Geoderma 325 (2018) 90–101

Contents lists available at ScienceDirect

Geoderma

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

Identification of soil profile classes using depth-weighted visible–near-infrared spectral reflectance

Xian-Li Xie[a,*], An-Bo Li[b,c]

[a] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, 71 East Beijing Road, 210008 Nanjing, China
[b] Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing Normal University, 1 Wenyuan Road, 210023 Nanjing, China
[c] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing Normal University, 1 Wenyuan Road, 210023 Nanjing, China

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:
Visible–near-infrared diffuse reflectance spectroscopy
Chinese Soil Taxonomy
Depth-weighted
Random forest
Synthetic minority over-sampling technique

ABSTRACT

Soil visible–near-infrared (Vis-NIR) diffuse reflectance, which can be collected rapidly and cheaply, contains rich information of the properties that are useful for soil taxonomic diagnosis. Vis-NIR spectroscopy has the potential to identify soil taxonomic classes in an efficient and cost-effective way. This study explores the potential of Vis-NIR spectroscopy to identify soil profile classes at the order, suborder, group, and subgroup levels of the Chinese Soil Taxonomy (CST).

The authors measured the Vis-NIR (350–2500 nm) diffuse reflectance of 2260 legacy soil samples, which were sampled by genetic horizon and collected from 527 soil profiles (75% for training and 25% for validation). To represent the overall spectral pattern for each soil profile, the depth-weighted average was used to combine the spectral reflectance of genetic horizons. The synthetic minority over-sampling technique (SMOTE) was used to obtain balanced training data. Principal component analysis (PCA) was performed to extract spectral predictors used for random forest (RF) modeling in soil classification.

Vis–NIR spectral reflectance exhibited acceptable overall performance for identification of soil orders and suborders, with overall validation accuracies of 0.63 and 0.62, respectively, but low overall performance at group and subgroup levels, with overall validation accuracies of 0.40 and 0.28, respectively. The overall performance at different taxonomic levels was affected by the number of soil classes and the class distribution of the soil profiles. Soil classes with pedogenic processes associated closely with spectrally active soil properties or with characteristic profile patterns were most accurately identified, even if they were minority classes. The results show that the Vis–NIR spectral pattern of soil profile can be used to identify soil profile classes at higher taxonomic levels in the CST system. Combined with machine-learning techniques, the soil Vis–NIR spectral library will serve as an efficient tool for digital soil survey mapping and updating with the use of legacy soil samples and the reduction of conventional laboratory analyses.

1. Introduction

Soil is a critical component of the Earth's ecosystem and plays important roles in agricultural food production, global climate change, hydrologic cycle, biological diversity, and environmental protection. In recent decades, due to the decline in the quantity and quality of soil (FAO and ITPS, 2015), the need to monitor and assess soil on different temporal and spatial scales has increased. Soil visible reflectance, or color, is one of the most useful attributes for delineating differences among soils and describing the characteristics of a soil profile (Baumgardner et al., 1986). Notably, the remote sensing reflectance of soils, measured from aerial or space platforms, is difficult to assess soil properties directly due to the complexity of environmental factors (McBratney et al., 2003), especially in hilly regions covered with dense vegetation. Under this background, many researchers have worked on rapid, cheap, and reliable soil analysis technologies as alternatives or complements to conventional laboratory soil analyses, which are laborious, time-consuming, and expensive.

Diffuse reflectance spectroscopy is an effective, convenient, and sensitive technique for the analysis of opaque powdered materials, such as soil samples. Visible and near-infrared (Vis-NIR) diffuse reflectance spectroscopy, often used in laboratories on dried and ground samples, has been widely explored to assess various soil properties, quantitatively and qualitatively, since the 1980s due to its time and cost saving

* Corresponding author.
E-mail address: xlxie@issas.ac.cn (X.-L. Xie).

https://doi.org/10.1016/j.geoderma.2018.03.029
Received 2 November 2017; Received in revised form 19 March 2018; Accepted 23 March 2018

potential (Krishnan et al., 1980; Dalal and Henry, 1986; Ben-Dor and Banin, 1995; Chang et al., 2001; Shepherd and Walsh, 2002; Brown et al., 2006; Terhoeven-Urselmans et al., 2008; Viscarra Rossel and Behrens, 2010; Viscarra Rossel et al., 2016). The spectrally active soil components in the Vis-NIR region are mainly iron oxide minerals, organic matter, clay minerals, carbonates, and water, as summarized by Baumgardner et al. (1986), Ben-Dor (2002) and Stenberg et al. (2010). The spectral absorption features in the 380–1000 nm region are mainly associated with electronic transitions of Fe^{3+} and Fe^{2+} ions, and the absorption features in the 1000–2500 nm region are associated with overtones and combinations of fundamental vibrations of organic and inorganic functional groups (e.g. C−H, N−H, C=O, O−H, metal-OH, CO_3^{2-}, and H_2O). The soil properties that have been reported in applications of Vis-NIR spectroscopy are comprehensive, covering the physical, chemical, and biological properties of soil. The capability of Vis-NIR spectroscopy to assess soil properties is related with the selective absorptions in the Vis-NIR region of spectrally active soil components; while for inactive soil components it is based on intercorrelation with other spectrally active components (Ben-Dor and Banin, 1995).

Most applications of Vis-NIR spectroscopy for soil analysis are focused on assessing individual soil properties, mainly for surface soils. In recent years, a few studies have explored the potential of using Vis-NIR spectroscopy in the integrative assessment of soils, such as evaluating soil tillage (Demattê et al., 2004), soil degradation (Awiti et al., 2008), soil quality (Askari et al., 2015; Paz-Kagan et al., 2015), soil horizon type (Viscarra Rossel and Webster, 2011), and soil profile class (Viscarra Rossel and Webster, 2011; Vasques et al., 2014; Ogen et al., 2017). Traditionally, soil profile class is identified based on the soil's morphological, physical, and chemical characteristics, according to the diagnostic criteria of a soil classification system. Soil Vis-NIR spectroscopy contains integrated information of the properties used for soil taxonomic diagnosis, including soil color, iron oxide minerals, organic matter, clay minerals, and carbonates. Hence, Vis–NIR spectroscopy can provide an efficient and cost-effective method to identify soil profile classes for digital soil survey mapping and updating by the use of legacy soil samples and the reduction of conventional laboratory analyses. Baumgardner et al. (1986) stated that “soil visible reflectance, or color, is a differentiating characteristic for many classes in all modern soil classification systems and is an essential part of the definitions for both surface and subsurface diagnostic horizons.” A limited number of studies have confirmed the ability of Vis-NIR diffuse reflectance for soil classification: Ben-Dor et al. (2008) developed an optical method for classifying soil profiles by using a soil spectral profiler to quantitatively describe soil profiles in situ; Viscarra Rossel and Webster (2011) successfully discriminated between the orders of the Australian Soil Classification system for 1697 soil profiles using Vis-NIR spectra, and they concluded that Vis–NIR spectroscopy could be used to define and identify classes in an effective system of soil classification; Vasques et al. (2014) classified 291 soil profiles, 202 in calibration and 89 in validation, in the Brazilian Soil Classification System using Vis-NIR spectroscopy, and obtained 67% of correct predictions for the validation set at the order level, 48% at the suborder level, and 24% at the suborder plus textural classification level; Ogen et al. (2017) demonstrated a hyper-depth spectral classification method that achieved satisfactory results for obtaining soil profile types in a local *catena*.

A critical issue when using Vis-NIR spectroscopy to identify soil profile classes is how to combine Vis-NIR spectral reflectance of soil horizons for individual profiles. In previous studies, Viscarra Rossel and Webster (2011) simplified genetic horizons by classifying them as topsoil and subsoil for each profile, averaged the Vis–NIR spectra for topsoil and subsoil, and used these spectra to determine soil profile class; Vasques et al. (2014) combined the Vis-NIR spectra of three depth layers (0–20, 40–60 and 80–100 cm) sequentially to create a pseudo multi-depth soil spectroscopy for soil profiles; Ogen et al. (2017) used hyper-depth spectral data, interpolated from 15 spectra for each soil profile, 10 from the first 1 m (10 cm each) and 5 from 1 to 2 m depth (20 cm each). Since distinct differences in soil properties generally occur between genetic soil horizons (Schoeneberger et al., 2012), soil sampling in most pedological studies is by genetic horizon, not by fixed-depth, and most legacy soil samples were collected by genetic horizon. In this study, the authors tested the depth-weighted spectral reflectance from genetic horizons to identify soil profile classes in the Chinese Soil Taxonomy (CST) system. The CST, developed in the mid-1980s, is a classification system based on quantitative diagnostic horizons and characteristics (CSTRGISSCAS and CRGCST, 2001). In recent years, the CST has been promoted in China, and soil taxonomists have been conducting soil series surveys across the country (Zhang et al., 2013).

Machine learning algorithms are useful for recognizing the spectral patterns of soil taxonomic classes and making predictions from the discovered patterns. Random forests (RF) (Breiman, 2001), a combination of tree predictors, is a relatively new algorithm and one of the most powerful tools for classification and regression purposes. RF preserves most of the appealing features of decision trees, and the prediction performance is among the best (Svetnik et al., 2003). RF has been applied in soil science, mainly for digital soil mapping (Grimm et al., 2008; Ließ et al., 2012; Heung et al., 2014; Hengl et al., 2015; Brungard et al., 2015), and is currently being used in soil spectroscopy as well (Gholizadeh et al., 2013). Viscarra Rossel and Behrens (2010) compared multiple data mining algorithms, including RF, to estimate soil properties by using soil diffuse reflectance spectra. Overall, RF has not been widely used for soil spectral classification. When using machine learning algorithms to predict soil classes, imbalanced class distribution in training data commonly occurs, which influences classification accuracy. Most learning algorithms are overwhelmed by majority classes and ignore minority classes (Chawla et al., 2002). To overcome this problem, the synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002) was used.

The objective of this study is to explore the feasibility of using Vis-NIR spectroscopy to identify soil profile classes at the order, suborder, group, and subgroup levels of the CST system. Soil profile spectral reflectance was determined as the depth-weighted average of the reflectance values of genetic horizons. RF was performed to recognize and identify the spectral patterns of soil profile classes. The dataset in this study showed highly imbalanced class distribution, and SMOTE was used to obtain balanced training data.

2. Materials and methods

2.1. Soil samples

A total of 2260 legacy soil samples from 527 profiles were used in this study. The soil samples were collected from a soil series survey (2009–2013) in six central-eastern provinces of China: Jiangsu, Anhui, Zhejiang, Fujian, Hubei, and Guangdong. In the survey, soil profiles were sampled by genetic horizons down to 1.2–2 m depth or bedrock. The profiles that were used in this study had complete archived samples of their genetic horizon sequence and were properly classified in the CST system. The profile locations in the survey areas are shown in Fig. 1.

These soil profiles belong to 6 orders, 10 suborders, 20 groups, and 35 subgroups in the CST. Table 1 lists the number of profiles in each class. According to CSTRGISSCAS and CRGCST (2001), Anthrosols in the CST correspond to Anthrosols in the World Reference Base (WRB) (ISSS/ISRIC/FAO, 1998); Halosols correspond to Aridisols, Alfisols, and Inceptisols in the American Soil Taxonomy (ST) (Soil Survey Staff, NRCS, USDA, 1999) and Solonchaks and Solonetz in the WRB; Ferrosols mostly correspond to Ultisols in the ST and Acrisols in the WRB; Argosols mostly correspond to Alfisols in the ST and Luvisols in the WRB; Cambosols mostly correspond to Inceptisols in the ST and Cambisols in the WRB; and Primosols correspond mostly to Entisols in the ST and Fluvisols, Leptisols, Arenosols, Regosols, and Cryosols in the WRB.