Geophysical Research Abstracts
Vol. 13, EGU2011-12699-2, 2011
EGU General Assembly 2011
© Author(s) 2011

Spiking as strategy to extend NIR models: effects of the type of samples added and weight in model

César Guerrero (1) and Raúl Zornoza (2)

(1) University Miguel Hernández, Agrochemistry and Environment, Elche, Spain (cesar.guerrero@umh.es), (2) Technical University of Cartagena, Department of Agrarian Science and Technology, Cartagena, Spain

The near infrared reflectance (NIR) spectroscopy technique needs models (calibrations) relating the NIR spectra with the analytical data. These calibrations should contain the variability of the target sites soils on which the models are to be used. Many times this premise is not easy to fulfil. A classical way to solve this problem is by the spiking of model, which is the addition of a few samples from the target sites, and the subsequent recalibration of the model. But there are scarce studies focused on what type of samples should be used for model spiking. We hypothesised that each sample added could cause a different impact on the model accuracy. For these reasons, we considered interesting to evaluate what type samples could be more useful to spike a model when it is going to be applied to a new target site (TS).

Two sets of samples were used in this study, being the soil organic carbon (SOC) the target parameter:
i) Spectral library (SL): set of 365 soil samples. Four different models for SOC were constructed (PLS regression) using different samples from this set.
ii) Target site (TS) set: set of 155 samples from an area not previously sampled (not included in SL).
In both sets, the SOC contents (Walkley-Black), and the NIR spectra (830-2630 nm) of the soil samples were measured.
In order to study the effect of type of samples added for spiking, thirteen different methods (or strategies) of TS sample selection were compared. These thirteen strategies were designed, in essence, on the basis of the SOC values of TS set, the spectral characteristics of TS set and the spectral relationships between models and TS samples (Mahalanobis distance). In each strategy, 8 and 16 TS samples were selected for the spiking of these four models.
In general, the SOC estimations in the TS set were clearly less accurate (in terms of RMSEP) using unspiked than using spiked models. As expected, the increase in the accuracy was dependent of the strategy (or method) of sample selection (i.e. type of TS samples added), since there were important differences in accuracy between strategies. The magnitude of the increase in the accuracy was also dependent of the model type, but interestingly, the worst and the best strategies (in terms of accuracy) were the same across models. This suggest that the initial characteristics of models are important, but in a lesser degree than type of TS sample added. As expected, the higher the number of samples used for spiking the higher the accuracy of the spiked models.

We also studied the changes in the accuracy of the spiked models when the weight of the TS added samples was increased. In order to increase the weight of the TS added samples in the models, several copies of the TS added samples were included in the spiked models. An increase of the weight of those TS added samples can increase their importance in the model and as consequence, the model should be more adequate (accurate) for this type of samples. If these samples are representative of the overall TS set, consequently the model should be more accurate for the overall target set. So, we hypothesised that an increase in the weight could be an inexpensive way to enhance the accuracy of the spiked models. The results confirmed that the increase of the weight had positive effects on the accuracy of the spiked models. In some of the strategies of TS sample selection, the increase of the weight had a similar (positive) effect on accuracy than the duplication of the number of TS samples added (from 8 to 16).

The selection of the most adequate samples for an optimal spiking of models should be a relevant topic of research in the future due to large differences in results. The increase in the weight of the added samples is an inexpensive way to increase the accuracy of the spiked models.

Authors acknowledge to “Bancaja-UMH” for the financial support of the project “NIR-PROS”.