Science of the Total Environment 490 (2014) 422–434

Contents lists available at ScienceDirect

Science of the Total Environment

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

Applications of stochastic models and geostatistical analyses to study sources and spatial patterns of soil heavy metals in a metalliferous industrial district of China

Buqing Zhong, Tao Liang *, Lingqing Wang, Kexin Li

Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

HIGHLIGHTS

- Conditional inference tree can identify variables controlling metal distribution.
- Finite mixture distribution model can partition natural and anthropogenic sources.
- Geostatistics with stochastic models can delineate soil contaminated area.

ARTICLE INFO

Article history:
Received 16 February 2014
Received in revised form 29 April 2014
Accepted 29 April 2014
Available online xxxx

Editor: F.M. Tack

Keywords:
Heavy metals
Soil pollution
Conditional inference tree
Finite mixture distribution model
Geostatistical analysis

ABSTRACT

An extensive soil survey was conducted to study pollution sources and delineate contamination of heavy metals in one of the metalliferous industrial bases, in the karst areas of southwest China. A total of 597 topsoil samples were collected and the concentrations of five heavy metals, namely Cd, As (metalloid), Pb, Hg and Cr were analyzed. Stochastic models including a conditional inference tree (CIT) and a finite mixture distribution model (FMDM) were applied to identify the sources and partition the contribution from natural and anthropogenic sources for heavy metal in topsoils of the study area. Regression trees for Cd, As, Pb and Hg were proved to depend mostly on indicators of anthropogenic activities such as industrial type and distance from urban area, while the regression tree for Cr was found to be mainly influenced by the geogenic characteristics. The FMDM analysis showed that the geometric means of modeled background values for Cd, As, Pb, Hg and Cr were close to their background values previously reported in the study area, while the contamination of Cd and Hg were widespread in the study area, imposing potentially detrimental effects on organisms through the food chain. Finally, the probabilities of single and multiple heavy metals exceeding the threshold values derived from the FMDM were estimated using indicator kriging (IK) and multivariate indicator kriging (MVIK). The high probabilities exceeding the thresholds of heavy metals were associated with metalliferous production and atmospheric deposition of heavy metals transported from the urban and industrial areas. Geostatistics coupled with stochastic models provide an effective way to delineate multiple heavy metal pollution to facilitate improved environmental management.

1. Introduction

Soil pollution with heavy metals has been an ever growing concern due to the potential threat to food safety and its detrimental effects on human and animal health. Heavy metals are introduced into soils and environment through both natural and anthropogenic sources. The natural inputs of heavy metals in soils are attributed to geological parent materials. On the other hand, the anthropogenic sources of heavy metals have become more complex in recent decades, including metalliferous mining and smelting, chemical industry, fossil fuel combustion, waste incineration, agricultural activities, and atmospheric deposition (Alloway, 2013; Cheng, 2003).

It is essential to identify pollution sources and to map high risk areas before pollution control actions are taken (Chen et al., 2009). Multivariate statistics combined with geostatistics have been applied to identify the sources and spatial distribution patterns of heavy metals in soils (Li et al., 2013; Lu et al., 2012; Maas et al., 2010; Martin et al., 2006; Saby et al., 2009). Relationships between variables, however, may be strongly nonlinear and involve high-order interactions, whereas it could be hard for the commonly used multivariate statistical techniques to find meaningful patterns of metal concentrations in soils (De'ath and Fabricius,

* Corresponding author at: Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11 A Datun Road, Beijing 100101, China. Tel.: +86 10 64859781; fax: +86 10 82521857.
E-mail address: liangt@igsnrr.ac.cn (T. Liang).

http://dx.doi.org/10.1016/j.scitotenv.2014.04.127
0048-9697/© 2014 Elsevier B.V. All rights reserved.

2000). To overcome this problem, stochastic models can be useful tools for pollution source identification, based on the statistical analysis of sampled data in combination with corresponding environmental parameters. A stochastic method combining artificial neural networks and Monte Carlo simulations (ANN–MCS), which can help deal with uncertainties arising from limited data quality and measurement errors in samples, has been used to predict the phytoavailability of copper in contaminated soils based on the variability of soil input parameters (Hattab et al., 2013). A classification and regression tree (CART) has been used in a few studies to identify various sources attributed to soil pollution by heavy metals and persistent organic pollutants (Kubosova et al., 2009; Vega et al., 2009; Zhang et al., 2008). Furthermore, a conditional inference tree (CIT) has been developed to consider different types of predictor variables measured at arbitrary scales and overcome the problems of over-fitting and biased variable selection (Hothorn et al., 2006; Hu and Cheng, 2013). Another stochastic approach, a finite mixture distribution model (FMDM), which describes a pooled population as a single distribution or a mixture of two or more distributions, has been used to estimate the proportion of natural and anthropogenic sources for heavy metals in soils (Hu and Cheng, 2013; Yang and Chang, 2005). A combination of FMDM with geostatistical models provide further insight into identification of pollution sources and high risk areas of heavy metals in soils and particulate matter in atmospheric environment, which generally consists of determining FMDM cutoff values of pollutant concentrations and mapping probabilities of contaminated areas (Chu et al., 2012; Lin et al., 2010).

The study area is located in the northwest of Guangxi Zhuang Autonomous Region (China), and is a metalliferous industrial district with abundant mineral resources such as tin, lead, zinc and antimony. As one of the well-developed karst areas, the shortage of arable land and the low restoration rate of wastelands have made soils even more valuable there (Li et al., 2007). Previous studies have focused on soil pollution around mining sites, areas surrounding smelters (Li et al., 2009; Xiang et al., 2011; Zhang et al., 2012), as well as agricultural soil pollution along riversides due to the mining activities upstream (Liu et al., 2010; Wang et al., 2009; Zhou et al., 2005). However, few studies have been carried out on soil pollution sources and risk mapping in the study area. In this study, an extensive survey was conducted for soil heavy metal contamination in topsoils of this area. The CIT and FMDM models were applied to identify the sources and to estimate the proportions of contributions from natural and anthropogenic sources for heavy metal contamination in the topsoils. Finally, indicator kriging and multivariate indicator kriging were implemented to map the spatial probabilities of contamination calculated based on the thresholds obtained from FMDM model. This work will provide a scientific basis for environmental management of heavy metal pollution in soils.

2. Material and methods

2.1. Study area, sampling and chemical analysis

The study area, including one district and ten counties, is located in the northwestern part of Guangxi on the southern end of the Yunnan–Guizhou Plateau, with a total area of 33,500 km^2 and a population of 3.99 million. The area has a monsoon-influenced humid subtropical climate with an annual average rainfall around 1090–1920 mm which is concentrated between May and August. The annual average temperature is 17 °C–20 °C. It has a widely distributed Karst landform which makes up 67% of the study area (Yang et al., 2011). Generally, the depth of Karst soil is around 100 cm in flat areas and 20–40 cm on slopes (W. et al., 2006; Yang et al., 2011). The dominant wind comes from the east–north-east and the north-east, and to a lesser extent, from the west direction. The study area is reputed to be a town of non-ferrous metals in China and a rare polyparagenetic ore district in the world. As a consequence of the long period of mining activities, high concentrations of heavy metals have been gradually accumulated during the mining exploration, mineral concentration and smelting processes, especially lead, arsenic, mercury, cadmium and chromium, which are widely known for their deleterious effects on the environment and human health.

A total of 597 topsoil samples (0–20 cm depth) were systematically collected from a grid of 5 × 5 km^2 in the four north counties and a grid of 15 × 15 km^2 in the other seven counties (Fig. 1). For each sampling sites, three to five sub-samples were collected and mixed to form one composite sample using a stainless steel auger. Geographic coordinates of the sampling points were recorded using a global positioning system. After the air-drying and manual removal of root and plant materials, the soil samples were sieved to pass a 2-mm mesh. Portions of the soil samples were ground and passed through a 0.1-mm sieve and stored in plastic bags at 4 °C for chemical analysis. The soil samples were then digested with a mixture of HNO_3–$HClO_4$–HF. The total concentrations of Pb, Cd and Cr in the digested solutions were determined by flame atomic absorption spectrometry (Tao, 1995). The concentrations of As and Hg were detected using atomic fluorescence spectrometry after digestion with a H_2SO_4–HNO_3–$HClO_4$ mixture for As and a H_2SO_4–HNO_3–$KMnO_4$ mixture for Hg (Lin et al., 2010).

2.2. Data collection and preparation

The CIT was used to identify the contributing factors associated with the spatial distribution of heavy metals in topsoils of the study area. A subset of available factors affecting the distribution of heavy metals was chosen for the CIT model. The variables used in the study were classified into four categories (See details in Table 1): (1) soil type (Soil.type), which is the only available map related to the original parent materials (Alloway, 2013; Wang et al., 2012); (2) land use type (Landuse), which was expected to affect the redistribution and accumulation of heavy metals in soils (Nicholson et al., 2003); (3) climatic conditions, such as annual average precipitation (Rain), which could remove airborne metal particulates on land surfaces (Garnaud et al., 1999); (4) indicators of anthropogenic activities, including road conditions (i.e. Road.dens and Class.road), industry types within a buffer zone of industrial plants (Type.industry) and a distance from urban areas (Dist.city), which were reported as the most important sources of heavy metal in topsoils (Biasioli et al., 2006; Li et al., 2001).

Eight main soil types were identified in the study area (Fig. 2a): ferric acrisols (ACf), haplic acrisols (ACh), humic acrisols (ACu), haplic alisols (ALh), cumulic anthrosols (ATc), rendzic leptosols (LPk), haplic luvisols (LVh), and dystric regosol (RGd), according to the Harmonized World Soil Database (FAO and ISRIC, 2010). The land use type was compiled from Landsat 5 TM images and finally divided into four categories, including paddy field (PF), dry land (DL), shrub land (SL) and forest land (FL) as shown in Fig. 2b. Annual average precipitation was collected from the hydrological yearbooks (Ministry of Water Resources, 2012) and interpolated using the inverse distance weighted method. Here, road conditions were represented by the total lengths and classes of roads within a 5 km buffer zone surrounding the soil samples. These roads are divided into primary and secondary road classes (Fig. 2c). The types of industrial plants within a 2500 m buffer zone were selected to account for their influences on heavy metal accumulation, based on a research on heavy metal concentrations of topsoils around the antimony–lead smelter of the study area (Xiang et al., 2011). Three types of industries were determined in the field investigation, including metallurgical plants (MP), arsenic chemical plants (AP) and tailing ponds of metallurgical plants (TP) (Fig. 2d).

2.3. Conditional inference tree (CIT)

The CIT was applied to estimate regression relationships between predictor variables and heavy metal concentrations by recursive binary partitioning in a conditional inference framework (Hothorn et al., 2006). In this procedure, an assumption was made that the conditional