Estimating heavy metal concentrations in suburban soils with reflectance spectroscopy

Hang Cheng[a,b,c], Ruili Shen[d], Yiyun Chen[a,b,e,*], Qijin Wan[c,**], Tiezhu Shi[f], Junjie Wang[f], Yuan Wan[g], Yongsheng Hong[a], Xiaocui Li[a]

[a] School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China
[b] State Key Laboratory of Soil and Sustainable Agriculture, Chinese Academy of Sciences, Nanjing 210008, China
[c] School of Chemistry and Environmental Engineering, Wuhan Institute of Technology, Wuhan 430073, China
[d] Hubei Academy of Environmental Sciences, Wuhan 430072, China
[e] Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China
[f] College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China
[g] College of Urban and Environmental Sciences, Hubei Normal University, Huangshi 435002, China

ARTICLE INFO

ABSTRACT

Soil contamination by heavy metals has become a serious environmental issue worldwide. Rapidly and reliably obtaining heavy metal concentrations in soil is vital for soil monitoring and remediation. Visible and near-infrared reflectance (VNIR) spectroscopy provides a promising method for the estimation of heavy metal concentrations over large areas. Ninety-three soil samples were collected from a suburb of Wuhan City, Hubei Province, China, and their reflectance spectra were measured in the laboratory. This study aimed to (i) examine the feasibility of using soil reflectance spectra to estimate the concentrations of Cd, Pb, As, Cr, Cu and Zn in suburban soils; (ii) compare the performances of different spectral pretreatments and (iii) explore the mechanism underlying the estimation of heavy metal concentration from VNIR spectra. In particular, we proposed a strategy for the mechanism investigation that combined PCA biplot analysis, correlation and partial correlation analyses. Partial least-square regression was adopted to calibrate the VNIR model. Results showed that the VNIR model provided acceptable estimation accuracies for Cr, As and Cd concentrations with the ratio of the performance to deviation (RPD) values of 2.70, 1.81 and 1.63, respectively, but unsatisfactory estimation accuracies for Pb, Cu and Zn concentrations with the RPD values of 0.70–1.03. Savitzky–Golay smoothing outperformed other spectral pretreatments. The mechanisms underlying the estimation of the six studied heavy metals varied on a case-to-case basis. Specifically, the spectral estimation of Cd (Group I) concentration was attributed to its close correlations with soil organic matter (SOM). Cr and As (Group II) concentrations could be estimated by the VNIR model on the basis of their close correlations with Fe. Pb, Cu and Zn (Group III) concentrations, however, had weak correlations with neither SOM nor Fe, resulting in poor estimations. The proposed strategy on mechanism investigation for heavy metals could be transferred to other study areas. In summary, VNIR spectroscopy combined with the PLSR model is an alternative method for the rapid monitoring of some heavy metal pollution in suburban soils.

1. Introduction

Soil contamination with heavy metals has become an increasingly severe environmental problem (Shi et al., 2018). Suburban soils are a spatial transition zone between urban and rural areas and maintain the balance between the local environment and the ecosystem (Yang et al., 2009). However, soil heavy metal concentrations in suburban soils may accumulate more quickly in suburban soils than in rural soils because of rapid urbanisation and industrialisation (Meng et al., 2014). Therefore, quick access to information on heavy metal concentrations in suburban soils may facilitate the environmental monitoring of the local areas.

Conventional approaches to investigating soil heavy metal contamination involve intensive field sampling and laboratory analysis (Leenaers et al., 1990), which are time consuming and expensive (Shi

et al., 2014; Sun et al., 2014). Visible and near-infrared reflectance (VNIR, 350–2500 nm) spectroscopy has been widely applied in soil science because it is rapid, non-destructive and cost effective (Sun et al., 2017). Most previous studies on the estimation of heavy metals through VNIR spectroscopy have mainly focused on soils that were collected from mining areas (Choe et al., 2008; Sun and Zhang, 2017; Zhuang et al., 2009), floodplains (Kooistra et al., 2001; Liu and Chen, 2012a, 2012b), agricultural soils (Shi et al., 2014, 2017), sediment (Choe et al., 2008; Jiang et al., 2018; Malley and Williams, 1997; Moros et al., 2009) and urban areas (Pandit et al., 2010). However, save for two studies (Wu et al., 2005, 2007) conducted 10 years ago, the estimation of heavy metal concentrations in suburban soils from VNIR spectral information has received limited attention and thus warrants further investigation.

Partial least-squares regression (PLSR) is the most commonly adopted strategy in VNIR modeling (Viscarra Rossel et al., 2006). Different spectral preprocessing techniques have been adopted to enhance spectral features and establish a robust estimation model. For example, Shi et al. (2014) combined Savitzky–Golay smoothing (SG), first-derivative (FD) and mean-centring (MC) spectral pretreatments to construct an optimal calibration model for estimating soil As concentrations from the spectra of rice plants. Chen et al. (2015) investigated the use of PLSR models to estimate soil Cd from raw and preprocessed spectra and found that the orthogonal signal correction method can effectively remove irrelevant information, thus improving estimation accuracy. Therefore, the application of these spectral pretreatments may be essential in the estimation of heavy metal concentrations in suburban soils.

Several works have explored the mechanism underlying the estimation of heavy metal concentrations from the VNIR spectra of soil samples. For example, by applying Pearson correlation analysis, Wang et al. (2014) found that the spectrally featureless heavy metals (i.e. Pb, Zn and Cu) had a significant correlation with soil organic matter (SOM). Moreover, Liu and Chen (2012b) reported that the mechanism for estimating Cu concentration lay in its correlation with Fe concentration. Both SOM and Fe have been regarded as spectrally active soil components (Chen et al., 2015). Either SOM or Fe could be the confounder that influences the accuracy of estimating heavy metal concentrations from soil spectra. To the best of our knowledge, the confounding effects of these factors have seldom been considered in the estimation of heavy metal concentrations from the VNIR spectra of soil samples. Partial correlation analysis is commonly applied to reveal confounding effects in the analysis of financial markets (Kenett et al., 2015; Li et al., 2017), gene networks (Freudenberg et al., 2009) and meteorology (Zhou et al., 2000).

Principal component analysis (PCA) can be used to clarify the complex interrelationships among different heavy metals (Wu et al., 2007). PCA is a valid tool for extracting the most important information from data, such as soil spectra, and for compressing datasets, such as the spectral matrix, into different factors or principal components (Mishra et al., 2017). Through PCA, Wu et al. (2005) obtained a component-loading plot that showed correlations between heavy metal and Fe concentrations. Shi et al. (2017) also employed PCA to extract the relevant spectral features for the spectroscopic diagnosis of As contamination in agricultural soils. A biplot is a two-dimensional scatter diagram that depicts the scattering pattern of properties on the basis of principal factors identified through PCA (Aslam et al., 2017). Principal component biplot analysis (PCA biplot) is commonly used in biomedical analyses and plant genotypic selection (Pasqualoto et al., 2007; Ukalski and Klisz, 2016). However, this method has rarely been used in research on soil contamination by heavy metals. PCA biplot, correlation and partial correlation analyses may be combined to obtain an interesting and effective approach that may better elucidate the mechanism underlying the estimation of heavy metal concentrations from the VNIR spectra of soil samples than other approaches.

The objectives of this study were to (i) examine the feasibility of estimating Cd, Pb, As, Cr, Cu and Zn concentrations in suburban soils of Wuhan City from VNIR spectra; (ii) compare the heavy metal estimation accuracies of PLSR models subjected to different spectral pretreatments and (iii) propose a comprehensive strategy that combines PCA biplot, correlation and partial correlation analyses for exploring the mechanism that underlies the estimation of heavy metal concentrations from VNIR spectra.

2. Materials and methods

2.1. Study area and sampling

The study area is located in the west of Wuhan City, Hubei Province, China. This area is characterised by a subtropical humid monsoon climate with abundant sunlight and rainfall and distinct seasons (Hong et al., 2017). Wuhan has experienced rapid urbanisation over the past several decades (Jiao et al., 2017). In recent years, contamination caused by urbanisation has drastically affected the soil ecological environment, including that in suburban areas.

In October 2013 and May 2014, 93 topsoil samples (depth: 0–20 cm) were collected in accordance with the uniform grid method with a sampling density of one sample per km^2. The location of each sampling site was recorded by a handheld global position system (GPS). The locations of sampling sites are shown in Fig. 1. Stones, plant residues and other large debris were removed from each fresh sample, which was then stored in a labelled plastic bag. The 93 soil samples were collected from areas with the land use types of wheat (*Triticum aestivum*), paddy and vegetable lands. In accordance with the Chinese Soil Taxonomic Classification, soil samples were classified as paddy soils and fluvo-aquic soils.

2.2. Chemical analysis of soil properties

All soil samples were air-dried at room temperature for three days and sieved through a 2 mm polyethylene sieve to remove stones and other debris (Chen et al., 2016). Afterwards, the samples were ground into fine particles and passed through a 100-mesh sieve (0.149 mm) for further analysis (Wang et al., 2014). The soil samples were preprocessed through electric heating board acid ($HCl - HNO_3 - HClO_4$) digestion and then subjected to inductively coupled plasma atomic emission spectroscopy for the determination of Pb, As, Cr, Cu and Zn concentrations (Wu et al., 2007). Cd concentrations were determined through inductively coupled plasma mass spectrometry after digestion with HNO_3 and $HClO_4$ (Chen et al., 2015). The SOM concentrations of all the soil samples were determined through wet oxidation at 180 °C with a mixture of potassium dichromate and sulphuric acid (Liu et al., 2014). Total Fe concentrations were determined through power $X - ray$ fluorescence spectrometry (Liu and Chen, 2012a).

2.3. Spectral measurement and pretreatments

Soil spectra were measured by an ASD Field Spec® 3 portable spectroradiometer (Analytical Spectral Devices, Inc., USA) which covers a spectral range of 350–2500 nm. Its spectral resolution is 3 nm in the 350–1000 nm range and 10 nm in the 1000–2500 nm range. Its sampling interval is 1.4 nm in the 350–1000 nm range and 2 nm in the 1000–2500 nm range (Shi et al., 2014). Spectral measurements were collected in a dark room to minimise the influence of external light. The stable light source was a 50 W halogen lamp matched with a spectroradiometer and placed 30 cm away from the sample with a 45° incident angle (Hong et al., 2017, 2018). An optical probe was installed approximately 12 cm above the centre of the sample surface with a 10° field of view. Each soil sample was uniformly tiled in a petri dish and was successively scanned 10 times (Shi et al., 2014). The spectroradiometer was calibrated before spectral measurements and recalibrated after every 10 samples by using a standardised white $BaSO_4$ panel (Chen et al., 2016).