Geophysical Research Abstracts
Vol. 13, EGU2011-9801-1, 2011
EGU General Assembly 2011

Extending the geographical validity of NIR models by spiking: can hybrid spectra act as bridges between sets? A case study for soil organic carbon

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 site (TS), and the subsequent recalibration of the model. The disadvantage of spiking is the need to analyse these few TS samples using the reference (or classical) laboratory methods. Thus, the number of samples added should be as low as possible in order to maintain the advantages of NIR spectroscopy. As consequence, the proportion of added TS samples respect total samples in model is low. If the spectral characteristics of both sets (model and TS) are clearly different, there could be a great spectral distance between these sets. Since the number of added TS samples is low in comparison with the total number of samples in model, the TS added samples (i.e. used for spiking) could be located at the boundaries of the spectral space of the model or even recognized as outliers. As consequence, the accuracy of the spiked model in the rest of TS samples could be low.
For this reason we hypothesised that an increase in the density of spectra located in the spectral space between both sets (model and TS) could improve the accuracy of the spiked model. In order to increase the spectral density in that space, we suggest the use of hybrid spectra for this purpose. We defined 'hybrid spectrum' to the averaged spectrum of two spectra of different samples, and assuming that concentration of the averaged spectrum is also the average of the original concentration values.
In this study, hybrid spectra were obtained averaging spectra from both sets (TS and model). Several hybrid spectra were derived from each of the TS sample used to spike the models. Interestingly, the hybrid spectra share characteristics of the samples from the model and also characteristics of TS samples (i.e. both sets). As consequence, we hypothesised that hybrid spectra in spiked models could increase the density in the spectral space located between sets, acting as 'bridges', and improving the accuracy of the estimations in the overall TS set.
In order to test these hypotheses, two different NIR models were constructed with samples from a library (n=365) and were used to predict the soil organic carbon (SOC %) content in samples from a new target site (TS) (n=155). One of these models was constructed with all the samples of the library (n=365 samples). The other model was constructed using only the most dissimilar samples (respect TS set), in order to generate the most adverse scenario (n=192 samples); in this case both sets (model and TS) were clearly different. These models were spiked with 8 TS samples, which represents around the 5% of the overall TS set. Thirteen different methods were used to select these 8 TS samples used for spiking. Moreover, several hybrid spectra were also added to these spiked models.
In general, the estimations of SOC in the TS set were less accurate (in terms of RMSEP) using unspiked than using spiked models. The improvement in RMSEP was dependent of the type of TS added (i.e. dependent of the TS selection method) and also dependent of the different characteristics of the models. As expected, the presence of hybrid spectra improved the accuracy of the spiked models. In some cases, the improvement in accuracy was due to a reduction bias and in other cases was due to lower SEP values.
Our results suggest the use of this new strategy which can improve the accuracy of NIR spiked models, allowing the extension of their geographical validity. More studies are required to confirm if the patterns observed here are similar in other regions and other soil properties, and if it can be extrapolated to other scales.

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