Geoderma 230–231 (2014) 9–21

Contents lists available at ScienceDirect

Geoderma

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

Identification of spatial distributions and uncertainties of multiple heavy metal concentrations by using spatial conditioned Latin Hypercube sampling

Yu-Pin Lin [a,*], Wei-Chih Lin [a], Meng-Ying Li [b], Yen-Yu Chen [a], Li-Chi Chiang [c], Yung-Chieh Wang [a]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC
[b] Taiwan GIS Center, 6F-7 Roosevelt Road Sec. 1, Taipei 10092, Taiwan, ROC
[c] Department of Civil and Disaster Prevention Engineering, National United University, Miaoli 36003, Taiwan, ROC

ARTICLE INFO

Article history:
Received 20 November 2012
Received in revised form 15 March 2014
Accepted 17 March 2014
Available online 18 April 2014

Keywords:
Spatial sampling
Conditioned Latin hypercube sampling (cLHS)
Experimental variogram
Sequential indicator simulation (SIS)
Soil pollution

ABSTRACT

This work develops spatial conditioned Latin hypercube sampling (scLHS), a novel advance on conditioned Latin hypercube sampling (cLHS) [Minasny and McBratney, Computers & Geosciences (2006) 1378–1388]. The difference between cLHS and scLHS is that the latter introduces variograms of ancillary variables into the objective function of the optimization procedure that selects sampling locations. The improvement of scLHS was evaluated by applying both scLHS and cLHS to simulated samples of multiple heavy metals (Cr, Cu, Ni, and Zn) in the soil of Changhua County in Taiwan. Simulation was done by using sequential indicator simulation (SIS), which generated 1000 realizations of spatial distributions of the heavy metals, based on existing sample data, to represent the real distributions of heavy metals. The results of the Ripley K analysis show that the locations of samples obtained by both approaches were not significantly spatially segregated. The sampling results show that the declustering of sample locations may dominate the optimization process in cLHS and scLHS. Statistical analysis indicated that compared with cLHS, the means, standard deviations and contamination proportions of scLHS samples captured more efficiently the variability of the SIS realizations. Moreover, the experimental variograms of scLHS samples, especially for small sample sizes, captured the experimental variograms of the SIS realizations more efficiently. Therefore, the use of the scLHS approach is recommended as a novel alternative sampling approach without the need for a reconnaissance survey to increase the efficiency of capturing the spatial structures of soil heavy metals and delineating contaminated sites.

1. Introduction

Sampling involves the selection of a subset of individuals from a total population using various measurements (Pennock et al., 2007). Designing a spatial sampling program for environmental purposes (such as soil heavy-metal remediation) should focus on collecting spatial data of sufficient quality and quantity to support decisions such as identifying a polluted area for remediation (Byrnes, 2008). Such as a spatial sampling program can be undertaken using various sampling approaches, including random, systematic, stratified, ranked set, adaptive cluster, composite, and sequential or nested sampling (Byrnes, 2008; Fortin and Edwards, 2001). Sampling designs vary with the research objectives of a particular study (Pennock et al., 2007). In soil pollution studies, some of them have developed digital soil mapping applications (Kerry and Oliver, 2004, 2007; Lin et al., 2011b; Minasny and McBratney, 2010; Simbahan and Dobermann, 2006).

Sampling schemes determine the variation of environmental properties of interest. As a stratified random procedure, Latin hypercube sampling (LHS) is an efficient means of sampling multivariate distributions (Carré et al., 2007; McKay et al., 1979; Minasny and McBratney, 2006). Minasny and McBratney (2006) proposed the sampling of soil at locations determined by 'conditioned' LHS (cLHS), and their approach has been widely used. However, cLHS does not consider the spatial distribution of the variables (Lin et al., 2011b). Simbahan and Dobermann (2006), on the other hand, improved their sampling approach using multiple environmental variables and experimental variograms (spatial structures) as secondary information. Integrating cLHS with the variogram of auxiliary (environmental) data may improve spatial sampling.

Heavy metals persist in all parts of soil since they cannot be degraded (Adriano et al., 2004; Garau et al., 2007; Lin et al., 2011a). Anthropogenic activities have dominated heavy-metal pollution of soil (Kien et al., 2010; Lin et al., 2011a; Rehman et al., 2008; Wang et al., 2003). In central Taiwan, these activities include those of industrial plants and irrigation systems (Lin et al., 2011a). Chemical and metallurgical industries are the major sources of heavy metals in the environment (Lin et al., 2011a; Suciu et al., 2008), especially in the industrialized areas of

* Corresponding author. Tel./fax: +886 2 3366 3467.
E-mail address: yplin@ntu.edu.tw (Y.-P. Lin).

http://dx.doi.org/10.1016/j.geoderma.2014.03.015

central Taiwan. Information about anthropogenic activities and the environment can be used in soil sampling to increase the efficiency with which the spatial distributions of heavy metals in the soil are captured. If relevant auxiliary (environmental) information is available, then a sampling approach should incorporate spatial relationships among the auxiliary data (Falk et al., 2011). Such auxiliary (environmental) data could be used in cLHS as prior knowledge of the sampling area. Logistic regression (LR) can be used to interpret how soil pollution relates to the ancillary data (Lee et al., 2009; Lin et al., 2011a; Papritz and Reichard, 2009; Twarakavi and Kaluarachchi, 2005).

Geostatistical simulation can generate multiple realizations of a regionalized variable, conditional on a model of spatial variability in the form of variogram parameters (Lin et al., 2011b). Here, we are particularly interested in sequential indicator simulation (SIS). As a non-Gaussian stochastic simulation method, SIS has been used to map heavy-metal concentrations in soil (Lin et al., 2011b; Van Meirvenne and Meklit, 2010; Zhou and Xia, 2010; Zhuang et al., 2009). By incorporating a Monte Carlo approach, SIS generates values on all nodes of a simulation grid, which are visited sequentially on a random path. SIS realizations vary significantly in space with respect to the original data and all previously simulated values (Deutsch and Journel, 1992). Several SIS realizations can represent the real distributions of the heavy metals in an area, and the proportion of heavy-metal contamination.

This study describes the spatial variability of heavy metals in Changhua County, central Taiwan, by comparing the abilities of cLHS and a modified version of cLHS that is referred to herein as spatial cLHS (scLHS). Additionally, the sampling efficiency is evaluated by using both scLHS and cLHS approaches to sample different sample sizes of multiple heavy metals (Cr, Cu, Ni, and Zn) in the soil of Changhua County.

2. Material and methods

In scLHS, a set of multivariate soil samples was determined from selected auxiliary (herein environmental) variables — describing human activities and soil properties — and their experimental variograms, by modifying cLHS. The selected variables were verified by analysis of existing heavy-metal sample data with logistic regressions (LR). Based on the logistic regressions, a binary variable, such as whether the heavy metal concentrations exceeded allowed values (pollution thresholds) was predicted from the environmental variables. The statistical significances of the coefficients in logistic regression models were tested using the Wald chi-squared test. Based on the sample data, SIS was used to generate 1000 realizations of spatial distributions of heavy metals in soil in Changhua County, central Taiwan. These 1000 SIS realizations were then used to verify the efficiencies of cLHS and scLHS for capturing spatial structures (experimental variograms) of heavy metals and their statistical characteristics (mean of concentration, standard deviation of concentration and proportion of contamination) of heavy metals in the study area.

2.1. Study area and soil sampling data

The study area is Changhua County, a major agricultural region in central Taiwan. In the eastern area of the county is Changhua City, while in the western area is Lugang town. Most industrial plants in the study area are metalworking, electroplating, textile or metal surfacing-treatment plants (Figure 1). The study area is 2.33 km × 1.63 km (3.80 km^2), and is centered at 24.05°N and 120.30°E. The industrial plants have been suspected of discharging wastewater into irrigation channels in the study area (Lin et al., 2010, 2011a); as a result, 946 topsoil (0–15 cm) Cr, Cu, Ni, and Zn concentrations were obtained from an investigation of heavy metals in soil, undertaken by the Environmental Protection Administration of Taiwan from February to August 2002 (Figure 1). The sampling density was approximately one sample per 1.45 ha and sampling was performed on the irregularly shaped farmland in the study area. The coordinates of the sampling locations were recorded by GPS. Approximately 1 kg of soil was collected at each location using a stainless steel spade and a plastic scoop. Soil

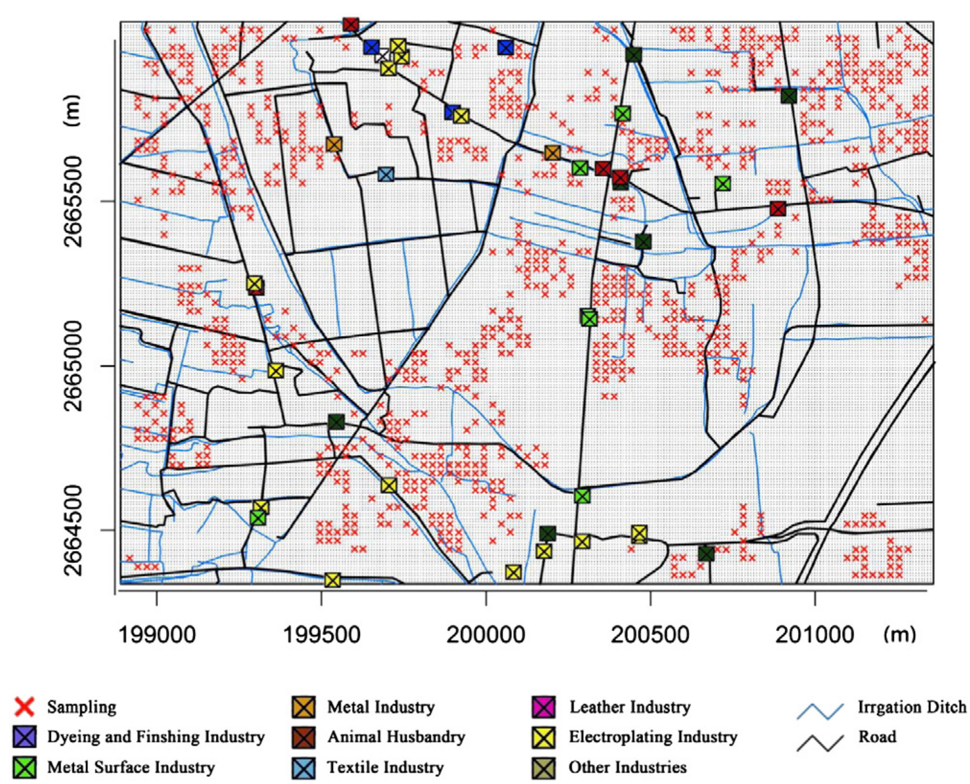

Fig. 1. Study area, soil sampling data and industrial factories.