



ELSEVIER

Contents lists available at ScienceDirect

Geoderma

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

Using rule-based regression models to predict and interpret soil properties from X-ray powder diffraction data



Benjamin M. Butler^{a,*}, Sharon M. O'Rourke^b, Stephen Hillier^{a,c}

^a The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, UK

^b UCD School of Biosystems and Food Engineering, University College Dublin, Dublin 4, Ireland

^c Department of Soil and Environment, Swedish University of Agricultural Sciences (SLU), Uppsala SE-75007, Sweden

ARTICLE INFO

Handling editor: Edward A Nater

Keywords:

Soil mineralogy
X-ray powder diffraction
Soil properties
Cubist

ABSTRACT

Data mining is often used to derive calibrations for soil property prediction from diffuse reflectance spectroscopy, facilitating inference of organic and mineral contributions to given properties. In contrast to spectroscopy, X-ray powder diffraction (XRPD) offers a more direct probe into the complexities of soil mineralogy. Here a national scale XRPD dataset of Scottish soils is used in combination with the rule-based regression algorithm 'Cubist' for prediction of eight soil properties (total carbon and nitrogen, cation exchange capacity, pH, aqua regia extractable potassium, and the sand, silt and clay size fractions), and interpretation of soil property–mineralogy relationships. Precision sample preparation methods prior to XRPD analysis eliminated effects of preferred orientation, creating reproducible data appropriate for data mining. For direct comparison, Cubist was also applied to an equivalent dataset of near infrared spectroscopy (NIRS) measurements.

In terms of predictive performance, XRPD surpassed NIRS for prediction of six of the eight soil properties investigated. Notably, diffuse scattering from X-ray amorphous organic matter facilitated relatively accurate predictions of total carbon and nitrogen from XRPD. Aqua regia extractable potassium was predicted with substantial accuracy and confirmed to reflect the phyllosilicate potassium. The particle size fractions were predicted with moderate-substantial agreement using combinations of quartz, phyllosilicate and feldspar variables. This approach introduces the value of XRPD datasets in enhancing the understanding of soil mineralogy–property relationships whilst contributing to soil mineralogy's advance into the digital soil typing paradigm.

1. Introduction

There is a growing base of literature detailing the application of data mining algorithms to the prediction of soil properties from spectroscopic data (Minasny and McBratney, 2008, 2016; Nocita et al., 2015; Pérez-Fernández and Robertson, 2016; Reeves and Smith, 2009; Viscarra Rossel and Webster, 2012; Viscarra Rossel et al., 2016; Viscarra Rossel and Behrens, 2010). This chemometric analysis of soil spectral data is accepted as a rapid and cost-effective form of analysis where multiple soil properties can be derived from a single measurement. The approach facilitates attainment of greater spatial and temporal resolution (Sanchez et al., 2009), and also allows identification of specific soil constituents that contribute to each soil property (Viscarra Rossel et al., 2009). To date, visible-near infrared spectroscopy (vis-NIRS) is the most widely applied analytical technique for soil property prediction (Soriano-Disla et al., 2014; Viscarra Rossel et al., 2016).

The prediction of soil properties from visible-near infrared

spectroscopy is based on the intensity of absorbance bands that relate to functional groups in the organic and mineral components of a soil sample (Viscarra Rossel and Behrens, 2010). In terms of mineralogy, most soils contain a mixture of primary minerals derived from the soil's parent materials, and secondary minerals such as clay minerals, the occurrence of which is often controlled by weathering (Newman, 1984; Dixon and Schulze, 2002). As a whole therefore, soil mineralogy displays substantial heterogeneity on a variety of scales. Additionally, the suite of minerals present in a given soil are variously distributed amongst the sand, silt and clay size fractions, whilst also varying widely in chemical composition, crystal structure, surface area and solubility (Dixon and Schulze, 2002). Soil minerals are thus intimately related both directly and indirectly to many of the physical, chemical and biological properties of a soil (Newman, 1984; Andrist-Rangel et al., 2006). Given the complexity of soil however, it has been notoriously challenging to systematise soil property–mineralogy relationships (Newman, 1984).

* Corresponding author.

E-mail address: benjamin.butler@hutton.ac.uk (B.M. Butler).

Though vis-NIR spectra may contain considerable information on soil mineralogy (Viscarra Rossel et al., 2009), X-ray powder diffraction (XRPD) provides a more direct mineralogical probe because diffraction data is fundamentally related to the crystal structure and crystal chemistry of minerals in soils (Schulze, 1989). Rather than analysing a sample at a range of wavelengths, XRPD most commonly uses a monochromatic X-ray beam. The resulting signal (a series of peaks) as a function of diffraction angle (2θ) is related to atomic spacings in the ordered crystalline lattice by the 'Bragg' equation:

$$n\lambda = 2d \sin \theta \quad (1)$$

where n is an integer, λ is the monochromatic X-ray wavelength (angstroms, Å), d is the atomic spacing (d-spacing, Å) between planes of atoms, and θ is the angle between the incident rays and the plane of atoms. As such, XRPD produces diffraction patterns, or 'diffractograms', rather than 'spectra'. Chemical information about minerals is also encoded in the relative intensities of the various diffractogram peaks relating to a given mineral, and variation in mineral chemical composition may also alter d-spacings (and hence peak positions). Phase identification from these peaks can be readily achieved using comprehensive databases such as the Powder Diffraction File (ICDD, 2016). Further to the discrete 'Bragg' peaks derived from crystalline materials like minerals, the presence of amorphous phases within a sample results in diffuse scattering of X-rays across a wide 2θ range. Therefore amorphous soil constituents (e.g. organic matter and volcanic glass) typically result in broad maxima that are often considered as 'background' and consequently ignored or removed during analysis.

XRPD is particularly useful for analysing complex mineral mixtures, such as soil (Ulery and Drees, 2008), and with appropriate sample preparation (Hillier, 1999) can be used to accurately quantify soil mineral concentrations (Chipera and Bish, 2002; Omotoso et al., 2006). Mineral identification and quantification, however, are time consuming and often challenging undertakings- and the limited number of samples measured by XRPD in many soil science studies reflects these difficulties [e.g. Andrist-Rangel et al. (2010, 2013), Nagra et al. (2017), Jones and McBratney (2016), Kramer et al. (2017)]. Aside from these challenges, advances in sample preparation [e.g. spray drying, Hillier (1999)] now facilitate reproducible high-throughput XRPD analysis where hundreds or thousands of soil samples can be analysed for a single project (Towett et al., 2015; Barr et al., 2009). As such, a national scale dataset of Scottish soils has recently been analysed by XRPD with the aim of advancing the understanding of soil property–mineralogy relationships. At the same time, such datasets may contribute to aligning soil XRPD with the data-driven, 'digital', approaches widely applied in soil spectroscopy (Nocita et al., 2015).

Since it is currently impractical to apply accurate quantitative mineral analysis to high-throughput soil XRPD datasets, alternative techniques to relate mineralogy to properties must be applied. Here this alternative takes the form of data mining, which to our knowledge has not been previously applied to soil XRPD data. Specifically, this investigation aims to illustrate how a national scale XRPD dataset in combination with data mining can be used to predict and interpret soil (< 2 mm) properties, whilst also deriving descriptions for the way these properties are defined by, and linked to, soil mineral composition.

1.1. Hypotheses

A national dataset of Scottish soil properties and their corresponding XRPD measurements were used to investigate the following hypotheses:

- i) Data mining of XRPD data can be used to predict mineral soil properties.
- ii) Information derived from the models of predicted soil properties can be used for interpretation of soil property–mineralogy relationships.

2. Materials and methods

2.1. Soil dataset and sample selection

The National Soil Inventory of Scotland (NSIS) is an objective dataset of Scottish soils consisting of samples collected from two sampling campaigns. The first collection represents samples obtained between 1978 and 1988 from 721 sites defined by a 10 km grid across Scotland (hereafter NSIS 1). The second collection was obtained between 2007 and 2009 (hereafter NSIS 2), where a quarter of the original locations (184 sites) were re-sampled (Chapman et al., 2013; Pérez-Fernández and Robertson, 2016). During sampling for both NSIS 1 and NSIS 2, soils were taken from each of the main soil horizons at all sites, resulting in a combined archive of 3936 soil samples.

From the NSIS soil archive, all NSIS 2 samples ($n = 700$) were selected for analysis by XRPD (Section 2.3), whilst NSIS 1 samples from corresponding profiles were also retrieved and analysed if sufficient sample was available ($n = 546$). The same selection was previously applied for NIRS analysis [see Pérez-Fernández and Robertson (2016) and Section 2.3], thus permitting direct comparison of XRPD to a more established chemometric technique. For the purpose of this investigation the dataset of 1246 samples was refined further by selection of mineral horizons only (as recorded in the field). The mineral horizons were selected for two reasons: first, soil properties were found to display bimodal distributions in terms of the presence of mineral and organic 'clusters' (Fig. 1), which can create misleading performance parameters of predictive models; second, information derived from XRPD in this context is dominated by the mineral components of each soil sample, as opposed to organic and amorphous phases. In addition to removing organic horizons, calcareous soils (an uncommon soil type in the Scottish context) were identified using the Powder Diffraction File database (ICDD, 2016) in Bruker EVA software and were also removed from the dataset ($n = 9$).

Overall, the selection criteria resulted in a dataset comprised of 854 mineral soil samples measured by XRPD and NIRS. This encompassed 7 major soil groups (Fluvisols, Cambisols, Gleysols, Histosols, Podzols, Leptosols and Regosols; Table 1) and 32 primary rock types in the parent material (defined in the field, Table 2). The site locations of the selected samples, and the number of soil samples at each site ($n = 184$), are displayed in Fig. 2.

2.2. Soil properties

For each soil sample in the NSIS dataset, a wide variety of properties were measured that might be related to the XRPD and NIRS measurements using data mining. This study focuses upon eight of these properties: total carbon (C_T), total nitrogen (N_T), pH in water ($\text{pH}_{\text{H}_2\text{O}}$), cation exchange capacity (CEC), aqua-regia extractable potassium (K_{aqr}), and particle size distribution (sand, silt and clay). Together these span organic, textural, and chemical properties of soil. Most of these properties are also relevant to the Global soil Map project (Sanchez et al., 2009).

C_T and N_T were measured by mass spectrometry using the method described in Chapman et al. (2013). $\text{pH}_{\text{H}_2\text{O}}$ was determined with a combination electrode on the supernatant of a 3:1 mixture of distilled water (volume, cm^3) and soil (weight, g). To determine CEC, exchangeable base cations were displaced from the soil exchange sites using a neutral solution of ammonium acetate (1 M) and analysed by inductively coupled plasma atomic emission spectroscopy (ICP-OES). The CEC was then estimated as the potential cation exchange at pH 7.0, calculated as the sum of exchangeable Na^+ , Mg^{2+} , K^+ and Ca^{2+} concentrations (in $\text{cmol}_c \text{kg}^{-1}$) (Andrist-Rangel et al., 2010). K_{aqr} was determined by digesting 0.5 g of ground sample using the procedure of McGrath and Cunliffe (1985) as modified by McGrath (1987) (3:1 of 50% HCl:concentrated HNO_3 by vol.), with the digest being made up into 100 ml of 12.5% HNO_3 prior to analysis by ICP-OES. Particle size

Download English Version:

<https://daneshyari.com/en/article/8893929>

Download Persian Version:

<https://daneshyari.com/article/8893929>

[Daneshyari.com](https://daneshyari.com)