NIR associated to PLS and SVM for fast and non-destructive determination of C, N, P, and K contents in poultry litter

Flavia Chiamulera Borsatti Bedin [a], Mateus Vinicius Faust [b], Giovanni Alfredo Guarneri [b], Tangriani Simioni Assmann [a], Cintia Boeira Batista Lafay [c], Lisiane Fernandes Soares [d], Paulo Armando Victória de Oliveira [e], Larissa Macedo dos Santos-Tonial [c,*]

[a] Universidade Tecnológica Federal do Paraná (UTFPR), Programa de Pós-Graduação em Agronomia (PPGAG) - câmpus Pato Branco, PR, Brazil
[b] Universidade Tecnológica Federal do Paraná (UTFPR), Programa de Pós-Graduação em Engenharia Elétrica (PPGEE) - câmpus Pato Branco, PR, Brazil
[c] Universidade Tecnológica Federal do Paraná (UTFPR), Departamento Acadêmico de Química (DAQUI) - câmpus Pato Branco, PR, Brazil
[d] Universidade Tecnológica Federal do Paraná (UTFPR), Departamento Acadêmico de Agronomia (DAGRO) - câmpus Pato Branco, PR, Brazil
[e] Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) Suínos e Aves, Concórdia, SC, Brazil

ARTICLE INFO

ABSTRACT

Using near-infrared (NIR) spectroscopy for poultry litter characterization can be a rapid, non-destructive, and low-cost alternative. This study aims to estimate the C, N, P, and K content in poultry litter samples using for first time NIR spectroscopy. For these purposes, the building models were carried out using Partial Least Squares (PLS) and Support Vector Machines (SVM) methods. A total of 160 litter samples were analyzed in poultry houses of different rearing systems, seeking the highest possible variability in their chemical composition. NIR spectroscopy, combined with PLS and SVM methods, is an alternative method for non-destructive C, N, P, and K determination in poultry samples. The regression models using SVM provide better accuracy for all elements, laying the basis for the nonlinear regression approach's application. The K determination on poultry litter using NIR was possible only by the SVM model ($R^2 = 0.8620$ and RPD = 2.7330). Conclusively, the predictive ability was improved using the SVM method.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The near-infrared (NIR) spectroscopy has proved to be an interesting alternative for analyze samples of food [1,2,3], pharmaceuticals [4,5], soils [6,7], plants [8], agroindustrial compost [9], and other samples. Very few studies described the use of the NIR spectroscopy in analytical poultry litter characterization. The characterization of industrial waste by NIR spectroscopy is the highlight of this work, which is due to the poultry meat sector's potential importance in the Brazilian economy. As a consequence, there is a high demand for analysis of industrial waste from this sector.

Globally, Brazil represents one of the largest producers and exporters of poultry meat [10]. In 2017, 13,607,352 tons of chicken meat were produced in Brazil [11]. Many farmers use poultry manure as a soil amendment. However, the farmers apply poultry manure without knowing the nutrients contents of this material. Most countries use the nitrogen (N) content in soil for determining the amount of fertilizer to be applied. The knowledge of nutrients contents of poultry manure is vital to meet the crop nutritional demands and not cause environmental contamination from excess manure application.

In Brazil, chemical analyses of poultry manure are realized traditionally [12]. However, fast, non-destructive and low-cost analyses methodologies such as NIR spectroscopy [13,14] are not in use nowadays. Additionally, the characterization of the poultry manure by NIR spectroscopy allows the elimination of potentially dangerous reagents, as sulfuric acid applied in the carbon (C) (*Walkley-Black method*), N, phosphorus (P) and potassium (K) determination (*Kjeldahl method*), and the reduced residue generation such as sulphochromic solution [15].

The NIR spectroscopy is based on overtone and combination bands of the fundamental vibrations relating to C—H, N—H, O—H and C—O functional groups in the spectral range from 12,500 to 4000 cm^{-1} [16,17,18,19,20]. The specificity of vibrational effects in NIR spectroscopy, different from those observed in mid-infrared, creates a large amount of independent spectral information of high value for physical chemistry. Due to this, a large number of overlapping vibrational contributions influenced by anharmonic effects creates complex patterns of spectral dependencies, which in many cases, hinder our comprehension of NIR spectra. Besides that, NIR spectral analysis remains prone to ambiguities which translate to convoluted spectral changes [21].

* Corresponding author.
E-mail address: larissasantos@utfpr.edu.br (LM. dos Santos-Tonial).

Nevertheless, the analytic information present in NIR spectra is very complex, self-correlated, and usually not promptly available for analytic purposes. Most of the time, the information is dispersed all over the spectrum [22]. Thus, it is necessary to use chemometric methods to process information [16,23].

With the use of multivariate calibration, it is possible to create a regression curve that searches the relationship between the multiple NIR spectroscopy data obtained from poultry bed samples and the concentration of the chemical elements of interest (C, N, P, and K). There is plenty of regression analysis techniques that study and allow us to realize this relationship among the variables. These techniques are distinguished mainly by using linear or non-linear methods [24].

Although carbon is not a nutrient, the reasons for including it in this study are two. First, it takes an essential role in nutrients availability, especially concerning N. High C/N ratio, in general, results in mineralization slowdown. Consequently, the availability of the nutrients, primarily N and P, is going to be retarded. Finally, the poultry litter has been often applied in the soil not only as a final disposal alternative but also in supplying the soil with carbon [25]. Thus, resuming carbon determination in poultry litter is essential to have an idea when nutrients are gone be available, and the impact of applications on soil physical and biological properties.

Thus, for the modeling of quantitative analysis curves of poultry litter residues, two techniques were chosen: *Partial Least Squares* (PLS) and *Support Vector Machines* (SVM). The first one belongs to the set of methods that use linear methods and is chosen because it is considered one of the standard techniques when using NIR spectroscopy [26]. The last one belongs to the group that uses non-linear methods, which may be more appropriate for quantitative analysis.

This study aims to estimate, in a non-destructive manner, C, N, P, and K content in poultry litter samples using for the first time NIR spectroscopy. For these purposes, the models building were carried out using PLS and SVM methods.

2. Materials and methods

2.1. Poultry litter samples

A set of 160 poultry litter samples were collected between October 2014 and September 2015 in 72 poultry farms of 40 properties. The properties are localized in the state of Santa Catarina and Paraná. Ten litter samples were collected randomly from each aviary from where one composite sample was obtained, homogenized, packed, and identified. In total, approximately 4000 g of each litter samples. The samples analyzed comprised composed of samples, collected from ten random points, mixed and homogenized. The substrate used as poultry litter was wood shaving. The broiler production system were: (1) *Broiler*

(42d, 3 kg) — the most tradable in Brazil; (2) *Griller* (28d, 1.5 kg) — for export; (3) *Fowl* (before reproduction phase) — the housing of these chickens was 1 day until 23 weeks; (4) *Hen* — the duration of housing of these hens was 42 weeks. The litter collection was achieved in the different stages of the development of chickens.

2.2. Sample preparation

Before NIR analysis, the samples were dried in an oven at 40 °C for 7 days, ground, and sieved through a 20 mesh steel sieve. After, the dry samples stored were hermetically sealed.

2.3. Traditional analysis

The C concentration was determined using the elemental analyzer. The total N content was determined by the Kjeldahl method after sulfuric digestion of the samples. The P and K were determined using ultraviolet-visible absorption spectroscopy and flame photometry, respectively [27].

2.4. NIR spectra collecting

NIR spectra of 20 cm^3 of the sample were recorded in diffuse reflectance mode with the integrating sphere and sampling rotator using a Bruker MPA FT-NIR spectrometer (Bruker Optics Inc., Ettlingen, Germany). The system was operated using Spectral Acquisition and Processing Software (OPUS 7.2, Bruker Optics, Germany). Data were recorded from 12000 to 4000 cm^{-1} at a resolution of 16 cm^{-1}, and 64 scans per spectrum. Three replicate spectra were collected for each sample.

2.5. Regression models

The process of modeling the regression curves is defined in two steps: *calibration* and *validation*. The first one, also called training, is the step in which the regression method is established. The last one, known as *test*, is the moment in which the predictions are made with the model proposed in the previous step, comparing the predicted values with the measured results of the samples to validate the constructed model. The construction of the PLS and SVM regression models proposed in this work followed the methodology shown in Fig. 1. All these procedures were developed using the MATLAB software (MathWorks, Natick, USA) with libraries and additional functions not available in its basic version.
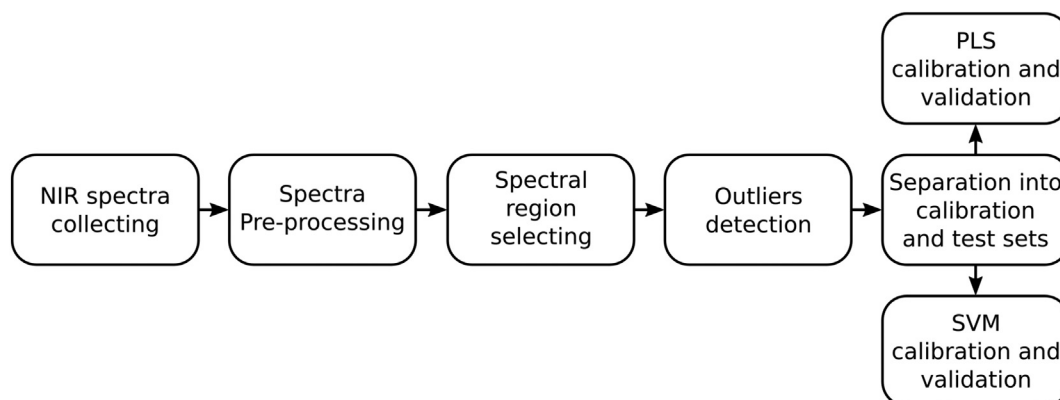

Fig. 1. Diagram showing the methodology used to create the regression models for C, N, P, and K by PLS and SVM.

2.5.1. Spectra preprocessing

The first step after collecting the spectra of the samples was preprocessing. In this work, two techniques were used for spectra preprocessing: mean centering and applying the differentiation of Savitzky-Golay (SG) [28].

Mean centering is commonly used as the first step in preprocessing. Each variable is subtracted from the mean value of all bins of a NIR spectrum, for each recorded band region as defined in [29]. This practice aims to make the variables to be interpreted according to the variation concerning the mean [30], making clear the correlation between them. We implemented this procedure in MATLAB from the definition given by [29].

The SG differentiation has the objective of increasing the signal referring to the absorption of energy about the noises present in the spectrum. Thus, it is possibly better to assimilate the most present infrared band region during spectrometry, since the peaks are typical to this process. However, when applied, it may be noted that as the height of the spectral wave rises, the noise is also high, though less expressive. Therefore, the addition of another technique, such as mean centering, can help to achieve better results. To perform this step, we used the function deriv from the *Data processing package for modeling with the NIR data* developed by Yonghuan [31].

2.5.2. Spectral region selecting

After the spectral preprocessing, the noises present in the spectrum, which do not bring information relevant to the models' construction, may be amplified. Thus, the procedure for selecting the spectral region is indicated to minimize the effects caused by the noise. In the literature, it is possible to find two techniques that are popularly used to find the best wavelength region within the total spectrum. The first is the region-by-region assignment of the estimated element [32]. However, this technique does not have the desired outcome for the spectra treated in this work due to the effect of overlaps and combinations.

To optimize the choice of the best spectral region for a better calibration of the models, Nørgaard et al. [33] introduced an algorithm specially designed for this stage, interval PLS (iPLS). The iPLS method splits the spectrum into subspaces of equal size, where for each of these spaces. Then, a "reduced" version of a PLS regression is performed by cross-validation. Thus, the region of the smallest prediction error is selected because this range represents the region in which the observed component has the most significant influence on the spectroscopic characterization [34]. We used the iPLS function available in *The iToolbox for MATLAB* that implements the iPLS method [35].

2.5.3. Outliers detection

An outlier is defined as a sample, from a data set, which does not belong to the studied model [36]. Thus, we can consider as outliers the spectra which show a behavior very different from the others, or that its measure of concentration is far from the other of the set. To best optimize the data before the calibration process is carried out, it is advisable to check and, consequently, to remove the outliers.

There are several techniques for outliers detection. However, we have chosen the statistical method of *Monte Carlo* [37]. This method's choice was due to its effectiveness in other papers involving the study of spectroscopy, mainly in the NIR region [32]. We used the mcs function available in *Tomcat: the MATLAB toolbox for multivariate calibration techniques* [38].

2.5.4. Separation into calibration and test sets

The procedures for calibrating the regression models by PLS and SVM techniques need to divide the set of poultry farm bed samples into two groups: *training* and *test* sets. The first one, which comprises the majority of samples, is the set used to model the regression curves. The last one, which is composed of the remainder of the samples, is used to carry out the verification of the designed model. From this, it is possible to observe and analyze prediction errors' values, determining if the methods used resulted in a fair prediction curve.

However, to obtain a good regression model after training a PLS and SVM machines, it is necessary to ensure that the training set samples contain variables that represent the whole set [36]. Kennard and Stone [39] developed an algorithm that is widespread in chemometrics to solve this problem. The Kennard-Stone (KS) algorithm performs the selection of samples for the calibration set according to the spatial distance between them [40]. We carried out the separation of the sets by the KS algorithm using the kenstone function available in *Tomcat: the MATLAB toolbox for multivariate calibration techniques* [38].

2.5.5. PLS calibration

The method known as PLS was conceived by Herman Wold [41] as an econometric technique. However, it was in the chemical field that this technique became popular, being considered the usual method when searching the calibration of spectrometry [42]. The result of PLS regression is a robust prediction model that does not undergo significant changes when new samples are added or removed from the training set, and by spectral errors either. However, PLS is suitable for the calibration of models in which the relations between input and output variables are linear [43]. To calibrate the PLS regression models of this work in MATLAB, we use the library libPLS (version 1.95) [44].

Before creating a model by the PLS method, it is necessary to identify how many *latent variables* are needed to define it. The choice of an optimal amount of latent variables is essential for the PLS model to generalize on inputs never presented during the training process. When we set fewer latent variables than the optimal value, we get an *underfitting* problem, which causes a low fidelity model. If the number is higher than the optimal one, the *overfitting* issue appears, responsible for raising the interference of some noises in the final model, causing a model sensitive to some unforeseen distortions [36]. In PLS model calibration, the leave-one-out cross-validation optimizes the number of latent variables. The optimum number of latent variables corresponds to the lowest RMSEC values. We got ten latent variables after applying this procedure.

2.5.6. SVM calibration

The SVM is a highly flexible and powerful computational algorithm for solving machine learning problems [45]. The SVM was created in the 1990s, initially for the development of classification models, its primary focus being the binary classification. However, it was quickly adapted to solve regression problems, where this technique presented a robust model that performs a minimization of the effects caused by outliers.

Currently, its use has been growing in the middle of the work involving applications with real data, these being mostly of non-linear nature, thus expanding the capacity of the previously developed prediction models, to which only linear techniques were applied [46]. The solution proposed by SVM to work with non-linear data was the addition of the *kernel* function, which maps the variables to higher dimensions, where it is possible to characterize the same data linearly [47]. Another advantage that comes from the use of the model by SVM is its flexibility. With the introduction of a gap variable, or also called *cost coefficient* (C), it is possible to classify variables that are considered outliers that did not present characteristics that contributed to the development of the model. The library libSVM (version 3.22) [48] was used to calibrate the SVM models in MATLAB.

Like PLS, the SVM also needs to set some parameters before the calibration process. These parameters are the C and the type of kernel with its respective factors. However, since it is necessary to determine two or more parameters, the technique called *grid search* is indicated [49]. This procedure showed the Gaussian kernel (RBF) as more suitable for our application. Using the kernel RBF adds the γ parameter as a user choice during SVM training. The values of C and γ chosen after applying the grid search are displayed in Table 1.

Table 1
Parameter values defined for the SVM training by the grid search method. C is the cost coefficient, and γ is the parameter of the RBF kernel.

Element	C	γ
Nitrogen	4616	5653
Carbon	26,760	8144
Phosphorus	469,200	150,900
Potassium	29,125,000	21,668

2.6. Statistical parameters of models

The calibrated PLS and SVM regression models for N, P, K, and C contends in poultry litter samples were compared. The statistical parameters used to compare the performances of models were the coefficient of determination (R^2), the root-mean-square error (RMSE) and the root-mean-square error (RMSE) defined in Eqs. (1), (2) and (3), respectively.

$$R^2 = 1 - \frac{\sum_{m=1}^{M}(y_i - \widehat{y}_m)^2}{\sum_{m=1}^{M}(y_m - \overline{y})^2}, \tag{1}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(y_m - \widehat{y}_m)^2}, \tag{2}$$

$$RPD = \frac{s_y}{RMSE} \tag{3}$$

where, M is the number of samples, y_m is the element content value measured for a sample m by traditional analysis, $\overline{y}$ is the mean of all samples, $\widehat{y}_m$ is the predicted value of the sample m and, s_y is the sample standard deviation of all samples y. The RMSE value was calculated for all samples in the calibration set (RMSEC) and all samples in the validation set (RMSEP).

3. Results

3.1. Nutrient content in poultry manure

The minimum and maximum values of C, N, P, and K observed in the 160 poultry litter samples collected in 72 poultry farms from Santa Catarina, and Paraná states are presented on Table 2.

3.2. Spectral interpretation

The NIR spectra of 160 poultry litter samples without preprocessing are shown in Fig. 2.

3.3. Preprocessing

Table 3 presents the results for the spectral region selection, outliers detection, and calibration and validation sets separation according to Sections 2.5.2 to 2.5.4 to build the N, P, K and, C estimation models.

Table 2
The carbon, nitrogen, phosphorus, and potassium maximum, minimum, average, and standard deviation values in the poultry litter samples.

	Carbon	Nitrogen	Phosphorus	Potassium
	g kg^{-1}	g kg^{-1}	mg kg^{-1}	mg kg^{-1}
Minimum value	206.1	1.1	130	578
Maximum value	464.1	42.3	25,800	33,700
Average	298.8	22.9	9790	17,100
Standard deviation	51.7	7.2	5257	6330

$n = 160$.

3.4. Histograms

The histograms showing the distribution of C, N, P, and K contents in poultry litter samples collected in 72 poultry farms from Santa Catarina and Paraná (Fig. 3).

3.5. Predicted vs. measured

Scatter plots of NIR predicted vs. measured values of C, N, P, and K using PLS and SVM are presented in (Fig. 4).

3.6. Statistical parameters of models

Tables 4 and 5 summarize the ability of NIR spectroscopy through PLS and SVM methods to predict the N, P, K and C contend in poultry litter. In this work, the performance of the models obtained was evaluated using the R^2, RPD, RMSEC, and RMSEP values.

4. Discussion

4.1. Poultry litter composition

The chemical composition of poultry manure varies because of several factors: the source of manure, feed of animals, age and condition of animal, storage, and handling of manure and litter used. If the real value is not knowing, this variation can cause severe issues such as plant under-nutrition or environmental contamination.

On our samples the C values ranged from 206.1 to 464.1 g kg^{-1}, with an average value of 298.8 g kg^{-1}. The N presents values ranging from 1.1 to 42.3 g kg^{-1}, with an average value of 22.9 g kg^{-1}. In 60 samples of broiler litter, Qafoku et al. [50] reported N concentrations from 23.8 to 47 g kg^{-1}. P values were between 130 and 25,800 mg kg^{-1}, with an average value of 9793 mg kg^{-1} and K values ranging from 578 to 33,714 mg kg^{-1}, with an average value of 17,087 mg kg^{-1}.

It was observed that there is a differentiation in the chemical composition of the litter according to the type of breeding used, where residues of poultry farms of matrices (rearing and egg production) presented levels of P that differentiate them from the others and that the systems of raising of chicken (*Griller* and *Broiler*) present more significant differentiation as a function of the K contents. Poultry litter is widely used as an excellent source of P fertilizer for agricultural crops and pasture in United States [51,52], Brazil [53], and other countries.

Therefore, if this material is used in agriculture without its proper chemical characterization, the risk of environmental contamination and nutritional deficiency of plants due to application with nutritional rates below that demanded by them is imminent.
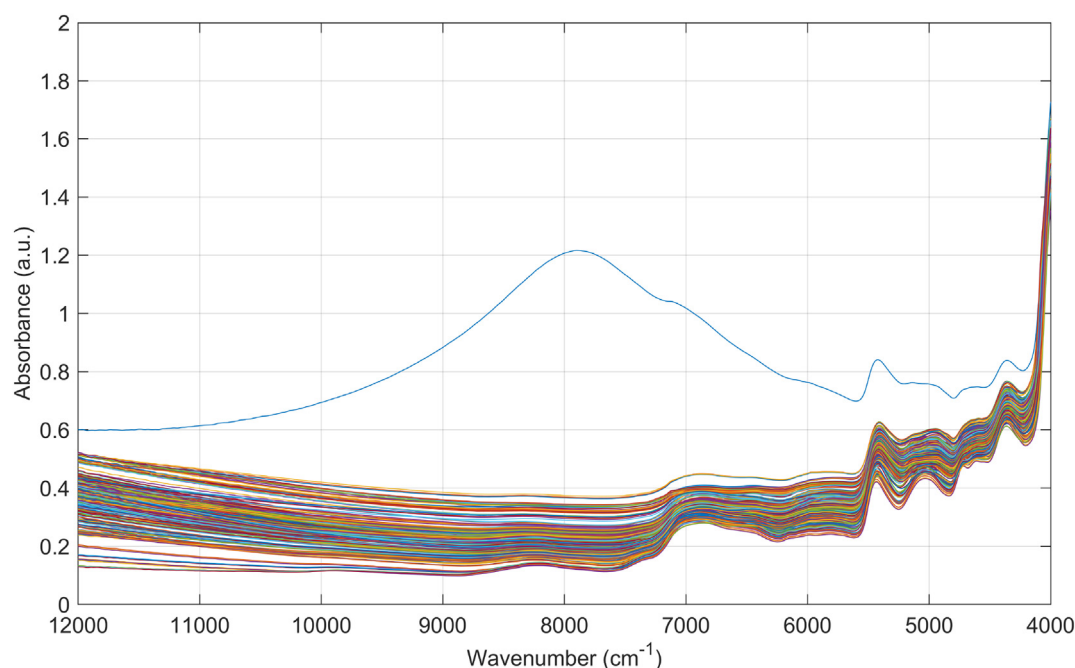
4.2. Spectral interpretation

In general, the NIR spectra of 160 poultry litter samples are very similar, presenting systematic variations of the baseline (Fig. 2).

The spectral region between 12,000 and 9000 cm^{-1} did not present any important information; that is, there was no significant absorption since no absorbance variations were observed in this wave interval. However, all samples presented important absorbance bands around 8300, 6800, 5700, 5200, 4700 and 4200 cm^{-1} (Fig. 2).

Similar results were obtained by [32] for organic waste. These signals can be related to the presence of functional groups such as C—H, O—H, and N—H chemical bonds [32]. The region from 6100 to 5600 cm^{-1} contains first overtones of the C—H stretching modes of CH$_3$, CH$_2$, and aromatic CH groups [54]. The signal located around 8300 cm^{-1} is associated with overtones of C—H stretch vibration. The peaks around at 6900 cm^{-1} (first overtone with O—H stretching) and 5150 cm^{-1} (combination band of O—H stretching and O—H deformation), are usually observed in fresh samples of organic matrices [9], and can be attributed to water [3]. Thus, owing to the overlap of bands, quantitative

Fig. 2. NIR spectra of 160 poultry litter without preprocessing. The blue line is the spectrum of a sample that has been detected as an outlier, as described in Section 2.5.3.

predictions are difficult by direct interpretations from the NIR spectra, and multivariate analyses are needed to discern the response of soil properties from spectral characteristics [55].

4.3. Spectral region selecting

The regions above 7986 cm^{-1} did not significantly improve the performance of the C, N, and P models. Similar results were observed to the region above 5583 cm^{-1} and the K model (Table 3).

4.4. Histograms interpretation

The histogram of the C, N, P, and K contents shows the distribution of these elements in the litter poultry samples. The histogram analysis method was applied to assess the homogeneity of the poultry litter samples used in this study.

The C was the element that presented smaller variability in the poultry litter samples. The results show that most samples have C content between 240 and 350 g kg^{-1} (Fig. 3(a) and (b)).

The N histogram indicates that the most litter poultry samples applied in this work presents N content above 17.5 mg kg^{-1}. The result demonstrated that N content distribution was relatively heterogeneous in the poultry litter samples, it is due to the small number of samples with N contents <17.5 g kg^{-1} and >32.5 g kg^{-1} (Fig. 3(c) and (d)).

Heterogeneous distributions were also observed for P and K contents. The most poultry litter samples presented P content between

7000 and 14,000 mg kg^{-1} (Fig. 3(e) and (f)), and K content between 10,000 and 21,000 mg kg^{-1} (Fig. 3(g) and (h)).

The heterogeneity observed in the C, N, P, and K contents of the poultry litter samples may also be due to the variability in the number of lots and animals. In general, the C, N, P, and K contents in the poultry litter samples confirm the potential of this residue as organic fertilizer [56].

4.5. Predicted vs. measured

Based on Fig. 4, we found that the PLS and SVM regression models have good results for estimating C, N, P, and K content in poultry litter samples. The best R^2 results were obtained between the measured and predicted values using the SVM model (Tables 4 and 5). The solid red line in the graphs in Fig. 4 indicates an individual correspondence between predicted and measured values. The closer to this line, the more sample points, the more accurate the model [20].

The SVM method can model linear and non-linear relationships between NIR spectral data and chemical content data. Liang et al. [57] states that the SVM prediction models are a suitable tool for non-linear regression analysis. Looking at Fig. 4, where the SVM performs is slightly better than the PLS method for C, N, and P content, and it is much better for K content.

The better performance obtained by the SVM method in estimating C, N, P, and K contents may be associated with several factors involving the non-linearity of the detection system and/or characteristics of the sample constituents [26].

According to Pasquini [26], SVM involves learning algorithms based on statistical learning theory and has been introduced in chemometrics recently with success in applications using NIR spectroscopy with superior performance to reference algorithms such as PLS [58].

These results suggest that the SVM method as an emerging alternative for NIR calibration for C, N, P, and K quantification in poultry litter. This method may present some advantages when non-linearity between the spectral data and the quantitative information of interest exists.

Many substances were showing a linear relationship between concentration and absorbance, although others present a non-linear relationship. The P and K content can be predicted by NIR spectroscopy, although it does not have specific vibrational bands in the NIRS.

Table 3

Preprocessing results. The "Spectrum's region" column shows the result of the procedure describes in Section 2.5.2; the "Useless" column presents the outliers detected according to Section 2.5.3; and the last two columns indicate the samples' quantity in calibration and validation sets.

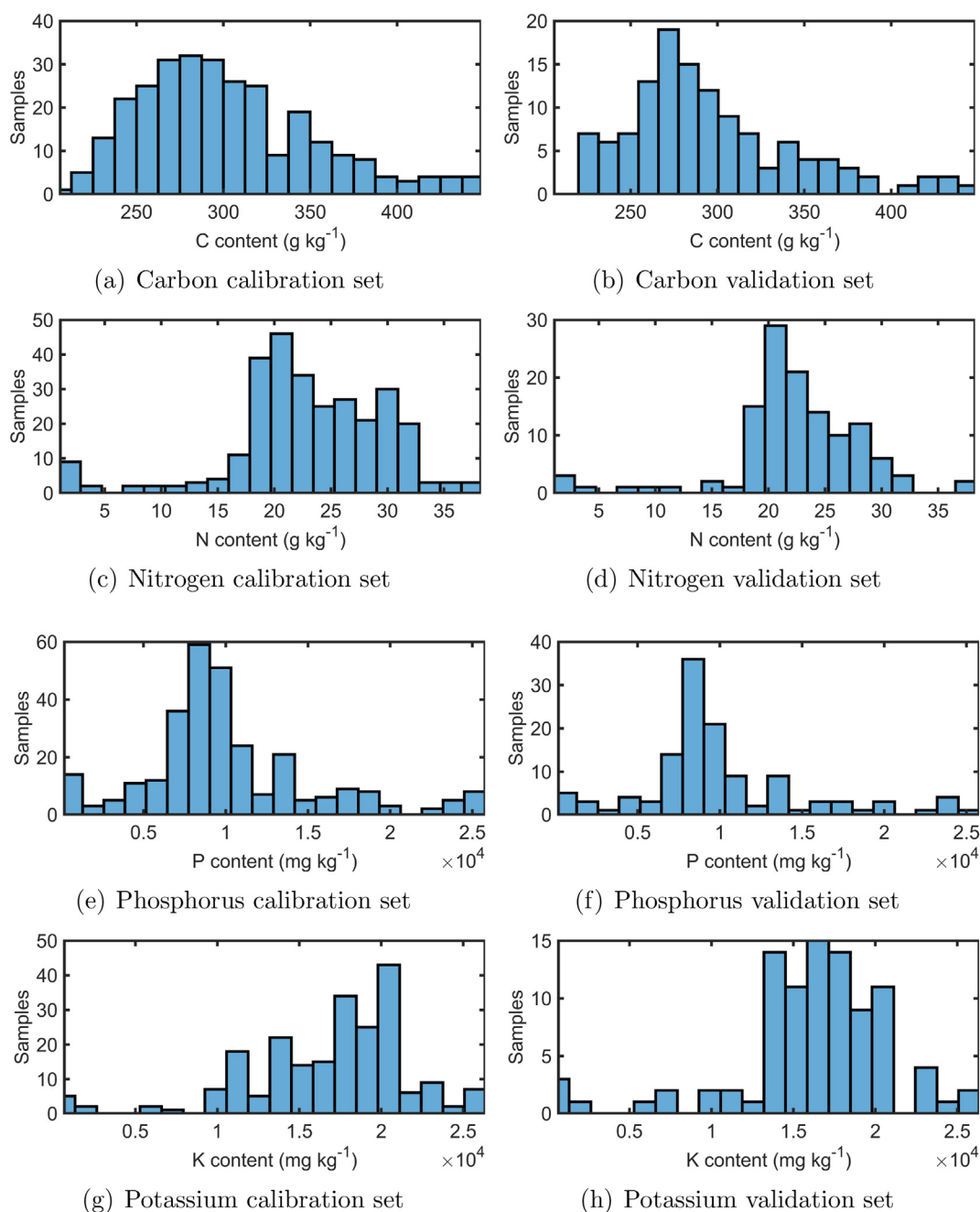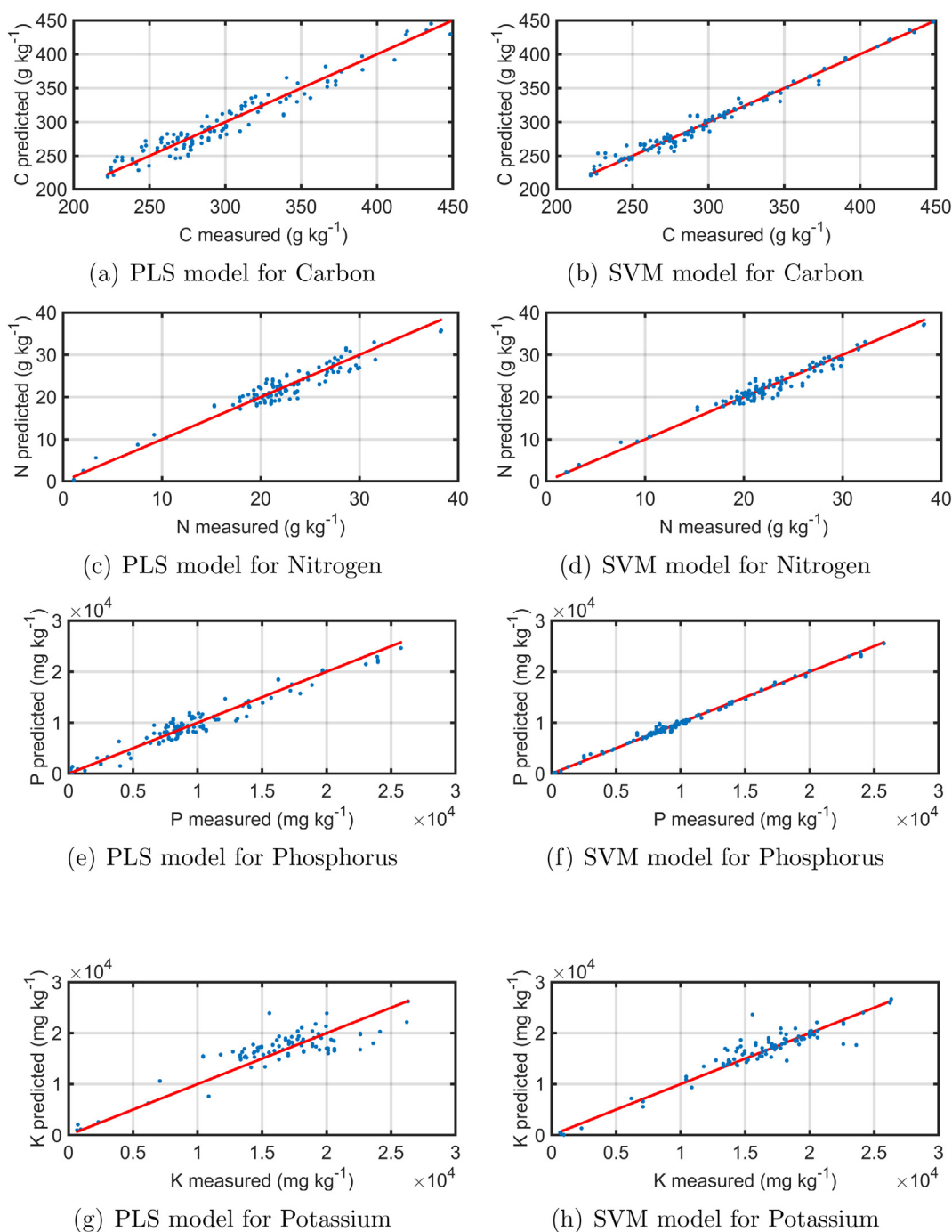| Element	Spectrum's region (cm^{-1})	Useless	Calib. set	Valid. set
Carbon	7986 to 4000	70	287	123
Nitrogen	7986 to 4000	72	286	122
Phosphorus	7986 to 4000	67	289	124
Potassium	5583 to 4000	170	217	93

Fig. 3. Histograms of the (a–b) Carbon, (c–d) Nitrogen, (e–f) Phosphorus, and (g–h) Potassium contents in the poultry litter samples.

The K and P models built using NIR spectroscopy data are indirect models for quantifying these elements. The possibility of predicting the concentration of K and P with accuracy in poultry litter samples has been attributed to the association of these elements with the organic components (organic complexes) of the poultry litter constituents.

In general, the poultry litter consists of a mix of bedding material (wood chips), excreta (feces and urine), waste feed, and feathers [59]. It is usually composed of organic compounds such as cellulose, hemicellulose, and lignin [60]. Thus, the P and K could be indirectly detected through their linkage or interaction with organic compounds in poultry litter. Therefore, a quantitative determination of K and P by NIR spectroscopy is achieved by "indirect analysis" using the interaction of these elements and organic groups [61].

The indirect analysis by NIR spectroscopy has been demonstrated in the literature. Abdi et al. [62] showed that the NIR spectroscopy could be used for the indirect determination of P concentrations in soils from Canada. Success in predictions of P may be attributed to their relationship to soil organic matter. de Oliveira et al. [63] showed that NIR spectroscopy could be successfully applied for the determination of micro (B, Cu, Fe, Mn, and Zn) and macronutrients (N, P, K, Ca, and Mg) in Brachiaria samples. Coutinho et al. [64] demonstrated the acceptable performance of NIR spectroscopy for estimating P and K availability in soils.

According to Chang et al. [65], NIR spectroscopy can be used to simultaneously estimate several secondary soil properties (such as extractable Ca, K, Mn, and others). It is possible due to the correlations of these properties with soil properties having a primary response in the

(a) PLS model for Carbon

(b) SVM model for Carbon

(c) PLS model for Nitrogen

(d) SVM model for Nitrogen

(e) PLS model for Phosphorus

(f) SVM model for Phosphorus

(g) PLS model for Potassium

(h) SVM model for Potassium

Fig. 4. Graphics of the PLS and SVM regression models after the validation process comparing the actual values to the predicted ones.

Table 4
Statistical evaluation parameters of the regression models were designed by the PLS technique for Carbon, Nitrogen, Phosphorus, and Potassium's determination.

Element	Calibration set			Validation set		
	R^2	RPD	RMSEC	R^2	RPD	RMSEP
C	0.9282	3.7382	1.3400	0.9294	3.7790	1.2981
N	0.9403	4.1005	0.1672	0.9123	3.3916	0.1752
P	0.9407	3.9897	1276.2	0.9344	3.7701	1271.5
K	0.7284	1.6413	2550.3	0.7206	1.6927	2537.0

Table 5
Statistical evaluation parameters of the regression models were designed by the SVM technique for Carbon, Nitrogen, Phosphorus and, Potassium's determination.

Element	Calibration set			Validation set		
	R^2	RPD	RMSEC	R^2	RPD	RMSEP
C	0.9815	7.3611	0.6805	0.9733	6.1468	0.7981
N	0.9793	6.9571	0.0986	0.9560	4.7884	0.1241
P	0.9975	20.081	260.56	0.9926	11.446	426.01
K	0.9361	3.8335	1236.5	0.8620	2.7330	1783.1

NIR region. Abdi [66] found a moderately useful prediction for total P ($0.70 \leq R^2 \leq 0.80$) by NIR spectroscopy for 192 gravely sandy loam soil samples from an experiment site in eastern Canada, due in part to its relationship to, especially active soil OM.

4.6. Statistical parameters of models

The R^2 is often used to judge the adequacy of a regression model, related to the amount of variability in the data explained or accounted for by the regression model. Its value ranges from 0 to 1. If R^2 is 1, the model accounts for 100% of the variability in the data [67]. According to Williams [68], a value for R^2 between 0.66 and 0.81 indicates approximate quantitative predictions. Whereas, a value for R^2 between 0.82 and 0.90 reveals good prediction. Calibration models having a value for R^2 above 0.91 are considered to be excellent.

The prediction capacity was excellent for C, N, and P by the PLS technique with $R^2 \geq 0.9282$ (calibration set), and $R^2 \geq 0.9123$ (validation set) (Table 4). Similar results were obtained by the SVM technique with $R^2 \geq 0.9793$ (calibration set), and $R^2 \geq 0.9560$ (validation set) (Table 5). Additionally, an excellent regression model should have high RPD and low RMSEC, and RMSEP.

There are in the literature several model assessment methodologies using RPD values [55,62,65,69,70,71,72,73,74]. Zornoza et al. [55] suggested that RPD values between 1.5 and 2.0 are insufficient for applications; values between 2.0 and 2.5 indicate approximate quantitative predictions; values between 2.5 and 3.0 provide good prediction models and; above 3.0 are considered excellent prediction models. Therefore, the higher RPD value indicates the best predictive capacity of the model.

The performance of C, N, and P regression models was showed RPD values higher than 3.3916 (Tables 4 and 5). These RPD values indicate an excellent predictive ability of models for quantitative estimating C, N, and P contents in poultry litter. The regression model for K prediction using PLS showed RPD values around 1.6, indicating insufficient for applications. However, the SVM model for estimating K showed RDP = 2.7330, which indicates that it is a good predictor.

In this study, the higher RPD values were observed for all the models using SVM, indicating higher prediction ability of these models in the C, N, P, and K contents determination in poultry litter samples (Table 5) [20].

The RMSEC indicates the modeling error or calibration variance. The RMSEP shows the prediction error or validation variance, thus the imprecision (quality) of the regression model [75].

The lower values of RMSEC were obtained by the SVM technique. This behavior was observed for all elements. Similar results were found for RMSEP. The SVM technique gives better results for modeling C, N, P, and K contents in poultry litter samples.

5. Conclusions

The poultry litter samples presented variation in the C, N, P, and K contents. This differentiation in the litter's chemical composition is mainly due to the type of breeding used (Broiler, Griller, Fowl, and Hen). The number of lots and animals and the localization of poultry farms may also have contributed to the heterogeneity of samples analyzed. This result confirms the importance of the chemical characterization of the poultry litter before the final disposal.

In general, the NIR spectra of poultry litter samples are very similar. The samples presented meaningful absorbance bands around 8300, 6800, 5700, 5200, 4700, and 4200 cm^{-1}. The regions above 7986 cm^{-1} did not provide any significant improvement in the performance of the C, N, and P models, and the zone above 5583 cm^{-1} to the K model.

In general, NIR spectroscopy, combined with PLS and SVM methods, is confirmed to be an alternative for non-destructive C, N, P, and K determination in poultry litter samples within an excellent or acceptable range of predictability according to R_2, RPD, RMSEC, and RMSEP.

The regression models using SVM provides better accuracy for all elements, laying the basis for the application of the non-linear regression approach. Only the employment of the SVM method made it possible the use of NIR for K determination on poultry litter.

It is also noticeable that the prediction capability of the C, N, P, and K contents in poultry litter by NIR spectroscopy, PLS, and SVM are reported for the first time in the literature.

CRediT authorship contribution statement

Flavia Chiamulera Borsatti Bedin: Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing. **Mateus Vinicius Faust:** Methodology, Validation, Formal analysis, Investigation, Data curation, Visualization. **Giovanni Alfredo Guarneri:** Conceptualization, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Tangriani Simioni Assmann:** Conceptualization, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Supervision. **Cíntia Boeira Batista Lafay:** Methodology, Validation, Formal analysis, Investigation, Resources, Writing - original draft. **Lisiane Fernandes Soares:** Resources. **Paulo Armando Victória de Oliveira:** Resources. **Larissa Macedo dos Santos-Tonial:** Conceptualization, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Supervision, Project administration.

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.

References

[1] J.U. Porep, D.R. Kammerer, R. Carle, On-line application of near infrared (NIR) spectroscopy in food production, Trends Food Sci. Technol. 46 (2015) 211–230.

[2] P. Sirisomboon, NIR spectroscopy for quality evaluation of fruits and vegetables, Mater. Today: Proceedings 5 (2018) 22481–22486.

[3] J. Véstia, J.M. Barroso, H. Ferreira, L. Gaspar, A.E. Rato, Predicting calcium in grape must and base wine by FT-NIR spectroscopy, Food Chem. 276 (2019) 71–76.

[4] O.Y. Rodionova, A.V. Titova, N.A. Demkin, K.S. Balyklova, A.L. Pomerantsev, Qualitative and quantitative analysis of counterfeit fluconazole capsules: a non-invasive approach using NIR spectroscopy and chemometrics, Talanta 195 (2019) 662–667.

[5] L.M. Kandpal, J. Tewari, N. Gopinathan, J. Stolee, R. Strong, P. Boulas, B.K. Cho, Quality assessment of pharmaceutical tablet samples using Fourier transform near infrared spectroscopy and multivariate analysis, Infrared Phys. Technol. 85 (2017) 300–306.

[6] J.B. Carra, M. Fabris, J. Dieckow, O.R. Brito, P.R.S. Vendrame, L. Macedo Dos Santos Tonial, Near-infrared spectroscopy coupled with chemometrics tools: a rapid and non-destructive alternative on soil evaluation, Commun. Soil Sci. Plant Anal. 50 (2019) 421–434.

[7] R. Recena, V.M. Fernández-Cabanás, A. Delgado, Soil fertility assessment by Vis-NIR spectroscopy: predicting soil functioning rather than availability indices, Geoderma 337 (2019) 368–374.

[8] M.T. Sánchez, J.A. Entrenas, I. Torres, M. Vega, D. Pérez-Marín, Monitoring texture and other quality parameters in spinach plants using NIR spectroscopy, Comput. Electron. Agric. 155 (2018) 446–452.

[9] L. Galvez-Sola, R. Moral, M.D. Perez-Murcia, A. Perez-Espinosa, M.A. Bustamante, E. Martinez-Sabater, C. Paredes, The potential of near infrared reflectance spectroscopy (NIRS) for the estimation of agroindustrial compost quality, Sci. Total Environ. 408 (2010) 1414–1421.

[10] D. Caro, S.J. Davis, E. Kebreab, F. Mitloehner, Land-use change emissions from soybean feed embodied in brazilian pork and poultry meat, J. Clean. Prod. 172 (2018) 2646–2654.

[11] FAO, Faostat online database, http://faostat3.fao.org/download/Q/QL/E, 2019. Accessed: Set 2019.

[12] L.J. Chen, L. Xing, L.J. Han, Quantitative determination of nutrient content in poultry manure by near infrared spectroscopy based on artificial neural networks, Poult. Sci. 88 (2009) 2496–2503.

[13] M. Kamruzzaman, D.W. Sun, G. ElMasry, P. Allen, Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis, Talanta 103 (2013) 130–136.

[14] M. Blanco, I. Villarroya, NIR spectroscopy: a rapid-response analytical tool, TrAC Trends Anal. Chem. 21 (2002) 240–250.

[15] V.E. Dabkiewicz, S. de Mello Pereira Abrantes, R.J. Cassella, Development of a non-destructive method for determining protein nitrogen in a yellow fever vaccine by near infrared spectroscopy and multivariate calibration, Spectrochim. Acta A Mol. Biomol. Spectrosc. 201 (2018) 170–177.

[16] H. Cen, Y. He, Theory and application of near infrared reflectance spectroscopy in determination of food quality, Trends Food Sci. Technol. 18 (2007) 72–83.

[17] F. Zapata, M. Ferreiro-González, C. García-Ruiz, Interpreting the near infrared region of explosives, Spectrochim. Acta A Mol. Biomol. Spectrosc. 204 (2018) 81–87.

[18] Y. Huang, T.M. Rogers, M.A. Wenz, A.G. Cavinato, D.M. Mayes, G.E. Bledsoe, B.A. Rasco, Detection of sodium chloride in cured salmon roe by sw-nir spectroscopy, J. Agric. Food Chem. 49 (2001) 4161–4167.

[19] V.T. Svensson, H.H. Nielsen, R. Bro, Determination of the protein content in brine from salted herring using near-infrared spectroscopy, LWT Food Sci. Technol. 37 (2004) 803–809.

[20] P.C. Williams, Implementation of near-infrared technology, in: P.C. Williams, K. Norris (Eds.), Near-Infrared Technology in the Agricultural and Food Industries, 2 editionAmerican Association of Cereal Chemists 2001, pp. 145–169.

[21] K.B. Beć, C.W. Huck, Breakthrough potential in near-infrared spectroscopy: spectra simulation. A review of recent developments, Front. Chem. 7 (2019).

[22] C. Pasquini, Near infrared spectroscopy: a mature analytical technique with new perspectives–a review, Anal. Chim. Acta 1026 (2018) 8–36.

[23] L. Xu, Y.P. Zhou, L.J. Tang, H.L. Wu, J.H. Jiang, G.L. Shen, R.Q. Yu, Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration, Anal. Chim. Acta 616 (2008) 138–143.

[24] D.-W. Sun, Infrared Spectroscopy for Food Quality Analysis and Control, Academic Press, 2009.

[25] D.B. Watts, H.A. Torbert, S.A. Prior, G. Huluka, Long-term tillage and poultry litter impacts soil carbon and nitrogen mineralization and fertility, Soil Sci. Soc. Am. J. 74 (2010) 1239–1247.

[26] C. Pasquini, Near infrared spectroscopy: fundamentals, practical aspects and analytical applications, J. Braz. Chem. Soc. 14 (2003) 198–219.

[27] M.J. Tedesco, C. Gianello, C.A. Bissani, H. Bohnen, S.J. Volkweiss, et al., Análises de solo, plantas e outros materiais, volume 174, Ufrgs Porto Alegre 1995.

[28] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem. 36 (1964) 1627–1639.

[29] P.E. Gemperline, S. Brown, A. Walmsley, J. Kalivas, M. Hubert, R. Tauler, K. Booksh, M. Maeder, Y.M. Neuhold, C. Davidson, A. de Juan, B. Lavine, K. Stoyanov, Pratical Guide To Chemometrics, CRC Press, Boca Raton, 2006.

[30] B.M. Nicolai, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, J. Lammertyn, Nondestructive measurement of fruit and vegetables quality by means of NIR spectroscopy: a review, Postharvest Biol. Technol. 46 (2007) 99–118.

[31] Y. Yonghuan, Data Processing for Modeling with the NIR Data, 2014.

[32] M. Sisouane, M. Cascant, S. Tahiri, S. Garrigues, M.E. Krati, G.E.K. Boutchich, M. Cervera, M. de la Guardia, Prediction of organic carbon and total nitrogen contents in organic wastes and their composts by infrared spectroscopy and partial least square regression, Talanta 167 (2017) 352–358.

[33] L. Nørgaard, A. Saudland, J. Wagner, J. Nielsen, L. Munck, S. Engelsen, Interval partial least-squares regression (ipls): a comparative chemometric study with an example from near-infrared spectroscopy, Appl. Spectrosc. 54 (2000) 413–419.

[34] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in partial least squares regression, Chemom. Intell. Lab. Syst. 118 (2012) 62–69.

[35] L. Nørgaard, R. Leardi, The itoolbox for Matlab, 2005.

[36] S. Da-Wen, Infrared Spectroscopy for Food Quality Analysis and Control, 2009.

[37] Q.-S. Xu, Y.-Z. Liang, Monte carlo cross validation, Chemom. Intell. Lab. Syst. 56 (2001) 1–11.

[38] M. Daszykowski, S. Serneels, K. Kaczmarek, P. Van Espen, C. Croux, B. Walczak, Tomcat: a matlab toolbox for multivariate calibration techniques, Chemom. Intell. Lab. Syst. 85 (2007) 269–277.

[39] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148.

[40] M. Daszykowski, B. Walczak, D. Massart, Representative subset selection, Anal. Chim. Acta 468 (2002) 91–103.

[41] H. WOLD, Estimation of principal components and related models by iterative least squares, Multivar. Anal. (1966) 391–420.

[42] R.D. Tobias, An introduction to partial least squares regression, Proceedings of the Twentieth Annual SAS users group international conference (1995) 1250–1257.

[43] S. Wold, M. Sjöström, L. Eriksson, Pls-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130 PLS Methods.

[44] H.-D. Li, Q.-S. Xu, Y.-Z. Liang, libpls: an integrated library for partial least squares regression and discriminant analysis, Chemom. Intell. Lab. Syst. 176 (2018) 34–43.

[45] S.S. Haykin, et al., Neural Networks and Learning Machines/Simon Haykin, Prentice Hall, New York, 2009.

[46] A. Mammone, M. Turchi, N. Cristianini, Support vector machines, Wiley Interdiscip. Rev. Comput. Stat. 1 (2009) 283–289.

[47] W.S. Noble, What is a support vector machine? Nat. Biotechnol. 24 (2006) 1565.

[48] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 27:1–27:27.

[49] A. Aggarwal, T. Sahay, A. Bansal, M. Chandra, Grid search analysis of nu-svc for text-dependent speaker-identification, in: India Conference (INDICON), Annual IEEE, IEEE, 2015 1–5.

[50] O. Qafoku, M. Cabrera, W. Windham, N. Hill, Rapid methods to determine potentially mineralizable nitrogen in broiler litter, J. Environ. Qual. 30 (2001) 217–221.

[51] T.D. Ranatunga, S.S. Reddy, R.W. Taylor, Phosphorus distribution in soil aggregate size fractions in a poultry litter applied soil and potential environmental impacts, Geoderma 192 (2013) 446–452.

[52] A. Ashworth, P. Owens, F. Allen, Long-term cropping systems management influences soil strength and nutrient cycling, Geoderma 361 (2020) 114062.

[53] J.J. Frazão, V. de Melo Benites, J.V.S. Ribeiro, V.M. Pierobon, J. Lavres, Agronomic effectiveness of a granular poultry litter-derived organomineral phosphate fertilizer in tropical soils: soil phosphorus fractionation and plant responses, Geoderma 337 (2019) 582–593.

[54] H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise, Near-infrared Spectroscopy: Principles, Instruments, Applications, John Wiley & Sons, 2008.

[55] R. Zornoza, C. Guerrero, J. Mataix-Solera, K.M. Scow, V. Arcenegui, J. Mataix-Beneyto, Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils, Soil Biol. Biochem. 40 (2008) 1923–1930.

[56] F.S. Dalólio, J.N. da Silva, A.C.C. de Oliveira, I.d.F.F. Tinôco, R.C. Barbosa, M. de Oliveira Resende, L.F.T. Albino, S.T. Coelho, Poultry litter as biomass energy: a review and future perspectives, Renew. Sust. Energ. Rev. 76 (2017) 941–949.

[57] Y. Liang, Q.-S. Xu, H.-D. Li, D.-S. Cao, Support Vector Machines and Their Application in Chemistry and Biotechnology, CRC Press, 2016.

[58] J.C.L. Alves, R.J. Poppi, Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM), Talanta 104 (2013) 155–161.

[59] D. Lynch, A.M. Henihan, B. Bowen, D. Lynch, K. McDonnell, W. Kwapinski, J.J. Leahy, Utilisation of poultry litter as an energy feedstock, Biomass Bioenergy 49 (2013) 197–204.

[60] K. Wilkinson, E. Tee, R. Tomkins, G. Hepworth, R. Premier, Effect of heating and aging of poultry litter on the persistence of enteric bacteria, Poult. Sci. 90 (2011) 10–18.

[61] F. Liu, W. Cai, X. Shao, Selective determination of mercury (ii) ion in water by near-infrared diffuse reflection spectroscopy with the aid of preconcentration and multivariate calibration, Vib. Spectrosc. 68 (2013) 104–108.

[62] D. Abdi, B.J. Cade-Menun, N. Ziadi, G.F. Tremblay, L.É. Parent, Visible near infrared reflectance spectroscopy to predict soil phosphorus pools in chernozems of Saskatchewan, Canada, Geoderma Regional 7 (2016) 93–101.

[63] D.M. de Oliveira, L.M. Fontes, C. Pasquini, Comparing laser induced breakdown spectroscopy, near infrared spectroscopy, and their integration for simultaneous multi-elemental determination of micro-and macronutrients in vegetable samples, Anal. Chim. Acta 1062 (2019) 28–36.

[64] M.A. Coutinho, F.d.O. Alari, M.M. Ferreira, L.R. do Amaral, Influence of soil sample preparation on the quantification of npk content via spectroscopy, Geoderma 338 (2019) 401–409.

[65] C.-W. Chang, D.A. Laird, M.J. Mausbach, C.R. Hurburgh, Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties, Soil Sci. Soc. Am. J. 65 (2001) 480–490.

[66] H. Abdi, Partial least squares (pls) regression, Wiley Interdiscip. Rev. Comput. Stat. 2 (2010) 97–106.

[67] D.C. Montgomery, G.C. Runger, Applied statistics and probability for engineers, John Wiley & Sons, 2010.

[68] P. Williams, Near-infrared technology getting the best out of light, PDK Grain, Nanaimo, Canada, 2003.

[69] E. Di Iorio, L. Circelli, R. Lorenzetti, E.A. Costantini, S.P. Egendorf, C. Colombo, Estimation of andic properties from vis-nir diffuse reflectance spectroscopy for volcanic soil classification, CATENA 182 (2019) 104109.

[70] B.M. Nicolai, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, J. Lammertyn, Nondestructive measurement of fruit and vegetable quality by means of nir spectroscopy: A review, Postharvest Biol. Technol. 46 (2007) 99–118.

[71] D. Cozzolino, A. Moron, The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics, J. Agric. Sci. 140 (2003) 65–71.

[72] D. Malley, P. Williams, Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter, Environ. Sci. Technol. 31 (1997) 3461–3467.

[73] B. Dunn, G. Batten, H.G. Beecher, S. Ciavarella, The potential of near-infrared reflectance spectroscopy for soil analysis—a case study from the riverine plain of south-eastern australia, Aust. J. Exp. Agric. 42 (2002) 607–614.

[74] J. Moros, M.J. Martnez-Sánchez, C. Pérez-Sirvent, S. Garrigues, M. de la Guardia, Testing of the region of murcia soils by near infrared diffuse reflectance spectroscopy and chemometrics, Talanta 78 (2009) 388–398.

[75] E.D. Louw, K.I. Theron, Robust prediction models for quality parameters in Japanese plums (Prunus salicina L.) using NIR spectroscopy, Postharvest Biol. Technol. 58 (2010) 176–184.