ARTICLE IN PRESS

STOTEN-21150; No of Pages 10

Science of the Total Environment xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Science of the Total Environment

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

Modeling and mapping of cadmium in soils based on qualitative and quantitative auxiliary variables in a cadmium contaminated area

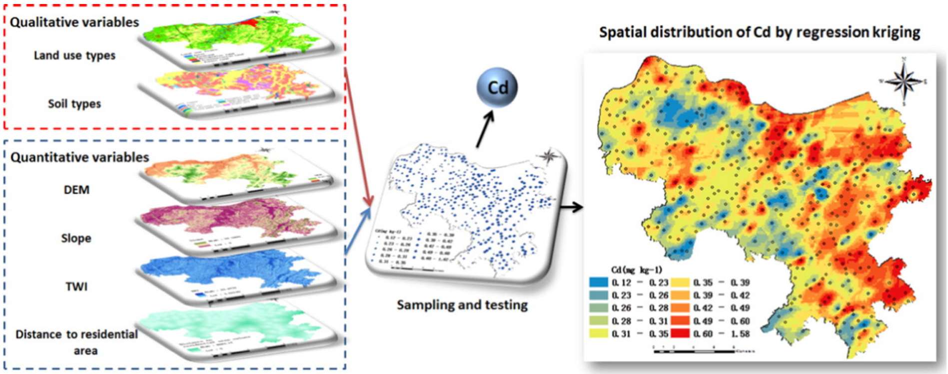
Shanshan Cao [a,b], Anxiang Lu [a,b,c], Jihua Wang [a,b,*], Lili Huo [d]

[a] *Beijing Research Center for Agricultural Standards and Testing, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China*
[b] *Beijing Municipal Key Laboratory of Agriculture Environment Monitoring, Beijing 100097, China*
[c] *Collaborative Innovation Center for Key Technology of Smart Irrigation District in Hubei, Yichang 443002, China*
[d] *Agro-Environmental Protection Institute of Ministry of Agriculture, Tianjin 300191, China*

HIGHLIGHTS

- Regression kriging (RK) was used to map the spatial distribution of Cd in soils.
- Relationship between auxiliary variables and cadmium in soils was analyzed.
- Introduction of qualitative variables can improve the interpolation accuracy of Cd.
- Appropriate auxiliary variables were essential to improve the accuracy of RK.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 25 July 2016
Received in revised form 25 September 2016
Accepted 12 October 2016
Available online xxxx

Editor: D. Barcelo

Keywords:
Heavy metals
Spatial prediction
Environmental variable
Variability
Regression kriging

ABSTRACT

The aim of this study was to measure the improvement in mapping accuracy of spatial distribution of Cd in soils by using geostatistical methods combined with auxiliary factors, especially qualitative variables. Significant correlations between Cd content and correlation environment variables that are easy to obtain (such as topographic factors, distance to residential area, land use types and soil types) were analyzed systematically and quantitatively. Based on 398 samples collected from a Cd contaminated area (Hunan Province, China), we estimated the spatial distribution of Cd in soils by using spatial interpolation models, including ordinary kriging (OK), and regression kriging (RK) with each auxiliary variable, all quantitative variables (RKWQ) and all auxiliary variables (RKWA). Results showed that mapping with RK was more consistent with the sampling data of the spatial distribution of Cd in the study area than mapping with OK. The performance indicators (smaller mean error, mean absolute error, root mean squared error values and higher relative improvement of RK than OK) indicated that the introduction of auxiliary variables can improve the prediction accuracy of Cd in soils for which the spatial structure could not be well captured by point-based observation (nugget to sill ratio = 0.76) and strong relationships existed between variables to be predicted and auxiliary variables. The comparison of RKWA with RKWQ further indicated that the introduction of qualitative variables improved the prediction accuracy, and even weakened the effects of quantitative factors. Furthermore, the significantly different relative improvement with similar R^2 and varying spatial dependence showed that a reasonable choice of auxiliary variables and analysis of spatial structure of regression residuals are equally important to ensure accurate predictions.

---

[*] Corresponding author.
E-mail address: wangjh3488@163.com (J. Wang).

http://dx.doi.org/10.1016/j.scitotenv.2016.10.088
0048-9697/© 2016 Published by Elsevier B.V.

Please cite this article as: Cao, S., et al., Modeling and mapping of cadmium in soils based on qualitative and quantitative auxiliary variables in a cadmium contaminated area, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.10.088

1. Introduction

The contamination of heavy metals in soils has received particular attention by governments and regulatory bodies. Cadmium (Cd) had been recognized as a high biotoxicity heavy metal because of its strong chemical activity, high persistence toxicity and enrichment effect in soils, and particularly for its cumulative effect in human beings (Gupta and Gupta, 1998). Obtaining efficient and accurate information on spatial variation of heavy metals in soils has become increasingly important because such information is critical to sustainable land use and remediation (Florinsky et al., 2002).

Spatial interpolation techniques are widely used to predict the spatial distribution of heavy metals in soils from data based on discrete samples. However, the spatial variation of heavy metals in soils is usually strong and complex because of the combined effects of the various surface environmental factors. Numerous studies have shown significant improvement in prediction accuracy using methods that consider the influence of environmental factors by introducing auxiliary variables, compared with commonly used interpolation methods (inverse distance weighting, Spline and ordinary kriging) that are based on discrete samples only (Li et al., 2013; Phachomphon et al., 2010; Shi et al., 2011). Regression kriging (RK), which combines a generalized regression model with ordinary kriging of the regression residuals (Hengl et al., 2004), was demonstrated to be one of the most representative hybrid geostatistical method because of its ease of use and high accuracy (Lin et al., 2011; Pásztor et al., 2016). In most cases, the interpolation precision of RK was superior to ordinary kriging (McBratney et al., 2000; Minasny and McBratney, 2007; Simbahan et al., 2006) and cokriging (Knotters et al., 1995).

Correlations between dependent and independent variables have a marked effect on the prediction accuracy of RK. In general, all of the auxiliary variables (topographic indexes, distance from pollution sources and density of different land use types) that were involved in the spatial interpolation model to predict the distribution of heavy metals in soil are quantitative variables (Ahmed et al., 2011; Dragovic et al., 2014; Lado et al., 2008; Lin et al., 2011). However, research into the application of qualitative variables as auxiliary factors in the geostatistical interpolation model of heavy metals in soil is limited, despite significant differences in the spatial distribution of heavy metals in different land use types, soil texture and land covers having been examined (Xia et al., 2011; Yong Yang and Christakos, 2015).

Hunan is a leading province in rice crop area and yield in the south of China. However, the problem of exceeding the quality standard for Cd in rice has aroused widespread concerns. As a province associated with high Cd pollution incidence, soil remediation and non-producing areas division of Hunan have been considered imminent (Wang et al., 2016). Thus, high accuracy interpolation of spatial distribution of Cd in soils is essential.

This study aimed to determine a combination of easy-to-obtain auxiliary variables that can improve the prediction accuracy of Cd in soils using RK, based on the combined consideration of spatial structure and the relationship with available auxiliary variables. Using the least-significant difference (LSD) method and Pearson correlation analysis, relationships between Cd in soils and environmental variables were analyzed systematically and quantitatively to obtain the most efficient auxiliary variables. In addition, ordinary kriging and regression kriging with each auxiliary variable, all quantitative variables and all auxiliary variables were built to evaluate the relative performances of mapping heavy metals based on the auxiliary information and the structure of the regression model was considered.

2. Materials and methods

2.1. Study area and calculation of environmental factors

The study area (112°24′47″–113°03′34″E, 27°20′13″–28°04′53″N) is located in the central eastern part of Hunan Province in south China, which is known as the "hometown" of nonferrous metals, and near to the river valleys known to be subject to high soil heavy metal pollution around the Xiangjiang River (Guo et al., 2008) (Fig. 1). The main landform features included in this 522.76 km^2 area are downlands, mountains, gentle plains and hills with elevations between 26 and 269 m above sea level and slopes between 0 and 23°.

Besides terrain factors that commonly used to predict the spatial distribution of soil properties in previous studies, the auxiliary environmental variables introduced in this study include soil types that determined the physical and chemical properties of the soil so as to affect the accumulation and migration of heavy metals in soil, land use types and distance to residential area that reflect the influence of human factors on the spatial distribution of heavy metals in soil, (Dragovic et al., 2014). It has been proved that these variables have significant impacts on the spatial distribution of heavy metals in soil, and can be obtained easily by using spatial analysis function of geographic information system based on the basic geographic data (such as digital elevation model data, soil map and land utilization map, etc.) of the study area, without the need of additional monitoring cost (Xu et al., 2013). These variables can be divided into two categories: qualitative variables containing land use types and soil types, and quantitative variables containing terrain factors and distance to residential area. All of them were generated using the spatial analyst extension of ArcGIS 10.1(ESRI, 2010).

Soil types of the investigated area, obtained from 1:50,000 soil maps for the Xiangtan County, are mainly red soil, purple soil and paddy soil (Fig. 2a) (Soil Survey Office of Xiangtan City, 1984). Of these, paddy soil, which occupies 46.75% of the total area, can be further divided into ore poisoned, gleyed, percogenic, submerged and hydromorphic paddy soil.

Major land uses that were extracted from the Xiangtan land utilization map at a cartographic scale of 1:50,000 include farmland, forest land and residential land occupying 41.88%, 53.03% and 5.09% of the total area, respectively (Fig. 2b).

Terrain factors adopted in this study were elevation (h, m, above mean sea level), slope (β, degrees, 0–90°) and the topographic wetness index (WTI). Elevation and slope were derived from a 30-m grid digital elevation model (DEM) of the study area using the spatial analytical tools of ArcGIS directly (Fig. 1 and Fig. 2c). The WTI is likely to exhibit a significant relationship with the soil properties because it has been recognized as an accurate description of topographic change and its influence on the surface runoff (Beven and Kirkbj, 1979; Pei et al., 2010) (Fig. 2d). The two independent variables used to calculate the WTI were acquired from the DEM and applied as follows:

$$\mathrm{WTI} = \ln\left(\frac{A_c}{\tan\beta}\right) \quad (1)$$

where A_c is catchment area per unit contour length that flows through the surface of the point to be computed and β is slope gradient. Both of the independent variables that used to calculate WTI were acquired from the DEM.

Distance to residential area (D_{res}) reflects the intensity of human activities. Land that is located close to the residential areas has more serious accumulation of heavy metals caused by human activities. "Residential area" in this study refers to the large-scale residential land (town and village). D_{re} was calculated by the Euclidean distance between sampling points and their nearest residential area using the spatial analysis methods of the Geographic Information System (Fig. 2e).

2.2. Soil sampling and chemical analysis

In consideration of the distribution of sampling sites among various areas, land use types, soil types, topographical positions, as well as distance to residential area, 398 soil samples were collected in July 2013

Please cite this article as: Cao, S., et al., Modeling and mapping of cadmium in soils based on qualitative and quantitative auxiliary variables in a cadmium contaminated area, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.10.088