



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Estimation of parameters in sewage sludge by near-infrared reflectance spectroscopy (NIRS) using several regression tools

Luis Galvez-Sola^a, Javier Morales^b, Asunción M. Mayoral^b, Concepción Paredes^a,
María A. Bustamante^{a,*}, Frutos C. Marhuenda-Egea^c, J. Xavier Barber^b, Raúl Moral^a

^a Department of Agrochemistry and Environment, Miguel Hernandez University, EPS-Orihuela, ctra. Beniel Km 3.2, 03312-Orihuela, Alicante, Spain

^b Applied Statistical Unit, Operations Research Center, Miguel Hernandez University, Elche, Alicante, Spain

^c Department of Agrochemistry and Biochemistry, University of Alicante. Ap. 99, E-03080 Alicante, Spain

ARTICLE INFO

Article history:

Received 31 August 2012

Received in revised form

30 January 2013

Accepted 5 February 2013

Available online 13 February 2013

Keywords:

NIRS

Biosolids

Partial least square regression (PLSR)

Penalized signal regression (PSR)

Chemical properties

Heavy metals

ABSTRACT

Sewage sludge application to agricultural soils is a common practice in several countries in the European Union. Nevertheless, the application dose constitutes an essential aspect that must be taken into account in order to minimize environmental impacts. In this study, near infrared reflectance spectroscopy (NIRS) was used to estimate in sewage sludge samples several parameters related to agronomic and environmental issues, such as the contents in organic matter, nitrogen and other nutrients, metals and carbon fractions, among others. In our study (using 380 biosolid samples), two regression models were fitted: the common partial least square regression (PLSR) and the penalized signal regression (PSR). Using PLSR, NIRS became a feasible tool to estimate several parameters with good goodness of fit, such as total organic matter, total organic carbon, total nitrogen, water-soluble carbon, extractable organic carbon, fulvic acid-like carbon, electrical conductivity, Mg, Fe and Cr, among other parameters, in sewage sludge samples. For parameters such as C/N ratio, humic acid-like carbon, humification index, the percentage of humic acid-like carbon, the polymerization ratio, P, K, Cu, Pb, Zn, Ni and Hg, the performance of NIRS calibrations developed with PLSR was not sufficiently good. Nevertheless, the use of PSR provided successful calibrations for all parameters.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The encouragement of the wastewater purification processes has aggravated the storage and management problems of the sludge generated after wastewater treatment. After reducing the moisture content and treating sewage sludge, when it is appropriate, the main destinations of this waste are the agricultural use, incineration with energy recovery and removal in refuse dump.

Sludge application to agricultural soils can imply several benefits, provided that this application is in accordance with the current legislation and a proper fertilization. For this, it is necessary to know the sludge composition and the crop needs to calculate the application dose. The incorporation of sewage sludge to agricultural soils must be done properly, paying special attention to the form and timing of the application, in order to

improve soil properties, such as soil structure or nutrient content and because of the positive effects on soil microbial activity. Soil microorganisms are key players in the turnover of plant nutrients and organic material and have important roles in maintaining soil structure [1]. But, despite the benefits of this application on agricultural soil, there are some disadvantages, for example, the risk for humans and animals due to the presence of xenobiotics and pathogenic microorganisms [2]. Furthermore, pollution problems may arise from toxic heavy metals that are mobilized into the soil solution and are either taken up by plants or transported in drainage water [3]. After the biosolids have been separated from wastewater, land applied sludges must be treated to reduce pathogens through one of a number of processes including anaerobic digestion, lime stabilization or composting. Each of these processes has effects on the fate of both pathogens and the organic contaminants in the sludge [4], so the contamination potential is lower after a treatment process.

Sewage sludge production in Spain was about 1205,124 t (in dry matter) in 2009, 82.6% of this amount being used for agricultural purposes [5]. The best strategy to decide the application rate should be based on the specific composition of the biosolids in each area, in order to verify the maximum sewage sludge application. However, in many cases, sewage sludge is

* Corresponding author. Tel.: +34 968 39 6395.

E-mail addresses: mabm@cebas.csic.es,
marian.bustamante@umh.es (M.A. Bustamante).

¹ Department of Soil and Water Conservation and Organic Waste Management, Centro de Edafología y Biología Aplicada del Segura, CSIC, PO Box 164, 30100 Murcia, Spain.

applied to soils without this information, due to the cost and time taken for a full characterization.

A promising alternative could be the estimation of different parameters in the sewage sludge samples, based on the use of near infrared reflectance spectroscopy (NIRS). There are several studies using NIRS for estimating some parameters in sewage sludge or in composted sewage sludge samples. Albrecht et al. [6] reported the use of NIRS technology to determine carbon and nitrogen contents in sewage sludge and green waste compost and Vergnoux et al. [7] studied physico-chemical parameters, organic matter, total organic carbon and concentrations of different nitrogen forms in sewage sludge compost. The heavy metal contents have been studied in sewage sludge samples [8] and in compost from sewage sludge [9,10]. Also, Galvez-Sola et al. [11] evaluated the nitrogen content in sewage sludge samples.

On the other hand, different mathematical tools, such as the use of multivariate calibration [12], have been considered to establish a model that helps to understand the signal obtained from NIRS and to optimize the output of information. Several aspects raised in this framework are non-trivial and include: [a] the number of regressors (p wavenumbers) can widely exceed training observations (n); [b] the regressors are highly correlated and thus, it appears a very ill-conditioned statistical problem [13]. There are two general approaches that can be used to solve these aspects: (1) a reduction of the regression bases, including partial least squares, principal components, or projection onto bases of splines [14] and (2) penalized regression, such as ridge regression or penalized signal regression [15]. With smooth spectra, penalized signal regression constitutes a very promising approach, since it forces the vector of regressor coefficients to vary smoothly with wavenumber.

Therefore, the aim of this study was to explore the potential of NIRS technique for the estimation of several parameters in sewage sludge samples, using different regression tools: the commonly used partial least square regression (PLSR) and the promising penalized signal regression (PSR).

2. Materials and methods

2.1. Sewage sludge samples

Three hundred and eighty sewage sludge samples were collected from 82 different wastewater treatment plants located in southern Spain, during the period 2001–2008. The accumulated equivalent population of the considered area is over 400,000 inhabitants. The treated volume of wastewater per year was about 200 hm³, with a sludge production of 80,000 t, approximately. All samples were dried at 50 °C, ground, dried again at 105 °C and stored prior to their analysis.

2.2. Analytical methods

The electrical conductivity (EC) and pH of the sewage sludge samples were determined in a 1:10 (w/v) water extract; total organic matter content (TOM) was determined by loss on ignition at 540 °C for 4 h after the extraction of inorganic matter with HCl [16]. Total organic carbon (TOC) and total nitrogen (TN) were determined by dry combustion at 950 °C using elemental analyzer (Truspec CN, Leco, St. Joseph, Mich., USA) according to Navarro et al. [17] and Paredes et al. [18]. After microwave digestion with HNO₃, P was assessed colorimetrically as molybdovanadate phosphoric acid [19], K and Na were determined by flame photometry (Jenway PFP7 Flame Photometer, Jenway Ltd., Felsted, Dunmow, Essex, UK) and Ca, Mg, Fe, Mn, Cd, Cr, Cu, Pb, Zn, Ni and Hg by atomic absorption (Varian 220FS Atomic Absorption Spectrometer, Varian Inc.,

Melbourne, Australia). Water-soluble carbon and nitrogen fractions (C_{WS} and N_{WS}) and organic carbon fractions were determined by automatic microanalysis [17] (TOC-V_{CSN}, Shimadzu Corporation, Kyoto, Japan), as were the 0.1 M NaOH-extractable organic carbon (C_{EX}) and fulvic acid-like carbon (C_{FA}), the latter after precipitation of the humic acid-like carbon (C_{HA}) at pH 2.0 [20]. The C_{HA} was calculated by subtracting the C_{FA} from the C_{EX}. The humification ratio (HR), the humification index (HI) and the percentage of humic acid-like carbon (P_{HA}) were calculated as (C_{EX}/TOC) × 100; (C_{HA}/TOC) × 100 and (C_{HA}/C_{EX}) × 100, respectively. The polymerization ratio (PR) was calculated as C_{HA}/C_{AF}. All the analytical determinations were done in quadruplicate.

2.3. NIRS analysis

NIRS analyses were performed using a Fourier transform NIR spectrometer (MPA, Bruker Optik GmbH, Germany) in the range of wavenumber from 12,000 to 3800 cm⁻¹ (830–2600 nm) with a step of 8 cm⁻¹, producing a spectrum with 2126 data points per sample. Each sample was poured in a glass plate and scanned three times using Opus software, version 6.0 (Bruker Optik) remixing the individual sample between scans. The three spectra of each sample were averaged obtaining a new one used for calibration and validation. Fig. 1 shows the NIRS spectra of the sewage sludge samples. Partial least square regression (PLSR) was used in the calibration step. The validation step was carried out using the full cross-validation method, following the leave-one-out procedure. To ensure a good correlation between the spectral data and the concentration values and find the best one, different pre-treatments of the spectra and combinations of them were tested

- (1) Vector normalization (VN): normalizes a spectrum by an initial calculation of the average intensity value and subsequent subtraction of this value from the spectrum. Then, the addition of the squared intensities is calculated and the spectrum is divided by the square root of this addition.
- (2) Minimum–maximum normalization ($N_{\min-\max}$): first subtracts a linear offset and then sets the y -maximum to a value of 2 by multiplication with a constant. Used similar to the vector normalization.
- (3) First derivative (FD): it calculates the first derivative of the spectrum. This method emphasizes steep edges of a peak. It is used to emphasize pronounced, but small features over a broad background. Spectral noise is also enhanced.
- (4) Second derivative (SED): similar to the first derivative, it is another treatment carried out separately from the first derivative and that provides a more drastic result.
- (5) Multiplicative scatter correction (MSC): performs a linear transformation of each spectrum for it to best match the mean spectrum of the whole set.
- (6) Straight line subtraction (SLS): fits a straight line to the spectrum and subtracts it. This accounts for a tilt in the recorded spectrum.
- (7) Linear offset subtraction (LOS): shifts the spectra in order to set the y -minimum to zero.

To evaluate the estimation, several statistical parameters were performed, such as: R^2 (coefficient of determination for calibration); RMSEE (root mean square error of estimation); F (number of factors or principal components); r^2 (coefficient of determination for validation); RMSECV (root mean square error of cross validation) and RPD (calculated as the standard deviation of the reference data of the validation set divided by the standard error of prediction). Malley et al. [21] suggested a guideline for describing the performance of calibrations for environmental

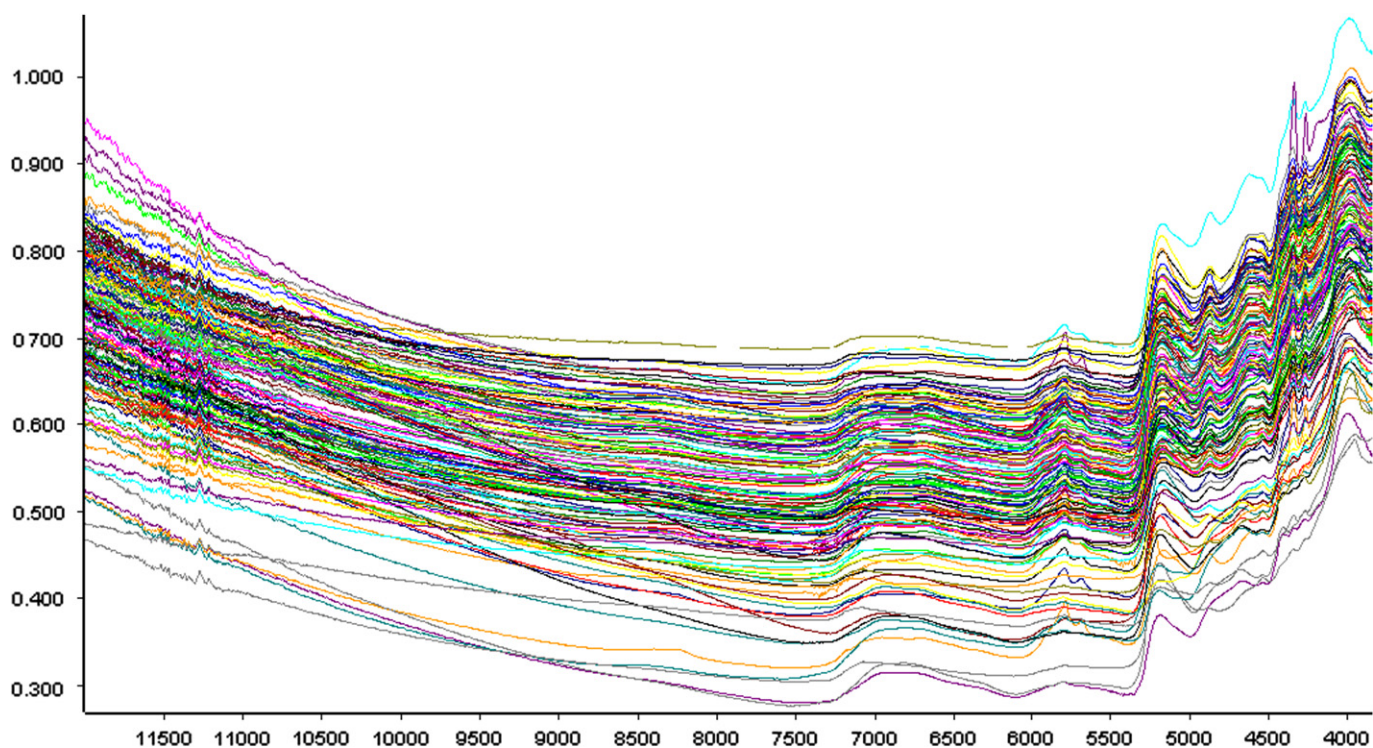


Fig. 1. NIRS spectra of sewage sludge samples (wavenumber (cm^{-1}) in abscissa and absorbance in ordinate).

samples, as follows: excellent calibrations $R^2 > 0.95$, $\text{RPD} > 4$; successful, $R^2 = 0.9-0.95$, $\text{RPD} = 3-4$; moderately successful, $R^2 = 0.8-0.9$, $\text{RPD} = 2.25-3$; and moderately useful, $R^2 = 0.7-0.8$, $\text{RPD} = 1.75-2.25$. Some calibrations with $R^2 < 0.7$ may be useful for screening purposes.

2.4. Mathematical model

A penalized signal regression (PSR) procedure was used to establish a prediction model for all the parameters from the absorbance signal given by the NIRS, similar to that proposed by Eilers and Marx [15]. This is the notation of the model

Y_i , $i = 1, \dots, n$, the value of the parameters measured by the analytic procedure in every sample.

$X = [x_{ij}]$, $i = 1, \dots, n$ y $j = 1, \dots, p$ is the absorbance matrix observed from the spectra, where p denotes the number of wavelengths considered and n the number of samples.

Then, the model corresponding to the concentration of each parameter can be formulated by

$$\log(\sqrt{Y_i}) = \alpha_0 + \sum_{j=1}^p x_{ij} \times \alpha_{ji}, \quad (1)$$

where α_0 represents intercept and α_{ji} gives the weight of the wavenumber w_j to predict the concentration of the parameter of interest. However, this formulation encounters a conditioning problem as the number of samples is quite fewer than the number of the wavelengths considered. To solve this problem, Eilers and Marx [15] proposed to project α onto a known basis of smooth functions, so making use of the ordered feature of the wavenumbers.

In this model, the number of initial predictors p is given by the number of wavelengths considered. To reduce that number, a fixed

number of nodes are established on the wavelength range considered, v_w , small enough to solve the problem of estimation and large enough to capture the general behavior of the spectrum. The proposed modelization states that the coefficients α of the prediction model (1) can be written in terms of a B-spline basis conformed from wavenumbers as

$$\alpha = \sum_{m=1}^{v_w} B_m \gamma_m = \mathbf{B}\Gamma,$$

where $\Gamma = [\gamma_m]$ is the vector of unknown coefficients associated with the base of B-splines. Then, (1) can be rewritten as

$$\log(\sqrt{Y_i}) = \alpha_0 + \mathbf{X}\mathbf{B}\alpha = \alpha_0 + \mathbf{U}\Gamma,$$

with $\mathbf{U} = \mathbf{X}\mathbf{B}$, where \mathbf{B} is the base of cubic B-splines presented along the axis defined by the wavenumber considered.

Galvez-Sola et al. [22] used a similar model to estimate the phosphorous content in compost samples. To verify the predictive ability of the model proposed, the set of samples was divided into two groups. The first one, with the 66% of the samples was used to adjust the proposed model, while the other 33% of the samples was used in the validation step.

The analysis has been resolved with the R package, version 2.9.2 [23]. The functions used are based on those programmed by Brian D. Marx and are available at <http://www.stat.lsu.edu/faculty/marx>. To obtain the value of α and Γ the criteria of the smallest prediction mean square was used.

3. Results and discussion

Table 1 shows the range of values of the different parameters studied in the sewage sludge samples. A wide range was obtained in most of the parameters considered, as a result of the high variability among the different wastewater treatment systems, but mainly due to the nature of the influent water quality [24]. In spite of a wide range in the constituent values as compared to the

Table 1
Range of the different parameters studied in the sewage sludge samples.

Parameter	Range	Mean value	SD	Parameter	Range	Mean value	SD
TOM (%)	14.0–82.0	53.3	10.9	P (%)	0.24–2.35	0.93	0.21
TOC (%)	12.0–46.3	33.2	6.1	K (%)	0.05–1.27	0.33	0.18
TN (%)	1.16–8.40	4.80	1.43	Ca (%)	1.01–25.40	7.52	3.69
C/N	1.9–14.3	5.8	2.0	Mg (%)	0.03–5.17	0.80	0.65
C _{WS} (%)	0.43–20.88	4.17	3.52	Na (%)	0.02–4.66	0.26	0.37
C _{EX} (%)	1.46–17.23	7.94	2.73	Fe (mg kg ⁻¹)	1007–150549	12684	23998
C _{FA} (%)	1.14–13.61	5.84	2.21	Mn (mg kg ⁻¹)	29–836	128	89
C _{HA} (%)	0.07–8.73	2.10	1.37	Cd (mg kg ⁻¹)	0.20–189.00	7.08	15.62
N _{WS} (%)	0.06–5.77	0.91	0.91	Cr (mg kg ⁻¹)	0.5–32662	535.4	2935.9
HR	9.30–54.92	23.65	6.14	Cu (mg kg ⁻¹)	26–4912	421	444
HI	0.30–32.31	6.23	3.78	Pb (mg kg ⁻¹)	0.5–1119	122	108
P _{HA} (%)	36.4–98.2	74.1	12.4	Zn (mg kg ⁻¹)	152–24634	927	1346
PR	0.02–1.75	0.39	0.26	Ni (mg kg ⁻¹)	2–1500	46	113
pH	3.92–7.73	6.51	0.45	Hg (mg kg ⁻¹)	0.01–7.00	0.69	0.87
EC (dS/m)	0.50–19.40	2.36	1.66				

TOM: total organic matter; TOC: total organic carbon; TN: total nitrogen; C_{WS}: water-soluble carbon; C_{EX}: extractable organic carbon; C_{FA}: fulvic acid-like carbon; C_{HA}: humic acid-like carbon; N_{WS}: water-soluble nitrogen; HR: humification ratio; HI: humification index; P_{HA}: percentage of humic acid-like carbon; PR: polymerization ratio; EC: electrical conductivity; SD: standard deviation.

Table 2
NIRS calibration and validation results for organic-matter related parameters.

Parameter ^a	Calibration ^b			Validation ^c					PSR validation	
	R ²	RMSEE	RPD	r ²	RMSECV	RPD	Bias	Factors	r ²	RMSECV
TOM (%)	0.84	3.76	2.46	0.78	4.27	2.13	-0.017	14	0.997	0.401
TOC (%)	0.88	1.79	2.93	0.87	1.88	2.73	0.005	13	0.999	0.219
TN (%)	0.89	0.43	2.99	0.87	0.45	2.77	0.001	12	0.995	0.156
C/N	0.68	0.97	1.73	0.58	1.06	1.54	-0.003	18	0.983	0.315
C _{WS} (%)	0.80	1.24	2.26	0.76	1.35	2.03	-0.001	16	0.935	0.519
C _{EX} (%)	0.85	0.93	2.62	0.80	1.05	2.26	-0.008	16	0.987	0.323
C _{FA} (%)	0.85	0.87	2.59	0.78	1.04	2.12	-0.002	17	0.986	0.283
C _{HA} (%)	0.69	0.65	1.79	0.58	0.73	1.55	-0.003	17	0.918	0.416
N _{WS} (%)	0.85	0.25	2.59	0.78	0.30	2.14	-0.007	16	0.922	0.438
HR	0.76	2.74	2.06	0.68	3.11	1.76	0.033	17	0.921	0.704
HI	0.64	1.96	1.66	0.51	2.23	1.43	0.006	14	0.995	0.593
P _{HA} (%)	0.67	7.01	1.71	0.51	8.12	1.43	0.035	20	0.906	0.192
PR	0.57	0.15	1.53	0.46	0.16	1.36	-0.001	17	0.922	0.265

^a TOM: total organic matter; TOC: total organic carbon; TN: total nitrogen; C_{WS}: water-soluble carbon; C_{EX}: extractable organic carbon; C_{FA}: fulvic acid-like carbon; C_{HA}: humic acid-like carbon; N_{WS}: water-soluble nitrogen; HR: humification ratio; HI: humification index; P_{HA}: percentage of humic acid-like carbon; PR: polymerization ratio.

^b R²: Coefficient of determination for calibration; RMSEE: root mean square error of estimation; RPD: calculated as the standard deviation divided by the standard error of prediction.

^c r²: Coefficient of determination for validation; RMSECV: root mean square error of cross validation; RPD: calculated as the standard deviation divided by the standard error of prediction.

error of individual measurements is a requirement for NIRS to maximize the correlation coefficient and minimize the error of prediction [25], we proceeded to the elimination of some samples detected as outliers by the software application, depending on the studied parameter, because of the high standard deviation in some elements, e.g., Fe and heavy metals.

3.1. Parameters related to organic matter fractions

In Table 2 are shown the NIRS calibration and validation results for the main parameters related to organic matter in the sewage sludge samples. For TOM, the coefficient of determination was 0.84 for calibration and 0.79 for cross-validation (Table 2). RMSEE was acceptable and the RPD obtained in the calibration process was higher than 2.25, so the results for this parameter were moderately successful, using the SLS pre-treatment of the absorbance signal. However, using the PSR method, the R² was improved (R²=0.997) and the RMSECV was lower, obtaining a better predictive model using this penalized signal regression. The determination of this parameter is important, not only for

what it represents, also because it is one of the parameters that, at least, should be analyzed in sewage sludge for agriculture, according to the Spanish legislation [26].

On the other hand, the results obtained for TOC were better; R² was 0.88 for the calibration process, with RPD=2.93, and similar results were obtained in the cross-validation step. Albrecht et al. [27] obtained a value of r²=0.87 for the calibration of C in samples of sewage sludge and green waste compost, using the same regression but in the Visible-NIR region. Our results were moderately successful and could be improved using the PSR model, increasing the R² to 0.999 with RMSECV=0.219. Similar results to those obtained for TOC were obtained for TN, according to the coefficients of determination and RPD. Galvez-Sola et al. [11], in a previous study using NIRS to evaluate the nitrogen content in sewage sludge samples, reported a R²=0.94 for TN using the PLSR model with the normalization as signal transformation but with higher error of estimation. Albrecht et al. [27] obtained an r²=0.89 in the Visible-NIR region in composted sewage sludge samples, while Vergnoux et al. [7] obtained a R²=0.98. For both TOC and TN, the best results were obtained

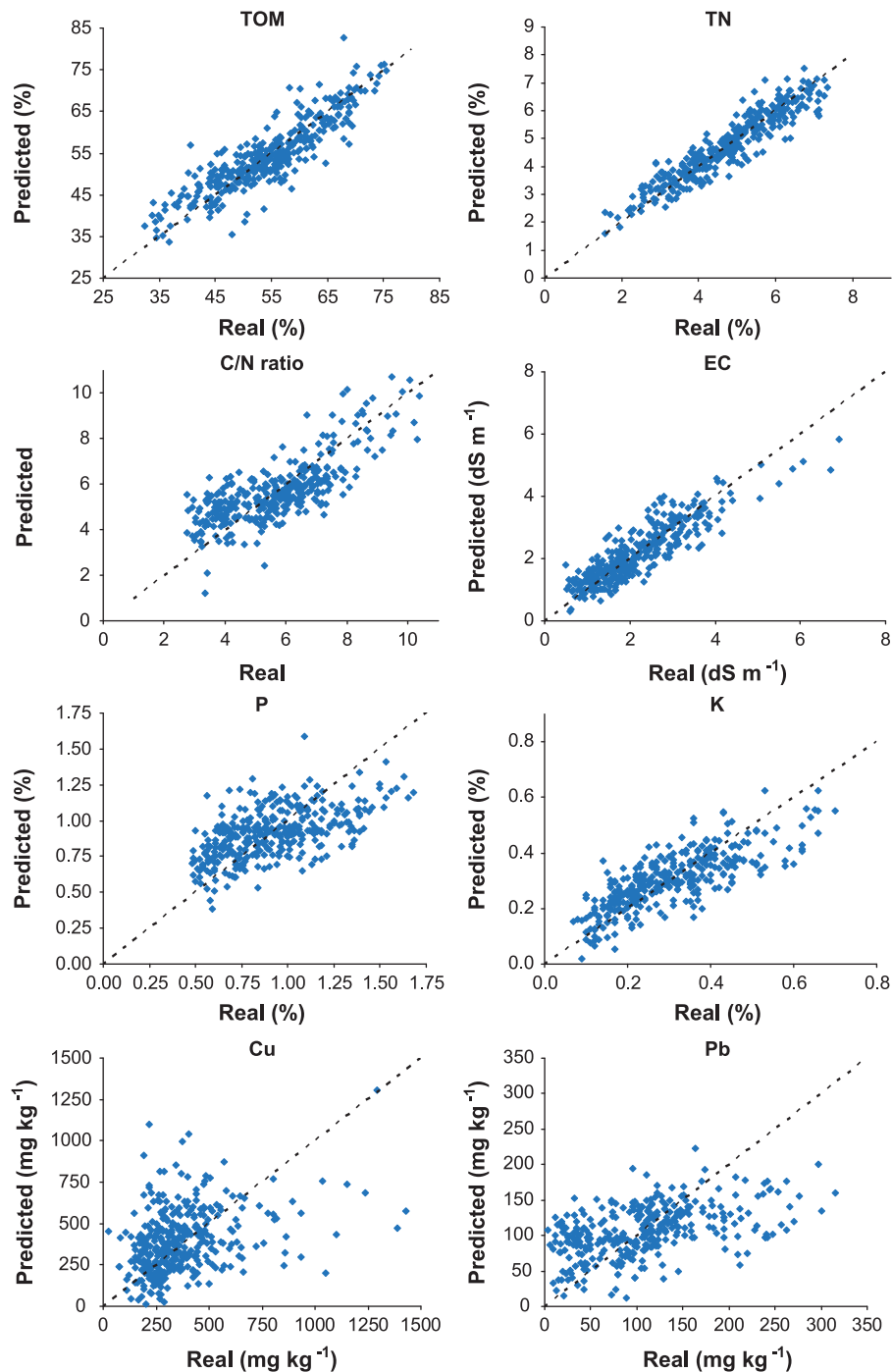


Fig. 2. NIRS cross-validation plots between predicted values (in ordinate) and real values (in abscissa) using PLSR for total organic matter (TOM), total nitrogen (TN), C/N ratio, electrical conductivity (EC), P, K, Cu and Pb (dotted lines represent perfect correlations).

with no pre-treatments of the absorbance signal. On the other hand, for C/N ratio, results were not satisfactory, because of the low determination coefficients and RPD and the high residual errors. This situation could be arranged using the PSR model, obtaining a $R^2=0.983$ for this ratio. Fig. 2 shows the calibration plots for TN and C/N ratio using PLSR and Fig. 3 shows the calibration plots for the same parameters, but using the PSR calibration method. Analyzing the nitrogen content and the C/N ratio is a requirement of the Spanish legislation associated to the application of sewage sludge in agriculture [28].

Regarding the water-soluble carbon and nitrogen fractions, C_{WS} calibration was moderately successful ($R^2=0.80$, $RPD=2.26$)

using the PLSR method. Nevertheless, calibration for C_{EX} was better, reaching R^2 the value of 0.85 and RMSE value was 0.92. Similar results were obtained for C_{FA} after combining the NV and FD as pre-treatment of the signal obtained by NIRS. No successful results were obtained for C_{HA} , due to the high prediction error and the low determination coefficients in both calibration and validation processes, using in this case the LOS pre-treatment. Finally, calibration for N_{WS} was moderately successful. With the use of the PSR method the determination coefficients obtained were better for all the parameters previously mentioned and higher than 0.91, with lower prediction error (except for N_{WS}), demonstrating the good goodness of fit achieved with this regression

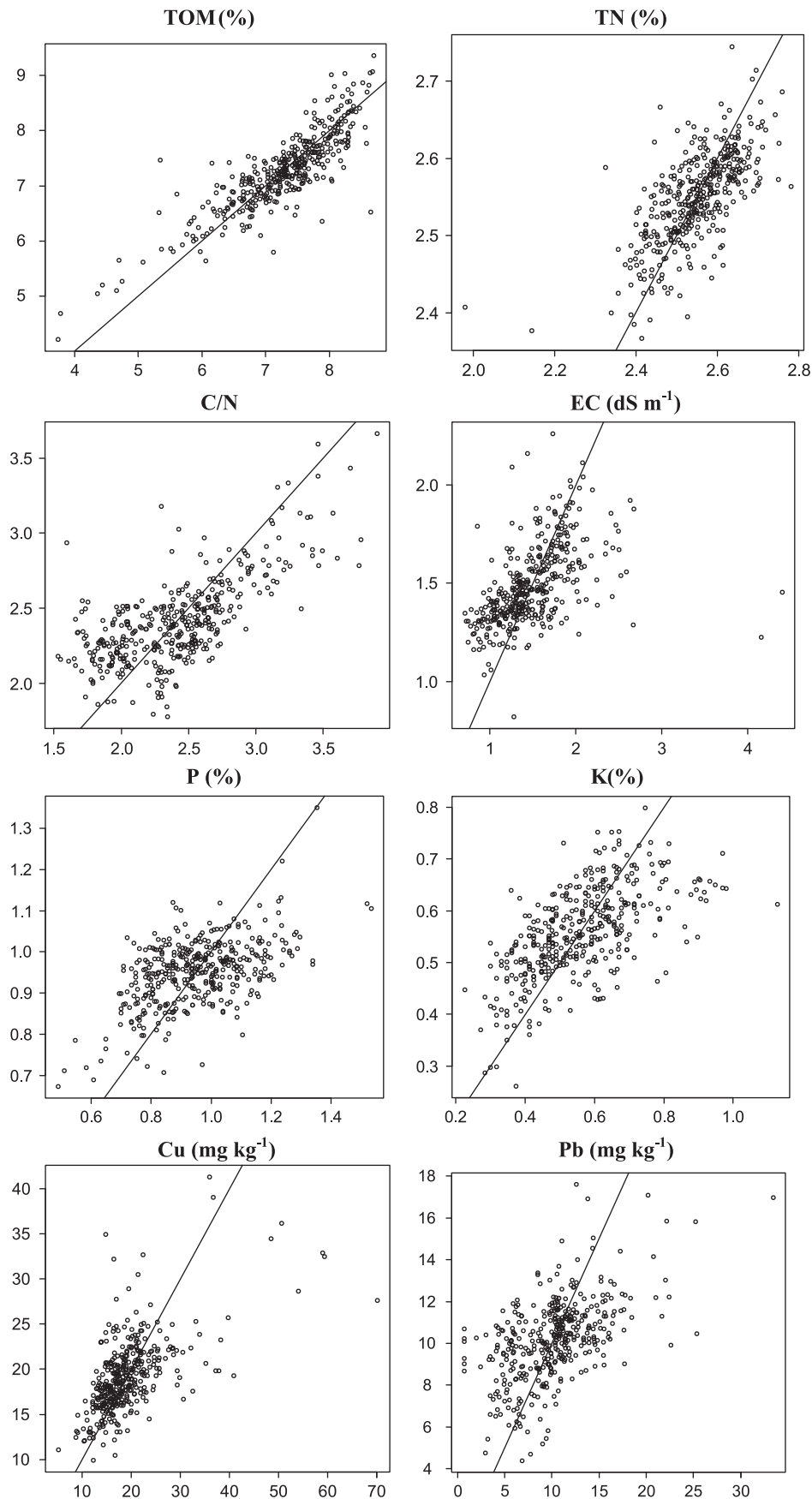


Fig. 3. PSR validation results for total organic matter (TOM), total nitrogen (TN), C/N ratio, electrical conductivity (EC), P, K, Cu and Pb (square root of the measured values in abscissa and square root of the predicted values in ordinate).

Table 3
NIRS calibration and validation results for chemical parameters and elements.

Parameter ^a	Calibration ^b			Validation ^c					PSR Validation	
	R ²	RMSEE	RPD	r ²	RMSECV	RPD	Bias	Factors	r ²	RMSECV
pH	0.79	0.19	2.18	0.71	0.22	1.87	−0.001	17	0.999	0.065
EC (dS m ^{−1})	0.85	0.42	2.55	0.77	0.5	2.09	0.004	18	0.942	0.370
P (%)	0.49	0.19	1.39	0.31	0.22	1.21	0.001	8	0.981	0.134
K (%)	0.64	0.08	1.66	0.60	0.08	1.57	0.001	9	0.96	0.114
Ca (%)	0.77	1.51	2.09	0.74	1.58	1.95	0.009	13	0.978	0.404
Mg (%)	0.89	0.18	3	0.86	0.19	2.68	0.002	12	0.923	0.249
Na (%)	0.79	0.10	2.16	0.65	0.12	1.7	0.019	18	0.834	0.208
Fe (mg kg ^{−1})	0.84	9840	2.49	0.79	1110	2.17	−34.80	15	0.868	42.727
Mn (mg kg ^{−1})	0.75	27.9	2.00	0.68	30.7	1.77	0.294	18	0.970	1.962

^a EC: Electrical conductivity.

^b R²: Coefficient of determination for calibration; RMSEE: root mean square error of estimation; RPD: calculated as the standard deviation divided by the standard error of prediction.

^c r²: Coefficient of determination for validation; RMSECV: root mean square error of cross validation; RPD: calculated as the standard deviation divided by the standard error of prediction.

Table 4
NIRS calibration and validation results for heavy metal content.

Parameter	Calibration ^a			Validation ^b					PSR Validation	
	R ²	RMSEE	RPD	r ²	RMSECV	RPD	Bias	Factors	r ²	RMSECV
Cd (mg kg ^{−1})	0.73	3.53	1.91	0.62	4.09	1.62	−0.040	13	0.829	1.102
Cr (mg kg ^{−1})	0.89	981	3.07	0.81	1280	2.3	9.140	17	0.732	1211.2
Cu (mg kg ^{−1})	0.67	258	1.74	0.44	334	1.33	9.060	8	0.925	5.638
Pb (mg kg ^{−1})	0.28	56.7	1.18	0.21	58.6	1.13	−0.095	8	0.891	3.648
Zn (mg kg ^{−1})	0.56	188	1.51	0.47	204	1.37	0.304	11	0.924	8.418
Ni (mg kg ^{−1})	0.60	72.4	1.58	0.24	98.7	1.15	0.621	6	0.787	3.148
Hg (mg kg ^{−1})	0.42	0.48	1.31	0.32	0.51	1.21	0.004	7	0.745	0.418

^a R²: Coefficient of determination for calibration; RMSEE: root mean square error of estimation; RPD: calculated as the standard deviation divided by the standard error of prediction.

^b r²: Coefficient of determination for validation; RMSECV: root mean square error of cross validation; RPD: calculated as the standard deviation divided by the standard error of prediction.

method. The performed calibrations for humification rates were not successful by PLSR, except for HR ($R^2=0.76$). However, with the PSR model, we obtained good correlations for all rates with determination coefficients higher than 0.90 (Table 2).

3.2. Chemical parameters and elements

The NIRS calibration and validation results for chemical parameters and elements are shown in Table 3. In relation to the chemical parameters, the calibration obtained for pH was moderately good. The resulting RPD was 2.18, with a $R^2=0.79$. However, in case of EC, the performed model was better, obtaining a RPD higher than 2.25 and a $R^2=0.85$ and pre-treating the signal using the SLS pre-treatment. Once again, with the PSR model, measures of goodness of fit improve, reaching more accurate results. R^2 for pH and EC were 0.999 and 0.942, respectively, and RMSECV were low enough for both parameters, as it is shown in Table 3.

For P, the resulting calibration was not satisfactory using the PLSR model, this fact being confirmed by the low values obtained for R^2 . For K, although the statistics were better than for P (Table 3), these were not good enough, even when using the model for monitoring purposes. Using the PSR model, the coefficients of determination were 0.98 and 0.96, respectively (Fig. 3). On the other hand, results for Ca were moderately useful, according to the values of R^2 and RPD obtained using PLSR, which were higher than 0.75 and 1.75, respectively. In addition, no signal pre-treatment was needed for K and Ca. With PSR, the r^2 was 0.978, improving the goodness of fit of the prediction and

decreasing the estimation error. Similar results were performed for Mn. For Na, the coefficient of determination obtained with PSR was also better than the obtained with PLSR, which constituted a moderately useful calibration with this regression. However, better results were obtained for Mg ($R^2=0.89$). This calibration was moderately successful with PLSR, improving the results with PSR. For Fe, R^2 was 0.84 for calibration and $r^2=0.79$ for the cross-validation process with the minimum–maximum normalization pre-treatment. So, the result obtained was moderately successful for this element, but with less estimation error using the calibration performed with the penalized regression tool. The analyses of P, K, Ca, Mg and Fe are essential from the point of view of plant nutrition; also, the Spanish legislation about the use of sewage sludge in agriculture requires that the determination of these parameters prior to sludge application [28].

3.3. Heavy metal content

The estimation of the heavy metal contents in sewage sludge samples using NIRS is complex, as it was reported by Moral et al. [8] in a study to evaluate the efficiency of NIRS to ascertain heavy metal contents in sewage sludge samples. Table 4 shows the NIRS calibration and validation results for the heavy metal contents in the sewage sludge samples studied. Using PLSR, R^2 for Cd was 0.73 with RPD=1.91, but this result can improve with the PSR model, increasing the coefficient of determination up to 0.83. On the other hand, the prediction of Cr using PLSR and the VN pre-treatment was possible, because the results were moderately successful, with $R^2=0.89$ and RPD=3.07. The cross-validation

process confirmed the good results obtained. In this case, despite the value of r^2 is lower using PSR, it is also lower using RMSECV and therefore, the method PSR supposes lower estimation error. The coefficient of determination was 0.73 and the estimation error was higher (Table 4).

For Cu, the estimation model was better with PSR, with a coefficient of determination higher than 0.90 and obtaining no successful results using PLSR (Figs. 2 and 3). Similar results were obtained for Pb and Zn, these results being successful and moderately successful with the PSR model respectively and not successful using PLSR. For Ni, the coefficient of determination by PSR was moderately useful ($r^2=0.79$ and RMSECV=3.148 mg kg⁻¹) and not successful with PLSR using the VN pre-treatment. Similar outcomes were obtained for the estimation of Hg in the sewage sludge samples using the PSR model ($r^2=0.75$ and RMSECV=0.418 mg kg⁻¹). In relation to the calibrations for the heavy metal content using PLSR, the results in this study were similar to those obtained in a previous research by Galvez-Sola et al. [29], but improving in this case the coefficient of determination for Cd and Hg and reducing the RMSE for all heavy metals except for Cr. As it has been previously mentioned for other elements, these heavy metals must be analyzed in sewage sludge for agriculture, according to the Spanish legislation.

As it was shown, PSR calibrations in this study were better than the typical PLSR calibrations widely used for NIRS prediction models. PSR attacks the high dimensionality and correlation in the regressors by projecting the regression coefficients onto a fewer basis of smooth functions, linearly related to them. Moreover, penalties are included in order to smooth the curve by avoiding neighbour regressors. Details and comparisons are explained in Marx and Eilers [14] for a similar example on NIRS signals.

4. Conclusions

Our results indicated that near infrared reflectance spectroscopy is a feasible tool for estimating total organic matter, total organic carbon, total nitrogen, water-soluble carbon, extractable carbon, fulvic acid-like carbon, electrical conductivity, Mg, Fe and Cr in sewage sludge samples using the PLSR tool. Besides, with the same regression tool, for pH, Ca, Na, Mn, Cd and also the humification ratio, the goodness of fit in the validation process was lower but useful. For C/N ratio, humic acid-like carbon, humification index, the percentage of humic acid-like carbon, the polymerization ratio, P, K, Cu, Pb, Zn, Ni and Hg, the goodness of fit between observed and predicted values of NIRS with PLSR was not good enough. Using PSR, all the parameters can be estimated with better goodness of fit than with PLSR, with coefficient of determination in the validation step higher than 0.90 for most of the studied parameters.

Acknowledgements

This work has been financed by the Ministerio de Ciencia e Innovación (currently Economía y Competitividad) of Spain and European Regional Development Funds (FEDER, “Una manera de

hacer Europa”) (Project AGL2009-12371-C02-01) and also by the Generalitat Valenciana (ACOMP/2010/177).

References

- [1] M. Odlare, V. Arthurson, M. Pell, K. Svensson, E. Nehrenheim, J. Abubaker, Appl. Energy 88 (2011) 2210–2218.
- [2] P.G. Gaspard, J. Schwartzbrod, Int. J. Hyg. Environ. Health 206 (2003) 117–122.
- [3] J.Y. Dai, L. Chen, J.F. Zhao, N. Ma, J. Environ. Sci. 18 (2006) 1094–1100.
- [4] H.R. Rogers, Sci. Total Environ. 185 (1996) 3–26.
- [5] Perfil Ambiental de España. Spanish Ministry of Agriculture, Food and Environment. 2010. Available at: <<http://www.magrama.gob.es/es/calidad-y-evaluacion-ambiental/temas/informacion-ambiental-indicadores-ambientales>>.
- [6] R. Albrecht, R. Joffre, R. Gros, J. Le Petit, G. Terrom, C. Périsol, Bioresour. Technol. 99 (2008) 448–455.
- [7] A. Vergnoux, M. Guiliano, Y. Le Dréau, J. Kister, N. Dupuy, P. Doumenq, Sci. Total Environ. 407 (2009) 2390–2403.
- [8] Moral, R., Galvez-Sola, L., Moreno-Caselles, J., Perez-Murcia, M.D., Perez-Espinosa, A., Paredes, C., A. Kungolos, K. Aravossis, A. Karagiannidis, P. Samaras (Eds.), First Conference on Environmental Management, Engineering, Planning and Economics, Skiathos Island, Greece, 2007, pp. 1683–1688.
- [9] L. Galvez-Sola, F.C. Marhuenda-Egea, M.D. Perez-Murcia, A. Pérez Espinosa, J. Moreno-Caselles, J. Morales, R. Moral, in: L. Shengcai, W. Yajun, C. Fengxia, H. Ping, Z. Yao (Eds.), Progress in Environmental Science and Technology, vol. II, Science Press, Beijing, China, 2009, pp. 643–648.
- [10] L. Galvez-Sola, R. Moral, M.D. Perez-Murcia, A. Perez-Espinosa, M.A. Bustamante, E. Martinez-Sabater, C. Paredes, Sci. Total Environ. 408 (2010) 14114–1421.
- [11] L. Galvez-Sola, R. Moral, J. Moreno-Caselles, M.D. Perez-Murcia, A. Perez-Espinosa, M.A. Bustamante, C. Paredes, Commun. Soil Sci. Plant Anal. 40 (2009) 726–735.
- [12] P.H.C. Eilers, L. Bin, B.D. Marx, Chemometr. Intell. Lab. Syst. 96 (2009) 196–202.
- [13] I.E. Frank, J.H. Friedman, Technometrics 35 (1993) 109–148.
- [14] B.D. Marx, P.H.C. Eilers, Technometrics 41 (1999) 1–13.
- [15] P.H.C. Eilers, B.D. Marx, Chemometr. Intell. Lab. Syst. 66 (2003) 159–174.
- [16] MAPA. Métodos Oficiales de Análisis. Tomo III; Secretaría General Técnica. Ministerio de Agricultura, Pesca y Alimentación, Madrid, Spain, 1994.
- [17] A.F. Navarro, J. Cegarra, A. Roig, M.P. Bernal, Commun. Soil Sci. Plant Anal. 22 (1991) 2137–2144.
- [18] C. Paredes, M.P. Bernal, A. Roig, J. Cegarra, M.A. Sánchez-Monedero, Int Biodeterior. Biodegrad. 38 (1996) 205–210.
- [19] R.E. Kitson, M.G. Mellon, Ind. Eng. Chem. Anal. Ed. 16 (1944) 379–383.
- [20] M.A. Sánchez-Monedero, A. Roig, C. Martínez-Pardo, J. Cegarra, C. Paredes, Bioresour. Technol. 57 (1996) 291–295.
- [21] D.F. Malley, Near Infrared Spectroscopy in Agriculture, in: C. Robert, J. Workman, J.B. Reeves (Eds.), American Society of Agronomy, Soil Science Society of America and Crop Science Society of America, Madison, WI, USA, 2004, pp. 729–784.
- [22] L. Galvez-Sola, J. Morales, A.M. Mayoral, F.C. Marhuenda-Egea, E. Martinez-Sabater, M.D. Perez-Murcia, M.A. Bustamante, C. Paredes, R. Moral, Chemosphere 78 (2010) 13–21.
- [23] R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2009. <<http://www.R-project.org>>.
- [24] J. Moreno-Caselles, M.D. Pérez-Murcia, A. Pérez-Espinosa, R. Moral, Fresenius Environ. Bull. 6 (1997) 519–524.
- [25] D.F. Malley, P.C. Williams, Environ. Sci. Technol. 31 (1997) 3461–3467.
- [26] BOE, Real Decreto 1310/1990, de 29 de Octubre, por el que se regula la utilización de lodos de depuración en el sector agrario, Boletín Oficial del Estado 262 (1990) 32339–32340.
- [27] R. Albrecht, R. Joffre, J. Le Petit, G. Terrom, C. Périsol, Environ. Sci. Technol. 43 (2009) 804–811.
- [28] BOE, Orden de 26 de Octubre de 1993 sobre utilización de lodos de depuración en el sector agrario, Boletín Oficial del Estado 265 (1993) 31073–31076.
- [29] Galvez-Sola, L., Marhuenda-Egea F.C., Perez-Murcia M.D., Pérez Espinosa A., Moreno-Caselles J., Morales, J., Moral R., in: Proceedings of the Second International Symposium on Agricultural and Agroindustrial Waste Management—II SIGERA, vol. II, D.D. Junior, G. Dalmédico, J.C.P. Palhares, V. Fracasso (Eds.), Foz do Iguaçu, Brazil 2008.