Spatial pattern of heavy metals accumulation risk in urban soils of Beijing and its influencing factors

Rui Liu, Meie Wang, Weiping Chen*, Chi Peng

State Key Laboratory for Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

Article info

Article history:
Received 17 June 2015
Received in revised form 23 November 2015
Accepted 24 November 2015
Available online 21 December 2015

Keywords:
Beijing
Urbanization
Soil pollution
Regression tree
Heterogeneity

Abstract

Accumulations of heavy metals in urban soils are highly spatial heterogeneity and affected by multiple factors including soil properties, land use and pattern, population and climatic conditions. We studied accumulation risks of Cd, Cu, Pb and Zn in urban soils of Beijing and their influencing based on the regression tree analysis and a GIS-based overlay model. Result shows that Zinc causes the most extensive soil pollution and Cu result in the most acute soil pollution. The soil's organic carbon content and CEC and population growth are the most significant factors affecting heavy metal accumulation. Other influence factors in land use pattern, urban landscape, and wind speed also contributed, but less pronounced. The soils in areas with higher degree of urbanization and surrounded by intense vehicular traffics have higher accumulation risk of Cd, Cu, Pb, and Zn.

1. Introduction

Urbanization processes not only changed inherent properties of the affected soils, such as their pH, texture, cation exchange capacity, and bulk density but also inadvertently caused harmful substances such as heavy metals to deposit in the soils (Vega et al., 2004; WRB, 2006; Obrador et al., 2007). The pollutant accumulations often led to environmental ills (Pouyat and McDonnell, 1991; Bermea et al., 2009; Chabukdhara and Nema, 2013). In China, the Cd, Cu, Pb and Zn concentrations of urban soils were frequently found to exceed the baseline and the levels were rising (Wei and Yang, 2010; Wang et al., 2012a,b).

Besides, urban development altered the natural landscapes and microclimatic features, created man-made topographies, and concentrated the human populations and activities that in turn would influence the pollutant deposition processes (Lin et al., 2002; Xia et al., 2011). Industrial establishments, transportation networks, residential communities, and other support systems were integrated across the urban horizons to make depositions of heavy metal in urban soils a complicated undertaking (Legret and Pagotto, 2006; Han et al., 2006; Mmolowa et al., 2011). Statistical methods, such as the correlation and regression analyses, were utilized to empirically link heavy metals deposited in the urban soils to potential causative urban factors (Zeng et al., 2011; Palumbo et al., 2000). The methods had not been entirely successful as the outcomes were simplistic and failed to consider the complex spatial and interactive nature of the causative factors. For a complicated system such as the urban metropolises, the classification and regression tree (CART) analysis would be an efficient approach to deduce relationships between heavy metal depositions and causative factors and to distinguish influences of the factors (Kheir et al., 2014; Greve et al., 2012). The CART was capable to search for the non-additive and non-linear relationships, and to uncover the hidden structures in complex data matrices (Breiman, 2001; Henderson et al., 2005; Razi and Athappilly, 2005).

The geo-statistical tool Kriging would convert the soil pollutant levels measured at point locations into spatial distribution of pollution across the urban landscape. The outcomes, however, tended to gloss over the distinctive local highs and lows because of the smoothed extrapolation technique (Journel et al., 2000) thus failed to preserve the highly variable nature and skewed pollutant distributions of the actual urban environment and reduced the accuracy of assessments (Goovaerts, 2000); especially for the highly heterogeneous urban environment. A more robust approach to evaluate the spatial variation of heavy metal contaminations in urban area would be the GIS-based model that linked and integrated the spatial distribution of heavy metals with the spatially referenced data of causative factors (Desmet and Govers, 1996).
We hypothesized that the soil, land use, demographic, and climatic factors influenced the spatial distribution of heavy metals in urban soils. Used the CART analysis and a GIS-based model to illustrate how causative factors including soil properties, urban land use, demographics, and microclimatic conditions would affect the distribution of Cd, Cu Pb and Zn in the urban soils throughout Beijing.

2. Materials and methods

2.1. Study area

The urban built-up areas of Beijing (Fig. 1), which encompassed areas inside of the 5th ring road of Beijing and included the Dongcheng and Xicheng Districts and parts of the Haidian, Chaoyang, Fengtai, and Shijingshan Districts were the study template. The area belonged to warm temperate semi-humid continental monsoon climate and the prevalent winds were northwest and south in directions. The cinnamon and fluvo-aquic were the dominant soils with parent materials consisted of weathering rocks and loose quaternary sediment. The city in recent decades experienced drastic and rapid urban renewal and developments. From 1949 to 2012, the population increased from 1.8 to 20.7 million, while the built-up areas increased from 109 to 670 km².

2.2. Soil sampling

To ensure a uniform distribution of sampling sites, the study area was divided by 1 km x 1 km sized grids on Google Earth, and then sample site was selected from each grid based on the land use and topographic conditions (some grids were unavailable for sampling). Finally, 232 sample sites were selected (Fig. 1). Surface soil (0–20 cm depth) samples were collected. Crumbled and free of roots and other organic debris, the specimens were then air-dried and crushed passed through a sieve of 2 mm x 2 mm openings. Each specimen was then subdivided with three quarter of the material preserved for analyzing pH and soil particle size, and the remaining one quarter of the material further ground to pass a sieve of 0.15 mm x 0.15 mm openings to be used in analysis of soil organic carbon and cation exchange capacity. In addition, 100 cm³ intact soil cores were obtained in metal cylinders made from 5 cm ID straight brass tubing and were kept at the field moisture content in cold storage. Later, they were used to determine the field capacity and bulk density of the soils.

2.3. Chemical analysis

The soil pH was determined in 1:2.5 weigh vs. volume aqueous soil suspension via glass electrodes potentiometry. The soil particle size distribution was determined by a laser particle size analyzer and the outcomes reported according to the United States Department of Agriculture (USDA) soil classification scheme in which the clay content (CLAY) denoted the percentage of particles with sizes less than 0.002 mm. The soil organic carbon content was measured using a C–N–S elemental analyzer, provided the soil aliquot was pretreated with 1 mol/L HCL to remove the inorganic
carbon. The soil’s cation exchange capacity (CEC) was determined by the sodium acetate extraction method according to the U. S. Environmental Protection Agency (EPA) Method 9081.

The Cd, Cu, Pb, and Zn contents of soils were obtained by first digesting a 0.25 g soil aliquot (ground to pass a sieve of 0.1 mm openings) in 10 ml HCl overnight, then heated to reduce down to 3 ml and then continued the heating with 5 ml HNO₃, 5 ml HF, and 3 ml HClO₄ until 3 ml remained. The final suspension would be further digested (repeated if necessary) in 2 ml HNO₃, 2 ml of HF, and 1 ml HClO₄ until the solution turned clear. The digested extracts were diluted by ultrapure water at 1 to 50 for ICP-OES analysis of Cu and Zn, and at 1–250 ml for ICP-MS spectrometry analysis of Cd and Pb.

2.4. Quality assurance and quality control

The national registered standard reference materials, GSS-13 (Geochemical Standard soil) was included in chemical analyses. The mean recoveries were 83% for Cd, 71% for Cu, 91% for Pb and 98% for Zn. Duplicates were analyzed on 10% of the soil samples and the standard deviations were within ±5% of the mean. Reagent blanks were included with each batch of samples analyzed.

2.5. Grid transformation

To spatially synchronize the data on soil properties with urban geographical date for statistical analyses, the 231 soil sampling grids were consolidated and transformed into 81 3 km × 3 km cells (Fig. 1). The soil properties of a cell were represented by the mathematic averages of those within a cell. Accordingly, the population, land use, and meteorological data were also spatially transformed into 81 cells.

2.6. Data collection and spatialization

The land uses were delineated from the high-resolution remote sense images of IKONOS that were collected on July 29, 2012 and were classified into 9 types, including agricultural (AGR), commercial (COM), industrial (IND), parks and greenbelt (P&G), public (PUB, include land used by public facilities, such as school, hospital, stadium and research institution), residential (RES), traffic (TRA, include land used by different grade roads, railways and public transportation depots), water body (WAT, include artificial or natural lakes and rivers), and unused (UNU, include land that was under construction and those were currently idle). The proportions of land uses in each cell were then calculated. The urban landscapes were described by landscape metrics, which were computed by Fragstats version 4.2, after land use vector data were converted into raster data and separated by the 3 km × 3 km cells.

The population data were based on statistics of the 2000 and 2010 national census that included populations of the districts. The population of each 3 km × 3 km cell was visualized and calculated using ArcGIS 9.3 spatial analysis tools and the 2000 to 2010 average annual population growth rates were calculated.

The Urban Ecosystem Research Station of Beijing provided the meteorological data including wind direction and wind speed. Data were recorded at hour interval, from 2010 to 2013. To be compatible with other matrices, the 81 cell were categorized into 8 classes and assigned a numerical value based on their location in the study region, namely, north = 1, north-east = 2, east = 3, south-east = 4, south = 5, south-west = 6, west = 7, and north-west = 8. The annual mean speed of each cell was then calculated.

2.7. Calculation of geoaccumulation index (Igeo)

A modified Geoaccumulation index (Igeo) was used to reflect the influences made by human activities on soils heavy metals contamination. The equation is given as:

\[
I_{geo} = \log_2 \left( \frac{C_n}{B_n} \right)
\]

where \(C_n\) was the concentration of the examined element in the examined bottom sediment, \(B_n\) was the geochemical background of a given element in mudstone, in this study, the Beijing soil background values were used (Table 2). 1.5 was a modified index, which would multiply the concentration of geochemical background in order to allow content fluctuations of a given substance in the urban environment as well as very small anthropogenic influences. The geoaccumulation index of Cd, Cu, Pb and Zn were calculated separately.

2.8. Regression tree analysis

Two types of regression tree analysis were conducted to identify the influences of the 40 selected factors (independent variable) on soil heavy metal pollution (dependent variables). In the univariate regression tree (URT), individual concentration of Cd, Cu, Pb, and Zn in urban soils was set as dependent variable, while in the multivariate regression tree (MRT), the four metal concentrations in urban soils were set as one dependent variable (Table 1).

The regression trees were developed based on those data generated for the 81 spatial cells. Trees were constructed by binary splits, and the data were recursively partitioned into subsets, which was as homogeneous as possible. The splitting procedures would stop when trees reached the specified minimum number of samples allowed at the node. The trees were then pruned by the complexity parameter \(cp\) at any node, which was the proportion in prediction at each node.

The regression trees were written in the R programming language and “anova” and “mrt” methods were chosen to construct regression trees. The minimum count of samples (MinS) involved at any node was 5, \(cp\) at the any end node was 0.01, the fitness was defined by relative error (RE), and the prediction validity of model was expressed by 1-RE.

2.9. Risk assessment of heavy metals accumulation

Soil accumulation of heavy metals like Cd, Cu, Pb and Zn would harm the integrity of urban soil ecosystem. Their potential risks are dependent on spatial information of emission sources and topographic conditions (Wang et al., 2012b). To assess the risk of heavy metal accumulation/loss in soils across the urban landscape, we followed the mathematical form of the Universal Soil Loss Equation (USLE) which was original designed to estimate long-term annual erosion rates on agricultural fields by different factors (Wischmeier and Smith, 1978), and developed a cell-by-cell GIS-based overlay model. The model was given as:

\[
P = \sum W_i \times F_i
\]

where \(P\) is the soil accumulation risk of specific heavy metal, \(F\) is the selected significant influence factors for each heavy metal based on results of the regression trees analysis, and \(W\) is the corresponding weight for each factor.

The data of all selected factors in each cell were standardized before the calculation of the accumulation risk. The weight was the percentage of 1-RE for each factor with respect to the total 1-RE. Based on the calculated risks, study area were grouped into low,
middle, and high risk regions by the Quantile method and mapped by spatial analysis in ArcGIS. In this manner, the study region was equally classified into three groups and the regions which have relatively high risk of heavy metal accumulation without reference data were distinguished.

3. Results and discussion

3.1. Spatial distribution of the heavy metal geoaccumulation index

The heavy metal contents of urban soils in Beijing were spatially heterogeneous (Table 2). The means of Cd, Cu, Pb and Zn concentrations were 0.138, 20.810, 25.357, and 80.278 mg kg\(^{-1}\), respectively and the ranges of Cd, Cu, Pb and Zn concentrations varied from 0.066 to 0.486, 8.408 to 213.147, 11.178 to 135.371, and 26.711–225.211 mg kg\(^{-1}\), respectively. Among metals, the soil Cu showed the highest spatial variations with the coefficient of variation (CV) at 84%, followed by Pb, Cd, and Zn with CV at 57, 42, and 37%, respectively.

In terms of the geo-accumulation index, \(I_{geo}\), heavy metal pollution of the urban soils were at the level 2 (contaminated) or lower (Fig. 2). Area-wise, the soil pollution were primarily at level 0 (uncontaminated) to 1 (moderately contaminated) with soils in most of the area exhibited \(I_{geo} \leq 0\) or \(0 \leq I_{geo} \leq 1\) with respect to Cd, Pb, and Zn, while small and isolated parts exhibited \(I_{geo} > 1\). Spatially, 70% of the study area was deemed uncontaminated and 10% of the area was contaminated by 2 or more metal elements at the same time. Overall, urban soils of Beijing sustained low levels of Cd, Cu, Pb, and Zn pollution, while Zn pollution was area-wise the most extensive and Cu pollution was concentration-wise the most severe.

3.2. Regression trees establishment

Both the un-pruned univariate and multivariate regression trees showed good prediction as the 1-RE were greater than 0.6 (Table 3). However, the resulting trees involved 23 to 29 TNs, and Tvars, were 9–12 that making the outcomes more complicated to interpret. After pruning the redundant nodes based on the complexity parameter (cp), the total nodes, TN, in the pruned tree were reduced to 17 to 23 and total variable count, Tvvar, were reduced to 6 to 8. The total nodes were reduced by 4, 8, 10, 4, and 6 and total count of variables, Tvars, were smaller and less complicated than the un-pruned ones in terms of analyses yet were equal in terms of predicting the outcomes.

3.3. Relationships between heavy metals and influence factors

The regression trees illustrated the relationships between the soils’ heavy metals content (dependent variables) and the influence factors (independent variables) including variables representing soil properties, land use type, landscape metrics, population, and meteorological data (Figures 15 and 25). The more an independent variable showed up in the tree, the closer was its relationship to the corresponding dependent variable (White et al., 2005).

The primary factors influencing the accumulations in soils were metal specific. For Cd, the factors included soil organic carbon (SOC, 1-RE = 0.313), annual growth rate of population (PGR, 1-RE = 0.152), wind speed (SPD, 1-RE = 0.053), landscape patterns (IJI, 1-RE = 0.085 and PAFRAC, 1-RE = 0.030), and the area of commercial land use (COM, 1-RE = 0.019). For Cu, the factors were soil’s cation exchange capacity (CEC, 1-RE = 0.206), population growth (PGR, 1-RE = 0.068), landscape patterns (SHDI, 1-RE = 0.131 and LSI, 1-RE = 0.075), areas of public and commercial land use (PUB, 1-RE = 0.065 and RES, 1-RE = 0.02), and wind speed (SPE, 1-
Factors influencing the Pb and Zn in the soils might be teased out in the same manners. Undoubtedly, the sources of emission were not the same for each metal and once emitted their fate and transport were affected by multitude of factors including the soil properties and land use patterns. For four heavy metals (Cd, Cu, Pb, Zn) combined, the accumulations were influenced by the soil’s organic carbon (SOC, 1-RE = 0.301) and cation exchange capacity (CEC, 1-RE = 0.071), population growth (PGR, 1-RE = 0.045), landscape patterns (NP, 1-RE = 0.072, PRD, 1-RE = 0.011, and SHDI, 1-RE = 0.054), and areas of traffic and public land use (TRA, 1-RE = 0.016 and PUB, 1-RE = 0.090). The soil’s organic carbon and cation exchange capacity were the key soil attributes governing the resulting chemical forms of the deposits and the metal sorption capacities of urban soils (Davis, 1984; Zeng et al., 2011). The soil’s pH (changed from 7.44 to 8.90, with coefficient of variation at 3.217%) and clay content (changed from 2.512% to 17.669%, with coefficient of variation at 35.608%) showed rather small spatial variations and they did not exhibit any significant contribution to the heavy metal accumulation in regression trees of Cd, Cu, Pb, Zn and MRT (Vega et al., 2004; Obrador et al., 2007).

Anthropogenic activities also significantly impact the accumulation of heavy metal in urban soils (Tessier et al., 1979). In the resulting regression trees, land uses including residential, commercial, public, traffic and industrial as well as demographics including population density and growth were closely significantly related to accumulation patterns.

The urban landscape as a result of human activities was defined more in terms of regularity and shape and less by the land uses and fragmented small-scale patches and the urban landscape might be

<table>
<thead>
<tr>
<th>Trees</th>
<th>Un-pruned</th>
<th>Pruned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TN  MinS  1-RE</td>
<td>Tvar  TN  MinS  1-RE</td>
</tr>
<tr>
<td>Cd</td>
<td>23  5  0.668</td>
<td>9  17  5  0.652</td>
</tr>
<tr>
<td>Cu</td>
<td>25  5  0.620</td>
<td>9  17  5  0.590</td>
</tr>
<tr>
<td>Pb</td>
<td>29  5  0.696</td>
<td>12  19  5  0.677</td>
</tr>
<tr>
<td>Zn</td>
<td>27  5  0.703</td>
<td>11  23  5  0.685</td>
</tr>
<tr>
<td>MRT</td>
<td>27  5  0.692</td>
<td>10  21  5  0.662</td>
</tr>
</tbody>
</table>

Note: TN means the total nodes of regression tree; MinS means the minimum count of samples (MinS) involved at any node; RE means relative error, and the prediction validity of model was expressed by 1-RE; Tvar means total variable count for a tree; cp means complexity parameter for total nodes.
characterized in various manners. The metrics such as perimeter-area fractal dimension (PAFRAC), metric and landscape shape index (LSI) reflected the characteristics of urban shapes. The largest patch index (LPI) and number of patches (NP) were metrics indicating the fragmentation of urban landscape. The Shannon’s diversity index (SHDI) and patch richness density (PRD) described the diversity of land uses. Interspersion and juxtaposition (IJI) indicated the contagion of the land uses. They were included in the regression tree analyses and IJI and PAFRAC were significant factors on Cd and SHDI and LSI were significant factors on Cu accumulations in urban soils. The spatial distribution (IJI) and shape (PAFRAC, LSI) of land use would influence the traffic condition, while the land use type diversity would link with the degree of urban development and intension of human activities, therefore, landscape metrics could be significant factors for some heavy metals accumulation.

The microclimatic factors would influence atmospheric fallouts and distributions of the emitted heavy metals in urban soils (McGrath et al., 2004; Juang et al., 2001; Lado et al., 2008). The speed and direction of wind were important factors in spreading and distributing airborne pollutants (Wu et al., 2008; Chen et al., 2010; Nezhad et al., 2015). According to the regression trees (Figure 1S), the wind speed (SPE) exhibited pronounced effects on Cd and Cu accumulations in urban soils.

The factors exerted different influences either accelerating or slowing down the heavy metal accumulation processes in the urban soil (Figures 1S and 2S) and each factor exhibited a different

### Table 4

Relative importance of influence factors in the constructed regression tree.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Cd Weigh</th>
<th>Cu Weigh</th>
<th>Pb Weigh</th>
<th>Zn Weigh</th>
<th>Comprehensive Weigh</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOC</td>
<td>0.480</td>
<td></td>
<td></td>
<td></td>
<td>0.456</td>
</tr>
<tr>
<td>PGR</td>
<td>0.233</td>
<td>0.349</td>
<td>0.115</td>
<td>0.127</td>
<td>0.230</td>
</tr>
<tr>
<td>IJI</td>
<td>0.130</td>
<td>0.155</td>
<td></td>
<td>0.127</td>
<td>0.155</td>
</tr>
<tr>
<td>SPEED</td>
<td>0.081</td>
<td>0.046</td>
<td>0.011</td>
<td></td>
<td>0.081</td>
</tr>
<tr>
<td>PAFRAC</td>
<td>0.009</td>
<td>0.019</td>
<td></td>
<td>0.046</td>
<td>0.016</td>
</tr>
<tr>
<td>COM</td>
<td>0.029</td>
<td></td>
<td>0.042</td>
<td></td>
<td>0.029</td>
</tr>
</tbody>
</table>

Fig. 3. Spatial distributions of accumulation risks.
cause — effect context. For examples, soil properties such as the organic carbon and cation exchange capacity exerted positive impacts to increase the heavy metal accumulation due to their abilities to chemically immobilize metals in the soils (Davis, 1984). The population growth (PCG) exerted negative influences yet the population density (10POP) exerted positive influences on the accumulations. The areas of high population density logically exhibited lower rate of population growth. Because of higher population per unit area, the pollutant emissions of the higher population area would be proportionally increased therefore higher metal accumulations in the surrounding soils. Wind speed negatively influenced the Cd and positive influenced the Cu accumulations in urban soils as Cd was emitted primarily by the mobile sources of vehicular traffic thus susceptible to atmospheric transport while the Cu was emitted primarily from stationary industrial production sources thus tended to staying local (Wang et al., 2012b). All of the land use related factors involved in the regression trees (i.e. RES, PUB, IND and TRA) showed negative impacts while all landscape pattern related factors involved in regression trees (i.e. PAFRAC, SHDI, LSI, LPI, NP, PRD and IIJ) showed positive influences on the heavy metal accumulation in the urban soils. Unlike the previously identified physical factors, we were unable to fully comprehend the complicated interrelations between these descriptive parameters and the heavy metal accumulations in urban soils (Table 4).

3.4. Assessment of heavy metals accumulation risks

The heavy metal accumulation risks of urban soils could be evaluated based on outcomes of the regression trees and Eq. (2). For Beijing, the risks varied from low to high and the spatial patterns of risks were similar among Cd, Cu, Pb, and Zn. The regions with areas of high risks were concentrated at the central-east and west sections and the areas of low risks were concentrated at the south sections of the city (Fig. 3).

For Cd and Cu, high risk regions were more concentrated at east-central, while for Pb and Zn, high risk regions were more concentrated at north—central. Comparing with single heavy metal accumulation risks, the distribution of multiple accumulation risk for 4 heavy metals was more decentralized. High risk regions were generally distributed in the north margin, east-central and parts of west, while most low risk regions were distributed in south and north—west.

Judging by the spatial distributions of accumulation risks of individual metals, the accumulations were related to the extent of urbanization, high risks regions were associated with high level of urbanization. For four metals combined accumulation risks, in addition to high degree of urbanization, the regions were surrounded with intense vehicular traffics. Intensive human activities were key factors affecting heavy metals (Cd, Cu, Pb and Zn) accumulation in urban soils of Beijing.

4. Conclusions

We divided the metropolitan Beijing into north—south and east—west grids with 81 cells. The Cd, Cu, Pb and Zn contents of soils in each cell along with metrics characterizing the soil properties, land use pattern, demographics, and microclimatic conditions were determined. The Cd, Cu, Pb and Zn distributions in urban soils were spatially heterogeneous with coefficients of variation (CV) of 42, 54, 57, and 37%, respectively. The soils, based on the geo-accumulation indices, were characterized as slightly polluted in terms of Cd, Cu, Pb and Zn accumulations in the soils. Zinc accumulations were the most extensive and Cu accumulations were the most acute. The regression tree analyses showed that the accumulations of Cd, Cu, Pb and Zn in urban soils were caused by different sets of influence factors, suggesting that their emission sources and environmental behaviors of each metal were unique. By far, the soil organic carbon was the most prominent factor affecting the accumulations regardless of the element and it was followed by population growth and the cation exchange capacity. Influence of wind speed on accumulation of Cd and Cu was significant. The metrics characterizing the land use pattern also showed significant influences on heavy metal accumulations. The relationships however were complicate and difficult to quantify.

The areas of high heavy metal accumulation risks were associated with high degrees of urbanization and were surrounded with intense vehicular traffics.

Acknowledgements

We acknowledge the financial support of the National Natural Science Foundation of China (#41173123).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.envpol.2015.11.044.

References


Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.envpol.2015.11.044.


