PREDICTION OF THE COMPOSITION OF LARGE POACEAE: *can multi-species calibrations be used for extrapolation to new species?*

L. Bonnal¹, D. Pot², J-F. Rami², J-Y. Hoarau³, P. Malvoisin⁴, D. Bastianelli^{1*}

¹ CIRAD, UMR SELMET, Baillarguet, F-34398 Montpellier Cedex 5, France

² CIRAD, UMR AGAP, F-34398 Montpellier Cedex 05, France

³ CIRAD, UMRAGAP - Ercane, F-97494 Sainte-Clotilde Cedex, La Réunion

⁴ AELRED, F-91058 Evry Cedex, France

Corresponding author: denis.bastianelli@cirad.fr

ABSTRACT

This study tested strategies for the prediction by NIRS of the chemical composition and nutritive value of large poaceae (sorghum, millet, sugarcane, miscanthus etc.). The main objective was to evaluate the possibility of prediction of samples when no specific calibration is available because the plant part of the species is not common enough.

The first strategy was to build a global calibration gathering several species. With a database of 750 samples from 8 species or plant parts, the precision of the models was close to the values usually recorded in single-species calibrations.

The second strategy was to test extrapolation capacity by removing one species from the previous database and validate on this species. The results depended on the species and on the parameters: while good results were obtained for protein content in most cases, results were less precise for more aggregative criteria (fibre, digestibility) particularly when removing species with extreme (high or low) values.

KEYWORDS: Chemical composition, Extrapolation, Crop residues, Near Infrared Spectroscopy

INTRODUCTION

Large poaceae such as sorghum, millet or sugarcane are widely cultivated in tropical areas, where their by-products (stems, straw, etc.) are often used for animal feeding. These products also have a potential as energy source or as organic matter source in soils.

In order to optimize the use of these biomass sources, it is important to characterize their chemical composition, degradability and/or nutritional value. For routine analysis of such resources, NIRS is a powerful analytical tool. Specific NIR calibrations exist for the most important species (e.g. whole crop maize, Cozzolino et al., 2001). By-products and less common species have a potential of use in some contexts but we often lack robust calibrations to characterize them. This study was designed to evaluate the possibility of predicting less common samples from a database gathering various species.

MATERIALS AND METHODS

Samples

A calibration database was created by gathering spectra from various large poaceae species: sorghum, millet, sugarcane, maize, Guinea grass (*Panicum maximum*), elephant grass (*Pennisetum purpureum*), miscanthus (*Miscanthus* sp.) from various tropical countries and from France. The samples (750) mainly concerned crop by-products or residues, or plants harvested at maturity (miscanthus).

Analytical parameters

The analytical parameters measured were: Dry matter (DM, 103°C), total minerals (ASH, 550°C), crude protein (CP, Kjeldahl; NF EN ISO 5983-2), Van Soest fiber fractions (NDF, ADF, ADL; Van Soest et al., 1991), in vitro dry matter digestibility determined by pepsine-cellulase method (IVDMD; Aufrère et al., 2007).

NIRS and calibration

Spectra were collected on dried and ground samples (1mm sieve) on a FOSS Nirsystem 5000 spectrometer (FOSSLaurel, MD, USA) in reflectance mode. Spectra were collected in duplicate in standard circular cups with quartz window (diameter 3.7cm) and averaged. A mathematical pretreatment was applied to spectra: second derivative, detrending and normalization (SNV). Calibration was performed with PLS regression, using the MPLS procedure in WINISI software. Cross-validation was performed on four subgroups selected at random.

The first approach was to test a general calibration including all species. A validation subset was created by choosing randomly 100 samples and calibrating with the 650 remaining samples. Calibration performance was evaluated by R^2 , standard error of calibration (SEC), cross validation (SECV). Prediction of the 100 validation samples led to the standard error of prediction (SEP), and the bias and slope of the regression between predicted and measured values. RPD was defined as the ratio between SD of the population and the SEP.

The second approach was to remove all the samples from one species (or plant part) from the calibration database and to use them as validation for equation derived from all other samples, in order to test an extrapolation capacity. This procedure was done successively for all the species.

RESULTS AND DISCUSSION

The statistical distance (Mahalanobis distance between groups) between the species showed a relatively homogenous cluster between sorghum, sugarcane, maize and millet. Miscanthus was clearly separated but it was closer from sugarcane than from maize and sorghum. Elephant grass, harvested at a less advanced stage, was closer from maize or millet than from sorghum or miscanthus.

First approach – global calibration

Calibration on the multi-species databases (Table 1) led to SECV values close to usual value for the same constituents in plant databases (e.g. Sabatier et al., 2012, Garcia and Cozzolino, 2006). The RPD values were quite high for NDF, ADF or IVDMD thanks to the high variability of the database. However the SEP values were higher than SECV values (ratio 1.2 to 1.8) showing that the calibrations were still not completely robust despite the large number of samples in the database.

These results indicate that for samples of the same nature as present in the database, composition can be predicted with an acceptable accuracy.

Table 1. Calibration statistics on the global (multispecies) calibration database

Constituent	Population (n=750)		Calibration (n=650)			Validation (n=100)			
	Mean	SD	R^2_{cal}	SEC	SECV	SEP	Bias	Slope	RPD
Ash (%DM)	5.2	3.3	0.97	0.58	0.68	1.02	- 0.15	0.94	3.2
Protein (%DM)	5.4	3.9	0.99	0.44	0.50	0.91	0.16	0.98	4.3
NDF (%DM)	67.6	16.1	0.99	1.87	2.02	2.50	- 0.07	0.99	6.5
ADF (%DM)	40.3	13.6	0.99	1.10	1.21	1.82	0.15	0.98	7.5
ADL (%DM)	6.2	3.5	0.97	0.56	0.66	0.85	0.00	0.95	4.1
IVDMD (%)	42.9	19.7	0.99	2.11	2.34	3.19	0.33	0.98	6.2

SEC: Standard error of calibration; SECV: Standard error of cross-validation; SEP: Standard error of prediction; Bias, slope: parameters of the regression between measured and predicted values; RPD=SD/SEP

Second approach – leave one species out

Extrapolation success depended on the block (species or plant part) used for validation, and on the analytical parameter considered.

Prediction of CP content (Figure 1a) was generally good, even in extrapolation, with SEP varying between 0.5% and 1.5% across the species. For some species (e.g. Panicum) a significant bias was observed. The regression derived from all samples was $CP_{nir}=0.93 \cdot CP_{lab}+0.14$ ($R^2=0.93$) with a RMSE of 0.96, close to the SEP found in the global calibration (Table 1).

Figure 1. Results of extrapolation prediction for CP, NDF, ADL and IVDMD. Each species is predicted with an equation built without this species

More aggregative criteria such as NDF or IVDMD had higher SEP values when validating with some of the datasets, especially sorghum or sugarcane. In the case of IVDMD the overall regression was satisfactory ($R^2=0.86$, slope=0.91) but within some of the groups the slopes were significantly different from 1. This was particularly well seen in *Miscanthus* (slope=2.48; bias=1.8) which is an extreme for low IVDMD values. Calibration for ADL was less precise, as it is generally the case, and significant biases were observed in some of the groups. Although this was not the objective of the current study, “local” calibration strategies, based on the closest neighbors of the sample to be predicted, could be tested for extrapolation in this type of databases (Berzaghi et al., 2000).

CONCLUSION

On a practical point of view, the trial showed that there is some extrapolation capacity in the “large poaceae” dataset, allowing to estimate the composition of samples from species not present in the calibration database (e.g. giant reed). Inclusion of a limited number of samples from an additional species in the database allows predicting samples from this species with a general database without requiring the development of a new specific database. This is particularly useful for less studied species for which no calibration is available.

Acknowledgments

This work was done in the framework of the project BFF (Biomass for the Future) funded by the French National Research Agency ANR, France

References

1. Aufrère, J.; Baumont, R.; Delaby, L.; Peccatte, J-R.; Andieu, J.; Andrieux, J-P.; Dulphy, J-P., Laboratory prediction of forage digestibility by the pepsin-cellulase method. The renewed equations. INRA Prod. Anim., 2007, 20 (2), 129–36.
2. Berzaghi, P.; Shenk, J.S.; Westerhaus, M.O., LOCAL prediction with near infrared multi-product databases. J Near Infrared Spectrosc., 2000, 8(1):1–9. doi:10.1255/jnirs.258
3. Cozzolino, D.; Fassio, A.; Gimenez, A., The use of near-infrared reflectance spectroscopy (NIRS) to predict the composition of whole maize plants. J. Sci. Food Agric., 2001, 81(1): 142-146.
4. Garcia, J.; Cozzolino, D., Use of near infrared reflectance (NIR) spectroscopy to predict chemical composition of forages in broad-based calibration models. Agric. Tecnica, 2006, 66(1), 41-47.
5. Sabatier, D.; Thuriès, L.; Bastianelli, D.; Dardenne, P., Rapid prediction of the lignocellulosic compounds of sugarcane biomass by near infrared reflectance spectroscopy: comparing classical and independent cross-validation. J. Near Infrared Spectrosc. 2012, 20, 371–85. doi:10.1255/jnirs.999
6. Van Soest PJ, Robertson JB, Lewis BA. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. J Dairy Sci, 1991, 74, 3583–97. doi:10.3168/jds.S0022-0302(91)78551-2