Spatial autoregression model for heavy metals in Beijing cultivated soils

Xiaoni Huo[A,C], Danfeng Sun[A] and Hong Li[B]

[A]Department of Land Resources and Management, College of Natural Resources and Environmental Science, China Agricultural University, Beijing, China, Email hxnsky@126.com, sundf@cau.edu.cn
[B]Beijing Academy of Agriculture and Forestry, Beijing, China, Email lihsdf@sina.com
[C]Department of Environmental Engineering, Taiyuan University, Taiyuan, China

Abstract
Soil heavy metals may be affected by many influencing factors, in order to find out the important factors leading to pollution in Beijing cultivated soils for risk evaluation and pollution control, the relationship between heavy metals and the influencing factors was analysed. Because of the significant spatial autocorrelation of heavy metals and the influencing factors, the significant positive spatial autocorrelation was detected for the residuals of a standard linear regression model. To consider the spatial autocorrelation fully, a spatial autoregression model was constructed. Taking Cr as an example, the spatial autoregression model for Cr had a better goodness-of-fit than the standard linear regression model, while the spatial autocorrelation of the residuals disappeared, indicating the spatial autoregression model could capture the spatial autocorrelation of data and explain the relationship between heavy metals and their influencing factors excellently. Spatial autoregression model showed that the important controlling factors for Cr in Beijing cultivated soils were soil parent rock and land use intensity.

Key Words
Spatial autoregression model, heavy metals, cultivated soils, Beijing.

Introduction
Heavy metals in cultivated soils are controlled by many factors, such as the overuse of fertilizer and pesticides, the mineral resource exploration and industry development. These factors can lead to the excessive accumulation of heavy metals, which may not only result in soil contamination but also affect the groundwater and food chains (Khan *et al.* 2008). Generally, the relationship between heavy metals and these factors analysed with standard statistical models, and the data analysed with the methods should be independent and identically distributed (Cliff and Ord 1981). However, Huo *et al.* (2009a) have showed that heavy metals had significant spatial autocorrelation, which contains some useful information. If the standard statistical models are used, they cannot capture all the spatial autocorrelation characteristics of data. To overcome the defect, Cliff and Ord (1981) provided a spatial autoregression model to deal with spatial data. Recently, the spatial autoregression model has been extensively applied in many studies, but the method is still not used by heavy metal distribtion. Therefore, the objectives of this study were to explore the relationship between heavy metals in Beijing cultivated soils and their influencing factors using a spatial autoregression model, and to find out the important factors for soil heavy metal risk evaluation and pollution control.

Methods
The study area is Beijing province. A total of 1018 soil samples were collected from Beijing cultivated land in 2006. Huo *et al.* (2009b) have provided the details of the study area, the methods for soil sample processing and analysis and the distribution of soil samples.

Spatial autoregression model
The most general formulation of a spatial autoregression model is Eq. (1) (Anselin, 1988):

$$\begin{aligned} y &= \rho w_1 y + x\beta + u \\ u &= \lambda w_2 u + \varepsilon \\ \varepsilon &\sim N\left(0, \sigma^2 I_n\right) \end{aligned} \tag{1}$$

where y is the dependent variable, x is the explanatory variable, β is a vector correlated with explanatory variable, ρ is a coefficient of the spatially lagged dependent variable, λ is a coefficient of the spatially correlated errors and w_1, w_2 are spatial weight matrices.
Setting $\rho = \lambda = 0$ produces a standard regression model, without considering spatial autocorrelation characteristics. Setting $\beta = \lambda = 0$ and $\rho \neq 0$ results a first-order spatial autoregression model, showing that the dependent variable is affected by the dependent variable of neighbouring units. Setting $\rho \neq 0$, $\beta \neq 0$ and $\lambda = 0$

produces a mixed regression-spatial autoregression model, indicating that the dependent variable is not only affected by the explanatory variable, but affected by the dependent variable of neighbouring units. This model is also called spatial lagged model. Setting $\beta \neq 0$, $\lambda \neq 0$ and $\rho = 0$ results a spatial error autoregression model. This model indicates that the dependent variable is affected by explanatory variable, dependent variable and explanatory variable of neighbouring units.

Measures of fit in spatial autoregression model

In the presence of spatial autocorrelation, the traditional R^2 measure of fit is not applicable to the spatial autoregression model. Instead, the so-called pseudo R^2 measure can be computed to asses fit. In the standard regression model, the pseudo R^2 is equivalent to the R^2, but in the spatial autoregression model it is not (Anselin 2002). So, the traditional R^2 and the pseudo R^2 cannot be compared, but it is possible to compare the pseudo R^2 of different spatial models. The measures for goodness-of-fit for the spatial model based on the likelihood function also include the value of the maximised log likelihood (LIK), the Akaike Information Criterion (AIC) and the Schwartz Criterion (SC). The model with the highest LIK, or with the lowest AIC or SC has the best goodness-of-fit. The LIK is not a standardised indicator like R^2, so it cannot be interpreted as an absolute value.

The selection of influencing factors

Soil heavy metals are affected by both natural processes and anthropogenic activities. All of the factors must be considered fully, while, these data should be obtained and quantified simply. The distance between the cultivated soil and residential area, the distance between the cultivated soil and industrial and mining establishment, the density of road, and the density of different soil types were used to show the effect of residential area, industrial and mining establishment, road, and soil types on heavy metals. The land use intensity was defined by the proportion of fertilizer and pesticide inputs. The land use intensity of orchard land was the strongest, followed by two-crops a year, one-crop a year, and then fallow land. All factors were quantified using geostatistical analyst extension of ArcGis 8.3, and the data were 100×100 m raster data.

Results

Table 1 shows the spatial autocorrelation of the residuals presented in the standard linear regression model for Cr, Ni, Zn, and Hg. Although the spatial autocorrelation of the residuals were less than those of heavy metals, they were still significant, thus, it is necessary to select the spatial autoregression model.

Table 1. Moran's I for soil heavy metals and their residuals of the standard linear regression model.

Heavy metals	The Moran's I of heavy metal	The Moran's I of the residuals
Cr	0.4801	0.3615
Ni	0.3173	0.2689
Zn	0.2924	0.2449
Hg	0.2725	0.1878

Taking Cr as an example, Table 2 illustrates the difference between the standard linear regression model and the spatial autoregression model for Cr. Spatial autoregression model 1 contained the same variables as the standard linear regression model. The regressive coefficient of variables in spatial autoregression model 1 were smaller than those in the standard regression model, and the significant of parameters also decreased and some variables were not longer significant (p<0.05), but the LIK of spatial autoregression model was higher than the LIK of the standard linear regression model, indicating that the spatial autoregression model has a better goodness-of-fit. The insignificant variables were dropped from the spatial autoregression model 1 to construct the spatial autoregression model 2. The pseudo R^2 and the LIK of the spatial autoregression model 2 decreased slightly, but the parameters were still significant, thus, the difference between the two spatial models was negligible. The Moran's I of the residuals in spatial autoregression model 2 was -0.0565, which tended to 0, indicating the spatial autocorrelation of the residuals disappeared. Thus, the spatial autoregression model could capture the spatial autocorrelation of Cr and the influencing factors and showed the relationship between Cr and the influencing factors.

Conclusion

The residuals of the standard linear regression model for Cr, Ni, Zn, and Hg showed significant and positive autocorrelation, indicating that the standard linear regression model failed to consider all the spatial autocorrelation characteristics in the data. Spatial autoregression model for Cr yielded residuals without spatial autocorrelation but with a better goodness-of-fit, while the coefficients of the spatially correlated errors were

Table 2. The model parameters for three models for Cr.

Regression model	Variables	Regression coefficient	Standard error	*t*-test value	*p*
Standard linear regression model	Constant	1.7515	0.0149	117.3316	0
	The distance to the residential area	-0.0744	0.0245	-3.0406	0.0024
	Land use intensity	0.0094	0.0034	2.799	0.0052
	The density of haplic fluvo-aquic soil	-0.0113	0.0054	-2.1077	0.0353
	The density of aquic-cinnamon soil	-0.0411	0.0145	-2.8453	0.0045
	The density of haplic-cinnamon soil	0.0725	0.0168	4.3055	0
	The density of meadow cinnamon soil	-0.149	0.0255	-5.8497	0
	The density of leached drab soil	0.0329	0.014	2.3512	0.0189
	Measures of fit in standard linear regression model: R2=0.115; LIK=794.474				
Spatial autoregression model 1	Constant	1.74	0.015	117.578	0
	The distance to the residential area	-0.004	0.021	-0.193	0.847
	Land use intensity	0.006	0.003	2.229	0.026
	The density of haplic fluvo-aquic soil	0	0.005	0.006	0.996
	The density of aquic-cinnamon soil	0.001	0.016	0.077	0.939
	The density of haplic-cinnamon soil	0.044	0.019	2.338	0.019
	The density of meadow cinnamon soil	-0.037	0.027	-1.367	0.172
	The density of leached drab soil	-0.014	0.012	-1.186	0.235
	λ	0.66	0.031	21.61	0
	Measures of fit in spatial autoregression model 1: pseudo R2=0.394; LIK=948.943				
Spatial autoregression model 2	Constant	1.7427	0.0117	148.4755	0
	Land use intensity	0.0065	0.0028	2.2977	0.0216
	The density of haplic-cinnamon soil	0.0452	0.0188	2.3991	0.0164
	λ	0.6631	0.0303	21.8649	0
	Measures of fit in spatial autoregression model 2: pseudo R2=0.392; LIK=946.657				

significant ($p<0.05$), indicating that the concentrations of Cr not only were affected by the influencing factors of the own region, but also were affected by the concentrations of heavy metals and influencing factors of the neighbouring regions. The regression coefficient of spatially correlated error was higher than those of the influencing factors, indicating that the concentrations of Cr and influencing factors in the neighbouring regions played an important role for Cr concentrations. Spatial autoregression model showed the important influencing factors for Cr in Beijing cultivated soil were soil parent rock and land use intensity.

References

Anselin L (1988) 'Spatial Econometrics: Methods and Models'. (Kluwer Academic Publishers: Dordrech).

Anselin L (2002) Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics* **27**, 247-267.

Cliff AD, Ord J K (1981) 'Spatial processes: Models and Applications'. (Pion Ltd: London)

Huo XN, Li H, Sun DF, Zhang WW, Zhou LD, Li BG (2009a) Spatial autocorrelation analysis of heavy metals in cultivated soils in Beijing. *Acta Scientiae Circumstantiae* **29**, 1339-1344.

Huo XN, Li H, Sun DF, Zhou LD, Li BG (2009b) Multi-scale spatial structure of heavy metals in cultivated soils in Beijing. *Environmental Monitoring and Assessment* DOI: 10.1007/s10661-009-0916-7

Khan S, Cao Q, Zheng YM, Huang YZ, Zhu YG (2008) Health risks of heavy metals in contaminated soils and food crops irrigated with wastewater in Beijing, China. *Environmental Pollution* **152**, 686-692.