Comparison between Principal Component Regression, Partial Least Squares Regression and Artificial Neural Network analyses of vis-NIR spectra for the prediction of selected soil properties

B. Kuang and A. Mouazen

(b.kuang@cranfield.ac.uk)

The selection of multivariate calibration method is one of the main components influencing the measurement accuracy with visible (vis) and near infrared (NIR) spectroscopy. Agricultural soils are very heterogeneous and complex materials, often imposes challenges to scientists involved in developing vis-NIR calibration models of soil properties. This paper aims at comparing the performance of three multivariate calibration methods, namely, Principle Component Regression (PCR), Partial Least Squares Regression (PLSR) and Back Propagation Neural Network (BPNN) to build calibration models for selected soil attributes, namely, organic carbon (OC), extractable phosphorous (P), sodium (Na), potassium (K) and magnesium (Mg). A total of 170 soil samples collected from Belgium and Northern France were used as the data set for the calibration-validation procedure. Sample scanning was done on fresh soil samples with a fibre-type, vis-NIR FieldSpec©3 (350-2500 nm) spectrophotometer from ASD (Colorado, USA) with a measurement range of 350 – 2500 nm. In the first calibration procedure, the entire data set was split randomly into 90% for calibration and 10% validation. The input of BPNN was the first 5 principal components (PCs) resulted from the Principle Component Analysis (PCA) or the optimal number of latent variables (LVs) obtained from PLSR. Both the leave-one-out cross validation and independent validation results showed that all BPNN models outperformed PCR models and PLSR models. Furthermore, BPNN models whose inputs were LVs had superior performance than corresponding BPNN models whose inputs were PCs. The prediction accuracy of BPNN LVs based modelling for OC ($R^2 = 0.88$ and ratio of prediction deviation (RPD) = 2.74) and Mg ($R^2 = 0.84$ and RPD = 2.61) was classified as good, whereas it was classified as to provide approximate quantitative prediction for K, P and Na ($R^2 = 0.71 – 0.77$ and RPD = 1.87 – 1.98). PLSR provided smaller accuracy than BPNN but showed better performance than PCR. It is recommended to adopt BPNN for vis-NIR spectral analysis of soil properties