Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Special Issue: Proximal Soil Sensing

Research Paper

How well can VNIR spectroscopy distinguish soil classes?

Rong Zeng [a,b,1], Gan-Lin Zhang [a,b,*], De-Cheng Li [a,2], David G. Rossiter [a,c,3], Yu-Guo Zhao [a,4]

[a] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, 210008, PR China

[b] University of the Chinese Academy of Sciences, Beijing, 100049, PR China

[c] Section of Soil and Crop Sciences, Cornell University, Ithaca, NY, 14853, USA

ARTICLE INFO

Article history:
Published online xxx

Keywords:
Reflectance spectroscopy
Soil classification
VNIR spectra
Taxonomic distance

Visible near-infrared (VNIR) spectra can provide rich information on soil physical and chemical properties, which implies the possibility of using soil spectra to aid in the discrimination of soil types. Pedological soil classification systems use a selected set of soil properties to identify diagnostic horizons and features, and to build a classification key. This research explored the application of VNIR spectra to classify typical soil profiles collected in Anhui province, China. The 279 soil profiles used are classified into five orders (Cambosols, Vertosols, Argosols, Primosols and Anthrosols), six suborders and 21 groups according to Chinese Soil Taxonomy. Soil spectra were collected within 350–2500 nm and principal component analysis (PCA) was applied to reduce data dimension. These principal components were used as independent variables in multinomial logistic regression for soil classification. Topsoil spectra, subsoil spectra and their combination were compared for prediction accuracy. Accuracy achieved at the level of suborder using spectra of topsoil, subsoil and combined horizons were 76.3%, 71.3% and 70.3% respectively, while the results for the level of soil group using the topsoil horizon was 40.5%. Since topsoil spectra alone achieved a prediction accuracy of more than 75%, reflectance spectroscopy can be judged a promising tool for soil classification. Taxonomic distances between classes calculated on the basis of physio-chemical properties and spectra were quite different, showing that the concept of distance between classes in feature space depends on the features chosen for evaluation. Taxonomic distances can serve as a supplement for better selection and evaluation of prediction models.

* *Corresponding author*. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, 210008, PR China. Tel.: +1 86 25 8688 1279; fax: +1 86 25 8688 1000.

E-mail addresses: rzeng@issas.ac.cn (R. Zeng), glzhang@issas.ac.cn (G.-L. Zhang), dcli@issas.ac.cn (D.-C. Li), dgr2@cornell.edu (D.G. Rossiter), ygzhao@issas.ac.cn (Y.-G. Zhao).

[1] Tel.: +86 25 8688 1300.

[2] Tel.: +86 25 8688 1358.

[3] Tel.: +1 607 254 5231.

[4] Tel.: +86 25 8688 1299.

http://dx.doi.org/10.1016/j.biosystemseng.2016.04.019

1. Introduction

Soil classification groups individual soil profiles (pedons) into clusters with a limited range of properties, functions, and interpretations. Despite the advances in digital soil mapping of individual properties which can be used in land surface modelling, soil classes remain holistic indicators of soil "personality": the whole is more than the sum of the parts. Hence there is continued interest in soil classification systems and their extensive use in products such as SOTER, the Harmonized World Soil Database, SoilGrids and soil series inventory projects in China (Hengl et al., 2014; FAO/IIASA/ISRIC/ISS-CAS/JRC, 2008; Oldeman & van Engelen, 1993; Zhang et al., 2013).

To classify a soil profile in modern systems such as Soil Taxonomy (ST) (Soil Survey Staff, 1999), World Reference Base for Soil Resources (WRB) (IUSS Working Group, 2006) and Chinese Soil Taxonomy (CST) (Chinese Soil Taxonomy Research Group, 2001) is a complicated and time-consuming process, which requires a detailed field description by an experienced pedologist, laboratory analysis of soil properties, and a detailed knowledge of diagnostic keys. Much research has confirmed the use of visible and near infrared (VNIR) spectra to assist in laboratory analysis by relating spectra to soil physio-chemical properties required to navigate the soil classification keys, such as pH, organic matter, iron contents, bulk density, cation exchange capacity and calcium carbonates (Brown, Shepherd, Walsh, Dewayne Mays, & Reinsch, 2006; Viscarra Rossel & Webster, 2012).

All current soil classification systems are semi-quantitative, in which the discrimination of soil classes are partly based on quantitative values, i.e., soil physical, chemical and morphological properties. The future development of soil classification systems is likely to convert from semi-quantitative to fully quantitative, since the idea of numerical classification have been proposed by soil scientists in the Universal Soil Classification project (Beaudette, Roudier, & O'Geen, 2013; Michéli, Láng, Owens, McBratney, & Hempel, 2016; Sarkar, Bidwell, & Marcus, 1966). Data from VNIR spectroscopy are a likely candidate for numerical classification. They have been shown to contain comprehensive information on soil properties, can be measured easily and rapidly, and don't require much sample pre-processing. Thus, spectra can provide a different perspective to understand soil genesis and soil classification as a supplement to the traditional viewpoint of soil properties. Thus, we propose the use of reflectance spectroscopy to directly predict soil classification.

Earlier research relating soil spectra to soil types dates back to Stoner and Baumgardner (1981), in which 485 surface soil samples were classified into five representative types according to their spectral curves. This research was mainly based on the visual interpretation of spectral patterns. Recently, Vasques, Dematte, Rossel, Ramirez-Lopez, and Terra (2014) combined spectra from three depth intervals (0–20, 40–60 and 80–100 cm) to predict soil classification, with an agreement rate from 24% to 67% at different classification levels.

However, not all misclassifications are equally serious; some classes are quite similar in properties and use, while other classes are radically different. So to evaluate the practical, as opposed to purely statistical, success of any classification, some method should be used to determine the importance of each misclassification. One method used already in soil science is the so-called soil taxonomic distance (Minasny & McBratney, 2007), as it is a measure of the similarity between soil classes in taxonomic space.

In this study, we explored the possibility of using reflectance spectroscopy to discriminate soil types among a limited set of soil types at the level of suborder and group, according to CST (Chinese Soil Taxonomy Research Group, 2001). We tried to answer the following research questions: (1) How well can soil spectroscopy distinguish soil types, using topsoil, subsoil or combined spectra respectively? (2) What is the difference between taxonomic distances calculated based on soil spectra and soil physio-chemical properties?

2. Methods and materials

2.1. Research area and datasets

A total of 279 soil profiles (1280 genetic horizons) were selected from a spectral library of Anhui province, east China. A detailed description of the spectral library, the sampling strategy, and the geographic context can be found in Zeng et al. (2016). The number of horizons per profile varied from two to six. All horizons were sampled up to a maximum depth of 120 cm. The 279 soil profiles were classified by professional soil classifiers into five orders, six suborders and 21 groups according to CST. Since the number of orders and suborders only differs by one, we worked only at the suborder and group levels.

The location of sampled soil profiles and their classification are presented in Fig. 1 and Table 1. Reflectance spectra were measured in the VNIR range of 350–2500 nm using a Cary 5000 spectrometer (Agilent Technologies, CA) under controlled laboratory conditions. Before spectra collection, soil samples were air-dried and ground to pass 100 mesh sieves and then oven-dried at 45 °C for 24 h. Detailed information about the Cary 5000 spectrometer and protocols for spectral measurements can be found in Zeng et al. (2016). The spectra were smoothed by Savitzky-Golay transformation (Savitzky & Golay, 1964) using a third-order polynomial across five bands and then transformed to absorbance. To reduce data redundancy, the absorbance data were reduced to 216 bands by averaging across 10 bands. Then the reduced 216 bands were transformed to 20 principal components (PC) by principal component analysis (Jolliffe, 2002) which were used in the subsequent modelling.

To facilitate comparison among profiles with different numbers of subsoil horizons, the following strategy was used to select one representative subsoil horizon for each profile:

(1) Except for Primosols which do not have a B horizon, a B horizon must be selected, because B horizons are, by definition, the product of pedogenesis.
(2) The Spectral Angle Mapper (SAM) method (Kruse et al., 1993) was used to calculate the spectral angles between the topsoil horizon and all B horizons within

Please cite this article in press as: Zeng, R., et al., How well can VNIR spectroscopy distinguish soil classes?, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.04.019