

Article

Application of Integrated Land Use Regression and Geographic Information Systems for Modeling the Spatial Distribution of Chromium in Agricultural Topsoil

Meng Cao ¹, Daoyuan Wang ² , Yichun Qian ¹, Ruyue Yu ³, Aizhong Ding ⁴ and Yuanfang Huang ^{3,*} 

¹ China International Engineering Consulting Corporation, Beijing 100048, China

² Department of Environmental Science and Engineering, Shanghai University, Shanghai 200444, China

³ College of Land Science and Technology, China Agricultural University, Beijing 100193, China

⁴ College of Water Sciences, Beijing Normal University, Beijing 100875, China

* Correspondence: yfhuang@cau.edu.cn

Abstract: Chromium (Cr) contamination is widely distributed in agricultural soil and poses a threat to agricultural sustainability. Developing integrated models based on soil survey data can be an effective measure to accurately predict the spatial distribution of Cr. Focused on an agriculturally dominated area, this study presents a novel hybrid mapping model that combines land use regression (LUR) and geostatistical methods to predict Cr distribution in topsoil and examines the influence of various influencing factors on Cr content. The LUR model was first adopted to screen the influencing factors for Cr predictions. Then LUR, was combined with ordinary Kriging (OK_LUR) and geographically weighted regression Kriging (GWRK_LUR) to describe the spatial distribution of Cr. Results showed that Cr distribution was profoundly influenced by soil Cu and Zn content, the distance between the soil sampling and livestock farm, orchard areas within 100 m, and population density within 1000 m. The developed GWRK_LUR model significantly improved the prediction accuracy of the OK_LUR and LUR models (by 9% and 16%, respectively). This model provides a novel route to account for the spatial distribution of Cr in agricultural topsoil at a regional scale, which has potential application in pollution remediation.

Keywords: chromium; land use regression; geographically weighted regression Kriging; spatial distribution; agricultural sustainability



Citation: Cao, M.; Wang, D.; Qian, Y.;

Yu, R.; Ding, A.; Huang, Y.

Application of Integrated Land Use

Regression and Geographic

Information Systems for Modeling the

Spatial Distribution of Chromium in

Agricultural Topsoil. *Sustainability*

2024, *16*, 5299. [https://doi.org/](https://doi.org/10.3390/su16135299)

[10.3390/su16135299](https://doi.org/10.3390/su16135299)

Academic Editor: Wentao Yang

Received: 22 May 2024

Revised: 14 June 2024

Accepted: 18 June 2024

Published: 21 June 2024



Copyright: © 2024 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license ([https://creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/)

[https://creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/)

4.0/).

1. Introduction

Heavy metal pollution of agricultural soil has received significant attention due to its great influence on food safety and human health [1,2]. Chromium (Cr) is widely distributed in agricultural soil, while hexavalent Cr is one of the most toxic heavy metals [3] and other valence states can be translated into hexavalent. Cr(VI) is not only more easily taken up by plants and affects the crop yield and quality, but also has harmful effects on humans and animals due to its carcinogenicity, mutagenicity, and genotoxicity. Human exposure to Cr is mainly derived from soil-crop systems because Cr is easily integrated into the food chain. Considering these potential risks to humans and plants, Cr contamination in soil has attracted worldwide attention [4].

It also has been reported that its distribution is influenced by a range of physicochemical properties of soil (e.g., soil type, geology, and pH) [5] and anthropogenic processes (e.g., industrialization, land use, traffic) [6]. Studies have confirmed that Cr pollution in agriculturally dominated areas (i.e., without the influence of industry) is mainly derived from the soil parent material [7]. Components within the soil parent material, including clay mineral colloids and functional groups, potentially exert a significant influence on the adsorption, desorption, migration, and transformation of Cr. Specifically, the mechanisms may differ considerably, depending on the type and inherent properties of the soil parent

material, environmental conditions, and the chemical character of the Cr present [8]. Both soil parent material and human activity can affect the distribution of Cr [9,10], making it challenging to evaluate Cr contamination in soils.

In soil Cr investigation, traditional surface sampling remains a reliable method for data collection [11]. However, given the high cost associated with soil surveys, various analytical techniques have been developed for Cr distribution mapping, such as multivariate statistical analysis and geostatistical methods [12,13]. Nevertheless, these methods have limitations. They potentially overlook fine-scale spatial changes in soil properties, and struggle to explain the specific factors that influence the spatial distribution of soil heavy metal contamination at a regional scale [14].

It is widely established that using correlated influencing factors as auxiliary variables can enhance prediction accuracy [14,15]. However, these models often fail to cover different variable types within a single model framework. Land use regression (LUR) employs ground observations and multiple geographic covariates, including land use type, to build a regression model for estimating concentrations in areas lacking monitoring data [16]. LUR has demonstrated its effectiveness in predicting the distribution of heavy metals [17]. Nevertheless, the relationship between these factors and heavy metal contents is rarely linear in nature [18], which may undermine the simulation performance of LUR models. Fortunately, geographically weighted regression (GWR) has a good ability to solve this problem. GWR can depict nonstationary spatial effects in the relationship between the dependent and independent variables through allowing model parameters to vary across space [19]. Geographically weighted regression Kriging (GWRK) integrates ordinary Kriging (OK) with GWR, enabling spatial prediction of unrecognized secondary factors contained in the GWR model.

Up to now, although researchers have adopted various methods to explore soil Cr distribution and affect factors, it is still unclear how Cr is distributed in agriculture dominated areas and which factors determine its distribution. Reasonable and high-precision mapping models for Cr spatial distribution are necessary but deficient. The objectives of this study were to identify the key factors influencing Cr distribution in agricultural topsoil in addition to the soil parent material, and to develop a spatial distribution model for Cr accounting for these influencing factors. We aimed to (1) develop a predictive mapping model for Cr concentrations in surface soil, (2) investigate the spatially variable relationship between the influencing factors and Cr concentrations, and (3) evaluate the overall distribution of Cr in topsoil in a typical agriculturally dominated area. The proposed model holds the potential not only to reveal the distribution of topsoil Cr in traditional agricultural areas but also to predict the spatial distribution of other soil pollutants at regional scales, thus enhancing our understanding and management of agricultural sustainability.

2. Materials and Methods

2.1. Soil Sampling and Chemical Analysis

The study area (116°55' E–117°24' E, 40°10' N–40°22' N, Figure 1) is situated in Pinggu district, Beijing, China. As a traditionally agriculture dominated region, it has a large proportion of agricultural land (66.26%) which is mainly cultivated as orchard, vegetable fields, and cropland. Fruit sales and crop production are major components of the local economy; meanwhile, the rapid development of the agricultural economy in this area may have concomitantly increased soil pollution levels. The mountainous terrain is dominant in this district, and the main soil parent material is alluvial–fluvial matter. The main soil texture types are cinnamon soil and fluvo-aquic soil.

The sampling design also considered the distribution of Cr among various areas, soil types, and topographical positions; 208 sampling sites were ultimately collected in agricultural soil (Figure 1a). The geographic coordinates of these sampling sites were recorded using a global positioning system.

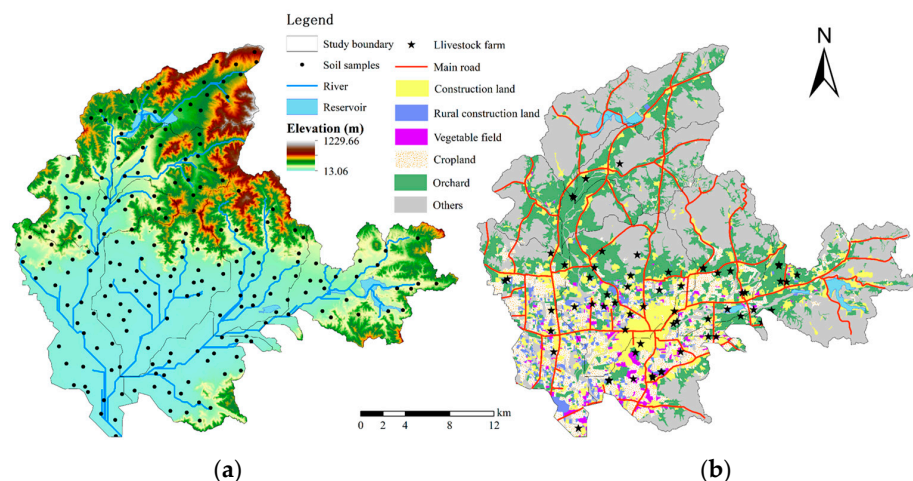


Figure 1. Location of soil sampling sites (a) and geo-environmental elements (b) in the study area.

At a depth of 0–25 cm, five subsamples surrounding the designated sampling location were collected from a 10 m × 10 m grid using a stainless-steel spade and were well mixed. After removing the stones and other debris, approximately 1 kg of soil was packaged in plastic bags and taken to the laboratory, then air-dried at ambient temperature and passed through a 100-mesh sieve. Cr contents were measured using graphite furnace atomic absorption spectrometry (PinAAcle 900 T, PerkinElmer, Shelton, CT, USA) after acid digestion (HCl, HNO₃, HF, and HClO₄).

2.2. Calculation of Potential Predictors

In addition to soil and terrain factors (e.g., pH, texture) that have commonly been used to predict the spatial distribution of heavy metals in previous studies [18], this study also incorporated auxiliary environmental variables in the development of the LUR model. Considering soil properties and auxiliary variables, we employed 84 unique covariates as the potential predictors (Table 1), including physicochemical characteristics of the soil (X1~X18), the shortest distance to a road/river/livestock farm (X19~X21), and the areas of land use types, road/river length, and population density within the circular buffer zones (X22~X84).

Table 1. Descriptions of the potential predictors.

Categories	NO.	Description	Factors (Units)
General categories	X1~X5	Soil physical index at the sample site	Texture, bulk density (g/cm ³), elevation (m), slope (°), aspect (°).
	X6~18	Soil chemical index at the sample site	pH, CEC (mol/kg), SOM (g/kg), TN (mg/kg), AP (mg/kg), AK (mg/kg), AB (mg/kg), AS (mg/kg), Fe (mg/kg), Mn (mg/kg), Zn (mg/kg), Cu (mg/kg), Mo (mg/kg).
	X19~X21	Shortest distance to river, road, and livestock farm	near_river (m), near_road (m), near_lf (m).
Specific categories	X22~X28	Vegetable field area in buffer	V20, V50, V100, V200, V350, V500, V1000 (m ²).
	X29~X35	Cropland area in buffer	C20, C50, C100, C200, C350, C500, C1000 (m ²).
	X36~X42	Construction land cropland area in buffer	JS20, JS50, JS100, JS200, JS350, JS500, JS1000 (m ²).
	X43~X49	Rural construction land cropland area in buffer	RJS20, RJS50, RJS100, RJS200, RJS350, RJS500, RJS1000 (m ²).
	X50~X56	Waterbody area in buffer	W20, W50, W100, W200, W350, W500, W1000 (m ²).
	X57~X63	Orchard area in buffer	O20, O50, O100, O200, O350, O500, O1000 (m ²).
	X64~X70	Road length in buffer	R20, R50, R100, R200, R350, R500, R1000 (m).
	X71~X77	River length in buffer	RI20, RI50, RI100, RI 200, RI350, RI500, RI1000 (m).
	X78~X84	Population density in buffer	P20, P50, P100, P200, P350, P500, P1000 (per km ²).

All land use maps, river and road network datasets, and data for the independent livestock farms in the study area were obtained from the National Science and Technology Infrastructure of China, National Earth System Science Data Sharing Infrastructure (<http://www.geodata.cn/>, accessed on 20 April 2020). Soil type maps and soil property were acquired from the regional Digital Soil System. Population grid data for 2015 with a 1 km resolution were provided by the Resource and Environmental Science Data Center of the Chinese Academy of Sciences (<http://www.resdc.cn/>, accessed on 20 April 2020). The shortest distance to the nearest livestock farm/river /road was accordingly calculated. These auxiliary variables were obtained via GIS based on the basic geographic data. The acreage of each land use type, the total road/river length, and population density were also calculated based on specific circular buffer zones; the buffer radius was 20, 50, 100, 200, 350, 500 and 1000 m.

2.3. Spatial Modeling Methods

2.3.1. LUR Model

The first step of the modeling was to screen the factors that exhibited significant correlations with soil Cr concentration, and then remove the partial variables that may have had collinearity ($r > 0.60$), to ensure the interpretability of the model parameters. The relationship between Cr content and environmental variables was analyzed using the nonparametric Spearman correlation test [20]. Secondly, based on the selection of variables via correlation analysis, a multivariate linear regression analysis with backward selections was performed to identify the most significant variables and build the LUR model [21]. Descriptive, relationship, and regression analyses were all conducted using SPSS. 22.0 software (SPSS Institute). The LUR model can be expressed as Equation (1):

$$y_{LUR} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \varepsilon = y_p + \varepsilon \quad (1)$$

where y_{LUR} is the Cr content in topsoil, $x_1, x_2, \dots$ and x_m refer to the relevant influence factors, β_0 is the intercept (constant value), $\beta_1, \beta_2, \dots$ and β_m are the regression coefficients, and ε is the residual (accidental error).

Thirdly, based on the assumption of independent residuals in linear regression, a standard diagnostic test was conducted to assess the spatial autocorrelation of the residuals using global Moran's I in ArcGIS 10.5 [22]. The two previous steps were iteratively repeated until the final LUR models encompassed only significant variables and satisfied all regression assumptions, including normally distributed residuals with low spatial autocorrelation.

2.3.2. Ordinary Kriging Combined with LUR Model (OK_LUR)

OK_LUR can consider the auxiliary variables at specific points for interpolating the outputs. This approach is grounded in the principle that the deterministic component of the target variable is explained through the regression methods. The predicted value is formed by the LUR model prediction and the OK prediction of the residual at a given point. The process of OK_LUR can be summarized as Equation (2):

$$y_i = \beta_i + \sum_{j=1} x_{ij} \beta_j + \varepsilon_i = y_{pi} + \varepsilon_i \quad (2)$$

where y_i represents the predicted value of the Cr content in topsoil at the i th location according to OK_LUR, β_i is the intercept, x_{ij} is the j th predictors' value of the i th observation, β_j is the j th predictors' coefficient, ε_i is the residual of the regression via semi-variogram and OK, and y_{pi} is the predicted value of the Cr contents at the i th location according to OK_LUR without accidental error. Geostatistical analysis and OK interpolation were conducted with ArcGIS 10.5 (ESRI, Redlands, CA, USA).

2.3.3. Geographical Weighted Regression Kriging Combined with LUR Model (GWRK_LUR)

In this phase, we constructed the GWRK_LUR model based on predictors identified in the LUR model. Specifically, this model represents a local LUR model based on the GWR method. While the predictors remained static, the variable coefficients instead of the constant coefficient varied over the study domain. The local coefficient estimates were obtained through weighting neighboring observations using a distance-adjusted kernel function [23]. Therein, the GWR tool in ArcGIS 10.5 was used, with the kernel type and bandwidth method parameters configured as ‘adaptive’ and ‘AICc’. Then, the OK method was used to interpolate residuals from the GWR. The GWRK_LUR model is specified according to Equation (3):

$$y_i^* = \beta_i^* + \sum_{j=1} x_{ij} \beta_{j(u_i, v_i)} + \varepsilon_i^* \quad (3)$$

where y_i^* represents the predicted value of the Cr topsoil content at the i th location as given by GWRK_LUR, β_i^* is the spatially varying intercept at the i th observation, x_{ij} is the j th predictors’ value of the i th observation, (u_i, v_i) is the location in geographical space of the i th observation, $\beta_{j(u_i, v_i)}$ is the j th predictors’ spatially varying coefficient at the i th location, and ε_i^* is the residual of the GWR according to OK.

2.4. Model Validations

The sensitivity of the models was examined via k-fold cross-validation (k-fold CV). Sampling sites were randomly divided into k parts, with $k - 1$ parts serving as the training set for model fitting, and the remaining part used for testing. The above procedure was repeated k times ($k = 10$) [24,25]. Then, the coefficient of determination (R^2), mean bias error (MBE), root mean squared error (RMSE), and normalized root mean squared error (NRMSE) were calculated to describe the accuracy of the models [26]. Generally, a higher CV- R^2 value, the coefficient of determination based on the k-fold CV, indicated better predictive ability of the Cr models; a positive value of MBE indicated that the observed value overestimated the simulated value and vice versa; lower RMSE and NRMSE values (closer to zero) meant more stable and accurate models.

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (4)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (5)$$

$$MBE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i) \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (7)$$

$$NRMSE = \frac{100}{\bar{y}_i} \times \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (8)$$

where y_i is the observed value of Cr contents at the i th location, $\hat{y}_i$ is the predicted value, $\bar{y}_i$ is the mean of the observed values, and N is the number of paired observed simulated values.

3. Results and Discussion

3.1. Descriptive Statistics for Topsoil Cr Concentrations

Cr concentrations in 208 soil samples ranged from 50.10 to 202.74 mg/kg with a standard deviation of 19.38, which was noted to be below the world average [27]. The mean concentration (68.59 mg/kg) was found to be lower than the upper threshold (150

mg/kg) according to the soil environmental quality standards of both China and the USA. Furthermore, none of the sampled data exceeded the limits (800 mg/kg) that could generate risk to the soil environment.

Moreover, Cr concentrations within the study area were not much different from soil background values, which were 50.60–163.00 mg/kg with a mean concentration of 68.10 mg/kg [28]. Agricultural activities in the past 30 years had not significantly impacted Cr concentrations in the topsoil within the study area.

The topsoil Cr concentrations displayed medium spatial variability with 28.25% coefficient of variations, indicating that Cr concentration was likely to have been influenced by natural factors such as soil genesis and parent material [9]. The measurements of all sampling points were highly skewed to the right (skewness = 3.97); therefore, they required transformation using the natural logarithm (Figure A1 in Appendix A). In addition, the topsoil Cr concentrations presented significant spatial autocorrelation with positive global Moran's I (0.15) and $p < 0.01$.

3.2. Evaluation of Environmental Factors Based on the Correlation between Predictors and Cr Concentrations

Based on the results of the Spearman correlation test, correlations between factors and Cr contents showed that the Cu factor had the largest correlation with Cr, followed by Zn, Mo, pH, population density within 1000 m (P1000), an orchard within 100 m (O100), cropland within 350 m (C350), and the shortest distance to a livestock farm (near_lf). Significant positive correlations between factors (Cu, near_lf, Zn, Mo, O100, C350) and Cr concentrations (Figure 2) indicated these factors could potentially impact the physicochemical process of Cr in soil.

The pH values of 208 soil samples ranged from 4.30 to 8.21 with a standard deviation of 0.76. The average value was 7.22, and most of the values were concentrated between 6.80 to 8.20. Soil pH can affect heavy metal leachability and bioavailability; it was utilized for a Cr-dependent criterion. The observed negative correlation between pH and Cr in our study was consistent with the results in paddy soil [29].

The soil organic matter (SOM) content of the sampling points ranged from 3.25 to 112.34 g/kg, with an average of 21.89 g/kg. The majority of these samples' content was mainly concentrated between 10 and 36 g/kg. Previous research has established that SOM significantly affects the distribution of Cr, primarily via complexation processes, redox reactions, and modulating the mobility of Cr [30]. However, in our study, no significant correlation was observed between SOM and Cr. This lack of correlation may be attributed to the interference of other factors at a regional scale level.

Based on the correlations between 84 factors (Table 1) and Cr concentrations, we determined 8 factors (Cu, near_lf, Zn, Mo, pH, O100, C350, P1000) that demonstrated significant correlation as potential predictors of Cr LUR models. The selected predictors within the buffer zones (O100, C350, and P1000) exhibited significant spatial heterogeneity (Figure 3) and effectively depicted the differences across the entire study area [31]. These predictors provided more information compared with commonly utilized variables in regression models.

3.3. Comparison of the LUR, OK_LUR and GWRK_LUR Models

In this study, the OK method was employed to interpolate the residuals derived from the LUR and GWR models. Based on the Kolmogorov–Smirnov (K-S) test, through natural logarithm transformation, it was found that the residuals of both the OK_LUR and GWRK_LUR models followed normal distribution. The Gaussian model offered the most optimal fit for the semivariogram (Table A1) of the OK_LUR residuals, while the exponential model proved the best fit for the GWRK_LUR residuals.

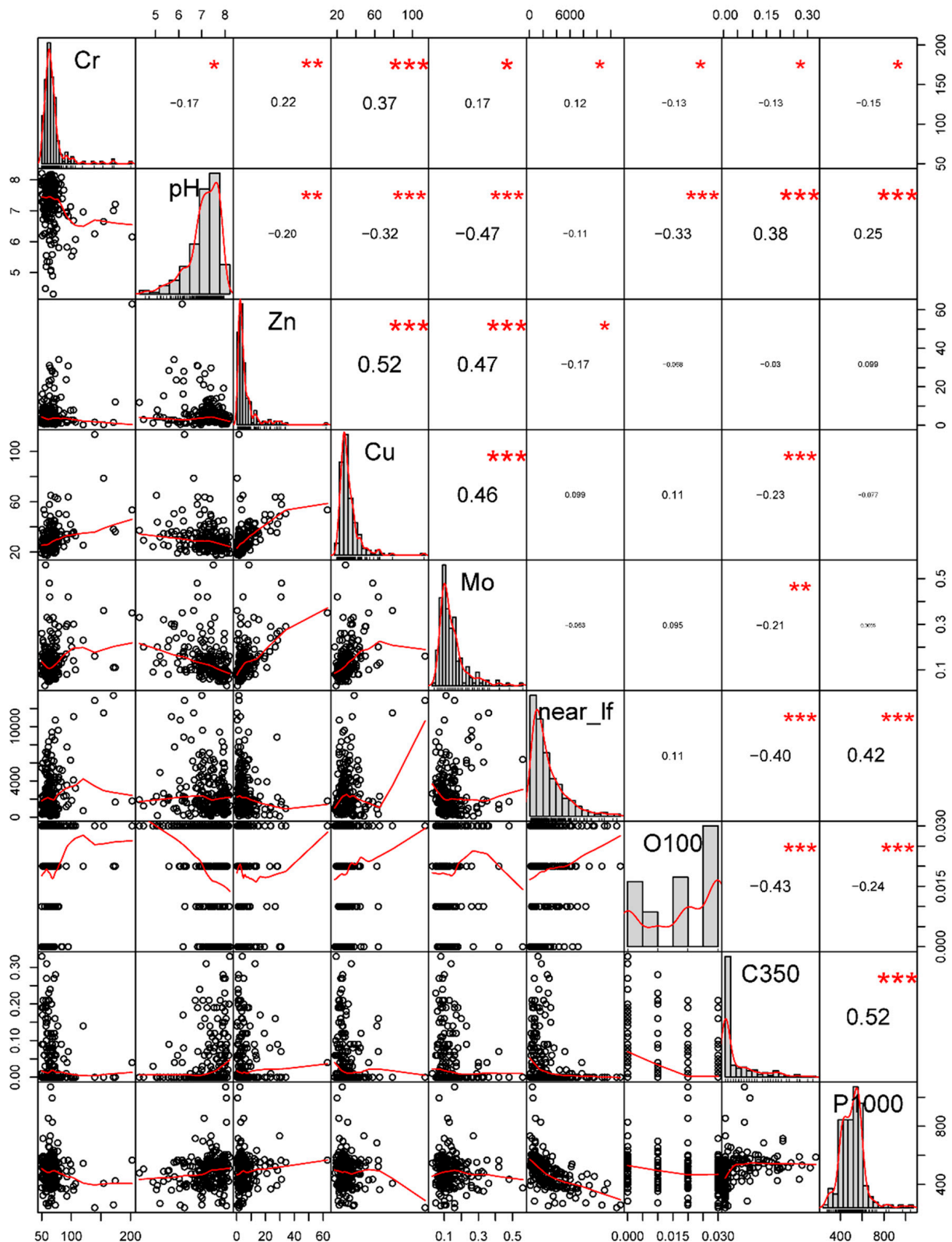


Figure 2. Correlations between predictors and Cr concentrations. Only the significant factors are presented in this figure. * p -value (i.e., significant level) less than 0.05; ** p -value below 0.01; *** p -value 0.00. Cr, Zn, Cu, and Mo are the contents (mg/kg) of Cr, Zn, Cu, and Mo in soil; near_lf is the shortest distance (m) between sampling site and a livestock farm; O100 is an orchard area (m²) with a circular buffer size of 100 m; C350 is the cropland area (m²) within 350 m; P1000 is population density (per km²) within 1000 m.

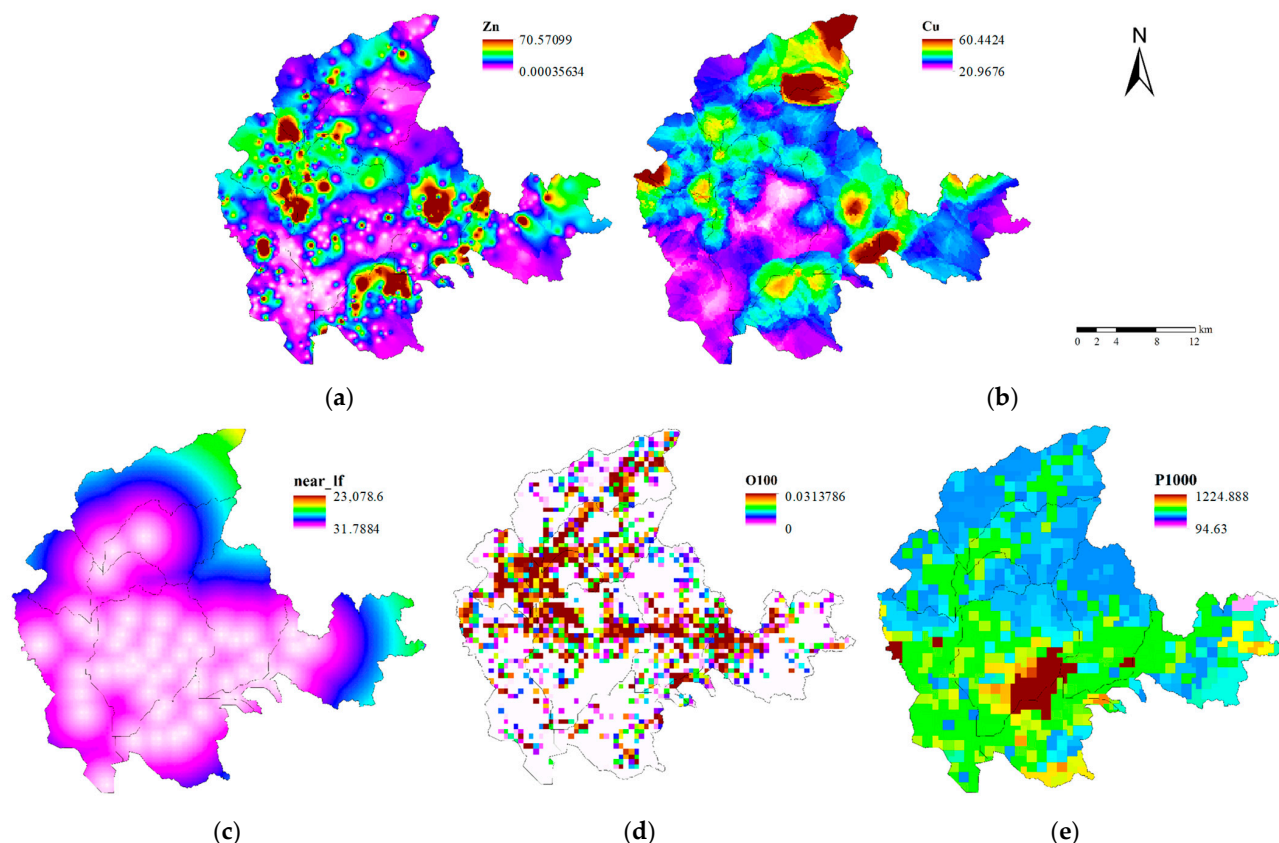


Figure 3. The distribution of predictors across the study area. (a) It is the contents (mg/kg) of Zn in the soil; (b) It is the contents (mg/kg) of Cu in the soil; (c) It is the shortest distance (m) between the sampling site and a livestock farm; (d) It is an orchard area (m^2) with a circular buffer size of 100 m; (e) It is population density (per km^2) within 1000 m. The color ranging from purple to red represents the value from small to large.

Table 2 compares the performances of the LUR, OK_LUR, and GWRK_LUR models in modeling the spatial distribution of Cr. These models considered Zn, Cu, near_lf, O100, and P1000 factors as predictor variables. The results indicated that the LUR, OK_LUR, and GWRK_LUR models captured 28%, 37%, and 44% of the variation in Cr, respectively. All models exhibited a certain degree of overestimation of the Cr concentration according to the MBE. In addition, the Moran's I value of the LUR model residuals was 0.03, which indicated that the LUR model did adequately reduce the spatial autocorrelation in the Cr data.

The OK_LUR and GWRK_LUR models improved the LUR model's explanation by 9% and 16%, respectively. This phenomenon was attributed to the information contained in the LUR model residuals [32]. The GWRK_LUR model showed better performance than OK_LUR. It benefited from the capacity to capture the spatial non-stationarity in the relationships through allowing coefficients to change in space [31]. In contrast, the OK_LUR model was unable to adequately describe local relationships. The coefficients of factors and Cr in the GWRK_LUR model further supported this observation (Table 2). The GWRK_LUR model not only had the highest R^2 , but also obtained optimal MBE, RMSE, and NRMSE compared with the LUR and OK_LUR models. However, the highest local R^2 (0.58) of the DWRK_LUR model appeared in the north area, and local R^2 in the central region was consistently lower than 0.3 (Figure A2). It can be concluded that the quality of the GWRK_LUR model performance was non-uniform in space. In particular regions, the predictive capability of land-use factors may have been overestimated when using global regression (LUR and OK_LUR models).

Table 2. Descriptive statistics for the Cr LUR, OK_LUR, and GWRK_LUR models.

Models	Factors	Coefficients	Coefficients with Z-Score Predictors	CV-R ²	MAE	MBE	RMSE	NRMSE
LUR	Intercept	34.676	68.591	0.28	0.15	0.48	16.41	23.91
	Zn	0.445	3.317					
	Cu	0.557	6.118					
	Near_lf	0.002	5.851					
	O100	108.637	1.289					
OK_LUR	P1000	0.012	1.369	0.37	0.04	1.48	15.70	22.89
	Intercept	34.676	68.591					
	Zn	0.445	3.317					
	Cu	0.557	6.118					
	Near_lf	0.002	5.851					
GWRK_LUR	O100	108.637	1.289	0.44	0.05	0.31	14.56	21.22
	P1000	0.012	1.369					
	Intercept	14.727–64.062	66.266–70.246					
	Zn	−0.617–1.562	−4.595–11.633					
	Cu	−0.358–0.860	−3.937–9.450					
	Near_lf	−0.001–0.004	−1.666–8.959					
	O100	−11.036–197.613	−0.131–2.344					
	P1000	−0.018–0.041	−2.002–4.605					

The coefficient between Cu and Cr was the highest ($r = 6.118$) according to the regression of the normalized parameters with the Z-score. This finding is consistent with reports that Cu and Cr have a major common origin in agricultural soil [33]. The Cr models incorporated only a single land-use factor (O100), consistent with previous findings that only a limited number of land-use factors could be included in the final LUR models [16].

The accuracy of the geostatistical models was greatly dependent on the quality of sampling [5]. Our models were based on 208 observations, exceeding the typical 40–80 samples recommended by Hoek and coworkers [22]. Although more observations could obtain influencing factors that might be neglected through modeling with fewer points, we also sacrificed the prediction accuracy presented by the model (R^2). Additionally, the barely satisfactory R^2 of these models also proved that the distribution of Cr in agricultural soil was significantly influenced by the soil parent material [33].

3.4. Spatial Distribution of Predicted Cr Concentrations

The ranges of the Cr concentrations' predicted values in agricultural topsoil with the DWRK_LUR and OK_LUR models were 46.36–162.67 mg/kg (Figure 4) and 48.35–113.99 mg/kg (Figure A3), respectively. The DWRK_LUR model yielded predictions that were closer to the real observations (50.10–202.74 mg/kg), indicating its superior capability in reflecting the spatial structure of the original regionalized variables. A Kriged surface is well known to be smoother than the regionalized variable predicts. Their strong smoothing and centralizing effects can lead to predicted values converging towards a narrow range [5].

The spatial distribution mapping results showed the continuous variety of characteristics of topsoil Cr across the area (Figure 4). Lower Cr concentrations were observed in the southern region, while elevated Cr concentrations emerged in the west, north, and northeast (Figure 4a). The Cr distribution exhibited a steeper gradient in the central region, coinciding with the distribution of Zn, Cu, O100, and P1000 (Figure 3). Higher Cr concentrations were predominantly found in the mountainous regions. This was chiefly because the soil from these regions was less sandy and retained the original soil state with fewer anthropogenic effects, including smaller populations and predominantly traditional patterns of agriculture [34]. The weathering and mineralization processes in the mountains provided a high quantity of Cr to the soil [33,35], while anthropogenic effects in the

mountainous regions may have contributed to a relatively lower transformation/transfer of Cr compared with more developed regions [19]. This result indicated the interactions of different sources of artificial and natural impacts on the distribution of Cr, aligning with previous research conducted in the Yellow River Delta [8]. Soil parent materials may have affected the 64.5% Cr concentration in soil, although was not considered for quantitative analysis in this study and should be improved in the future.

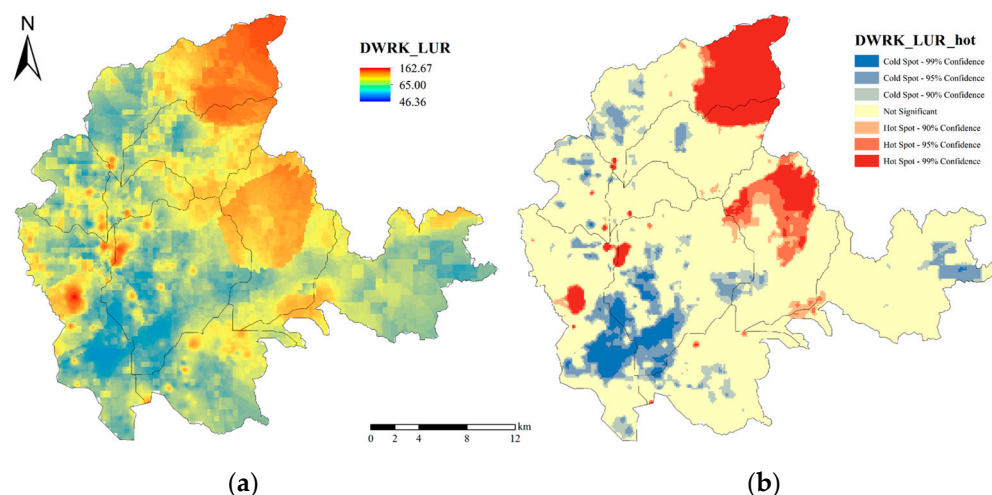


Figure 4. Spatial distribution (a) and cold-hot-spot pattern (b) of Cr content predicted via by land use regression combined with geographically weighted regression Kriging (DWRK_LUR).

The predicted Cr distribution exhibited a cold-hot-spot pattern, with high values clustered in the north area and low values clustered in the south (Figure 4b). Some areas had no significant cold-hot spots, indicating a lack of continuous high or low values. These localized variations in Cr distribution were attributed to a range of factors, including terrain, crop types, irrigation methods, and the distribution of livestock farms [7]. The Cr pattern was very similar to the Cu distribution (Figure 4b), and this result agreed with previous research on Chinese arable soil [36].

The modeling results showed the Cr contents were below the threshold (800 mg/kg) that could generate risk to the soil environment. They suggested that agricultural soil in the study area is not contaminated with Cr. While Cr from natural sources tends to have low bioavailability in its highly stable form [35], agricultural inputs represent other primary sources that could potentially cause risk to soil. Therefore, it is imperative to develop management practices and policies aimed at reducing pollutant inputs, to ensure the sustainable utilization of land resources and protect farmland soil from heavy metal contamination.

3.5. Effect of Predictors on Cr Concentrations Based on Coefficients

A significant positive coefficient between Zn and Cr was found in the southwest (Figure 5b), while the influence of Cu on Cr was negative in this region (Figure 5c). The Zn and Cu factors exhibited similar positively correlation with Cr content in their high-value region and were negatively correlated in their low-value region (Figure 3a,b). This observation in the high-value region suggests a potential common source of Zn, Cu, and Cr [33]. Agricultural practices using water irrigation, animal manure, and chemical fertilizers are contributory factors to the increased presence of not only Cu and Zn but also potentially Cr [36,37]. The observation in the low-value region indicated that the environment was conducive to Cr enrichment in topsoil. In the central area, the influence of Zn and Cu on Cr content was non-significant (Figure A4a,b), which could be attributed to their complex spatial heterogeneity (Figure 3) [6].

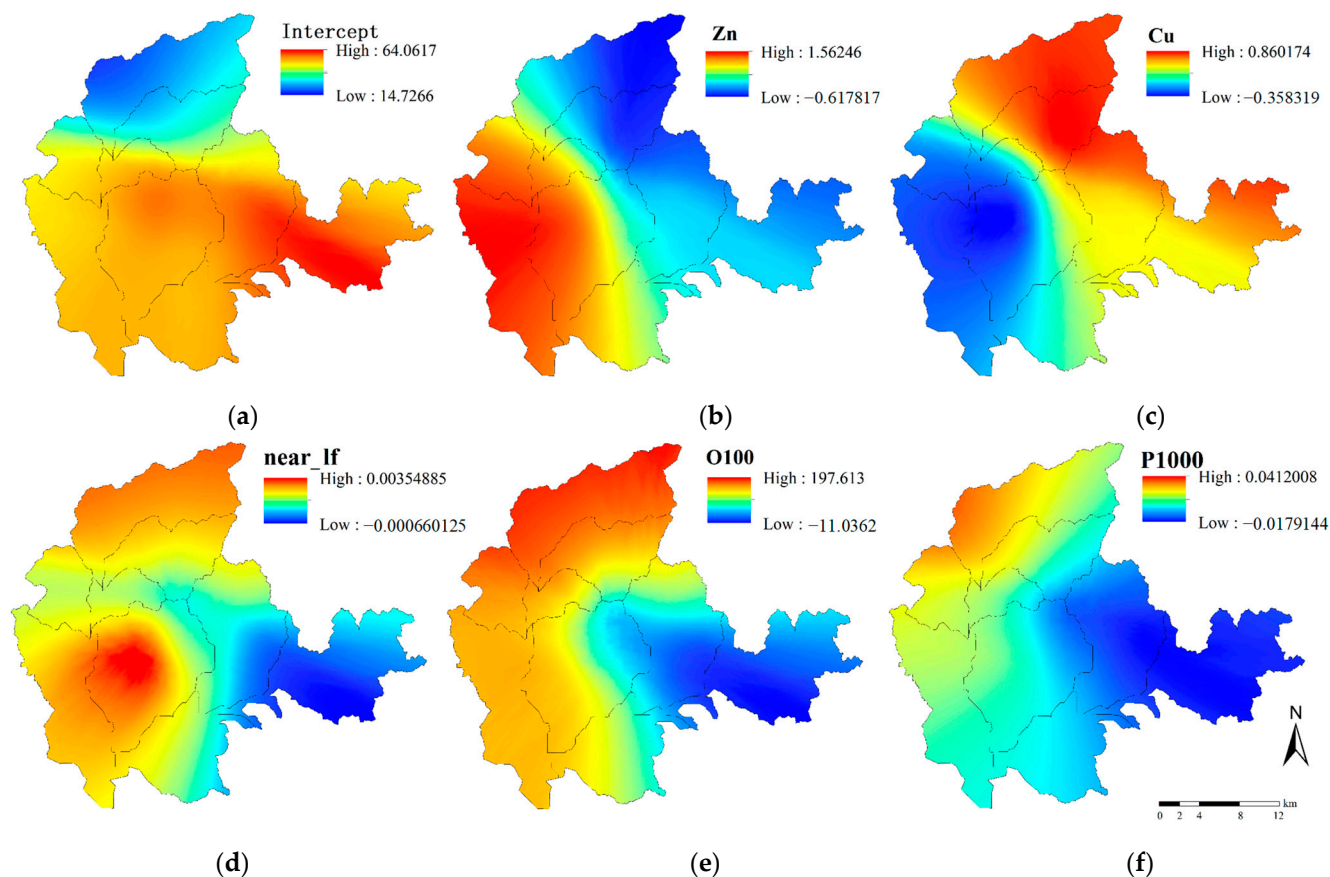


Figure 5. Effects of predictors on Cr across the study area. (a) It is the intercept of the GWRK_LUR model; (b) It is the contents (mg/kg) of Zn in the soil; (c) It is the contents (mg/kg) of Cu in the soil; (d) It is the shortest distance (m) between sampling sites and a livestock farm; (e) O100 is orchard area (m²) with a circular buffer size of 100 m; (f) P1000 is population density (per km²) within 1000 m. The color ranging from blue to red represents the value from small to large.

The shortest distance to a livestock farm (near_lf), orchard area within 100 m (O100), and population density within 1000 m (P1000) exhibited notably positive effects on Cr in the western region, whereas they displayed more negative effects in the eastern area (Figure 5d–f). The near_lf factor significantly influenced soil Cr in most areas (Figure A4), and the largest effect was observed in the southwest (Figure 5). This might be the dominant livestock farm existing in this region (Figure 3). The O100 factor had no statistically significant effect on Cr in a small area of the eastern region (Figure A4). The impact of P1000 factor on Cr varied spatially with irregular coefficients. As an ecological conservation development area, the eastern region has less agricultural land and fewer livestock farms (Figure 3). Less environmental impact could possibly explain the non-significant correlation between predictors and Cr concentration in this area [7,13].

The effect of predictors on Cr displayed a trend from low to high in the southeast (Figure 5). This weak prediction effect was attributed to the fact that this area is the administrative center of the study area and has the largest population density (Figure 3e), indicating that agricultural land was not the primary land-use type [6]. This phenomenon, namely the non-significant effect on Cr in the southeastern region, also caused a lower local R^2 in this area.

Overall, the GWRK_LUR model provided superior performance compared with the other models. The geographically varying coefficients of these factors provided evidence of the spatial non-stationarity in the relationship between the predictors and Cr contents [31]. Furthermore, the spatially discrepant effect of the predictors on Cr was also presented.

4. Conclusions

A hybrid mapping model integrating LUR and geostatistical methods was developed to predict the spatial distribution of topsoil Cr at a regional scale. The proposed GWRK_LUR model significantly improved the accuracy of prediction achieved with the OK_LUR and LUR models by 9% and 16%, respectively. These factors driving the Cr content variation in agricultural topsoil included the contents of Zn and Cu in soil, the distance between the soil sampling site and the nearest livestock farm, an orchard area within 100 m, and population density within 1000 m. These factors exhibited varying effects on Cr with spatial non-stationarity coefficients across the study area. The proposed method in this study not only enhances topsoil Cr spatial prediction and environmental risk assessment, but also contributes to the advancement of LUR combined with geostatistical methods for other soil pollutants and regions.

Author Contributions: Sample collection, software, and writing—original draft, M.C.; Investigation, writing—review and editing, D.W.; Writing—review and editing, Y.Q.; Validation and sample analysis, R.Y.; Writing—review and editing, A.D.; Methodology, writing—review and editing, funding acquisition, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number U20A20115.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Author Meng Cao was employed by the company China International Engineering Consulting Corporation. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

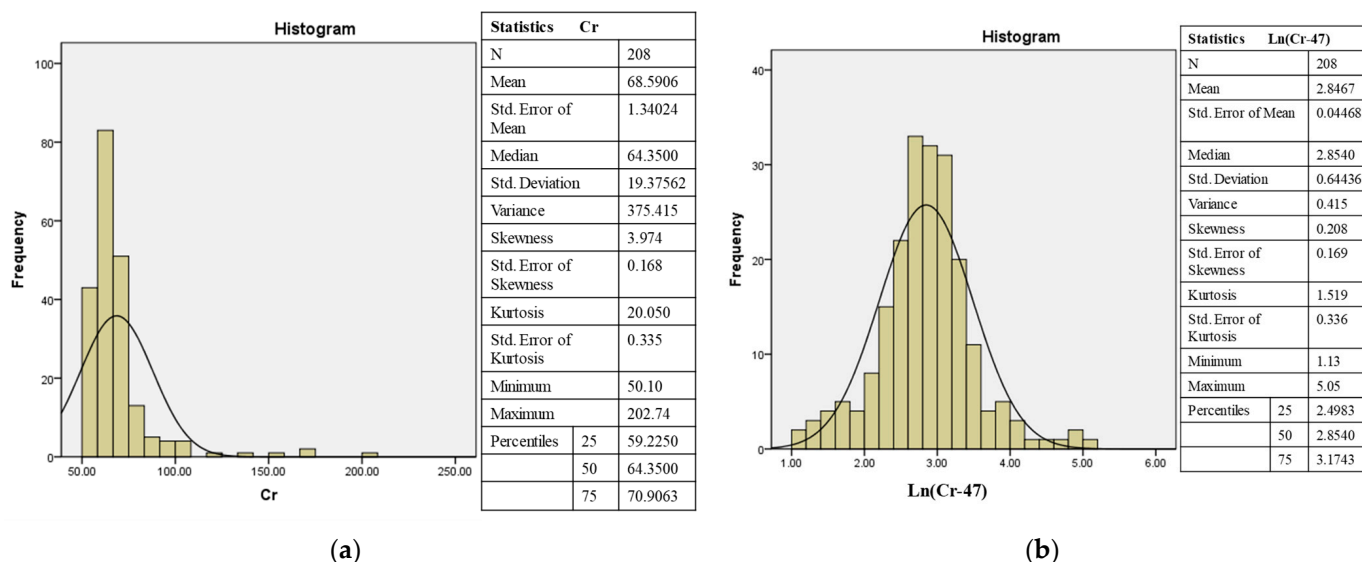


Figure A1. Data (a) and transformation information (b).

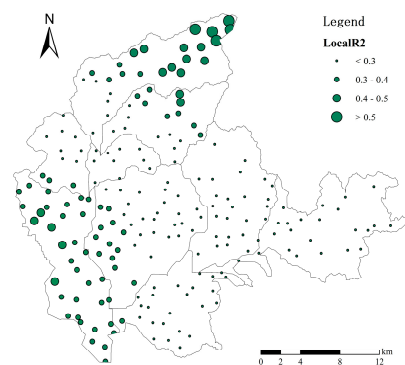


Figure A2. The distribution of local R2 for the Cr DWRK_LUR model.

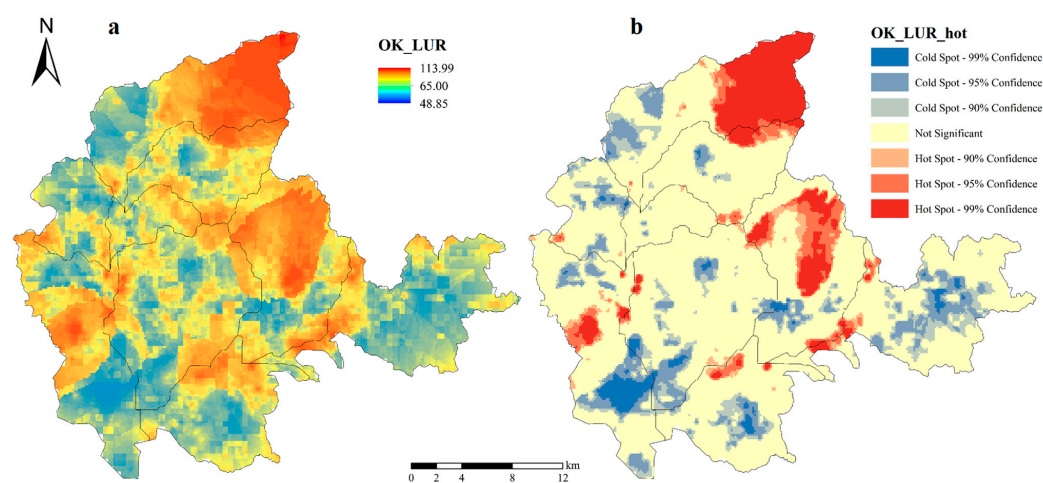


Figure A3. Spatial distribution (a) and cold-hot-spot pattern (b) of Cr concentrations predicted through ordinary Kriging combined with the LUR model (OK_LUR).

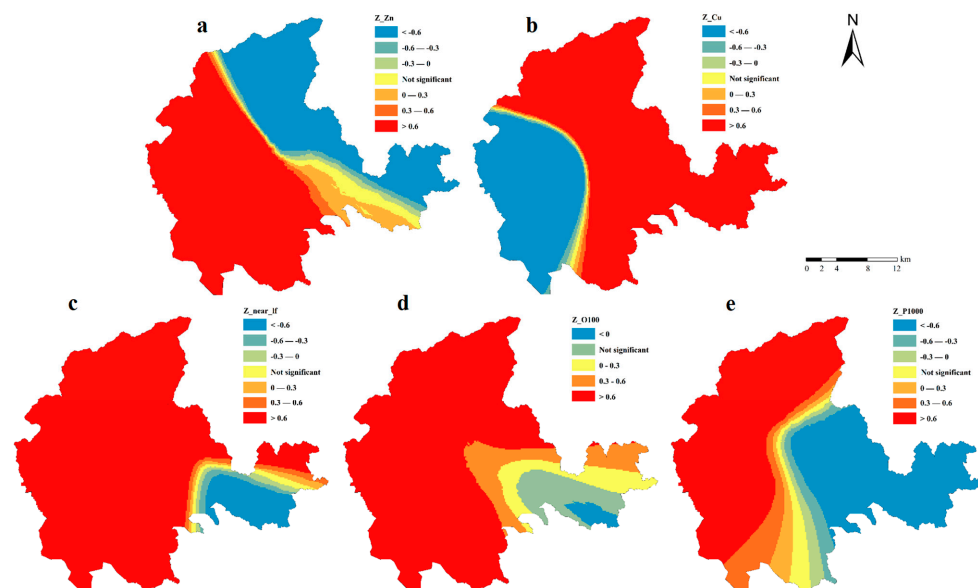


Figure A4. Effects of predictors with Z-score normalized on Cr across the study area. (a) It is the Z-score normalized of Zn in the soil; (b) It is the Z-score normalized of Cu in the soil; (c) It is Z-score normalized of the shortest distance between sampling site and a livestock farm; (d) It is the Z-score normalized of orchard area with a circular buffer size of 100 m; (e) It is the Z-score normalized of population density within 1000 m. Different colors in the images represent different ranges of numerical values.

Table A1. The results of geostatistical analysis of the residuals.

Item	Model	Range (m)	Nugget (C0)	Partial Sill (C1)	C0/(C0 + C1) (%)
Ln(Residuals _{OK_LUR})	Gaussian	2102	0.014	0.142	9.21%
Ln(Residuals _{GWK_LUR})	Exponential	2290	0.011	0.117	8.60%

References

- Ballabio, C.; Panagos, P.; Lugato, E.; Huang, J.-H.; Orgiazzi, A.; Jones, A.; Fernández-Ugalde, O.; Borrelli, P.; Montanarella, L. Copper distribution in European topsoils: An assessment based on LUCAS soil survey. *Sci. Total Environ.* **2018**, *636*, 282–298. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yang, W.; Chen, Y.; Yang, L.; Xu, M.; Jing, H.; Wu, P.; Wang, P. Spatial distribution, food chain translocation, human health risks, and environmental thresholds of heavy metals in a maize cultivation field in the heart of China's karst region. *J. Soils Sediments* **2022**, *22*, 2654–2670. [\[CrossRef\]](#)
- Song, Y.; Li, H.; Li, J.; Mao, C.; Ji, J.; Yuan, X.; Li, T.; Ayoko, G.A.; Frost, R.L.; Feng, Y. Multivariate linear regression model for source apportionment and health risk assessment of heavy metals from different environmental media. *Ecotoxicol. Environ. Saf.* **2018**, *165*, 555–563. [\[CrossRef\]](#)
- Ao, M.; Chen, X.; Deng, T.; Sun, S.; Tang, Y.; Morel, J.L.; Qiu, R.; Wang, S. Chromium biogeochemical behaviour in soil-plant systems and remediation strategies: A critical review. *J. Hazard. Mater.* **2022**, *424*, 127233. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lado, L.R.; Hengl, T.; Reuter, H.I. Heavy metals in European soils: A geostatistical analysis of the FOREGS Geochemical database. *Geoderma* **2008**, *148*, 189–199. [\[CrossRef\]](#)
- Hou, D.; O'Connor, D.; Nathanail, P.; Tian, L.; Ma, Y. Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review. *Environ. Pollut.* **2017**, *231*, 1188–1200. [\[CrossRef\]](#)
- Micó, C.; Recatalá, L.; Peris, M.; Sánchez, J. Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis. *Chemosphere* **2006**, *65*, 863–872. [\[CrossRef\]](#)
- Gan, Y.; Huang, X.; Li, S.; Liu, N.; Li, Y.C.; Freidenreich, A.; Wang, W.; Wang, R.; Dai, J. Source quantification and potential risk of mercury, cadmium, arsenic, lead, and chromium in farmland soils of Yellow River Delta. *J. Clean. Prod.* **2019**, *221*, 98–107. [\[CrossRef\]](#)
- Tume, P.; Roca, N.; Rubio, R.; King, R.; Bech, J. An assessment of the potentially hazardous element contamination in urban soils of Arica, Chile. *J. Geochem. Explor.* **2018**, *184*, 345–357. [\[CrossRef\]](#)
- Huang, J.; Wu, Y.; Sun, J.; Li, X.; Geng, X.; Zhao, M.; Sun, T.; Fan, Z. Health risk assessment of heavy metal(loid)s in park soils of the largest megacity in China by using Monte Carlo simulation coupled with Positive matrix factorization model. *J. Hazard. Mater.* **2021**, *415*, 125629. [\[CrossRef\]](#)
- Jiang, X.; Zou, B.; Feng, H.; Tang, J.; Tu, Y.; Zhao, X. Spatial distribution mapping of Hg contamination in subclass agricultural soils using GIS enhanced multiple linear regression. *J. Geochem. Explor.* **2019**, *196*, 1–7. [\[CrossRef\]](#)
- Xu, X.; Zhao, Y.; Zhao, X.; Wang, Y.; Deng, W. Sources of heavy metal pollution in agricultural soils of a rapidly industrializing area in the Yangtze Delta of China. *Ecotoxicol. Environ. Saf.* **2014**, *108*, 161–167. [\[CrossRef\]](#) [\[PubMed\]](#)
- Shi, T.; He, L.; Wang, R.; Li, Z.; Hu, Z.; Wu, G. Digital mapping of heavy metals in urban soils: A review and research challenges. *Catena* **2023**, *228*, 107183. [\[CrossRef\]](#)
- Sun, Y.; Lei, S.; Zhao, Y.; Wei, C.; Yang, X.; Han, X.; Li, Y.; Xia, J.; Cai, Z. Spatial distribution prediction of soil heavy metals based on sparse sampling and multi-source environmental data. *J. Hazard. Mater.* **2024**, *465*, 133114. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, J.; Han, C.; Peng, Y.; Wang, M.; Zhao, Y. Improved three-dimensional mapping of soil chromium pollution with sparse borehole data: Incorporating multisource auxiliary data into IDW-based interpolation. *Soil Use Manag.* **2023**, *39*, 933–947. [\[CrossRef\]](#)
- Hoek, G.; Eeftens, M.; Beelen, R.; Fischer, P.; Brunekreef, B.; Boersma, K.F.; Veeffkind, P. Satellite NO₂ data improve national land use regression models for ambient NO₂ in a small densely populated country. *Atmos. Environ.* **2015**, *105*, 173–180. [\[CrossRef\]](#)
- Nickel, S.; Hertel, A.; Pesch, R.; Schröder, W.; Steinnes, E.; Uggerud, H.T. Modelling and mapping spatio-temporal trends of heavy metal accumulation in moss and natural surface soil monitored 1990–2010 throughout Norway by multivariate generalized linear models and geostatistics. *Atmos. Environ.* **2014**, *99*, 85–93. [\[CrossRef\]](#)
- Cao, S.; Lu, A.; Wang, J.; Huo, L. Modeling and mapping of cadmium in soils based on qualitative and quantitative auxiliary variables in a cadmium contaminated area. *Sci. Total Environ.* **2017**, *580*, 430–439. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wu, S.-S.; Yang, H.; Guo, F.; Han, R.-M. Spatial patterns and origins of heavy metals in Sheyang River catchment in Jiangsu, China based on geographically weighted regression. *Sci. Total Environ.* **2017**, *580*, 1518–1529. [\[CrossRef\]](#)
- Henderson, S.B.; Beckerman, B.; Jerrett, M.; Brauer, M. Application of Land Use Regression to Estimate Long-Term Concentrations of Traffic-Related Nitrogen Oxides and Fine Particulate Matter. *Environ. Sci. Technol.* **2007**, *41*, 2422–2428. [\[CrossRef\]](#)
- Deschenes, S.; Setton, E.; Demers, P.A.; Keller, P.C. Modelling Arsenic and Lead Surface Soil Concentrations using Land Use Regression. *E3S Web Conf.* **2013**, *1*, 08004. [\[CrossRef\]](#)
- Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561–7578. [\[CrossRef\]](#)

23. Stewart Fotheringham, A.; Charlton, M.; Brunsdon, C. The geography of parameter space: An investigation of spatial non-stationarity. *Int. J. Geogr. Inf. Syst.* **1996**, *10*, 605–627. [[CrossRef](#)]
24. Wang, M.; Beelen, R.; Eeftens, M.; Meliefste, K.; Hoek, G.; Brunekreef, B. Systematic Evaluation of Land Use Regression Models for NO₂. *Environ. Sci. Technol.* **2012**, *46*, 4481–4489. [[CrossRef](#)] [[PubMed](#)]
25. Wolf, K.; Cyrus, J.; Harciníková, T.; Gu, J.; Kusch, T.; Hampel, R.; Schneider, A.; Peters, A. Land use regression modeling of ultrafine particles, ozone, nitrogen oxides and markers of particulate matter pollution in Augsburg, Germany. *Sci. Total Environ.* **2017**, *579*, 1531–1540. [[CrossRef](#)] [[PubMed](#)]
26. Lee, M.; Brauer, M.; Wong, P.; Tang, R.; Tsui, T.H.; Choi, C.; Cheng, W.; Lai, P.-C.; Tian, L.; Thach, T.-Q.; et al. Land use regression modelling of air pollution in high density high rise cities: A case study in Hong Kong. *Sci. Total Environ.* **2017**, *592*, 306–315. [[CrossRef](#)] [[PubMed](#)]
27. Tóth, G.; Hermann, T.; Szatmári, G.; Pásztor, L. Maps of heavy metals in the soils of the European Union and proposed priority areas for detailed assessment. *Sci. Total Environ.* **2016**, *565*, 1054–1062. [[CrossRef](#)]
28. Chen, J.; Wei, F.; Zheng, C.; Wu, Y.; Adriano, D.C. Background concentrations of elements in soils of China. *Water Air Soil Pollut.* **1991**, *57*, 699–712. [[CrossRef](#)]
29. Zeng, F.; Ali, S.; Zhang, H.; Ouyang, Y.; Qiu, B.; Wu, F.; Zhang, G. The influence of pH and organic matter content in paddy soil on heavy metal availability and their uptake by rice plants. *Environ. Pollut.* **2011**, *159*, 84–91. [[CrossRef](#)]
30. Jing, H.; Yang, W.; Chen, Y.; Yang, L.; Zhou, H.; Yang, Y.; Zhao, Z.; Wu, P.; Zia-ur-Rehman, M. Exploring the mechanism of Cd uptake and translocation in rice: Future perspectives of rice safety. *Sci. Total Environ.* **2023**, *897*, 165369. [[CrossRef](#)]
31. Song, W.; Jia, H.; Li, Z.; Tang, D.; Wang, C. Detecting urban land-use configuration effects on NO₂ and NO variations using geographically weighted land use regression. *Atmos. Environ.* **2019**, *197*, 166–176. [[CrossRef](#)]
32. Pásztor, L.; Szabó, K.Z.; Szatmári, G.; Laborcz, A.; Horváth, Á. Mapping geogenic radon potential by regression kriging. *Sci. Total Environ.* **2016**, *544*, 883–891. [[CrossRef](#)] [[PubMed](#)]
33. Alloway, B.J. Sources of Heavy Metals and Metalloids in Soils. In *Heavy Metals in Soils: Trace Metals and Metalloids in Soils and Their Bioavailability*; Alloway, B.J., Ed.; Springer: Dordrecht, The Netherlands, 2013; pp. 11–50.
34. Davis, H.T.; Marjorie Aelion, C.; McDermott, S.; Lawson, A.B. Identifying natural and anthropogenic sources of metals in urban and rural soils using GIS-based data, PCA, and spatial interpolation. *Environ. Pollut.* **2009**, *157*, 2378–2385. [[CrossRef](#)]
35. Jin, Y.; O'Connor, D.; Ok, Y.S.; Tsang, D.C.W.; Liu, A.; Hou, D. Assessment of sources of heavy metals in soil and dust at children's playgrounds in Beijing using GIS and multivariate statistical analysis. *Environ. Int.* **2019**, *124*, 320–328. [[CrossRef](#)] [[PubMed](#)]
36. Niu, L.; Yang, F.; Xu, C.; Yang, H.; Liu, W. Status of metal accumulation in farmland soils across China: From distribution to risk assessment. *Environ. Pollut.* **2013**, *176*, 55–62. [[CrossRef](#)] [[PubMed](#)]
37. Chen, T.B.; Wong, J.W.C.; Zhou, H.Y.; Wong, M.H. Assessment of trace metal distribution and contamination in surface soils of Hong Kong. *Environ. Pollut.* **1997**, *96*, 61–68. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.