A methodological framework for identifying potential sources of soil heavy metal pollution based on machine learning: A case study in the Yangtze Delta, China

Abstract

It is a great challenge to identify the many and varied sources of soil heavy metal pollution. Often little information is available regarding the anthropogenic factors and enterprises that could potentially pollute soils. In this study we use freely available geographical data from a search engine in conjunction with machine learning methodologies to identify and classify potentially polluting enterprises in the Yangtze Delta, China. The data were classified into 31 separate and five integrated industry types by five different machine learning approaches. Multinomial naive Bayesian methods achieved an accuracy of 86.5% and Kappa coefficient of 0.82 and were used to classify the geographic data from more than 250 000 enterprises. The relationship between the different industry classes and measurements of soil cadmium and mercury concentrations was explored using bivariate local Moran's I analysis. The analysis revealed areas where different industry classes had led to soil pollution. In the case of cadmium, elevated concentrations also occurred in some areas because of natural sources. This study provides a new approach to investigate the interaction between anthropogenic pollution and natural sources of soil heavy metals to inform pollution control and planning decisions regarding the location of industrial sites.

1. Introduction

Rapid economic and industrial development has led to the accumulation of heavy metals in the soil of impacted sites across the world [1]. Heavy metals generally have persistent bioavailability, long residence times (commonly exceeding decades), and often low concentration thresholds indicate toxicity [2]. The excessive accumulation of heavy metals can hence disrupt the usual biochemical processes which occur in soils, leading to deterioration of soil quality, reduced agricultural productivity and quality and human health risks [3-4]. According to the National Soil Pollution Condition Investigation Communique of China, 16.1% of soil samples are contaminated with heavy metals [5] and therefore detailed studies of soil contamination in China are required.

Human enterprises such as industry, transportation and agriculture can be the source of substantial quantities of heavy metals in soils [1]. According to the Statistical Yearbook of China in 2014, the number of registered and bankrupt enterprises in China were approximately 3.7 million and 3.0 million respectively (in combination, almost 30% of the enterprises in China). It is extremely difficult to make timely investigations and reports of the pollution effects of different enterprises across large regions using traditional methods, especially when the region is large and dispersed. The traditional source apportionment methods mainly include principal component analysis (PCA), isotope ratio analysis, positive matrix factorization (PMF) and stochastic models [6-8]. For example, Hu et al [9] analyzed seven environment variables relevant to the source and behavior of heavy metals using stochastic models, and Ma et al [10] researched the major potential source of soil heavy metals and human health risk using PCA in high population density area. These methods analyze the contribution of different sources to soil heavy metal pollution, but ignore the spatial distribution and characteristics of these sources [11-12]. Furthermore, the model mechanisms and data collection requirements of diffusion models of source apportionment for soil heavy metals are very complex, which is not convenient for wide uptake across large-scale regions. Exhaustive information regarding the location and type of enterprises within a region is rarely available.

In this study, we use freely available geographic information from a search engine to build an inventory of enterprises within the Yangtze Delta region of China. This geographic data does not specify the type of industry. We therefore survey a subset of the enterprise locations and form a training dataset of enterprise types. We test five different machine learning approaches to build a classifier of enterprise type and apply the best performing method to the full geographic dataset. We illustrate how these derived dataset might be utilized by using the bivariate local Moran's I method to analyze the spatial correlation between these enterprises and elevated soil metal concentrations and thus provide effective guidance and assistance for the management and control of these anthropogenic sources of pollution [13-15].

2. Materials and Methods

2.1 Study area

The study area (27° 2′ -31° 11′ N, 118° 01′ -123° 10′ E) is located in the Yangtze Delta of China, which covers 105 500 km^2 and has a population of 55.9 million. The study area possesses a typical subtropical monsoon climate, which is mild and humid with annual average temperature of 16.5 ℃ and annual average precipitation of 1 575 mm. The western, eastern and southern parts of study area are mainly red soil and yellow soil, and the southeast coastal and northern parts are mainly paddy soil. The industries in study area mainly include textile industry, chemical industry, metalwork industry. The Yangtze River Delta is one of the most developed regions in China and the concentrations of soil heavy metals are also remarkably high. According to Soil Pollution Condition Investigation Communique of the study area in 2013, the proportion of samples contaminated by the chromium (Cr), lead (Pb), cadmium (Cd), mercury (Hg), arsenic (As) elements were 0.87%, 0.24%, 15.63%, 10.94% and 1.03%, respectively. This study mainly focused on the source apportionment of Cd and Hg.

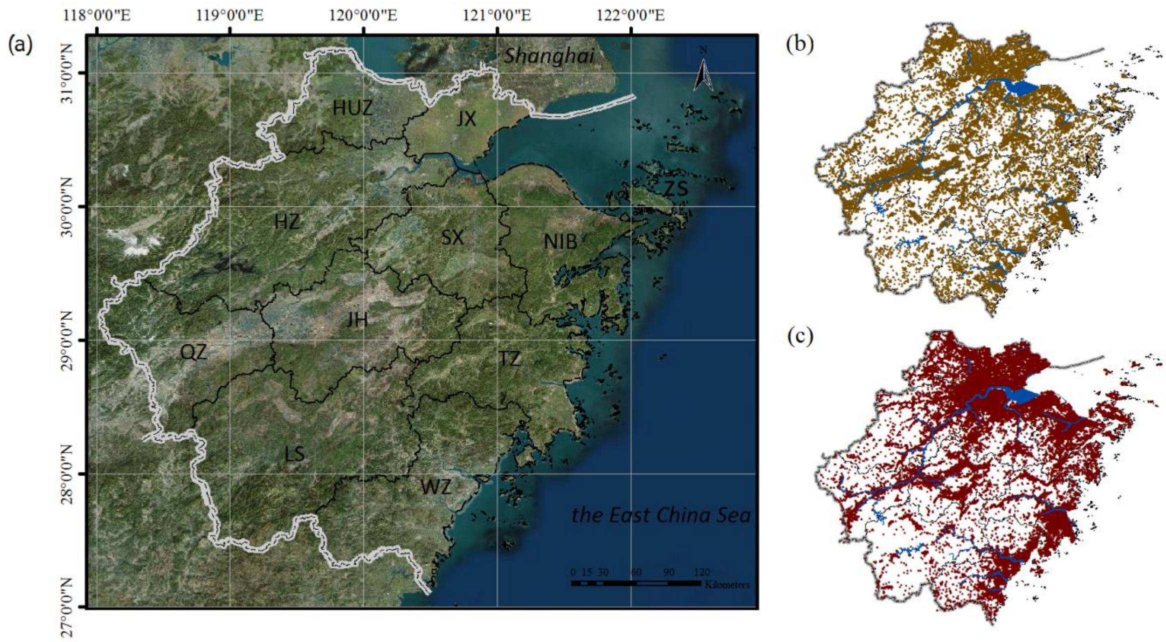

Figure 1. Maps showing the location of the study area, soil sampling and enterprise sites in the Yangtze Delta of China. a: HUZ, HZ, QZ, LS, JH, SX, JX, ZS, NIB, TZ, WZ were respectively the English abbreviations of the 11 provincial cities in the study area, b: the yellow points and the blue polygon respectively represented the 14801 soil samples and the river system, c: the red points represented the 264098 enterprises.

2.2 Soil sampling

A total of 14801 topsoil samples were collected from the study area in 2013 by the method of systematic grid sampling, making sample locations distributed as evenly as possible. Each soil sample was the bulked combination of five subsamples from five locations within five meters. All soil subsamples were collected at a depth of 0-20 cm using a stainless steel shovel. Fresh soil samples were transported to the laboratory, air-dried, ground, passed through a 2 mm sieve, and stored at room temperature. Soil pH was measured in H_2O with the soil and solution ratio of 1:2.5 (m/v) using the glass electrode method. The Cd element in soils was digested by HF-HNO_3-$HCLO_4$ and measured by an inductively coupled plasma-mass spectrometer (ICP-MS, Agilent 7500a, Palo Alto, CA, USA). The Hg element in soils was digested by HNO_3-HCl and determined by an atomic fluorescence spectrometer (Atomic Fluorescence Spectrometry, AFS) [5]. For quality control and quality assurance, blank control, duplicate samples, and standard reference soils were used in chemical analysis.

2.3 Data collection and preprocessing

Information including the latitude and longitude, potential contaminants, enterprise name and industrial category of 7 643 potentially polluting enterprises was collected by field investigation. Google search API data consisting of latitude and longitude and enterprise name was acquired for 264 098 sites using the keyword 'enterprise'. This search information did not include industry type which is likely to be a critical factor controlling the degree of soil pollution. Machine learning methodologies were therefore adopted to classify the industry types (as recorded in the field survey) using the Google search data. The degree of pollution in each soil sample was calculated using the single pollution index (SPI), which was calculated based on the national standard values of different soil heavy metal elements as the evaluation criterion.

2.4 Industrial classification

The main steps in performing classification of industry types, based primary on the enterprises name, were: 1) Word segmentation. The word segmentation, based on a hidden Markov model divided the text into words and the word corpus originating from the training (i.e. field investigation) samples was used for the segmentation of the unlabelled samples. 2) Feature vectorization. The feature vectorization consisted of the feature extraction and the feature selection. Feature extraction is required to remove noise, stop words and other irrelevant text and then present the text in vector form to the classification models. Feature selection leads to improved classification efficiency and reduces the computational complexity. The information gain method based on entropy was used to process this step in this study. The results were analyzed and evaluated by using the Kappa coefficient. Large Kappa coefficients indicate an accurate model. 3) Classification modelling. The classification models considered were Support Vector Machine (SVM), Naive Bayes (NB) and Artificial Neural Network (ANN) algorithm [16-17]. The SVM algorithm is a machine learning method based on statistical learning theory, to seek the best compromise between the model complexity and the learning ability using the limited sample information based on the principle of structural risk minimization. The NB algorithm, which is based on the probabilistic theory, estimates the classification according to the joint probability of the feature and classification. It has a strong theoretical foundation. The ANN algorithm is based on psychology, modern neurology and other specialties to simulate the behavioral characteristics of biological neural network and carries out distributed parallel information processing. It has the advantages of self-learning, nonlinear mapping, and flexible network structure. Details of the industrial classification procedure are illustrated in Figure 2.

2.5 Bivariate spatial correlation analysis

Traditional statistical analysis methods usually focus on statistical relationship between different variables recorded at the same site. However, pollution from an enterprise can potentially extend over a wider area. To overcome this gap, bivariate spatial correlation analysis was conducted to identify spatial association patterns of the industry type and soil pollution data. According to the number and spatial distribution of soil sampling sites, our study area was divided into nearly 5,000 (5 km × 5 km) grid cells. The bivariate local Moran's I was used for the spatial autocorrelation analysis of the grid data (Equation 1).

$$I_i^{ab} = X_i^a \sum_{j=1, j\neq i}^{n} w_{ij} X_i^b \tag{1}$$

where X_i^a is the value of the variable a at location i; X_i^b is the value of the variable b at location i; and w_{ij} is a weight which can be defined as the inverse of the distance d_{ij} among locations i and j [18]. When the value of I_i^{ab} is significantly positive or negative, it shows that the variable a at the grid i is observably correlated with the variable b in the adjacent area; if not, it means that there be no obvious correlation between them.

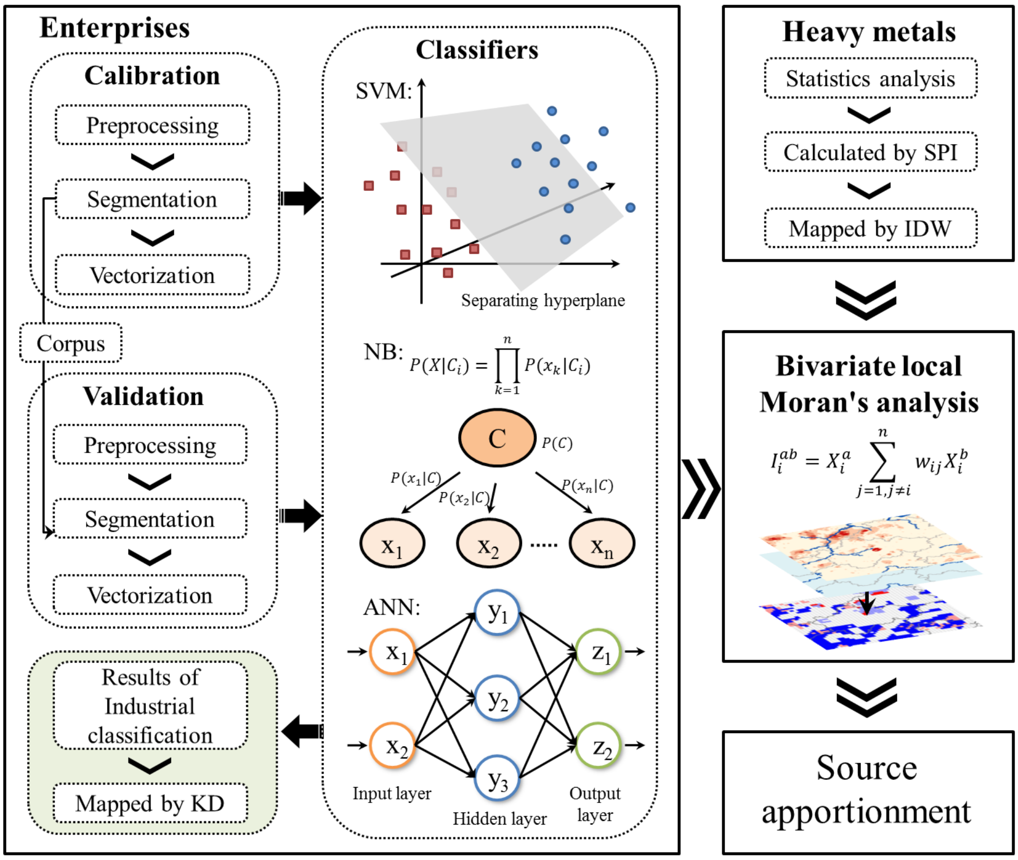

Figure 2. Workflow of the source apportionment in this study. KD: the Kernel Density method, SPI: the single pollution index, IDW: the Inverse Distance Weighted method, SVM: the Support Vector Machine method, NB: the Naive Bayes method, ANN: the Artificial Neural Network method.

3. Results and Discussion

3.1 Descriptive statistics of heavy metals and enterprises

The summary statistics regarding the concentration of the Cd and Hg elements in soils are shown in Table 1. The observed range in the concentrations of Cd and Hg were respectively 0.00-114.00 and 0.01-7.00 mg/kg. The mean concentrations of Cd and Hg were respectively 0.26 and 0.18 mg/kg, which are both higher than the soil background concentrations in the study area (0.07 and 0.09 mg/kg) and in China (0.10 and 0.07 mg/kg) [19]. The coefficient of variation (CV) of Cd and Hg with the values of 403.85% and 133.33% indicated the presence of extremely large concentrations of each element possibly as a result of anthropogenic activities.

Table 1. Descriptive statistics for the concentrations of the heavy metals Cd and Hg in soils.

Element	Mean	Median	SD	Skew	Min	Max	CV	SBC_1	SBC_2
Cd (mg/kg)	0.26	0.18	1.05	88.21	0.00	114.00	403.85%	0.07	0.10
Hg (mg/kg)	0.18	0.11	0.24	8.24	0.01	7.00	133.33%	0.09	0.07

SD: the standard deviation; CV: the coefficient of variation; SBC_1: the soil back concentrations in the study area; SBC_2: the soil back concentrations in China.

The training dataset included 31 industrial categories. Almost 80% of the sites belonged to textile industry (29.6%), chemical industry (28.9%) and metalwork industry (18.9%). The other 29 industrial categories accounted for only 22.65% of the whole dataset, and the proportion of any single industry type was never more than 4%. The data classified according to the complete set of 31 industrial types is referred to as the separated dataset. We also formed an integrated dataset where the textile industry, chemical industry and metalwork industry classes were retained whilst the remainder were combined into a single class. In order to analyze the classification accuracy of different machine learning models, the training samples were divided into a calibration dataset (1 148 samples) and a validation dataset (6 495 samples).

3.2 Industrial classification of enterprises

The radial basis function kernel and linear kernel were used within the SVM classification models. Multinomial NB and Bernoulli NB classified enterprises by adopting different strategies for calculating the likelihood probability of characteristics. The ANN model was a simple network model with only one hidden layer. The prediction results using different classification models are shown in Table 2. These five models had good predictive ability with high accuracy on both the separated and integrated datasets. The average accuracies of prediction results in calibration and validation dataset were 97% and 84% respectively. Overall, the accuracy of models using integrated samples was superior to those using separated samples. SVM, NB and ANN were improved by 1.17%, 2.46% and 1.94%, respectively. The SVM with linear kernel performed best on the calibration dataset with accuracies of 99% and 99% for the calibration dataset. However, by the comprehensive consideration of the results in different datasets, Multinomial NB was chosen to classify the enterprises since it had the highest accuracies of 87% in the integrated validation dataset.

Table 2. Correct rates of different indust classification models in calibration and validation datasets.

Dataset	Separation					Integration				
	SVM_a	SVM_b	NB_a	NB_b	ANN	SVM_a	SVM_b	$\mathbf{NB_a}$	NB_b	ANN
Calibration (%)	98.04	99.17	94.55	92.16	99.29	98.38	99.17	**94.27**	94.32	99.17
Validation (%)	82.32	83.54	83.62	81.62	81.36	85.28	85.71	**86.50**	84.67	85.37

SVM_a and SVM_b: the Support Vector Machine model respectively with radial basis function kernel and linear kernel; NB_a and NB_b: the Naive Bayes model respectively using the Multinomial and Bernoulli theorem; ANN: the Artificial Neural Network model; Separation: the 31 industrial classifications; Integration: the 4 industrial classifications.

For Multinomial NB model, the numbers of enterprise samples predicted correctly in validation dataset, of which industrial classifications were textile industry, chemical industry, metalwork industry and the other industry, were respectively 178, 274, 348 and 193 (Table 3). The average values of the prediction classification accuracy and the method classification accuracies in validation dataset were respectively 88% and 86%. The prediction classification accuracies in validation dataset were, from high to low, metalwork industry, chemical industry, textile industry and the other industry, respectively. Furthermore, the metalwork industry was also most accurately classified in the validation dataset. The prediction classification accuracies in validation dataset followed the order: metalwork industry>chemical industry>textile industry>the other industry. The Kappa coefficient of the classification matrix was 0.82 that meant that the predicted results of industrial classification of pollution enterprises by Multinomial NB were almost identical with the actual results.

Table 3. Comparison for industrial classification results of Multinomial NB model with the observed results.

Actual / Predicted	Textile industry	Chemical industry	Metalwork industry	The other industry	Total	Method Classification Accuracy
Textile industry	**178**	3	1	3	185	96.22%
Chemical industry	4	**274**	11	41	330	83.03%

Metalwork industry	17	22	**348**	26	413	84.26%
The other industry	7	16	4	**193**	220	87.73%
Total	206	315	364	263	1148	
Prediction Classification Accuracy	86.41%	86.98%	95.60%	73.38%		**Kappa Coefficient: 0.82**

3.3 Applicability analysis of classification models

The SVM approach was least sensitive to the number of industry classes and hence more widely applicable. The NB approach required a prior probability value in the classification process and hence was more sensitive to the distribution of categories of data than the other models. By comparing the classification results of separated and integrated samples, the model accuracies of SVM, NB and ANN were respectively improved by 1.17%, 2.46% and 1.94% after data integration. Therefore, SVM with linear kernel and Multinomial NB, which were the models with the highest accuracies in validation dataset, having an improvement of 1.04% and 1.89% after data integration. In this study, the training samples and unlabelled samples had similar distributions of industrial types since they were collected from the same research area. The spatial correction between the soil heavy metals and the main industries were analyzed, using only the textile industry, metalwork industry and chemical industry classes. Therefore, Multinomial NB was chosen for the industrial classification of unlabelled samples, as it had the highest accuracy of 87%. However, SVM with linear kernel had the best applicability ability. In the future work, when try to apply the classification model on the national scale, it is necessary to consider the applicability ability of models and adopt SVM with linear kernel.

3.4 Spatial distribution of heavy metals and enterprises

The search engine data classified to either the textile industry, metalwork industry or chemical industry were retained, accounting for 9.97%, 28.42% and 41.55% of the total content, respectively. The spatial distribution of enterprise density and soil heavy metal pollution degree are shown in Figure 3. The textile industry, metalwork industry and chemical industry was distributed mainly in the eastern part of study area and near rivers and lakes. The number of the enterprises belonged to textile industry was less than that of the other industries and mainly distributed in the HZ and JX districts. The enterprises in metalwork industry and chemical industry had the similar distribution, which were mainly in the HZ, NIB, WZ, TZ districts. The region seriously contaminated by Cd was located in the QZ and HZ districts, and Hg contaminated region was mainly located in the SX and NIB districts. The WZ district included Cd and Hg pollution in soils, where the contamination degree and area was relatively low.

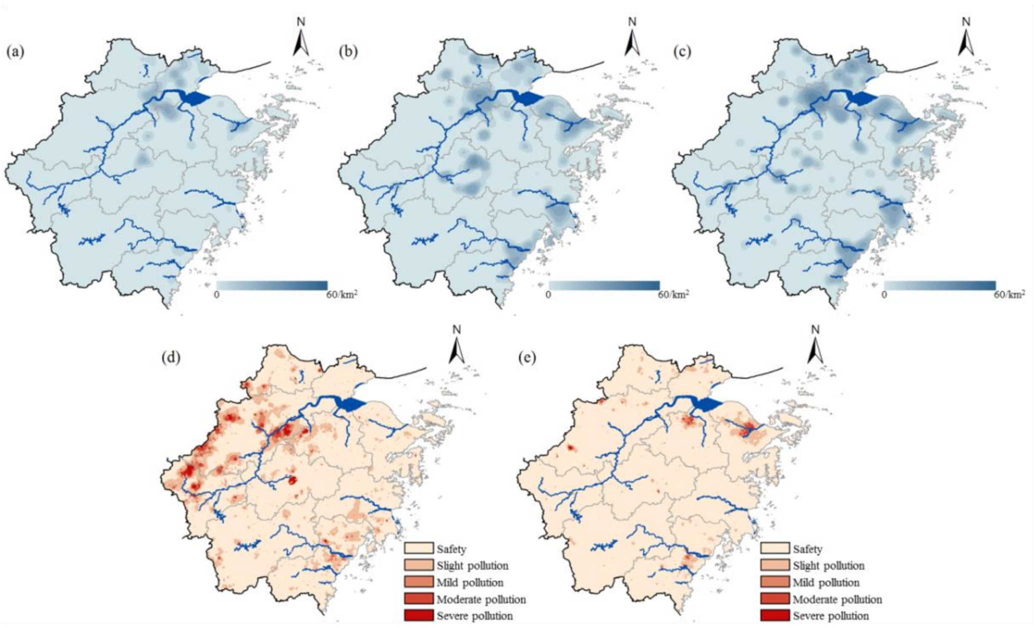

Figure 3. Spatial distribution of enterprise density and soil heavy metal pollution degree. a: textile industry, b: metalwork industry, c: chemical industry, d: Cd element, e: Hg element. The density of pollution enterprises and the pollution degree of Cd and Hg elements were respectively mapped by the Kernel Density method and the Inverse Distance Weighted method.

3.5 Source apportionment of soil heavy metal pollution

The degree of spatial correlation between the soil concentrations of the different elements and the different enterprises as calculated from the bivariate local Moran's I analysis is shown in the Figure 4. High-high and high-low indicate areas with high soil metal concentrations. In the former case this is likely to be the result of pollution from the enterprises whereas in the latter case it is more likely to result from natural factors. Low-high and low-low indicate uncontaminated areas. In low-high area, although the enterprises were distributed densely, the pollution prevention measures were better implemented and no soil pollution resulted. According to the results of bivariate spatial correlation analysis, Cd pollution in soils was mainly unrelated to the enterprises and located in the QZ and HZ districts, while Hg pollution basically belonged to enterprise pollution that was mainly distributed in the JX, SX, NIB and WZ districts. Considering the Cd pollution, the LS, WZ and TZ districts mainly had high-low area and a few high-low areas were sparsely located in the TZ district, meanwhile the JH, SX, WZ and TZ districts contained a small number of scattered high-high areas. In the case of Cd pollution, the textile industry led to almost no pollution in the TZ district and the chemical industry had almost no pollution in the JH district. The metalwork industry caused Cd pollution more seriously than the other industries. In the SX district, the Hg pollution was caused mainly by the textile industry and chemical industry, and in the WZ district by the metalwork industry and chemical industry. Moreover, the chemical industry had the largest high-high area in Hg pollution.

The average high-low area of Cd in the different enterprise analysis was 4277.3 km^2, which was 4.05% of the whole study area, while the average area of Hg was 106.8 km^2, 0.10% of the study area. The areas of Cd pollution mainly caused by the textile industry, metalwork industry and chemical industry were respectively 907.8, 1575.3 and 1161.5 km^2, while the areas of Hg pollution were respectively 1441.8, 1716.8 and 1903.7 km^2. The high-high distribution of Hg was relatively agglomerated compared with Cd. According to results of the field survey of 7643 enterprise contaminants, we found that the proportions of Cd contaminant in the textile industry, metalwork industry and chemical industry were respectively 1.45%, 17.57% and 9.75%, meanwhile the proportions of Hg contaminant were respectively 1.45%, 12.16% and 16.53%. In the three industries, the metalwork industry was the main source of Cd contaminant and the chemical industry mainly produced Hg contaminant. These results agreed with conclusion of our study using the search data to classify the industrial classes.

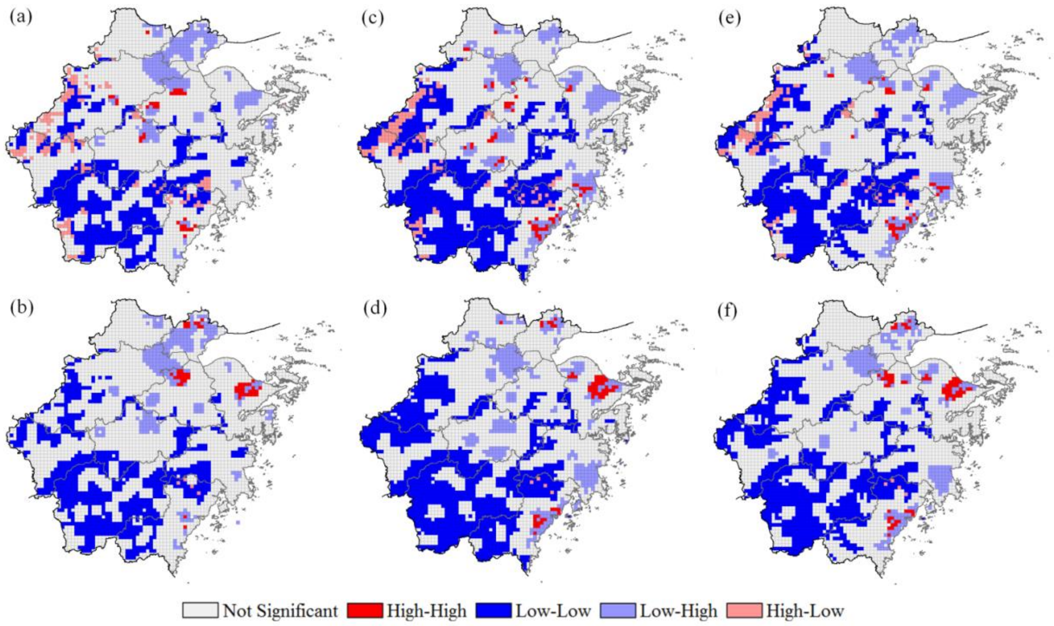

Figure 4. Source apportionment of soil heavy metal pollution by bivariate local Moran's I model using the data of soil Cd and Hg pollution degree and pollution enterprises. a: textile industry & Cd element, b: textile industry & Hg element, c: metalwork industry & Cd element, d: metalwork industry & Hg element, e: chemical industry & Cd element, f: chemical industry & Hg element.

In the late 1970s and early 1980s, the background values of heavy metal elements in soils of the study area were studied by the environmental protection department of the Zhejiang University. Table 5 shows the derived background values of Cd and Hg elements in soils (0-20 cm). According to the Figure 4, the high-low area of Cd pollution was mainly distributed in the HUZ district. The background value of Cd element in the HUZ district was 0.34 mg/kg, which was obviously higher than the other districts of the study area. The high background value of Cd element appears to be the cause of high concentrations in this area.

Table 5. Descriptive statistics for the background values of the Cd and Hg elements in soils.

Element	Statistics	HUZ	HZ	JH	JX	NIB	QZ	SX	TZ	WZ
Cd	Mean	0.14	0.23	0.18	0.11	0.11	**0.34**	0.17	0.15	0.13
(mg/kg)	SD	0.13	0.42	0.14	0.06	0.03	**0.33**	0.08	0.08	0.08

Hg	Mean	0.15	0.23	0.10	0.20	0.22	0.12	0.21	0.16	0.20
(mg/kg)	SD	0.10	0.28	0.08	0.09	0.21	0.08	0.33	0.12	0.13

HUZ, HZ, JH, JX, NIB, QZ, SX, TZ, WZ were respectively the English abbreviations of the 9 provincial cities in the study area; SD: the standard deviation.

Acknowledgements

This work was founded by Key Research and Development Project of Zhejiang Province (2015C02011). The project was planned and designed by Zhou S. and Lianqing Z.; the research data was provided by Youwei Z.; the model was constructed and analysed by Xiaolin J., Bifeng H. and Ye X.; the paper was constructed by Xiaolin J.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

Reference:

(1) Facchinelli, A.; Sacchi, E.; Mallen, L. Multivariate statistical and gis-based approach to identify heavy metal sources in soils. *Environmental Pollution* **2001**, 114(3), 313-324.

(2) Jiang, Y. X.; Chao, S. H.; Liu, J. W.; Yang, Y.; Chen, Y. J.; Zhang, A. C.; et al. Source apportionment and health risk assessment of heavy metals in soil for a township in Jiangsu Province, China. *Chemosphere* **2017**, 168, 1658.

(3) Zawadzka, M.; Łukowski, M. I. The content of Zn, Cu, Cr in podzolic soils of Roztocze National Park at the line of metallurgical and sulphur and the highway. *Acta Agrophysica* **2010**, 16 (2), 459-470.

(4) Dudzik, P.; Sawicka-Kapusta, K.; Tybik, R.; Pacwa, K. Assessment of environmental pollution by metals, sulphure dioxide and nitrogen in Wolinski National Park. *Nat. Environ. Monit.* **2010**, 11, 37-48.

(5) Hu, B. F.; Jia, X. L.; Hu, J.; Xu, D. Y.; Xia, F.; Li, Y. Assessment of heavy metal pollution and

health risks in the soil-plant-human system in the Yangtze River Delta, China. *International Journal of Environmental Research & Public Health* **2017**, 14(9), 1-18.

(6) Qu, M.; Li, W.; Zhang, C.; Wang, S.; Yang, Y.; He, L. Source apportionment of heavy metals in soils using multivariate statistics and geostatistics. *Pedosphere* **2013**, 23, 437-444.

(7) Luo, X. S.; Ip, C.; Li, W.; Tao, S.; Li, X. D. Spatial-temporal variations, sources, and transport of airborne inhalable metals (PM_{10}) in urban and rural areas of northern China. *Atmos. Chem. Phys. Discuss.* **2014**, 14, 13133-13165.

(8) Wang, C.; Yang, Z.; Zhong, C.; Ji, J. Temporal-spatial variation and source apportionment of soil heavy metals in the representative river-alluviation depositional system. *Environ. Pollut.* **2016**, 216, 18-26.

(9) Hu, Y. A.; Cheng, H. F. Application of stochastic models in identification and apportionment of heavy metal pollution sources in the surface soils of a large-scale region. *Environment Science & Technology* **2013**, 47, 3752-3760.

(10) Ma, W. C.; Tai, L. Y.; Qiao, Z.; Zhong L.; Wang, Z.; Fu, K. X.; et al. Contamination source apportionment and health risk assessment of heavy metals in soil around municipal solid waste incinerator: A case study in North China. *Science of The Total Environment* **2018**, s631-632, 348-357.

(11) Guan, Q. Y.; Wang, F. F.; Xu, C. Q.; Pan, N. H.; Lin, J. K.; Zhao, R.; et al. Source apportionment of heavy metals in agricultural soil based on PMF: A case study in Hexi Corridor, northwest China. *Chemosphere* **2018**, 193, 189-197.

(12) Duodu, G. O.; Goonetilleke, A.; Ayoko, G. A. Potential bioavailability assessment, source apportionment and ecological risk of heavy metals in the sediment of Brisbane River estuary, Australia. *Marine Pollution Bulletin* **2017**, 117(1–2), 523-531.

(13) Salles, T.; Rocha, L.; Mourão, F.; Gonçalves, M.; Viegas, F.; Jr, W. M. A two-stage machine learning approach for temporally-robust text classification. *Information Systems* **2017**, 69, 40-58.

(14) Piroonsup, N.; Sinthupinyo, S. Analysis of training data using clustering to improve semi-supervised self-training. *Knowledge-Based Systems* **2017**, 143, 65-80.

(15) Zhou, Y.; Xu, J.; Cao, J.; Xu, B.; Li, C. Hybrid attention networks for chinese short text classification. *Computacion Y Sistemas* **2017**, 21(4), 759-769.

(16) Gong, Z.; Yu, T. Chinese web text classification system model based on Naive Bayes. *International Conference on E-Product E-Service and E-Entertainment* **2010**, 1-4.

(17) Siolas, G.; D'Alche-Buc, F. Support Vector Machines based on a semantic kernel for text categorization. *Ieee-Inns-Enns International Joint Conference on Neural Networks* **2002**, 5205.

(18) Zhang, C.; Luo, L.; Xu, W.; Ledwith, V. 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* **2008**, 398(1), 212-221.

(19) Hu, B. F.; Chen, S. C.; Hu, J.; Xia, F.; Xu, J. F.; Li, Y.; et al. Application of portable xrf and vnir sensors for rapid assessment of soil heavy metal pollution. *Plos One* **2017**, 12(2), e0172438.