation of four NIR spectrometers in the analysis of cattle slurry. *Biosyst. Eng.* 133, 1–13. <https://doi.org/10.1016/j.biosystemseng.2015.02.011>.
- Cassotti, M., Grisoni, P., 2012. Variable selection methods: an introduction (Tutorial 6). In: *Molecular Descriptors*. <https://doi.org/10.13140/RG.2.1.1009.5129>.
- Chen, L.J., Xing, L., Han, L.J., 2010. Influence of data preprocessing on the quantitative determination of nutrient content in poultry manure by near infrared spectroscopy. *J. Environ. Qual.* 39 (5), 1841. <https://doi.org/10.2134/jeq2009.0449>.
- Cobbinah, E., Generalao, O., Lageshetty, S.K., Adrianto, I., Singh, S., Dumanacas, G.G., 2022. Using near-infrared spectroscopy and stacked regression for the simultaneous determination of fresh cattle and poultry manure chemical properties. *Chemosensors* 10 (10), 410. <https://doi.org/10.3390/chemosensors10100410>.
- Derikx, P., van de Kooij, B., Heskamp, H., Rozijn, M., 2021. NIRS and Manure Composition (NIMACO) (Wageningen; WFSR report 2021.012.).
- DIN e.V., 2001. Characterization of Sludges - Determination of Dry Residue and Water Content; German Version of EN 12880:2000. Beuth Verlag GmbH, Berlin, Germany.
- DIN e.V., 2005. Water Quality - Determination of Ammonium Nitrogen - Method by Flow Analysis (CFA and FIA) and Spectrometric Detection (CFA und FIA) und spektrometrischer Detektion (ISO 11732:2005); German Version of EN ISO 11732:2005. Beuth Verlag GmbH, Berlin, Germany.
- DIN e.V., 2009. Water Quality - Determination of Selected Elements by Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES) (ISO 11885:2007); German version of EN ISO 11885:2009. Beuth Verlag GmbH, Berlin, Germany.
- DIN e.V., 2012a. Sludge, Treated Biowaste and Soil - Determination of Kjeldahl Nitrogen; German Version of EN 16169:2012. Beuth Verlag GmbH, Berlin, Germany.
- DIN e.V., 2012b. Sludge, Treated Biowaste and Soil - Determination of pH; German Version of EN 15933:2012. Beuth Verlag GmbH, Berlin, Germany.
- Drücker, H., 2020. Organische Düngung. <https://doi.org/10.24355/dbbs.084-202001201530-0>.
- Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. *Ann. Stat.* 32 (2), 407–499. <https://doi.org/10.1214/009053604000000067>.
- Fan, J., Lv, J., 2014. Sure independence screening. In: Balakrishnan, N., Colton, T., Everitt, B., Piegorsch, W., Ruggieri, F., Teugels, J.L. (Eds.), *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, Chichester, UK, pp. 1–8.
- Flynn, K., Spiegel, S., Kleinman, P.J., Meinen, R.J., Smith, D.R., 2023. Manure management to overcome longstanding nutrient imbalances in US agriculture. *Resour. Conserv. Recycl.* 188, 106632. <https://doi.org/10.1016/j.resconrec.2022.106632>.
- Foged, H., Flotats, X., Bonmatí, A., Palatsi, J., Magrí, A., Schelde, K., 2011. Inventory of manure processing activities in Europe. Brussels, Belgium. In: *Technical Report No. I to the European Commission, Directorate-General Environment*. Project reference: ENV.B.1/ETU/2010/0007 (138 pp.).
- Fornaro, B., Burini, D., Biczysko, M., Barone, V., 2015. Hydrogen-bonding effects on infrared spectra from anharmonic computations: uracil-water complexes and uracil dimers. *Chem. Eur. J.* 119 (18), 4224–4236. <https://doi.org/10.1021/acs.jpca.5b01561>.
- Friedrich, L., Simon, P., Hoffmann, A., Krieger, K.-L., 2022. Classification of manure type and their influence on the accuracy of nutrient determination in organic manures using NIR spectroscopy. In: *VDI Wissensforum GmBH (Ed.), VDI-Berichte: International Conference on Agricultural Engineering, AgEng-LAND. TECHNIK, November 22nd and 23rd 2022, Berlin, Germany*. VDI Verlag GmbH, Duesseldorf, Germany, pp. 359–364.
- Geladi, P., MacDougall, D., Martens, H., 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* 39 (3), 491–500. <https://doi.org/10.1366/0003702854248656>.
- Harland, C.E., 1994. Ion-exchange: theory and practice. Royal Society of Chemistry, London.
- Hastie, T., Tibshirani, R., Friedman, J.H., 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York, USA.
- Ho, T.K., 1995. Random decision forests. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*; 14–16 Aug. 1995. IEEE Comput. Soc. Press, Montreal, Que., Canada.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: applications to nonorthogonal problems. *Technometrics* 12 (1), 69–82. <https://doi.org/10.2307/1267352>.
- Horf, M., Bönecke, E., Gebbers, R., Kling, C., Kramer, E., Rühlmann, J., et al., 2022a. Utility of visible and near-infrared spectroscopy to predict base neutralizing capacity and lime requirement of quaternary soils. *Precis. Agric.* <https://doi.org/10.1007/s11119-022-09945-9>.
- Horf, M., Vogel, S., Drücker, H., Gebbers, R., Olf, H.-W., 2022b. Optical spectrometry to determine nutrient concentrations and other physicochemical parameters in liquid organic manures: a review. *Agronomy* 12 (2), 514. <https://doi.org/10.3390/agronomy12020514>.
- Huang, Y., 2022. *Chemometric methods in analytical spectroscopy technology*. In: Chu, X., Huang, Y., Yun, Y.-H., Bian, X. (Eds.), *Chemometric Methods in Analytical Spectroscopy Technology*. Springer Nature, Singapore, Singapore, pp. 1–29.
- Indahl, U.G., Liland, K.H., Naes, T., 2009. Canonical partial least squares—a unified PLS approach to classification and regression problems. *J. Chemom.* 23 (9), 495–504. <https://doi.org/10.1002/cem.1243>.
- Kitamura, R., Sugiyama, C., Yasuda, K., Nagatake, A., Yuan, Y., Du, J., et al., 2021. Effects of three types of organic fertilizers on greenhouse gas emissions in a grassland on andosol in southern Hokkaido, Japan. *Front. Sustain. Food Syst.* 5. <https://doi.org/10.3389/fsufs.2021.649613>.
- Malley, D., Martin, P., Dettman, L., 2001. *Analysis of Minor Elements and Metals in Hog Manure by Field-portable Near-infrared Spectroscopy: Results for the Zeiss Corona® Spectrometer: Final Report (3/3) to Manitoba Livestock Manure Management Initiative Inc. Manitoba Livestock Manure Management Initiative Inc. (MLMMI), Winnipeg, Canada*.
- Malley, D.F., Yesmin, L., Eilers, R.G., 2002. Rapid analysis of hog manure and manure-amended soils using near-infrared spectroscopy. *Soil Sci. Soc. Am. J.* 66 (5), 1677. <https://doi.org/10.2136/sssaj2002.1677>.
- Martens, H., Jensen, S.A., Geladi, P., 1983. Multivariate linearity transformation for near-infrared reflectance spectrometry. In: *Proceedings of the Nordic Symposium on Applied Statistics*.
- Mevik, B.-H., Wherrens, R., 2022. Introduction to the pls package. <https://cran.r-project.org/web/packages/pls/vignettes/pls-manual.pdf>.
- Minasny, B., McBratney, A., 2013. Why you don't need to use RPD. *Pedometron* 33, 14–15.
- Park, J., Kang, T., Heo, Y., Lee, K., Kim, K., Lee, K., et al., 2020. Evaluation of short-term exposure levels on ammonia and hydrogen sulfide during manure-handling processes at livestock farms. *Saf. Health Work* 11 (1), 109–117. <https://doi.org/10.1016/j.shaw.2019.12.007>.
- Piepel, M.-F., Olf, H.-W., 2023. Development of a physicochemical test kit for on-farm measurement of nutrients in liquid organic manures. *Agriculture* 13 (2), 477. <https://doi.org/10.3390/agriculture13020477>.
- Piepel, M.-F., Dittert, K., Olf, H.-W., 2022. Evaluation of physicochemical on-farm quick tests for estimating nutrient concentrations in pig slurry and development of an application for mobile devices. *Agronomy* 12 (11), 2809. <https://doi.org/10.3390/agronomy12112809>.
- Reeves, J.B., 2001. Near-infrared diffuse reflectance spectroscopy for the analysis of poultry manures. *J. Agric. Food Chem.* 49 (5), 2193–2197. <https://doi.org/10.1021/jf0013961>.
- Rosero-Vlasova, O.A., Pérez-Cabello, F., Montorio Llovería, R., Vlassova, L., 2016. Assessment of laboratory VIS-NIR-SWIR setups with different spectroscopy accessories for characterisation of soils from wildfire burns. *Biosyst. Eng.* 152, 51–67. <https://doi.org/10.1016/j.biosystemseng.2016.06.011>.
- Saeyns, W., Mouazen, A.M., Ramon, H., 2005. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosyst. Eng.* 91 (4), 393–402. <https://doi.org/10.1016/j.biosystemseng.2005.05.001>.
- Saeyns, W., Nguyen Do Trong, N., van Beers, R., Nicolai, B.M., 2019a. Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: a review. *Postharvest Biol. Technol.* 158. <https://doi.org/10.1016/j.postharvbio.2019.110981>.
- Saeyns, W., Watté, R., Postelmans, A., 2019b. Role of NIR spectroscopy in precision fertilisation with manure. In: *Proceedings of the International Fertilizer Society (IFS)*, Colchester, UK, pp. 1–21.
- Sanford, J.R., Larson, R.A., Digman, M.F., 2020. Assessing certified manure analysis laboratory accuracy and variability. *Appl. Eng. Agric.* 36 (6), 905–912. <https://doi.org/10.13031/aea.14214>.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36 (8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- Schirrmann, M., Gebbers, R., Kramer, E., 2013. Performance of automated near-infrared reflectance spectrometry for continuous in situ mapping of soil fertility at field scale. *Vadose Zone J.* 12 (4), vzj2012.0199. <https://doi.org/10.2136/vzj2012.0199>.
- Sen, A., Srivastava, M., 1990. *Regression Analysis: Theory, Methods and Applications*. John Wiley & Sons, New York, USA.
- Shen, X., Yang, Z., Huang, G., Han, L., 2014. Near infrared spectroscopy detection of copper in pig manure and the spectral basis of the analysis. *J. Near Infrared Spectrosc.* 22 (4), 305–312. <https://doi.org/10.1255/jnirs.1115>.
- Sørensen, L.K., Sørensen, P., Birkmose, T.S., 2007. Application of reflectance near infrared spectroscopy for animal slurry analyses. *Soil Sci. Soc. Am. J.* 71 (4), 1398. <https://doi.org/10.2136/sssaj2006.330>.
- Souza, G.B. de, Nogueira, A.R.d.A., 2021. Predicted standard deviation of proficiency testing performance in animal nutrition methods. *J. Phys. Conf. Ser.* 1826 (1), 12033. <https://doi.org/10.1088/1742-6596/1826/1/012033>.
- Tavakoli, H., Correa, J., Sabetizade, M., Vogel, S., 2023. Predicting key soil properties from Vis-NIR spectra by applying dual-wavelength indices transformations and stacking machine learning approaches. *Soil Tillage Res.* 229, 105684. <https://doi.org/10.1016/j.still.2023.105684>.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B. Methodol.* 58 (1), 267–288. <http://www.jstor.org/stable/2346178>.
- Waldrup, H.M., Pagliari, P.H., He, Z., 2020. *Animal Manure: Production, Characteristics, Environmental Concerns, and Management*. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI, USA.
- Williams, P., 2019. Karl H. Norris, the father of near-infrared spectroscopy. *NIR News* 30 (7–8), 25–27. <https://doi.org/10.1177/0960336019875883>.
- Williams, P., Eising, E., Malley, D.F., 2020. Industrial-scale continuous on-line analysis of liquid hog manure by NIRS. *NIR News* 31 (7–8), 25–29. <https://doi.org/10.1177/0960336020978714>.
- Wold, H., 1975. Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. *J. Appl. Probab.* 12 (S1), 117–142. <https://doi.org/10.1017/S0021900200047604>.

- Wold, S., Antti, H., Lindgren, F., Öhman, J., 1998. Orthogonal signal correction of near-infrared spectra. *Chemom. Intell. Lab. Syst.* 44 (1), 175–185 (<http://www.sciencedirect.com/science/article/pii/S0169743998001099>).
- Wolfrum, E.J., Payne, C., Schwartz, A., Jacobs, J., Kressin, R.W., 2020. A performance comparison of low-cost near-infrared (NIR) spectrometers to a conventional laboratory spectrometer for rapid biomass compositional analysis. *Bioenergy Res.* 13 (4), 1121–1129. <https://doi.org/10.1007/s12155-020-10135-6>.
- Xu, Y., Liu, J., Sun, Y., Chen, S., Miao, X., 2023. Fast detection of volatile fatty acids in biogas slurry using NIR spectroscopy combined with feature wavelength selection. *Sci. Total Environ.* 857 (Pt 1), 159282 <https://doi.org/10.1016/j.scitotenv.2022.159282>.
- Zhang, H., Schroder, J., 2014. Animal manure production and utilization in the US. In: He, Z., Zhang, H. (Eds.), *Applied Manure and Nutrient Chemistry for Sustainable Agriculture and Environment*. Springer, Netherlands, Dordrecht, The Netherlands, pp. 1–21.
- Zhang, X., Davidson, E.A., Zou, T., Lassaletta, L., Quan, Z., Li, T., et al., 2020. Quantifying nutrient budgets for sustainable nutrient management. *Glob. Biogeochem. Cycles* 34 (3). <https://doi.org/10.1029/2018GB006060>.
- Zimmermann, A., Hartung, E., 2009. Abschlußbericht zum BLE-Projekt: Nährstoffgesteuerte Gülleausbringung: -05UM003-. Institut für Landwirtschaftliche Verfahrenstechnik. Christian-Albrechts-Universität zu Kiel.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67 (2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.