The use of spatial autocorrelation analysis to identify PAHs pollution hotspots at an industrially contaminated site

Geng Liu · Rutian Bi · Shijie Wang · Fasheng Li · Guanlin Guo

Abstract The identification of contamination "hotspots" are an important indicator of the degree of contamination in localized areas, which can contribute towards the re-sampling and remedial strategies used in the seriously contaminated areas. Accordingly, 114 surface samples, collected from an industrially contaminated site in northern China, were assessed for 16 polycyclic aromatic hydrocarbons (PAHs) and were analyzed using multivariate statistical and spatial autocorrelation techniques. The results showed that the PCA leads to a reduction in the initial dimension of the dataset to two components, dominated by Chr, Bbf&Bkf, Inp, Daa, Bgp, and Nap were good representations of the 16 original PAHs; Global Moran's I statistics indicated that the significant autocorrelations were detected and the autocorrelation distances of six indicator PAHs were 750, 850, 1,200, 850, 750, and 1,200 m, respectively; there were visible high–high values (hotspots) clustered in the mid-bottom part of the site through the Local Moran's I index analysis. Hotspot identification and spatial distribution results can play a key role in contaminated site investigation and management.

Abbreviations

PAH	Polycyclic aromatic hydrocarbon
Nap	Naphthalene
Acy	Acenaphthylene
Ace	Acenaphthene
Fle	Fluorine
Phe	Phenanthrene
Ant	Anthracene
Fla	Fluoranthene
Pyr	Pyrene
Baa	Benzo(a)anthracene
Chr	Chrysene
Bbf&Bkf	Benzo(b,k)fluoranthene
Bap	Benzo(a)pyrene
Daa	Dibenzo(a,h)anthracene
Bgp	Benzo(g,h,i)perylene
Inp	Indeno(1,2,3-c,d)pyrene
PCA	Principal component analysis

G. Liu · S. Wang · F. Li · G. Guo (✉)
State Key Laboratory of Environmental Criteria and Risk
Assessment, Chinese Research Academy of Environmental
Sciences, Anwai Dayangfang 8, Beijing 100012,
People's Republic of China
e-mail: guoguanlin09@yahoo.com.cn

G. Liu · R. Bi
School of Resources and Environment,
Shanxi Agricultural University, Taigu 030801,
People's Republic of China

Introduction

Soil contamination of industrially contaminated sites has received wide attention in China in recent years

(Luo et al. 2009; Wang et al. 2010). The process of industrial production and anthropogenic emissions is the major source of PAHs in the atmosphere and contaminated soils seriously in sites (Wang et al. 2003). It has been confirmed that soils in large-scale coking contaminated sites have been severely contaminated by PAHs (Wang et al. 2012). The PAHs present in contaminated sites not only directly affects soil physicochemical properties and the environment, but also threaten human health in the contaminated area (Blake et al. 2007; Mostert et al. 2012). The aims of removing PAHs from contaminated soils are urgently required in the risk management and remediation of contaminated sites. In some large-scale abandoned industrially contaminated sites in China, the historical production layout, storage, management, and other factors have led to extreme contamination values, and they have varying levels of contamination across the sites. Industrial site contamination of PAHs is typically characterized by significant variability in contaminant concentrations. This means that it is often difficult to describe spatial patterns and identify pollution hotspots.

Following remedial investigations of contaminated sites, hotspots of serious contamination have been found and these have an important role in the formulation of remediation strategies (Sinha et al. 2007). Hotspot areas also need to be identified in order to develop effective management practices for soil contamination (Zhang et al. 2008). There are many definitions of what is meant by "hotspots" (Cressie 1991; Komnitsas et al. 2009) and several methods have been proposed to identify hotspots (Kulldorff 1997; Carson 2001; Patil et al. 2001). Spatial autocorrelation analysis approaches may also be useful tools for identifying hotspots and spatial patterns of pollution. A few spatial autocorrelation analysis methods have been suggested for hotspot or spatial cluster pattern identification, such as Moran's I (Jing et al. 2010; Prasannakumar et al. 2011), Geary's C (Tiefelsdorf et al. 1997; Barreto-Neto et al. 2004), Getis' G (Lai et al. 2009; Bohórquez et al. 2011), and Join Count analysis (Kabos et al. 2002; Epperson 2003), but Moran's I index seems to be the most widely used (Getis et al. 1996). Moran's I method has been used in a number of research fields, including environmental management (Zhang et al. 2004; Brody et al. 2006; Castillo et al. 2011; Su et al. 2011). However, within the geoscientific variables of a large-scale industrially contaminated site, this method is relatively untried.

This study used Global Moran's I and Local Moran's I indices of spatial autocorrelation to analyze the statistical structure characteristics of samples from a contaminated site. Global Moran's I was used to describe the correlation distance and spatial pattern of soil PAHs. The standardized spatial autocorrelation coefficients of different spatial lags were arranged in order and mapped on to standardized spatial correlograms, which were used to show spatial dependence. The Local Moran's I index was used to calculate the relationship between each sample and its neighbors, which allowed hotspots and cool spots for soil PAHs to be identified at the contaminated site. The research has provided important information about contamination at the site that can be used at other sites to aid remediation investigations and the creation of remediation strategies for large-scale industrially contaminated sites.

Materials and methods

Site characteristics

The plant is located in the north of China and has a total land area of 1.35 km^2. It was one of China's large-scale independent coking chemical industrial enterprises that had been in operation for over 40 years. The major products were coke, coke oven gas, tar, benzene, ammonium sulfate, asphalt and naphthalene, amongst a total of 40 chemical products. Serious environmental pollution was caused by poor production technology and pollution controls during its initial period, and carcinogenic substances and harmful emissions have damaged the area taken up by the plant and its surrounding environment. The plant took in coal as a raw material to produce gas and coke and extracted coal-based chemicals from crude tar. On site, the facility consisted of a former coal preparation factory, coking factory, sieve coke factory, gas purification system, tar factory, and a gas factory. Spills and leakages during gas purification and refining, tar processing, storage, transport and production processes, as well as planned and unplanned emissions, were the major causes of the field pollution characteristics shown by the polycyclic aromatic hydrocarbons and associated compounds. Soil contamination at the site is serious, based on site reconnaissance and relevant data analysis. It is now imperative to undertake a site risk assessment, contamination management, and other related work.

Sampling and chemical analyses

Based on the plant's historical production, management and technological layout, a total of 64 monitoring points were first sampled in an effort to pick soil sampling points in areas of potentially contaminated ground and to confirm known contamination areas, contamination depth, and contaminant species. A second sampling grid and monitoring points were combined with the first in order to provide a comprehensive picture of contamination at the site. The monitoring points and first analysis focused on the seriously polluted areas. A total of 114 effective soil sampling points were used in both sampling visits.

For the extraction of PAHs, 10 g of soil sample was extracted with acetone/dichloromethane (1/1, v/v) by ASE-300 (Dionex, Beijing, China). The extracted PAHs were concentrated by organomation and eluted with approximately 30 mL dichloromethane/n-hexane (2/1, v/v) and concentrated to 1 mL for analysis (Grimalt et al. 2004). PAHs in the extracts of all samples were analyzed by gas chromatography–mass spectrometry [Agilent, 6890N GC, 5975B mass spectrometric detector, USA] equipped with an HP-5MS capillary column (30 m, 0.25 mm inner diameter×0.25 mm film thickness, Agilent, USA). In this method, the identification of 16 priority PAHs was performed by gas chromatography–mass spectrometry, and quantification analysis was based on peak area external reference of 16 PAH standard sample (Supelco Co., USA) containing Nap, Acy, Ace, Fle, Phe, Ant, Fla, Pyr, Baa, Chr, Bbf&Bkf, Bap, Daa, Bgp, Inp, the analytical procedure was comprehensively evaluated against quality control acceptance criteria (USEPA 2007a), the linear quantitative equation was obtained with $r^2>0.99$ (USEPA 2007b), and the method detection limits ranged from 10 to 15 µg kg^{-1}, while the recoveries were 50 % to 105 % with relative standard deviation lower than 11 %.

Spatial autocorrelation

Spatial autocorrelation refers to the correlation of the same variable in different spatial positions and the relationship between a location variable value and nearby values. Ordinary correlation refers to the mutual relationship between two or more variables. Spatial autocorrelation includes spatial clusters and spatial outliers. A spatial cluster refers to a variable with high concentration surrounded by variables that also have high concentrations. In contrast, a spatial outlier means a high variable concentration surrounded by variables with low concentrations. Once the variables show a certain spatial regularity, there is spatial autocorrelation between them.

Spatial autocorrelation uses global and local indicators. Global spatial autocorrelation describes the overall distribution of a variable in order to determine if spatial clusters of this variable exist over a larger area and uses a single value to reflect the region's degree of autocorrelation. Local spatial autocorrelation computes each spatial unit in relation to neighboring units on a property. There have been many methods proposed for the calculation of spatial autocorrelation, such as Moran's I, Geary's C, Getis' G, Join Count et al.'s Express Global spatial autocorrelation characteristic and Moran's I and Getis' G et al.'s Express Local spatial autocorrelation characteristics. Moran's I seems to be the most often used, so this study identified pollution hotspots at the large coking contaminated site in north China using Moran's I index.

Global Moran's I index

Global Moran's I was introduced by Moran in 1948 and examines whether spatial correlation exists or not over an entire region. The Global Moran's I statistic for spatial autocorrelation is given as (Cliff et al. 1981):

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

where n is the number of the variable value for the whole region, x_i and x_j are the variable values at locations: i and j, $\bar{x}$ is the mean of x, C is the spatial weights matrix and C_{ij} is the spatial weight between locations: i and j. The value of Moran's I usually varies between −1 and 1. A value that approximates to −1 indicates negative autocorrelation and a value that approximates to 1 indicates positive autocorrelation. No autocorrelation results in a value close to 0.

Generally, normal approximation is a precondition, so the Global Moran's I index can be standardized to $Z(I)$ as follows (Cliff et al. 1981):

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}$$

where
$$E(I) = -\frac{1}{(n-1)}; Var(I) = \frac{1}{w_0^2(n^2-1)}\left(n^2 w_1 - n w_2 + 3 w_0^2\right) - E^2(I)$$
$$w_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}; w_1 = 0.5\sum_{i=1}^{n}\sum_{j=1}^{n}\left(w_{ij}+w_{ji}\right)^2; w_2 = \sum_{i=1}^{n}\left(w_{i*}+w_{*i}\right)^2$$

where w_{i*} is the sum of all weights located in the row i and w_{*i} is the sum of all weights in column i. Following significance testing on the I value (at 0.05 level), if $Z(I)$ is greater than 1.96 or smaller than -1.96, then it signifies an existing significant positive or negative spatial autocorrelation, respectively.

Local Moran's I index

Local Moran's I is the index used to describe the locations of spatial clusters and spatial outliers. This study used the Local Moran's I index to identify spatial clusters and outliers. The function is as follows:

$$I_i = \frac{\left(x_i - \bar{x}\right)}{S^2}\sum_{i=1}^{n} C_{ij}\left(x_i - \bar{x}\right)$$
$$S^2 = \frac{1}{n-1}\sum_{\substack{j=1 \\ j=i}}^{n} C_{ij}\left(x_i - \bar{x}\right)^2$$

The notations of the methods are similar to that for Global Moran's I. Local Moran's I can be standardized, and the significance of a spatial autocorrelation can be tested. The methods are similar with that used for the Global Moran's I.

Data analyses using computer software

The calculation of the spatial autocorrelation index was conducted using OpenGeoDa 1.0.1 and STARS 0.8.2 software. The maps were produced using OriginPro 7.5 and ArcGis 10.0 software. Multivariate and statistical analyses on the sample data were performed using SPSS14.0.

Results and discussion

Correlation analysis for PAHs

Correlation analysis can be used to test similarity among PAH sample concentration data. The correlation analysis of different pollutants helps analyze the cause of soil contamination and identify pollutant sources and typical pollutants. The data for PAHs in soils from the contaminated coking plant site used in this study varied greatly and were highly skewed, which resulted in a non-normal distribution, so the Spearman statistical test was used to analyze the correlation. Table 1 presents the coefficients between soil PAHs. The results showed that all the PAHs were highly correlated (correlation is significant at the 0.05 level). The higher correlations between soil PAHs may indicate that these PAHs had similar sources. Although significantly positive correlations existed between PAHs, the correlation coefficients were different. For example, the correlations for Nap and Acy and Nap and Ant were significant at $r = 0.814$ and $r = 0.674$, respectively (at the 0.05 level) but different. A high correlation coefficient represents a greater correlation between the contaminants.

Principal component analysis of PAHs

PCA is a useful statistical technique, which can reduce the initial dimension of a large dimension dataset and is a common technique for finding patterns in a dataset. The pollution caused by the 16 different PAHs analyzed in this study was not the same, and the sample numbers that exceeded the threshold concentration were also different. Two principal components (PC), with eigenvalues higher than 1, were extracted. Their eigenvalues were 9.92 and 2.51. The PCA method led to a reduction in the original 16 PAHs to two components that could explain 84.52 % of the data variation. Representation of the variables by the two components was adequate, considering the commonalities obtained in the analysis. In the rotated principal component analysis of the matrix (Table 2), the first principal component was dominated by Chr, Bbf&Bkf, Inp, Daa and Bgp, and the second principal component was dominated by Nap, which meant that these six PAHs were representative of the site. In the scattergram of "factor scores" (Fig. 1), principal component 1 is the X-axis and principal component 2 is the

Table 1 Correlation analysis results for PAHs

		Nap	Acy	Ace	Fle	Phe	Ant	Fla	Pyr	Baa	Chr	Bbf&Bkf	Bap	Inp	Daa	Bgp
Nap	Spearman	1.000														
	Sig. (2-tailed)	–														
Acy	Spearman	.830*	1.000													
	Sig. (2-tailed)	.000	–													
Ace	Spearman	.814*	.744*	1.000												
	Sig. (2-tailed)	.000	.000	–												
Fle	Spearman	.812*	.808*	.870*	1.000											
	Sig. (2-tailed)	.000	.000	.000	–											
Phe	Spearman	.809*	.851*	.713*	.782*	1.000										
	Sig. (2-tailed)	.000	.000	.000	.000	–										
Ant	Spearman	.828*	.922*	.775*	.810*	.891*	1.000									
	Sig. (2-tailed)	.000	.000	.000	.000	.000	–									
Fla	Spearman	.674*	.796*	.580*	.648*	.886*	.836*	1.000								
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	–								
Pyr	Spearman	.746*	.858*	.666*	.702*	.925*	.928*	.912*	1.000							
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	–							
Baa	Spearman	.659*	.781*	.576*	.620*	.862*	.850*	.967*	.920*	1.000						
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	–						
Chr	Spearman	.743*	.871*	.670*	.706*	.906*	.920*	.898*	.967*	.907*	1.000					
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	–					
Bbf&Bkf	Spearman	.691*	.833*	.585*	.631*	.887*	.880*	.890*	.951*	.892*	.952*	1.000				
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	–				
Bap	Spearman	.698*	.804*	.611*	.644*	.849*	.840*	.925*	.877*	.936*	.861*	.868**	1.000			
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	–			
Inp	Spearman	.699*	.812*	.607*	.659*	.833*	.833*	.912*	.859*	.920*	.856*	.857**	.981*	1.000		
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	–		
Daa	Spearman	.707*	.864*	.649*	.717*	.825*	.904*	.812*	.893*	.832*	.923*	.895**	.836*	.847*	1.000	
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	–	
Bgp	Spearman	.675*	.816*	.560*	.644*	.873*	.849*	.858*	.918*	.874*	.913*	.939**	.871*	.873*	.898*	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	–

*Correlation is significant at the 0.05 level, two-tailed

Y-axis. The scattergram shows that for division number 2 for sample numbers 13, 27, 39 and 89, principal component 2 gave the highest scores and principal component 1 gave relatively low scores. The corresponding concentration for Nap seriously exceeded acceptable levels. For division number 1 of the sample numbers, principal component 1 gave the highest scores and principal component 2 gave relatively low scores, which showed that the first component's five PAHs showed similar levels of contamination.

Status of PAH concentrations and identification of outliers

Table 3 shows a number of statistical parameters, including skewness, kurtosis, SD and CV, for the concentrations of PAHs in soils at the contaminated site. The concentration ranges were quite wide. For example, Nap concentrations in the 114 samples ranged from 0.01 to 4,100.00 mg kg^{-1}, with a mean of 100.91 mg kg^{-1}. The data had high CVs and skewness. This implied that the original data included extreme values and were not normally or lognormally distributed, according to the Kolmogorov–Smirnov (K-S) test, because all the K-S *P* parameters were 0.00. In contaminated site investigation and remediation, only extremely large values have a major effect on the statistics and are of concern. In this study, the box-and-whiskers plot method was used to identify extreme values (extreme values were defined as those which were greater than the sum of the mean values and at least three times the standard deviation.).

Global spatial autocorrelation analysis

In general, the higher the absolute value of Moran's I, the stronger the spatial autocorrelation; the larger the absolute value of standardized Moran's I, the more significant the spatial structure is (Huo et al. 2012). Figure 2 shows the standardized spatial correlograms for the soil PAHs. The *Y*-axis is the standardized Moran's I and the *X*-axis is the lag distance. The values of the spatial autocorrelation for regionalized variables can be obtained from the standardized Moran's I map. The data can also be used to estimate if spatial clusters or spatial outliers exist in the study area and to compare the significant spatial patterns of the variables. Generally, there were more than two positive autocorrelations in the standardized Moran's I map (the nearer distance represents the spatial correlation distance of the variable). Figure 2 provides the standardized Moran's I values for the six PAHs, and the results show the standardized Moran's I values for

Table 2 Principal component matrix analysis for PAHs

| | Component | | Rotated Component | |
	PC1	PC2	PC1	PC2
Nap	0.567	0.668	0.180	0.858
Acy	0.658	0.474	0.352	0.731
Ace	0.121	0.438	−0.102	0.443
Fle	0.405	0.634	0.053	0.751
Phe	0.765	0.525	0.422	0.726
Ant	0.810	0.470	0.488	0.800
Fla	0.961	0.113	0.791	0.558
Pyr	0.984	−0.002	0.766	0.468
Baa	0.982	−0.101	0.728	0.380
Chr	0.981	−0.153	0.935	0.334
Bbf&Bkf	0.932	−0.336	0.979	0.150
Bap	0.841	−0.325	0.694	0.115
Inp	0.904	−0.387	0.979	0.091
Daa	0.911	−0.361	0.973	0.118
Bgp	0.837	−0.450	0.950	0.004
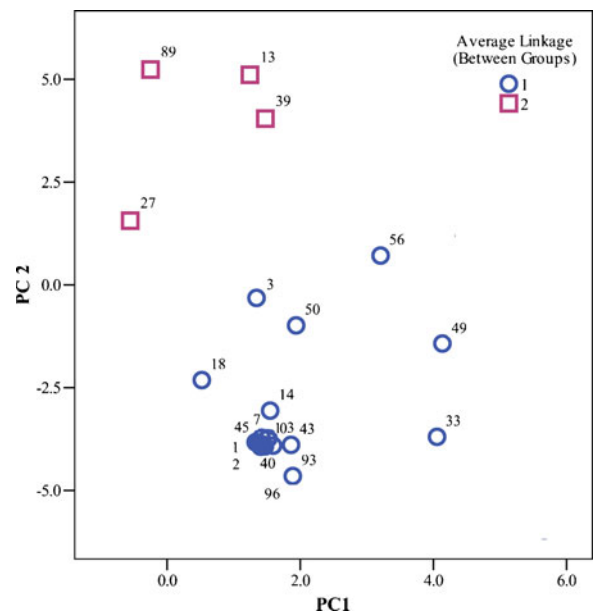
Fig. 1 PCA scattergram of "factor scores" for each sample location

Table 3 Summary statistics for the original PAH data

	Nap	Chr	Bbf&Bkf	Inp	Daa	Bgp
Minimum	0.01	0.01	0.01	0.01	0.01	0.01
Maximum	4,100.00	175.00	393.00	144.00	45.70	160.00
Mean	100.91	6.69	12.51	4.65	1.57	4.81
Median	0.04	0.25	0.44	0.26	0.09	0.20
Skewness	6.07	4.96	5.51	5.59	5.25	5.88
Kurtosis	40.48	26.31	35.17	36.22	31.30	39.72
CV	5.01	3.83	3.92	3.83	3.78	4.04
SD	505.93	25.63	49.04	17.82	5.93	19.42
K-S test P	0.00	0.00	0.00	0.00	0.00	0.00

Concentration, milligrams per kilogram, $n=114$

SD standard deviation, *CV* coefficients of variation, *K-S* Kolmogorov–Smirnov test

Nap, which were positive at a distance from 350 to 750 m and 1,850 to 2,100 m. This meant that spatial clusters for Nap concentrations existed in these ranges. The standardized Moran's I values for Nap were negative at a distance from 800 to 1,750 m, which implied spatial outliers existed in this range. Similarly, five other PAHs showed spatial clusters and spatial outliers in the study area. With the exception of Bgp, the amplitudes of the spatial outliers were larger than they were for the spatial clusters, which meant that negative spatial autocorrelation existed in the study area. The autocorrelation distances for Nap, Chr, Bbf&Bkf, Inp, Daa, and Igp were 750, 850, 1,200, 850, 750, and 1,200 m, respectively.
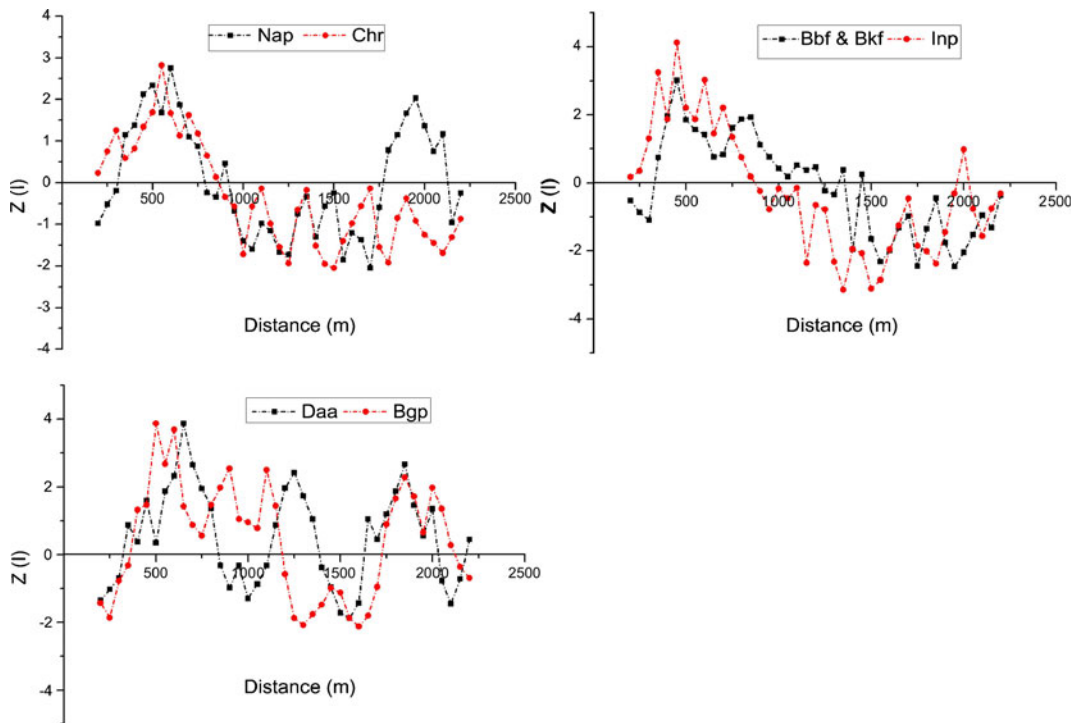
Fig. 2 Standardized spatial correlograms for soil PAHs

Using the Local Moran's I index to identify hotspots

Hotspots and cool spots for soil PAH distribution can be identified using the Local Moran's I index method.

Cool spots are often considered as clean areas, but soil PAH pollution studies of this particular contaminated site focused on the identification of high pollution risk areas, so it was important to identify hotspots if the
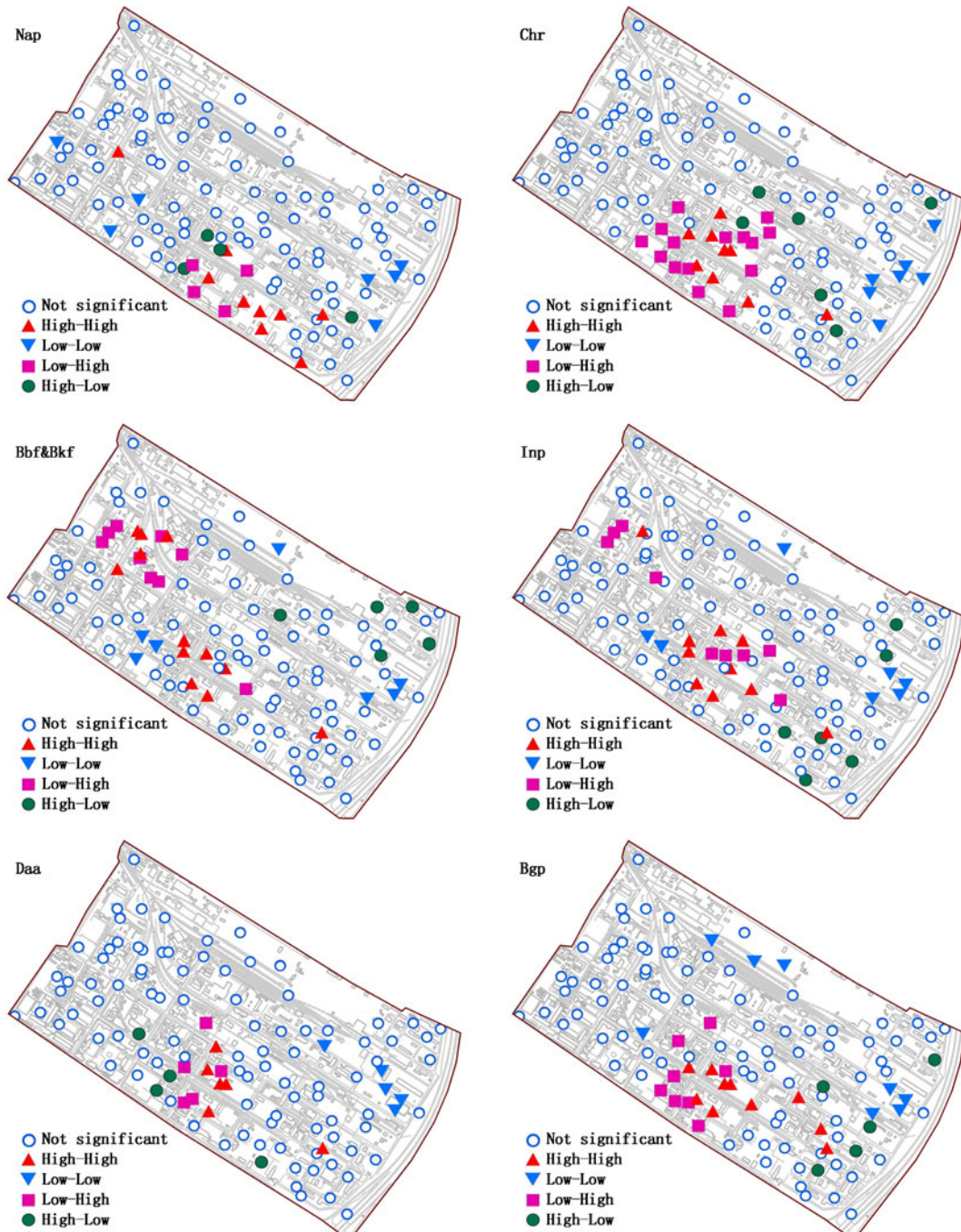

Fig. 3 Local spatial cluster characteristics for soil PAHs

degree of pollution at the site was to be quantified. The Local Moran's I index was calculated based on a distance band of 300 m. The results are shown in Fig. 3. It was found that all soil PAHs showed significant spatial clusters and spatial outliers. The high–high values showed hotspots and low–low values showed cool spots. For Nap, there were nine high–high values, clustered in the mid-bottom part of the site and seven low-low spatial clusters. There were also several low–high and high–low outliers identified in the mid-bottom part of the site. For Chr, the number of high–high values and low–low values were also nine and seven, but the number of low–high spatial outliers increased to 15 and most of samples were located in the eastern part of the site. Bbf&Bkf and Daa had 12 and six high-high values that were clustered, respectively. Bgp and Daa, which had eight and five low–low values that were clustered, respectively, were the most and fewest low–low clusters of the six PAHs.

All the PAHs had four kinds of spatial clusters. Although they had different distributions, pollution hotspots were visible in the mid-bottom part of the site, where soil PAH samples with high concentrations were surrounded by samples with similarly high PAHs concentrations. Taking the hotspot distribution results into consideration, it seems that the hotspot distribution results were consistent with the contamination that would have been caused by past processing operations at the contaminated site. Overall, areas at the mid-bottom part of the site were mainly where the production process workshops for coking, gas purification, tar products, etc. were situated and were expected to be seriously contaminated. It was to be expected that nearly all of the high–high values would be found in these areas. A relationship can be established between where the pollution hotspots were found and the influencing factors based on the occurrence and distribution of processes around the contaminated site. The areas of the site mainly contained coal preparation areas and gas purification workshops, and these areas did not show serious contamination. Relatively low–low values were found in these areas.

Zhang et al. (2008) used Local Moran and GIS results to identify pollution hotspots for Pb in urban soils. He found that different distance bands produced different results, so he suggested that a number of factors should be considered when choosing suitable distance bands. A distance band of 300 m was used in this study. The distance between samples was 80–200 m. The spatial interpolation of un-sampled areas at a contaminated site is often based on the weighted average of the nearest neighbors, so 300 m was chosen as weight Matrix for this study.

Pollution hotspots in large-scale abandoned industrially contaminated sites need to be identified for remediation investigations. The knowledge of hotspots and the statistical characteristic of hotspots are useful for site environmental management. This study showed that the Global Moran's I and Local Moran's I indices of spatial autocorrelation are reliable tools for classifying spatial clusters and spatial outlier characteristics of PAH concentration data sets, and also for identifying hotspots of PAH pollution in contaminated sites. The research has provided important information about hotspot identification for the environmental risk management of large-scale industrially contaminated sites.

Conclusions

Principal component analysis and correlation analysis were applied to 114 samples, in which the total concentrations of 16 PAHs were measured at a contaminated large-scale independent coking chemical industrial site in China. Two significant components were extracted by PCA, which explained 84.52 % of the total variance. The analysis showed that Chr, Bbf&Bkf, Inp, Daa, Bgp, and Nap were representative of the whole site.

The results of the global spatial autocorrelation analysis showed that soil PAH distribution in soil at the contaminated site was not random, but there were some significant spatial autocorrelation features. The spatial positive correlation ranges for Nap, Chr, Bbf&Bkf, Inp, Daa, and Igp were found at 750, 850, 1,200, 850, 750, and 1,200 m, respectively.

Different significant spatial structures existed for the PAH distribution in the target site according to the Local Moran's I index analysis. The pollution hotspots were located in the mid-bottom part of the site, which contained workshops of coking, gas purification, tar products, which were suspected to be the seriously contaminated areas in the site. Pollution hotspots at this contaminated site needed to be identified so that efficient environmental remediation strategies can be developed.

References

Barreto-Neto, A. A., & Silva, A. B. D. S. (2004). Methodological criteria for auditing geochemical data sets aimed at selecting anomalous lead–zinc–silver areas using geographical information systems. *Journal of Geochemical Exploration, 84*(2), 93–101.

Blake, W. H., Walsh, R. P., Reed, J. M., Barnsley, M. J., & Smith, J. (2007). Impacts of landscape remediation on the heavy metal pollution dynamics of a lake surrounded by non-ferrous smelter waste. *Environmental Pollution, 148*(1), 268–280.

Bohórquez, L., Gómez, I., & Santa, F. (2011). Methodology for the discrimination of areas affected by forest fires using satellite images and spatial statistics. *Procedia Environmental Sciences, 7*, 389–394.

Brody, S. D., Highfield, W. E., & Thornton, S. (2006). Planning at the urban fringe: an examination of the factors influencing nonconforming development patterns in southern Florida. *Environment and Planning B: Planning and Design, 33*(1), 75–96.

Carson, J. H. (2001). Analysis of composite sampling data using the principle of maximum entropy. *Environmental and Ecological Statistics, 8*(3), 201–211.

Castillo, K. C., Körbl, B., Stewart, A., Gonzalez, J. F., & Ponce, F. (2011). Application of spatial analysis to the examination of dengue fever in Guayaquil, Ecuador. *Procedia Environmental Sciences, 7*, 188–193.

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: models and applications*. London: Pion.

Cressie, N. (1991). *Statistics for spatial data*. New York: Wiley.

Epperson, B. K. (2003). Covariances among join-count spatial autocorrelation measures. *Theoretical Population Biology, 64*(1), 81–87.

Getis, A., & Ord, J. K. (1996). Local spatial statistics: an overview. In P. Longley & M. Batty (Eds.), *Spatial analysis: modelling in a GIS Environment*. Cambridge: GeoInformation International.

Grimalt, J. O., van Drooge, B. L., Ribes, A., Fernández, P., & Appleby, P. (2004). Polycyclic aromatic hydrocarbon composition in soils and sediments of high altitude lakes. *Environmental Pollution, 131*, 13–24.

Huo, X. N., Li, H., Sun, D. F., Zhou, L. D., & Li, B. G. (2012). Combining geostatistics with Moran's I analysis for mapping soil heavy metals in Beijing, China. *International Journal of Environmental Research and Public Health, 9*(3), 995–1017.

Jing, N., & Cai, W. X. (2010). Analysis on the spatial distribution of logistics industry in the developed East Coast Area in China. *The Annals of Regional Science, 45*(2), 331–350.

Kabos, S., & Csillag, F. (2002). The analysis of spatial association on a regular lattice by join-count statistics without the assumption of first-order homogeneity. *Computers & Geosciences, 28*(8), 901–910.

Komnitsas, K., & Modis, K. (2009). Geostatistical risk estimation at waste disposal sites in the presence of hot spots. *Journal of Hazardous Materials, 164*(2–3), 1185–1190.

Kulldorff, M. (1997). A spatial scan statistic. *Commun. Statist. Theory Methods, 26*(6), 1481–1496.

Lai, D. J. (2009). Testing spatial randomness based on empirical distribution function: A study on lattice data. *Journal of Statistical Planning and Inference, 139*(2), 136–142.

Luo, Q., Catney, P., & Lerner, D. (2009). Risk-based management of contaminated land in the UK: lessons for China? *Journal of Environmental Management, 90*(2), 1123–1134.

Mostert, M. M. R., Ayoko, G. A., & Kokot, S. (2012). Multi-criteria ranking and source identification of metals in public playgrounds in Queensland, Australia. *Geoderma, 173–174*, 173–183.

Patil, G. P., & Taillie, C. (2001). Use of best linear unbiased prediction for hot spot identification in two-way compositing. *Environmental and Ecological Statistics, 8*(2), 163–169.

Prasannakumar, V., Vijith, H., Charutha, R., & Geeth, N. (2011). Spatio-temporal clustering of road accidents: GIS based analysis and assessment. *Procedia - Social and Behavioral Sciences, 21*, 317–325.

Sinha, P., Lambert, M. B., & Schew, W. A. (2007). Evaluation of a risk-based environmental hot spot delineation algorithm. *Journal of Hazardous Materials, 149*(2), 338–345.

Su, S. L., Jiang, Z. L., Zhang, Q., & Zhang, Y. (2011). Transformation of agricultural landscapes under rapid urbanization: a threat to sustainability in Hang-Jia-Hu region, China. *Applied Geography, 31*(2), 439–449.

Tiefelsdorf, M., & Boots, B. (1997). A note on the extremities of Local Moran's I is and their impact on Global Moran's I. *Geographical Analysis, 29*(3), 248–257.

USEPA. SW-846 method 8270D. (2007a). Semivolatile organic compounds by gas chromatography/mass spectrometry (GC/MS). http://www.epa.gov/osw/hazard/testmethods/sw846/pdfs/8270d.pdf. Accessed February 2007.

USEPA. SW-846 method 8272. (2007b). Parent and alky polycyclic aromatic in sediment pore water by solid -phase microextraction and gas chromatography/mass spectrometry in selected ion monitoring mode. http://www.epa.gov/osw/hazard/testmethods/pdfs/8272.pdf. Accessed December 2007.

Wang, J., Zhu, L. Z., & Shen, X. Y. (2003). PAHs pollution in air of coke plant and health risk assessment. *Environment Science, 24*(1), 135–138 (in Chinese).

Wang, J. G., Zhan, X. H., Zhou, L. X., & Lin, Y. S. (2010). Biological indicators capable of assessing thermal treatment efficiency of hydrocarbon mixture-contaminated soil. *Chemosphere, 80*(8), 837–844.

Wang, F. S., Xing, S. L., & Huo, X. R. (2012). Study on the distribution pattern of PAHs in the coking dust from the coking environment. *Procedia Engineering, 45*, 959–961.

Zhang, C. S., & McGrath, D. (2004). Geostatistical and GIS analyses on soil organic carbon concentrations in grassland of southeastern Ireland from two different periods. *Geoderma, 119*(3–4), 261–75.

Zhang, C. S., Luo, L., Xu, W. L., & Ledwith, V. (2008). Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Science of the Total Environment, 398*(1–3), 212–221.