Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 285 (2023) 121922

Contents lists available at ScienceDirect

Spectrochimica Acta Part A:
Molecular and Biomolecular Spectroscopy

journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

Optimisation of the predictive ability of NIR models to estimate nutritional parameters in elephant grass through LOCAL algorithms

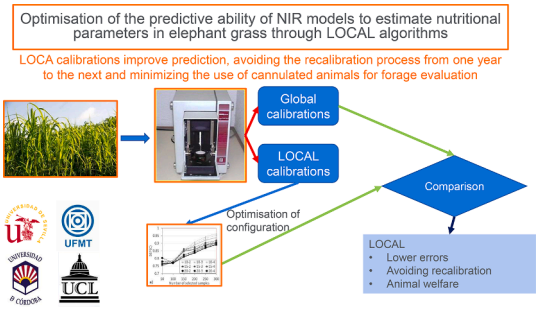
Víctor M. Fernández-Cabanás [a,*], Dolores C. Pérez-Marín [b], Tom Fearn [c], Joadil Gonçalves de Abreu [d]

[a] Urban Greening and Biosystems Engineering Research Group, Dpto. Agronomía, Universidad de Sevilla, ETSIA, Ctra. Utrera km.1, 41013 Seville. Spain
[b] Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain
[c] Department of Statistical Science, University College London, London WC1E 6BT, UK
[d] Universidade Federal de Mato Grosso/UFMT, Faculdade de Agronomia e Zootecnia, Cuiabá, MT, Brazil

HIGHLIGHTS

- NIR LOCAL models improve accuracy in estimating nutritional value of elephant grass.
- Adequate configuration of parameters for LOCAL calibrations reduces computing time.
- Appropriate graph display allows decisions about LOCAL algorithm configuration.
- Nutritive evaluation of forages by NIRS could avoid the use of fistulated animals.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Near infrared
LOCAL regression
Pennisetum purpureum

ABSTRACT

Elephant grass is a tropical forage widely used for livestock feed. The analytical techniques traditionally used for its nutritional evaluation are costly and time consuming. Alternatively, Near Infrared Spectroscopy (NIRS) technology has been used as a rapid analysis technique. However, in crops with high variability due to genetic improvement, predictive models quickly lose accuracy and must be recalibrated. The use of non-linear models such as LOCAL calibrations could mitigate these issues, although a number of parameters need to be optimized to obtain accurate results. The objective of this work was to compare the predictive results obtained with global NIRS calibrations and with LOCAL calibrations, paying special attention to the configuration parameters of the models.

The results obtained showed that the prediction errors with the LOCAL models were between 1.6 and 17.5 % lower. The best results were obtained in most cases with a low number of selected samples (n = 100–250) and a high number of PLS terms (n = 20). This configuration allows a reduced computation time with high accuracy, becoming a valuable alternative for analytical determinations that require ruminal fluid, which would improve the welfare of the animals by avoiding the need to surgically prepare animals to estimate the nutritional value of the feeds.

* Corresponding author.
E-mail addresses: victorf@us.es (V.M. Fernández-Cabanás), dcperez@uco.es (D.C. Pérez-Marín), t.fearn@ucl.ac.uk (T. Fearn), joadil.abreu@ufmt.br (J. Gonçalves de Abreu).

https://doi.org/10.1016/j.saa.2022.121922
Received 20 May 2022; Received in revised form 28 July 2022; Accepted 20 September 2022
Available online 24 September 2022

1. Introduction

Determining the nutritional characteristics of the feed supplied to livestock is essential for decision-making when planning feed management on farms. However, this activity entails frequent analyses, especially for forages, since their composition can undergo significant variations depending on the growth stage, type of management carried out, climate conditions, season and fertilization [1].

Elephant grass (*Pennisetum purpureum* Schum.) is a type of forage widely produced in countries with a tropical climate in Africa and South America, where it is consumed by grazing animals or provided by cut-and-carry system, or preserved as silage [2]. This species is even used for biofuel production [3]. The main parameters used for evaluating the quality of elephant grass forage are crude protein (CP), acid detergent fibre (ADF), neutral detergent fibre (NDF) and total digestible nutrients (TDN) [4], together with *in vitro* dry matter digestibility (IVDMD). Most of those analytical determinations are expensive, time consuming and require the use of chemical reagents, which could lead to environmental and operator health risks. Furthermore, it is worth mentioning that efforts have been made by the scientific community to develop methods that reduce the time and cost of research, and at the same time minimize the use of cannulated animals in the rumen, in order to ensure their welfare. In this sense, the use of predictive models that allow estimating the nutritional value of feed accurately, without having to use ruminal fluid extracted from fistulated animals, would be of great help to improve the animal welfare of experimental animals. As an alternative, NIRS technology (near infrared reflectance spectroscopy) could be considered an interesting analytical method for the quality of this forage as it presents a high analysis speed, lower cost, it is not polluting since it does not use chemical reagents, in addition to making it possible to dispense with the use of surgically prepared animals usually required for indigestible neutral detergent fibre (iNDF) or IVDMD determinations [5].

The basis of this spectroscopic method is the absorption of radiation at specific wavelengths by certain molecular bonds in the near infrared (NIR) region (1100–2500 nm): O—H, C—H, N—H, and C—O. Due to the presence of these kind of molecular bonds in agricultural products, NIRS technology has been successfully used for the evaluation of quality of food [6,7], forage [5,8] and, particularly, for nutritive value and/or ethanol conversion yield of kikuyu and elephant grass [3,9,10].

Most of the published works on the use of NIRS in forage analysis have been developed with global calibrations, using linear models (such as Partial Least Squares Regression, PLSR) that work properly when the calibration group for the studied product is homogeneous and the number of samples is relatively low. For this purpose, Shenk et al. [11] recommended creating a training data set including all possible sources of variation that might be encountered during prediction. However, although this strategy generally increases the robustness of the calibrations, it tends to decrease their precision if the collective is highly variable and/or very broad. This decrease in the precision of the calibrations is especially significant for highly variable crops due to breeding programs, such as elephant grass, where year after year the concentrations of the constituents of interest increase, resulting in a non-linear adjustment for the highest values [12]. In this sense, Barton et al. [13] proposed three approaches to deal with highly variable and large sample populations: 1) accept the loss of precision or develop specific calibrations for subsets within the general population, 2) use artificial neural networks for regression purposes dealing with non-linearity, 3) use of local approaches based on the development of specific calibrations for each one of the samples to be predicted. Regarding the third of these mentioned options, Berzaghi et al. [14] stated that the use of the LOCAL algorithm [11] would avoid the need to choose between precision and robustness of a calibration.

The LOCAL algorithm was developed to manage very large spectral libraries. In a first step, it looks for similar spectra to the unknown sample using the correlation between the spectrum of that sample and the rest of spectra in the library and, then, it develops a specific PLS regression for the sample, dealing with non-linearity. This algorithm has proven to be very useful for improving the precision in calibrations developed for a series of agricultural products and analytical parameters, among which we can highlight compound feeds [15–17], fruits [18–22], vegetables [12,23], multi-product data bases [14] and even forages [24,25].

However, for the correct operation of the LOCAL algorithm, it is necessary to configure a series of previous settings, not only for the math pretreatments that are applied to spectral data, such as derivatives, but also for the minimum number of samples that will be selected for the development of calibrations, the maximum number of PLS terms, and the number of first PLS factors to be removed [17,26]. In this sense, one of the outstanding drawbacks of this method is the need to perform a regression with some selected samples to predict each unknown sample. It is possible that not enough samples similar to the sample to be predicted will be found to provide an acceptable estimate, which can cause the NIRS application to fail [15]. Despite the high power of modern computers used in calculations, when the number of samples in the spectral library is very high, the time consumed in processing the information can be significantly prolonged [26]. Therefore, it would be very useful to have precise information on the most appropriate range of these parameters to obtain precise estimates for the desired parameters in various types of products. For this reason, the present study was conducted in which the prediction errors of models developed to estimate nutritive values in elephant grass with global calibrations and with the LOCAL algorithm will be compared, selecting the most appropriate configuration to optimize the results with the non-linear model. A graphical approach will be tested for decision making on the optimization of the configuration parameters of the LOCAL algorithm.

2. Materials and methods

2.1. Characterization of sites, experiments and samples

A total of 993 elephant grass forage samples from different trials (1, 2 and 3) were used. The trials were carried out in two locations during two different seasons: rainy (October to April) and dry (May to September). The first and third trial were carried out at the Experimental Farm of the Federal University of Mato Grosso, in Santo Antônio de Leverger-MT, located at 15°51′ South Latitude and 56°04′ West Longitude of Greenwich, 140 m altitude. The second experiment was carried out in the Experimental Field of the Mato Grosso Rural Research, Assistance and Extension Company (EMPAER), in Cáceres MT, located at 16°09′04″ South Latitude; 57°38′03″ West Longitude; altitude of 157 m. The climate in both places, according to the Köppen classification, is of the Aw type, that is, a megathermic tropical climate, characterized by two well-defined seasons: rainy and dry.

In trial 1 (2017–2018), the cultivar BRS Canará, launched by Empresa Brasileira de Pesquisa Agropecuária (Embrapa) in 2012, was used. There, 168 forage samples were collected from weekly cuts at 21 ages (from 21 to 161 days), along two seasons (rainy and dry), with four repetitions.

In trial 2 (2018–2019), 13 cultivars were used (BRS Canará, Cameroon, Napier, Vruckwona, Guaçu, Taiwn A-25, Porto Rico, Cubano Pinda, Piracicaba 241, Cuba 116, Mercker, CNPGL clones 93–41-1 and 91–25-1), making cuts in the rainy season at 15 ages (16; 30; 45; 58; 74; 86; 98, 105; 115; 130; 143; 171; 186; 200; 215 days) with three repetitions, totalling 585 samples.

In trial 3 (2018–2019), five cultivars (Vruckwona, BRS Capiaçu, Napier, BRS Canará and Cameron) were used, making cuts in the rainy and dry seasons at 6 ages (30; 45; 60; 75; 90; 105 days) with four repetitions, totalling 240 samples.

Each parcel was 5.00 m long and 4.00 m wide, with a line spacing of 1.00 m. The usable area was the central 4.00 m of the two lines in the middle of each plot (Fig. 1). The cultivars were harvested at the

recommended ages, and the cut was made with a machete near the ground. After cutting, the material was chopped to a size of 2 cm. Then, forage samples were placed in paper bags and dried in a forced ventilation oven at 55 ± 5 °C for 72 h. The previously dried samples were weighed and ground in a Wiley mill with a 1.0 mm screen, and stored in polyethylene containers to evaluate their bromatological composition.

The contents in mineral matter (Ash) and CP were evaluated, according to Helrich [27]; neutral (NDIP) and acid detergent insoluble protein (ADIP), according to Licitra et al. [28]; NDF and ADF, according to Van Soest [29]; iNDF, according to Cochran et al. [30]; neutral and acid detergent fiber corrected for ash and protein (NDFap; ADFap) and ash insoluble in neutral and acid detergent (NDFA; ADFA), according to Detmann et al. [31]. The hemicellulose content was determined by the difference between NDF and ADF.

The contents of TDN, net lactation energy (NLE) and *in vitro* organic matter digestibility (IVOMD) were estimated, according to Cappelle et al. [32] (Equations 1, 2 and 3):

$$TDN = 88.9 - (ADF \times 0.779); \quad (1)$$

$$NLE = 2.39 - (ADF \times 0.028). \quad (2)$$

$$IVOMD = 84.86 - (NDF \times 0.41). \quad (3)$$

2.2. Spectra collection

Reflectance spectra of dried and ground forage samples were obtained on a Foss-NIRSystems 6500 SY-II monochromator (Foss NIRSystems, Silver Spring, MD), from 400 to 2498 nm, every 2 nm (spectral bandpass 10 nm ± 1 nm). Analysis was performed using a transport module. Samples were scanned using a ¼ rectangular cup (4.6 cm wide and 5.7 cm long), being the spectrum of each sample the average of two forage sub-samples.

All spectra were manipulated and processed, and all calibration equations were obtained using WINISI software version 1.5 (Infrasoft International, State College, PA).

2.3. Selection of training and external validation data sets

Appropriate selection of representative samples for calibration and validation is crucial, as both of them should include all the variables affecting spectral attributes, they have to be placed at boundary of the category and filling the group space uniformly [33]. When the LOCAL regression method is used, the stability of the predictions is achieved using large spectral libraries that cover all the expected variability for the product and parameter analysed [26].

In this work, samples for training and validation sets were chosen using a structured selection based on the Global Mahalanobis (GH) distance of each sample to the centre of the spectral population, after spectral pre-treatment with Standard Normal Variate (SNV) and Detrend and first derivative 1,5,5,1 – where the first digit indicates the order of the derivative; the second is the derivative gap; the third is the smoothing segment and the fourth is the second smoothing segment [34]. Then, once all samples were sorted from lower to higher GH values, it was selected one of every-four samples for the validation set and the remaining samples for calibration. The composition of the training and validation sets is showed in Table 1.

2.4. Development of global calibration models based on Visible-NIR and external validation

The study was performed in the Vis-NIR spectral range (400–2498 nm). Predictive models for each variable were developed using Modified Partial Least Squares Regression (MPLSR) [35] with the PLS1 algorithm for comparison purposes. Cross validation was used to determine the ideal number of factors for the regression models and to avoid overfitting. Validation errors were combined to obtain a standard cross validation error (SECV).

Spectral dispersion was corrected using the Standard Normal Variate (SNV) and Detrending (Dt) spectral pre-treatments [36]. Six spectral derivatives (1,5,5,1; 1,10,5,1; 1,10,10,1; 2,5,5,1; 2,10,5,1; 2,10,10,1) were also applied.

Outliers were identified and removed during the calibration process, as they could affect model performance and decrease precision for most samples. A maximum of two outliers deletion passes (T and H) were performed before completing the final calibration [37]. T outliers corresponded to samples with significant differences between their predicted and laboratory values, while H outliers were samples with extreme spectral distances (H greater than 3) to the center of the calibration group [35].

MPLSR models were evaluated in terms of their calibration statistics, namely standard error of calibration (*SEC*), calibration determination coefficient (R^2), SECV and determination coefficient for cross validation (r^2).

All equations were then evaluated by external validation, through the determination of the bias and the standard error of performance bias corrected (SEP(C)).

In this study, the following limits were used for the evaluation of calibrations developed with more than 100 samples, with validation sets containing more than 9 samples [34]:

$$\text{Limit Control for SEP(C)} = 1.30 \times \text{SEC}$$

$$\text{Limit control for Bias} = \pm 0.60 \times \text{SEC}$$

Another statistic used was the RPD or ratio of the standard deviation

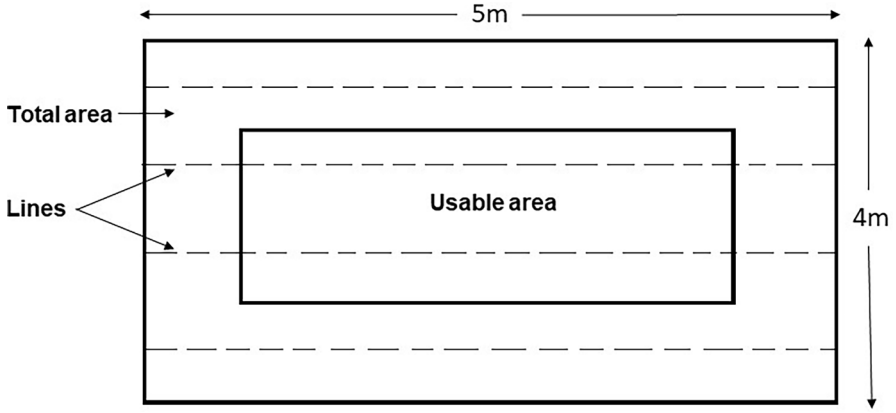


Fig. 1. Plot design used in the study.

Table 1
Composition of calibration (n = 705) and validation (n = 235) data sets.

	Calibration data set				Validation data set			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Ash (%)	3.29	21.18	9.97	2.70	3.19	18.89	10.18	2.64
CP (%)	1.00	16.59	5.44	3.24	1.79	15.88	5.59	3.50
NDF (%)	54.98	87.74	71.54	4.64	56.52	82.30	71.37	4.62
ADF (%)	30.08	63.35	49.67	6.41	29.32	62.81	49.30	6.71
HEM (%)	5.58	39.56	21.87	5.65	11.18	39.96	22.06	6.07
TDN (%)	39.55	62.13	49.78	4.18	39.97	62.13	49.97	4.25
NLE ($Mcal.kg^{-1}$)	0.62	1.55	1.00	0.18	0.63	1.57	1.01	0.19
IVOMD (%)	41.35	63.56	51.53	4.30	41.76	63.56	51.72	4.37
NDIP (%)	0.69	3.74	1.63	0.64	0.70	3.52	1.60	0.62
ADIP (%)	0.45	2.67	1.07	0.43	0.39	2.57	1.06	0.43
NDFA (%)	0.11	6.85	3.40	1.29	0.72	6.60	3.45	1.20
ADFA (%)	0.63	5.60	2.49	0.87	0.69	4.56	2.46	0.82
NDFap (%)	51.98	82.54	66.56	4.90	53.12	79.95	66.26	4.50
ADFap (%)	27.83	59.02	45.59	6.32	27.58	59.12	44.93	6.50
iNDF (%)	12.62	60.20	32.08	8.32	14.74	54.30	32.62	8.41

of the original data (SD) to SEP [38]. Best calibrations were selected based on the higher r^2 and RPD values and lower SEP.

2.5. Development and validation of prediction models using the LOCAL algorithm

In this study, LOCAL algorithm described by Shenk et al. [11] was compared with global calibrations using MPLS regression. As it was stated before, different parameters must be evaluated in order to optimize the LOCAL algorithm. In this work, an optimization design was set up by varying the number of samples to select from the spectral library (k) from 50 to 300 in steps of 50; the maximum number of PLS terms to be used (l) (10, 15 and 20); and the number of predicted values generated with the first PLS terms to be excluded from the calculation of final predicted values (m) (2, 3 and 4). In total, 54 combinations of configuration parameters were applied to each analytical determination and calculated in batch mode.

Furthermore, for each parameter analysed, the same mathematical signal pre-treatments tested with global calibrations were evaluated with LOCAL, and results obtained were evaluated using the same statistics described above for the validation of global equations.

3. Results

3.1. Calibration and validation data sets

Selection of samples for the validation data set based on the GH values of samples led to a similar distribution for the mean, minimum, maximum and standard deviation values for all the analytical parameters studied, as shown in Table 1. This configuration is very convenient for an accurate estimation of the predictive capacity of the global and LOCAL models developed in this work using external validation.

3.2. Features of the developed global calibrations

Global calibrations were obtained based on the visible and NIR spectra (400–2500 nm) for the estimation of the bromatological composition of elephant grass forage (Table 2). Although there is no derivative that stands out for its best results in general terms, it can be observed that the best predictive models have been obtained with second derivatives (2,5,5,1 and 2,10,5,1).

For the evaluation of the performance of the models developed, the determination coefficients for cross validation showed excellent values for CP and iNDF (0.95 and 0.89, respectively), while for Ash, ADIP, NDFA, NDIP and ADFap good values of r^2 were obtained (0.85, 0.79, 0.77, 0.76 and 0.71, respectively).

Lower r^2 values were found for the estimation of ADF and NDF and for the parameters calculated from them (Hem, TDN, NLE and IVOMD); as well as for ADFA and NDFap. Nevertheless, all of them are above 0.50 and, according to Shenk and Westerhaus [39], these results enable to distinguish between low, medium and high values of these parameters.

According to RPD values [38], calibration for CP is classified as fair, and could be used for screening purposes; Ash and iNDF are classified as poor, and could be used for very rough screening; and the rest are classified as very poor, with not recommended use. However, this classification was designed for calibrations developed for very homogeneous products, and should be less restrictive for heterogeneous products such as forages.

3.3. Configuration of the LOCAL algorithm for elephant grass calibrations

Statistics for selected LOCAL calibrations are displayed in Table 3. The number of selected samples (k) ranged between 100 and 300, the lowest number being associated with the best predictive models.

The maximum number of PLS terms to be used was the highest possible value for most of the selected models ($l = 20$). Although this is a relatively low number of terms, excluding 3–4 of the first terms was found to improve predictive results. The number of PLS terms to be excluded in the calculations (m) in about half of the selected models were 4, although most of them (except iNDF) were the ones with the lowest predictive capacity.

In order to delve into the effect of the configuration parameters of the LOCAL algorithm (k, l and m) on the predictive ability of the developed models, a graphical approach has been performed for selected parameters with excellent (CP), moderate (iNDF and Ash) and poor predictive results (NDF).

Fig. 2 shows the evolution of the SEP(C) values for the estimate of CP as a function of the configuration parameters of the LOCAL algorithm (k, l and m) and of the derivatives used in the calculation. In most cases, and particularly for second derivatives and with l values of 20 PLS terms, a minimum value was observed for $k = 100$. Above this number of selected samples, the error values increased significantly. For this parameter, one of the best estimated with LOCAL and global calibrations, lower errors were observed for high l and low m values.

The models developed for estimation of Ash showed a different behaviour (Fig. 3). It is clear that these are more complex models that require a greater number of PLS terms to be able to make correct predictions, although it was also necessary to exclude a greater number of the first PLS terms in order to optimize results. For this constituent, depending on the derivative used, minimum error values were obtained by selecting between 100 and 200 samples.

The models developed for iNDF (Fig. 4) had a different response to those already described. Regardless of the derivative used, the starting point was high error values for a low number of selected samples and,

Table 2
Calibration statistics for selected global equations.

Constituent	N	Derivative	Mean	SD	SEC	R^2	SECV	r^2	RPD
ASH (%)	668	2,5,5,1	10.05	2.54	0.89	0.88	0.99	0.85	2.6
CP (%)	675	2,10,5,1	5.33	3.19	0.64	0.96	0.69	0.95	4.6
NDF (%)	679	2,5,5,1	71.73	4.29	2.72	0.60	2.91	0.54	1.5
ADF (%)	702	2,10,5,1	49.69	6.39	3.63	0.68	3.91	0.63	1.6
HEM (%)	685	2,5,5,1	21.84	5.42	3.47	0.59	3.80	0.51	1.4
TDN (%)	696	2,10,5,1	49.72	4.13	2.68	0.58	2.88	0.52	1.4
NLE (Mcal.kg-1)	696	1,5,5,1	1.00	0.18	0.10	0.69	0.11	0.64	1.7
IVOMD (%)	696	2,10,5,1	51.47	4.26	2.65	0.61	2.84	0.56	1.5
NDIP (%)	571	2,5,5,1	1.62	0.62	0.28	0.79	0.31	0.76	2.0
ADIP (%)	637	2,5,5,1	1.03	0.40	0.16	0.83	0.18	0.79	2.2
NDFA (%)	586	1,10,5,1	3.43	1.27	0.54	0.82	0.60	0.77	2.1
ADFA (%)	533	2,5,5,1	2.47	0.80	0.50	0.61	0.52	0.58	1.5
NDFap (%)	511	1,5,5,1	66.60	4.70	2.73	0.66	3.09	0.57	1.5
ADFap (%)	553	1,5,5,1	45.60	6.24	3.13	0.75	3.39	0.71	1.8
iNDF (%)	641	1,10,10,1	31.98	8.05	2.58	0.90	2.68	0.89	3.0

Table 3
Configuration parameters and validation statistics for selected LOCAL calibrations.

Constituent	Deriv.	*k*	*l*	*m*	SEP	Bias	SEP(C)	Slope	R^2	GH	NH
Ash (%)	1,5,5,1	100	20	3	1.00	0.07	1.00	0.94	0.86	0.89	0.45
CP (%)	2,10,5,1	100	20	2	0.74	−0.10	0.73	1.01	0.96	0.87	0.35
NDF (%)	1,5,5,1	150	20	3	3.11	−0.19	3.11	0.98	0.55	0.94	0.55
ADF (%)	2,5,5,1	300	20	4	3.70	−0.06	3.71	1.07	0.70	0.96	0.46
HEM (%)	1,5,5,1	100	10	4	3.78	−0.07	3.79	1.02	0.61	0.92	0.29
TDN (%)	2,5,5,1	300	20	4	2.81	0.00	2.81	1.07	0.57	0.98	0.48
NLE ($Mcal.kg^{-1}$)	2,5,5,1	300	20	4	0.10	0.00	0.10	1.08	0.70	0.96	0.46
IVOMD (%)	2,5,5,1	300	20	4	2.77	−0.01	2.77	1.05	0.60	0.98	0.47
NDIP (%)	2,5,5,1	150	20	2	0.30	0.00	0.31	0.98	0.76	1.02	0.41
ADIP (%)	1,5,5,1	200	15	3	0.22	−0.02	0.22	0.97	0.75	1.68	0.22
NDFA (%)	2,10,5,1	100	20	2	0.63	−0.04	0.63	0.95	0.73	1.26	0.62
ADFA (%)	1,5,5,1	150	20	4	0.54	−0.04	0.54	0.93	0.56	1.14	0.65
NDFap (%)	1,10,5,1	150	20	4	3.29	−0.32	3.29	0.93	0.47	1.36	0.81
ADFap (%)	1,5,5,1	100	10	2	3.34	−0.19	3.34	1.06	0.74	1.06	0.28
iNDF (%)	2,5,5,1	250	20	4	3.32	0.59	3.27	0.99	0.85	1.16	0.40

K: number of samples to select from the spectral library; *l*: maximum number of PLS terms to be used; *m*: number of predicted values generated with the first PLS terms to be excluded from the calculation of final predicted values; GH: average global H value; NH: average neighbourhood H value.

above 100 samples, a level was reached where the variations were minimal, with a slight upward trend with the first derivatives. With *k* values greater than 150, lower errors were obtained when the number of PLS terms was higher and a greater number of terms were excluded ($m > 3$).

The results for the models developed for NDIP (data not shown) were similar to those already described for CP, while those obtained for ADIP were similar to those discussed for iNDF.

The magnitude of the differences found in the errors, as a function of the LOCAL configuration parameters used, were much higher when the predictive capacity of the models was lower, as can be seen for the calibrations obtained for NDF (Fig. 5). For this constituent, the best results were obtained with first derivatives, $k = 150$, $l = 20$ and $m > 2$.

3.4. External validation of global and LOCAL models

External validation of global calibrations (Table 4) confirmed that for some of the parameters with the worst calibration statistics (NDF, Hem, TDN, IVOMD and NDFap) the quality criteria defined by Shenk et al. [34] for coefficient of determination values were not met. However, the specific quality criteria (maximum value for SEP(C) = 1.30 × *SEC* and Bias limits = ± 0.60 × *SEC*) did not indicate any nonconformity for the selected equations.

Validation statistics for best LOCAL models selected from the 54 combinations of configuration parameters for each of the 6 derivatives implemented are showed in Table 3. The influence on the predictive results of the pretreatments used as well as the LOCAL configuration parameters is particularly relevant. In general, there was no coincidence in the derivative that provided better results in global and local calibrations, although most models with better statistics in LOCAL were usually obtained with a second derivative (2,5,5,1).

It was evident the decrease in predictive errors with LOCAL calibrations compared to global ones. In Table 5 it can be observed that the improvements imply reductions in the SEP(C) between 1.62 and 17.55 %, with an average value of 9.53 %.

4. Discussion

Results obtained with global calibrations developed for elephant grass forage in this study were similar to those reported by Lobos et al. [40]. These authors evaluated the quality of samples of several permanent pasture species (*Lolium perenne, Agrostis sp, Holcus lanatus, Bromus valdivianus, Dactylis glomerata, Medicago sativa, Trifolium pratense* and *T. repens*), using NIRS and obtaining values of coefficient of determination suitable for the estimation of the CP content ($R^2 = 0.99$). However, SECV values obtained in the present work were lower than those reported for kikuyu grass (SECV = 1.14) [41] and for tropical forages (SECV = 0.8) [42] and similar to those obtained for Tifton 85 grass (*SEC* = 0.60) [43].

Another of the constituents that was estimated with adequate precision with global calibrations was Ash, despite the fact that it is assumed that minerals do not absorb in the NIR region. For this parameter, the results could only be compared with those obtained by Serafim et al. [43] for Tifton 85 grass (*Cynodon* spp.), where lower *SEC* values were reported than those of this work (*SEC* = 0.38; $R^2 = 0.84$).

The statistics of global calibrations obtained for ADF, NDF and the

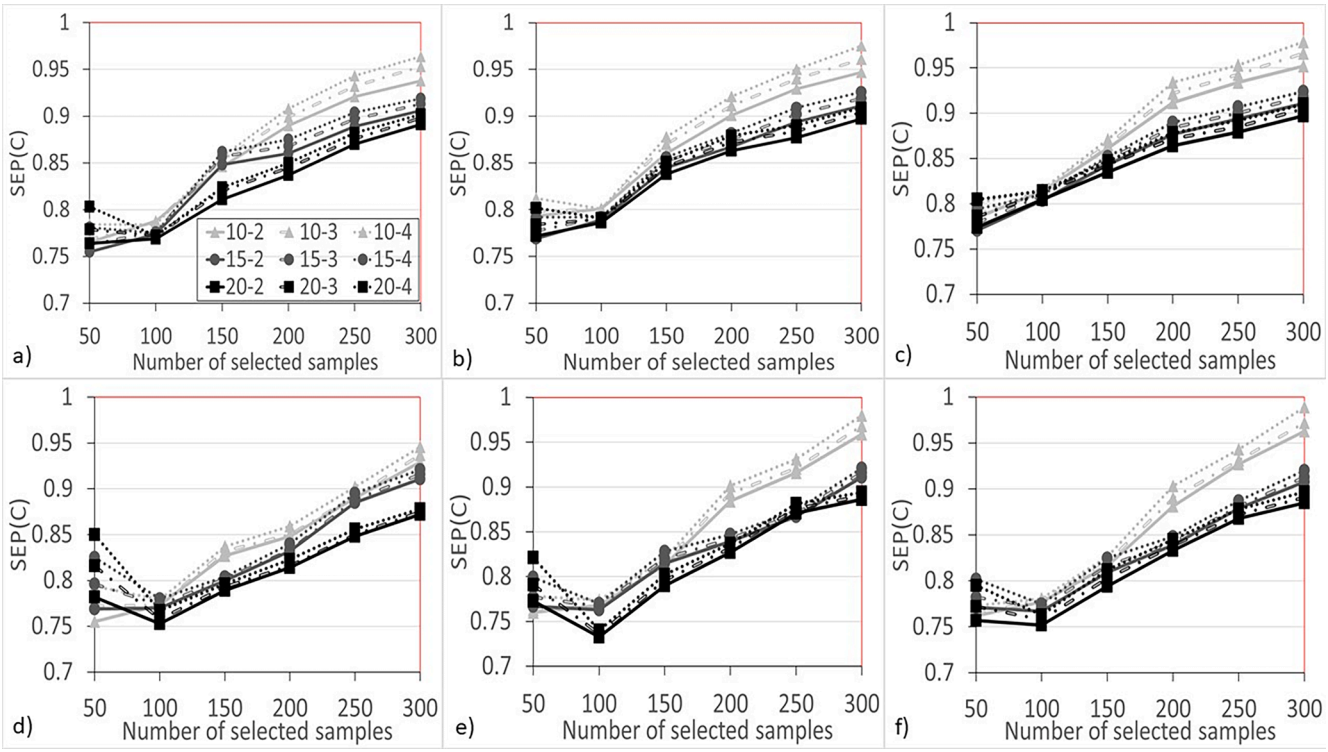


Fig. 2. SEP(C) values for LOCAL calibrations developed for CP according to maximum number of PLS terms and number of excluded terms (*l*-*m*) and number of selected samples (*k*) (horizontal axis). a) Derivative 1,5,5,1; b) derivative 1,10,5,1; c) derivative 1,10,10,1; d) derivative 2,5,5,1; e) derivative 2,10,5,1 and f) derivative 2,10,10,1.

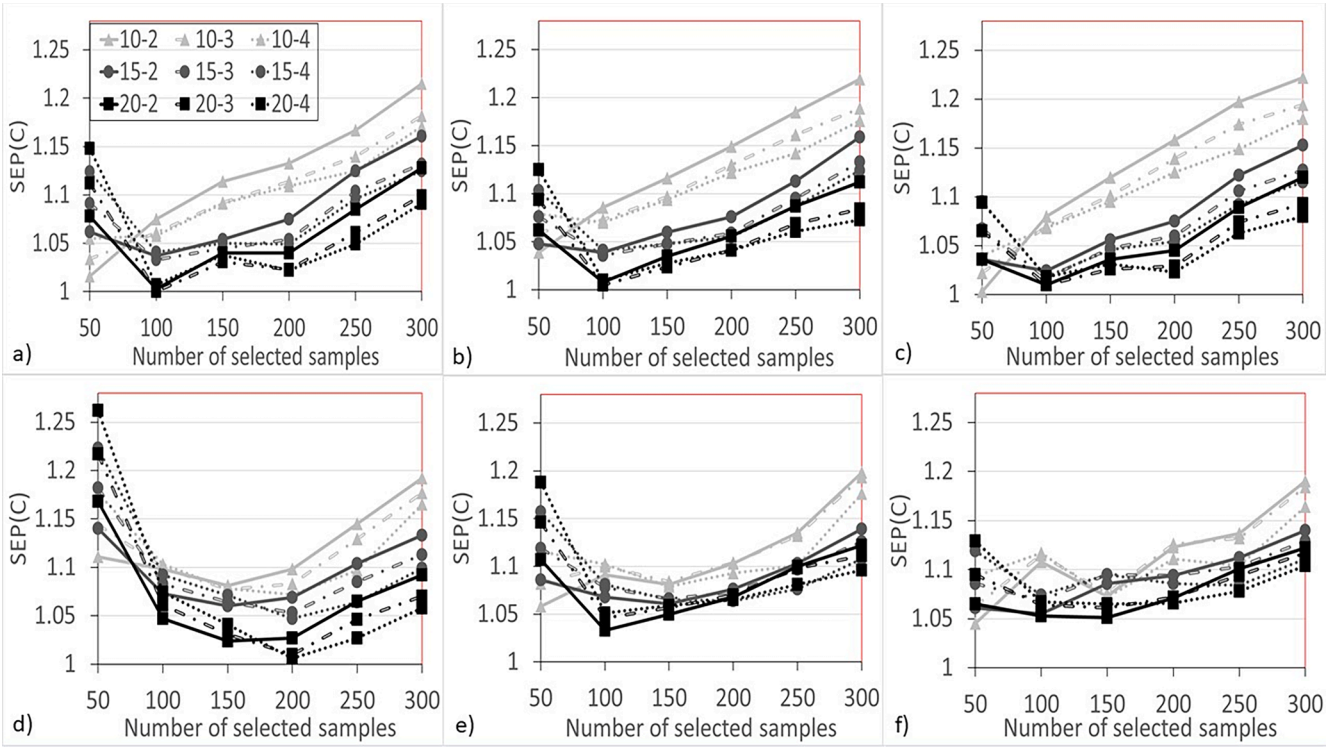


Fig. 3. SEP(C) values for LOCAL calibrations developed for Ash according to maximum number of PLS terms and number of excluded terms (*l*-*m*) and number of selected samples (*k*) (horizontal axis). a) Derivative 1,5,5,1; b) derivative 1,10,5,1; c) derivative 1,10,10,1; d) derivative 2,5,5,1; e) derivative 2,10,5,1 and f) derivative 2,10,10,1.

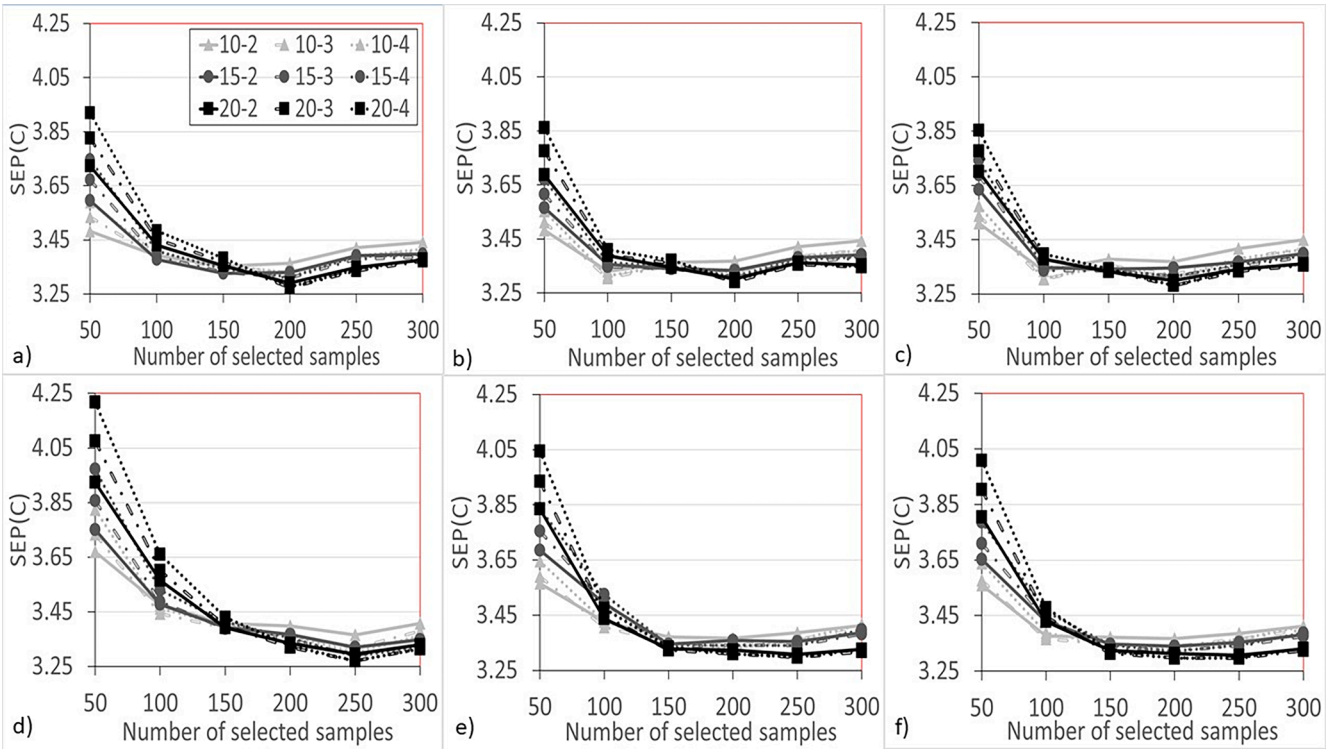


Fig. 4. SEP(C) values for LOCAL calibrations developed for iNDF according to maximum number of PLS terms and number of excluded terms (*l*-*m*) and number of selected samples (*k*) (horizontal axis). a) Derivative 1,5,5,1; b) derivative 1,10,5,1; c) derivative 1,10,10,1; d) derivative 2,5,5,1; e) derivative 2,10,5,1 and f) derivative 2,10,10,1.

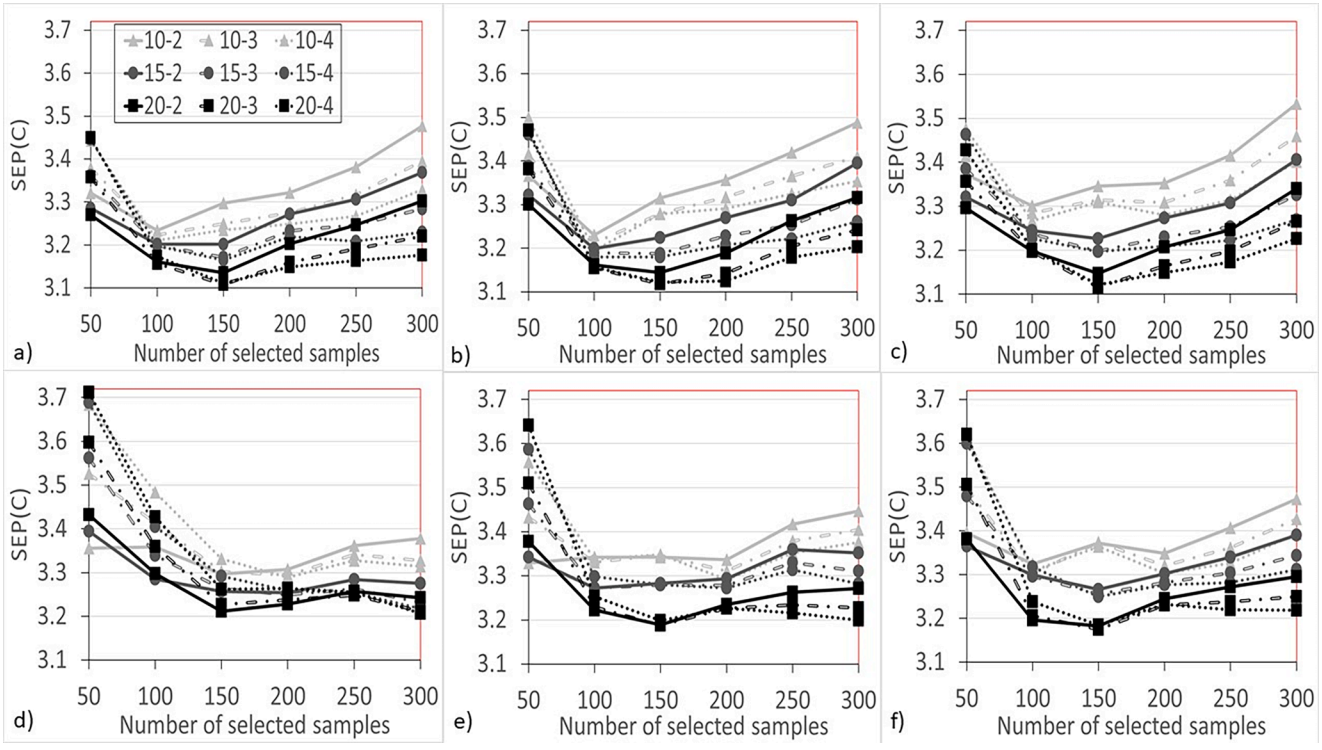


Fig. 5. SEP(C) values for LOCAL calibrations developed for NDF according to maximum number of PLS terms and number of excluded terms (*l*-*m*) and number of selected samples (*k*) (horizontal axis). a) Derivative 1,5,5,1; b) derivative 1,10,5,1; c) derivative 1,10,10,1; d) derivative 2,5,5,1; e) derivative 2,10,5,1 and f) derivative 2,10,10,1.

Table 4
Validation statistics for selected global calibrations (n = 235. Average global H = 1.01. Average neighbourhood H = 0.24).

Constituent	SEP	Bias	SEP(C)	Slope	R^2
Ash (%)	1.20	−0.01	1.20	0.92	0.80
CP (%)	0.89	−0.04	0.89	0.99	0.94
NDF (%)	3.42	−0.21	3.42	0.94	0.45*
ADF (%)	4.00	−0.17	4.01	1.03	0.64
HEM (%)	4.36	0.13	4.37	1.01	0.48*
TDN (%)	3.01	0.07	3.02	1.00	0.50*
NLE (Mcal.kg^{-1})	0.11	0.00	0.11	1.01	0.65
IVOMD (%)	2.96	0.08	2.97	1.00	0.54*
NDIP (%)	0.36	−0.01	0.36	0.94	0.67
ADIP (%)	0.23	0.00	0.23	0.90	0.72
NDFA (%)	0.73	−0.05	0.73	0.85	0.65
ADFA (%)	0.60	−0.05	0.60	0.91	0.47
NDFap (%)	3.53	−0.28	3.53	0.81	0.41*
ADFap (%)	3.53	−0.20	3.54	1.01	0.70
iNDF (%)	3.34	0.40	3.33	0.96	0.85

* Validation statistics that exceed validation limits.

Table 5
Reduction (%) in prediction errors (SEP and SEP(C)) with LOCAL vs global calibrations.

Constituent	**% SEP**	**% SEP(C)**
Ash (%)	16.94	16.81
CP (%)	17.55	16.89
NDF (%)	9.18	9.18
ADF (%)	7.46	7.52
HEM (%)	13.25	13.27
TDN (%)	6.73	6.75
NLE (Mcal.kg-1)	7.14	8.04
IVOMD (%)	6.54	6.58
NDIP (%)	14.80	14.85
ADIP (%)	5.60	5.60
NDFA (%)	14.27	14.27
ADFA (%)	9.36	9.36
NDFap (%)	6.94	6.80
ADFap (%)	5.57	5.57
iNDF (%)	1.62	0.75
Average improvement	**9.53**	**9.48**

parameters calculated from them (Hem, TDN, NLE and IVOMD) were lower than those reported in other studies [42,43], while performance for NDIP and ADIP models resulted quite similar. These results contrast with those published in previous studies for *Brachiaria brizantha* forage samples (240 from marandu grass and 120 from piatã grass), where R^2 values of 0.94, 0.88, 0.86, 0.88 and 0.88 were found for CP, NDF, ADF, IVOMD and TDN, respectively [3].

Surprisingly, the iNDF estimate was good ($r^2 = 0.89$; SECV = 2.68). This is an analytical determination subject to great variability, since it depends on the type of fistulated animal, the kind of grass used and the time for collection of the bags within the rumen. Furthermore, incubation times to obtain indigestible fractions are quite varied in the rumen (*in situ*) or *in vitro*, ranging from 144 h [44] to 288 h [31], which implies a great advantage for NIRS by reducing the time needed to estimate this constituent for elephant grass forage. Moreover, the use of NIRS to estimate iNDF can be considered as an alternative to the *in vitro* method, which would allow avoiding the use of surgically prepared animals, and therefore significantly contributing to animal welfare.

Regarding the configuration parameters of the predictive models developed with LOCAL, one of those that has shown a greater influence on the predictive errors is the number of samples to select (*k*), something already highlighted in other published works [26]. For most constituents, a number of samples close to 100 provided the best results. In general, it is quite convenient that models developed from a relatively low value of *k* perform better, since the computation time required for calculations is minimized. However, in no case the lowest number of samples (k = 50) was elected as the most appropriate.

As previously mentioned, the best validation results were obtained in most cases for the highest value of the number of PLS terms in the LOCAL models (*l* = 20). It is likely that the configuration of the LOCAL models allowing the inclusion of a higher number of PLS terms would have allowed the results to be improved even more, as reported in a study carried out with compound feed in which a maximum of 35 terms was reached [16].

The improvement in predictive results obtained in validation with LOCAL models is in line with that described by Berzaghi et al. [14], who reported improvements of 10 to 30 %, depending on the spectral diversity of the samples. The use of LOCAL in this case provides better precision and accuracy, and for the further expansion of the models is easier to manage in comparison with global models, since in this case it is only necessary to add the samples to the spectral library, avoiding recalibrations. In the future, the use of multi-products libraries – i.e. different types of forages – could be also explored and implemented.

5. Conclusions

The predictive models developed for the estimation of nutritional parameters in elephant grass provided excellent results for CP, good for Ash and iNDF and reasonable for ADIP and ADFap with global and LOCAL calibrations.

The graphical display of the SEP(C) values obtained based on the configuration parameters of the LOCAL algorithm allows decision making that optimizes the predictive ability of the models.

An adequate configuration of the spectroscopic derivative and the adjustment parameters of the LOCAL models (number of samples to select, number of PLS terms and PLS terms to be excluded from the models (k, l and m) allowed improvements in the external validation errors between 1.6 and 17.5 %.

Taking all these considerations into account, a rapid nutritional assessment of elephant grass samples could be performed with NIRS LOCAL models, avoiding the recalibration process from one year to the next and minimizing the use of cannulated animals for forage evaluation.

CRediT authorship contribution statement

Víctor M. Fernández-Cabanás: Conceptualization, Formal analysis, Writing – original draft, Visualization. **Dolores C. Pérez-Marín:** Investigation, Writing – original draft, Writing – review & editing. **Tom Fearn:** Writing – review & editing. **Joadil Gonçalves de Abreu:** Methodology, Writing – review & editing.

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.

Acknowledgements

To the financial support of the Tordesillas Group and the Carolina Foundation for the financing of the Postdoctoral Scholarship of Dr. Gonçalves de Abreu, according to Call C.2019 (Teaching Mobility Program Universities of the Tordesillas Group). To the Escuela Técnica Superior de Ingeniería Agronómica of the University of Seville for the opportunity to do a Postdoctorate research stay in NIRS Technology.

References

[1] Â.F. Rech, Amostragem de alimentos para análise bromatológica, Agropecuária Catarinense. 31 (2018) 33–36, https://doi.org/10.22491/rac.2018.v31n1.1.

[2] R.C. Amaral, B.F. Carvalho, D.M. Costa, M.J.F. Morenz, R.F. Schwan, C.L. da S. Ávila, Novel lactic acid bacteria strains enhance the conservation of elephant grass silage cv. BRS Capiaçu, Anim. Feed Sci. Technol. 264 (2020) 114472, doi: 10.1016/j.anifeedsci.2020.114472.

[3] W.F. Anderson, B.S. Dien, S.D. Masterson, R.B. Mitchell, Development of near-infrared reflectance spectroscopy (NIRS) calibrations for traits related to ethanol conversion from genetically variable Napier Grass (Pennisetum purpureum Schum.), Bioenergy Res. 12 (2019) 34–42, https://doi.org/10.1007/s12155-018-9946-8.

[4] J.D. Cleland, E. Johnson, P.C.H. Morel, P.R. Kenyon, M.R. Waterland, Mid-infrared reflectance spectroscopy as a tool for forage feed composition prediction, Anim. Feed Sci. Technol. 241 (2018) 102–111, https://doi.org/10.1016/j.anifeedsci.2018.04.022.

[5] V. Decruyenaere, P. Lecomte, C. Demarquilly, J. Aufrere, P. Dardenne, D. Stilmant, A. Buldgen, Evaluation of green forage intake and digestibility in ruminants using near infrared reflectance spectroscopy (NIRS): developing a global calibration, Anim. Feed Sci. Technol. 148 (2009) 138–156, https://doi.org/10.1016/j.anifeedsci.2008.03.007.

[6] 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, https://doi.org/10.1016/j.tifs.2006.09.003.

[7] L.S. Oliveira, A.S. Franca, Applications of near infrared spectroscopy (NIRS) in food quality evaluation, in: D.A. Medina, A.M. Laine (Eds.), Food Qual. Control. Anal. Consum. Concerns, Nova Science Publishers, Inc., 2011, pp. 131–180.

[8] G.C. Marten, J.S. Shenk, F.E. Barton II, Near Infrared Reflectance Spectroscopy (NIRS): analysis of forage quality. U:S: Departmen of Abriculture, Agriculture Handbook Nr. 643, U.S. Dep. Agric. Agric. Handb. 643 (1985) 96.

[9] M. Herrero, N.S. Jessop, R.H. Fawcett, I. Murray, J.B. Dent, Prediction of the in vitro gas production dynamics of kikuyu grass by near-infrared reflectance spectroscopy using spectrally-structured sample populations, Anim. Feed Sci. Technol. 69 (1997) 281–287, https://doi.org/10.1016/S0377-8401(97)81642-2.

[10] F. Mejía, I. Yoplac, W. Bernal, W. Castro, Evaluación de modelos de predicción de composición química y energía bruta de kikuyo (Pennisetum clandestinum) usando espectroscopía en infrarrojo cercano (NIRS), Rev. Investig. Vet. Del Perú. 30 (2019) 1068–1076, https://doi.org/10.15381/rivep.v30i3.16598.

[11] J.S. Shenk, M.O. Westerhaus, P. Berzaghi, Investigation of a LOCAL calibration procedure for near infrared instruments, J. Near Infrared Spectrosc. 5 (1997) 223–232, https://doi.org/10.1255/jnirs.115.

[12] F. Davrieux, D. Dufour, P. Dardenne, J. Belalcazar, M. Pizarro, J. Luna, L. Londoño, A. Jaramillo, T. Sanchez, N. Morante, F. Calle, L.A. Becerra Lopez-Lavalle, H. Ceballos, LOCAL regression algorithm improves near infrared spectroscopy predictions when the target constituent evolves in breeding populations, J. Near Infrared Spectrosc. 24 (2016) 109–117, https://doi.org/10.1255/jnirs.1213.

[13] F.E. Barton, J.S. Shenk, M.O. Westerhaus, D.B. Funk, The development of near infrared wheat quality models by locally weighted regressions, J. Near Infrared Spectrosc. 8 (3) (2000) 201–208.

[14] P. Berzaghi, J.S. Shenk, M.O. Westerhaus, LOCAL prediction with near infrared multi-product databases, J. Near Infrared Spectrosc. 8 (2000) 1–9, https://doi.org/10.1255/jnirs.258.

[15] E. Zamora-Rojas, A. Garrido-Varo, F. Van den Berg, J.E. Guerrero-Ginel, D. C. Pérez-Marín, Evaluation of a new local modelling approach for large and heterogeneous NIRS data sets, Chemom. Intell. Lab. Syst. 101 (2010) 87–94, https://doi.org/10.1016/j.chemolab.2010.01.004.

[16] E. Fernández-Ahumada, T. Fearn, A. Gómez-Cabrera, J.E. Guerrero-Ginel, D. C. Pérez-Marín, A. Garrido-Varo, Evaluation of local approaches to obtain accurate Near- Infrared (NIR) equations for prediction of ingredient composition of compound feeds, Appl. Spectrosc. 67 (2013) 924–929, https://doi.org/10.1366/12-06937.

[17] D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero, Implementation of LOCAL algorithm with near-infrared spectroscopy for compliance assurance in compound feedingstuffs, Appl. Spectrosc. 59 (2005) 69–77, https://doi.org/10.1366/0003702052940585.

[18] I. Torres, M.T. Sánchez, M.J. de la Haba, D. Pérez-Marín, LOCAL regression applied to a citrus multispecies library to assess chemical quality parameters using near infrared spectroscopy, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 217 (2019) 206–214, doi: 10.1016/j.saa.2019.03.090.

[19] M.T. Sánchez, I. Torres, M.J. De la Haba, D. Pérez-Marín, First steps to predicting pulp colour in whole melons using near-infrared reflectance spectroscopy, Biosyst. Eng. 123 (2014) 12–18, https://doi.org/10.1016/j.biosystemseng.2014.04.010.

[20] M.T. Sánchez, M.J. De la Haba, J.E. Guerrero, A. Garrido-Varo, D. Pérez-Marín, Testing of a local approach for the prediction of quality parameters in intact nectarines using a portable NIRS instrument, Postharvest Biol. Technol. 60 (2011) 130–135, https://doi.org/10.1016/j.postharvbio.2010.12.006.

[21] V. González-Caballero, D. Pérez-Marín, M.I. López, M.T. Sánchez, Optimization of NIR spectral data management for quality control of grape bunches during on-vine ripening, Sensors. 11 (2011) 6109–6124, https://doi.org/10.3390/s110606109.

[22] R.G. Dambergs, D. Cozzolino, W.U. Cynkar, L. Janik, M. Gishen, The determination of red grape quality parameters using the LOCAL algorithm, J. Near Infrared Spectrosc. 14 (2006) 71–79, https://doi.org/10.1255/jnirs.593.

[23] M.J. De la Haba, D. Pérez-Marín, D. Rial-Huerta, M.T. Sánchez, Texture prediction in intact green asparagus by near infrared (NIR) spectroscopy, assaying linear and non-linear regression strategies, Food Anal. Methods 7 (2014) 606–615, https://doi.org/10.1007/s12161-013-9662-0.

[24] D. Andueza, F. Picard, W. Martin-Rosset, J. Aufrère, Near-infrared spectroscopy calibrations performed on oven-dried green forages for the prediction of chemical composition and nutritive value of preserved forage for ruminants, Appl. Spectrosc. 70 (2016) 1321–1327, https://doi.org/10.1177/0003702816654056.

[25] B. Godin, R. Agneessens, J. Delcarte, P. Dardenne, Prediction of chemical characteristics of fibrous plant biomasses from their near infrared spectrum: Comparing local versus partial least square models and crossvalidation versus independent validations, J. Near Infrared Spectrosc. 23 (2015) 1–14, https://doi.org/10.1255/jnirs.1138.

[26] D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero, Non-linear regression methods in NIRS quantitative analysis, Talanta 72 (2007) 28–42, https://doi.org/10.1016/j.talanta.2006.10.036.

[27] K. Helrich, A. of O.A. Chemists., Official methods of analysis of the Association of Official Analytical Chemists, Association of Official Analytical Chemists, Arlington, Va., 1990.

[28] G. Licitra, T.M. Hernandez, P.J. Van Soest, Standardization of procedures for nitrogen fractionation of ruminant feeds, Anim. Feed Sci. Technol. 57 (1996) 347–358, https://doi.org/10.1016/0377-8401(95)00837-3.

[29] P.J. Van Soest, Nutritional Ecology of the Ruminant, Comstock Pub, Ithaca, 1994.

[30] R.C. Cochran, D.C. Adams, J.D. Wallace, M.L. Galyean, Predicting digestibility of different diets with internal markers: evaluation of four potential markers, J. Anim. Sci. 63 (1986) 1476–1483, https://doi.org/10.2527/jas1986.6351476x.

[31] E. Detmann, M.A. de Souza, S. de C. Valadares Filho, A.C. de Queiroz, T.T. Berchielli, E. de O.S. Saliba, L. da S. Cabral, D. dos S. Pina, M.M. Ladeira, J.A.G. Azevedo, Métodos para análise de alimentos, 2012.

[32] E.R. Cappelle, S.D.C. Valadares Filho, J.F.C. Da Silva, P.R. Cecon, Estimativas do Valor Energético a partir de Características Químicas e Bromatológicas dos Alimentos, Rev. Bras. Zootec. 30 (2001) 1837–1856, https://doi.org/10.1590/s1516-35982001000700022.

[33] L. Pieszczek, H. Czarnik-Matusewicz, M. Daszykowski, Identification of ground meat species using near-infrared spectroscopy and class modeling techniques – aspects of optimization and validation using a one-class classification model, Meat Sci. 139 (2018) 15–24, https://doi.org/10.1016/j.meatsci.2018.01.009.

[34] J.S. Shenk, M.O. Westerhaus, S.M. Abrams, Protocol for NIR calibrations: monitoring analysis results and recalibration, II near Infrared Spectrosc, Anal. Forage Qual. Agric. Handb. 643 (1989) 104–110.

[35] J.S. Shenk, M.O. Westhaus, Analysis of agriculture and food products by near infrared reflectance spectroscopy. Monograph NIR systems, Infrasoft International, NIR Systems, Inc. Silver Spring, MD, USA., 1995.

[36] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777, https://doi.org/10.1366/0003702894202201.

[37] 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, https://doi.org/10.1016/j.geoderma.2018.09.049.

[38] P. Williams, Implementation of Near-infrared Technology, in: P.C. Williams, K. Norris (Eds.), Near-Infrared Technol. Agric. Food Ind., Second Edi, American Assoc. of Cereal ChemiSts Inc., St, Paul, Minnesota, USA, 2001, p. 143.

[39] J.S. Shenk, M.O. Westerhaus, Calibration the ISI way, in: A.M.C. Davis, P. C. Williams (Eds.), Near Infrared Spectrosc. Futur. Waves, NIR publications, Chichester, U.K, 1996, pp. 198–202.

[40] I. Lobos, P. Gou, S. Hube, R. Saldaña, M. Alfaro, Evaluation of potential nirs to predict pastures nutritive value, J. Soil Sci. Plant Nutr. 13 (2013) 463–468, https://doi.org/10.4067/S0718-95162013005000036.

[41] M. Herrero, I. Murray, R.H. Fawcett, J.B. Dent, Prediction of the in vitro gas production and chemical composition of kikuyu grass by near-infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 60 (1996) 51–67, https://doi.org/10.1016/0377-8401(95)00924-8.

[42] H. Tran, P. Salgado, P. Lecomte, Species, climate and fertilizer effects on grass fibre and protein in tropical environments, J. Agric. Sci. 147 (2009) 555–568, https://doi.org/10.1017/S0021859609008752.

[43] C.C. Serafim, G.L. Guerra, I.Y. Mizubuti, F.A.B.d. Castro, O.P. Prado-Calixto, S. Galbeiro, A.R.P. Parra, V.H. Bumbieris Junior, S.F.N. Pértile, F.C.d.A. Rego, Use of near-infrared spectroscopy for prediction of chemical composition of Tifton 85 grass, Semin. Agrar. 42 (3) (2021) 1287–1302.

[44] T.T. Berchielli, S.G. Oliveira, A.V. Garcia, Application of techniques for intake, diet composition and digestibility studies, Arch. Vet. Sci. 10 (2005) 29–40.