Prediction of the composition of large poaceae : can multi-species calibrations be used for extrapolation to new species ?

L. Bonnal[a], D. Pot[b], J-F. Rami[b] , J-Y. Hoareau[c], P. Malvoisin[d], D. Bastianelli[a]

[a]CIRAD, UMR SELMET Animal feed laboratory, 34398 Montpellier Cedex 5, France. [b]CIRAD, UMR AGAP, Montpellier, France. [c]CIRAD, UMR AGAP – ercane, La Réunion. [d]AELRED, Evry, France

Contact: laurent.bonnal@cirad.fr

Large poaceae as sorghum, millet or sugarcane are widely cultivated in tropical areas, where their byproducts (stems / straw, etc.) are often used for animal feeding, but also have a potential as energy source or organic matter source in soils. It is important to characterize their chemical composition, degradability and nutritional value. Specific NIR calibrations exist for the most important species (e.g. whole crop maize). Byproducts 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

A calibration database was created by gathering spectra from various large poaceae species: sorghum, millet, sugarcane, maize, elephant grass, miscanthus from various tropical countries and southern France. The samples (750) mainly concerned crop by products or plants harvested at maturity (miscanthus). Spectra were collected on dried and ground samples on a FOSS Nirsystem 5000 spectrometer in reflectance mode. Calibration was performed with PLS regression, based on spectra with mathematical pretreatment (2nd derivative + SNV and detrending).

The parameters measured were : Dry matter (DM), total minerals (ASH), crude protein (CP), Van Soest fiber fractions (NDF, ADF, ADL), *in vitro* dry matter digestibility (IVDMD). The first approach was to test a general calibration. A validation subset was created by choosing randomly 100 samples and calibrating with the 650 remaining samples. The second approach was to remove one species (or plant part) of the calibration database and use it as validation for equation derived from the other samples, in order to test an extrapolation capacity.

Results and discussion

The statistical distance between the species showed a relatively homogenous cluster between sorghum, sugarcane, maize and millet. Miscanthus 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.

Calibration on the multi-species databases (Table 1) showed SECV values close to usual value for the same constituents in plant databases. However the SEP values higher than SECV showed that the calibration is still not completely robust. However the RPD (=SD/SEP) values were quite high thanks to the high variability of the database.

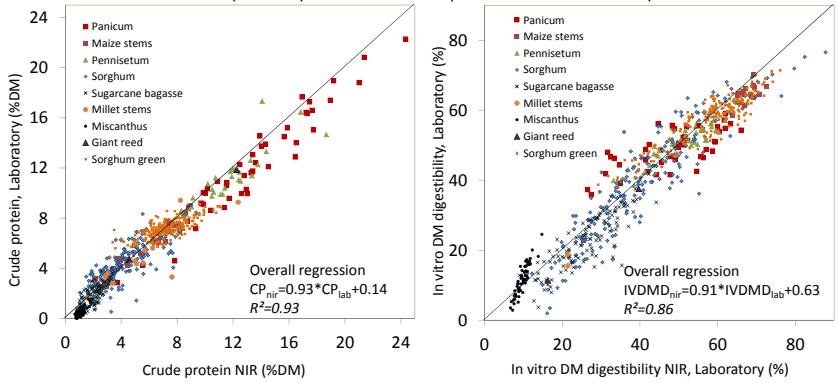
These results indicate that for samples of the same nature as present in the database, composition can be predicted with an acceptable accuracy.

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

	Population (n=750)		Calibration (n=650)			Validation (n=100)			
Constituent	*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	-	0.95	4.1
IVDMD (%)	42.9	19.7	0.99	2.11	2.34	3.19	0.33	0.98	6.2

SD: standard deviation; SEC: standard error of calibration; R^2cal: coefficient of determination of calibration ; SECV: standard error of cross validation; SEP: standard error of prediction; RPD=SD/SEP. NDF: Neutral detergent fiber; ADF: Acid detergent fibre; ADL: Acid detergent lignin; IVDMD: *In vitro* DM digestibility.

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



Extrapolation success depended on the dataset used for validation, and on the parameter considered. Prediction of CP content (Figure 1) was generally good, even in extrapolation, with SEP varying between 0.5% and 1.5%. For some species (e.g. Panicum) a significant bias was observed

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 groups the slopes were significantly different from 1.

"Local" calibration strategies, based on the closest neighbors of the sample to be predicted, could also be used in this type of databases.

Conclusion

On a practical point of view, the trial showed that there is some extrapolation capacity in the "large poaceae" dataset, allowing estimating 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 without requiring the development of a new specific database.

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