Received: 26 October 2020 | Accepted: 23 October 2020 | Published online: 6 April 2021

DOI: 10.1002/saj2.20189

Soil Science Society of America Journal

NUTRIENT MANAGEMENT & SOIL & PLANT ANALYSIS

Improving the spatial prediction accuracy of soil alkaline hydrolyzable nitrogen using GWPCA-GWRK

Jian Chen[1,2] | **Mingkai Qu**[1,2] | **Jianlin Zhang**[1] | **Enze Xie**[1,2] | **Yongcun Zhao**[1,2] | **Biao Huang**[1,2]

[1] Key Laboratory of Soil Environment and Pollution Remediation, Institute of Soil Science, Chinese Academy of Sciences, East Beijing Road 71, Nanjing 210008, China

[2] University of Chinese Academy of Sciences, Yuquan Road 19, Beijing 100049, China

Correspondence
Mingkai Qu, Key Laboratory of Soil Environment and Pollution Remediation, Institute of Soil Science, Chinese Academy of Sciences, East Beijing Road 71, Nanjing 210008, China.
Email: qumingkai@issas.ac.cn

Assigned to Associate Editor Carl Bolster.

Funding information
National Natural Science Foundation of China, Grant/Award Number: 41771249; National Key Research and Development Program of China, Grant/Award Number: 2018YFC1800104; Institute of Soil Science, Chinese Academy of Sciences, Grant/Award Number: ISSASIP1623; Youth Innovation Promotion Association, CAS, Grant/Award Number: 2018348

Abstract

Principal component analysis-multiple linear regression (PCA-MLR) is usually used to weaken the multi-collinearity effects among auxiliary variables in a regression prediction. However, both PCA and MLR in this model are only built on variable space rather than geographical space. When used in the spatial prediction of soil properties, PCA-MLR usually cannot effectively capture the spatially non-stationary structures among auxiliary variables and spatially non-stationary relationships between the target variable and principal component scores. Moreover, PCA-MLR may ignore the potentially valuable regression residual. To address these limitations, this study first proposed geographically weighted principal component analysis-geographically weighted regression kriging (GWPCA-GWRK) for the spatial prediction of soil alkaline hydrolyzable nitrogen (AN) in Shayang County, China. Then, the spatial prediction accuracy of GWPCA-GWRK was compared with those of the following five models: ordinary kriging (OK), co-kriging (CoK), PCA-MLR, PCA-graphically weighted regression (PCA-GWR), and GWPCA-GWR. Results showed that (i) eight variables were determined as auxiliary data by a geodetector; (ii) the spatially non-stationary relationships among the eight auxiliary variables were revealed by the results of the local correlation analysis, Monte Carlo test, and GWPCA; (iii) GWPCA-GWRK provided the lowest prediction error (RMSE = 18.80 mg kg^{-1}, MAE = 12.79 mg kg^{-1}) and highest Lin's concordance correlation coefficient (LCCC; 0.75); (iv) relative improvement accuracies over the traditionally-used OK were 19.74% for GWPCA-GWRK, 16.42% for GWPCA-GWR, 8.09% for

Abbreviations: ACu, available copper; AFe, available iron; AMn, available manganese; AN, alkaline hydrolyzable nitrogen; AS, available sulfur; ASi, available silicon; AZn, available zinc; CEC, cation exchange capacity; CV, coefficient of variation; GWPCA, geographically weighted principal component analysis; GWR, geographically weighted regression; LCCC, Lin's concordance correlation coefficient; MAE, mean absolute error; MLR, multiple linear regression; PCA, principal component analysis; RI, relative improvement accuracy; RMSE, root mean square error; SD, standard deviation; SOM, soil organic matter; TK, total potassium; TN, total nitrogen; TP, total phosphorus.

PCA-GWR, −3.67% for PCA-MLR, and 4.70% for CoK. It is concluded that the proposed GWPCA-GWRK model is an effective spatial predictor, which can adequately extract the main information of the multiple auxiliary variables in a large-scale area.

1 | INTRODUCTION

In the agricultural ecosystem, soil alkaline hydrolyzable nitrogen (AN) is one of the most important indicators of soil quality, and its content is the key to precision fertilization (Huang et al., 2007; Zhu, Wen, & Freney, 1997). Excessive fertilization will lead to the accumulation of N in the soil, and surplus soil N may migrate with surface runoff into the waterbodies, thereby negatively affecting aquatic ecosystems (Howarth & Marino, 2006; Wang, Zhang, & Huang, 2009). Therefore, an accurate spatial distribution pattern of soil AN is important for implementation of precision fertilization and environmental protection.

When multiple environmental factors are closely related to the target variable, these environmental factors usually can be used as auxiliary information for spatial prediction in the models, such as multiple linear regression (MLR), regression kriging (RK) (e.g., Hengl, Heuvelink, & Stein, 2004; Piccini, Marchetti, & Francaviglia, 2014), CoK (e.g., Knotters, Brus, & Voshaar, 1995), GWR (e.g., Qu, Wang, Huang, & Zhao, 2018; Song et al., 2016; Wang, Zhang, & Li, 2013), and GWRK (e.g., Kumar, Lal, & Liu, 2012a; Yang et al., 2019). For CoK, combining multiple auxiliary variables usually cannot further improve the spatial prediction accuracy because of the influence of the redundant information between auxiliary variables, cumulative fitting errors of variograms, and so on (Goovaerts, 1997). The regression models are usually adopted to incorporate multiple auxiliary variables to enhance the spatial prediction of the target variable. When there are multi-collinearity effects among the auxiliary variables, PCA-MLR, also known as principal components regression, is often used (e.g., Guo et al., 2018; Zhai et al., 2018). Principle component analysis (PCA) can transform the raw auxiliary variables into several sets of orthogonal principal component (PC) scores, thereby weakening the information redundancy among the raw auxiliary variables (Philippi, 1993). However, in the actual large-scale field environment, the relationships among the auxiliary variables generally vary with spatial location because of the effects of natural (e.g., terrain, soil types, and parent material) and human factors (e.g., land management) (Schleus, Wu, & Blume, 1998; Zhao et al., 2010). Traditionally-used PCA, only built on variable space, usually cannot effectively extract the information about spatially non-stationary relationships among the multiple auxiliary variables, and may further affect the spatial prediction accuracy of PCA-MLR.

Geographically weighted principal component analysis (GWPCA) is a spatial form of PCA, where the local variance-covariance matrix of the auxiliary variables is calculated based on the multivariate auxiliary dataset in the vicinity of each calibration location (Fotheringham, Brunsdon, & Charlton, 2002; Harris, Brunsdon, & Charlton, 2011). Therefore, the geographically-weighted PC (GWPC) scores, as part of the outputs of GWPCA, contain the structured information about the local relationships among the raw auxiliary variables at each location. The GWPCA was first used by Lloyd (2010) to investigate the spatial structures among the population in Northern Ireland. However, to the best of the authors' knowledge, GWPCA has rarely been used for improving the spatial prediction accuracy of soil properties. Given the spatially non-stationary relationships among the multiple auxiliary variables, GWPCA may be a more effective tool than PCA for extracting the spatially structured information among the multiple auxiliary variables prior to the regression prediction.

Traditionally used MLR, also only built on variable space, cannot effectively capture the spatially non-stationary relationships between the target variable and independent variables in geographical settings. GWR is a spatial local regression technique, and its local regression coefficients are calculated based on the multivariate auxiliary dataset in the vicinity of the calibration locations (Fotheringham et al., 2002). Therefore, compared with MLR, GWR may be more effective in revealing the spatially non-stationary relationships. Since it is impossible to incorporate all factors related to the target variable into the regression prediction, there may be some valuable information in the regression residual. In this case, GWRK, which is the sum of the GWR-predicted value and the kriging-predicted residual, may be a valuable tool (Kumar et al., 2012a; Yang et al., 2019).

Moreover, the selection of high-quality auxiliary variables is critical to the spatial prediction. Geodetector, a spatial statistical model, can test the spatial correlation between the target variable and potential auxiliary variable without linear correlation assumption (Wang et al., 2010). The implicit assumption in geodetector is the following: if the potential auxiliary variable X is associated with target variable Y, then Y would show a similar spatial distribution to X. Moreover, geodetector has no normal distribution requirements for data (Wang et al., 2010). Therefore, geodetector provides an effective tool for the selection of auxiliary variables.

In this study, the specific objectives were to (i) reveal the spatially non-stationary relationships among the auxiliary variables; (ii) propose GWPCA-GWRK for the spatial prediction of soil AN; and (iii) compare the spatial prediction accuracy of GWPCA-GWRK with those of the following five models: OK, CoK, PCA-MLR, PCA-GWR, and GWPCA-GWR. The goal is to recommend a high-precision spatial prediction model that can effectively incorporate multiple auxiliary variables with spatially varying relationships.

Core Ideas

- Spatially varying relationships among the auxiliary data were revealed by GWPCA.
- GWPCA-GWRK was proposed to utilize auxiliary data fully in the spatial prediction.
- GWPCA-GWRK obtained higher prediction accuracy than OK, CoK, PCA-MLR, PCA-GWR, and GWPCA-GWR.

2 | MATERIALS AND METHODS

2.1 | Study area and data sources

The study area is Shayang County (30°23′–30°55′N, 112°02′–112°42′E), a typical agricultural region in Central China, with an area of 2044 km^2 (Figure 1). The terrain of this county is mainly dominated by plains, and only parts of the north are low hills. Due to long-term paddy cultivation, the dominant soil type in Shayang County is Hydragric Anthrosols (World Reference Base Soil Taxonomy) (IUSS Working Group WRB, 2014), mainly located in the central and western part of this county.

Six hundred and fifty-eight topsoil samples (0–20 cm) were collected in this county in October and November 2007 (Figure 1). Each sample was comprised of five subsamples within an area of approximately 0.1 ha surrounding a specific sampling location, and the diameter of each soil core was about 10 cm. All sampling sites were geo-located using a hand-held global position system receiver. Soil samples were air-dried, ground, and sieved for chemical analysis.

Soil AN was determined by the method of Lu (2000). Briefly, 1.0 mol L^{-1} NaOH was used to hydrolyze the soil, and H_3BO_3 was used to absorb the distilled NH_3. The AN concentration was determined by back titration of H_3BO_3 with dilute H_2SO_4. In addition to soil AN, twelve other soil properties that may be potentially related to soil AN are also analyzed. Soil organic matter (SOM) was measured using the Walkley-Black method (Lu, 2000). Soil total nitrogen (TN) was determined using the Kjeldahl method (Lu, 2000). Soil pH was determined in a ratio of 1:2.5 soil/water suspension with a pH meter (pHs-3C). Cation exchangeable capacity (CEC) was measured by titrating distilled ammonium in soils in which cations were exchanged with ammonium acetate (Lu, 2000). For soil total potassium (TK) and phosphorus (TP), the soil samples were digested using an acid mixture of HF-$HClO_4$-HCl, then TK and TP in solution were determined using the flame photometry and colorimetric method, respectively (Lu, 2000). Soil available copper, zinc, iron, and manganese (ACu, AZn, AFe, and AMn) were first extracted with diethylenetriaminepentaacetic acid (DTPA) (Lindsay & Norvell, 1978), and then their concentrations were determined with flame atomic absorption spectrometry. Soil available silicon (ASi) was extracted with acid sodium acetate (Lu, 2000). Soil available sulfur (AS) was extracted with phosphate (Lu, 2000).

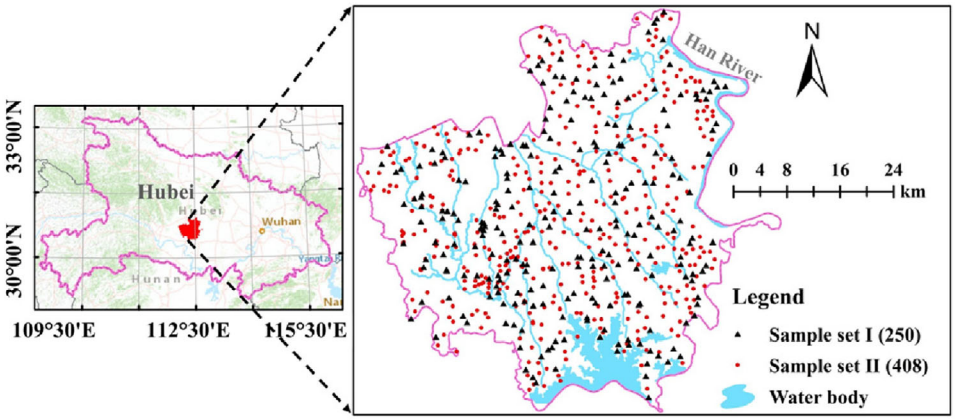

FIGURE 1 Study area and sample sites in Shayang County, Hubei Province, China

Note. Soil alkaline hydrolyzable nitrogen (AN) in the Sample Set I was used for model calibration; soil AN in the Sample Set II was used for model validation

2.2 | Modeling process and data utilization

Firstly, geodetector was used to determine the auxiliary variables for spatial prediction of soil AN. Secondly, GWPCA was applied on the total sample points ($n = 658$) to obtain the corresponding GWPC scores, then the scores on the soil sample points were interpolated into the surfaces of GWPC scores. Thirdly, GWPCA-GWRK was constructed to predict the spatial distribution of soil AN based on soil AN in the Sample Set I ($n = 250$) and the surfaces of GWPC scores. Lastly, the spatial prediction accuracy of GWPCA-GWRK was compared with those of OK, CoK, PCA-MLR, PCA-GWR, and GWPCA-GWR based on 408 pairs of the predicted and measured soil AN concentration in the sample set II.

2.3 | Geodetector

Geodetector, a spatial variance analysis method, has been widely applied to test the spatial correlation between target variable (i.e., soil AN) and potential auxiliary variables (i.e., 12 other soil properties in this study) (Wang, Zhang, & Fu, 2016). The degree of spatial correlation can be measured by the q-statistic, which is related to the ratio of the weighted sum of local variance (weighted by the number of samples in each strata) to the global variance (Wang et al., 2010):

$$q_x = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2}, \tag{1}$$

where L is the number of strata of a potential auxiliary variable x; N is the number of soil samples in the entire study area; N_h is the number of soil samples in strata h; σ^2 is the variance of soil AN in the entire region; σ_h^2 is the variance of soil AN in strata h. If the spatial distribution of the potential auxiliary variable x is completely consistent with that of soil AN, local variance is 0 and $q_x = 1$; If the spatial distribution of a potential auxiliary variable x is completely irrelevant to that of soil AN, local variance is 1 and $q_x = 0$. In general, the larger the q value, the more similar the spatial distribution of soil AN and that of the impact factors is.

2.4 | Geographically weighted principal component analysis-geographically weighted regression kriging (GWPCA-GWRK)

GWPCA-GWRK is the hybrid model of GWPCA and GWRK. First, GWPCA was conducted to explore the spatially varying structures among the multiple auxiliary variables through the percentage of total variation (PTV) and winning variables, and further transform the multiple auxiliary variables into the orthogonal GWPC scores. Then, the GWPC scores were interpolated into auxiliary surfaces for spatial prediction. This model mainly involves the following steps:

(i) to calculate the geographically weighted (GW) variance-covariance matrix. GWPCA involves regarding a vector of auxiliary variables x_i as conditional on its position (u_i, v_i) and conceptualizing the mean vector and variance-covariance matrix as position functions. That is, $\boldsymbol{\mu}(u_i, v_i)$ and $\boldsymbol{\Sigma}(u_i, v_i)$ are the GW mean vector and the GW variance-covariance matrix, respectively. Geographically weighted variance-covariance matrix is expressed as (Fotheringham et al., 2002):

$$\boldsymbol{\Sigma}\left(u_i, v_i\right) = \mathbf{X}^{\mathrm{T}}\mathbf{W}\left(u_i, v_i\right)\mathbf{X}, \tag{2}$$

where $\mathbf{X}$ is the $n \times m$ matrix of auxiliary variables, with n the number of sampling sites within the bandwidth and m the number of the auxiliary variables; $\mathbf{W}(u_i, v_i)$ is the diagonal matrix of the spatial weighted matrix generated by the bi-square weight function with the adaptive bandwidth.

(ii) to obtain the GW scores matrix. The variance-covariance matrix was decomposed to obtain GW eigenvalues and GW eigenvectors. The GWPC can be written as (Fotheringham et al., 2002):

$$\mathbf{L}\left(u_i, v_i\right)\mathbf{V}\left(u_i, v_i\right)\mathbf{L}\left(u_i, v_i\right)^{\mathrm{T}} = \boldsymbol{\Sigma}\left(u_i, v_i\right), \tag{3}$$

where $\mathbf{L}(u_i, v_i)$ and $\mathbf{V}(u_i, v_i)$ represent the local eigenvectors and the diagonal matrix of local eigenvalues, respectively. Then, the GWPC scores matrix $\mathbf{S}(u_i, v_i)$ can be calculated based on the local eigenvector matrix as follows:

$$\mathbf{S}\left(u_i, v_i\right) = \mathbf{X}\mathbf{L}\left(u_i, v_i\right), \tag{4}$$

where the ith row of GWPC scores were obtained by the product of the ith row of the auxiliary variable's matrix with the local eigenvectors for the ith location. At each target location for GWPCA with m variables, there are m components, m eigenvalues, m sets of loadings (each of size $m \times m$), and m sets of GWPC scores (each of size $658 \times m$ in this study). For the m sets of GWPC scores, the scores fully corresponding to their location were extracted as the independent variables in GWPCA-GWR and GWPCA-GWRK.

(iii) to run GWR with soil AN as the dependent variable and the GWPC scores as the independent variables. GWPCA-GWR is expressed as:

$$y_{\mathrm{GWPCA-GWR}}\left(u_i, v_i\right) = \beta_0\left(u_i, v_i\right) + \sum_{k=1}^{p} \left[\beta_k\left(u_i, v_i\right) \times \mathrm{GWPCS}_k\left(u_i, v_i\right)\right], \tag{5}$$

where $y_{GWPCA-GWR}(u_i, v_i)$ and $GWPCS_k(u_i, v_i)$ are the dependent variable and the kth independent variable, respectively, at the location (u_i, v_i); $\beta_0(u_i, v_i)$ and $\beta_k(u_i, v_i)$ are the local intercept and the kth regression coefficient, respectively. In matrix form, the local regression coefficients of GWPCA-GWR are estimated using the weighted least squares model (Fotheringham et al., 2002):

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = \left[(\mathbf{GWPCS})^{\mathrm{T}}\mathbf{W}(u_i, v_i)(\mathbf{GWPCS})\right]^{-1} (\mathbf{GWPCS})^{\mathrm{T}}\mathbf{W}(u_i, v_i)\mathbf{Y}, \quad (6)$$

where **GWPCS** and **Y** represent $m \times (p+1)$ the design matrix of the independent variables (i.e., GWPC scores) and the $(n \times 1)$ design vector of the dependent variable (i.e., soil AN), respectively. The kernel function $W(u_i, v_i)$ used in GWR is the same as that used in GWPCA.

(iv) to combine the OK-interpolated regression residual. The GWPCA-GWRK model is as follows:

$$y^*_{GWPCA-GWRK}(u_i, v_i) = y^*_{GWPCA-GWR}(u_i, v_i) + r^*_{OK}(u_i, v_i), \quad (7)$$

where $y^*_{GWPCA-GWRK}(u_i, v_i)$ and $y^*_{GWPCA-GWR}(u_i, v_i)$ are the soil AN data predicted by GWPCA-GWRK and GWPCA-GWR, respectively, at (u_i, v_i); $r^*_{OK}(u_i, v_i)$ is the regression residual predicted by OK.

For GWPCA-GWRK, GWPCA can be replaced by basic PCA, and GWRK can be replaced by basic GWR or MLR. Details about GWPCA, GWRK, and GWR were described in Harris et al. (2011), Kumar et al. (2012a), and Fotheringham et al. (2002), respectively.

2.5 | Classic geostatistics

As a classic geostatistical model, OK has been widely adopted in the spatial prediction of soil properties, and it was used as a reference model in this study. When the auxiliary variable with a higher sample density is significantly correlated with the target variable, the prediction accuracy of CoK is usually higher than that of OK (Knotters et al., 1995). Therefore, CoK was also adopted to predict the spatial distribution of soil AN in this county. A detailed description of OK and CoK can be found in Goovaerts (1997).

2.6 | Evaluation criteria

To evaluate the spatial prediction accuracies of the six models (i.e., OK, CoK, PCA-MLR, PCA-GWR, GWPCA-GWR, and GWPCA-GWRK), this study calculated LCCC (Lin, 1989), mean absolute error (MAE), and root mean square error (RMSE) based on 408 pairs of the measured and predicted AN concentration (i.e., the sample set II) (Figure 1). The calculation equations above three indices were as follows:

$$\mathrm{MAE} = \frac{1}{408}\sum_{i=1}^{408}\left|z(u_i, v_i) - z^*(u_i, v_i)\right|, \quad (8)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{408}\sum_{i=1}^{408}\left(z(u_i, v_i) - z^*(u_i, v_i)\right)^2}, \quad (9)$$

$$\mathrm{LCCC} = \frac{2r\sigma_{obs}\sigma_{pred}}{\left(\overline{z} - \overline{z^*}\right)^2 + \sigma^2_{obs} + \sigma^2_{pred}}, \quad (10)$$

where $z(u_i, v_i)$ and $z^*(u_i, v_i)$ are the measured and predicted concentration of soil AN, respectively, at validation location (u_i, v_i); $\overline{z}$ and $\overline{z^*}$ are the mean of $z(u_i, v_i)$ and $z^*(u_i, v_i)$, respectively; r is the Pearson correlation coefficient between the measured and predicted concentration of soil AN; σ_{obs} and σ_{pred} are the variances of measured and predicted concentration of soil AN, respectively. Greater LCCC and lower MAE and RMSE mean higher spatial prediction accuracy for the model. The relative improvement (RI) accuracy of a model over OK is calculated using the following equation (Mishra, Lal, Liu, & Van Meirvenne, 2010):

$$\mathrm{RI} = \frac{\mathrm{RMSE}_{OK} - \mathrm{RMSE}_{EM}}{\mathrm{RMSE}_{OK}}, \quad (11)$$

where RMSE_{OK} and RMSE_{EM} are the RMSE for OK and the evaluated models (i.e., CoK, PCA-MLR, PCA-GWR, GWPCA-GWR, or GWPCA-GWRK), respectively.

In this study, all spatial geocomputations were performed on a regular grid with the cell size of 200 m × 200 m. The "geodetector" package in R 3.3.3 software was used for geodetector analysis (Wang et al., 2010); the "GWmodel" package for GW correlation analysis, Monte Carlo randomization test, GWPCA, and GWR (Gollini, Lu, Charlton, Brunsdon, & Harris, 2015; Lu, Harris, Charlton, & Brunsdon, 2014); the "gstat" package for the fitting of variograms (Pebesma, 2004); ArcGIS (version 10.3) for mapping.

3 | RESULTS AND DISCUSSION

3.1 | Descriptive statistics and auxiliary variables selection

The descriptive statistics of soil AN and the potentially related soil properties are summarized in Table 1. The average soil

TABLE 1 Descriptive statistics of soil alkaline hydrolyzable nitrogen (AN) and potential auxiliary variables in Shayang County, China ($n = 658$)

Soil properties	*n*	Minimum	Maximum	Mean	SD[a]	CV[b] (%)
TN (g kg^{-1})	658	0.13	2.35	1.45	0.32	22.07
SOM (g kg^{-1})	658	4.89	40.54	23.18	5.35	23.08
pH	658	5.10	8.10	6.62	0.66	9.97
CEC (cmol kg^{-1})	658	1.08	31.50	12.79	4.18	32.68
TK (g kg^{-1})	658	7.33	35.69	19.93	3.38	16.96
TP (g kg^{-1})	658	0.03	15.80	0.73	0.66	90.41
ACu (mg kg^{-1})	658	0.61	5.89	2.71	0.89	32.84
AFe (mg kg^{-1})	658	6.03	355.77	120.25	89.21	74.19
AZn (mg kg^{-1})	658	0.12	8.51	1.40	0.74	52.86
AMn (mg kg^{-1})	658	5.79	551.20	33.12	25.48	76.93
ASi (mg kg^{-1})	658	0.00	542.30	121.72	81.26	66.76
AS (mg kg^{-1})	658	5.79	551.20	33.10	25.53	77.13
AN (mg kg^{-1})	658	13.00	273.00	135.40	29.66	21.90
AN for validation	408	13.00	272.00	134.81	29.44	21.84
AN for calibration	250	45.00	273.00	136.35	29.98	21.99

[a]SD, standard deviation.
[b]CV, coefficient of variation.

pH was 6.62, indicating topsoil was slightly acidic in this county. The coefficients of variation (CVs) for soil TP, AMn, and AS were about 90%, 76%, and 77%, respectively, all of which showed relatively high variability. The mean of soil AN concentration was 135.40 mg kg^{-1}, significantly higher than that obtained by the second national soil survey in Shayang County in 1984 (i.e., 84.33 mg kg^{-1}) (Jingmen Municipal Office of Soil Survey, 1984). Such results showed that the AN concentration in the topsoil had significantly increased in the past few decades. Meanwhile, there was no obvious difference between the validation set and calibration set in terms of statistical indices (e.g., mean, standard deviation, and CV) for soil AN (Table 1).

The results of geodetector analysis showed that the spatial distribution of SOM ($q = 0.64$) and TN ($q = 0.62$) has the best consistency with that of soil AN. Moreover, soil pH ($q = 0.41$), ASi ($q = 0.34$), AFe ($q = 0.24$), CEC ($q = 0.23$), ACu ($q = 0.22$), and TK ($q = 0.21$) were also significantly related to soil AN. However, the q-statistic of AZn, AS, TP, and AMn were very close to 0, indicating the minimal explanatory power of these four variables for soil AN. Thus, eight auxiliary variables, namely SOM, TN, pH, ASi, AFe, CEC, ACu, and TK, were retained as auxiliary variables for the spatial prediction. Here, SOM was used as the secondary variable in CoK due to the best consistency with the spatial distribution of soil AN.

3.2 | Local correlation analysis and randomization test

The results of the global correlation analysis showed that most of the eight auxiliary variables were highly correlated with each other (Figure 2). However, in geographical settings, the strength of the correlation among the auxiliary variables may vary spatially. In this study, the GW Pearson correlation coefficients (GW r), which is a local form of r (Harris & Brunsdon, 2010), were used to illustrate such non-stationary relationships. For example, the local correlation between soil pH and SOM was strong in the northeast and relatively weak in the southwest of the study area (Figure 2b); the local correlation between soil TN and AFe was strong in the north and weak in the south (Figure 2c); the local correlation between CEC and ASi was stronger in the east of the study area (Figure 2d).

Monte Carlo randomization test was used to evaluate whether the eigenvalues generated by GWPCA varied significantly throughout the study area (Harris, Clarke, Juggins, Brunsdon, & Charlton, 2015). The results showed that the p value of the SD of the local eigenvalues for GWPC1 to GWPC4 were all below 0.05 (Figure 3). Therefore, the null hypothesis of the Monte Carlo test was rejected, which means that the spatial non-stationarity of the relationships among the eight auxiliary variables is statistically significant.

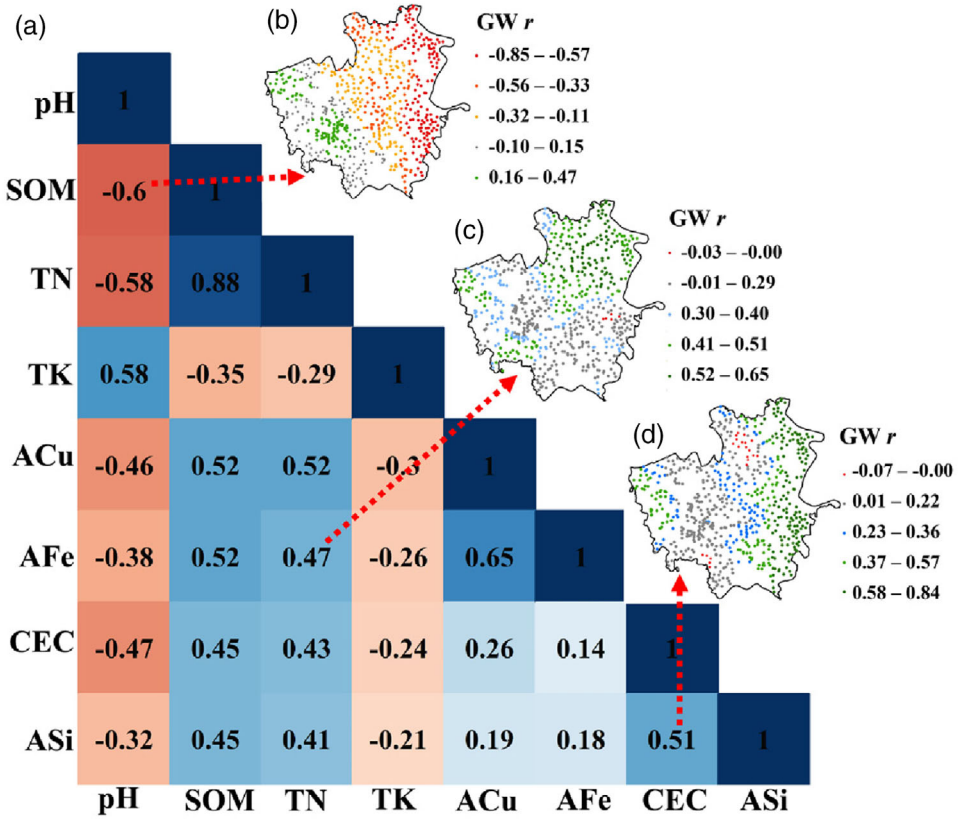

FIGURE 2 Maps of global correlation coefficient matrix (a) and geographically weighted Pearson correlation coefficients (GW *r*) (b, c, and d) among the auxiliary variables

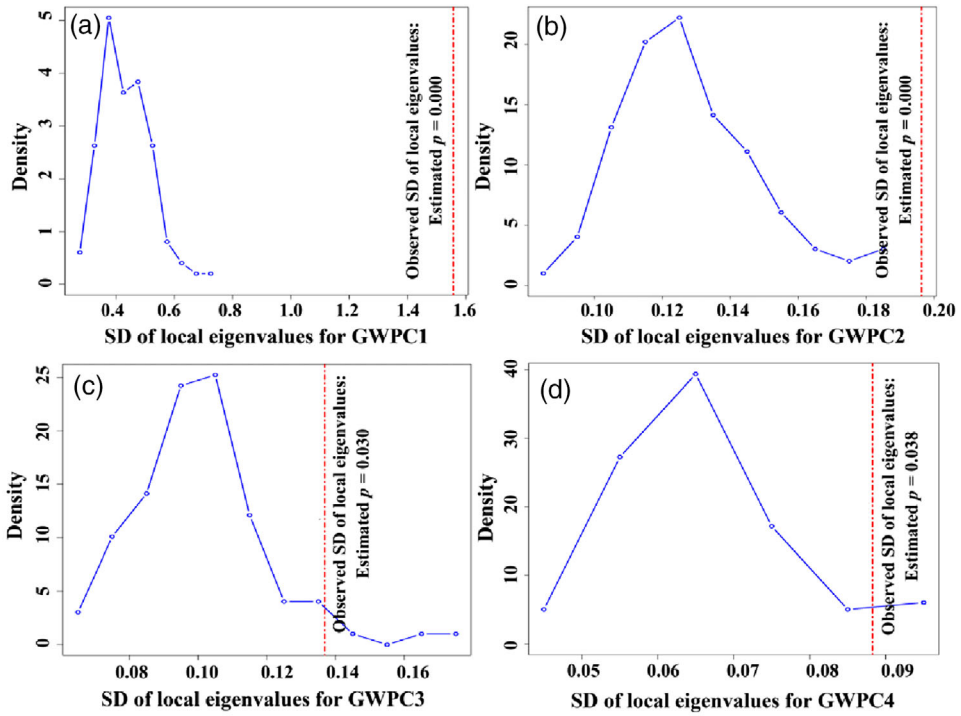

FIGURE 3 Randomization tests (99 times) for local eigenvalue non-stationarity for (a) GWPC1, (b) GWPC2, (c) GWPC3, and (d) GWPC4, respectively. GWPC, geographically weighted principal component

Note. The bandwidth used in the randomization test was the same as that in the GWPCA. The null hypothesis of the randomization test is that the local eigenvalues are stationary, and the alternative hypothesis is that the local eigenvalues are non-stationary

TABLE 2 Principal component analysis for the eight auxiliary variables in Shayang County, China ($n = 658$)

	PC1	PC2	PC3	PC4
Eigenvalue	1.91	1.73	1.60	1.14
PTV[a] (%)	23.90	21.64	20.03	18.43
CPTV[b] (%)	23.90	45.54	65.57	84.00
Loadings				
pH	-0.48	-0.21	-0.24	**0.58**
SOM	**0.84**	0.32	0.28	-0.19
TN	**0.88**	0.29	0.24	-0.14
TK	-0.07	-0.15	-0.10	**0.83**
ACu	0.25	**0.83**	0.12	-0.19
AFe	0.23	**0.88**	0.03	-0.11
CEC	0.25	0.03	**0.79**	-0.12
ASi	0.16	0.11	**0.87**	-0.04

Note. The principal components (PC) with the eigenvalue > 1 were retained, and the absolute loading values > 0.5 were presented in bold type.

[a]PTV, percentage of the total variation.

[b]CPTV, cumulative PTV.

3.3 | Principal component analysis and geographically weighted principal component analysis

The results of PCA showed that the percentages of total variation (PTV) were 23.90% for PC1, 21.64% for PC2, 20.03% for PC3, and 18.43% for PC4 (Table 2). For the PCA loadings, PC1 mainly represented SOM and TN; PC2 mainly represented ACu and AFe; PC3 mainly represented CEC and ASi; PC4 mainly represented pH and TK. Although the traditional PCA decomposed multiple auxiliary variables into four sets of orthogonal PCs, this non-spatial dimension reduction technology might not adequately extract the main information of the auxiliary variables with the spatially varying relationships.

The results of GWPCA showed that the cumulative PTV (CPTV) for the first four GWPCs (i.e., GWPC1–GWPC4) ranged from 78.17% to 88.89%, with high values in the east and low values in the west of the study area (Figure 4). The variables with the highest loadings (i.e., winning variable) for the first four GWPCs were visually displayed in Figure 5. It could be seen that the winning variables varied with the spatial location and presented certain clustering characteristics. For example, TN and SOM, which had the highest loadings in GWPC1, were mainly distributed in the southeast of the study area, and did not dominate the entire research area as global PCA showed (Figure 5a and Table 2). In addition, soil AFe and ACu appear to play an important part in defining the local structures among eight auxiliary variables in the midwest of the study area (Figure 5). The reason may be that the land-use types in the midwest of the study area are mainly dominated by paddy fields with relatively strong acidity and submerged condition, and the availability of soil copper and iron is relatively high. Hence, the relationships between soil AFe (or ACu) and other soil properties in this subarea were different from those in other subareas in Shayang County. Furthermore, the soil samples with pH as the winning variable in GWPC1 were mainly distributed in the northeast of the study area. The reason may be that soil parent materials in this subarea are mainly carbonate-rich river alluvial deposits due to the influence of Han River. Therefore, the soil in this subarea was alkaline, which is significantly different from other areas in this county.

The map of winning variables for GWPC2 (Figure 5b), GWPC3 (Figure 5c), and GWPC4 (Figure 5d) also showed the spatial change in the complex interrelationships of the eight auxiliary variables. It was worth noting that the relationships among the auxiliary variables are more likely to be spatially non-stationary in this county. The possible reasons are as follows: (i) soil properties usually have strong spatial variability in large-scale areas (Granger et al., 2017); (ii) soil properties are affected by multiple endogenous factors such as soil parent material and pedogenic processes, and the above endogenous factors are usually spatially non-stationary in large-scale areas (Qu, Chen, Huang, & Zhao, 2020; Zhao et al., 2010); (iii) soil properties in Shayang County are also strongly affected by exogenous factors such as land-use patterns and fertilization, and the effects of the above exogenous factors on different soil properties also vary from place to place (Schleus et al., 1998; Xie et al., 2019). Therefore, the relationships among the soil properties were not constant in space, and PCA built on variable space could not effectively capture such non-stationary relationships.

3.4 | Geographically weighted regression and residual analysis

The local regression coefficients between soil AN and the first four GWPC scores generated by GWPCA-GWR and GWPCA-GWRK were displayed in Figure 6, suggesting that the relationships between soil AN and GWPC scores were spatially non-stationary. This may be because the GWPC scores were a linear combination of the original auxiliary variables, and the relationships between the original auxiliary variables and soil AN were also spatially non-stationary.

The global Moran's I was further computed to explore the spatial dependence of the residuals generated by PCA-MLR, PCA-GWR, and GWPCA-GWR. The residuals generated by GWPCA-GWR had less spatial dependency (Moran's $I = -0.12$, $p < 0.001$) than that generated by PCA-MLR (Moran's $I = 0.38$, $p < 0.001$) and PCA-GWR (Moran's $I = 0.31$, $p < 0.001$). Such a result indicated that PCA-MLR and PCA-GWR may ignore the crucial information about the relationships among the auxiliary variables. It is worth noting that Moran's I of the regression residual generated by

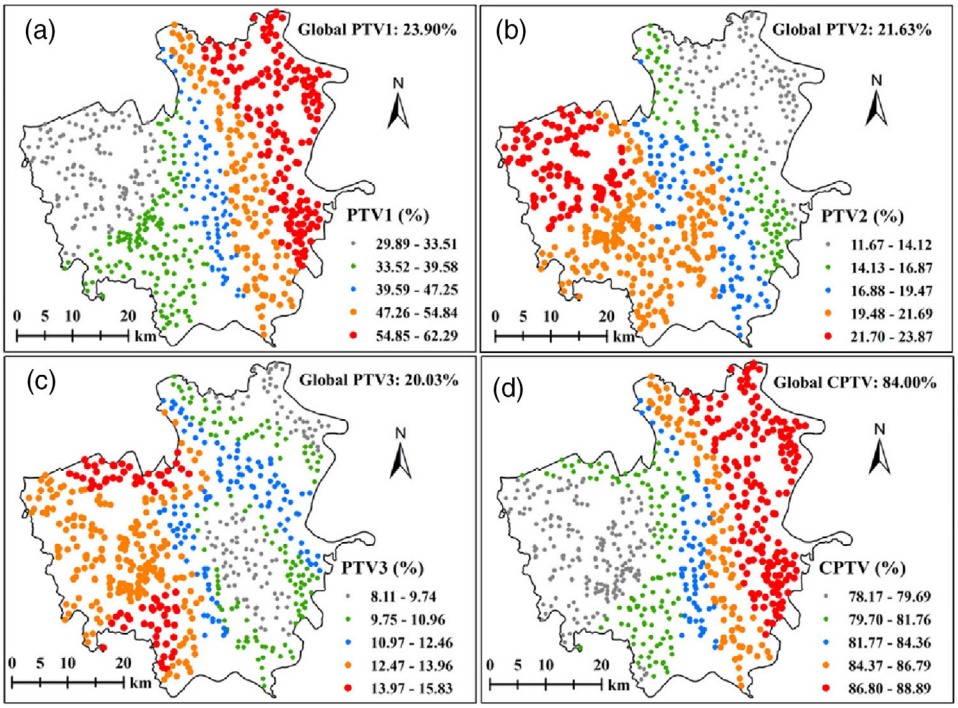

FIGURE 4 Maps of the percentage of the total variance (PTV) for (a) GWPC1, (b) GWPC2, (c) GWPC3, and (d) cumulative PTV (CPTV) for the first four geographically weighted principal component (GWPC)

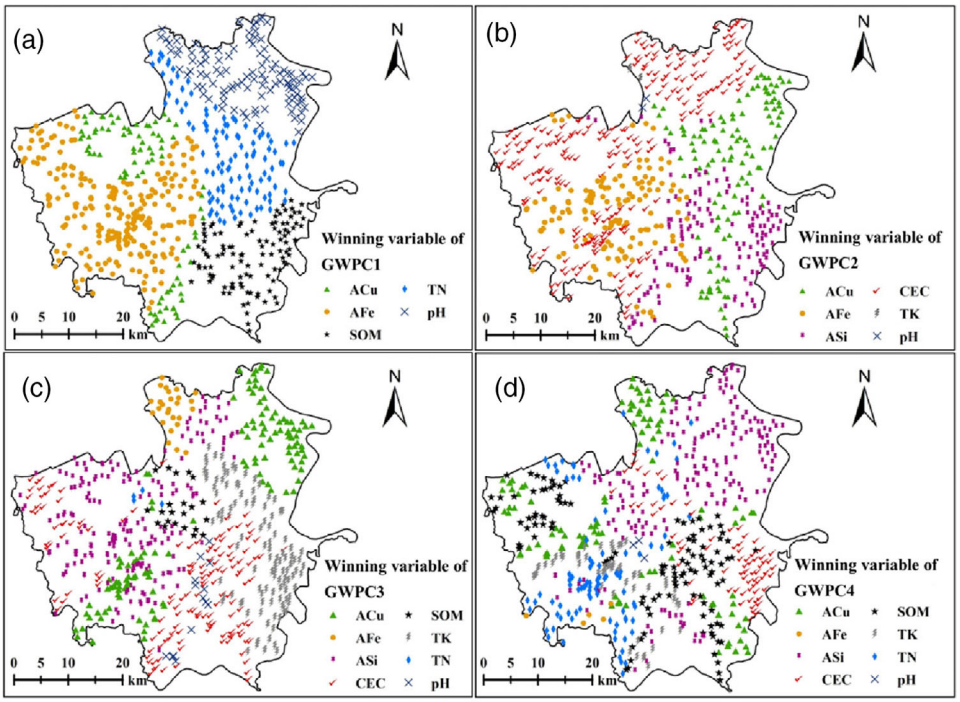

FIGURE 5 Maps of the winning variables (i.e., the variables with the highest loadings) for (a) GWPC1, (b) GWPC2, (c) GWPC3, and (d) GWPC4. GWPC, geographically weighted principal component

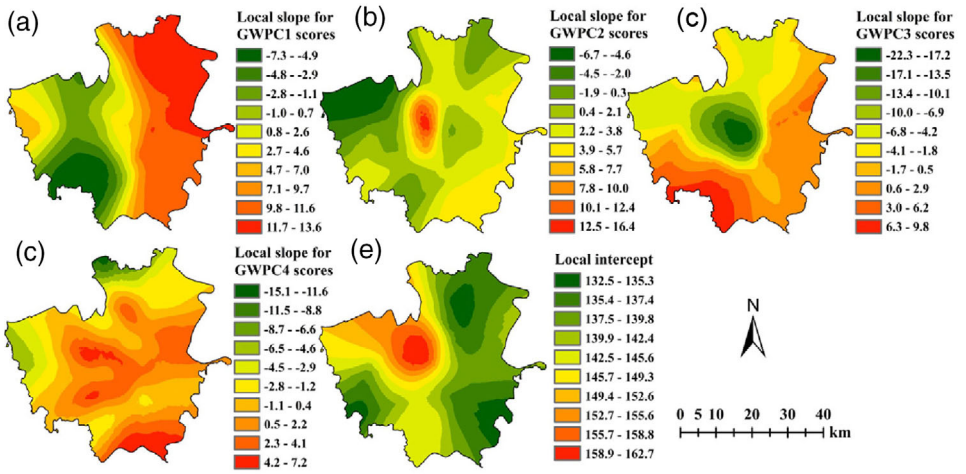

FIGURE 6 Local regression coefficient maps generated by geographically weighted principal component analysis-geographically weighted regression (GWPCA-GWR): (a) the local regression coefficient for the first geographically weighted principal component (GWPC1); (b) the local regression coefficient for GWPC2; (c) the local regression coefficient for GWPC3; (d) the local regression coefficient for GWPC4; and (e) the local intercept

TABLE 3 Parameters of variograms for soil alkaline hydrolyzable nitrogen (AN) and regression residual

Variograms	Model	Nugget	Sill	Nugget/Sill	Range (km)
Auto-variogram					
AN[a]	Spherical	510.86	998.69	0.51	14.17
SOM[b]	Gaussian	8.74	29.33	0.30	18.47
Residuals[c]	Spherical	77.88	89.52	0.87	4.47
Cross-variogram					
(AN × SOM)[d]	Spherical	403.51	876.99	0.46	15.96

[a]AN, auto-variogram fitted by 250 soil AN data;
[b]SOM, auto-variogram fitted by 658 SOM data;
[c]Residuals, auto-variogram fitted by 250 regression residual data generated by GWPCA-GWR;
[d](AN × SOM), cross-variogram fitted by 250 soil AN data and 658 SOM data.

GWPCA-GWR was small, but still statistically significant. Therefore, it is necessary to combine the residual information using GWPCA-GWRK for the spatial prediction of soil AN in this county.

3.5 | Spatial distribution of soil alkaline hydrolyzable nitrogen

Parameters of variograms of soil AN and the regression residual generated by GWPCA-GWR were well fitted by the Spherical model (Table 3). The spatial distribution maps of soil AN concentration predicted by the six models are displayed in Figure 7. In general, the six models produced similar spatial distribution patterns for soil AN, with high concentration mainly in the north of the county. Therefore, in these subareas, some measures, such as reducing the input of N fertilizer and establishing ecological interception ditches, could be implemented to reduce the eutrophication risk in the surrounding water bodies caused by the loss of soil AN (Zhang et al., 2020). Besides, these maps all showed that the soil AN concentration in the east was significantly lower than its average level in the entire county. The reason may be that soil in this subarea has been affected by the side seepage water of Han River, and soil AN in the topsoil has been leached and lost.

However, there were some differences for the spatial prediction maps. For example, the maps predicted by CoK, PCA-MLR, PCA-GWR, GWPCA-GWR, and GWPCA-GWRK were rougher than that predicted by OK (Figure 7). The main reason may be that more auxiliary information was incorporated in the above five models than in OK.

3.6 | Comparison of spatial prediction accuracy

The results of the model comparison are shown in Table 4. GWPCA-GWRK had the highest LCCC (0.75), followed

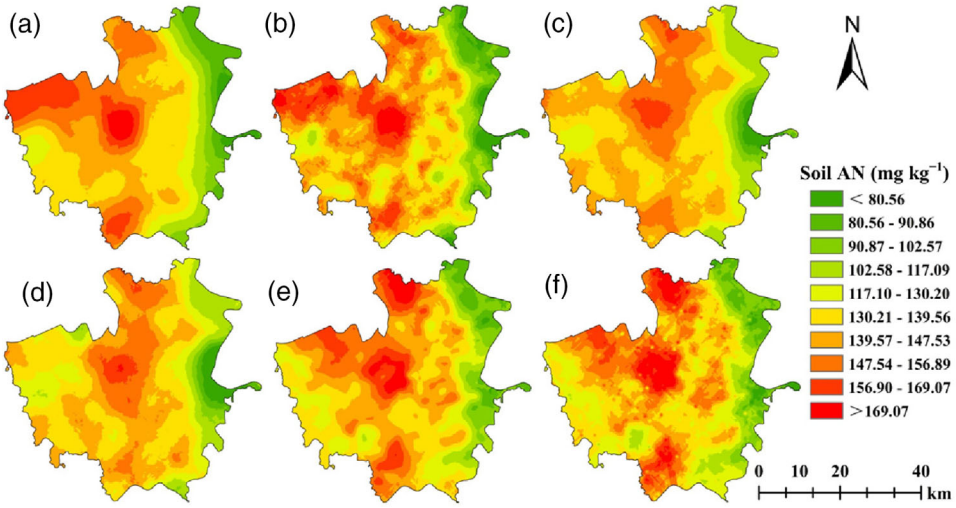

FIGURE 7 Spatial distribution patterns of soil alkaline hydrolyzable nitrogen (AN) generated by different models: (a) ordinary kriging (OK); (b) co-kriging (CoK); (c) principal component analysis-multiple linear regression (PCA-MLR); (d) principal component analysis-geographically weighted regression (PCA-GWR); (e) geographically weighted principal component analysis- geographically weighted regression (GWPCA-GWR); and (f) geographically weighted principal component analysis-geographically weighted regression kriging (GWPCA-GWRK)

TABLE 4 Comparison of the spatial prediction accuracies of six models for soil alkaline hydrolyzable (AN) concentration (mg kg^{-1}) in Shayang County, China ($n = 408$)

	Indices[a]			
Models	**LCCC**	**MAE (mg kg^{-1})**	**RMSE (mg kg^{-1})**	**RI (%)**
OK	0.59	17.64	23.43	—
CoK	0.63	16.53	22.33	4.70
PCA-MLR	0.48	18.46	24.29	-3.67
PCA-GWR	0.63	15.86	21.53	8.09
GWPCA-GWR	0.72	12.80	19.58	16.42
GWPCA-GWRK	0.75	12.79	18.80	19.74

[a]LCCC, Lin's concordance correlation coefficient; MAE, mean absolute error; RMSE, root mean square error; RI, relative improvement accuracy with OK as a reference.

by the GWPCA-GWR model (0.72); CoK and PCA-GWR had approximately equal LCCC (0.63); PCA-MLR had the lowest LCCC (0.48), even lower than that of OK (0.59). According to MAE (mg kg^{-1}) and *RMSE* (mg kg^{-1}), the spatial prediction performance of the six models followed the order of PCA-MLR (MAE = 18.46, RMSE = 24.29) < OK (MAE = 17.64, RMSE = 23.43) < CoK (MAE = 16.53, RMSE = 22.33) < PCA-GWR (MAE = 15.86, RMSE = 21.53) < GWPCA-GWR (MAE = 12.80, RMSE = 19.58) < GWPCA-GWRK (MAE = 12.79, RMSE = 18.80). Taking OK as the reference, the relative improvement accuracies were 19.74% for GWPCA-GWRK, 16.42% for GWPCA-GWR, 8.09% for PCA-GWR, −3.67% for PCA-MLR, and 4.70% for CoK. Therefore, GWPCA-GWRK had the highest prediction accuracy, followed by GWPCA-GWR.

Compared with the other five models, PCA-MLR performed worst in terms of LCCC (0.48), MAE (18.46 mg kg^{-1}), and RMSE (24.29 mg kg^{-1}). The reason may be that (i) PCA cannot capture the spatially non-stationary structures among the auxiliary variables; (ii) MLR cannot capture the spatially non-stationary relationships between the target variable and PC scores; and (iii) PCA-MLR ignored the valuable residual information. Therefore, even if PCA-MLR combined multiple auxiliary variables, it did not obtain higher spatial prediction accuracy. Compared with PCA-MLR, PCA-GWR also combined the auxiliary information of the PC scores, and the spatial prediction accuracy had been improved (RI = 8.09%). The main reason is that GWR can reveal the spatially non-stationary relationships between the target variable and explanatory variables (Qu, Li, Zhang, Huang, & Zhao, 2014; Yang et al., 2019).

Different from the traditional PCA-GWR, the GWPCA-GWR model used GWPCA to extract the local information of auxiliary variables. Previous studies suggested that GWPCA is more effective than PCA in geographic data processing (Fernández, Cotos-Yáñez, Roca-Pardiñas, & Ordóñez, 2018; Granger et al., 2017; Kumar, Lal, & Lloyd, 2012b). In this study, in the northeast of Shayang County (close to Han River), soil pH and CEC had higher loadings, while in the west of Shayang County (the paddy field area), AFe had higher loadings (Figure 5). These spatially non-stationary structures were effectively captured by GWPCA but ignored by PCA. Moreover, this model incorporated GWR to ensure that spatially non-stationary relationships between the target variable and GWPC scores were used for spatial prediction. Comber, Harris, and Tsutsumida (2016) showed that using GW logistic regression with GWPC scores as explanatory variables greatly improved the accuracy of land cover classification, which was similar to the results of this study.

Compared with GWPCA-GWR, GWPCA-GWRK made slight improvements in the spatial prediction accuracy (RI = 16.42% for GWPCA-GWR, RI = 19.74% for GWPCA-GWRK). Kumar et al. (2012a) illustrated that when the regression residual has spatial autocorrelation, it is worthwhile to combine the kriging-interpolated residuals to construct GWRK. In this study, the residuals generated by GWPCA-GWR had significant spatial dependence (Moran's $I = -0.12$, $p < 0.001$). This part of the residual information made the spatial prediction accuracy of GWPCA-GWRK higher than that of GWPCA-GWR.

3.7 | The application prospect of GWPCA-GWRK

GWPCA-GWRK also provides an optional tool for spatial prediction of other soil properties. For the spatial prediction of soil properties in low-relief areas, the relevant soil properties in the historical database may be valuable auxiliary variables for GWPCA-GWRK. For the spatial prediction of soil properties in large-scale areas with complex landscapes, environmental factors that are easily accessible (e.g., slope, elevation, and vegetation) may be valuable auxiliary variables for GWPCA-GWRK. It is worth noting that the auxiliary variables should have significantly spatial correlation with the target variables. Moreover, if the auxiliary information is raster data, it should have a higher spatial resolution; if the auxiliary information is soil sample data, its sample density should be much higher than that of the target variable. Finally, the fine calculation grid may bring some computational burden in the calculation and decomposition of the spatially varying covariance for GWPCA-GWRK.

4 | CONCLUSIONS

In this study, eight soil properties (i.e., SOM, TN, pH, ASi, AFe, CEC, ACu, and TK) were first determined by geodetector as the auxiliary variables. Then, GWPCA-GWRK was proposed for the spatial prediction of soil AN. Results showed that (i) the relationships among the eight auxiliary variables were spatially non-stationary; (ii) GWPCA-GWRK provided the lowest prediction error (RMSE = 18.80 mg kg^{-1}, MAE = 12.79 mg kg^{-1}) and highest LCCC (0.75); (iii) relative improvement accuracies of GWPCA-GWRK, GWPCA-GWR, PCA-GWR, PCA-MLR, and CoK over OK were 19.74%, 16.42%, 8.09%, −3.67%, and 4.70%, respectively. Compared with the traditional spatial prediction model, GWPCA-GWRK has the following advantages: (i) GWPCA can adequately extract the information about the spatially non-stationary structures among the auxiliary variables; (ii) besides combing the valuable information in the residual, GWRK can effectively capture the spatially non-stationary relationships between the target variable and the independent variable. It is concluded that GWPCA-GWRK is an effective tool for spatial prediction of regional soil properties, which can effectively incorporate multiple auxiliary variables with spatially varying relationships.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China [41771249]; the National Key Research and Development Program of China [2018YFC1800104]; the Knowledge Innovation Program of Institute of Soil Science, Chinese Academy of Sciences [ISSASIP1623]; and the Youth Innovation Promotion Association, CAS [2018348].

AUTHOR CONTRIBUTIONS

Jian Chen: Conceptualization; Data curation; Formal analysis; Methodology; Writing-original draft. Mingkai Qu: Conceptualization; Funding acquisition; Investigation; Methodology; Resources; Validation; Writing-review & editing. Jianlin Zhang: Investigation; Resources,Validation; Visualization. Enze Xie: Software; Validation; Yongcun Zhao: Project administration; Supervision; Visualization. Biao Huang: Project administration; Supervision; Validation.

CONFLICT OF INTEREST

The authors have no conflicts of interest.

ORCID

Mingkai Qu https://orcid.org/0000-0003-3974-3242
Yongcun Zhao https://orcid.org/0000-0001-6747-1665

REFERENCES

Comber, A. J., Harris, P., & Tsutsumida, N. (2016). Improving land cover classification using input variables derived from a geographically weighted principal components analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *119*, 347–360. https://doi.org/10.1016/j.isprsjprs.2016.06.014.

Fernández, S., Cotos-Yáñez, T., Roca-Pardiñas, J., & Ordóñez, C. (2018). Geographically weighted principal components analysis to assess diffuse pollution sources of soil heavy metal: Application to rough mountain areas in Northwest Spain. *Geoderma*, *311*, 120–129. https://doi.org/10.1016/j.geoderma.2016.10.012.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The Analysis of spatially varying relationships*. Chichester: Wiley.

Gollini, I., Lu, B. B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: An R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, *63*(17), 1–50. https://doi.org/10.18637/jss.v063.i17.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford Univ. Press.

Granger, S. J., Harris, P., Peukert, S., Guo, R. R., Tamburini, F., Blackwell, M. S. A., … Mcgrath, S. (2017). Phosphate stable oxygen isotope variability within a temperate agricultural soil. *Geoderma*, *285*, 64–75. https://doi.org/10.1016/j.geoderma.2016.09.020.

Guo, L., Luo, M., Zhangyang, C. S., Zeng, C., Wang, S. Q., & Zhang, H. T. (2018). Spatial modelling of soil organic carbon stocks with combined principal component analysis and geographically weighted regression. *The Journal of Agricultural Science*, *156*(6), 774–784. https://doi.org/10.1017/s0021859618000709.

Harris, P., & Brunsdon, C. (2010). Exploring spatial variation and spatial relationships in a freshwater acidification critical load data set for Great Britain using geographically weighted summary statistics. *Computers & Geosciences*, *36*(1), 54–70. https://doi.org/10.1016/j.cageo.2009.04.012.

Harris, P., Brunsdon, C., & Charlton, M. (2011). Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, *25*(10), 1717–1736. https://doi.org/10.1080/13658816.2011.554838.

Harris, P., Clarke, A., Juggins, S., Brunsdon, C., & Charlton, M. (2015). Enhancements to a geographically weighted principal component analysis in the context of an application to an environmental data set. *Geographical Analysis*, *47*(2), 146–172. https://doi.org/10.1111/gean.12048.

Hengl, T., Heuvelink, G. B. M., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, *120*(1–2), 75–93. https://doi.org/10.1016/j.geoderma.2003.08.018.

Howarth, R. W., & Marino, R. (2006). Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: Evolving views over three decades. *Limnology and Oceanography*, *51*(1), 364–376. https://doi.org/10.4319/lo.2006.51.1_part_2.0364.

Huang, B., Sun, W. X., Zhao, Y. C., Zhu, J., Yang, R. Q., Zou, Z., … Su, J. P. (2007). Temporal and spatial variability of soil organic matter and total nitrogen in an agricultural ecosystem as affected by farming practices. *Geoderma*, *139*(3–4), 336–345. https://doi.org/10.1016/j.geoderma.2007.02.012.

IUSS Working Group WRB (2015). *World reference base for soil resources 2014, update 2015, international soil classification system for naming soils and creating legends for soil maps*. Rome: FAO.

Jingmen Municipal Office of Soil Survey. (1984). *Soil annals of Jingmen City*. (In Chinese.) Wuhan, China: Jingmen Municipal Office of Soil Survey.

Knotters, M., Brus, D. J., & Voshaar, J. H. O. (1995). A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, *67*(3–4), 227–246. https://doi.org/10.1016/0016-7061(95)00011-C.

Kumar, S., Lal, R., & Liu, D. S. (2012a). A geographically weighted regression kriging approach for mapping soil organic carbon stock. *Geoderma*, *189*, 627–634. https://doi.org/10.1016/j.geoderma.2012.05.022.

Kumar, S., Lal, R., & Lloyd, C. D. (2012b). Assessing spatial variability in soil characteristics with geographically weighted principal components analysis. *Computational Geosciences*, *16*, 827–835. https://doi.org/10.1007/s10596-012-9290-6.

Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255–268. https://doi.org/10.2307/2532051.

Lindsay, W. L., & Norvell, W. A. (1978). Development of a DTPA soil test for zinc, iron, manganese and copper. *Soil Science Society of America Journal*, *42*(3), 421–428. https://doi.org/10.2136/sssaj1978.03615995004200030009x.

Lloyd, C. D. (2010). Analysing population characteristics using geographically weighted principal components analysis: A case study of Northern Ireland in 2001. *Computers, Environment and Urban Systems*, *34*(5), 389–399. https://doi.org/10.1016/j.compenvurbsys.2010.02.005.

Lu, B. B., Harris, P., Charlton, M., & Brunsdon, C. (2014). The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science*, *17*(2), 85–101. https://doi.org/10.1080/10095020.2014.917453.

Lu, R. K. (2000). *Soil analytical methods of agronomic chemical*. (In Chinese.) Beijing: China Agricultural Science and Technology Press.

Mishra, U., Lal, R., Liu, D. S., & Van Meirvenne, M. (2010). Predicting the spatial variation of the soil organic carbon pool at a regional scale. *Soil Science Society of America Journal*, *74*(3), 906–914. https://doi.org/10.2136/sssaj2009.0158.

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, *30*(7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012.

Philippi, T. E. (1993). Multiple regression: Herbivory. In S. M. Scheiner & J. Gurevitch (Eds.), *Design and analysis of ecological experiments* (pp. 183–210). New York: Oxford Univ. Press.

Piccini, C., Marchetti, A., & Francaviglia, R. (2014). Estimation of soil organic matter by geostatistical methods: Use of auxiliary information in agricultural and environmental assessment. *Ecological Indicators*, *36*, 301–314. https://doi.org/10.1016/j.ecolind.2013.08.009.

Qu, M. K., Chen, J., Huang, B., & Zhao, Y. C. (2020). Enhancing apportionment of the point and diffuse sources of soil heavy metals using robust geostatistics and robust spatial receptor model with categorical soil-type data. *Environmental Pollution*, *265*, 114964. https://doi.org/10.1016/j.envpol.2020.114964.

Qu, M. K., Li, W. D., Zhang, C. R., Huang, B., & Zhao, Y. C. (2014). Spatially nonstationary relationships between copper accumulation in rice grain and some related soil properties in paddy fields at a regional scale. *Soil Science Society of America Journal*, *78*(5), 1765–1774. https://doi.org/10.2136/sssaj2014.02.0067.

Qu, M. K., Wang, Y., Huang, B., & Zhao, Y. C. (2018). Source apportionment of soil heavy metals using robust absolute principal component scores-robust geographically weighted regression (RAPCS-RGWR) receptor model. *Science of The Total Environment*, *626*, 203–210. https://doi.org/10.1016/j.scitotenv.2018.01.070.

Schleus, U., Wu, Q. L., & Blume, H. (1998). Variability of soils in urban and periurban areas in Northern Germany. *Catena*, *33*(3), 255–270. https://doi.org/10.1016/S0341-8162(98)00070-8.

Song, X. D., Brus, D. J., Liu, F., Li, D. C., Zhao, Y. G., Yang, J. L., & Zhang, G. L. (2016). Mapping soil organic carbon content by geographically weighted regression: A case study in the Heihe River Basin, China. *Geoderma*, *261*, 11–22. https://doi.org/10.1016/j.geoderma.2015.06.024.

Wang, J. F., Li, X. H., Christakos, G., Liao, Y. L., Zhang, T., Gu, X., & Zheng, X. Y. (2010). Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *International Journal of Geographical Information Science*, *24*(1), 107–127. https://doi.org/10.1080/13658810802443457.

Wang, J. F., Zhang, T. L., & Fu, B. J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators*, *67*, 250–256. https://doi.org/10.1016/j.ecolind.2016.02.052.

Wang, K., Zhang, C. R., & Li, W. D. (2013). Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and cokriging. *Applied Geography*, *42*, 73–85. https://doi.org/10.1016/j.apgeog.2013.04.002.

Wang, Y. Q., Zhang, X. C., & Huang, C. Q. (2009). Spatial variability of soil total nitrogen and soil total phosphorus under different land uses in a small watershed on the Loess Plateau, China. *Geoderma*, *150*(1–2), 141–149. https://doi.org/10.1016/j.geoderma.2009.01.021.

Xie, E. Z., Zhao, Y. C., Li, H. D., Shi, X. Z., Lu, F. Y., Zhang, X., & Peng, Y. X. (2019). Spatio-temporal changes of cropland soil pH in a rapidly industrialising region in the Yangtze River Delta of China, 1980–2015. *Agriculture, Ecosystems & Environment*, *272*, 95–104. https://doi.org/10.1016/j.agee.2018.11.015.

Yang, S. H., Liu, F., Song, X. D., Lu, Y. Y., Li, D. C., Zhao, Y. G., & Zhang, G. L. (2019). Mapping topsoil electrical conductivity by a mixed geographically weighted regression kriging: A case study in the Heihe River Basin, northwest China. *Ecological Indicators*, *102*, 252–264. https://doi.org/10.1016/j.ecolind.2019.02.038.

Zhai, L., Li, S., Zou, B., Sang, H. Y., Fang, X., & Xu, S. (2018). An improved geographically weighted regression model for $PM_{2.5}$ concentration estimation in large areas. *Atmospheric Environment*, *181*, 145–154. https://doi.org/10.1016/j.atmosenv.2018.03.017.

Zhang, J. L., Wang, Y., Qu, M. K., Chen, J., Yang, L. F., Huang, B., & Zhao, Y. C. (2020). Source apportionment of soil nitrogen and phosphorus based on robust residual kriging and auxiliary soil-type map in Jintan County, China. *Ecological Indicators*, *119*, 106820. https://doi.org/10.1016/j.ecolind.2020.106820.

Zhao, Y. C., Wang, Z. G., Sun, W. X., Huang, B., Shi, X. Z., & Ji, J. F. (2010). Spatial interrelations and multi-scale sources of soil heavy metal variability in a typical urban-rural transition area in Yangtze River Delta region of China. *Geoderma*, *156*(3–4), 216–227. https://doi.org/10.1016/j.geoderma.2010.02.020.

Zhu, Z. L., Wen, Q. X., & Freney, J. R. (1997). Nitrogen management and environmental and crop quality. In M. L. Shan (Ed.), *Nitrogen in soils of China* (pp. 314–315). Dordrecht: Kluwer Academic Publishers.

How to cite this article: Chen J, Qu M, Zhang J, Xie E, Zhao Y, Huang B. Improving the spatial prediction accuracy of soil alkaline hydrolyzable nitrogen using GWPCA-GWRK. *Soil Sci Soc Am J*. 2021;*85:*879–892. https://doi.org/10.1002/saj2.20189