Geoderma 330 (2018) 244–253

Contents lists available at ScienceDirect

Geoderma

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

Importance of the spatial extent for using soil properties estimated by laboratory VNIR/SWIR spectroscopy: Examples of the clay and calcium carbonate content

C. Gomez*, G. Coulouma

LISAH, Univ Montpellier, IRD, INRA, SupAgro, F-34060 Montpellier, France

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:
Laboratory VNIR/SWIR spectroscopy
Clay
Calcium carbonate
Soil
Partial least square regression
Spatial extent

ABSTRACT

Visible, near-infrared and short-wave infrared (VNIR/SWIR, 400–2500 nm) laboratory soil spectrometry is now considered to provide accurate estimations of primary soil properties (clay, calcium carbonate, iron, soil organic carbon, etc.). The performances of primary soil property prediction models are evaluated in regard to figures of merit calculated over calibration and validation databases but not in regard to the spatial extent of predicted soil samples. The objective of this study was to analyze regional model performances for soil property prediction at regional and within-field extents within contrasted representative geopedological situations. This study used a database of 240 soil samples collected over eight vineyard fields located in the Languedoc Region (southern France) (between 20 and 36 soil samples per field) for which VNIR/SWIR laboratory spectra were acquired and two soil physico-chemical properties (clay and calcium carbonate) were measured. Soil property prediction models were built using the classical partial least square regression (PLSR) method, which links the VNIR/SWIR laboratory spectra and the physico-chemical soil property. Our results showed that both clay and calcium carbonate prediction models are accurate at the regional extent, whereas prediction model performances at the within-field extent depend on the model robustness. Therefore, primary soil properties predicted by VNIR/SWIR laboratory spectra must be used with care at different extents.

1. Introduction

Soil has an integral role to play in the global environmental sustainability challenges of food security, water security, energy sustainability, climate stability, biodiversity, and ecosystem service delivery (Herrick, 2000; McBratney et al., 2014). So adequate decisions must be made both at regional and local levels for preserving soil functions to preserve ecosystem services and goods. Assisting these decisions requires precise spatially referenced soil information systems at several spatial extents and resolutions.

For two decades, visible, near-infrared and short-wave infrared (VNIR/SWIR, 400–2500 nm) laboratory soil spectrometry has been proven to be a good alternative to physical and chemical laboratory soil analyses (e.g., Viscarra Rossel et al., 2006; Stenberg et al., 2010; Gholizadeh et al., 2013). This technique is used to estimate several soil properties, such as the contents of soil organic carbon (SOC) (e.g., Chang and Laird, 2002; Islam et al., 2003), iron (e.g., Islam et al., 2003; Cozzolino and Moron, 2003), clay (e.g., Islam et al., 2003; Shepherd and Walsh, 2002; Lagacherie et al., 2008) and calcium carbonate ($CaCO_3$) (e.g., Ben-Dor and Banin, 1990, 1994; Lagacherie et al., 2008). For the past decade, on-line VNIR/SWIR spectral measurements sensors have been in development for mapping soil properties at the within-field extent and with spatial resolutions between 1 and 30 m (e.g., Mouazen et al., 2005; Christy, 2008; Knadel et al., 2011). These developments are encouraging for predicting soil properties, such as soil moisture content (e.g., Mouazen et al., 2005) and SOC (e.g., Knadel et al., 2011; Nocita et al., 2011; Rodionov et al., 2015). In parallel and for also one decade, VNIR/SWIR hyperspectral imaging data (acquired from airborne or satellite platforms) have been successfully used for mapping soil properties over spatial extents up to 400 km^2 and with spatial resolutions between 4 m (using airborne VNIR/SWIR hyperspectral imaging sensors) and 30 m (using the HYPERION sensor on the EO-1 satellite, Folkman et al., 2001). Successful studies were performed using VNIR/SWIR hyperspectral imaging data for predicting soil properties, such as SOC (e.g., Stevens et al., 2010; Lu et al., 2013), iron (e.g., Bartholomeus et al., 2007; Gomez et al., 2008), clay (e.g., Selige et al., 2006; Gomez et al., 2008) and $CaCO_3$ (e.g., Lagacherie et al., 2008; Gomez et al., 2008).

Several methodologies have been successfully tested to link the VNIR/SWIR soil spectra (laboratory, on-line and imaging spectra) to the

* Corresponding author.
E-mail address: cecile.gomez@ird.fr (C. Gomez).

https://doi.org/10.1016/j.geoderma.2018.06.006
Received 6 March 2018; Received in revised form 7 June 2018; Accepted 10 June 2018
0016-7061/

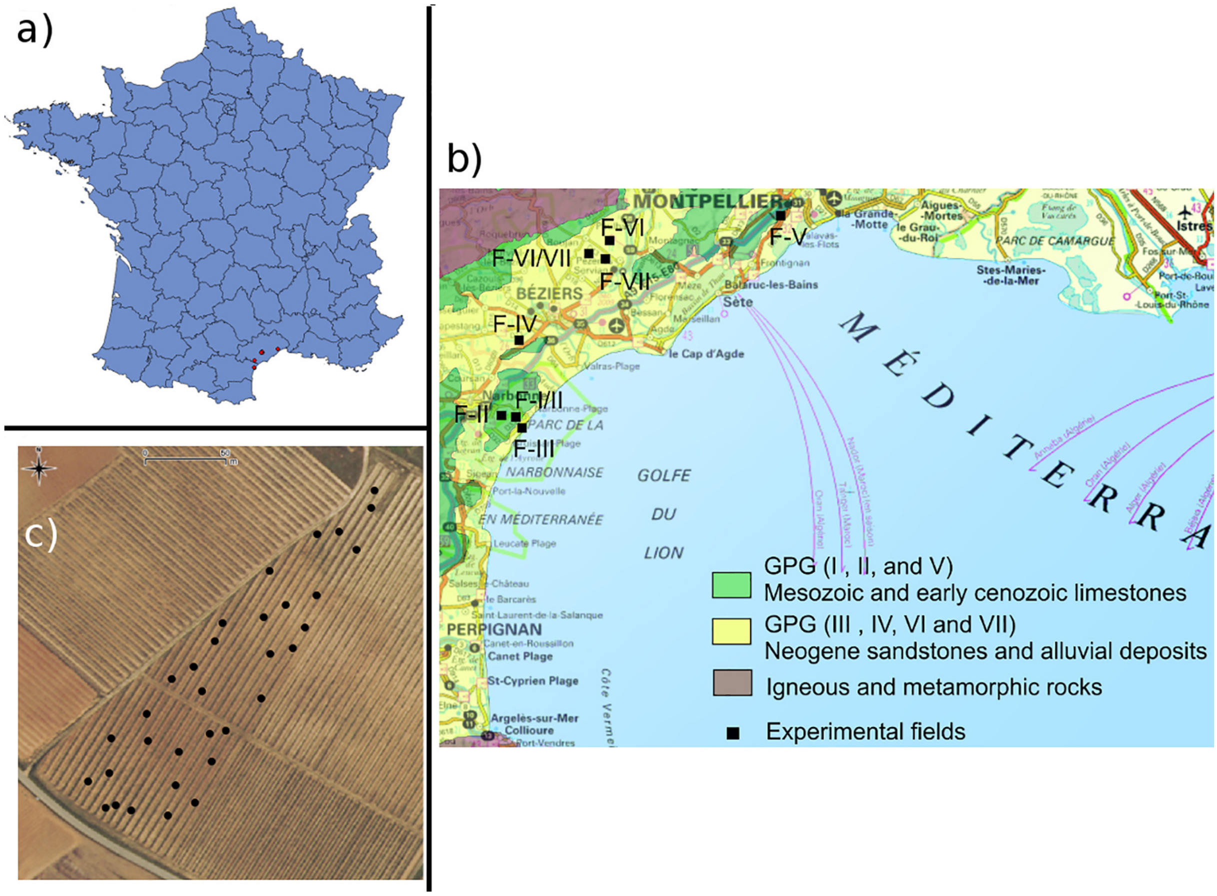

Fig. 1. The location of the eight vineyard plots a) in Southern France (red points), b) over the Languedoc Region (black squares) and c) location of the 33 soil samples collected over the F-VI/VII field (black points plotted over an aerial orthoimage). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

soil property, including partial least square regression (PLSR; e.g., Chang and Laird, 2002; Cozzolino and Moron, 2003), spectral indices (e.g., Lagacherie et al., 2008; Gomez et al., 2008; Bartholomeus et al., 2008), principal component regression (PCR; e.g., Chang et al., 2001; Islam et al., 2003) and stepwise multiple linear regression (SMLR; e.g., Shibusawa et al., 2001; Bartholomeus et al., 2012; Ben-Dor and Banin, 1995). Regardless of the regression method, the soil property, the number of samples in the calibration database and the spatial extent of collected soil samples, the soil property prediction models performances and efficiencies (i.e., estimations quality) are expressed using numerical expressions, also called figures of merit, such as the standard error of calibration (SEC), standard error of prediction (SEP), coefficient of determination (R^2) or the ratio of performance deviation (RPD) (Bellon-Maurel et al., 2010). In the case of VNIR/SWIR hyperspectral imaging data, variograms of soil property estimations are occasionally studied to estimate the accuracy of prediction models to reproduce the spatial structures of the properties in the study area (e.g., Gomez et al., 2012, 2018; Steinberg et al., 2016).

As far as we know, the performances of soil property prediction models have not been analyzed regarding different spatial extents (where a spatial extent can be defined as the smallest shape which totally encloses all of the referenced data, e.g., a field, a farm, a catchment, a region, etc.). This lack in the performance analysis process may be due to the limited soil sampling (used in the calibration and validation step) due to the cost of sampling in terms of money and time. Therefore, the soil sampling rarely allows the consideration of several spatial extents. For example, Kuang and Mouazen (2011) built individual farm models (samples collected over each farm are used to build an individual model) and a general model (samples collected over the three farms are used together) for several soil property predictions, including organic carbon, total carbon and nitrogen. The performances of the general model were not analyzed regarding to each within-field extent but regarding the validation dataset, which included samples collected over the three fields. As another example, Peng et al. (2014) built a general model of SOC prediction using 298 soil samples collected over three study areas from at least 200 km from each other. They studied the performances of this general model in regard to the figures of merit, independent of the spatial extent and location of the samples.

The purpose of the present work is to analyze the use of regional models for soil property prediction at regional and within-field extents. This study used a database of 240 soil samples collected over eight vineyard fields for which VNIR/SWIR laboratory spectra were acquired and two soil physico-chemical properties (clay and $CaCO_3$) were measured. The soil properties prediction models were built using the PLSR method, as PLSR seems to be the most appropriate approach to demonstrate that performances of regional models have to be studied carefully at several spatial extents because of two reasons. First PLSR is the most common regression model in soil spectroscopy as it is a powerful method, very well documented in computing software and statistic books, and with large references in literature to compare PLSR results. And secondly as PLSR is considered as powerful for many researchers, PLSR could become a very classical method for which prediction model performances would be studied superficially due to an excessive confidence on it. The eight vineyard fields were chosen to represent a large range of geopedological situations according to the World Reference Base (IUSS Working Group WRB, 2014) which are among the most significantly observed in the Mediterranean area (The Soil Map of Europe, 2014).