Geoderma 303 (2017) 78–84

Contents lists available at ScienceDirect

Geoderma

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

How accurately can soil classes be allocated based on spectrally predicted physio-chemical properties?

Rong Zeng[a,b], David G. Rossiter[a,c], Fan Yang[a,b], De-Cheng Li[a], Yu-Guo Zhao[a], Gan-Lin Zhang[a,b,*]

[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

Keywords:
Soil classification
Reflectance spectroscopy
Uncertainty
PLSR modelling

ABSTRACT

Soil class maps are useful representations of the landscape distribution of holistic soil functions. However these are often only available as generalized classes at small cartographic scales. One reason is that allocating a soil profile to a class in most current soil classification system requires laboratory determination of many diagnostic soil properties. The advantage of reflectance spectroscopy along with the development of spectral libraries can provide a relatively low-cost solution to this problem. Reflectance spectroscopy has demonstrated its ability to rapidly predict soil physio-chemical properties; however prediction accuracy varies among soil properties. When properties predicted with different accuracies are used to substitute for traditional laboratory determinations in allocating a soil profile to a class, the resulting reliability of the allocation is questionable. The objective of this research is to explore whether the soil properties predicted by reflectance spectroscopy can be used to correctly allocate soil profiles into soil taxa at different hierarchical levels. Two hundred and six soil profiles were allocated to eight Orders, 12 Suborders, 23 Groups and 49 Subgroups according to Chinese Soil Taxonomy, with the help of ten soil properties predicted by spectra using ten-fold cross-validated PLSR modelling. The overall allocation accuracy at Order, Suborder, Group and Subgroup level was 98.5%, 98.5%, 87.7% and 76.0% respectively. These results show that soil reflectance spectroscopy can assist in allocation of profiles. When predicted soil properties with varying accuracy are used for soil allocation, propagation of prediction errors and model uncertainties must be considered. We propose the use of multiple indicators (RPD, confidence intervals, comparison of RMSE and threshold requirements) to evaluate the allocation results.

1. Introduction

As holistic indicators of soil "personality" and the result of definite pathways of pedogenesis, soil classes remain a focus of interest by agencies which are dealing with soil survey, soil databases and soil quality assessment (Grunwald, 2009). However, in many areas soil class maps are only available as generalized classes at small cartographic scales. For example, in China the most detailed nationwide soil map is at the scale of 1:1 million, as a generalization of the second national soil survey of China (1979–1994) (Shi et al., 2004). Soil maps at finer scales are needed to support different applications such as agricultural production, land evaluation, and land use planning. Despite the rapid development of soil sensors, most soil surveys are still following the procedures of purposive sample site selection, pit digging, field observation, soil sampling, laboratory analysis of soil properties and soil allocation by classifiers. Properly allocating a soil profile in all modern soil classification systems requires the laboratory determination of many diagnostic soil properties to navigate the soil classification keys. Where funds or relevant measurement techniques are constrained, only a few profiles can be sampled and often only some soil properties can be measured.

Thus, cost-effective methods are needed. The most promising has proven to be reflectance spectroscopy and the development of spectral libraries (Viscarra Rossel et al., 2016). Reflectance spectroscopy has demonstrated its ability to rapidly predict soil physio-chemical properties (Viscarra Rossel et al., 2006). However the prediction accuracy varies among soil properties (Soriano-Disla et al., 2014), due to the intrinsic spectral characteristics of the target property, the type of regression models, the calibration data size, and the specificity of the study area. Much research has been devoted to improving the prediction accuracy of the target properties (Gogé et al., 2014; Guerrero et al., 2014), but there is often an apparent limit to prediction accuracy. For

* Corresponding author at: State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, PR China.
E-mail addresses: rzeng@issas.ac.cn (R. Zeng), dgr2@cornell.edu (D.G. Rossiter), alexyang@issas.ac.cn (F. Yang), dcli@issas.ac.cn (D.-C. Li), ygzhao@issas.ac.cn (Y.-G. Zhao), glzhang@issas.ac.cn (G.-L. Zhang).

http://dx.doi.org/10.1016/j.geoderma.2017.05.011
Received 25 July 2016; Received in revised form 21 March 2017; Accepted 5 May 2017
Available online 13 May 2017

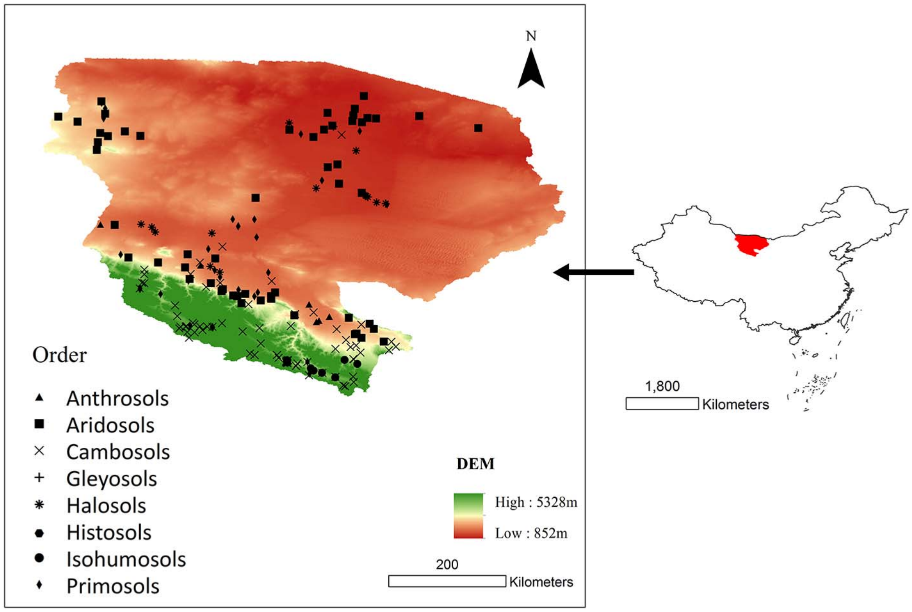

Fig. 1. Distribution of soil profiles.

example, decreased prediction accuracy was reported for very high soil organic carbon (SOC) contents (organic soils) due to spectral saturation, and the prediction performance of SOC was influenced by the presence of high level of iron contents (Stevens et al., 2013) due to spectral feature interference.

These variable prediction accuracies of soil properties then propagate through the allocation process. Since, for better or worse, monothetic soil classification systems use rigid numerical limits (Webster, 1968), this raises the question of how much imprecise property predictions affect allocation to soil classes. Soil classifiers or soil mappers are more concerned whether these predicted properties can assist in correct soil allocation, rather than the numerical prediction accuracy of different models. Thus, there is a gap between “prediction accuracy of soil properties by spectra” and “allocation accuracy to soil classes”. If reflectance spectroscopy can be proved to assist in soil allocation with sufficient accuracy, this should help improve the efficiency of soil classification as part of soil survey.

Research has been reported on using soil spectra and classifiers (e.g., logistic regression, neural networks) to directly allocate soil profiles to soil classes (Vasques et al., 2014; Zeng et al., 2016a) determined by traditional allocation of a training set of soil profiles. This direct method is straightforward and utilizes only spectral information. However, the disadvantage of such a machine learning method is that it is a “black box”, making it difficult to interpret the relationship between spectra and soil classes. In addition, there are likely serious misallocations of soil profiles due to the presence of “similar soil, different spectra” or “different soil, similar spectra”. An allocation requires many steps that do not depend on laboratory results, e.g., thickness of horizons and soil structure. So this direct method may be applicable in limited geographic areas with extensive training samples, but not in more general cases.

All modern soil classification systems are semi-quantitative (property and morphology based) monothetic diagnostic systems, in which soil profiles with properties falling within the same set of property ranges are allocated to the same class. Thus soil classifiers are more concerned about whether the properties can be predicted within the required range or threshold rather than the prediction precision as such. For example the threshold for SOC content is an average 6 g kg^{-1} weight for separating Mollic from Ochric epipedons in most soils in (USA) Soil Taxonomy (Soil Survey Staff, 1999), Chinese Soil Taxonomy (CST) (Cooperative Research Group on Chinese Soil Taxonomy, 2001), and the World Reference Base for Soil Resources (WRB) (IUSS Working Group, 2006). Therefore, if the confidence limits for SOC do not cross this limit, the classifier can safely allocate the profile to the appropriate diagnostic horizon, and from then to the appropriate soil class.

The objective of this research is to determine to what extent the soil properties predicted by reflectance spectroscopy can be used to correctly allocate soil profiles into soil taxa at several categorical levels (Order, Suborder, Group and Subgroup) according to CST as a representative of a typical modern semi-quantitative monothetic hierarchical classification system. Since reflectance spectroscopy can only partially assist in soil classification, this study is based on the following premise that: for soil properties or morphological features or ancillary environmental information which cannot be predicted by reflectance spectroscopy, e.g. soil structure, soil moisture regime, soil temperature regime etc., we take these as available information for allocation of soil profiles. The second objective of this study is to determine how to evaluate the reliability of the allocation results when the predicted soil properties with varied accuracy are used for soil allocation.

2. Material and methods

2.1. Research area and datasets

The soil profiles were selected from a spectral library of soil samples collected in the Heihe river basin in northwestern China (between 96°8′–104°11′ E and 37°42′–43°19′ N), which covers an area of 271,000 km^2 (Fig. 1). The basin is divided into three geomorphological units from south to north: the Qilian Mountains in the upper reaches, the middle Hexi Corridor, and the Alxa high plain in the lower reaches (Wang et al., 2006). The Qilian mountains range from 2000 m to 5500 m a.s.l., with diversified landscapes such as high mountain glacier and snow zone, permafrost zone, the mountain vegetation zone with typical steppe and desert steppe (Li et al., 2001). The dominant soil Orders in the Qilian Mountains are Cambosols, Isohumosols, Aridosols and Primosols according to CST. Most of the artificial oases in the basin are located in the middle reach due to a well-developed irrigation system. The major soil Orders here are Anthrosols and Cambosols. The lower reach is extremely arid with annual precipitation of < 50 mm, while the potential evaporation exceeds 3000 mm. The predominant soil Orders in this region are Halosols, Ariodosols and Primosols. In addition there are small areas of Histosols and Gleysols. The eight soil