Catena 151 (2017) 118–129

Contents lists available at ScienceDirect

Catena

journal homepage: www.elsevier.com/locate/catena

Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques

Said Nawar [a,*], Abdul M. Mouazen [a,b]

[a] Cranfield Soil and AgriFood Institute, School of Water, Energy and Environment, Cranfield University, Cranfield MK43 0AL, UK
[b] Department of Soil Management, Ghent University, Coupure 653, 9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 8 July 2016
Received in revised form 4 December 2016
Accepted 16 December 2016
Available online 22 December 2016

Keywords:
Diffuse reflectance spectroscopy
Spectral library
Soil properties
Data mining
Spiking

ABSTRACT

The development of accurate visible and near infrared (vis-NIR) spectroscopy calibration models for selected soil properties based on mobile measurements is essential for site specific soil management at fine sampling scale. The objective of the present study was to compare the mobile and laboratory prediction performance of vis-NIR spectroscopy for total nitrogen (TN), total carbon (TC) and soil moisture content (MC) of field soil samples based on single field (SFD), two-field dataset (TFD), UK national dataset (UND) and European continental dataset (ECD) calibration models developed with linear and nonlinear data mining techniques including spiking. Fresh soil samples collected from fields in the UK, Czech Republic, Germany, Denmark and the Netherlands were scanned with a fibre-type vis-NIR spectrophotometer (tec5 Technology for Spectroscopy, Germany), with a spectral range of 305–2200 nm. After dividing spectra into calibration (75%) and validation (25%) sets, spectra in the calibration set were subjected to three multivariate calibration models, including the partial least squares regression (PLSR), multivariate adaptive regression splines (MARS) and support vector machines (SVM), with leave-one-out cross-validation to establish calibration models of TN, TC and MC. Results showed that the best model performance in cross-validation was obtained with MARS methods for the majority of dataset scales used, whereas the lowest model performance was obtained with the SFD. The effect of spiking was significant and the best model performance in general term was obtained when local samples collected from two target fields in the UK were spiked with the ECD, with coefficients of determination (R^2) values of 0.96, 0.98 and 0.93, root mean square error (RMSE) of 0.01, 0.1 and 1.75, and ratio of performance to interquartile distance (RPIQ) of 7.46, 6.57 and 3.98, for TC, TN and MC, respectively. Therefore, these results suggest that ECD vis-NIR MARS calibration models can be successfully used to predict TN, TC and MC under both laboratory and mobile measurement conditions.

1. Introduction

Visible and near-infrared (vis-NIR) diffuse reflectance spectroscopy has attracted increasing interest among soil scientists in recent times, and has been proposed as a possible method of soil analysis. It provides higher soil sampling density for mapping purposes compared with conventional laboratory analysis (Shepherd and Walsh, 2002; Wetterlind et al., 2010). This technique also allows for in field (in situ) non mobile (Viscarra Rossel and Chen, 2011; Brodský et al., 2013) and mobile measurement with high soil-sampling resolution (Maleki et al., 2008; Kuang and Mouazen, 2013). Literature show that Vis-NIR spectroscopy has been used successfully for modelling and mapping of soil properties, under both mobile and non-mobile measurement conditions, (i.e., Shibusawa et al., 2001; Mouazen et al., 2005; Kuang and Mouazen, 2013; Kuang et al., 2015). However, compared to laboratory spectral measurements that is done under controlled conditions, mobile spectroscopy analyses are affected by environmental factors such as ambient light, soil moisture content, soil structure, temperature, contamination by stones and excessive residues (Mouazen et al., 2007; Stenberg et al., 2010). One way to overcome these negative influences is by adopting effective spectral data preprocessing and advanced data mining techniques when developing calibration models for mobile spectroscopy.

Precision agriculture aims at optimizing management of within field variability for sustainable increase in land productivity (Bongiovanni and Lowenberg-Deboer, 2004). Variable-rate fertilizer application, which requires reliable soil information at a high spatial resolution, is required to achieve this goal (Wetterlind et al., 2010; Mouazen and Kuang, 2016). Successful implementation of the mobile vis-NIR

---

* Corresponding author.
E-mail address: s.m.nawar@cranfield.ac.uk (S. Nawar).

http://dx.doi.org/10.1016/j.catena.2016.12.014

spectroscopy was reported for sensor based and map-based variable rate phosphorus fertilization (Maleki et al., 2008; Mouazen and Kuang, 2016) and variable rate N-fertilizer application (Halcro et al., 2013). However, accurate recommendation maps for variable rate applications will depend on accurate prediction of soil properties with vis-NIR spectroscopy, which can be achieved by minimizing error in reference and spectra measurements, minimizing influences of ambient conditions and finally by adopting effective spectra pre-processing and advanced data mining techniques.

In order to gain the full advantage of the use of vis-NIR, different techniques have been applied to the development of calibration models, such as spiking of samples from target site or dataset into existing spectral library (Brown, 2007; Sankey et al., 2008; Wetterlind et al., 2010; Kuang and Mouazen, 2011, 2013; Guerrero et al., 2014). The geographical scale of soil samples collected was reported to have influence on model performance (Sudduth and Hummel, 1996). This will be especially evident when predicting variations on a small scale (Brown, 2007). Combining global and local samples by adding a few local ones to a more general soil library (spiking) and recalibrating was proposed by Brown (2007) as another way to increase the accuracy of soil organic carbon (SOC) prediction, as opposed to local-sample calibrations. Sankey et al. (2008) also reported improved prediction results for clay content, SOC and inorganic carbon, using the same global calibration set spiked with local samples from three highly variable landscape study sites in Montana, US, compared with global or local calibrations alone. Kuang and Mouazen (2013) used spiking technique with different dataset ratios for mobile vis-NIR modelling at European field scale. They achieved good results using partial least square regression (PLSR) models for soil moisture content (MC), SOC, and total nitrogen (TN) with residual prediction deviation (RPD) (calculated as standard deviation of measured soil properties divided by root mean square error of prediction) of 2.76 to 3.96, 1.88 to 2.38, and 1.96 to 2.52, respectively. Limited works have been reported on the combined effect of spiking and samples scale on model predictive performance in soil analysis (Sankey et al., 2008; Guerrero et al., 2010, 2014; Wetterlind and Stenberg, 2010), particularly for mobile collected vis-NIR spectral data, where no reports could be found in the literature.

As a linear multivariate analysis, PLSR is the most commonly used technique for soil spectral analysis (Conforti et al., 2013, 2015). However, the accuracy of linear-regression techniques in spectroscopic analysis tends to decrease due to the non-linear nature of the relationship between spectral data and the dependent variable (Araújo et al., 2014). Data-mining techniques, such as artificial neural network (ANN) (Mouazen et al., 2010; Kuang et al., 2015), multivariate adaptive regression splines (MARS) (Bilgili et al., 2010; Nawar et al., 2015) and support-vector machines (SVM) (Morellos et al., 2016; Nawar et al., 2016), were reported to improve the accuracy of the calibration models. As a nonlinear method based on the machine learning theory, SVM was proposed by Vapnik (1998) to be capable of modelling linear and nonlinear relationships and solving calibration problems with high performance (Suykens and Vandewalle, 1999). SVM has been used successfully for modelling soil properties based on reflectance spectroscopy (e.g. Viscarra Rossel and Behrens, 2010; Vohland et al., 2011; Peng et al., 2014; Nawar et al., 2016), and has gained extensive application in soil spectroscopy, because of its advantages and high performance (Viscarra Rossel and Behrens, 2010; Vohland et al., 2011). As nonparametric method MARS estimates complex nonlinear relationships among independent and dependent variables (Friedman, 1991), and has been effectively applied in different fields (Luoto and Hjort, 2005; Bilgili et al., 2010; Felicísimo et al., 2012; Samui, 2012) and generally exhibits higher performance results for modelling soil properties (e.g. Shepherd and Walsh, 2002; Bilgili et al., 2010; Nawar et al., 2016). However, these linear and non-linear modelling methods were not compared in the literature for soil analyses at different geographical scales including spiking. This is particularly true for modelling of mobile collected vis-NIR soil spectra.

The aim of this paper was to compare the predictive performance of the vis-NIR spectroscopy of TN, TC and MC for field samples based on single field, two-field, national and continental scales, using PLSR and two multivariate data-mining techniques, namely, SVM and MARS. The calibrations were made using laboratory and mobile collected soil spectra for predicting within-field variation in named soil properties.

2. Materials and methods

2.1. Experimental sites

The experimental wok was carried out in two fields in Yorkshire, UK. Hagg field is located at Cawood, north Yorkshire with longitudes of −1.172° and −1.166° W, and latitudes of 53.936° and 53.941° N with total area of about 21 ha (Fig. 1). This field is characterised by fosters cooler summer conditions with regular rainfall (average annual 600 mm), and the mean air temperatures range from 1 °C to 6 °C in winter and from 8 °C to 18 °C in summer, allowing more consistent crop growth. The field is cultivated with vegetables crops (leeks, cabbage, carrots and onions). The soil type is a slightly acidic sandy loam with impeded drainage to the south and eastern margins of the field. The soil organic matter (SOM) ranges between 0.7 and 1.95% at 0–30 cm depth. Hessleskew field is located in Sancton, north Yorkshire between longitudes −0.590° and −0.586° W, and latitudes 53.844° and 53.844° N and with total area of about 12 ha (Fig. 1). Elevations in the area barely reach 212 m above sea level. The annual rainfall and air temperature are in the same range of the Hagg field. The field is characterised by freely draining with texture varies between clay loam to clay. The SOM ranges between 0.9 and 2.1%, and it is cultivated with cereal crops in rotation.

2.2. Mobile soil measurement and collection of soil samples

The mobile measurement system designed and developed by Mouazen (2006) was used to measure both fields. It consists of a subsoiler, which penetrates the soil to the required depth, making a trench, whose bottom is smoothened by the downwards forces acting on the subsoiler (Mouazen et al., 2005). The subsoiler was retrofitted with the optical unit and attached to a frame. This was mounted onto the three point linkage of the tractor (Mouazen et al., 2005). An AgroSpec mobile, fibre type, vis–NIR spectrophotometer (Tec5 Technology for Spectroscopy, Germany) with a measurement range of 305–2200 nm was used to measure soil spectra in diffuse reflectance mode. The sampling interval of the instrument was 1 nm. A deferential global positioning system (DGPS) (EZ-Guide 250, Trimble, USA) was used to record the position of mobile measured spectra with sub-meter accuracy. Each scan consisted of three spectra that were averaged in one spectrum representing about 1.2 m travel distance. Outlier spectra were visually detected and removed from further analyses. These outlier spectra represented 2.5 and 4% out of 12,766 and 14,276 soil spectra collected from Hessleskew and Hagg fields, respectively. During the measurement at each line, 3 or 4 soil samples were collected from the bottom of the trench and the sampling positions were carefully recorded with a DGPS. The locations of soil samples were selected to cover the soil variation within both fields. A total of 122 and 149 soil samples were collected during the mobile measurement form Hessleskew and Hagg fields, respectively (Fig. 1). Measurements were carried out in 2015 and 2016 for the former and latter fields, respectively, pulling the sensor at 12 m gap between adjacent transects (Fig. 1).

2.3. Laboratory chemical and optical measurements

Each sample was divided into two parts; one part was dried for 24 h at 105 °C and the other part was left fresh (wet). The dried soil sample was analyses for total carbon (TC) according to the British Standard BS 7755 Section 3.8 (1995) using combustion method, which is identical to ISO 10694:1995. Total nitrogen was determined by the Dumas method, where the soil samples are heated to 900 °C in the presence of