Prediction of soil parameters using the spectral range between 350 and 15,000 nm: A case study based on the Permanent Soil Monitoring Program in Saxony, Germany

Frank Riedel[a,*], Michael Denk[a], Ingo Müller[b], Natalja Barth[b], Cornelia Gläßer[a]

[a] Department of Remote Sensing and Cartography, Institute of Geosciences and Geography, Martin Luther University Halle-Wittenberg, Halle, Germany
[b] Saxon State Office for Environment, Agriculture and Geology, Freiberg, Germany

ARTICLE INFO

ABSTRACT

In this study, we tested the potential of visible-near infrared (VNIR, 350–2500 nm) and mid-infrared spectroscopy (MIR, 2500–15,000 nm) for quantification and prediction of soil parameters to support the Saxon Permanent Soil Monitoring Program. As Saxony is characterised by a large variety of soil metal concentrations which can negatively affect essential soil functions, the presented study is strongly focused on the forecast of soil metal contents. As data basis, a total of 203 soil samples of the Saxon Permanent Soil Monitoring Program, collected between 1998 and 2013 at 48 representative locations with respect to soil type, parent material, land use and climate conditions, were used. The chemical analysis provided information regarding total soil organic carbon content (TOC) and pH-value as well as element concentrations (Al, As, Ca, Cu, Fe, K, Mn, Na, Ni, Pb, Zn). VNIR-spectra were collected utilising an ASD FieldSpec Pro FR while an Agilent 4300 Handheld FTIR spectrometer was applied to cover the MIR wavelength region. Spectra pre-processing comprised the application of multi scatter correction (MSC), standard normal variate (SNV), continuum removal (CR) and the first and second derivatives. To model the relationship between soil spectral and chemical properties, and to predict element concentration, the spectra and chemical data were used as input for Partial Least Square Regression (PLSR) models. The definition of training data was realised on Kennard-Stone sampling algorithm and we selected 103 samples for model calibration and 100 samples for model validation. In general, the MIR-spectra and the MSC- and SNV-pre-processing improved the model performance. We obtained promising model results for TOC, Al, Fe, K and Ni with R^2-values between 0.70 and 0.88. Moderate results were obtained for Ca ($R^2 = 0.61$) and Mn ($R^2 = 0.43$). We conclude that VNIR and MIR spectroscopy has the potential to quickly provide reliable information regarding major soil parameters and metal contents and is thus a promising alternative approach to support soil analysis within the Saxon Permanent Soil Monitoring Program.

1. Introduction

Soils represent a dynamic, open and complex ecosystem and a fundamental natural resource for plants, animals and humans at the same time. The chemical, biological and physical properties, as well as the structure and composition, determine their balance and functionality (Viscarra Rossel et al., 2006; Shi et al., 2014). Although many metals represent a natural soil component as they are constituents of minerals and the liquid soil phase, the mentioned soil balance can be influenced and disturbed by unnaturally high levels of metal concentrations (Slonecker and Fischer, 2014). During the last years, such critical thresholds were exceeded worldwide in some regions as a

consequence of the rapid economic development and urbanisation processes. Other metal sources due to anthropogenic activities are extensive mining operations, including mining waste or water affected by acid mine drainage (Kemper and Sommer, 2002; Wu et al., 2005; Ren et al., 2009). Furthermore, besides hampering the quality of soils and their habitat function, metals can be assimilated by plants and thus enter the food chains and cause negative effects on human health (Song et al., 2012). In this context, long-term soil monitoring programs were developed in Germany to observe the current soil conditions, to act as warning systems for soil pollution and degradation and to derive reliable forecasts for its future development (Pälchen, 2001). Such a program was implemented in Saxony, a federal state with a long mining

history (Siems, 2002; Zimmermann et al., 2003). Permanent Soil Monitoring Sites were implemented and distributed over representative locations all over Saxony. The aim of these sites is the determination of background values and their spatiotemporal shift for specific soil types to estimate the existing soil pollution and furthermore the identification of natural and anthropogenic pollutant sources. To meet these requirements, the derivation of various soil physical- and chemical properties is necessary, which includes, for example, the pH-value, plant available nutrients or the total content of metals and metalloids like Al, Cd, Fe, Pb and As. Conventional methods for determining metal contents in soils are usually time-consuming and expensive and their quantification is often based on field sampling and different laboratory analyses (Wu et al., 2007). For this reason, the estimation and assessment of soil properties and metal contents based on reflectance spectroscopy have become an alternative tool in the field of soil monitoring (e.g. Ben-Dor and Banin, 1995; Malley and Williams, 1997; Janik et al., 1998; Chang et al., 2001; Kemper and Sommer, 2002; Udelhoven et al., 2003; Siebielec et al., 2004; Brown et al., 2006).

In the last decades, the derivation of soil parameters from spectral variations associated with metals and furthermore the binding forms of metals by spectrally active soil properties was well described by several authors (e.g. Xia et al., 2007; Choe, 2008; Ji et al., 2010; Scheffer and Schachtschabel, 2010). Also, the application of multivariate calibration techniques based on spectrometric measurements for the quantitative estimation of soil metal contents was successfully demonstrated. In most studies, soil constituents were modelled based on reflectance spectra covering the visible-near infrared (VNIR) (350–2500 nm) utilising regression models, like partial least square regression (PLSR), multiple linear regression (MLR) or principal component regression (PCR). It became obvious, that the quantitative estimation of metals within this spectral range is usually based on iron oxides, clay minerals and organic soil components (Shi et al., 2014). For example, Chang et al. (2001) successfully utilised PCR based on a dataset of 800 soil samples to predict Ca with a R^2 of 0.80. Kooistra et al. (2001) used PLSR to estimate Cd- and Zn concentrations of 69 floodplain soil samples and obtained model results with R-values > 0.90. Kemper and Sommer (2002) analysed 214 soil samples, which were affected by the Aznacollar mining accident in Spain, and could predict As, Fe, Hg, Pb contents with high accuracies (R^2 = 0.72–0.96) based on MLR. With PLSR, Wu et al. (2005) achieved R-values > 0.87 for Ni and Cr for 120 soil samples with anthropogenic metal input. Choe et al. (2009) investigated the metal contents of 22 soil samples from a former mining area and with MLR they reached promising results for the prediction of Cu (R^2 = 0.81). In contrast to previous studies, Ren et al. (2009) examined the contamination levels of 33 agricultural soils samples and based on PLSR, it was possible to quantify the concentrations of Fe (R = 0.88). The contamination level of floodplain soils was in the focus of the study of Vohland et al. (2009). Based on 149 samples, the highest PLSR prediction accuracies for metals were achieved for Cu (R^2 0.71). Also with PLSR, Zheng et al. (2011) successfully predicted As (R^2 = 0.71) and Lian et al. (2015) obtained promising results with MLR for Al (R^2 = 0.81) for 122 soil samples from a mining region.

Besides the numerous and successful applications of VNIR-spectroscopy as a tool for the rapid and non-destructive estimation of soil metal contents, in the last years an increasing number of studies focused on soil parameter modelling based on spectral range between 2500 and 15,000 nm, which is defined as the mid-infrared-region (MIR, according to Viscarra Rossel et al., 2006; Hecker et al., 2010; Song et al., 2012). In this spectral region, quartz, calcite, kaolinite and smectite and other OH-bearing minerals, as well as organic compounds, can lead to absorption features (Viscarra Rossel et al., 2006). Here, a few studies tested combined approaches with spectra from both ranges as model input to investigate the different binding forms and prediction possibilities. As presented by Siebielec et al. (2004), PLSR-model calibrations with the MIR-spectra showed substantially better results based on 70 soil samples from a mining region. In contrast, Moros et al. (2009)

described the PLSR-models based on VNIR-spectra as more robust and consistent, but the authors did not advocate a certain spectral range. For 117 sediment samples, the metal-specific prediction results were comparable. However, based on 111 agricultural soil samples, Dong et al. (2011) concluded that the PLSR results were slightly better for Zn, Pb and Cr using the MIR-spectra (R = 0.56–0.95). On the other hand, the PLSR-models derived by Song et al. (2012) showed better prediction capabilities for selected metals with the VNIR-spectral range.

As shown above, spectroscopic approaches in combination with regression methods are a feasible tool to estimate different soil parameters and especially soil metal contents. Despite the increasing application of such techniques, contrasting results for VNIR- and MIR-spectra regarding their prediction capability exists. Moreover, the achieved results can hardly be generalised as the soil properties, the analysed soil type, the number of elements and their concentration level differ for each study site. Furthermore, the majority of the soil samples were taken at a specific date or within a year, thus, the regression models were calibrated for a relatively short time period. Therefore, there is a demand to investigate how far multivariate calibration techniques can be applied without further limitations. This study focuses on how far well-established methods in form of PLSR-models based on VNIR- and MIR-soil spectra can provide reliable information about soil constituents and metal contents for heterogeneous soil samples from a wide study area, regardless of the soil type, sample depth and sampling time. For this, 203 soil samples and related analytical values, representing different locations and soil types, were provided by the Saxon State Office for Environment, Agriculture and Geology as part of their Permanent Soil Monitoring Program. The applicability of spectral soil information for the quantification and prediction of soil parameters will be tested. The study will demonstrate for the first time, how far spectral measurements can support and be connected with the concept and analytics of a long time German soil monitoring program.

2. Materials and methods

2.1. Study sites and soil samples

The 203 soil samples used in this study were collected between 1998 and 2013 at 48 representative locations all over Saxony, with respect to soil type, parent material, land use and climate conditions (Fig. 1). These sites are part of the Saxon Permanent Soil Monitoring Program (Barth and Forberg, 2015), which is characterised by performing periodic sampling and analysis including long-term retained sample storage. As a consequence of the huge study area, predominant soil types according to FAO (2015) are Cambisols, Luvisols, Gleysols, Fluvisols, Podzols, and Stagnosols with various subtypes and specifications. Land cover is ranging from typical agricultural sites to grassland areas and uncultivated regions. Depending on how the soil horizons developed on each site, two to seven samples were taken from soil depths between 15 and 180 cm, but no site was repeatedly sampled between 1998 and 2013 or rather we were using data from the first sampling cycle of the locations. The large variation of metal concentrations in these soils is related to bedrock and ore mineralisation in the southern part of Saxony, namely the Ore Mountains (Rank et al., 2009). Europe's maybe earliest mining activities could be drawn back to the Ore Mountains and mining really started to bloom when the first silver ore was discovered in its central region at Freiberg in 1168 CE (Asrih, 2013). Up to the 16th century, the Ore Mountains became the centre of Central Europe mining industry's development. As a consequence of > 800 years of mining and ore processing generated a widespread environmental pollution. Besides tailings and former mining and ore processing sites, large areas of agriculturally used land and alluvial plains which were located a hundred kilometres north were found to be affected by huge loads of various metals, mainly As, Pb and Cd (SMI, 2013). A recent publication from the Saxon geochemical soil