

Using robust kriging and sequential Gaussian simulation to delineate the copper- and lead-contaminated areas of a rapidly industrialized city in Yangtze River Delta, China

Yongcun Zhao · Xianghua Xu · Biao Huang ·
Weixia Sun · Xuexin Shao · Xuezheng Shi ·
Xinling Ruan

Received: 23 October 2006 / Accepted: 29 January 2007 / Published online: 1 March 2007
© Springer-Verlag 2007

Abstract A total of 540 topsoil samples (0–15 cm), 188 subsoil samples (20–40 cm), and four individual soil profiles were collected in this study for mapping the Cu- and Pb-contaminated areas in soils of Zhangjiagang city, an industrialized city in the Yangtze River Delta region of China. Robust geostatistical methods were applied for identifying possible spatial outliers of Cu and Pb data, and then a sequential Gaussian simulation was employed for delineating the potential areas where Cu or Pb concentration was affected by diffuse pollution. The results showed that the spatial outliers of Cu and Pb were strongly associated with various types of factories. The anthropogenic input of Cu to soils at local hotspots was closely related to emissions of printing and dyeing, metallurgical, and chemical factories, whereas a lead oxide factory and a chemical factory resulted in a considerable increase of Pb in the topsoil of the study area. Approximately 30% of the total land area of the study was at potential risk from the Cu or Pb diffuse pollution resulting from rapid industrialization of the area over the past 20 years.

Keywords Robust geostatistics · Sequential Gaussian simulation (SGS) · Heavy metal · Spatial outlier · Diffuse pollution

Introduction

With the quick progress of industrialization and urbanization in developing countries, the dispersion of heavy metals into the environment through poorly treated waste-water discharge, solid waste disposal, atmospheric deposition and the use of agrochemicals has gained great concerns due to their potential impacts on the environment and on public health. As the core of the superficial ecosystem of the earth and the pivot for material and energy exchanges among the atmosphere, hydrosphere, biosphere, and lithosphere, soil is being affected by the human activities which are stronger than ever before. Consequently, soils are subjected to a number of pollutants due to different human activities. Among the soil and environmental pollutants, heavy metals have received considerable attention over the last few decades (Franssen et al. 1997) due to their accumulation with time and their mobility resulting from changing environmental conditions or by saturation beyond the buffering capacity of the soil (Facchinelli et al. 2001).

Information and knowledge about the spatial patterns of heavy-metal contaminated areas and hotspots are important for decision makers to develop effective management recommendations. To delineate the heavy-metal contaminated areas, the kriging estimation, which provides the best linear unbiased prediction at unsampled locations, has been widely used for quantifying the spatial patterns of heavy metals (Atteia et al. 1994; Zhang and Selinus 1997; Meuli et al. 1998; Van Meirvenne and Goovaerts 2001). In practice, however, it is difficult to accurately characterize heavy-metal polluted areas because of their complex spatial pattern, high coefficient of variation, and occurrence of hotspots with locally strong-contaminated soils (Franssen et al. 1997). The kriging estimator is a smoothing interpolator characterized as

Y. Zhao · B. Huang (✉) · W. Sun · X. Shao ·
X. Shi · X. Ruan
State Key Laboratory of Soil and Sustainable Agriculture,
Institute of Soil Science, The Chinese Academy of Sciences,
Nanjing 210008, China
e-mail: bhuang@issas.ac.cn

X. Xu
Department of Earth Sciences, Nanjing University,
Nanjing 210093, China

locally accurate but globally inaccurate, and kriging estimation cannot reproduce basic statistics of the input data. Another drawback of the kriging estimator is that the smoothing effect depends on the local data configuration. For these reasons, the kriged maps should not be used for applications sensitive to the presence of extreme values and their patterns of continuity (Goovaerts 1999). Therefore, uncertainty associated with the delineation of heavy-metal contaminated areas by the kriging estimator needs to be considered in decision making for future management (Goovaerts 2001) because such uncertainty can be propagated into the subsequent environment modelling and has fundamental impacts on the ultimate results of the model. The stochastic simulation method was recently proposed to overcome the shortcomings of kriging variance (Goovaerts 1999) and the limitations inherent in indicator kriging for assessing this kind of uncertainty (Deutsch and Journel 1998; Juang et al. 2004). The equiprobable realizations generated by conditional stochastic simulations allow to establishing models to quantify the uncertainty. The stochastic simulation provides either an estimated mean value of heavy-metal concentration or the probability of exceeding a given threshold level (Cattle et al. 2002), thus it can be used for determining the heavy-metal contaminated areas and associated uncertainty through setting a given critical probability and calculating the joint probability of the obtained areas. In studies by Mowrer (1997) and Wang et al. (2001), sequential Gaussian simulations were applied to generate estimations of variables for old-growth forest conditions and to assess the uncertainty of the soil erodibility factor for a revised universal soil-loss equation, respectively. Juang et al. (2004) assessed the uncertainty of delineating heavy-metal contaminated soils using a sequential indicator simulation and discussed the local uncertainty and spatial uncertainty of mapping Cu concentrations.

The heavy-metal data in polluted areas often contain outliers; the existence of outliers will strongly influence the estimation of polluted areas. Spatial patterns of heavy metals in soils affected by industrialization process involve spatially continuous variations arising from parent material and the effects of diffuse pollution. Superimposed on this is point contamination at local hotspots (Rawlins et al. 2005). Sites influenced by point contamination are likely to be spatial outliers. In the studies of Lark (2002) and Rawlins et al. (2005), a robust geostatistical method was applied for detecting the spatial outliers of heavy metals. An observation will be called a spatial outlier if its value is notably different compared to those distributed at the sites around it. In geostatistics, the spatial outliers are interesting by themselves because they can provide additional information about the model (Militino et al. 2006) and potential

point sources of contamination. The removal of spatial outliers can also eliminate the influences of spatial outliers on conventional geostatistical analysis using Matheron's estimator of the variogram (Lark 2000). Therefore, the detection and removal of spatial outliers before mapping the potential "areas" of heavy-metal contamination will help identify the point sources of contamination and improve the estimation of areas that are under the influences of non-point sources of contamination.

The Yangtze River Delta region is one of the most developed areas in China. Soil quality in the area has been greatly impacted by rapid industrialization, urbanization and intense agricultural uses. In this area, industrialization has occurred even in small towns and villages with limited land area and many small factories are intermixed with farmlands. Consequently, soil properties and pollutants vary dramatically even across very short distances due to scattered industrialization and human activities. Identification and delineation of heavy-metal contaminated areas and local hotspots in this area are important because they can serve as a basis for planning management strategies to achieve better environmental quality.

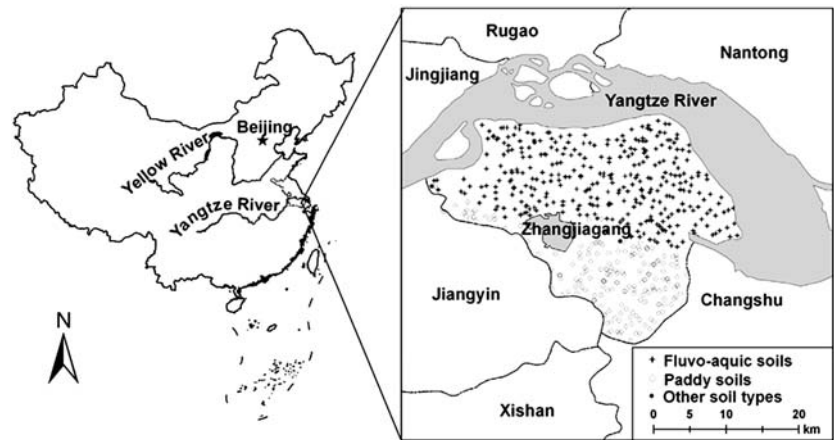
The aim of this study is to combine robust geostatistical analysis, conditional stochastic simulation and sampling analysis of topsoil (0–15 cm), subsoil (20–40 cm), and individual soil profiles to delineate the possible areas of Cu and Pb contamination in topsoil of a rapidly industrialized city, Zhangjiagang City, in the Yangtze River Delta region of China. The specific objectives of this study were: (1) to identify local hotspot of Cu and Pb contamination in the study area; (2) to delineate the potential areas where the Cu or Pb concentration were affected by diffuse pollution; and (3) to assess the uncertainties associated with the delineation of Cu/Pb-contaminated areas.

Materials and methods

Study area

The study area is Zhangjiagang city, which is situated on a flat alluvial plain on the Yangtze River Delta (Fig. 1). The total area of Zhangjiagang city is 999 km², of which 799 km² are terrestrial areas, inhabited by a population of 860,200. With a north sub-tropical monsoon climate, Zhangjiagang has a mean annual temperature 15.2°C and an annual rainfall of 1039.3 mm. The main soil types in the study area are fluvo-aquic soils (Aquic Cambosols) and paddy soils (Stagnic Anthrosols) (SSOSC 1984; STRG 2001). Prevailing in the fields along the Yangtze River in the northern part of the city are slightly alkaline fluvo-aquic soils, most are grey fluvo-aquic soils (Anthrostagnic-Dark-Cambosols), which were derived from alluvial parent

Fig. 1 The geographic location of Zhangjiagang city and the sampling distribution patterns



materials with medium-light loamy texture. Derived from fairly clayed, neutral and slightly acidic lagoon-phase alluvial materials, paddy soils were mainly distributed in flat fields in the southern part of the city; most are identified as waterloggogenic paddy soils (Typ-Fe-accumuli-Anthrosols).

Since the 1980s, Zhangjiagang has become one of the most rapidly economically developing areas in the Yangtze River Delta region. With a greater than 60% share of the city’s total GDP, industries in the city consist of chemical, metallurgical, electroplating, printing and dyeing, paper making, breeding, etc.

Sampling and chemical analysis

A total of 540 topsoil samples (0–15 cm) were collected in 2004 (Fig. 1), of which 379 samples throughout the study area were collected according to the soil types and land use, as well as an even coverage of the whole study area. When sampling, soils in the top layer of 6–8 points in each site of an area of about 0.17 ha were collected and then fully mixed, and finally divided into parts of approximately 1–2 kg each. Only one of the parts was packed in a bag and brought back for laboratory analysis. In 27 out of the 379 sampling sites, samples for subsoil (20–40 cm) were also collected. One hundred sixty-one samples around various factories (the factory locations with pollutant discharging were provided by the local environmental authority) were also collected; sampling sites were mainly decided by waste discharging forms of industries. When sampling around a waste-water discharging industry, topsoil (0–15 cm) and subsoil (20–40 cm) at 6–8 points distributed in an area of about 0.08 ha at points 50–100 m away in the downstream direction from the water-discharging outlet were collected and mixed, respectively. Of the mixed soils, 1–2 kg were packed and taken back to the laboratory. For

atmospheric discharging cases, topsoil and subsoil samples were collected at sites at 50–100 m away from discharging outlets in the windward or counter-windward direction. For solid-waste disposal sites, samplings (topsoil and subsoil) were taken 50–100 m away from the disposal site. Furthermore, four profiles were also selected and sampled according to their degree of surface contamination and the soil types. All sampling sites were recorded by global positioning system (GPS) and the related information such as land-use history, vegetation, soil type, and present and potential pollutant sources were also recorded in detail.

The collected soil samples were air dried and then ground to 100 meshes for chemical analysis. The soil samples were digested by reverse aqua regia (HCl/HNO₃ = 1/3) and the total concentrations of Cu and Pb were determined by flame atomic absorption spectrometry (AAS) (Burt et al. 2003). Quality control was based on the use of certified samples (GBW 07413, ASA-2 and GBW 07415, ASA-4) and duplicates of the analysis.

Detection of high-value spatial outliers

The robust method proposed by Lark (2000; 2002) was employed in this study to detect the probable spatial outliers in Cu and Pb data. This mainly involved two procedures:

- (1) Define a proper aggregation function and neighborhood set to estimate the expected value of the spatial point using its neighbors’ non-spatial value.

Here, the kriging estimator was selected as the aggregation function and the range of the variogram for Cu (Pb) was set as the boundary of the neighborhood set. The key to procedure was to select the proper estimator of the variogram. Matheron’s estimator and three robust estimators, Cressie and Hawkins’s (1980) estimator, Dowd’s (1984) estimator, and Genton’s (1998) estimator, were applied to

determine whether the robust estimator was necessary or which robust estimator was best (see Lark 2000, 2002 for detailed information about the equations for calculating the four estimators of the variogram). The experimental semi-variances of each estimator and the corresponding fitted models were calculated by using R (R Development Core Team 2005), GeoR (Ribeiro Jr. and Diggle 2001) and qn packages for R (Furrer and Genton 1999).

A variogram model for each estimator was then used in kriging to estimate each observation in the data set by cross-validation, and the variogram models were compared with respect to the median of $\theta(x)$. The statistic $\theta(x)$ can be used for evaluation of the variogram used in the kriging (Lark 2000).

$$\theta(x_i) = \frac{\{\hat{Z}(x_i) - z(x_i)\}^2}{\sigma_{K,i}^2}, \quad (1)$$

where $\hat{Z}(x_i)$ denotes the estimate of ordinary kriging, $z(x_i)$ is the observed value, and $\sigma_{K,i}^2$ represents the kriging variance. The best variogram models for each of the two metals can be judged by comparing the median of $\theta(x)$ with 0.455, if the median of $\theta(x)$ is close to 0.455, then proper, otherwise not.

- (2) Define a proper difference function to evaluate the contrast of the estimated value and the measured value, and then define a statistical test to identify spatial outliers by examining the difference function.

The difference function was defined as $\hat{Z}(x_0) - z(x_0)$, and the standardized kriging error $\varepsilon_s(x)$ was calculated for identifying spatial outliers.

$$\varepsilon_s(x) = \frac{\hat{Z}(x_0) - z(x_0)}{\sigma_{K,0}}. \quad (2)$$

A datum was classified as a high-value spatial outlier if its standardized kriging error was less than -1.96 , i.e., if it fell below the lower 95% confidence interval for a standard normal variate (Rawlins et al. 2005).

Modeling the contaminated areas and associated uncertainty

The Cu (Pb) data without spatial outliers and the conditional sequential Gaussian simulation (SGS) were used to generate 1,000 equiprobable realizations (cell size 50 m × 50 m) for obtaining the potential areas where the Cu (Pb) concentration was mainly affected by diffuse pollution.

The SGS algorithm (Deutsch and Journel 1998) involves the following steps: (1) A random path visiting each node of the regularly spaced grid covering the study area was

established; (2) For each unsampled location, the normal-score-transformed data values were used to estimate the expected value of the multivariate Gaussian probability distribution function for that grid node using ordinary kriging; (3) The conditional probability density function for the grid node could be determined because of the assumption of Gaussian spatial behavior; (4) A simulated value was randomly drawn from that probability distribution as one possible simulation of the variable at that location; this simulated value is then added to the conditioning data set and the calculation proceeds to the successive node along the whole random path through the grid. All nodes at the random path are visited and each node received a simulated value will obtain a realization of SGS. Repeating L times of SGS, and each time using the different path to visit all nodes of the grid defined over the study area will result in L equiprobable realizations. A set of realizations generated by SGS can be used to calculate the probability that the unknown Cu or Pb concentration $z(x')$ at location X' is greater than a given background value z_t of Cu or Pb, it can be denoted as $\text{Prob}[z(x') > z_t]$, and this probability can be calculated based on the following equation:

$$\text{Prob}[z(x') > z_t] = \frac{n(x')}{1,000}. \quad (3)$$

The SGS was carried out 1,000 times and $n(x')$ is the number of realizations that the simulated Cu (Pb) values generated by SGS were greater than the background value in each of the 1,000 realizations.

With a given critical probability P_c , the areas where Cu and Pb concentrations were probably affected by diffusion pollution process can be obtained using the rule $\text{Prob}[z(x') > \text{background value}] \geq P_c$, and the reliability of the obtained areas can be assessed using joint probability. Suppose there are n locations, m_1, m_2, \dots, m_n , in area A , the probability of Cu (Pb) concentration of n locations in area A all being greater than the threshold z_t , namely the joint probability, can be calculated based on the following equation:

$$\begin{aligned} & \text{Prob}[z(m_1) > z_t, z(m_2) > z_t, \dots, z(m_n) > z_t] \\ &= \frac{n(m_1, m_2, \dots, m_n)}{1,000}, \end{aligned} \quad (4)$$

where 1,000 is the simulation times, $n(m_1, m_2, \dots, m_n)$ is the number of realizations that have all simulated Cu (Pb) values of n locations in area A all being greater than the threshold value z_t in each realization in the 1,000 realizations.

The *sgsim* and *postsim* subroutines in the software package *GSLIB* (Deutsch and Journel 1998) were used to

perform the SGS and the post processing of the realizations.

Results

Descriptive statistical analysis

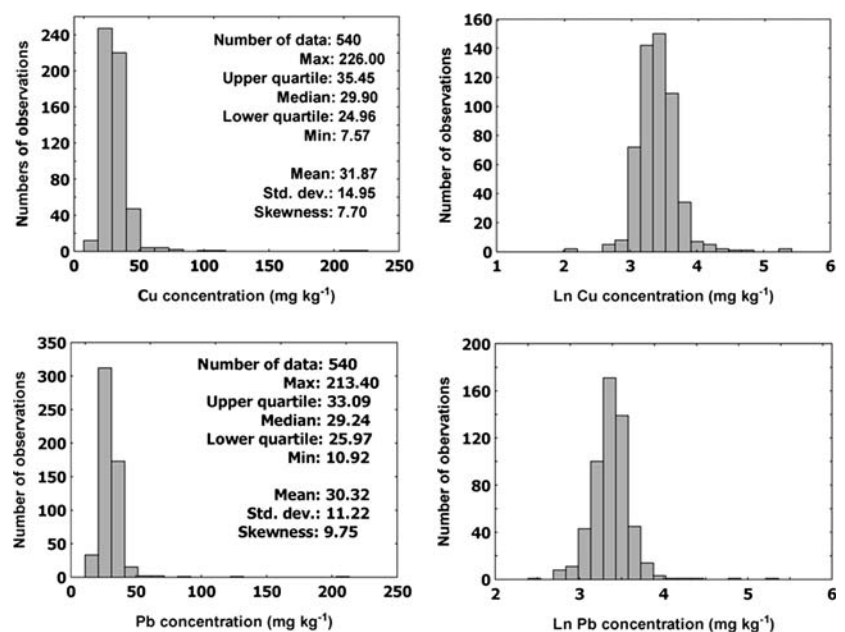
Figure 2 is the histogram and descriptive statistics of the raw data on soil Cu and Pb concentrations and the same data transformed to natural logarithms. The descriptive statistics showed that the mean values of Cu and Pb concentrations in topsoil of the study area were 31.87 and 30.32 mg kg⁻¹, respectively. The ranges between the minimum and maximum of Cu and Pb concentrations were estimated at 218.43 and 202.48 mg kg⁻¹, respectively. The Cu and Pb concentrations observed in this study highlighted the large variation of Cu and Pb data, and the skewness is 7.70 for Cu data and 9.75 for Pb data. The histograms of Cu and Pb raw data showed a long right-tail distribution, indicating that there existed some higher Cu and Pb concentration values and that the spatial distribution of Cu and Pb concentrations in topsoil of the study area is not homogeneous. The skewness of Cu and Pb data decreased after the log transformation (natural logarithm) was applied (Fig. 2), and the log-transformed histograms exhibited the similar shapes to a normal distribution excepted that several high values of Cu and Pb data existed in the right tails, indicating that outliers of Cu and Pb concentrations might exist in the data.

Identification of spatial outliers

The isotropic variograms of the log-transformed Cu and Pb concentrations estimated by Matheron’s estimator and the three robust estimators of variograms are shown in Fig. 3. Compared with the result from Matheron’s classical estimator, the nugget (C_0) and sill estimated by the three robust estimators of the variogram were significantly lower, whereas the ranges were larger than the result of Matheron’s estimator, which could be attributed to the fact that the robust estimators of the variograms are of great resistances to extreme values.

Table 1 presents the median values of $\theta(x)$ calculated from the cross-validations based on four variogram models. It could be inferred that the median value of $\theta(x)$ obtained by using Matheron’s estimator was significantly smaller than 0.455 and Matheron’s estimator appears to overestimate the variogram. All the robust variogram estimators give median values of $\theta(x)$ in the 95% confidence interval about 0.455 (the expected confidence interval for the median of $\theta(x)$ is 0.338–0.572 for 540 samples), with Dowd’s estimator for Cu and Pb closest to this target value, thus Dowd’s robust estimator of variogram was selected to compute the standardized kriging error of the log-transformed data for identifying the spatial outliers in Cu and Pb data. The kriged maps of Cu and Pb, derived from the back-transform of ordinary kriging estimates of the log-transformed data (Webster and Oliver 2000), were also presented for giving a first idea of the Cu and Pb distribution trends before and after the spatial outliers were removed.

Fig. 2 The descriptive statistics and histograms of Cu and Pb in topsoil of Zhangjiagang city



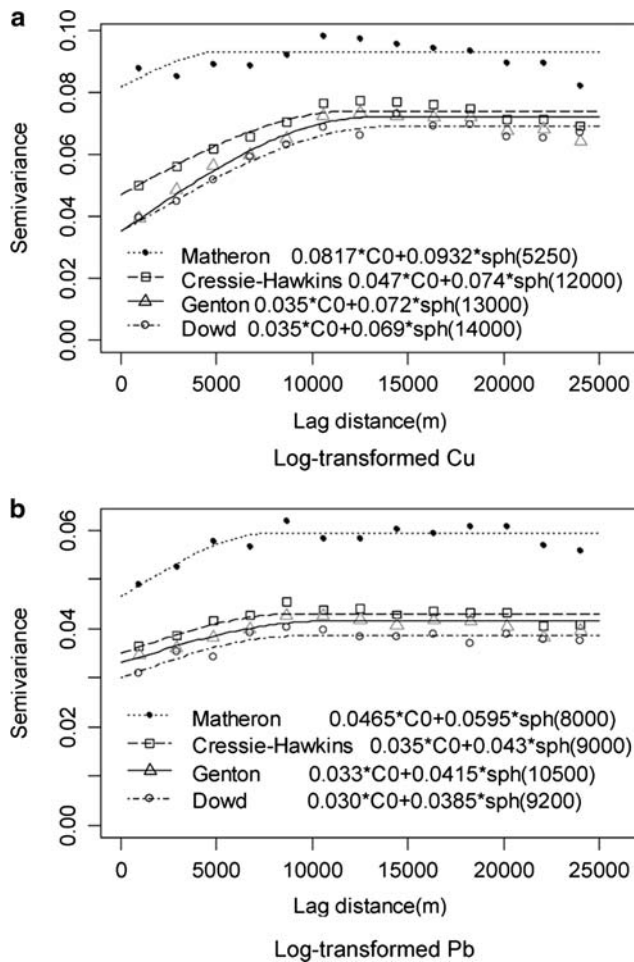


Fig. 3 The variograms and fitted models of log-transformed Cu and Pb data estimated by four estimators of variograms

Table 1 The median values of $\theta(x)$ from cross-validation based on four variogram models

	Matheron	Cressie–Hawkins	Genton	Dowd
Log-transformed Cu data	0.202	0.338	0.444	0.451
Log-transformed Pb data	0.258	0.348	0.369	0.404

A total of 35 unique sample locations were identified as high-value spatial outliers, the number of high-value spatial outliers for Cu (Fig. 4) and Pb (Fig. 5) were 21 and 17, respectively. There are three locations where Cu and Pb concentrations were all spatial outliers. The three locations were distributed around an electroplating factory, a non-ferrous metal solvent factory and a paper making factory, respectively, according to the field investigation.

Delineation of contaminated areas and assessment of associated uncertainty

The diagnosis of Cu and Pb diffuse accumulation in soils requires information of the pedogeochemical background of the elements in soil. The Ministry of Agriculture of China (Agricultural Environment Background Value Research Group 1997) provided the background values of heavy metals in soils of Suzhou city (Table 2) and the study area, Zhangjiagang city, is part of Suzhou city. Fortunately, this background value survey project was conducted from 1978 to 1985 and this period was immediately before the industrialization of the study area. The dominant soil types in Suzhou are waterloggogenic paddy soils and grey fluvo-aquic soils, and these two types of soils covered more than 85% of total land area of Suzhou, which were similar to the study area, thus these background values might be used for determining the areas where Cu or Pb were mainly affected by non-point sources of pollution (diffuse pollution).

The mean values of the background values of Cu and Pb concentrations in topsoil of Suzhou city (Table 2) were selected to calculate the probability of Cu or Pb concentration being larger than the background values to map the potential areas affected by diffuse pollution. The upper concentration limit of the Cu and Pb background values were also applied for obtaining the most probable areas. Figure 6 shows the calculated probabilities of Cu and Pb being greater than the corresponding background values and upper concentration limit of background values, respectively. It demonstrated that areas where Cu concentrations were probably affected by diffuse pollutions were

Fig. 4 The identified spatial outliers and ordinary kriging estimates of Cu concentrations (figures labeled on corners of the map are coordinates of Albers projection; ‘East (m)’ and ‘North (m)’ indicate the direction of the coordinate system, and the coordinate unit is meter)

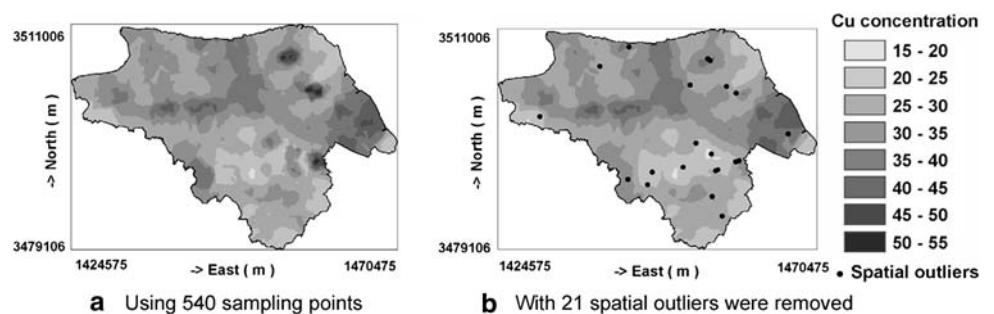


Fig. 5 The identified spatial outliers and ordinary kriging estimates of Pb concentrations

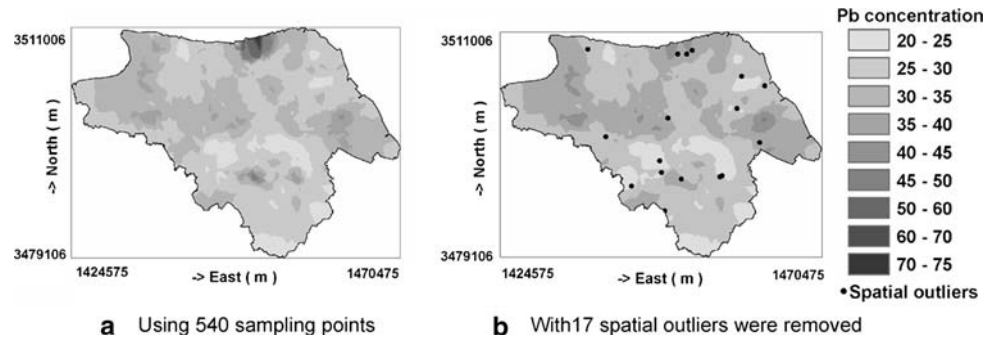


Table 2 The background values of copper and lead in topsoil of Suzhou city

	Number of samples	Coefficient of variation (%)	Mean (mg kg ⁻¹)	Lower-upper concentration limit (mg kg ⁻¹)
Cu	45	19.5	24.88	15.35–34.41
Pb	45	25.3	25.48	12.86–38.10

mainly located in the center of northern parts of the study area (Fig. 6a), whereas for Pb the dispersed areas covered most of the study area (Fig. 6c).

With a given critical probability P_c , the areas where Cu and Pb concentrations were probably affected by diffuse pollution were obtained from Fig. 6 based on the rule $\text{Prob}[z(x') > \text{background value}] \geq P_c$. Here, 0.95 was selected as the critical probability. The obtained areas are presented in Fig. 7a, b, respectively. The black zones in Fig. 7 indicate that the Cu (Pb) was affected by diffuse pollution more so than the gray zones due to application of

the upper concentration limit of background values in the former. Here, the black zones may be considered the most possible areas affected by diffuse pollution whereas the gray zones indicate the potential areas. The potential areas were mainly distributed in the northern part of the study area where metallurgical, chemical and dyeing factories were dense and intermixed with farmlands. Table 3 shows that approximately 30% of the total land area is at potential risk resulting from the diffuse pollution.

However, the probability distributions in Fig. 6 cannot provide any measure of the reliability of the obtained areas in Fig. 7, because the conditional cumulative distribution function obtained by SGS only provides a measure of single-point uncertainty, and a series of single-point conditional cumulative distribution functions do not provide any measure of spatial uncertainty (Goovaerts 2001). The joint probability obtained from the realizations generated by SGS can be used to assess the reliability of obtained areas affected mainly by diffuse pollutions. Table 3 indicates that the obtained areas by SGS are of relatively high reliability because the joint probability of the obtained

Fig. 6 Probability of Cu/Pb being greater than the background value

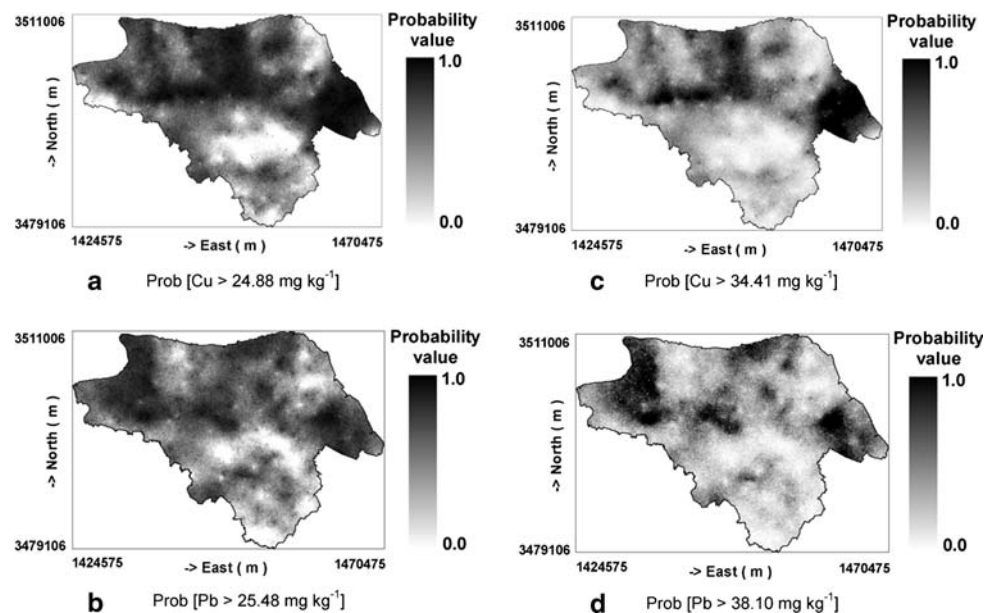


Fig. 7 The areas where Cu or Pb was affected by diffuse pollution

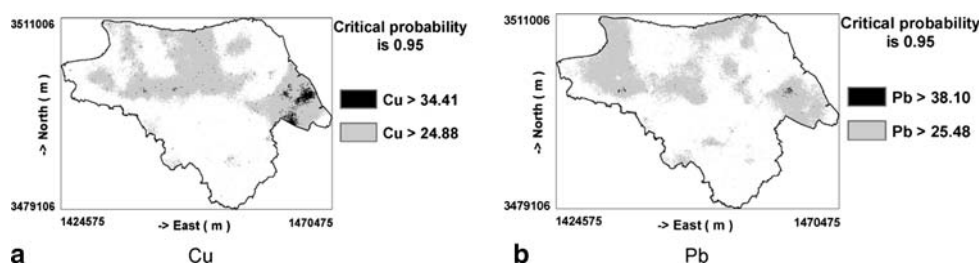


Table 3 The joint probability for areas where heavy-metal concentration is greater than the background value

Heavy metal	Background value (mg kg ⁻¹)	Area percentage (%)	Joint probability
Cu	34.41	1.30	0.79
	24.88	33.3	0.72
Pb	38.10	0.14	0.84
	25.48	29.2	0.70

The critical probability is 0.95

areas were all greater than 0.7; therefore, the spatial uncertainty of the obtained areas is small.

Discussion

Relationship between spatial outliers and the type of factories

The high nugget-to-sill ratio of the variograms coupled with the long range would cause a great deal of smoothing. Therefore, a high value with few sample points near it would likely appear as an outlier, not because it is indeed a hotspot, but because too few samples were taken near it. However, no such situations for the identified spatial outliers were observed in this study. The spatial outliers of Cu and Pb were strongly associated with various types of factories. Figure 8 shows the relationships between Cu/Pb spatial outliers and the type of factories closest to each outlier. It could be inferred from Fig. 8 that the anthropogenic input of Cu to soils was mainly connected to emissions of printing and dyeing, metallurgical, and chemical factories, whereas the lead oxide factory and chemical factory resulted in the significant increase of Pb in the topsoil of the study area. Based on the Chinese Environmental Quality Standard for Soils (GB 15618-1995) (State Environmental Protection Administration of China 1995), the Cu concentrations of five samples exceeded the Cu guide value (the guide value is 50 mg kg⁻¹ when pH < 6.5, and 100 mg kg⁻¹ when pH > 6.5) (Fig. 8, samples with concentrations exceed the guide value were

labeled by dashed rectangles). As for Pb in topsoils of the study area, no samples fell outside the guide value (the guide value is 250 mg kg⁻¹ when pH < 6.5, and 300 mg kg⁻¹ when pH > 6.5). There are also samples around breeding farms or even in farmland without potential influences of factories that were identified as spatial outliers (Fig. 8), these could be related to the intensive use of agricultural soils in the study area, and Cu and Pb are also present in some chemical fertilizers and organic or inorganic pesticides, as well as in domestic sewage sludge.

The concentrations of Cu and Pb in topsoil were significantly higher than those in subsoil (Fig. 8), and the maximal ratio of metal concentration in topsoil to those in subsoil for Cu and Pb were 6.2 and 7.6, respectively, indicating that Cu and Pb accumulated mainly in topsoil. The vertical distributions of Cu and Pb in two profiles near the spatial outliers with the highest Cu and Pb values are presented in Fig. 9. To obtain detailed information about the vertical distribution of Cu or Pb, the profiles were sampled carefully by fixed-depth intervals ranging from 2 to 10 cm without taking into account the podogenic horizons. Figure 9 also shows that Cu and Pb concentrations decreased abruptly from the surface downward, indicating that the accumulations of Cu and Pb were mainly restricted to the upper 20 cm whereas no significant changes in concentrations were observed below 20 cm. The highest Cu concentration was found at about 13 cm, which may be the result of anthropogenic disturbances, e.g., the farmland is plowed every year and generally the ploughed depth would be less than 20 cm. Thus, it could be inferred that Cu and Pb in soils of the study area might have a very low geochemical mobility. Furthermore, various reports on the vertical distribution of metals in soil also showed a superficial accumulation of non-indigenous metals, whether these originate from the spreading of waste or from atmospheric fallout of industrial contamination (Kuo et al. 1983; Chang et al. 1984; Li and Shuman 1996; Sterckeman et al. 2000). It may be insufficient, however, to evaluate the depth that has been reached by the contamination because the profile depths examined in this study did not exceed 1 m.

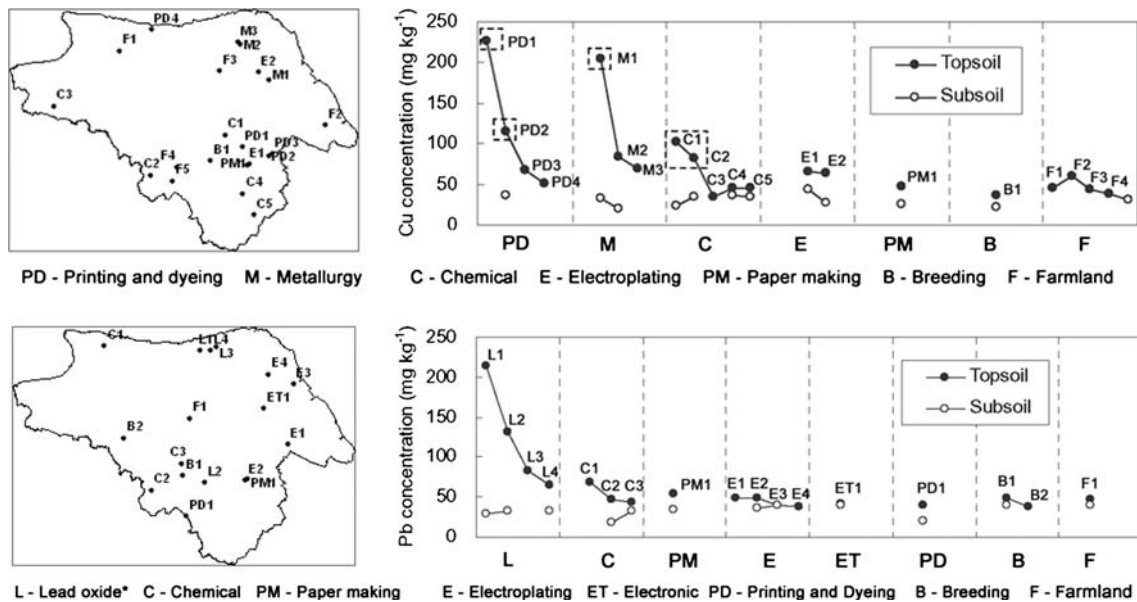


Fig. 8 The comparison of Cu and Pb concentrations of spatial outliers between topsoil and subsoil and factory types (*Here, the lead oxide factory was separated for facilitating to identify the potential point-source pollution)

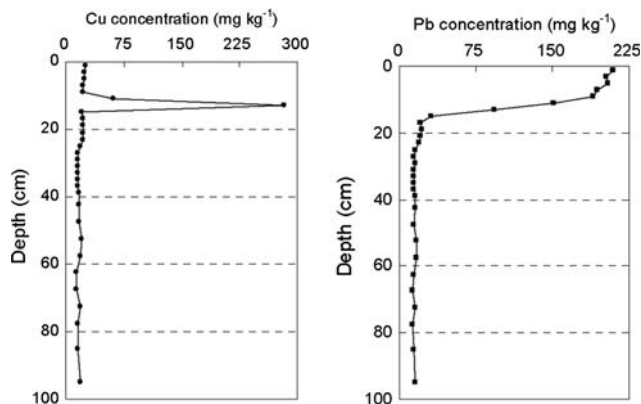


Fig. 9 Vertical distributions of Cu and Pb in profiles around spatial outliers of Cu and Pb with the highest concentrations (grey fluvo-aquic soils, near chemical factory)

Representing background concentrations of Cu and Pb before industrialization

The pedogeochemical background of Cu and Pb in soils of the study area may also be determined by analyzing the corresponding horizons of non-contaminated soils of the same type (Sterckeman et al. 2000). In terms of pedogenic influences, the two main soil types in the study area, grey fluvo-aquic soils and waterloggogenic paddy soils, were derived from different parent materials. As the underlying lithology is not readily accessible, the influence of parent material was studied through the analysis of the C (parent material) and G (gley) horizons of two profiles taken from

the two main soil types. Figure 10 shows that the Cu concentrations in the C horizon of the grey fluvo-aquic soil profile and the G horizon of the waterloggogenic paddy soil profile were 25.39 and 26.62 mg kg⁻¹, respectively, and the two figures were approximately equal to the average background values of Cu in Suzhou city (Table 2). The Pb concentration in the G horizon of the paddy soil profile was slightly higher than that in the C horizon of the grey fluvo-aquic soil (Fig. 10), indicating that the parent materials might have considerable influences on the Pb concentrations in soils of the study area. However, when quantifying the potential areas of Pb contamination due to the industrialization process over the past 20 years, the average background value of Pb of Suzhou city (Table 2) was still reliable in representing the Pb background levels because this value was obtained immediately prior to the industrialization of the study area. The profile used for analysis here is just one profile for each of the two main soil types and the search for a suitable profile site without influences of factories in or near the study area was very difficult due to the factories being intermixed with farmland and residential areas in this region.

It should be pointed out that this study only dealt with the spatial patterns of total Cu and Pb concentrations in soils. However, for better understanding the sources of heavy-metal pollution in the area, studies on element bio-availability in soils should be intensified because element bioavailable parts are a primary factor affecting element pollution in plant and water bodies in the environment instead of the total components.

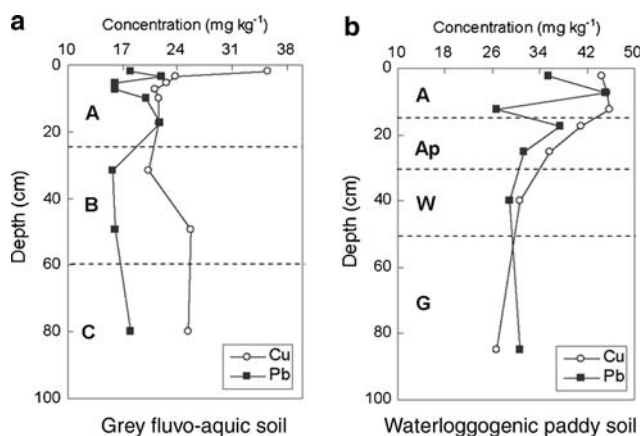


Fig. 10 Vertical distribution of Cu and Pb in profiles of grey fluvo-aquic soils and waterloggogenic paddy soils

Conclusions

Great variations exist in Cu and Pb concentrations in topsoils of the study area. The variability of Cu and Pb concentrations at the local hotspots was strongly associated with various types of factories, and the anthropogenic input of Cu into soils around the hotspots is closely related to the emissions of printing and dyeing and metallurgical factories, whereas the lead oxide factories and chemical factories resulted in the significant increase of Pb in topsoils. Approximately 30% of the total land area of the study area is of potential risk resulting from the non-point sources of contamination of Cu and Pb due to the rapid industrialization over the past 20 years.

It should be noted that the economy of the study area is still rapidly expanding and a decrease of soil pH was also observed in the area over the past 20 years (Shao et al. 2006), which may increase the concentrations of available Cu in soils, thereby enhancing the accumulations of Cu in crops or vegetables. Moreover, although Pb in soils absorbed by crops or vegetables was very limited, the accumulations of Pb in soils may also increase the risks of people exposed directly to soils. Therefore, effective managements and powerful control measures should be taken immediately and persistently in the area.

Acknowledgments Funding provided by the National Natural Science Foundation of China (40601039), the National Key Basic Research Support Foundation of China (2002CB410810) and the Knowledge Innovation Program of Chinese Academy of Sciences (ISSASIP0604). We gratefully thank Dr Murray Lark (Rothamsted Research) and Dr Barry Rawlins (British Geological Survey) for their help in providing reprints of the related papers.

References

Agricultural Environment Background Value Research Group (1997) Studies on agricultural environment background values of China

(in Chinese). Shanghai Science and Technology Press, Shanghai, pp 165

- Atteia O, Dubois JP, Webster R (1994) Geostatistical analysis of soil contamination in the Swiss Jura. *Environ Pollut* 86:315–327
- Burt R, Wilson MA, Mays MD, Lee CW (2003) Major and trace elements of selected pedons in the USA. *J Environ Qual* 32:2109–2121
- Cattle JA, McBratney AB, Minasny B (2002) Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. *J Environ Qual* 31:1576–1588
- Chang AC, Warneke JE, Page AL, Lund LJ (1984) Accumulation of heavy metals in sewage sludge-treated soils. *J Environ Qual* 13:87–91
- Cressie N, Hawkins D (1980) Robust estimation of the variogram. *Math Geol* 12:115–125
- Deutsch CV, Journel AG (1998) GSLIB, Geostatistical software library and user's guide. Oxford University Press, New York
- Dowd PA (1984) The variogram and kriging: robust and resistant estimators. In: Verly G, David M, Journel AG, Marechal A (eds) *Geostatistics for natural resources characterization (Part 1)*. Reidel, Dordrecht, pp 91–106
- Facchinelli A, Sacchi E, Mallen L (2001) Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. *Environ Pollut* 114:313–324
- Franssen HJWMH, van Eijnsbergen AC, Stein A (1997) Use of spatial prediction techniques and fuzzy classification for mapping soil pollutants. *Geoderma* 77:243–262
- Furrer R, Genton MG (1999) Robust spatial data analysis of lake Geneva sediments with S+SpatialStats. *Int J Syst Res Inf Sci Spec Issue Spat Data Anal Model* 8:257–272
- Genton MG (1998) Highly robust variogram estimation. *Math Geol* 30:213–221
- Goovaerts P (1999) Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89:1–45
- Goovaerts P (2001) Geostatistical modeling of uncertainty in soil science. *Geoderma* 103:3–26
- Juang KW, Chen YS, Lee DY (2004) Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils. *Environ Pollut* 127:229–238
- Kuo S, Heilman PE, Baker AS (1983) Distribution and forms of copper, zinc, cadmium, iron, and manganese in soils near a copper smelter. *Soil Sci* 135:101–109
- Lark RM (2000) A comparison of some robust estimators of the variogram for use in soil survey. *Eur J Soil Sci* 51:137–157
- Lark RM (2002) Modelling complex soil properties as contaminated regionalized variables. *Geoderma* 106:173–190
- Li Z, Shuman LM (1996) Heavy metal movement in metal-contaminated soil profiles. *Soil Sci* 161:656–666
- Meuli R, Schulin R, Webster R (1998) Experience with the replication of regional survey of soil pollution. *Environ Pollut* 101:311–320
- Militino AF, Palacios MB, Ugarte MD (2006) Outliers detection in multivariate spatial linear models. *J Stat Plan Inference* 136:125–146
- Mowrer H T (1997) Propagating uncertainty through spatial estimation processes for old-growth subalpine forests using sequential Gaussian simulation in GIS. *Ecol Modell* 98:73–86
- R Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- Rawlins BG, Lark RM, O'Donnell KE, Tye AM, Lister TR (2005) The assessment of point and diffuse metal pollution of soils from an urban geochemical survey of Sheffield, England. *Soil Use Manag* 21:353–362

- Ribeiro PJ Jr, Diggle PJ (2001) GeoR: A package for geostatistical analysis. R-NEWS, vol 1, No 2, pp 15–18. ISSN 1609–3631
- Shao XX, Huang B, Gu ZQ, Qian WF, Jin Y, Bi KS, Yan LX (2006) Spatial-temporal variation of pH values of soils in a rapid economic developing area in the Yangtze River Delta region and their causing factors (in Chinese). *Bull Miner Petrol Geochem* 25(2):143–149
- Soil Survey Office of Shazhou County (SSOSC) (1984) Description of Soils in Shazhou County (in Chinese). Agriculture Bureau of Suzhou City, Jiangsu Provincial Office for Soil Survey
- Soil Taxonomy Research Group (STRG) (2001) Key to Chinese Soil Taxonomy (in Chinese). University of Science and Technology of China Press, Hefei
- State Environmental Protection Administration of China (1995) Chinese Environmental Quality Standard for Soils (GB 15618-1995) <http://www.zhb.gov.cn/eic/650208300142493696/19951206/1023470.shtml>
- Sterckeman T, Douay F, Proix N, Fourier H (2000) Vertical distribution of Cd, Pb and Zn in soils near smelters in the north of France. *Environ Pollut* 107:377–389
- Van Meirvenne M, Goovaerts P (2001) Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. *Geoderma* 102:75–100
- Wang GX, Gertner G, Liu XZ, Anderson A (2001) Uncertainty assessment of soil erodibility factor for revised universal soil loss equation. *Catena* 46:1–14
- Webster R, Oliver M (2000) *Geostatistics for environmental scientists*. Wiley, New York
- Zhang CS, Selinus O (1997) Spatial analyses for copper, lead and zinc contents in sediments of the Yangtze River basin. *Sci Total Environ* 204:251–262