Environmental Pollution 214 (2016) 400–409

Contents lists available at ScienceDirect

Environmental Pollution

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

A method for apportionment of natural and anthropogenic contributions to heavy metal loadings in the surface soils across large-scale regions☆

Yuanan Hu [a], Hefa Cheng [b,*]

[a] School of Water Resources and Environment, China University of Geosciences (Beijing), Beijing 100083, China
[b] MOE Laboratory for Earth Surface Processes, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China

ARTICLE INFO

Article history:
Received 27 December 2015
Received in revised form
8 April 2016
Accepted 8 April 2016
Available online 22 April 2016

Keywords:
Sequential extraction
Source identification
Conditional inference tree
Random forest
Soil metal loading
Non-residual fractions

ABSTRACT

Quantification of the contributions from anthropogenic sources to soil heavy metal loadings on regional scales is challenging because of the heterogeneity of soil parent materials and high variability of anthropogenic inputs, especially for the species that are primarily of lithogenic origin. To this end, we developed a novel method for apportioning the contributions of natural and anthropogenic sources by combining sequential extraction and stochastic modeling, and applied it to investigate the heavy metal pollution in the surface soils of the Pearl River Delta (PRD) in southern China. On the average, 45–86% of Zn, Cu, Pb, and Cd were present in the acid soluble, reducible, and oxidizable fractions of the surface soils, while only 12–24% of Ni, Cr, and As were partitioned in these fractions. The anthropogenic contributions to the heavy metals in the non-residual fractions, even the ones dominated by natural sources, could be identified and quantified by conditional inference trees. Combination of sequential extraction, Kriging interpolation, and stochastic modeling reveals that approximately 10, 39, 6.2, 28, 7.1, 15, and 46% of the As, Cd, Cr, Cu, Ni, Pb, and Zn, respectively, in the surface soils of the PRD were contributed by anthropogenic sources. These results were in general agreements with those obtained through subtraction of regional soil metal background from total loadings, and the soil metal inputs through atmospheric deposition as well. In the non-residual fractions of the surface soils, the anthropogenic contributions to As, Cd, Cr, Cu, Ni, Pb, and Zn, were 48, 42, 50, 51, 49, 24, and 70%, respectively.

1. Introduction

Identifying the sources of heavy metals in soils and quantifying the exact contributions from natural and anthropogenic sources can be important for developing appropriate pollution prevention and control regulations. Heavy metals in soils can originate from both natural processes (e.g., volcano eruptions and weathering of rocks) and human activities (e.g., mining and fossil fuel combustion), thus their presence at high levels in the surface soils of a particular area does not necessarily point to pollution by human activities (Hooda, 2010; Shuman, 1991). Source identification and apportionment of heavy metal pollution in surface soils is particularly challenging for large-scale regions because of the high spatial variability of heavy metal contents in surface soils caused by both heterogeneous parent materials and widespread human activities. Multivariate statistical analyses, such as correlation matrix, principal component analysis, and cluster analysis, were widely used to investigate the interrelationships of different heavy metals, which can help distinguish the contributions from natural and anthropogenic sources (Facchinelli et al., 2001; Huang et al., 2015; Mamat et al., 2014). Mapping based on geographical information system (GIS) and spatial analysis are also popular tools to understand the spatial distribution patterns and identify the likely sources of metals in surface soils. In general, these methods can only provide descriptive information and identify the pollution sources roughly. Recently, data mining methods, including decision tree and random forest, have been used to identify the impact factors for soil metal contents, to understand the influence of soil properties and human activities on metal contents, and to quantify the contributions from natural and anthropogenic sources (Hu and Cheng, 2013; Lin et al., 2010; Wang et al., 2015; Zhang et al., 2008).

☆ This paper has been recommended for acceptance by B. Nowack.

* Corresponding author.
E-mail address: hefac@umich.edu (H. Cheng).

http://dx.doi.org/10.1016/j.envpol.2016.04.028
0269-7491/

Although the total contents of heavy metals in soils can provide useful information on the distributions and sources of pollution, the chemical fractionation of them, i.e., the forms or phases in which they occur, can reveal much more details on the mineralogy and chemistry of the soils, as well as their release potentials under various conditions (Tack and Verloo, 1995; Templeton et al., 2000; Tessier et al., 1979). According to Shuman (1991), heavy metals are found in soils in a variety of forms, including (i) soluble form in the soil solution; (ii) ionic form bound at the exchangeable sites of the inorganic constituents; (iii) adsorbed form on the surface of the inorganic constituents; (iv) precipitated/co-precipitated form on the surface of the inorganic constituents; (v) sorbed form bound to the soil organic matter; (vi) components of the secondary minerals; and (vii) components of the primary minerals. The chemical fractionation of heavy metals in soils and sediments is experimentally determined by sequential extraction, in which a series of selective extracting agents are used to remove the heavy metals bound to different fractions or phases (Maiz et al., 1997; Tessier et al., 1979; Usero et al., 1998). Among the several sequential extraction methods that have been established, the modified BCR procedure, in which the metals are separated as the acid soluble, reducible, oxidizable, and residual fractions, has been proven to be a useful and widely accepted method (Favas et al., 2011; Kubova et al., 2008; Kumar et al., 2013; Pueyo et al., 2008; Rauret et al., 2000; Svete et al., 2001; Zhang et al., 2012).

Heavy metals present in the matrices of primary and secondary minerals are gradually released into the more bioavailable and mobile fractions during the weathering of soil parent materials and leaching of soils (Hooda, 2010). Meanwhile, deposition of heavy metals associated with the aerosol particles released from volcanic eruption and forest fires also results in their enrichment in surface soils, primarily in the relatively bioavailable and mobile fractions, while heavy metals introduced by human activities are also expected to be enriched in these fractions. Heavy metals released from anthropogenic sources (e.g., industrial wastewater and vehicular emissions) can bind to different soil phases and components, including organic matter, carbonates, and oxides through a range of processes, such as ion exchange, adsorption, co-precipitation, and complexation, which affect their fractionation, mobility, and bioavailability in soils (Carrillo-Gonzalez et al., 2006; Garcia et al., 2005; Hooda, 2010). In contrast, heavy metals in the residual fraction mainly exist in the lattices of the primary and secondary soil minerals, and are barely subjected to the influence of human activities in general (Shuman, 1991). Thus, chemical fractionation of heavy metals can reveal more detailed information and offer valuable insights on their sources, and the natural and anthropogenic contributions to soil heavy metals can be more evident from evaluating the non-residual fractions compared to the total contents.

The Pearl River Delta (PRD), which is located in Guangdong province of southern China (N 21°48′−22°27′, E 113°03′−114°19′), is one of the world's largest manufacturing bases. Widespread contamination of surface soils by a range of pollutants, including heavy metals, has occurred in the region during the course of rapid industrialization and urbanization over the past three decades. The total contents of major heavy metals (As, Cd, Cr, Cu, Ni, Pb, and Zn) in the surface soils of the PRD were measured, and their sources were identified using a combination of geospatial analysis, multivariate statistics, and stochastic modeling in our previous work (Hu and Cheng, 2013; Hu et al., 2013). However, for As, Cr, and Ni, which are predominantly of lithogenic origin, the relatively small contributions of anthropogenic inputs could not be accurately quantified (Hu and Cheng, 2013).

This study aimed to develop a novel method that can improve the accuracy of source apportionment of soil heavy metals, particularly for the species dominated by the soil parent materials, by combining sequential extraction and stochastic modeling. The main hypothesis was that the contributions from human activities to soil heavy metals could be identified and quantified more accurately in the non-residual fractions, which are contributed by both natural and anthropogenic sources (in contrast, metals in the residual fraction are predominantly of lithogenic origin), using stochastic modeling techniques, such as decision tree and random forest, compared to the analysis based on the total metal contents. The performance of this method was tested by applying it to investigate the heavy metal pollution in the surface soils of the PRD, and the results were validated by comparison with the estimations based on subtraction of regional soil background from soil metal loadings and the soil metal inputs through atmospheric deposition.

2. Materials and method

2.1. Soil samples and chemical analysis

A total of 227 soil samples were collected in the surface layer (0–10 cm) of the PRD region that covered approximately 6.81×10^4 km^2 of land area (Fig. S1), based on a hexagonal sampling scheme, as described in a previous work (Hu et al., 2013). The contents of total organic carbon (TOC) of these soil samples were measured by an elemental analyzer, while their total contents of heavy metals were determined by inductively coupled plasma-mass spectrometry (ICP-MS) after microwave-assisted acid digestion. Sequential extraction of the heavy metals from the soil samples were conducted following the modified BCR procedure. Details of the sample digestion, sequential extraction, and instrumental analysis are presented in the Supplementary Information.

2.2. Data preparation

Before construction of the decision trees, the outliers of the data sets were detected and removed based on interquartile range (IQR), which is the width of the box in the box-and-whisker plot. In statistical analysis, an outlier is the erroneous or “surprising” data that “does not follow the same model as the rest of the data” (Weisberg, 1985). Outliers affected the identification of split factors and the subsequent splitting procedures during the tree construction, and no meaningful impact factors could be identified in many cases when such extreme values were present. In this study, outliers were identified as any observations outside the range of $[Q_1-1.5IQR, Q_3+1.5IQR]$, where Q_1 and Q_3 are the lower quartile and upper quartile, respectively, and $IQR = Q_3 - Q_1$. It is worth noting that some nodes in the decision trees constructed contain extremely high values that prevent further split. Although such data could also be classified as outliers, they were kept in the analysis because some nodes had relatively small populations and over-fitting could result from removal of too many data points. While the metal concentrations did not follow normal or log normal distribution, log transformation of the data was not considered in this study because this would magnify the distribution of low concentration data (presumably “unpolluted” soil samples) but diminish that of the high concentration data (presumably soil samples that had been subjected to significant human influence).

2.3. Stochastic modeling

Conditional inference tree (CIT), which allows unbiased analysis and handles both categorical and continuous variables while overcoming the problems of over-fitting and biased variable selection (Hothorn et al., 2006; Strobl et al., 2007), was employed to