Geoderma 320 (2018) 12–22

Contents lists available at ScienceDirect

Geoderma

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

Allocate soil individuals to soil classes with topsoil spectral characteristics and decision trees

Xiaokang Zhang[a], Huanjun Liu[a,b,*], Xinle Zhang[a], Shengnan Yu[a], Xin Dou[a], Yahui Xie[a], Nan Wang[a]

[a] College of Resources and Environment Sciences, Northeast Agricultural University, Harbin 150030, China
[b] Northeast Institute of Geography and Agroecology Chinese Academy of Sciences, Changchun 130102, China

ARTICLE INFO

Keywords:
Soil classification
Topsoil spectral characteristics
Decision trees
K-means clustering
Neural network
Support vector machine

ABSTRACT

Spectral reflectance of soil is a function of physical and chemical characteristics and its internal structure. Spectral reflectance provides a novel approach for soil allocation. This paper presents a nondestructive, rapid, and low-cost soil allocation method using topsoil spectral characteristics to allocate soil at the level of soil great group within Genetic Soil Classification of China. We measured the spectral reflectance in the visible and near-infrared regions (400–2500 nm) of 148 soil samples from 4 soil classes in the Songnen Plain of northeast China. We extracted the spectral characteristic parameters with clear physiochemical meanings for the topsoil samples, and compared these to the principle component, first spectral derivative and Continuum Removal of soil reflectance. Models were built using the K-means Clustering (K-mean), Multi-layer Perceptron Neural Network (MLPNN), Support Vector Machine (SVM), and Decision Tree (DT) methods. The DTs allocation model based on topsoil spectral characteristic parameters had the highest allocation accuracy. Only the allocation accuracy of Cambisols was < 85%, because the spectral curve of Cambisols topsoil was similar to its adjacent soil due to soil erosion. This new method could simplify digital soil mapping, because topsoil spectra are easier to obtain than multilayer soil spectral data.

1. Introduction

Topsoil is the upper and outermost layer of soil, usually the top 2 in. (5.1 cm) to 8 in. (20 cm) (Miller et al., 2011). In soil classification systems, topsoil is known as the “O Horizon or A Horizon” (Soil Survey Staff, 1993). It is well known that crops generally concentrate their roots and obtain most of their vital nutrients from this layer, and to a large extent, the physical and chemical properties of this layer determine the growth of crops (Bouwman and Wbm, 2000; Ishaq et al., 2001; Passioura, 2002). Therefore, rapid and accurate classification of topsoil is an important prerequisite for crop. Soil types have traditionally been determined by physicochemical properties, diagnostic horizons, and forming processes based on a given classification system, which is a laborious and time consuming process (Xie et al., 2015). It is difficult to apply traditional soil classification methods to the digital mapping of soil. Visible/near-infrared (Vis-NIR) spectroscopy is a rapid, non-destructive, reproducible, and cost-effective analytical method (Reeves et al., 2010; Conforti et al., 2014) to determine clay mineral content, soil organic matter, pH, iron, bulk density, cation exchange capacity, calcium carbonates, and micronutrients in soil (Adeline et al., 2017; Chang et al., 2001; Dalal and Henry, 1986; Demattê and Terra, 2014; Du and Zhou, 2009; Hummel et al., 2001; Stenberg et al., 2010; Vasques et al., 2008 and Viscarra Rossel et al., 2006). Besides, Vis-NIR spectroscopy is useful for defining and identifying soil classes (Viscarra Rossel and Webster, 2011).

Soil spectral characteristics were used to classify soil samples into several groups based on the relationship between the shape characteristics of the spectral curves and soil physical-chemical properties, and those groups were not related to pre-existing soil classes. For example, Stoner and Baumgardner (1981) classified soil spectral reflectance curves into 5 representative types using the spectral reflectance characteristics of 485 soil samples from the U. S. and Brazil. Dai (1981) classified the spectral reflectance curves of 100 soil samples in China into 4 types using the spectral characteristics and slope curves of the samples. Shi et al. (2014) used 1581 soil samples collected from 14 provinces in China to study soil spectral classification. The spectra were compressed using Principal Component Analysis (PCA), and the fuzzy K-means method was used to calculate the optimal soil spectral classification. The soil samples were classified into 5 clusters.

There are also some studies that allocated soil samples into soil

* Corresponding author at: College of Resources and Environment Sciences, Northeast Agricultural University, Harbin 150030, China.
E-mail address: liuhuanjun@neigae.ac.cn (H. Liu).

https://doi.org/10.1016/j.geoderma.2018.01.023
Received 14 March 2017; Received in revised form 8 September 2017; Accepted 17 January 2018

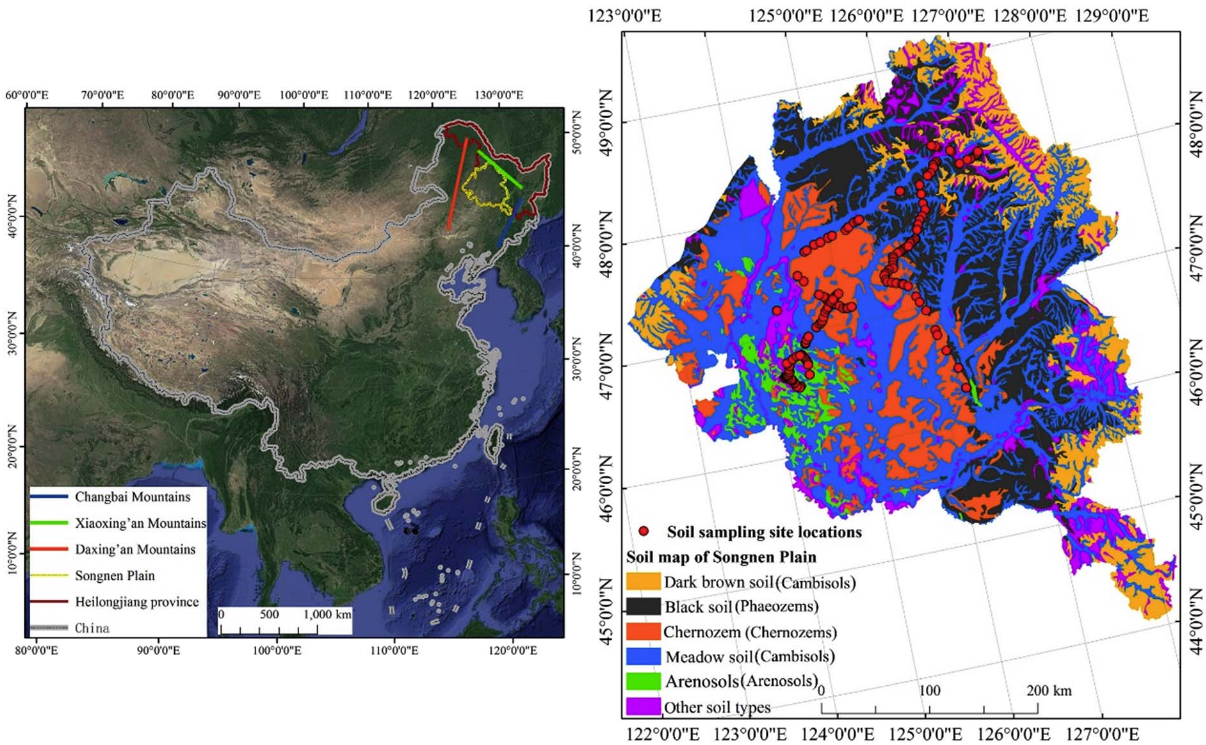

Fig. 1. Soil map of Songnen Plain and the distribution sampling points.

classes of the traditional classification system with soil spectral characteristics. Mouazen et al. (2007) used the visible and near-infrared (VNIR) spectroscopy as a new methodology on soil allocation. Liu et al. (2008) used the continuum removal of soil reflectance to extract allocation index and build the Artificial Neural Network model, and the allocation accuracy of the model was over 60%. Li et al. (2015) used the Genetic Algorithm to extract the spectral characteristics of soil samples from Guangdong province and built a soil allocation model based on the Support Vector Machine (SVM). The SVM has an accuracy of up to 59%. Some studies of soil spectral allocation have measured soil spectral reflectance of soil samples at different depths, just following the procedure of traditional soil classification (Wang et al., 1986; Vasques et al., 2014; Zeng et al., 2016). Vasques et al. (2014) collected Vis-NIR (400–2500 nm) spectra from 3 depth intervals (0–20, 40–60 and 80–100 cm) and combined them in sequence to generate a pseudo multi-depth spectral curve and derive a soil allocation model. The best allocation was obtained at the 67% order level coincidence rate. Zeng et al. (2016) explored the application of Vis-NIR spectra to allocate the typical soil profiles in the Anhui province. The 279 soil profiles were allocated into 5 orders, 6 suborders, and 21 groups using Chinese Soil Taxonomy (CST). The accuracy at the suborder level was 76.3% for topsoil, 71.3% for subsoil, and 70.3% for the combined horizons. Hence, using the topsoil spectra is a promising approach for soil allocation. Zeng et al. (2017) used ten soil properties predicted by spectra to build PLSR allocation model, the overall allocation accuracy at Order, Suborder, Group and Subgroup level was 98.5%, 98.5%, 87.7% and 76.0%, respectively.

Soil allocation based on soil spectral reflectance characteristics in the laboratory is the basis of soil allocation using satellite remote sensing. However, satellite remote sensing can only obtain the spectral information of topsoil. The spectral information of subsoil horizons cannot be obtained with remote sensing sensors, so it is difficult for remote sensing to be applied in soil allocation using soil reflectance spectra at different depths. It is more effective to use the spectral characteristics of the topsoil to study soil allocation.

Some studies have processed soil spectral data using PCA to build allocation models, including the SVM and Neural Network. There is, however, no obvious physical meaning of the variables obtained by the PCA, so the results of different studies cannot be compared.

There are many soil classification systems in the world such as Soil Taxonomy (ST), World Reference Base for Soil Resources (WRB) and Russia Soil Classification System (RSCS), which can be divided into the soil genetic classification and the soil taxonomy. However, there are two soil classification systems in China, namely, Genetic Soil Classification of China (GSCC) and CST. CSGC bases on soil forming conditions, soil forming processes and their attributes in the whole profiles, and includes order, suborder, great group, subgroup, family and series (Xi et al., 1998). In this paper, we used topsoil spectral characteristics to allocate soil great group based on the results of the Second National Soil Survey of China, which belongs to GSCC.

The objectives of this study are to verify whether soil allocation at the level of soil great group could be realized with topsoil spectral characteristics and whether the allocation results could be consistent with the traditional soil classification. Three steps were conducted to fulfill the objective: (1) construct spectral characteristic parameters with definite soil physiochemical meanings for topsoil samples, (2) compare the accuracy of the allocation models using PCA and spectral characteristic parameters as input, and (3) determine the optimal soil allocation model with topsoil spectral characteristics.

2. Material and methods

2.1. Study area

In this study, soils were allocated in the field according to Chinese Soil Classification System. The study area was the Songnen Plain in northeast China. It is located between the Daxing'an Mountains, Xiaoxing'an Mountains, and Changbai Mountains. The Songnen Plain has a flat terrain, fertile soil, and large farm areas. It is an important producer of commodity grain in China. This area has a temperate continental, semi-humid and semi-arid monsoon climate. It is cold and dry in the winter and warm and rainy in the summer. The main soil