Environmental Pollution 206 (2015) 217–226

Contents lists available at ScienceDirect

Environmental Pollution

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

Rapid identification of soil cadmium pollution risk at regional scale based on visible and near-infrared spectroscopy

Tao Chen [a, *], Qingrui Chang [a, **], J.G.P.W. Clevers [b], L. Kooistra [b]

[a] College of Resources and Environment, Northwest A&F University, Shaanxi, Yangling 712100, China
[b] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 17 March 2015
Received in revised form
4 July 2015
Accepted 7 July 2015
Available online xxx

Keywords:
Soil cadmium
Visible and near-infrared spectroscopy
Orthogonal signal correction
Back propagation neural network
Indicator kriging

ABSTRACT

Soil heavy metal pollution due to long-term sewage irrigation is a serious environmental problem in many irrigation areas in northern China. Quickly identifying its pollution status is an important basis for remediation. Visible-near-infrared reflectance spectroscopy (VNIRS) provides a useful tool. In a case study, 76 soil samples were collected and their reflectance spectra were used to estimate cadmium (Cd) concentration by partial least squares regression (PLSR) and back propagation neural network (BPNN). To reduce noise, six pre-treatments were compared, in which orthogonal signal correction (OSC) was first used in soil Cd estimation. Spectral analysis and geostatistics were combined to identify Cd pollution hotspots. Results showed that Cd was accumulated in topsoil at the study area. OSC can effectively remove irrelevant information to improve prediction accuracy. More accurate estimation was achieved by applying a BPNN. Soil Cd pollution hotspots could be identified by interpolating the predicted values obtained from spectral estimates.

1. Introduction

Soil heavy metal pollution has become an important environmental issue (Tchounwou et al., 2012). Amongst various heavy metal elements, cadmium (Cd) is identified as an extremely significant pollutant because of its high transfer rate from soil to plants and strong bio-toxicity (Das et al., 1997; Mclaughlin and Singh, 1999; Satarug et al., 2010). Excessive Cd uptake can result in plant growth inhibition and even death (Gussarsson et al., 1996; Haghiri, 1974). Worse is that it also can enter the human body through the food chain, threatening health, especially for children and pregnant women (Peralta-Videa et al., 2009). Many studies have shown that soil Cd is an important source of human Cd intake (Jarup, 2003). Therefore, quickly obtaining information on its pollution level and spatial distribution is very important for remediation.

However, the conventional method of determining Cd concentration in soil requires complex, multi-step laboratory analyses that involve strong acid digestion and concentration process (Gholizadeh et al., 2015). This method is time-consuming, expensive and requires excellent operation skills. Soil reflectance spectroscopy provides a promising tool for the efficient detection and monitoring of soil contaminants (Ben-Dor et al., 1997; Gholizadeh et al., 2015). Using reflectance spectroscopy measurements in the visible-near-infrared (VNIR 400–1100 nm) and shortwave-infrared (SWIR 1100–2500 nm) range, some direct and indirect soil properties can be extracted (Gholizadeh et al., 2015). These chemometric processes are termed NIRA (near-infrared reflectance analysis) (Stenberg and Rossel, 2010). Dalal and Henry (1986) used the NIRA method to predict soil moisture, organic carbon, and total nitrogen contents in Australia. Ben-Dor and Banin (1995) simultaneously evaluated soil clay content, surface area, cation exchange capacity, hygroscopic moisture, OM and $CaCO_3$ contents in Israel. As regards heavy metals, the first report on quantitative prediction by NIRA was published by Malley and Williams (1997). Hereafter, some similar studies were performed in different mining areas and floodplains (Choe et al., 2008; Moros et al., 2009; Siebielec et al., 2004). In recent years, more and more focus was put on estimating low concentrations of heavy metals in agricultural soils (Ren et al., 2009; Wang et al., 2014; Wu et al., 2005), which had a close relationship with food safety.

Among these existing studies, researchers found that certain specific information was not easy to be extracted and stripped from weak soil spectral signals. Due to the comparable size of the

* Corresponding author.
** Corresponding author.
E-mail addresses: davidlp77@zju.edu.cn (T. Chen), changqr@nwsuaf.edu.cn (Q. Chang).

http://dx.doi.org/10.1016/j.envpol.2015.07.009
0269-7491/

wavelengths in VNIR-SWIR electromagnetic radiation and particle sizes in soil samples, the reflectance spectroscopy is a battleground for undesired scatter effects (both baseline shift and non-linearities) that influence the sample spectra (Rinnan et al., 2009). By applying suitable pre-treatments, these effects can largely be eliminated. For example, Ben-Dor et al. (1997) used first and second derivative absorption spectra to enhance the spectral information. Stenberg and Rossel (2010) compared the effects of log(1/R) transformation, first derivative, and SNV-Detrending on soil reflectance spectra. Hively et al. (2011) compared 30 combinations of spectral math pre-treatments and found that first derivative worked well for nearly all analytes (including soil carbon, particle size distribution, and some agronomical elements). Generally speaking, these pre-treatments are divided into two categories: scatter-correction and spectral derivative (Buddenbaum and Steffens, 2012). No matter which pre-treatment technique is used, its spectral math manipulations are performed only based on the spectral matrix **X** without considering soil physical-chemical components **Y** (response variable). However, orthogonal signal correction (OSC), as a commonly used approach in food science, can purposefully remove irrelevant information from **X** that is mathematically orthogonal to target matrix **Y**, or as close to orthogonal as possible (Wold et al., 1998). The result is that more valuable spectral information is retained that is closely related to the target variable (Y). Unfortunately, so far no study was reported about the application of OSC in soil science, especially in heavy metal estimation.

In addition, choosing the appropriate calibration method can also help to achieve a more reliable prediction model in NIRA. At present, many studies used Multiple Linear regression (MLR) (Kemper and Sommer, 2002), Principal Component Regression (PCR) (Wu et al., 2005), or Partial Least Squares Regression (PLSR) (Kooistra et al., 2001) to estimate the concentrations of soil heavy metals, in which PLSR became one of the most popular calibration technologies. As a classic linear calibration method, it inevitably had some deficiencies when it was used to model some non-linear relationships. Artificial neural networks (ANN) overcomes this problems, which has the ability to model linear or non-linear relationships between a set of input and output variables (Ben-Dor et al., 1997). For example, Daniel et al. (2003) confirmed the capability of ANN to estimate soil macronutrients by solving difficulties incurred from high cross-channel correlations. Meanwhile, it has also been used to estimate soil salt (Farifteh et al., 2007), bulk density (Al-Asadi and Mouazen, 2014) and water infiltration rate (Goldshleger et al., 2012) etc. However, to our knowledge research on the application of ANN in estimating soil heavy metals is still rarely reported.

Sewage irrigation is an important pollution source of soil heavy metals. In the past few decades, sewage irrigation in northern China has become a common phenomenon, especially in the arid and semi-arid regions. In recent years, to improve the eco-environment, remediation of the pollution is undertaken by the government by prohibiting agricultural practices, covering with new soil or phytoextraction. In this process, rapid and repeated assessment of the pollution status is necessary in order to decide the moment that the pollution has reduced to acceptable levels. However, it is difficult for the conventional method to meet the practical needs due to its high costs and time-consuming characteristics.

Therefore, in this study, based on different spectral pre-treatments and calibration methods, soil Cd pollution in a typical sewage irrigation area was estimated using the NIRA technique in the VNIR region. On this basis, the spatial distribution and pollution hotspots of Cd were obtained. This case study aims to: i) identify most effective pre-processing and calibration methods by comparing the estimation accuracy of soil Cd; ii) reveal the estimation mechanism in the typical sewage irrigation area; and iii) explore the feasibility of rapidly identifying soil Cd pollution hotspots at regional scale based on the NIRA technique and geostatistics.

2. Materials and methods

2.1. Study area and sampling

This study was conducted in the Fenghui irrigation region, which is situated in the northwest of Xi'an City, Shaanxi Province, China (Fig. 1). Since the early 1950s sewage irrigation has started in this region and gradually has expanded. With the enhancement of environmental consciousness, sewage irrigation fully stopped in 2010. This region has distinctive seasons with an annual mean temperature of 13.4 °C and a mean annual precipitation of 580 mm (Chen et al., 2014). Local terrain is relatively flat, with an average elevation of 382 m. The dominant soil type is Loessial soil (Hapli-Ustic Cambosols according to the Chinese soil taxonomy), with an average pH of 8.5 and a silty soil texture. Due to long-term sewage irrigation, soil nutrient contents are generally higher than the mean standard of Shaanxi soils (Chen et al., 2014). The crop rotation is mostly wheat (*Triticum aestivum* L.) followed by maize (*Zea mays* L.).

Two soil sampling campaigns were carried out in May 2010 and May 2012. The first sampling consisted of 52 sites spread over the study area in a random design and in total 52 soil samples were collected (depth 0–20 cm). In the second sampling, 8 sampling sites were chosen according to the distance from the irrigation canal. At each site, three soil samples were collected, totaling 24 samples: 8 surface (depth 0–20 cm), 8 subsurface (20–40 cm), and 8 mixed (0–40 cm). The location of each sampling site was determined by GPS. As a result, a total of 76 samples were collected.

Soil samples were air-dried at room temperature, removed of stones or other debris, and passed through a 0.149 mm mesh. The samples were analyzed for soil organic matter (SOM), total Fe_2O_3, and Cd. In the chemical analysis, SOM was determined by the $K_2Cr_2O_7$–H_2SO_4 oxidation method and total Fe_2O_3 concentrations were determined by power X-ray fluorescence spectrometry. Concentrations of Cd were measured by an inductively coupled plasma mass spectrometer after digestion with nitric acid (HNO_3) and perchloric acid ($HClO_4$) (Agricultural Chemistry Committee of China, 1983). Duplicate tests were carried out, and quality control procedures were conducted by using certified reference materials (GSS-17 and GSS-19).

2.2. Spectral measurement and preprocessing

Soil spectra were measured by a portable FieldSpec® Spectrometer (ASD Inc., USA), which covers the VNIR region (325–1075 nm) and offers a spectral resolution of 3 nm at 700 nm. Spectral sampling interval is 1.4 nm and data were resampled to 1 nm output values. The spectral measurements were performed in a dark room. A 50 W halogen lamp pointing at an angle of 30° from nadir was mounted on a tripod (40 cm above samples) as the stable light source. A 7.5° field-of-view at nadir was chosen and the spectrometer was fixed at 40 cm above the center of the samples. Each soil sample was uniformly tiled in a petri dish and was scanned 10 times. The average spectral curve was calculated and further used for pre-treatment and subsequent modeling. The spectrometer was recalibrated after every 10 samples using a white $BaSO_4$ panel. In order to reduce noise and calculation time, the reflectance curves firstly were smoothed using a pseudo-Gaussian method, and the spectral bands in the intervals 325–399 nm and 1001–1075 nm were removed. Then, the reflectance values were resampled across a 9-nm wide window, which resulted in 67