Spatially Nonstationary Relationships between Copper Accumulation in Rice Grain and Some Related Soil Properties in Paddy Fields at a Regional Scale

Mingkai Qu
Key Lab. of Soil Environment and Pollution Remediation
Institute of Soil Science
Chinese Academy of Sciences
210008 Nanjing
China

Weidong Li
Chuanrong Zhang
Dep. of Geography and Center for Environmental Sciences and Engineering
Univ. of Connecticut
Storrs, CT 06269

Biao Huang*
Yongcun Zhao
Key Lab. of Soil Environment and Pollution Remediation
Institute of Soil Science
Chinese Academy of Sciences
210008 Nanjing
China

Previous studies on the relationships between trace metal accumulation in plants and related soil properties have commonly assumed spatial stationarity by using traditional (global) regression techniques such as ordinary least squares regression (OLS). In this study, geographically weighted regression (GWR) was used to explore the relationships between the Cu concentration in rice (Oryza sativa L.) grain and a set of perceived soil properties, including total Cu in the soil, pH, and soil organic matter in Jiaxing, China. It was found that GWR performed much better than OLS in characterizing the relationships, in terms of the coefficient of determination ($R^2$), corrected Akaike information criterion, ANOVA test, sum of squared residuals, and spatial autocorrelations of residuals. The GWR analysis showed that the relationships between grain Cu and the three related soil properties were not constant across space and there was great spatial non-stationarity. Results also indicated that GWR could reveal the spatial variations of the relationships ignored by OLS and thus could be used to explain the local causes of Cu accumulation in rice grain. This study demonstrated the advantages of local spatial modeling techniques in modeling the accumulation of hazardous materials in soil–plant systems at regional scales.

Abbreviations: AICc, corrected Akaike information criterion; CEC, cation exchange capacity; GWR, geographically weighted regression; OLS, ordinary least squares regression; SOM, soil organic matter.

With rapid industrialization and urbanization in recent decades, pollution by trace metals has become a global problem and consequently has been a focus of environmental studies (Adriano, 2001; Huang and Gobran, 2005). Unlike many organisms and radionuclides, trace metals in soils do not decay with time due to their characteristics of non-biodegradability and persistence (Raghunath et al., 1999; Adriano, 2001; Huang and Gobran, 2005; Liu et al., 2006; Shi et al., 2007; Chen et al., 2008). Trace metals in soils may be present in one or more of the following forms: (i) dissolved (in the soil solution), (ii) exchangeable (in organic and inorganic components), (iii) structural components of lattices of soil minerals, and (iv) insoluble precipitates combined with other soil components (Zalidis et al., 1999; Aydinalp and Marinova, 2003; Zeng et al., 2011). The first two forms are available to plants, while the other two are potentially available in the long term. Many studies have shown that the availability of trace metals is associated with soil properties, including pH, soil organic matter (SOM) content, cation exchange capacity (CEC), oxidation–reduction status (Eh), and contents of clay minerals, CaCO$_3$, and Fe and Mn oxides (Kashem and Singh, 2001; Antoniadis et al., 2008; Usman et al., 2008). Among these soil
properties, it was found that soil pH might play the most important role in determining metal speciation, solubilization from mineral surfaces, movement, and eventual bioavailability due to its strong effects on the solubility and speciation of metals both in soil bodies and particularly in soil solutions (Muhlbachova et al., 2005; Zeng et al., 2011). Apart from soil pH, SOM is also one of the most important soil properties affecting the availability of trace metals. Soil organic matter is a major contributor to the capability of soils to retain trace metals in an exchangeable form (Zeng et al., 2011). In addition, organic matter also supplies soil solutions with organic chemicals that can serve as chelates, thus increasing metal availability to plants (Adriano, 2001; Huang and Gobran, 2005).

Previous studies on the influence of soil properties, such as pH and organic matter content, on trace metal uptake by plants commonly assumed spatial stationarity by using traditional global regression techniques (Li et al., 1998; Wang et al., 2013a). Although global multivariate regressions such as ordinary least squares regression (OLS) were relatively well established, the regression analyses conducted for soil–plant systems were commonly nonspatial, thus ignoring the local characteristics of the relationships between a dependent variable and its explanatory variables (Foody, 2003; Zeng et al., 2011; Wang et al., 2013b). In fact, unlike universal physical laws, observed geographical and ecological patterns and processes in the natural world tend to be spatially variable (Dutilleul and Legendre, 1993). In other words, even though the underlying natural processes are universal, actual spatial patterns still vary with local conditions (Jetz et al., 2005). This phenomenon is often referred to as spatial non-stationarity. Generally, the concentration of a trace metal in plants and its impact factors (e.g., the metal’s total concentrations in the soil, soil pH, and SOM content) are all spatial variables, and their values with closer separation distances may have stronger correlation than those farther apart. Conventional regression analysis methods such as the OLS model are based on the assumption of independence among observations and thus fail to capture the spatial dependence of the observed data when they are applied to georeferenced data analyses.

Recently, a local spatial statistical method called geographically weighted regression (GWR) has emerged for evaluating how the relationships between a dependent variable and its explanatory variables change spatially (Brunsdon et al., 1998; Fotheringham et al., 2002). This method has been extensively used in human geography in the last decade, and it is becoming a widely used local regression technique in different fields, such as ecological modeling (Shi et al., 2006), soil mapping (Wang et al., 2013b), social policy (Farrow et al., 2005), and urban studies (Yu, 2006). In soil mapping, the applications of GWR are still relatively rare and studies began to appear only in the last few years (Zhang et al., 2009, 2011; Scull, 2010; Mishra et al., 2010; Mishra and Riley, 2012; Kumar and Lal, 2011; Kumar et al., 2012; Terrón et al., 2011; Wang et al., 2012, 2013b). In these soil studies, GWR was mainly used as an interpolator and generally achieved better results than OLS and/or kriging methods. The local parameters of GWR are estimated using a locally weighted least squares estimator through a distance decay function, which assumes that a regression point is more affected by near observation data than by farther ones (Fotheringham et al., 2002). Thus, they may be different for different regression points. These local parameters generated by GWR can be visualized in a series of maps, so that the spatial patterns of estimated parameters can be investigated. Therefore, GWR may be a useful tool to explore the spatially varying relationships between trace metal accumulation in plants and related soil properties.

Rice is the dominant agricultural product in China and ranks second by quantity in the world (Zhao et al., 2011). Trace metals in rice may accumulate in the human body through the food chain, bringing ultimately fatal hazards to human health (Coen et al., 2001; Williams et al., 2009). Therefore, it is imperative to estimate the effect of soil properties on the uptake of trace metals by rice so as to minimize translocation of trace metals to the food chain. In this study, a regional-scale survey in Jiaxing, China, was performed to investigate the regression relationships between Cu accumulation in rice grain and some related soil properties, which included total Cu in the soil, soil pH, and soil organic matter (SOM), under real field conditions. Our specific objectives were to: (i) understand the spatial variability of grain Cu, soil Cu, pH, and SOM and their relations; (ii) compare the performance of OLS and GWR in modeling the regression relationships between grain Cu and related soil properties in the study area; and (iii) explore the spatially non-stationary relationships between grain Cu and related soil properties using GWR. It should be noted that it was not our intention to perform a comprehensive survey of the regression relationships between trace metal accumulation in rice and all of those related soil properties but rather to use grain Cu and a set of perceived related soil properties in a GWR case study to demonstrate the importance of local spatial modeling techniques in studying the accumulation of hazardous materials in soil–plant systems at a regional scale.

MATERIALS AND METHODS

Study Area and Sampling

The study was performed in Jiaxing (120°18′–121°16′ E, 30°21′–31°2′ N), a part of the Taihu Plain in northeastern Zhejiang Province, China, with an area of 3915 km² and a population of 3,356,000 (at the end of 2006). Jiaxing is close to the East China Sea, with a warm and humid subtropical climate. The annual mean temperature and rainfall are 15.9°C and 1168 mm, respectively. Due to the distribution of a dense river network and abundant water resources, Jiaxing is a traditional agricultural area, where rice is the major food crop. According to the soil taxonomy of China, Anthrosol (Inceptisol), which derived from lacustrine alluvium through long-term paddy cultivation, is the major soil order in this area (Gong et al., 2003). Since the early 1990s, this area has experienced a remarkable development in industry. At present, companies and factories in various industrial sectors, including chemical engineering, textiles, and...
printing and dying as traditional sectors, electronic information and equipment manufacturing as pillar sectors, and clean energy, advanced materials, and bioengineering as high-tech sectors are distributed within this area.

At rice harvest time (October 2006), 160 pairs of samples of rice grain and surface soils were collected in rice fields in the study area (Fig. 1). The rice genotype in the study area was mainly the japonica rice cultivar Xiushui 63. The sampling locations were randomly arranged throughout the whole area. For each pair of samples, the rice grain and surface soil (0–15-cm depth) were collected from the same location. Each sample was comprised of four to five subsamples within a distance of 10 m surrounding a specific sampling location. The weights of soil and rice grain subsamples were 1.0 and 0.5 kg, respectively. The composite samples were then numbered, with each of them stored in a plastic bag. The coordinates of the sampling locations were georeferenced using a handheld GPS receiver (MAP60CSX, Garmin Ltd.). In this study, sampling was done only in paddy fields for aquatic rice production, while mountain and urban areas were avoided.

**Chemical Analysis**

Soil samples were first air dried at room temperature (20–22°C), crushed after stones and debris were removed, and then sieved using a 2-mm nylon mesh. A portion of each sample (about 50 g) was then ground in an agate grinder to a particle size of <0.149 mm (i.e., passing through a no. 100 sieve). The prepared soil samples were stored in polyethylene bottles for later use. Grain samples were oven dried at 38°C until constant weights were achieved. After hulling, dried grain samples were ground using a stainless steel grinder (<0.25 mm) for trace metal analysis. The prepared rice grain samples were then stored in polyethylene bottles for further analysis.

The total Cu concentrations in both soil and rice samples were measured according to the national standard methods in China (Agricultural Chemistry Committee of China, 1983). The total Cu concentrations of the soil samples were determined using flame atomic absorption spectroscopy after 0.1-g soil samples were digested by concentrated HF–HNO$_3$–HClO$_4$ and brought to 25 mL with dilute HNO$_3$. The total Cu concentrations of rice grain samples were determined with graphite furnace atomic absorption spectroscopy after 1-g grain samples were digested by concentrated HNO$_3$/HClO$_4$ (4:1) and brought to 25 mL with dilute HNO$_3$. The limits of detection were 1.25 and 0.01 mg kg$^{-1}$ for soil and grain, respectively. The pH values of the soil samples were measured in suspensions (1:2.5 soil/water) with glass pH electrodes (Agricultural Chemistry Committee of China, 1983). The organic matter contents of the soil samples were obtained by the potassium bichromate wet combustion procedure (Agricultural Chemistry Committee of China, 1983). For quality control purposes, duplicates, method blanks, and standard reference materials were also analyzed.

**Kriging Interpolation**

To identify the effects of different soil properties on rice Cu accumulation, we compared the spatial distribution patterns of grain Cu and three related soil properties (i.e., soil Cu, pH, and SOM). Ordinary kriging (OK), a widely used interpolation method for continuous variables, was used for this purpose. Ordinary kriging provides a method to predict the values of a spatial variable at unsampled locations from its values at sampled locations by taking into account the spatial correlation of the sampled points (Qu et al., 2012). It can minimize the variance of estimation errors and investigation costs (Saito et al., 2005). The estimate at an unsampled location $x_0$ in OK is given by a linearly weighted average of observation data in the neighborhood of the unsampled location being estimated:

$$\hat{y}(x_0) = \sum_{i=1}^{N} \lambda_i y(x_i)$$  \[1\]

where $\hat{y}(x_0)$ is the predicted value at location $x_0$; $y(x_i)$ is the measured value of the variable under study at location $x_i$; $\lambda_i$ is the corresponding weight obtained from the OK equation system with $\sum_{i=1}^{N} \lambda_i = 1$; and $N$ is the number of sample data within the neighborhood. A detailed description of kriging can be found in Goovaerts (1997).

**Global Regression and Geographically Weighted Regression**

Traditional regression methods such as OLS are global statistics, which assume that the relationships under study are constant across space. An OLS model can be expressed as

$$y_j = \beta_0 + \sum_{i=1}^{p} \beta_i x_{ij} + \varepsilon_j$$  \[2\]

![Fig. 1. The spatial distribution of sampling points.](image-url)
where $y_i$, $X_{ij}$, and $e_j$ represent the dependent variable, the independent variables, and the Gaussian error term at the $i$th spatial position with $j$ independent variables; $\beta_0$ is the model intercept, and $\beta_j$ is the regression coefficient (i.e., slope) for the $j$th independent variable. This model is aspatial, that is to say, geographical location information is not considered in the estimation of model parameters. Conventional regression analysis methods such as OLS are based on the assumption of observations being independent of each other, which fails to capture the spatial dependence of the data when applied to georeferenced data analysis.

However, GWR uses local parameters for estimation and relationship analysis at each model calibration location and can be expressed as

$$
y_j = \beta_0(u_j, v_j) + \sum_{i=1}^{n} \beta_j(u_j, v_j) X_{ij} + \epsilon_j$$

where $u_j$ and $v_j$ are the coordinates for each $j$th location, $\beta_0(u_j, v_j)$ is the intercept for the $j$th location, and $\beta_j(u_j, v_j)$ is the local parameter estimate for independent variable $x_i$ at the $j$th location. The vector of estimated regression coefficients at the $j$th location is obtained by

$$\hat{\beta}(u_j, v_j) = \left[ X^T W(u_j, v_j) X \right]^{-1} \times X^T W(u_j, v_j) Y $$

where $\hat{\beta}(u_j, v_j)$ represents the vector of unbiased estimates of the regression coefficients; $X$ is an $(m \times (n + 1))$ data matrix of the independent variables, with $m$ and $n$ being the numbers of observed data and independent variables, respectively; $Y$ is an $(m \times 1)$ data vector of the dependent variable; and $W(u_j, v_j)$ is the local weights matrix, which is calculated from a kernel function that places more weights on locations spatially closer to the calibration location (Fotheringham et al., 2002). The weighting, therefore, follows the assumption of spatial autocorrelation (Wheeler and Páez, 2009). Because the sample density is not uniform across the study area, we used the following adaptive bisquare kernel function form for weight estimates in the study:

$$w_{ij} = \begin{cases} 
\left( \frac{1-d_{ij}^2}{\theta_{(k)}^2} \right)^2 & d_{ij} < \theta_{(k)} \\
0 & d_{ij} > \theta_{(k)} 
\end{cases}$$

where $w_{ij}$ is the weight value of the observation at the $j$th location for estimating the coefficient at the $i$th location; $d_{ij}$ is the Euclidean distance between $i$ and $j$; and $\theta_{(k)}$ is an adaptive bandwidth size defined as the $k$th nearest neighbor distance. In the study, the corrected Akaike information criterion (AICc) as described by Fotheringham et al. (2002) was adopted to determine the optimal adaptive bandwidth, and 32 surrounding data points were finally used for calibrating the GWR model at each location.

### Comparison of Geographically Weighted and Ordinary Least Squares Regression

Several methods were used to assess whether GWR can improve the OLS model. First, three indices—the AICc (Akaike, 1973; Hurvich and Tsai, 1989), the coefficient of determination ($R^2$), and the sum of squared residuals—were calculated for the two models for comparison purposes. A lower AICc value means that the corresponding model has a better fit to the observed data and consequently better model performance. Usually, a model can be considered better than another model if its AICc value is at least three points lower. Second, an ANOVA (Leung et al., 2000; Fotheringham et al., 2002), which tests the null hypothesis that GWR has no improvement over the OLS model, was conducted. Third, the spatial variability of local parameter estimates was evaluated. The test of spatial non-stationarity is important to determine whether parameter estimates in GWR analysis significantly vary across the spatial area or not. If any one of independent variables shows spatial stationarity in the test, a mixed GWR model may be more appropriate (Fotheringham et al., 2002; Yu, 2006). To assess the variability of the $k$ coefficient, two GWR models are compared: a GWR model usually fitted by local data (usual GWR) and a GWR model with the $k$ coefficient being constant (mixed GWR). If the usual GWR model is better than the mixed GWR model, as judged by a model comparison criterion such as the AICc, we can claim that the $k$ coefficient varies spatially. The test routine repeats this comparison for each coefficient. Details about the test of geographical variability of coefficients were described by Nakaya et al. (2009). Moreover, the Moran’s indices (i.e., Moran’s $I$) of residuals from the OLS and GWR models were computed to compare their abilities in dealing with spatial autocorrelation. Moran’s $I$ is commonly used as an indicator of spatial autocorrelation, and its values range from $-1$ to $1$ (Moran, 1950). The larger the absolute value of Moran’s $I$, the more significant the spatial autocorrelation. A value of 0 means almost perfect spatial randomness.

### RESULTS AND DISCUSSIONS

#### Descriptive Statistics

Some descriptive statistics of sample data for the Cu concentration in rice grain and three soil properties (i.e., soil Cu, pH, and SOM) are listed in Table 1. Grain Cu, soil Cu, and SOM all have wide value ranges. The coefficient of variation (CV) implies low variability when it has a value of <10% and extensive variability when it is >90% (Zhang et al., 2007). The CVs for grain Cu, soil Cu, pH, and SOM were 36.8, 53.5, 9.6, and 35.9%, respectively. This indicates that grain Cu, soil Cu, and SOM have moderate variability in the study area, while pH has low variability. The relatively larger variations of the three

---

**Table 1**: Distribution of Cu concentration and soil properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Coefficient of Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain Cu</td>
<td>2.584</td>
<td>40.27</td>
<td>12.5</td>
<td>6.34</td>
<td>53.5%</td>
</tr>
<tr>
<td>Soil Cu</td>
<td>0.132</td>
<td>8.02</td>
<td>2.5</td>
<td>1.54</td>
<td>53.5%</td>
</tr>
<tr>
<td>pH</td>
<td>4.20</td>
<td>7.20</td>
<td>6.0</td>
<td>0.60</td>
<td>9.6%</td>
</tr>
<tr>
<td>SOM</td>
<td>0.001</td>
<td>5.60</td>
<td>0.3</td>
<td>0.20</td>
<td>35.9%</td>
</tr>
</tbody>
</table>
properties (grain Cu, soil Cu, and SOM) may result from some extrinsic factors that are more influential on these properties.

The background value of soil Cu concentration in Zhejiang Province, estimated by China National Environmental Monitoring Centre, is 30.54 mg kg\(^{-1}\) (China National Environmental Monitoring Centre, 1990). Based on this background value, the total Cu concentrations in 67.5% of the soil samples in our study exceeded the baseline. This means that Cu accumulation is widely present in agricultural soils of the study area. Based on Chinese Environmental Quality Standard for Soils (Ministry of Environmental Protection of China, 1995), however, only 15 samples have Cu concentrations exceeding the guideline value for total Cu in soil (50 mg kg\(^{-1}\) under the condition of soil pH < 6.5 and 100 mg kg\(^{-1}\) when pH is > 6.5). This indicates that paddy soils in most of the study area have total Cu concentrations below the contamination level. The benchmark of contamination for grain Cu recommended by the Ministry of Health of China is 10 mg kg\(^{-1}\) (Ministry of Health of China, 2005). In this study, no rice sample reached this level. Therefore, the rice produced by this area may be regarded as being safe in terms of only the grain Cu content. In fact, none of the 15 locations with the highest soil Cu concentrations had a grain Cu concentration higher than the 15th highest grain Cu concentration. This means that rice uptake of Cu from soil is not proportional to the total Cu content in the soil. This is normal because only available Cu in the soil can be absorbed by rice. However, the concentration of available Cu in the soil is related to many factors, including soil Cu concentration, soil pH, SOM content and type, CEC, and others.

Spatial Distribution Patterns of Grain Copper and Three Related Soil Properties

The spatial distribution patterns of the three explanatory variables were compared with that of grain Cu to identify the effects of different soil properties on rice grain Cu accumulation. Because no apparent anisotropy was found from the sample data, experimental variograms were estimated omni-directionally for each property. Parameters of each variogram model are presented in Table 2. The experimental variogram of pH was fitted by an exponential model, and the others were all fitted by spherical models. The \(C_0/(C_0 + C)\) ratios of fitted variogram models all fall between 25 and 75%, which means that all properties exhibited moderate spatial autocorrelation. According to Cambardella et al. (1994), this situation may be attributed to both intrinsic factors such as other soil properties and extrinsic factors such as human activities (e.g., land use and industrial impacts).

Ordinary kriging was used to map the spatial distributions of total Cu in rice grain and the three related soil properties (i.e., soil Cu, pH, and SOM), as presented in Fig. 2. The spatial pattern of grain Cu shows much similarity to those of soil Cu (except for the central part of the study area) and SOM (except for the middle of the western boundary strip), while pH exhibits approximately a reverse pattern. These comparisons indicate that the three soil properties may be important factors that can impact the accumulation of Cu in rice grain.

Global Effects of Soil Properties on Copper Accumulation in Rice Grain

To characterize the global influence of the three selected soil properties on Cu accumulation in rice grain, we calculated the Pearson’s nonparametric correlation coefficients between each soil property and grain Cu (Table 3). The results indicated that grain Cu is positively correlated with soil Cu (correlation coefficient \(r = 0.732\)) and SOM (\(r = 0.437\)) and negatively correlated with soil pH (\(r = 0.459\)), and the correlation degrees are moderate to high, all being significant at the 0.01 level. These correlation relationships indicate that soil Cu, soil pH, and SOM all play important roles in Cu accumulation in rice grain, and soil Cu is the most important factor among the three determinants because it has the highest correlation coefficient with grain Cu. However, the correlation coefficients among the three soil properties are not significant at the 0.01 level. Therefore, this study tried to build regression models of grain Cu on soil Cu, pH, and SOM.

A quantitative and statistically significant (\(p < 0.01\)) relationship between grain Cu and the three impact factors was established in an OLS model

\[
Y_{\text{grain Cu}} = 2.9163 + 0.0286X_{\text{soil Cu}} - 0.2849X_{\text{pH}} + 0.0189X_{\text{SOM}}
\]

with variance inflation factor VIF = 1.3 and \(R^2 = 0.45\). Here the VIF measures redundancy among explanatory variables. As a rule of thumb, if an explanatory variable has a VIF value > 7.5, it should be removed from the regression model. The VIF

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model</th>
<th>Nugget (C_0)</th>
<th>Partial sill C</th>
<th>(C_0/(C_0 + C))</th>
<th>Range</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain Cu conc.</td>
<td>spherical</td>
<td>0.4</td>
<td>0.7</td>
<td>33.3</td>
<td>17.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Soil Cu conc.</td>
<td>spherical</td>
<td>346.7</td>
<td>472.7</td>
<td>42.3</td>
<td>23.8</td>
<td>0.8</td>
</tr>
<tr>
<td>pH</td>
<td>exponential</td>
<td>0.2</td>
<td>0.4</td>
<td>38.1</td>
<td>20.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Soil organic matter</td>
<td>spherical</td>
<td>45.7</td>
<td>76.1</td>
<td>37.5</td>
<td>18.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>

† The unit of \(C_0\) and C is (mg kg\(^{-1}\))\(^2\) for grain Cu and soil Cu, (g kg\(^{-1}\))\(^2\) for SOM. The \(C_0\) and C of pH have no units.

Table 1. Summary statistics for Cu concentrations in rice grain and three related soil properties in paddy fields.

Table 2. Parameters of the variogram models of Cu concentrations in rice grain and soil and soil pH and organic matter content (SOM) in the study area.
values for all explanatory variables in Eq. [6] were 1.3, indicating that multicollinearity was low among these variables, which was further verified by the minimal correlations between the explanatory variables (Table 3). In addition, the OLS model has $R^2 = 0.45$, which means that about 45% of the variation in grain Cu concentrations in the study area can be explained by the three independent variables (i.e., soil Cu, pH, and SOM).

### Comparison of Model Performances of Ordinary Least Squares and Geographically Weighted Regression

The non-stationarity test showed that the usual GWR model was better than the mixed GWR models in terms of the AICc index (not shown). Thus we can conclude that all explanatory variables exhibited spatial non-stationarity, and the usual GWR model was adopted in this study. Comparison of OLS and GWR showed that the latter model had a better overall fit performance, characterized by increases in $R^2$ values and decreases in AICc and the sum of squared residuals (Table 4). To further compare the two models, an ANOVA (Fotheringham et al., 2002) was used to test whether the GWR model offered an improvement over the OLS model or not. The result of the test indicated that the improvement of the GWR model over the OLS model is statistically significant ($p < 0.01$) based on 160 sampling data (not shown). Therefore, as a candidate, the GWR model is a better choice for investigating the relationships between grain Cu and related soil properties.

In addition to the above comparisons, the spatial pattern of residual values in a regression analysis is an important index for evaluating the performance of model fit. In other words, the spatial randomness of residual values can be a diagnostic measure of the regression model. Whether the residual values are significantly clustered or not implies whether a key variable is missing from the regression model (Fotheringham et al., 2002). The global Moran’s $I$ was computed to examine the spatial autocorrelations of the residuals generated by the OLS and GWR models. The GWR model

---

**Table 3. Pearson’s correlation coefficients between rice grain Cu concentration and related soil properties.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Grain Cu conc.</th>
<th>Soil Cu conc.</th>
<th>pH</th>
<th>Soil organic matter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain Cu conc.</td>
<td>1</td>
<td>0.732 **</td>
<td>−0.459 **</td>
<td>0.437 **</td>
</tr>
<tr>
<td>Soil Cu conc.</td>
<td></td>
<td>−0.161 *</td>
<td>0.110</td>
<td></td>
</tr>
<tr>
<td>pH</td>
<td></td>
<td></td>
<td>1</td>
<td>0.071</td>
</tr>
<tr>
<td>Soil organic matter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.05 level (two-tailed)

**Correlation is significant at the 0.01 level (two-tailed).
gave a smaller global Moran’s index ($I = 0.15$) than the OLS model ($I = 0.35$), suggesting that the spatial pattern of the residuals returned by the GWR model was more random than that returned by OLS. This result indicates that the GWR model represents an improvement over the OLS model by reducing the spatial autocorrelation in the residuals. Although some clusters of residuals may still remain in a GWR analysis, the data autocorrelation problem is usually reduced. This issue may be worthy of further study because some other factors influencing the accumulation of Cu in rice grain may still be missing in the regression.

Global regression methods, such as OLS, have been widely used in investigating the relationships between trace metal accumulation and soil properties. However, this type of regression is based on the assumption of independence of field observations, failing to capture spatial data dependence when applied to georeferenced data. That is to say, the global relationships between rice grain Cu and related soil properties in paddy fields derived from an OLS regression analysis may deviate considerably from those observed locally and may in fact never provide an objective description of the real relationships at any site but rather an average impression of the relationships across an entire region. However, the local relationships (i.e., local regression coefficients) derived from a GWR analysis can provide a clear image of how rice grain Cu is correlated with (or impacted by) its environmental impact factors at different places. Using local regression coefficients rather than a set of global regression coefficients is the main reason for GWR to perform better than OLS. That is to say, local regression coefficients play a significant role in the GWR model, and the role of a specific environmental factor in the GWR model changes with the change in its regression coefficient value across space.

**Spatially Nonstationary Relationships**

The spatial variation maps of regression parameters (i.e., slopes and the intercept) from the GWR analysis of Cu in rice grain against the three explanatory variables together are presented in Fig. 3. These regression parameters estimated in the GWR analysis are not constant across space; that is to say, the relationships between grain Cu and the related soil properties are spatially non-stationary, varying greatly from one location to another across the study area (Fig. 3). The influences of the three

![Spatial variation maps of regression parameters from the geographically weighted regression analysis of Cu in rice grain against three explanatory variables: (a) model intercept, (b) regression coefficient of soil Cu concentration, (c) regression coefficient of soil pH, and (d) regression coefficient of soil organic matter (SOM).](image-url)

---

**Table 4. Comparison between ordinary least squares (OLS) and geographically weighted regression (GWR) models.**

<table>
<thead>
<tr>
<th>Model</th>
<th>AICc†</th>
<th>$R^2$</th>
<th>Moran’s I of residuals</th>
<th>Sum of squared residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>386.91</td>
<td>0.45</td>
<td>0.35</td>
<td>64.89</td>
</tr>
<tr>
<td>GWR</td>
<td>320.01</td>
<td>0.75</td>
<td>0.15</td>
<td>30.74</td>
</tr>
</tbody>
</table>

† Corrected Akaike information criterion.
environmental factors (i.e., soil Cu, pH, and SOM) on Cu accumulation in rice grain can be explained by their respective regression coefficients. A positive coefficient means a positive influence (i.e., correlation), while a negative coefficient indicates a negative influence. Similarly, a large absolute coefficient value means a strong correlation, while a small absolute coefficient value indicates a weak correlation. The influence of each environmental factor changes with the variation of its coefficient values across space. However, such spatial non-stationarity cannot be captured by global regression techniques because they use only one set of values of the model parameters for the whole study region. The single estimates of parameters derived from a conventional global regression method do not represent local conditions or even the condition at any site within the study area.

Metals exist mainly in four forms in soils: mineral, organic, absorbed (bound to the surface of soil clay particles), or dissolved. The majority of trace metals in soils is bound in minerals and passive SOM and therefore is unavailable to plants (Jones and Jacobsen, 2009). However, active SOM may enhance its availability to plants by increasing the CEC of the soil, providing metal chelates and increasing the solubility of nutrients in the soil solution (McCauley et al., 2009). Hence, the influence of SOM on trace metal accumulation in plants is dependent not only on its content but also on its components. Moreover, different types of metal minerals may not have the same bioavailability at the same pH level. The contents and components of soil Cu and SOM may not be the same across the study area. All of these can cause the effects of the selected soil properties (i.e., soil Cu, pH, and SOM) on the accumulation of Cu in rice grain to be nonconstant across space (i.e., show apparent spatial non-stationarity). Other factors, such as CEC, oxidation–reduction status (Eh), and contents of clay minerals, CaCO3, and Fe and Mn oxides, which were not selected or unavailable in this study, may also impact Cu accumulation in rice grain. The spatial variations of these hidden factors, if they have obvious influence, may also distort the relationships between grain Cu and the selected factors across space. Rice cultivars may also have an important influence on the accumulation of Cu in rice grain. In this study, a japonica rice cultivar Xiushui 63, one of the most important rice cultivars in Jiaxing, was used. If other rice cultivars were planted in the same study area, the relationships between Cu accumulation in rice grain and the related soil properties in paddy fields might not be exactly the same.

The relationships between grain Cu accumulation and related soil properties indicated in the OLS regression were significant with a $R^2$ value of 0.45. However, the GWR analysis highlighted local variations in the regression parameters and explained substantially more of the variance, with $R^2 = 0.75$. Notably, the explanatory power of the GWR model varies spatially, as demonstrated by the local estimates of $R^2$ for the model (see Fig. 4). The $R^2$ map for the GWR model has varying values from 0.58 to 0.98, all being larger than the global $R^2$ value (0.45) of OLS for the entire study area. This means that GWR performed better than OLS across all locations in the study area but with different degrees of improvement at different places. This finding and the spatial variations in the model parameter maps derived from the GWR analysis highlight the deficiency of using global parameters or global regression methods to account for the relationships between grain Cu and related soil properties. Note that in this study, adaptive bandwidth was used in GWR, and the bandwidth value changes with the density of sample points. As the bandwidth declines or enlarges, the analysis becomes increasingly local or global, revealing more or less geographical detail like some sort of spatial microscope (Fotheringham et al., 2002). The sparseness of sample points may be an important reason why the explanatory power of the GWR model is relatively weak in the central-northwestern zone of the study area (see Fig. 1 and 4).

**CONCLUSIONS**

This study investigated the relationships between Cu accumulation in rice grain and three related soil properties (i.e., soil Cu, pH, and SOM) in paddy fields in Jiaxing, China. The contents of soil Cu in 15 samples exceeded the guideline value based on the Chinese Environmental Quality Standard for Soils, although the grain Cu in the study area was still below the benchmark value of 10 mg kg$^{-1}$ recommended by the Ministry of Health of China. Variogram analysis showed that grain Cu, soil Cu, pH, and SOM all had strong spatial autocorrelations. The spatial distribution maps generated by ordinary kriging visually indicated that the three soil properties were correlated with Cu accumulation in rice grain, which was further verified by the Pearson’s correlation analysis. Comparison of the GWR model and the OLS model showed that GWR performed much better than OLS in characterizing the relationships between Cu accumulation in rice grain and the related soil properties in paddy fields at a regional scale in terms of some global indices, including the $R^2$, AICc, ANOVA, sum of the squared residuals, and spatial autocorrelations of residuals. Geographically weighted regression analysis further showed that the relationships between grain Cu and the related soil properties were not constant across space, indicating a great spatial non-stationarity. Geographically

![Fig. 4. The local estimate map of the coefficient of determination ($R^2$).](attachment:image.png)
weighted regression, therefore, could reveal the spatial variations of the relationships ignored by OLS and consequently might explain the local causes of Cu accumulation in rice grain. This study demonstrated the advantages of local spatial modeling techniques in modeling the accumulation of hazardous materials in soil–plant systems at regional scales.

Soil scientists and agricultural management agencies are generally concerned with how plant or soil properties are affected by different anthropogenic and natural factors, such as other soil properties, climate, and land use. Because these factors are spatial attributes and change across space, spatial non-stationarity may exist in their relationships with the target variable under study. A GWR model can capture such spatially non-stationary relationships and reveal their local characteristics so that decision makers may adopt appropriate management measures suitable to the local soil environment. Hopefully, GWR will be of great value to the general field of soil science.

ACKNOWLEDGMENTS

This work was supported by the Knowledge Innovation Program of Chinese Academy of Sciences (KZCX2-EW-QN404), the National Natural Science Foundation of China (41401523), and the Natural Science Foundation of Jiangsu Province (BK20141055).

REFERENCES


