

# Modeling spatial uncertainty of heavy metal content in soil by conditional Latin hypercube sampling and geostatistical simulation

Yu-Pin Lin · Hone-Jay Chu · Yu-Long Huang ·  
Bai-You Cheng · Tsun-Kuo Chang

Received: 14 September 2009 / Accepted: 12 March 2010  
© Springer-Verlag 2010

**Abstract** This study proposes the method of simulating spatial patterns and quantifying the uncertainty in multivariate distribution of heavy metals (Cr, Cu, Ni, and Zn) by sequential indicator simulation (SIS) combined with conditional Latin hypercube sampling (cLHS) in Changhua County, Taiwan. The cLHS is used for a sampling then for SIS mapping and assessing uncertainties of heavy metal concentrations. The indicator variogram results indicate that the 700 cLHS samples replicate statistical multivariate distribution and spatial structure of the 1,082 samples. Moreover, the SIS realizations based on 700 cLHS samples are more conservative and reliable than those based on 1,082 samples for delineating soil contamination by all heavy metals with the exception of Zn. Given adequate sampling, soil contamination simulation provides sufficient information for delineating contaminated areas and planning environmental management.

**Keywords** Heavy metals · Spatial uncertainty · Conditional Latin hypercube sampling · Conditional simulation · Pollution

---

Y.-P. Lin · H.-J. Chu (✉) · Y.-L. Huang · B.-Y. Cheng ·  
T.-K. Chang

Department of Bioenvironmental System Engineering, National  
Taiwan University, 1, sec. 4, Roosevelt Rd, Da-an District,  
Taipei 10617, Taiwan, ROC  
e-mail: honejaychu@gmail.com

Y.-P. Lin  
e-mail: yplin@ntu.edu.tw

Y.-L. Huang  
e-mail: morris0109@hotmail.com

B.-Y. Cheng  
e-mail: d91622007@ntu.edu.tw

T.-K. Chang  
e-mail: tknchang@ntu.edu.tw

## Introduction

Soil sampling data, although sparsely distributed across the sites cannot provide all suitable information for risk assessment and environmental management. The collected data may cause significant uncertainty, due to extremely complicated spatial patterns or errors in measuring the characteristics of the investigated pollution sources. Therefore, when delineating contaminated areas and assessing risk of soil pollution during the decision-making process, uncertainty assessment of sampling data and unsampled locations is essential (Goovaerts 2001; Van Meirvenne and Goovaerts 2001; Amini 2005; Schnabel 2004; Lin 2002; Hassan and Atkins 2007; Lin 2008). Using adequate samples and mapping techniques to reliably delineate soil contamination by heavy metals can improve the efficiency of environmental decision-making and action. Reliable data analysis of spatially distributed data requires appropriate statistical tools and sampling strategies (Fortin and Edwards 2001; Lin et al. 2008). Spatial sampling schemes have been developed to determine the sampling schemes that cover the variation of environmental variables in a given area (Minasny and McBratney 2006). Soil sampling and mapping potentially contamination areas are often costly and time-consuming, especially if expensive multiple laboratory analyses are required or if the investigated area is large (Zhao et al. 2008). Moreover, implementing efficient sampling methods to understand the spatial distribution of heavy metals is essential.

Probability theory and geostatistics provide methodologies for modeling data and process uncertainty and then propagating that uncertainty all the way to a stochastic conclusion (Journel 1996). The stochastic simulation can estimate either the mean value of heavy metal concentrations or the probability of exceeding a given threshold level.

The geostatistical simulation builds the local probability distribution of a regionalized variable for each unsampled location. The simulated value is then randomly drawn from the probability distribution. The conditional cumulative distribution function (ccdf) can be modeled using either parametric (e.g., multivariate Gaussian distribution) or non-parametric (indicator) approaches (Goovaerts 2001). Unlike classical interpolation techniques, geostatistical techniques, such as stochastic conditional simulation can be applied to generate multiple realizations, including an error component (Lin et al. 2008). Recently, stochastic simulation techniques have been applied in delineating soil pollution areas and assessing risk of soil pollution in studies such as Deutsch and Cockerham (1994); Saito and Goovaerts (2000); Goovaerts (2001); Lin et al. (2001); Franco et al. (2006); Bourennane et al. (2007). Furthermore, uncertainty assessment is basically performed using stochastic simulation algorithms that are becoming more common in soil science (Pachepsky and Acock 1998; Goovaerts 2001; Van Meirvenne and Goovaerts 2001; Castrignano and Buttafuoco 2004; Zhao et al. 2005; Bourennane et al. 2007; Cherubini et al. 2009). As a result, the realizations match the sample statistics, and the conditioning data provide a visual and quantitative measure of local uncertainty and spatial uncertainty (Goovaerts 1996). However, to avoid long running times of hundreds or thousands of simulations, it is necessary to generate a relatively small set of conditional realizations capturing most of the variability. To reduce the number of conditional realizations needed to capture the spatial uncertainty, we introduced an effective sampling method (i.e. Latin hypercube sampling) into a geostatistical stochastic simulation algorithm. Latin hypercube sampling (LHS) is a stratified random procedure that efficiently samples variables from their multivariate distributions (McKay et al. 1979; Iman and Conover 1980; Xu et al. 2005; Minasny and McBratney 2006; Carre et al. 2007; Lin et al. 2008). For sampling existing data, traditional LHS cannot be directly applied to multivariate distribution because the samples selected by LHS may not exist in the real world (Minasny and McBratney 2006), particularly for soil sampling data. Conditional LHS (cLHS), which is based on the empirical distribution of original data, provides full coverage of each variable by maximally stratifying the marginal distribution and ensuring a good spread of sampling points (Minasny and McBratney 2006).

This study mapped and assessed spatial patterns and uncertainties of heavy metals (Cr, Cu, Ni, and Zn) using cLHS and stochastic conditional simulation (i.e. sequential indicator simulation (SIS)) in Changhua County, central Taiwan. The SIS was applied to map spatial patterns based on the set of cLHS samples. Finally, the local uncertainty and spatial uncertainty were determined on the basis of SIS realizations with regulatory thresholds of the heavy metals.

The realizations can be used to identify heavy metal contaminated areas and associated uncertainty by setting a given critical probability. Briefly, this study utilized cLHS to obtain 100, 300, 500, 700, and 900 samples from existing 1,082 heavy metal samples in the study area. Experimental indicator variograms of the selected samples were then calculated and compared with the experimental indicator variograms of the 1,082 samples based on same cut-off values (the 25th, 50th, and 75th percentiles of Cr, Cu, Ni, and Zn of 1,082 samples). The SIS was applied to simulate 1,000 realizations of heavy metals in 7,488 grid cells (25 × 25 m). Moreover, the realizations of heavy metal concentrations obtained by SIS with regulatory thresholds were mapped the spatial probability of contaminated areas and the uncertainties of soil contaminations.

## Materials and methods

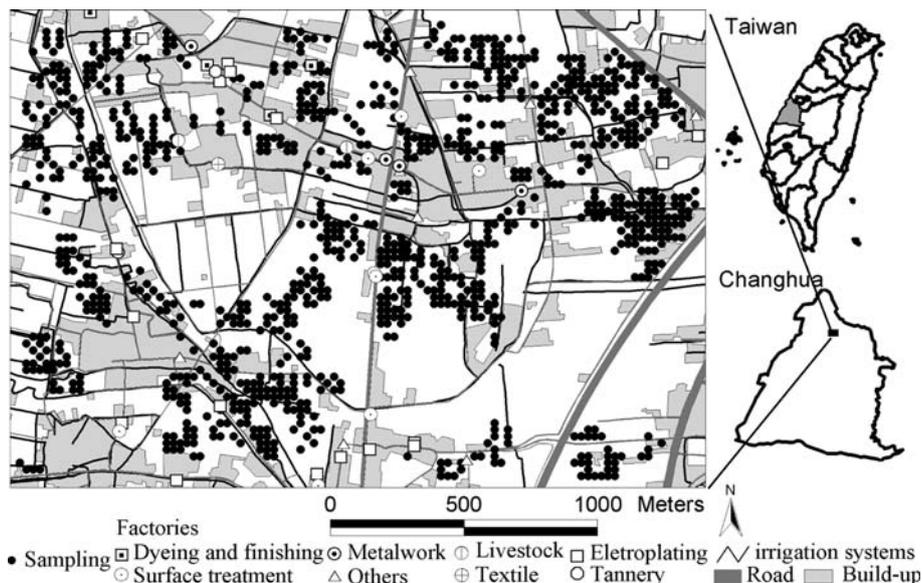
### Study area and soil sampling data

The study area is Changhua County which is a critical agricultural region in Taiwan. The east area is the Changhua city and the west one is the Lugang town. Most industrial plants in the study area comprise metalwork, electroplating, textile, and metal surface treatment industries (Fig. 1). The industrial plants have been suspected of discharging wastewater into irrigation channels in this study area (Lin et al. 2002a, b; 2010). In this study, the data of 1,802 topsoil (0–15 cm) samples' Cr, Cu, Ni, and Zn concentration were obtained from the soil heavy metal investigation project done by the Environmental Protection Administration (EPA) of Taiwan, between February and August 2002, and sampling sites as shown in Fig. 1. The sampling density was about 1 sample per 1.45 ha based on the irregular shape of farmland in the study area and the coordinates of the sampling locations were recorded by GPS. About 1 kg of each soil sample was collected using a stainless steel spade and a plastic scoop, and stored in a plastic food bag. After air drying at room temperature, 3 g soil samples was disaggregated, sieved to 0.85 mm (20 mesh) and ground to a fine 0.15 mm (100 mesh) powder. Each 3 g milled sample was then digested for 2 h at room temperature with 7 mL HNO<sub>3</sub> and 21 mL HCl (acqua regia 1:3) to slowly oxidize organic matter in the soil. The digest was filtered before analyzing the levels of Cr, Ni, Cu, and Zn in the sample which was determined by inductively coupled plasma-optical emission spectrometers (ICP-OES).

### Conditional Latin hypercube sampling

The cLHS procedure represents an optimization problem: given  $N$  sites with data ( $Z$ ), select  $n$  sample sites ( $n \ll N$ )

**Fig. 1** Locations of sampling sites, the factories, and irrigation systems in Changhua County of Taiwan



such that the sampled sites form a Latin hypercube. For  $k$  continuous variables, each component of  $Z$  (size  $N \times k$ ) is divided into  $n$  (sample size) equally probable strata based on their distributions, and  $z$  (size  $n \times k$ ) is a sub-sample of  $Z$ . The procedures of the cLHS algorithm (Minasny and McBratney 2006) are the following.

1. Divide the quantile distribution of  $Z$  into  $n$  strata, and calculate the quantiles for each variable  $q_j^i, \dots, q_j^{n+1}$ . Calculate the correlation matrix  $C$  of  $Z$ .
2. Pick  $n$  random samples from  $N$ ; calculate the correlation matrix of  $T$  of  $Z$ .
3. Calculate the objective function. The overall objective function is

$$O = w_1 O_1 + w_2 O_2 + w_3 O_3, \tag{1}$$

where  $O_1, O_2, O_3$  are different components of objective function to take into account continuous variables, categorical variables, and correlation of the elements, respectively and  $w$  is the weight given to each component of the objective function. For general applications, all  $w$  are set to 1 for all components of the objective function.

- (a) The components of objective function are defined as:

$$O_1 = \sum_{i=1}^n \sum_{j=1}^k \left| \eta(q_j^i \leq z_j \leq q_j^{i+1}) - 1 \right| \tag{2}$$

where  $\eta(q_j^i \leq z_j \leq q_j^{i+1})$  is the number of  $z_j$  that falls between quantiles  $q_j^i$  and  $q_j^{i+1}$

- (b) For categorical data, the objective function is to match the probability distribution for each class of

$$O_2 = \sum_{j=1}^c \left| \frac{\eta'(z_j)}{n} - k_j \right| \tag{3}$$

where  $\eta'(z_j)$  is the number of  $z$  that belongs to class  $j$  in sampled data, and  $k_j$  is the proportion of class  $j$  in  $Z$ .

- (c) To ensure that the correlation of the sampled variables will replicate the original data, component is added to the objective function, which is defined as:

$$O_3 = \sum_{i=1}^k \sum_{j=1}^k |c_{ij} - t_{ij}| \tag{4}$$

where  $c$  is the element of  $C$ , the correlation matrix of  $Z$ , and  $t$  is the equivalent element of  $T$ , the correlation matrix of  $z$ .

4. Perform an annealing schedule:  $M = \exp[-\Delta O/T]$ , where  $\Delta O$  is the change in the objective function, and  $T$  is a cooling temperature (between 0 and 1), which is decreased by a factor  $d$  during each iteration.
5. Generate a uniform random number between 0 and 1. If  $rand. < M$ , accept the new values; otherwise, discard changes.
6. Try to perform changes: generate a uniform random number  $rand$ . If  $rand. < P$ , pick a sample randomly from  $z$  and swap it with a random site from unsampled sites  $r$ . Otherwise, remove the sample from  $z$  that has the largest  $\eta(q_j^i \leq z_j \leq q_j^{i+1})$  and replace it with a random site from unsampled sites  $r$ .
7. Go to step 3. Repeat steps 3–7 until the objective function value falls beyond a given stop criterion or a specified number of iterations.

Sequential indicator simulation (SIS)

In the sequential indicator simulation algorithm, modeling of the  $N$ -point conditional cumulative distribution function (ccdf) is a sequence of  $N$  univariate ccdfs at each node (grid cell) along a random path (Kyriakidis 2001; Lin et al. 2009). The sequential indicator simulation algorithm requires the following steps (Deutsch and Journel 1992; Kyriakidis 2001; Lin et al. 2009):

1. Define a random path that visits each location of the domain once, in which all nodes  $\{x_i, i = 1, \dots, N\}$  discretizing the domain of interest domain. A random visiting sequence ensures that no spatial continuity artifact is introduced into the simulation by a specific path visiting  $N$  nodes.
2. At the first visited  $N$  nodes  $x_1$ :
  - (a) Model, using either a parametric or nonparametric approach, the local ccdf of  $Z(x_1)$  conditional on  $n$  original data  $\{Z(x_\alpha), \alpha = 1, \dots, n\}$ :
 
$$F_Z(x_1; z_1 | (n)) = \text{prob}\{Z(x_1) \leq z_1 | (n)\} \quad (5)$$
  - (b) Generate, via the Monte Carlo drawing relation, a simulated value  $z^{(l)}(x_1)$  from this ccdf  $F_Z(x_1 : z_1 | (n))$ , and add it to the conditioning data set, now of dimension  $n + 1$ , to be used for all subsequent local ccdf determinations.
3. At the  $i$ th node  $x_i$  along the random path:
  - (a) Model the local ccdf of  $Z(x_i)$  conditional on  $n$  original data and the  $i - 1$  near previously simulated values  $\{z^{(l)}(x_j), j = 1, \dots, i - 1\}$ :
 
$$F_Z(x_i; z_i | (n + i - 1)) = \text{prob}\{Z(x_i) \leq z_i | (n + i - 1)\} \quad (6)$$
  - (b) Generate a simulated value  $z^{(l)}(x_i)$  from this ccdf, and add it to the conditioning data set, now of dimension  $n + i$ .
4. Repeat step 3 until all  $N$  nodes along the random path are visited.

In the SIS, the indicator kriging estimator is used to model the prior ccdf at each unsampled location (Juang et al. 2004). Since modeling the prior ccdf at each unsampled location should use previously simulated values at other sampled locations, the simulated values for all unsampled locations are referred to as a joint realization (Goovaerts 1996; Juang et al. 2004). In this study the cutoff values for each soil heavy metal are the 25th, 50th, and 75th percentiles.

Indicator kriging estimates the probability of exceeding a specific threshold value  $z_k$  at a given location (Lin et al.

2002a). In indicator kriging, the data  $[z(x)]$  are transformed into an indicator as follows:

$$i(x, z_k) = \begin{cases} 1, & \text{if } z(x) \leq z_k \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

At an unsampled location ( $x_0$ ), probability of  $z(x_0) \leq z_k$  can be estimated using a linear combination of neighboring indicator variables. This ordinary indicator kriging estimator is

$$\text{Prob}[z(x_0) \leq z_k / (n)]^* = \sum_{\alpha=1}^n \lambda_\alpha i(x_\alpha; z_k) \quad (8)$$

where  $i(x_\alpha; z_k)$  represents the indicator values at  $x_\alpha$ ;  $\alpha = 1, \dots, n$ ;  $\lambda_\alpha$  is the kriging weight of  $i(x_\alpha; z_k)$  determined by solving the following kriging system.

$$\sum_{\beta=1}^n \lambda_\beta \gamma_i(x_\alpha - x_\beta; z_k) + \mu = \gamma_i(x_\alpha - x_0; z_k) \quad (9)$$

$$\sum_{\beta=1}^n \lambda_\beta = 1 \quad (10)$$

where  $\mu$  is the Lagrange multiplier;  $\gamma_i(x_\alpha - x_\beta; z_k)$  is the indicator variogram between indicator variables at the  $\alpha$ th and  $\beta$ th sampling points;  $\gamma_i(x_\alpha - x_0; z_k)$  is the variogram between the indicator variables and  $\alpha = 1, \dots, n$ .

Local and spatial uncertainty

Uncertainty of soil contamination with heavy metals at a single location ( $x'$ ) can be modeled by a probability model (Goovaerts 1999; Juang et al. 2004; Zhao et al. 2005; Cherubini 2009). Therefore, the probability of a soil heavy metal at  $x'$  exceeding the regulatory threshold ( $z_r$ ) can be denoted by  $\text{prob}[z(x') > z_r]$ :

$$\text{prob}[z(x') > z_r] = n(x') / 1,000 \quad (11)$$

where  $n(x')$  is the number of realizations if  $z(x')$  is higher than the threshold in the 1,000 realizations. In Taiwan, the regulatory thresholds ( $z_r$ ) for heavy metal Cr, Cu, Ni, and Zn concentrations in soil are 250, 200, 200, and 600 mg kg<sup>-1</sup>, respectively. Furthermore, the multi-location uncertainty, which is the jointly prevailing uncertainty at several specific locations, can be used to assess the reliability of delineation based on the probability  $\text{prob}[z(x') > z_r]$  (Juang et al. 2004). Therefore, uncertainty when mapping heavy metals at several locations simultaneously is spatial uncertainty, which can be determined by  $\text{prob}[z(x') > z_r] \geq P_c$ , where  $P_c$  is the critical probability (Juang et al. 2004; Zhao et al. 2005). The area is assumed with  $m$  locations ( $x'_1, x'_2, \dots, x'_m$ ). The joint probability ( $P_j$ ) of heavy metals in  $m$  locations of the

area exceeding threshold ( $z_r$ ) can be written as follows (Juang et al. 2004; Zhao et al. 2005):

$$P_j = \text{prob}[z(x'_1) > z_{r1}, z(x'_2) > z_{r2}, \dots, z(x'_m) > z_{rm}] = n(x'_1, x'_2, \dots, x'_m) / 1,000 \quad (12)$$

where 1,000 is the number of simulation, and  $n(x'_1, x'_2, \dots, x'_m)$  is the number of realizations in which all simulated heavy metal concentrations of  $m$  location in the area are greater than the threshold ( $z_r$ ) in 1,000 realizations.

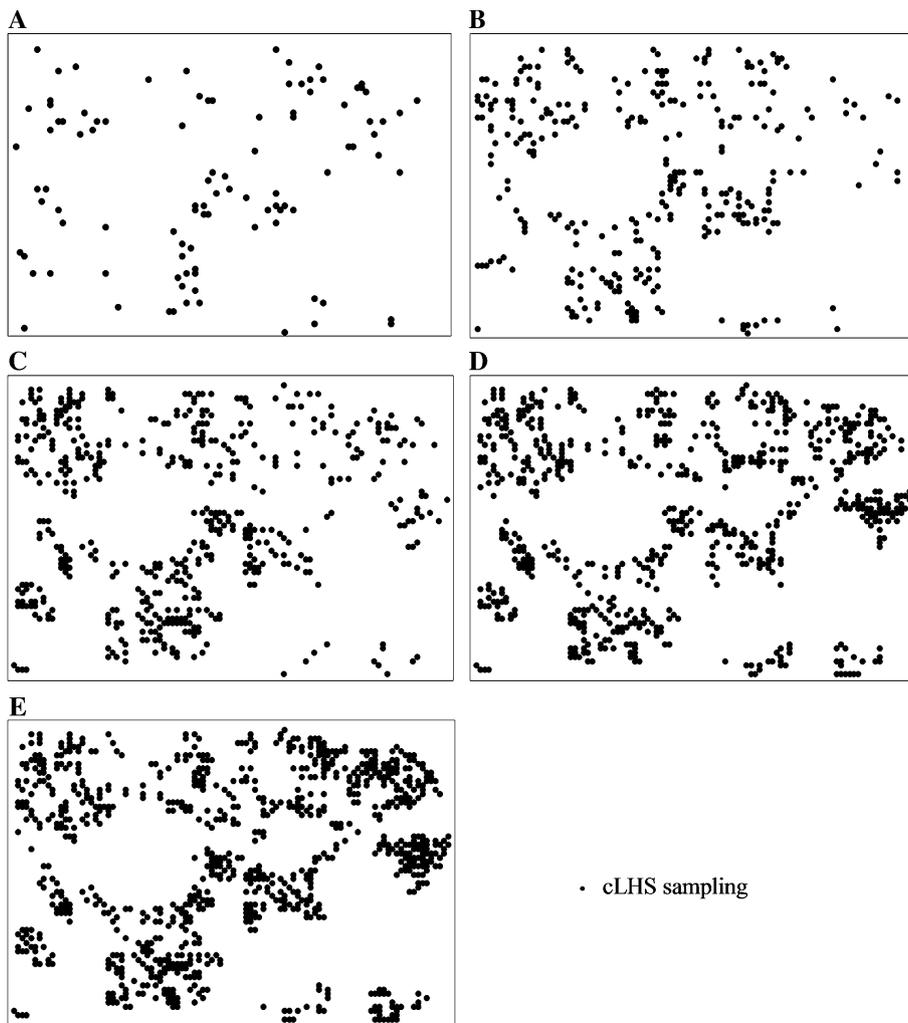
### Results and discussion

#### Statistics of the sampling data

Figure 2 shows the locations of the 100, 300, 500, 700, and 900 samples selected from the 1,082 original samples of Cr, Cu, Ni, and Zn by cLHS. Table 1 summarizes the descriptive statistics of Cr, Cu, Ni, and Zn for 100, 300, 500, 700, and 900 samples selected by cLHS from the

original samples. Descriptive statistics are used to describe the main features of samples. For example, the standard deviation represents that captures the variability of the sample values, the skewness is a measure of symmetry, the kurtosis is a measure of the peakedness of a probability distribution, Q25 is the first quartile, and Q75 is the third quartile. The results show that the sub-samples obtained by cLHS adequately represent 1,082 samples. As the number of cLHS samples exceeds 700 (64.7% of 1,082 samples), the distribution of studied heavy metals becomes similar to that of the original distribution of 1,082 samples, especially as far as the skewness and kurtosis are concerned. Moreover, the descriptive statistical results illustrate that cLHS is an efficient sampling method to catch multivariate distributions (Minasny and McBratney 2006; Lin et al. 2008). Because multiple heavy metals are sampled, considering the multivariate distribution of heavy metals by cLHS is useful for planning and assessing before taking further environmental action, such as monitoring, risk assessment, and remediation (Xu et al. 2005).

**Fig. 2** Locations of cLHS sampling sites in **a** 100, **b** 300, **c** 500, **d** 700, and **e** 900 samples

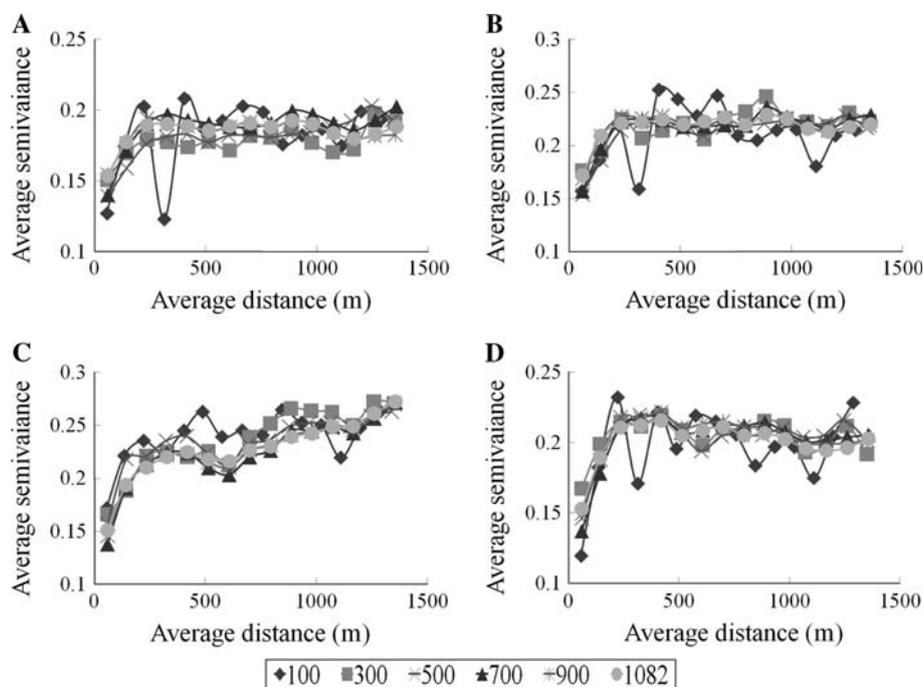


**Table 1** Descriptive statistics for the original samples and the cLHS samples

	No.	Mean	Med	SD	Min	Max	Q25	Q75	Skewness	Kurtosis
Cr	100	207.8	129.7	208.7	24.0	1,270.0	73.0	260.5	2.27	6.54
	300	213.9	127.4	266.6	22.6	3,037.0	76.1	240.9	5.27	45.13
	500	201.3	129.0	191.1	26.4	1,070.0	74.6	247.0	2.02	4.24
	700	210.6	123.8	235.7	22.6	3,070.0	72.6	248.1	4.01	31.62
	900	205.7	125.8	221.3	22.6	3,070.0	71.4	247.0	3.94	32.51
	1,082	205.4	126.4	217.7	22.6	3,070.0	73.8	247.2	3.81	31.22
Cu	100	201.4	122.0	200.1	15.0	1,210.0	71.4	269.0	2.29	6.78
	300	215.1	123.5	299.3	11.0	3,810.0	72.4	256.0	6.74	60.99
	500	198.7	120.0	203.7	11.0	1,380.0	73.3	260.0	2.50	7.90
	700	213.7	122.0	259.8	11.0	3,810.0	68.2	273.0	5.22	54.65
	900	207.0	122.3	244.2	15.0	3,810.0	69.7	260.0	5.18	55.87
	1,082	206.4	123.0	236.3	11.0	3,810.0	73.2	260.0	4.98	53.67
Ni	100	287.3	197.0	252.4	30.7	1,290.0	105.0	388.6	1.57	2.32
	300	284.7	197.0	308.1	22.5	4,020.0	112.0	367.7	6.38	70.41
	500	276.3	198.0	237.0	22.5	1,605.0	112.0	366.0	1.77	3.82
	700	297.6	205.0	292.6	22.5	4,020.0	109.0	398.0	4.09	38.52
	900	297.0	209.0	278.4	22.6	4,020.0	116.3	402.0	3.76	36.19
	1,082	296.9	209.0	274.2	22.5	4,020.0	121.0	392.5	3.56	32.79
Zn	100	595.7	370.0	644.3	82.4	4,540.0	219.0	682.0	3.23	14.13
	300	542.1	382.0	499.3	60.5	3,650.0	218.0	647.0	2.36	7.49
	500	539.8	395.0	473.5	60.5	3,350.0	221.5	654.0	2.24	6.63
	700	562.0	370.0	552.0	60.5	4,540.0	212.0	655.0	2.73	10.60
	900	559.1	368.0	555.1	62.2	4,540.0	207.1	665.0	2.74	10.41
	1,082	553.3	368.0	534.9	60.5	4,540.0	216.0	654.0	2.71	10.41

*Med* median, *Min* minimum, *Max* maximum, *Q25* the first quartile, *Q75* the third quartile

**Fig. 3** The indicator variogram of heavy metals: **a** Cr, **b** Cu, **c** Ni, **d** Zn based on 100, 300, 500, 700, 900, and 1,082 samples



Indicator variogram of sampling data

In this study, experimental indicator variograms of sub-samples were constructed using the same lag interval in GS + software to compare the spatial structures between the original soil samples and cLHS samples. Figure 3 shows the experimental indicator variograms of 100, 300, 500, 700, and 900 cLHS samples of Cr, Cu, Ni, and Zn, respectively. The experimental indicator variograms of 100, 300, and 500 sub-samples of Cr and Cu underestimate those of the original samples. The experimental indicator variograms of 100 sub-samples of Zn underestimate those of the original samples. Conversely, the experimental indicator variograms of 100 and 300 sub-samples of Ni overestimate those of the original samples. These experimental indicator variograms reveal that as the number of samples increases from 700 to 1,082, the ability of experimental indicator variograms to capture the spatial structure of the original data increases. Table 2 demonstrates indicator variogram models for the 25th, 50th, and 75th percentiles of heavy metals in 700, 900, and 1,082 samples. These indicator variography results also show that the cLHS approach can simultaneously select samples from multivariate distributions of heavy metals in soil to capture spatial structures of all heavy metals. Finally, the statistical and indicator variogram analyses of sub-samples illustrate that the cLHS approach can be applied to capture the spatial structures of multiple heavy metals from existing samples for further monitoring and risk assessment. In this study, the effective sampling does not lose information in basic statistics and spatial structure of the original data (Xu et al. 2005; Minasny and McBratney 2006).

Simulated spatial patterns of soil heavy metal concentration

In this study, SIS realizations are performed based on the indicator variogram models for the 25th, 50th, and 75th percentiles of the sample distribution (Table 2) of 500, 700, 900 cLHS samples and original samples for Cr, Cu, Ni, and Zn in the study area. Figures 4a–d show the SIS maps, respectively, of averages of 1,000 realizations of Cr, Cu, Ni, and Zn in 7,488 grid cells in the study area using the original data (1,082 samples). The spatial pattern of Cr shown in Fig. 4 reveals high concentrations near industrial plants, particularly the electroplating, dyeing and finishing, and metalworking plants (Lin et al. 2010). The spatial pattern also reveals high Cr concentrations in the northern, central, eastern, and some southwestern areas of the study area. The areas with high concentrations of Cu are in the central and eastern parts of the study area in the vicinity of the industrial plants, such as the surface treatment and metalworking plants (Fig. 4). Nickel concentrations are

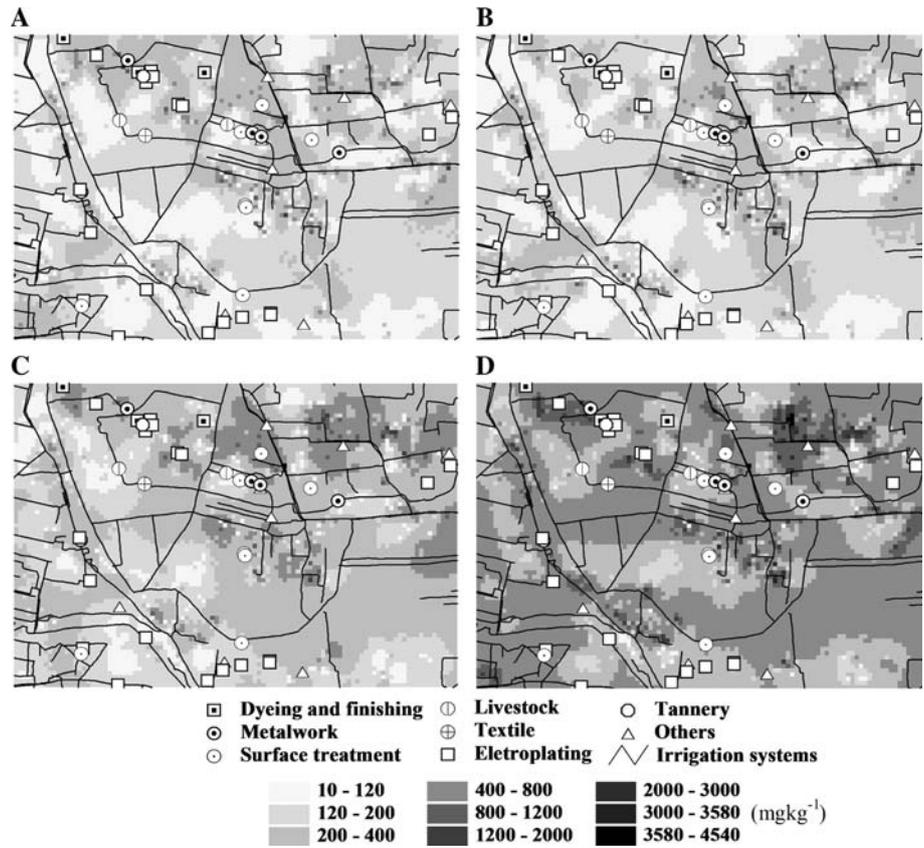
**Table 2** Indicator variogram models for the 25th, 50th, and 75th percentiles of heavy metals in 1,082, 900 and 700 samples

No.	Heavy metal	Model	Parameters			RSS	$r^2$
			C0	C0 + C	R (m)		
1,082	Cr	25% Exp.	0.020	0.184	216	1.730E-03	0.722
		50% Exp.	0.026	0.247	171	1.202E-03	0.807
		75% Exp.	0.025	0.190	120	2.075E-04	0.852
	Cu	25% Exp.	0.017	0.184	240	2.008E-03	0.737
		50% Exp.	0.025	0.247	186	7.016E-04	0.899
		75% Exp.	0.024	0.190	108	5.293E-04	0.663
	Ni	25% Exp.	0.015	0.179	222	2.614E-03	0.634
		50% Exp.	0.022	0.237	228	3.608E-03	0.671
		75% Exp.	0.018	0.183	159	5.723E-04	0.805
Zn	25% Exp.	0.024	0.190	222	1.464E-03	0.768	
	50% Exp.	0.028	0.250	171	3.795E-04	0.936	
	75% Exp.	0.021	0.189	144	8.077E-03	0.710	
900	Cr	25% Exp.	0.019	0.195	222	1.854E-03	0.747
		50% Exp.	0.026	0.247	177	1.359E-03	0.800
		75% Exp.	0.024	0.019	107	2.948E-04	0.794
	Cu	25% Exp.	0.019	0.184	237	1.816E-03	0.747
		50% Exp.	0.024	0.246	198	9.480E-04	0.882
		75% Exp.	0.022	0.191	120	6.752E-07	0.637
	Ni	25% Exp.	0.013	0.178	219	2.211E-07	0.669
		50% Exp.	0.124	0.219	260	7.409E-03	0.682
		75% Exp.	0.016	0.182	156	4.957E-04	0.825
Zn	25% Exp.	0.024	0.190	225	1.503E-03	0.760	
	50% Exp.	0.027	0.250	183	4.654E-04	0.934	
	75% Exp.	0.228	0.189	147	1.109E-03	0.654	
700	Cr	25% Exp.	0.018	0.190	177	1.211E-03	0.759
		50% Exp.	0.024	0.247	177	1.323E-03	0.824
		75% Exp.	0.025	0.197	165	3.978E-04	0.883
	Cu	25% Exp.	0.020	0.197	198	1.961E-03	0.709
		50% Exp.	0.025	0.247	180	8.824E-04	0.873
		75% Exp.	0.025	0.201	153	5.656E-04	0.818
	Ni	25% Exp.	0.017	0.193	183	2.418E-03	0.621
		50% Exp.	0.049	0.237	278	3.548E-03	0.735
		75% Exp.	0.022	0.186	237	9.884E-04	0.834
Zn	25% Exp.	0.023	0.191	177	9.684E-04	0.778	
	50% Exp.	0.026	0.250	168	5.214E-04	0.918	
	75% Exp.	0.023	0.193	207	7.342E-04	0.863	

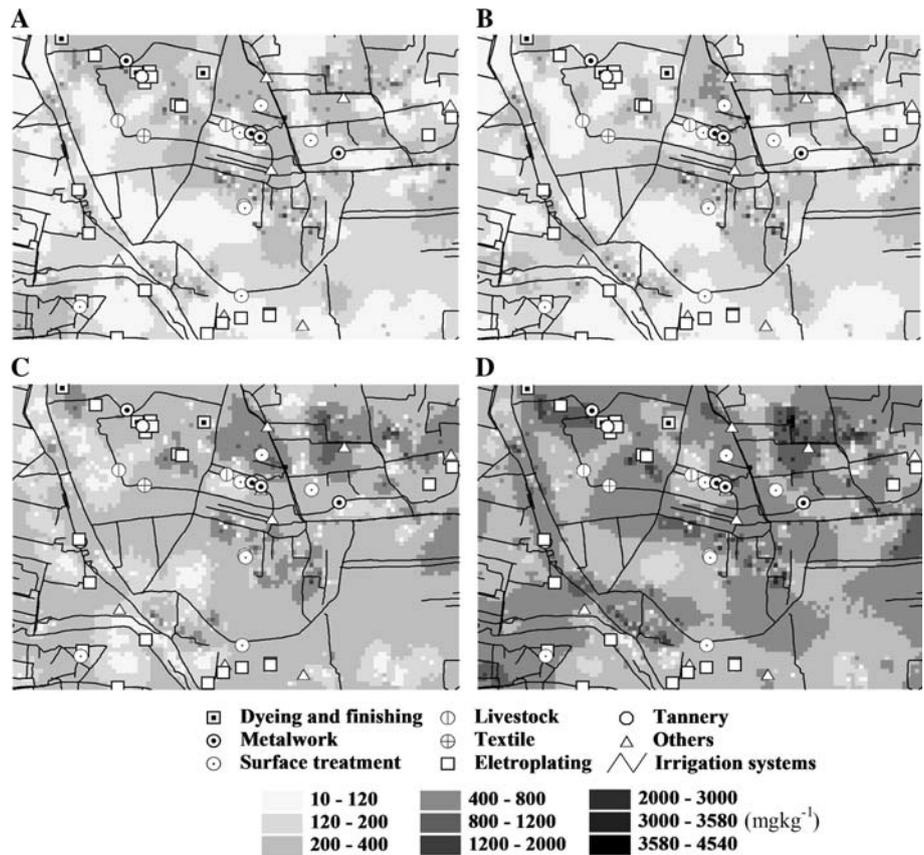
Exp. exponential model, C0 Nugget, C0 + C Sill, R range, RSS residual sums of squares

distributed throughout the studied area, except in the southwest area (Fig. 4). The areas with high concentrations of Zn are close to the industrial plants, such as the electroplating, dyeing and finishing, and metalworking plants in the northwest and near the surface treatment and electroplating plants in the southwest (Fig. 4). Most areas with high Cr, Cu, Ni, and Zn concentrations are located near

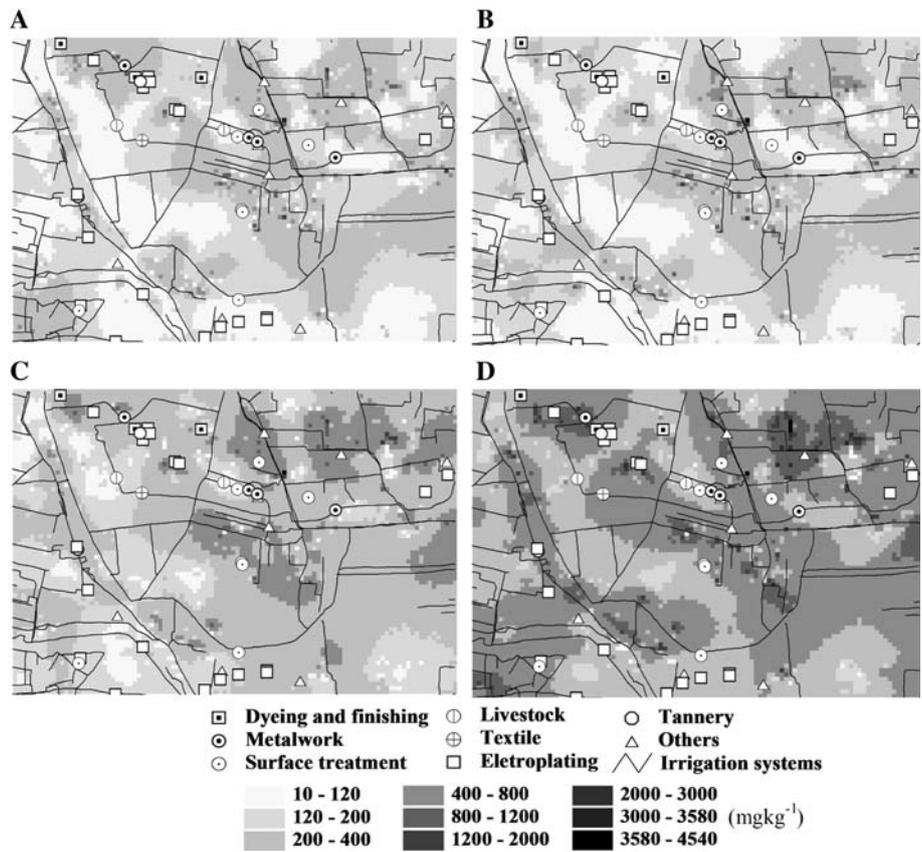
**Fig. 4** The contour map of average concentration in 1,000 realizations based on the 1,082 samples: **a** Cr, **b** Cu, **c** Ni, **d** Zn



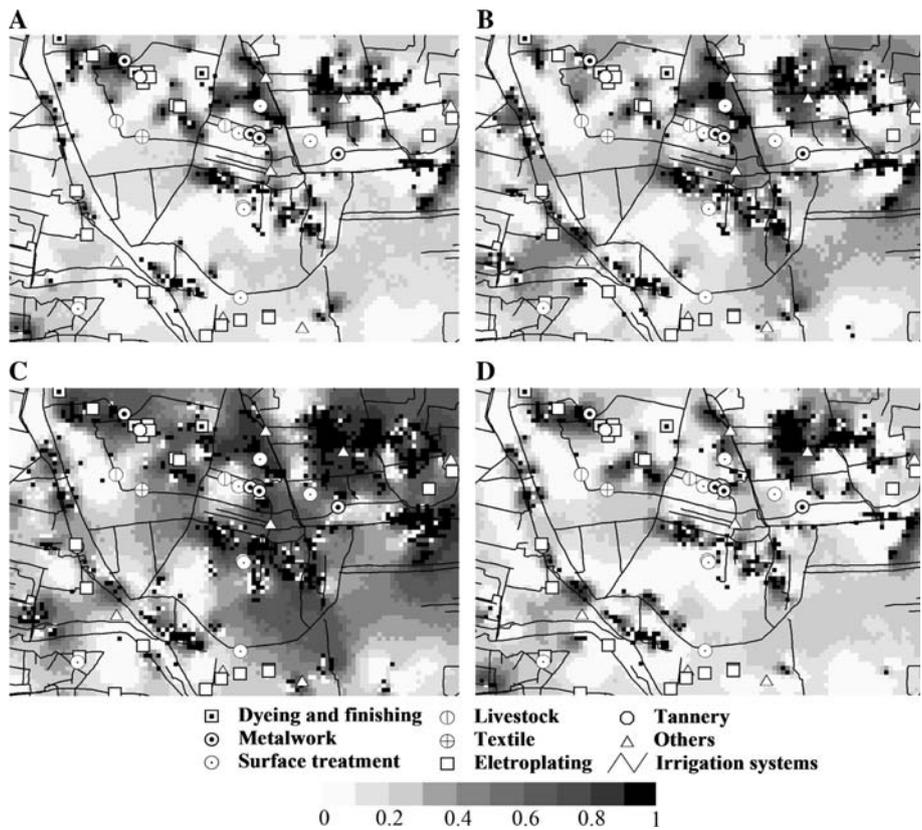
**Fig. 5** The contour map of average concentration in 1,000 realizations based on 900 samples: **a** Cr, **b** Cu, **c** Ni, **d** Zn



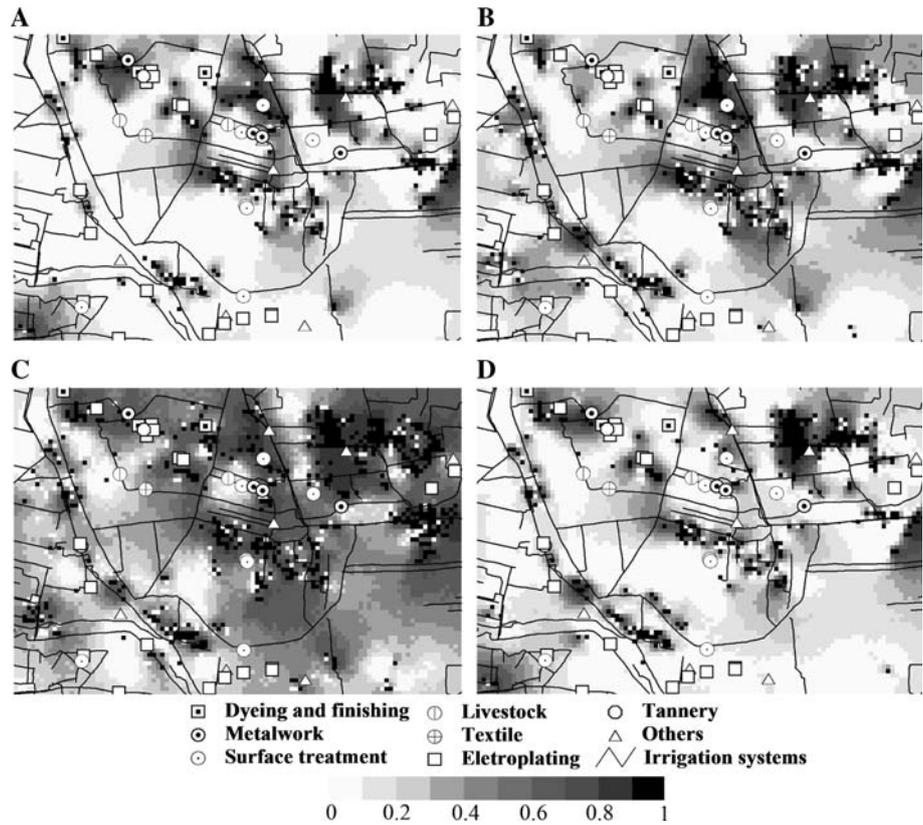
**Fig. 6** The contour map of average concentration in 1,000 realizations based on 700 samples: **a** Cr, **b** Cu, **c** Ni, **d** Zn



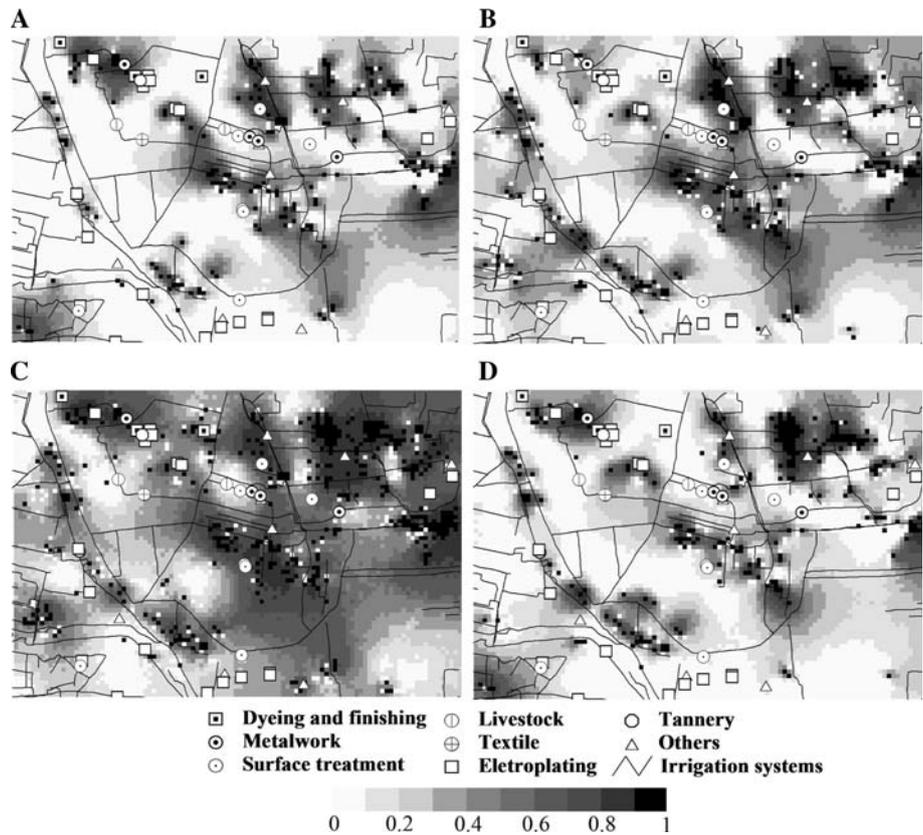
**Fig. 7** The probability map exceeding **a** Cr = 250 mg kg<sup>-1</sup>, **b** Cu = 200 mg kg<sup>-1</sup>, **c** Ni = 200 mg kg<sup>-1</sup>, **d** Zn = 600 mg kg<sup>-1</sup> of 1,000 realizations based on 1,082 samples



**Fig. 8** The probability map exceeding  
**a** Cr = 250 mg kg<sup>-1</sup>,  
**b** Cu = 200 mg kg<sup>-1</sup>,  
**c** Ni = 200 mg kg<sup>-1</sup>,  
**d** Zn = 600 mg kg<sup>-1</sup> of 1,000  
 realizations based on 900  
 samples



**Fig. 9** The probability map exceeding  
**a** Cr = 250 mg kg<sup>-1</sup>,  
**b** Cu = 200 mg kg<sup>-1</sup>,  
**c** Ni = 200 mg kg<sup>-1</sup>,  
**d** Zn = 600 mg kg<sup>-1</sup> of 1,000  
 realizations based on 700  
 samples



industrial plants and the irrigation systems of the study area, as Fig. 4 shows (Lin et al. 2002a, b; 2008). Figures 5 and 6 show the patterns of Cr, Cu, Ni, and Zn simulated by the SIS with 900 and 700 samples. The SIS yields similar spatial distributions of the simulated hot spots based on 700, 900, and 1,082 samples. Moreover, the SIS realizations based on 700 cLHS samples reveal slightly less areas with high heavy metal concentrations ( $300 \text{ mg kg}^{-1} > \text{Cr} > 150 \text{ mg kg}^{-1}$ ,  $300 \text{ mg kg}^{-1} > \text{Cu} > 150 \text{ mg kg}^{-1}$ ,  $600 \text{ mg kg}^{-1} > \text{Ni} > 300 \text{ mg kg}^{-1}$ ,  $600 \text{ mg kg}^{-1} > \text{Zn} > 300 \text{ mg kg}^{-1}$ ) than the simulations based on 1,082 samples do. However, in the areas with even higher heavy metal concentrations ( $\text{Cr} > 300 \text{ mg kg}^{-1}$ ,  $\text{Cu} > 300 \text{ mg kg}^{-1}$ ,  $\text{Ni} > 600 \text{ mg kg}^{-1}$ ,  $\text{Zn} > 600 \text{ mg kg}^{-1}$ ), the SIS realizations based on 700 cLHS samples identify less areas than SIS realizations based on 1,082 samples do. The simulated maps indicate that, for SIS realizations, the number of soil samples can be reduced by the cLHS for further soil monitoring and risk assessment in the study area.

Local and spatial uncertainty of heavy metals

Uncertainty assessment is a preliminary step in decision-making processes, such as delineation of hazardous area (Castrignanò et al. 2004; Cherubini 2009). Bourenane et al.(2007) argued that there are no validation criteria, such as for local uncertainty. The ccdf value, defined as the probability of being less than a threshold, could be referred to as local uncertainty for an unsampled location (Goovaerts 2001; Cattle et al. 2002; Amini 2005; Juang et al. 2004). This enables risk assessment when delineating locations, whether or not they are contaminated (Juang et al. 2004). Figure 7 shows the probability maps of soil contamination by Cr, Cu, Ni, and Zn exceeding the regulation thresholds ( $z_r$ ) calculated by local uncertainty equation (Eq. 11) based on the simulations using 1,082 samples. Figures 8 and 9 demonstrate the probability maps of Cr, Cu, Ni, and Zn contaminations exceeding the regulation thresholds based on the simulations using 900 and 700 cLHS samples. The probability maps indicate that the soil concentrations of heavy metals in areas near irrigation systems exceed regulatory thresholds (Figs. 7, 8, 9). According to the 1,000 realizations, the areas most likely to exceed regulation thresholds for Cr, Cu, Ni, and Zn content in the soil are those in the northern and central regions of the studied area. Moreover, the simulations with 700 and 900 samples obtained by cLHS reveal fewer areas with high probability of contamination than those with 1,082 samples.

The ccdf obtained by SIS only provides a measure of local uncertainty related to a single location, and a series of single-point ccdfs does not provide a measure of multi-point or spatial uncertainty (Goovaerts 2001, Juang et al. 2004). Therefore, given critical probabilities ( $P_c$ ), spatial

**Table 3** The joint probability ( $P_j$ ) of four heavy metals at m simulated locations in the contaminated area exceed regulatory thresholds on various critical probabilities ( $P_c$ ) in 1,000 realizations based on the 1,082, 900 and 700 samples

	$P_c$	Cr		Cu		Ni		Zn	
		m	$P_j$	m	$P_j$	m	$P_j$	m	$P_j$
No. 1,082	0.98	289	0.955	362	0.921	573	0.912	326	0.908
	0.97	292	0.905	366	0.829	589	0.636	331	0.821
	0.96	295	0.831	369	0.749	611	0.341	340	0.663
	0.95	298	0.734	373	0.649	631	0.187	347	0.510
No. 900	0.98	237	0.961	301	0.933	471	1.000	276	0.923
	0.97	241	0.869	312	0.771	475	0.912	282	0.809
	0.96	246	0.759	316	0.691	483	0.682	288	0.732
	0.95	248	0.704	327	0.456	494	0.445	295	0.604
No. 700	0.98	184	0.982	238	0.943	361	0.961	223	0.865
	0.97	185	0.958	247	0.784	363	0.925	231	0.745
	0.96	186	0.925	255	0.621	382	0.542	237	0.652
	0.95	190	0.799	266	0.439	408	0.230	241	0.575

uncertainty is required to assess reliability when delineating contamination (Juang et al. 2004; Zhao et al. 2005). The joint probabilities of given critical probabilities can be used to measure the reliability of delineated contaminations of heavy metals. Table 3 shows joint probabilities ( $P_j$ ) with given critical probabilities ( $P_c = 0.98, 0.97, 0.96$  and  $0.95$ ) when Cr, Cu, Ni, and Zn content in the soil exceeds regulatory thresholds. At a given critical probability, the higher the joint probability is, the more reliable the mapped contamination is. The risk could be low in mapping contamination when critical probabilities are greater than 0.98, particular in cLHS 900 samples. Moreover, the uncertainty analysis results indicate that spatial uncertainty is higher when delineating contaminations in the 1,082 samples than when delineating contaminations in cLHS 700 samples, except for the Zn contamination. The results of the analyses also indicate that delineating soil heavy metal contaminations using 700 samples are more reliable than using 1,082 samples for with the exception of Zn contamination. The uncertainty analyses results confirm the simulated maps indicating that the number of soil samples can be reduced by the cLHS for further soil monitoring, risk assessment, and remediation in the examined area.

Conclusions

Given a sufficient sample size, geostatistical stochastic simulation techniques are considered reliable for delineating contamination areas and for quantifying the uncertainty of heavy metal distributions in soil. This study combines cLHS and geostatistical simulation techniques in

multivariate distribution of heavy metal samples to select effective samples replicating the original data and to simulate spatial patterns and uncertainty of heavy metals. The nonparametric geostatistical simulation method (SIS) is used to map and delineate the spatial distributions and uncertainty of measured heavy metals. The cLHS approach is an effective approach for sampling multiple heavy metals from their multivariate distributions to replicate the statistical distributions and spatial structures of the heavy metals. The sampling results indicate that 700 soil heavy metal samples obtained by cLHS are sufficient for replicating the multivariate distribution and indicator variograms of the original heavy metal samples and can be used to monitor and delineate heavy metals in soil. The SIS combined with a sufficient number of cLHS samples can be used to simulate and map the spatial pattern and uncertainty of heavy metals. It seems that there is a correlation between the areas of high heavy metal concentration and industrial plants and irrigation systems in the study area. Based on regulatory thresholds, the proposed method is an effective approach for delineating heavy metal pollutions in soil sampling data when implementing environmental monitoring.

## References

- Amini M (2005) Mapping risk of cadmium and lead contamination to human health in soils of Central Iran. *Sci Total Environ* 347(1–3):64–77
- Bourennane H, King D, Couturier A, Nicoulaud B, Mary B, Richard G (2007) Uncertainty assessment of soil water content spatial patterns using geostatistical simulations: an empirical comparison of a simulation accounting for single attribute and a simulation accounting for secondary information. *Ecol Model* 205:323–335
- Carre F, McBratney AB, Minasny B (2007) Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. *Geoderma* 141(1–2):1–14
- Castrignano A, Buttafuoco G (2004) Geostatistical stochastic simulation of soil water content in a forested area of South Italy. *Biosyst Eng* 87:257–266
- Castrignanò A, Cherubini C, Di Mucci G, Giasi CI, Molinari M (2004) The application of spatial modelling of the variability for the interpretation of a case regarding arsenic pollution in groundwater. Integrated methods for assessing water quality. 21–22 October, Louvain-La-Neuve, Belgium
- Cherubini C (2009) A geostatistically based probabilistic risk assessment approach. In: *Evolving application domains of data warehousing and mining: trends and solutions*. Igi Global Publisher Idea Group Inc. ISBN: 978-1-60566-816
- Cherubini C, Musci F, Pastore N (2009) Stochastic geolithological reconstruction coupled with artificial neural networks approach for hydrogeological modeling. *Int J Math Models Methods Appl Sci* 2(3), ISSN: 1998-0140
- Deutsch CV, Cockerham PW (1994) Practical considerations in the application of simulated annealing to stochastic simulation. *Math Geol* 26(1):67–82
- Deutsch CV, Journel AG (1992) *GSLIB: geostatistical software library and user's guide*. Oxford University Press, New York, p 340
- Fortin MJ, Edwards G (2001) A cognitive view of spatial uncertainty. In: Hunsaker CT, Goodchild MF, Friedl MA, Case TJ (eds) *Spatial uncertainty in ecology: implications for remote sensing and GIS applications*. Springer, New York, pp 133–157
- Franco C, Soares A, Delgado J (2006) Geostatistical modelling of heavy metal contamination in the topsoil of Guadiamar river margins (S Spain) using a stochastic simulation technique. *Geoderma* 136(3–4):852–864
- Goovaerts P (1996) Stochastic simulation of categorical variables using a classification algorithm and simulated annealing. *Math Geol* 28:909–921
- Goovaerts P (1999) Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89(1–2):1–45
- Goovaerts P (2001) Geostatistical modelling of uncertainty in soil science. *Geoderma* 103(1–2):3–26
- Hassan MM, Atkins PJ (2007) Arsenic risk mapping in Bangladesh: a simulation technique of cokriging estimation from regional count data. *J Environ Sci Health Part A Toxic Hazard Subst Environ Eng* 42(12):1719–1728
- Iman RL, Conover WJ (1980) Small sample sensitivity analysis techniques for computer-models, with an application to risk assessment. *Commun Stat Part A Theory Methods* 9(17):1749–1842
- Journel AG (1996) Modelling uncertainty and spatial dependence: stochastic imaging. *Int J Geogr Inform Sci* 10(5):517–522
- Juang KW, Chen YS, Lee DY (2004) Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils. *Environ Pollut* 127(2):229–238
- Kyriakidis PC (2001) Geostatistical models of uncertainty for spatial data. In: Hunsaker CT, Goodchild MF, Friedl MA, Case TJ (eds) *Spatial uncertainty in ecology: implications for remote sensing and GIS applications*. Springer, New York, pp 175–213
- Lin YP (2002) Multivariate geostatistical methods to identify and map spatial variations of soil heavy metals. *Environ Geol* 42:1–10
- Lin YP (2008) Simulating spatial distributions, variability and uncertainty of soil arsenic by geostatistical simulations in Geographic information systems. *Open Environ Sci* 2:26–33
- Lin YP, Chang TK, Teng TP (2001) Characterization of soil lead by comparing sequential Gaussian simulation, simulated annealing simulation and kriging methods. *Environ Geol* 41(1):189–199
- Lin YP, Chang TK, Shi CW, Tsen CH (2002a) Factorial and indicator kriging method using geographic information system to delineate spatial variation and pollution sources of soil heavy metals. *Environ Geol* 42:900–909
- Lin YP, Teng TP, Chang TK (2002b) Multivariate analysis of soil heavy metal pollution and landscape pattern in Changhua County in Taiwan. *Landsc Urban Plan* 62(1):19–35
- Lin YP, Yeh MS, Deng DP, Wang YC (2008) Geostatistical approaches and optimal additional sampling schemes for spatial patterns and future sampling of bird diversity. *Glob Ecol Biogeogr* 17:175–188
- Lin YP, Cheng BY, Shyu GS, Chang TK (2010) Combining a finite mixture distribution model with indicator kriging to delineate and map the spatial patterns of soil heavy metal pollution in Chunghua County, Central Taiwan. *Environ Pollut* 158(1):235–244
- McKay MD, Beckman RJ, Conover WJ (1979) Comparison of 3 methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21(2):239–245
- Minasny B, McBratney AB (2006) A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput Geosci* 32(9):1378–1388

- Pachepsky Y, Acock B (1998) Stochastic imaging of soil parameters to assess variability and uncertainty of crop yield estimates. *Geoderma* 85:213–229
- Saito H, Goovaerts P (2000) Geostatistical interpolation of positively skewed and censored data in a dioxin-contaminated site. *Environ Sci Technol* 34(19):4228–4235
- Schnabel U (2004) Uncertainty assessment for management of soil contaminants with sparse data. *Environ Manage* 33(6):911–925
- Van Meirvenne M, Goovaerts P (2001) Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. *Geoderma* 102(1–2):75–100
- Xu CG, He HS, Hu YM, Chang Y, Li XZ, Bu RC (2005) Latin hypercube sampling and geostatistical modeling of spatial uncertainty in a spatially explicit forest landscape model simulation. *Ecol Model* 185(2–4):255–269
- Zhao YC, Shi XZ, Yu DS et al (2005) Uncertainty assessment of spatial patterns of soil organic carbon density using sequential indicator simulation, a case study of Hebei Province, China. *Chemosphere* 59(11):1527–1535
- Zhao Y, Xu X, Sun W et al (2008) Uncertainty assessment of mapping mercury contaminated soils of a rapidly industrializing city in the Yangtze River Delta of China using sequential indicator co-simulation. *Environ Monit Assess* 138:343–355