Characterizing the risk assessment of heavy metals and sampling uncertainty analysis in paddy field by geostatistics and GIS

Xingmei Liu, Jianjun Wu, Jianming Xu*

Institute of Soil and Water Resources and Environmental Science, Zhejiang University, Hangzhou 310029, China

Received 8 February 2005; accepted 19 August 2005

Probability maps gave insight into risk assessment of environmental metals in a rice paddy field.

Abstract

For many practical problems in environmental management, information about soil heavy metals, relative to threshold values that may be of practical importance is needed at unsampled sites. The Hangzhou–Jiaxing–Huzhou (HJH) Plain has always been one of the most important rice production areas in Zhejiang province, China, and the soil heavy metal concentration is directly related to the crop quality and ultimately the health of people. Four hundred and fifty soil samples were selected in topsoil in HJH Plain to characterize the spatial variability of Cu, Zn, Pb, Cr and Cd. Ordinary kriging and lognormal kriging were carried out to map the spatial patterns of heavy metals and disjunctive kriging was used to quantify the probability of heavy metal concentrations higher than their guide value. Cokriging method was used to minimize the sampling density for Cu, Zn and Cr. The results of this study could give insight into risk assessment of environmental pollution and decision-making for agriculture.

Keywords: Geostatistics; Spatial variability; Risk assessment; Heavy metals; Sampling density

1. Introduction

Soil heavy metals pollution is an urgent problem worldwide. Understanding the spatial distribution of pollutants is critical for environmental management and decision-making. Particular emphasis is placed on the study of risk assessment for heavy metals (Steiger et al., 1996; McGraph et al., 2004). Paddy field is widely distributed in the southeast region of China. In particular, the way paddy fields are cultivated and the growth characteristics of rice impart spatial variability of soil properties that are different from that of upland (Liu et al., 2004). As a dominating rice production area for Zhejiang Province, the accumulation of soil heavy metals in Hangzhou–Jiaxing–Huzhou (HJH) Plain could either directly endanger the natural soil functions, or indirectly endanger the biosphere by bioaccumulation and inclusion in the food chain, and ultimately endanger the health of people. With the rapid development of industrialization and urbanization, the soil heavy metal pollution in HJH Plain is presenting a trend of extending and therefore badly destroys the product quality. It was reported that about half of the soils in HJH plain were polluted by heavy metals to different degrees. And in Hangzhou, the soils for the dominant crop and vegetable producing area are also in the heavy metal pollution status. Accordingly, the crop yield decreases and the product quality descends, this further results in the economic loss. So it is urgent to understand the true situation of soil heavy metal pollution in HJH plain and provide scientific basis for agricultural policy making.

Geostatistics has been popularly applied in investigating and mapping soil pollution by heavy metals, in particular, in recent years (Steiger et al., 1996; White et al., 1997; Lin et al., 2001; Romic and Romic, 2003; McGraph et al., 2004). However, most studies were carried out at a small scale and rare researches have been completed at large scale. In fact, geostatistics has been proved useful in characterizing spatial variability and mapping a variety of soil properties even at the scale of the conterminous USA (White et al., 1997). Meanwhile, the research at
large area could provide the spatial patterns of soil heavy metal over long distance which a small scale study would fail to accomplish. And this would make it more convenient for macroscopic agricultural policy making. However, the implementation of soil sampling and analysis at large area are often limited by cost and time. So, how to minimize the sampling density has gradually become an interest.

Primary objectives of this research were to (1) analyze the spatial dependency and explain the variation mechanism of heavy metals in the paddy field at large scale; (2) map the spatial distribution of soil heavy metals and quantify their risk assessment; and (3) minimize the sampling density of soil Cu and Zn using cokriging.

2. Materials and methods

2.1. Study area

The HJH water-net Plain with large scale of 1:250,000 was selected in this study. It is located in the north of Zhejiang province, in the southeast of China and to the south of the Tai Lake. The region includes Jiaxing city, Jiashan county, Tongxiang county, Haining county, Haiyan county, Pinghu county, Huzhou city, Deqing county, Anji county, Changxing county, most part of Hangzhou city and part of Lin’an county (Fig. 1). As a typical region of the South Yangtze River, the plain is densely dotted with drainage ditches that form a network of waterway. HJH plain covers 6390.8 km², and it is a more concentrated agricultural area and the primary food product region in Zhejiang Province. Rice (Oryza sativa) has been a dominant crop in the area. Moreover, it is also one of the most developed areas, as to rural economy, in Zhejiang Province. Soils are mainly paddy soils.

2.2. Data sampling and analysis

Considering the uniformity of soil samples distribution and soil types in the study area, 450 soil samples were collected from different locations in paddy field in 2002 at an interval of 5 km. Distribution of sampling points is presented in Fig. 2. All soil samples were taken at a depth of 0–15 cm and air-dried to remove stones and coarse plant roots or residues. Samples were thoroughly mixed and ground to pass a 0.149-mm sieve, then stored in plastic bags prior to chemical analyzing. The mean value of soil pH is 5.81. After digesting with a mixture of nitric acid (HNO₃) and perchloric acid (HClO₄), Cu, Zn, Pb, Cr and Cd concentrations in the digestion liquid were determined by atomic absorption spectrometry (AAC) (Liu, 1996).

2.3. Geostatistical methods

The main application of geostatistics to soil science has been the estimation and mapping of soil attributes in unsampled areas (Goovaerts, 1999).
And the disjunctive kriging is the principal technique to estimate the probability that the true values of soil heavy metals at an unsampled location exceed a specified threshold. It is based on the assumption that our data are a realization of a process with a second-order stationary bivariate distribution. The assumption of second-order stationarity means that the covariance function exists and that the variogram is therefore bounded. It is assumed that the concentration of a heavy metal is a realization of a random variable \( Z(x) \), where \( x \) denotes the spatial coordinates in two dimensions. If a threshold concentration \( z_c \) is defined, marking the limit of what is acceptable, then the scale is dissected into two classes which is less and more than \( z_c \), respectively. The soil must belong to either class at any one place. The value 0 and 1, respectively, can be assigned with certainty. Elsewhere, one can at best estimate (predict) \( Z(x) \). In this study, the fitted linear model and exponential model were selected.

The linear function is:

\[
\gamma(h) = C_0 + Ah \quad h > 0
\]

The exponential function is:

\[
\gamma(h) = C_0 + C(1 - e^{-h/a}) \quad h > 0
\]

where \( C_0 \) is the nugget variance (\( h = 0 \)), represents the experimental error and field variation within the minimum sampling spacing. Typically, the semivariance increases with increasing lag distances to approach or attain a maximum value or sill (\( C_0 + C \)) equivalent to the population variance. \( C \) is the structural variance and \( a \) is the spatial range across which the data exhibit spatial correlation. \( A \) is the slope.

Cokriging, an interpolation method derived from kriging, was used in this study to minimize the sampling density of soil Cu and Zn using the covariable of soil organic matter. And the cross-variograms are calculated according to the following formula (Cressie, 1993):

\[
\gamma_{ij}(h) = \frac{1}{2N(h)} \sum_{i,j=1}^{Z(h)} [Z_i(x) - Z_j(x+h)] [Z_i(x) - Z_j(x+h)]
\]

where \( \gamma_{ij}(h) \) is the cross-semivariance of two random variables as a function of \( h \), and \( N(h) \) is the number of pairs of \( Z_i(x) \) and \( Z_j(x) \) at a separate distance \( h \).

3. Results and discussions

3.1. Descriptive statistics

Occurrence of exceptional values can lead to the data discontinuity and this would violate the geostatistics theory. In this study, we consider the data out of the extent of \( (A \pm 3s) \) as the exceptional values, where \( A \) denotes the average value.
for each heavy metal and $s$ is the standard deviation. Only data exceeding the value $(A + 3s)$ were found in the raw data set and we replaced them with the maximum value of data set without exceptional values.

Table 1 gives the summary statistics of the data sets for five metals. It is shown in Table 1 that the kurtosis and skewness values for Zn and Cr were low, however, that for Cu, Pb, and Cd were high and these three metals were not normally distributed. The kurtosis and skewness values for Cu, Pb, and Cd decreased after the raw data sets were logarithmically transformed and further analyzed using their logarithmically transformed values. The coefficients of variation of Pb and Cd were 147.8% and 125.3%, respectively, and higher than those of Cu, Zn and Cr, suggesting that Pb and Cd had greater variation among the soils.

Based on Chinese Environmental Quality Standard for Soils (GB 15618-1995) (State Environmental Protection Administration of China, 1995), the guide values for Cu, Zn, Pb, Cr and Cd were 50 mg kg$^{-1}$, 200 mg kg$^{-1}$, 250 mg kg$^{-1}$, 250 mg kg$^{-1}$ and 0.30 mg kg$^{-1}$, respectively, when soil pH value was less than 6.5. And the average value for soil pH in HJH plain was 5.81, which just fell on this scope. It was displayed that seven sample concentrations exceed the guide value as to Cu. The most distinguished were for Zn and Cd, 20 and 36 samples were found to exceed greatly their guide value, respectively. However, the mean concentrations of Pb and Cr were 13.54 mg kg$^{-1}$ and 68.39 mg kg$^{-1}$, respectively, and the maximum was 85.9 mg kg$^{-1}$ and 182.64 mg kg$^{-1}$, respectively. Judged against their guide value of 250 mg kg$^{-1}$, they present no threat to soil environmental quality.

### Table 1

<table>
<thead>
<tr>
<th>Soil properties</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>S.D.</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>C.V. (%)</th>
<th>Guide value</th>
<th>Number exceeding guide value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu (mg kg$^{-1}$)</td>
<td>28.36</td>
<td>28.08</td>
<td>8.84</td>
<td>67.94</td>
<td>8.26</td>
<td>4.07</td>
<td>1.13</td>
<td>29.12</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>Logarithm of Cu</td>
<td>1.45</td>
<td>1.45</td>
<td>0.95</td>
<td>1.83</td>
<td>0.13</td>
<td>1.08</td>
<td>−0.32</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Zn (mg kg$^{-1}$)</td>
<td>108.8</td>
<td>105.79</td>
<td>13.51</td>
<td>246.44</td>
<td>41.88</td>
<td>1.97</td>
<td>0.93</td>
<td>38.49</td>
<td>200</td>
<td>20</td>
</tr>
<tr>
<td>Pb (mg kg$^{-1}$)</td>
<td>13.54</td>
<td>7.3</td>
<td>1.11</td>
<td>85.9</td>
<td>20.02</td>
<td>6.83</td>
<td>2.87</td>
<td>147.8</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>Logarithm of Pb</td>
<td>0.93</td>
<td>0.86</td>
<td>0.05</td>
<td>1.93</td>
<td>0.34</td>
<td>2.8</td>
<td>1.7</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Cr (mg kg$^{-1}$)</td>
<td>68.39</td>
<td>67.81</td>
<td>1.58</td>
<td>182.64</td>
<td>31.77</td>
<td>1.56</td>
<td>0.8</td>
<td>46.46</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>Cd (mg kg$^{-1}$)</td>
<td>0.12</td>
<td>0.07</td>
<td>0.01</td>
<td>0.65</td>
<td>0.14</td>
<td>7.38</td>
<td>2.84</td>
<td>125.3</td>
<td>0.3</td>
<td>36</td>
</tr>
<tr>
<td>Logarithm of Cd</td>
<td>−1.12</td>
<td>−1.16</td>
<td>−2.26</td>
<td>−0.19</td>
<td>0.36</td>
<td>0.9</td>
<td>0.72</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

S.D.: standard deviation; C.V.: coefficient of variation.

3.2. Geostatistical analysis

Fig. 3 presents the semivariogram and fitted model for five metals. The attributes of the semivariograms for each soil heavy metal are summarized in Table 2.

Nugget variance represents the experimental error and field variation within the minimum sampling spacing. The Nug/Sill ratio can be regarded as a criterion to classify the spatial dependence of soil properties. If the ratio is less than 25%, the variable shows only weak spatial dependence; between 25% and 75%, the variable has moderate spatial dependence; and greater than 75%, the variable has strong spatial dependence; for linear model, however, the other three heavy metals in soil were all best fitted for an exponential model. And the Nug/Sill ratios of five metals were between 28.4% and 63.4%, suggesting moderate spatial dependence at the large scale of HJH plain, which showed that the extrinsic factors such as fertilization, ploughing and other soil management practices weakened their spatial correlation after a long history of cultivation. The semivariograms for Cu and Cr were similar and their ranges were all 97.4 km, much longer than that of Zn, Pb and Cd. Meanwhile, this result indicates the potential of precise environmental survey for soil heavy metals with a rational sampling distance within the ranges.

3.3. Spatial distributions and risk assessment

Paddy soils mainly distribute in plain and valley, and the parent materials are alluvial—lacustrine or coastal—alluvial aggradations. As a result, the paddy soil is easily enriched by heavy metals (Zhejiang Soil Survey Office, 1994). With the rapid development of industrialization in recent years, more and more township enterprises in HJH Plain present potential heavy metal pollution. Meanwhile, the pesticide and fertilizer application add to the heavy metal pollution.

Fig. 4 presents the spatial patterns of five heavy metals in soils of HJH plain generated from their semivariograms. From Fig. 4, it can be seen that all heavy metals had distinct geographical distribution. The spatial distribution maps for Cu and Cr showed similar geographical trends, with low concentrations in west area and high concentrations in east area. There possess of most quantities of pigs in Jiaxing, Jiashan, Pinghu, Haiyan and Tongxiang in 2001 (Zhejiang Provincial Bureau of Statistics, 2001), so the high Cu concentrations here should owe to much manure application. In terms of Cr, its background value in these counties was high (Zhejiang Soil Survey Office, 1994). The spatial patterns for Zn, Pb and...
Cd also showed similar geographical trend, especially for soil Pb and Cd, they were low in central HJH plain. The higher concentrations of Zn, Pb and Cd were mainly due to the pollution sources from industry. The similarities of the spatial distribution for Cu and Cr were consistent with that they had similar semivariograms, likewise, for Zn, Pb and Cd.

To understand the effect of soil property on heavy metals, the correlativity between five heavy metals and soil properties (grain size, pH, OM, CEC) was analyzed (Table 3). The result showed that soil Cu was in positive correlation with Cr ($p < 0.01$). The Zn was, meanwhile, in positive correlation with Pb and Cd ($p < 0.01$). In addition, Pb was significantly in positive correlation with Cd ($p < 0.01$), and the correlation coefficient achieve to 0.816. The correlativity between five metals was just in accordance with the similarities in their spatial distributions.

The estimated probability of excess for Cu, Zn and Cd, defined by the thresholds in Table 1, was kriged by disjunctive kriging and given in Fig. 5. For soil Cu, the map shows that the areas with high risk are mainly located in South Yuhang, East Hangzhou and Haiyan, etc., where the estimated probability ($U_{Cu_{R}}$) reached 0.05–0.12. According to the environmental quality standard for soils, it would contribute to the environmental pollution and ultimately threaten the health of plant and human when Cu concentration in soils exceeds the value of 50 mg/kg. However, the areas with low estimated probability can be regarded as safe for rice growth. It should be noted that the areas with high Cu concentrations were always used for vegetable cultivation. Therefore, the high pollution risk in these areas should be due to the much pesticide applying on vegetable.

The probability map of Zn exceeding the guide value 200 mg/kg exhibits many risk patches, where the pH values are meanwhile high, and this is just corresponding to the above result that the Zn has positive correlation with pH ($p < 0.01$) (Table 3). Especially, the highest risk areas are distributed in Yuhang and Lin’an, due to the sewage and sludge application and higher background value in these areas (Zhejiang Soil Survey Office, 1994), and this could give an insight into the crop cultivation structure adjustment.

For soil Cd, the polluted areas in HJH Plain are dominant. The highest estimated probability ($U_{Cd_{R}}$) mainly concentrated in Anji, Yuhang, Hangzhou and Haiyan, 

![Fig. 3. Experimental semivariograms of soil heavy metals with fitted models.](Image)
and the estimated probability achieved to 0.11—0.25, presenting severe environmental pollution risk. The sewage irrigation and industrial pollution in these areas could have caused the Cd contamination. In the later agricultural policy making, it would be inappropriate still cultivating crop here. Maybe developing horticulture in these areas could be a rational decision-making.

### 3.4. Sampling density minimization

Due to economic and practical reasons, it is difficult to collect a great deal of soil samples, especially at large scale. So how to minimize the sampling density has always been an issue of concern. Cokriging has been proved to be an advantageous method to minimize samples. The real advantage of cokriging is to improve estimates with additional information from variables that are easier and cheaper to measure (Danielsson et al., 1998; Chang et al., 1998). The root mean square error (RMSE) was used to validate the error of estimates:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [Z(x_i) - Z^*(x_i)]^2}$$

where $Z(x_i)$ and $Z^*(x_i)$ represented true and estimated values, respectively. Lower the RMSE value is, higher the correlation coefficient ($R^2$) between true and estimated values, the estimate error is smaller.
It is indicated in Table 3 that the correlation between five metals and grain size is not significant, however, Cu, Zn and Cr all have positive correlation with CEC. So CEC was selected as the covariable in this study to estimate Cu, Zn and Cr using cokriging. Table 4 shows the estimated RMSE and $R^2$ values by kriging and cokriging methods, using various sampling densities. For Cu, the RMSE estimated by kriging is lower and $R^2$ is higher than that by cokriging using original sampling densities. When samples decrease to 400, 370, and even 350, the RMSE is still lower and $R^2$ is higher than that estimated using 450 samples by kriging. However, when the samples decrease to 300, though the RMSE is still lower than 4.979, the $R^2$ value appears to decrease, nevertheless, presenting positive correlation significantly ($p < 0.01$). Therefore, the sampling

<table>
<thead>
<tr>
<th>Heavy metals</th>
<th>Method</th>
<th>Samples</th>
<th>RMSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>Kriging</td>
<td>450</td>
<td>5.979</td>
<td>0.3232**</td>
</tr>
<tr>
<td></td>
<td>Cokriging</td>
<td>450</td>
<td>5.204</td>
<td>0.4085**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>5.447</td>
<td>0.3878**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>370</td>
<td>5.599</td>
<td>0.3766**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>350</td>
<td>5.609</td>
<td>0.3814**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>5.884</td>
<td>0.3216**</td>
</tr>
<tr>
<td>Zn</td>
<td>Kriging</td>
<td>450</td>
<td>34.748</td>
<td>0.1271*</td>
</tr>
<tr>
<td></td>
<td>Cokriging</td>
<td>450</td>
<td>29.893</td>
<td>0.1919*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>31.986</td>
<td>0.1454*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>370</td>
<td>31.936</td>
<td>0.1408*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>350</td>
<td>33.636</td>
<td>0.1287*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>34.046</td>
<td>0.1264*</td>
</tr>
<tr>
<td>Cr</td>
<td>Kriging</td>
<td>450</td>
<td>26.253</td>
<td>0.2341**</td>
</tr>
<tr>
<td></td>
<td>Cokriging</td>
<td>450</td>
<td>25.392</td>
<td>0.2577**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>24.946</td>
<td>0.2490**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>370</td>
<td>25.018</td>
<td>0.2453**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>350</td>
<td>25.569</td>
<td>0.2347**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>25.879</td>
<td>0.2305**</td>
</tr>
</tbody>
</table>

* $p < 0.05$; ** $p < 0.01$. 

Fig. 5. The estimated probability maps of Cu, Zn and Cd.
density of soil Cu in HJH plain can be reduced to 150 samples when cokriging method was used.

For soil Zn, the estimated error is much smaller using 450 samples by cokriging than by kriging, which can be judged from the RMSE and $R^2$ values. When reducing 150 samples, RMSE is still lower compared to the value of 34.748, which was estimated using 450 samples by kriging. However, $R^2$ is 0.1246, lower than 0.1271. Therefore, 150 samples can be reduced when cokriging method was used to estimate the Zn concentration in HJH plain.

As for Cr, 150 soil samples can, likewise, be reduced using cokriging method when CEC acted as the covariable, which can be illuminated by the RMSE and $R^2$ values in Table 4.

### 4. Conclusions

These results showed that the spatial variability of the soil heavy metals in paddy field were apparent in this study. Over long history of land management, the spatial variability of heavy metals was determined not only based on the characteristics of the soil properties but also the human activity. Among the five metals, Cu, Zn and Cd had high risk for environment pollution and human health. These results indicated the potential of agricultural cultivation adjustments in the polluted areas.

Selecting CEC as the covariable, the overall estimation quality of Cu, Zn and Cr by cokriging was better than that by kriging. Furthermore, the soil sampling density could be reduced to 150 for three metals, which would not only decrease the cost, but also save time for us.

### Acknowledgements

This research was sponsored in part by the National Basic Research Program of China (2005CB121104), the National Natural Science Foundation of China (40371062), and the Natural Science Foundation of Zhejiang Province (Y504128).

### References


