Environmental Pollution 231 (2017) 1188–1200

Contents lists available at ScienceDirect

Environmental Pollution

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

Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review☆

Deyi Hou [a,*], David O'Connor [a], Paul Nathanail [b], Li Tian [c], Yan Ma [d]

[a] School of Environment, Tsinghua University, Beijing, 100084, China
[b] School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK
[c] Department of Urban Planning, School of Architecture, Tsinghua University, Beijing, 100084, China
[d] School of Chemical and Environmental Engineering, China University of Mining & Technology, Beijing 100083, China

A R T I C L E I N F O

Article history:
Received 4 April 2017
Received in revised form
4 July 2017
Accepted 7 July 2017
Available online 19 September 2017

Keywords:
GIS
Kriging
Multivariate statistical analysis
Principal component analysis
Cluster analysis

A B S T R A C T

Heavy metal soil contamination is associated with potential toxicity to humans or ecotoxicity. Scholars have increasingly used a combination of geographical information science (GIS) with geostatistical and multivariate statistical analysis techniques to examine the spatial distribution of heavy metals in soils at a regional scale. A review of such studies showed that most soil sampling programs were based on grid patterns and composite sampling methodologies. Many programs intended to characterize various soil types and land use types. The most often used sampling depth intervals were 0–0.10 m, or 0–0.20 m, below surface; and the sampling densities used ranged from 0.0004 to 6.1 samples per km^2, with a median of 0.4 samples per km^2. The most widely used spatial interpolators were inverse distance weighted interpolation and ordinary kriging; and the most often used multivariate statistical analysis techniques were principal component analysis and cluster analysis. The review also identified several determining and correlating factors in heavy metal distribution in soils, including soil type, soil pH, soil organic matter, land use type, Fe, Al, and heavy metal concentrations. The major natural and anthropogenic sources of heavy metals were found to derive from lithogenic origin, roadway and transportation, atmospheric deposition, wastewater and runoff from industrial and mining facilities, fertilizer application, livestock manure, and sewage sludge. This review argues that the full potential of integrated GIS and multivariate statistical analysis for assessing heavy metal distribution in soils on a regional scale has not yet been fully realized. It is proposed that future research be conducted to map multivariate results in GIS to pinpoint specific anthropogenic sources, to analyze temporal trends in addition to spatial patterns, to optimize modeling parameters, and to expand the use of different multivariate analysis tools beyond principal component analysis (PCA) and cluster analysis (CA).

1. Introduction

Whilst it is acknowledged that there is no authoritative definition of the term 'heavy metals' to be found in the relevant literature (Duffus, 2002), the present study uses the term as a group name for metals and semimetals (metalloids) that have been associated with soil contamination and potential toxicity or ecotoxicity. The heavy metals that have been most intensively studied within the reviewed publications include Pb, Zn, Cu, Ni, Cr, and Cd, listed in descending order of frequency. Heavy metal contamination in soil has become a serious issue globally (Järup, 2003; Jingling et al., 2016; Phoungthong et al., 2016). Harmful amounts of heavy metals can enter the human body from contaminated soil via exposure pathways such as direct or indirect ingestion, inhalation and dermal contact; potentially resulting in human health effects. Heavy metals can also exhibit ecotoxicity leading to inhibited ecological health in addition to bioaccumulation in the food chain. To address this issue, the fate and transport of heavy metals in soil, as well as the remediation of contaminated soils, has been intensively studied (Hou et al., 2016; Ma et al., 2015; Tsang and Lo, 2006; Tsang et al., 2007, 2009). It is also of upmost importance to be able to robustly discern the spatial distribution of heavy metals in soils at the regional scale, in order to enable sound assessment of human and ecological risks, and to implement efficient pollution

☆ This paper has been recommended for acceptance by Dr. Yong Sik Ok.
* Corresponding author.
E-mail address: houdeyi@tsinghua.edu.cn (D. Hou).

http://dx.doi.org/10.1016/j.envpol.2017.07.021
0269-7491/

mitigation actions where required (Maas et al., 2010). Techniques such as geostatistics have an important role to play in this task (Hooker and Nathanail, 2006).

Some particular challenges exist in addressing soil heavy metal contamination: i) heavy metals are non-destructible and will often naturally accumulate in soils rather than attenuate (Maas et al., 2010); ii) they cause a wide range of health effects, and the health risk is complicated by their oxidation state and associated bioavailability differences (Walker et al., 2003); iii) there are many diffusive sources of heavy metal contamination (Nriagu and Pacyna, 1988).

Understanding heavy metal concentrations at the regional scale is of particular relevance to policy makers. Regional soil studies help guide actions in combating pollutant linkages - managing risks rather than molecules. It is important to understand all of the uncertainties regarding contaminant concentration, form, spatial distribution and temporal change. Heavy metals deriving from natural or anthropogenic sources can display widely differing bioavailability levels and, hence, effective intake/dose and, therefore, risk levels. Appreciating these uncertainties is central to designing and implementing risk mitigation strategies, and only focusing on reducing soil concentrations when deemed necessary.

Statistical analysis has been used across various disciplinary boundaries to address soil contamination issues, including geosciences, soil science, atmospheric studies, environmental engineering, chemometrics (Mostert et al., 2010). Historically, most soil investigators have used classical univariate statistics for processing soil data. The use of multivariate statistical analysis was only observed in a small number of studies published in the 1980s (Hopke et al., 1980; Vogt et al., 1987). It was not until the current decade that larger numbers of soil studies were published that applied these techniques (Mostert et al., 2010). In recent years, a growing number of studies have used integrated geographical information systems (GIS) and multivariate analysis for regional soil quality assessmemt (Ali et al., 2016; Huang et al., 2015; Lin et al., 2016; Mihailović et al., 2015; Moore et al., 2016; Zhou et al., 2016). This is partly attributed to the usage of specialist software that can deal with large spatial data-sets presented in GIS. However, many statistical techniques suffer from a failure to recognize the role of spatial correlation. Matheron's theory of regionalized random variables renders invalid many classical statistical approaches and present day guidance often fails to appreciate this (e.g. CLAIRE/CIEH 2009). GIS and GIS-based geostatistics has proved to be a powerful tool in studying soil contamination (Facchinelli et al., 2001; Goovaerts, 1999) (Hooker and Nathanail, 2006) and a particularly powerful useful tool for understanding background levels of heavy metals in soil (Zhou and Xia, 2010).

This paper presents a critical review of 29 field studies conducted in 15 countries, which have examined the distribution of heavy metals in soil by a combination of GIS techniques and multivariate statistical analysis. The review focused on regional, rather than site specific studies. The review discusses methodologies for combining GIS and multivariate statistical analysis, summarizes geochemical and anthropogenic determinants that were found to correlate with soil heavy metal distribution, and proposes critical future research directions in this field.

2. Overview of soil heavy metal spatial distribution assessment

Although this review focusses on the integration of GIS and multivariate statistical analysis it is important to first provide an overview of the principal issues in this field. This section reviews several issues associated with soil heavy metal spatial distribution.

2.1. Sources of heavy metal contamination in soil

Geogenic heavy metals are naturally present in the Earth's crust and surficial soil. The spatial distribution of naturally occurring heavy metals is highly heterogeneous and significantly elevated concentrations may exist in soil at certain localities. Heavy metals in high concentration areas can be distributed to other areas by surface runoff (Herngren et al., 2005), groundwater flow (Mandal and Suzuki, 2002), weathering and atmospheric cycling (e.g. wind, sea salt spray, volcanic eruptions, deposition by rivers) (Nriagu, 1989). Typical anthropogenic sources of heavy metal contamination in urban soils include vehicle exhaust, waste disposal, sewage, industrial emission (Chen et al., 2005; Hou et al., 2012; Kelly et al., 1996; Li et al., 2001; Wei and Yang, 2010). Elevated heavy metal concentrations in rural soils typically derive from impurities in agrochemicals such as pesticide and fertilizer application, irrigation with contaminated water, surface runoff from localized industrial facilities, mineral ore extraction and subsequent waste disposal (Candeias et al., 2014; Li et al., 2014b), road dust, sewage sludge, waste disposal, and livestock manures, and atmospheric deposition (Ke-Lin et al., 2006; Micó et al., 2006; Nicholson et al., 2003; Pagotto et al., 2001).

2.2. Regional versus site specific assessment

Soil heavy metal contamination is usually studied at the regional scale or on a site specific basis. At the regional scale (typically ranging from approximately 10 km^2 to 10,000 km^2), investigations are carried out to establish geochemical background levels (Dung et al., 2013; Esmaeili et al., 2014; Reimann and de Caritat, 2017), source tracking (Li et al., 2014a), and public health protection (Chen et al., 2015). Many regional soil quality studies have been conducted, however, only within the past two decades have researchers applied GIS-based approaches to undertake geochemical interpretation of soil data (Facchinelli et al., 2001; Li et al., 2004). At the site-specific scale (typically ranging from 0.01 km^2 to 10 km^2) investigators typically aim to determine spatial extent, concentration, and the fate and transport of contamination in order to assess risks to human health and ecological systems, and to identify remediation alternatives (USEPA, 1988; Wu et al., 2015). Differences between regional and site specific assessments are also found in sampling depth, method, and density, as discussed in Section 2.3 below.

2.3. Sampling issues

It is imperative to establish a robust and efficient soil sampling regime to achieve the objectives associated with the assessment of soil heavy metal spatial distribution. A number of factors may affect the representativeness and efficiency of sampling programs. Of these, sampling location, depth, and density are the most important concerning data usefulness for multivariate statistical analysis and GIS-based geostatistic analysis.

2.3.1. Choosing sampling locations

For smaller scale site specific soil contamination investigations, sampling locations are often chosen based on historical knowledge and professional judgement, in order to determine the lateral and vertical extent of soil contamination with the least number of sampling locations. For larger scale regional soil surveys, the strategies employed in choosing sampling locations may differ from study to study, and often depend on survey specific objectives. As shown in Table 1, the three most widely used strategies for choosing soil sampling locations are: i) achieving a uniform distribution, thus optimizing the geostatistical fit; ii) representing