Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions
Open Access
EGU

Improved logistic regression model based on a spatially weighted technique (ILRBSWT v1.0) and its application to mineral prospectivity mapping

Daojun Zhang[1 2*], Na Ren[1], Xianhui Hou[1]

[1]College of Economics and Management, Northwest A&F University, Yangling 712100, China

[2]Center for Resource Economics and Environment Management, Northwest A&F University, Yangling 712100, China

*Corresponding author: cugzdj@gmail.com (Zhang, D)

Abstract: Due to complexity, multiple minerogenic stages, and superposition during geological processes, the spatial distributions of geological variables also exhibit specific trends and non-stationarity. For example, geochemical elements exhibit obvious spatial non-stationarity and trends because of the deposition of different types of coverage. Thus, bias may clearly occur under these conditions when general regression models are applied to mineral prospectivity mapping (MPM). In this study, we used a spatially weighted technique to improve general logistic regression and developed an improved model called the improved logistic regression model based on spatially weighted technique (ILRBSWT, version 1.0). The capabilities and advantages of ILRBSWT are as follows: (1) ILRBSWT is essentially a geographically weighted regression (GWR) model, and thus it has all its advantages when dealing with spatial trends and non-stationarity; (2) the current software employed for GWR mainly applies linear regression whereas ILRBSWT is based on logistic regression, which is used more commonly in MPM because mineralization is a binary event; (3) a missing data process method borrowed from weights of evidence is included to extend the adaptability when dealing with multisource data; and (4) the differences of data quality or exploration level can also be weighted in the new model as well as the geographical distance.

Keywords: anisotropy; geographical information system modeling; geographically weighted logistic regression; mineral resource assessment; missing data; trend variable; weights of evidence.

1 Introduction

The main distinguishing characteristic of spatial statistics compared with classical statistics is that the former has a location attribute. Before the development of geographical information systems, spatial statistical problems were often transformed into general statistical problems, where the spatial coordinates were more like a sample ID because they only had an indexing feature. However, even in non-spatial statistics, the reversal paradox or amalgamation paradox (Pearson et al., 1899; Yule, 1903; Simpson, 1951), which is commonly called Simpson's paradox (Blyth, 1972), has attracted much attention from statisticians and other researchers. In spatial statistics, some spatial variables usually exhibit certain trends and non-stationarity. Thus, it is possible for Simpson's paradox to occur when a global regression model is applied and the existence of unknown important variables may make this condition even worse. The influence of Simpson's paradox can be fatal. For example, due to the presence of cover and other factors that occur after mineralization, the ore-forming elements in Area I are generally much lower than those in Area II, but the actual probability of a mineral in Area I is higher than that in Area II, and more deposits may be discovered in Area I (Agterberg, 1971). In this case, a negative correlation will be obtained between the ore-forming elements and the mineralization according to the classical regression model, whereas a high positive correlation can be obtained in both areas if they are separated. Simpson's paradox is an extreme case of the bias caused by using a global model and it is usually not so severe in practice. However, this type of biased needs to be considered and we should take care when applying a classical regression model to a spatial problem. Several solutions to this issue have been proposed previously, which can be divided into three types.

(1) Locations are introduced as direct or indirect independent variables. Several studies have employed spatial trend variables (Agterberg, 1964; Agterberg and Cabilio, 1969; Agterberg, 1970; Agterberg and Kelly, 1971; Agterberg, 1971) to express linear or nonlinear trends in space by adding coordinate variables or their functions in predictive models. In these methods, the locations themselves are taken as independent variables as well as the normal independent variables. For example, Reddy et al. (1991) performed logistic regression by including trend variables for mapping the base-metal potential in the Snow Lake area, Manitoba, Canada. In addition, Casetti (1972) developed a spatial expansion method where the regression parameters are themselves functions of the x and y coordinates as well as their combinations.

(2) Using local models to replace global models, i.e., geographically weighted models (Fotheringham et al., 2002). Geographically weighted regression (GWR) is the most popular model among the geographically weighted models. GWR was first developed at the end of the 20th century by Brunsdon et al. (1996) and Fotheringham et al. (1996, 1997, 2002) for modeling spatially heterogeneous processes, and it has been used widely in the field of geography.

(3) Reducing the trends in spatial variables. For example, Cheng developed a local singularity analysis technique and spectrum-area (S-A) model based on fractal/multi-fractal theory (Cheng, 1997; Cheng, 1999). These methods can remove spatial trends and prevent the strong effects of the original high and low values of the variables on predictions, and thus they are used widely to weaken the effect of spatial non-stationarity to some degree (e.g., Zuo et al., 2016; Zhang et al., 2016; Xiao et al., 2017).

GWR can be readily visualized and understood, and it is particularly valid for dealing with spatial non-stationarity, thus it has been used widely in geography and other areas that require spatial data analysis. In general, GWR is a moving window-based model where instead of establishing a unique and global model for prediction, it makes a prediction for each current location using the surrounding samples, and a higher weight is given when the sample is located closer. The theoretical foundation of GWR is based on Tobler's observation that: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). In mineral prospectivity mapping (MPM), the dependent variables are binary and logistic regression is used instead of linear regression, and it is necessary to apply geographically weighted logistic regression (GWLR) instead. GWLR belongs to geographically weighed generalized linear regression model (Fotheringham et al. 2002) and it is included in the software module GWR 4.09 (Nakaya, 2016). However, GWLR can only deal with the data in the form of a tabular dataset containing the fields of dependent and independent variables, and the x-y coordinates. Therefore, the spatial layers must be re-processed into two-dimensional tables and the resulting data needs to be transformed back into a spatial form. Another problem with the application of GWR 4.09 for MPM is that it cannot deal with missing data (Nakaya, 2016). Weights of evidence (WofE) is a widely used model for MPM (Bonham-Carter et al., 1988, 1989; Agterberg, 1989; Agterberg et al., 1990), which can avoid the effect of missing data. However, WofE was developed based on the premise that an assumption of conditional independence is satisfied among the evidential layers with respect to the target layer; otherwise, the posterior probabilities will be biased and the number of estimated deposits will not be equal to the known deposits. Agterberg (2011) combined WofE with logistic regression and proposed a new model that can obtain an unbiased estimated of the number of deposits as well as avoiding the effect of missing data. In the present study, this concept is employed to deal with missing data and we propose the improved logistic regression model based on spatially weighted technique (ILRBSWT v1.0) for MPM. The main features of ILRBSWT include the following: (1) a spatial

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions
Open Access

t-statistics method (Agterberg et al., 1993) is introduced to determine the best binary threshold for independent variables, where binarization is performed based on a local window instead of the global level, which can increase the effect of indicating the independent variables to the target variable; and (2) a mask layer is included in the new model to deal with the data quality and exploration level differences among samples.

The idea of this research is origin from the first author's doctoral thesis (Zhang, 2015) in Chinese, which has been shown to have better efficiency for mapping intermediate and felsic igneous rocks (Zhang et al., 2017). The contribution of this research is to elaborate the principle of ILRBSWT, and provide a detailed algorithm for its design and implementation process with the code and software module attached. In addition, the processing of missing data is not covered by former researches. At last, the prediction of Au ore deposits in western Meguma Terrain, Nova Scotia, Canada, is chosen as case study to show the performance of ILRBSWT in MPM.

2 Models

Linear regression is commonly used for exploring the relationship between a response variable and one or more explanatory variables. However, in MPM and other fields, the response variable is binary or dichotomous, so linear regression is not applicable and thus a logistic model can be advantageous.

2.1 Logistic Regression

In MPM, the dependent variable(Y) is binary since Y can only take the value of 1 and 0, which means the mineralization occurs or not. Suppose that π represents the estimation of Y, 0≤π≤1, then a logit transformation of π can be made, i.e., logit (π) =ln(π/(1-π)). Logistic regression function can be obtained as following.

$$\text{Logit}\ \pi(X_1, X_2, \cdots, X_p) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \quad (1)$$

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

where $X_1, X_2, \cdots, X_p$, comprises a sample of p explanatory variables $x_1, x_2, \cdots, x_p$, β_0 is the intercept, and $\beta_1, \beta_2, \cdots, \beta_p$ are regression coefficients.

If there are n samples, we can obtain n linear equations with p+1 unknowns based on equation (1). Furthermore, if we suppose that the observed values for Y are $Y_1, Y_2, \cdots, Y_n$, and these observations are independent of each other, then a likelihood function can be established:

$$\mathrm{L}(\beta) = \prod_{i=1}^{n} (\pi_i^{Y_i} (1 - \pi_i)^{1-Y_i}), \tag{2}$$

where $\pi_i = \pi(X_{i1}, X_{i2}, \cdots, X_{ip}) = \frac{e^{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}}}$. The best estimate can be obtained if and only if equation (2) takes the maximum. Then the problem is converted into solving $\beta_1, \beta_2, \cdots, \beta_p$. Equation (2) can be further transformed into the following log-likelihood function.

$$\mathrm{lnL}(\beta) = \sum_{i=1}^{n} (Y_i \pi_i + (1 - Y_i)(1 - \pi_i)) \tag{3}$$

The solution can be obtained by taking the first partial derivative of β_i (i = 0 to p), which should be equal to 0.

$$\begin{cases} f(\beta_0) = \sum_{i=0}^{n} (Y_i - \pi_i) X_{i0} = 0 \\ f(\beta_1) = \sum_{i=0}^{n} (Y_i - \pi_i) X_{i1} = 0 \\ \vdots \\ f(\beta_p) = \sum_{i=0}^{n} (Y_i - \pi_i) X_{ip} = 0 \end{cases} \tag{4}$$

where $X_{i0} = 1$, i takes the value from 1 to n, and equation (4) is obtained in the form of matrix operations.

$$\mathbf{X}^{\mathbf{T}}(\mathbf{Y} - \boldsymbol{\pi}) = \mathbf{0} \tag{5}$$

The Newton iterative method can be used to solve the nonlinear equations:

$$\hat{\boldsymbol{\beta}}(t+1) = \hat{\boldsymbol{\beta}}(t) + \mathbf{H}^{-1}\mathbf{U} \;, \tag{6}$$

where $\mathbf{H} = \mathbf{X}^{\mathbf{T}}\mathbf{V}(t)\mathbf{X}$, $\mathbf{U} = \mathbf{X}^{\mathbf{T}}(\mathbf{Y} - \boldsymbol{\pi}(t))$, t represents the number of iterations, and $\mathbf{V}(t)$, $\mathbf{X}$, $\mathbf{Y}$, $\boldsymbol{\pi}(t)$, and $\hat{\boldsymbol{\beta}}(t)$ are obtained as follows:

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions
Open Access

$$\mathbf{V}(t) = \begin{pmatrix} \pi_1(t)(1-\pi_1(t)) & & & \\ & \pi_2(t)(1-\pi_2(t)) & & \\ & & \ddots & \\ & & & \pi_n(t)(1-\pi_n(t)) \end{pmatrix},$$

$$\mathbf{X} = \begin{pmatrix} X_{10} & X_{11} & \cdots & X_{1p} \\ X_{20} & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n0} & X_{n1} & \cdots & X_{np} \end{pmatrix}, \ \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \ \boldsymbol{\pi}(t) = \begin{pmatrix} \pi_1(t) \\ \pi_2(t) \\ \vdots \\ \pi_n(t) \end{pmatrix}, \text{ and } \hat{\boldsymbol{\beta}}(t) = \begin{pmatrix} \hat{\beta}_1(t) \\ \hat{\beta}_2(t) \\ \vdots \\ \hat{\beta}_n(t) \end{pmatrix}.$$

Hosmer et al. (2013) provided more information about the derivation from equations (1) to (6).

2.2 Weighted Logistic Regression

In practice, vector data is popularly used, and sample size (area) has to be considered. In this condition, weighted logistic regression modeling should be used instead of general logistic regression. In addition, it is preferable to use a weighted logistic regression model when a logical regression should be performed for large sample data, since weighted logical regression can greatly reduce the size of the matrix and improve the computational efficiency (Agterberg, 1992). Assuming that there are four binary explanatory variable layers and the study area consists of 1000×1000 grid points, the matrix size for normal logic regression modeling would be $10^6 \times 10^6$; however, if weighted logistic regression is used, the matrix size would be 32×32 at most. That is because sample classification process is contained in weighted logistic regression, and all samples are classified into the classes which own the same values at dependent and each independent variables. The samples with the same dependent and independent variables form certain continuous and discontinuous patterns in space, which are called “unique condition” units. Each unique condition unit is then treated as a sample, and the area (grid number) for it is taken as weight in weighed logistic regression. Thus, in the case of weighted logical regression, equations (2) to (5) in section 2.1 need to be changed as following Equations (7) to (10) respectively.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions
Open Access
EGU

$$\mathrm{L}_{new}(\beta) = \prod_{i=1}^{n}(\pi_i^{N_i Y_i}(1-\pi_i)^{N_i(1-Y_i)}), \tag{7}$$

$$\mathrm{lnL}_{new}(\beta) = \sum_{i=1}^{n}(N_i Y_i \pi_i + N_i(1-Y_i)(1-\pi_i)) \tag{8}$$

$$\begin{cases} f_{new}(\beta_0) = \sum_{i=0}^{n}(Y_i - \pi_i)X_{i0} = 0 \\ f_{new}(\beta_1) = \sum_{i=0}^{n}(Y_i - \pi_i)X_{i1} = 0 \\ \vdots \\ f_{new}(\beta_p) = \sum_{i=0}^{n}(Y_i - \pi_i)X_{ip} = 0 \end{cases} \tag{9}$$

$$\mathbf{X}^{\mathbf{T}}\mathbf{W}(\mathbf{Y} - \boldsymbol{\pi}) = \mathbf{0} \tag{10}$$

where N_i is the weight for the *i*-th unique condition unit, *i* takes the value from 1 to *n*, and *n* is the total number of grid points. And **W** is a diagonal matrix which can be expressed as following.

$$\mathbf{W} = \begin{pmatrix} N_1 & & & \\ & N_2 & & \\ & & \ddots & \\ & & & N_n \end{pmatrix}$$

Besides, new H and U should be used in equation (6) to perform Newton iterative under weighted logistic regression, i.e., $\mathbf{H}_{new} = \mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{V}(t)\mathbf{X}$, $\mathbf{U}_{new} = \mathbf{X}^{\mathbf{T}}\mathbf{W}(\mathbf{Y} - \boldsymbol{\pi}(t))$.

2.3 Geographically Weighted Logistic Regression

GWLR is a local window-based model because logistic regression is established at each current location in GWLR. The current location is changed using the moving window technique with a loop program. If we suppose that $\mathbf{u}$ represents the current location, which can be uniquely determined by a pair of column and row numbers, $\boldsymbol{x}$ denotes that *p* explanatory variables $x_1, x_2, \cdots, x_p$ take values of $X_1, X_2, \cdots, X_p$, respectively, and $\pi(\boldsymbol{x}, \mathbf{u})$ is the estimates of *Y*, i.e., the probability that *Y* takes a value of 1, then the following function can be obtained.

$$\mathrm{Logit}\,\pi(\boldsymbol{x}, \mathbf{u}) = \beta_{0i}(\mathbf{u}) + \beta_1(\mathbf{u})X_1 + \beta_2(\mathbf{u})X_2 + \cdots + \beta_p(\mathbf{u})X_p \ , \tag{11}$$

where $\beta_0(\mathbf{u})$, $\beta_1(\mathbf{u})$, $\cdots$, $\beta_p(\mathbf{u})$ denote that these parameters are obtained at the location of $\mathbf{u}$. The predicted probability for the current location can be obtained under the condition that the values of all the independent variables are known at the current location and all of the

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions
Open Access
EGU

parameters are also calculated based on the samples within the current local window. According to equation (6) in section 2.1, the parameters for GWLR can be estimated with equation (12):

$$\widehat{\boldsymbol{\beta}}(\mathbf{u})_{t+1} = \widehat{\boldsymbol{\beta}}(\mathbf{u})_t + (\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathbf{u})\mathbf{V}(t)\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathbf{u})(\mathbf{Y} - \boldsymbol{\pi}(t)), \quad (12)$$

where t represents the number of iterations; $\mathbf{X}$ is a matrix comprising the values of all the independent variable, and all of the elements in the first column are 1; $\mathbf{W}(\mathbf{u})$ is a diagonal matrix where the diagonal elements are geographical weights, which can be calculated according to distance, whereas the other elements are all 0; $\mathbf{V}(t)$ is also a diagonal matrix and the diagonal element can be expressed as $\pi_i(t)(1 - \pi_i(t))$; and $\mathbf{Y}$ is a column vector representing the values taken by the dependent variable.

2.4 Improved Logistic Regression Model based on Spatially Weighted Technique

If a diagonal element in $\mathbf{W}(\mathbf{u})$ is only for one sample (grid point in raster data), section 2.3 can be seen as the improvement of section 2.1, i.e. samples are weighted according to its location. If samples are reclassified firstly according to unique condition mentioned in section 2.2, and corresponding weights are then summarized according to each sample's geographical weight, we can obtain an improved logistic regression model considering both sample sizes and geographical distances. The new model can not only reflects the spatial distribution of samples, but also reduce the matrix size, and it is to be discussed in following section.

In addition to geographic factors, the degree considered in the study can affect the representativeness of a sample, e.g., differences in the level of exploration.

Suppose that there are n grid points in the current local window, S_i is the i-th grid, $W_i(g)$ is the geographical weight of S_i, and $W_i(d)$ represents the individual difference weight or non-geographical weight (in some cases, there may be differences in quality or the exploration level among samples, but $W_i(d)$ takes a value of 1 if there is no difference), where i takes a value from 1 to n. Furthermore, if we suppose that there are N unique

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

conditions after overlaying all of the layers ($N \leq n$) and C_j denotes the j-th unique condition unit, then we can obtain the final weight for each unique condition unit in the current local window:

$$W_j(t) = \sum_{i=1}^{n}[W_i(\mathrm{g}) * W_i(d) * \mathrm{df}_i], \tag{13}$$

where $\begin{cases}\mathrm{df}_i = 1 & \text{if } \mathrm{S}_i \in \mathrm{C}_j \\ \mathrm{df}_i = 0 & \text{if } \mathrm{S}_i \notin \mathrm{C}_j\end{cases}$, i takes a value from 1 to n, j takes a value from 1 to N, and $W_j(t)$ represents the total weight (by combining both $W_i(g)$ and $W_i(d)$) for each unique condition unit. We can use the final weight calculated in equation (13) to replace the original weight in equation (12), which is one of the advantages of ILRBSWT.

2.5 Missing data processing

Missing data is a problem existing in all statistics-related research fields. In MPM, missing data are also prevalent due to ground coverage, and limitations of exploration technique and measurement accuracy. Agterberg and Bonham-Carter (1999) once compared following commonly used missing data processing solutions: (1) removing variables containing missing data, (2) deleting samples with missing data, (3) using 0 to replace the missing data, and (4) replacing the missing data with the mean of the corresponding variable. From the point of utilization efficiency of existing data, both (1) and (2) are clearly not good solutions since more data will be lost. Solution (3) is superior to (4) for missing values due to the detection limit of the measuring instrument; with respect to the missing data caused by the limitation of geographical environment and the prospecting technique, solution (4) is obviously a better choice. Missing data is mainly caused by the latter in MPM, and Agterberg (2011) pointed out that missing data could be even better dealt with by performing WofE model. In WofE, the evidential variable takes the value of positive weight (W^+) if it is favorable for the happening of the target variable (e.g., mineralization); and the evidential variable takes the value of negative weight (W^-) if it is unfavorable for the happening of the target variable; and

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions
Open Access

automatically the evidential variable takes the value of 0 if missing data happens.

$$W^{+} = \ln\frac{\frac{D_1}{D}}{\frac{A_1-D_1}{A-D}} \quad (14)$$

$$W^{-} = \ln\frac{\frac{D_2}{D}}{\frac{A_2-D_2}{A-D}} \quad (15)$$

where A is an evidential layer, A1 means the area that A takes the value of 1, and A2 means the area that A takes the value of 0; A3 means the area with missing data, and A1+A2 is smaller than the total study area if missing data exists. D1, D2 and D3 are the area that the target variable takes the value of 1 in A1, A2 and A3 respectively. In fact, A3 and D3 are not used in equation (15) since no information is provided in area A3.

However, it is preferred to use 1 and 0 to represent the positive and negative condition of the independent variable in logistic regression model. In this case, equation (16) can be used to replace missing data in logistic regression modeling, which will cause an equivalent effect just as missing data are processed in WofE.

$$M = \frac{-W^{-}}{W^{+}-W^{-}} = \frac{\ln\frac{D}{A-D}-\ln\frac{D_2}{A_2-D_2}}{\ln\frac{D_1}{A_1-D_1}-\ln\frac{D_2}{A_2-D_2}} \quad (16)$$

3 Design of the ILRBSWT Algorithm

3.1 Local Window Design

A raster data set is used for ILRBSWT modeling. With a regular grid, the distance between any two grid points can be calculated easily and we can even obtain distance templates within a certain window scope, which is highly efficient for data processing. The circle and ellipse are used for isotropic and anisotropic local window designs, respectively.

(1) Circular Local Window Design

If we suppose that *W* represents a local circular window where the minimum bounding rectangle is *R*, then the geographical weights can be calculated only inside *R*. Obviously, the grid points inside of *R* but outside of *W* should be weighted as 0, and the weights for grid points inside *W* should be calculated according to the distances between themselves and the current location. *R* should be a square so we can also assume that there are *n* columns and rows in *R*, where *n* is an odd number. If we take east and south as the orientations of the *x*-axis and *y*-axis, respectively, and the position of the northwest corner grid is defined as (*x* = 1, *y* = 1), then a local rectangular coordinate system can be established and the position for the current location grid can be expressed as $O\ (x = \frac{n+1}{2},\ y = \frac{n+1}{2})$. The distance between any grid inside *W* and the current location grid can be expressed as $d_{o-ij} = \sqrt{\left(i - \frac{n+1}{2}\right)^2 + \left(j - \frac{n+1}{2}\right)^2}$, where i and j take values ranging from 1 to *n*. The geographical weight is a function of distance, so it is convenient to calculate w_{ij} with d_{o-ij} . Figure 1 shows the weight template for a circular local window with a half-window size of nine grid points.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

0	0	0	0	0	0	0	0	w30	0	0	0	0	0	0	0	0
0	0	0	0	0	w28	w27	w25	w24	w25	w27	w28	0	0	0	0	0
0	0	0	w29	w26	w23	w21	w20	w19	w20	w21	w23	w26	w29	0	0	0
0	0	w29	w25	w22	w18	w16	w15	w14	w15	w16	w18	w22	w25	w29	0	0
0	0	w26	w22	w17	w14	w13	w11	w10	w11	w13	w14	w17	w22	w26	0	0
0	w28	w23	w18	w14	w12	w9	w8	w7	w8	w9	w12	w14	w18	w23	w28	0
0	w27	w21	w16	w13	w9	w6	w5	w4	w5	w6	w9	w13	w16	w21	w27	0
0	w25	w20	w15	w11	w8	w5	w3	w2	w3	w5	w8	w11	w15	w20	w25	0
w30	w24	w19	w14	w10	w7	w4	w2	w1	w2	w4	w7	w10	w14	w19	w24	w30
0	w25	w20	w15	w11	w8	w5	w3	w2	w3	w5	w8	w11	w15	w20	w25	0
0	w27	w21	w16	w13	w9	w6	w5	w4	w5	w6	w9	w13	w16	w21	w27	0
0	w28	w23	w18	w14	w12	w9	w8	w7	w8	w9	w12	w14	w18	w23	w28	0
0	0	w26	w22	w17	w14	w13	w11	w10	w11	w13	w14	w17	w22	w26	0	0
0	0	w29	w25	w22	w18	w16	w15	w14	w15	w16	w18	w22	w25	w29	0	0
0	0	0	w29	w26	w23	w21	w20	w19	w20	w21	w23	w26	w29	0	0	0
0	0	0	0	0	w28	w27	w25	w24	w25	w27	w28	0	0	0	0	0
0	0	0	0	0	0	0	0	w30	0	0	0	0	0	0	0	0

Fig. 1 Weight template for a circular local window with a half-window size of nine grid points, where w1 to w30 represent different weight classes that decrease with distance and 0 denotes that the grid is weighted as 0. Gradient colors ranging from red to green are used to distinguish the weight classes for grid points.

If we suppose that there are T_n columns and T_m rows in the study area, and *Current* (T_i, T_j) represents the current location, where T_i takes values from 1 to T_n and T_j takes values from 1 to T_m, then the current local window can be established by selecting the range of rows $T\_i - \frac{n-1}{2}$ to $T\_i + \frac{n-1}{2}$ and columns $T\_j - \frac{n-1}{2}$ to $T\_j + \frac{n-1}{2}$ based on the total research area. Next, we establish a local rectangular coordinate system according to the steps in the last paragraph, where the *x* and *y* coordinates for the northwest corner are defined as the coordinate origin by subtracting $T\_i - \frac{n-1}{2}$ and $T_j - \frac{n-1}{2}$ from the *x* and *y* coordinates, respectively, for all of the grid points in the range. The corresponding relationship can then be established between the weight template and the current window. Global weights can also be included via the matrix product between the global weight layer and local weight template within the local window. In addition, special care should be taken when the weight template covers some area outside the study area, e.g., $T\_i - \frac{n-1}{2} < 0$, $T\_i + \frac{n-1}{2} > T\_n$, $T\_j - \frac{n-1}{2} < 0$, and $T\_j + \frac{n-1}{2} > T\_m$.

(2) Elliptic Local Window Design

In most cases, the spatial weights change to variable degrees in different directions and an elliptic local window may be better for describing the changes in the weights in space. In order to simplify the calculation, we can convert the distances in different directions into equivalent distances and an anisotropic problem then becomes an isotropic problem. For any grid, the equivalent distance is the semi-major axis length of the ellipse that passes through the grid and that is centered at the current location, where the parameters for the ellipse can be determined using the kriging method.

We still use W to represent the local elliptic window and a, r, and θ are defined as the semi-major axis, the ratio of the semi-minor axis relative to the semi-major axis, and the azimuth of the semi-major axis, respectively. Then, W can be covered by a square R, where the side length is 2a-1 and the center is the same as W. There are (2a –1) × (2a – 1) grid points in R. We establish the rectangular coordinates as described above and we suppose that the center of the top left grid in R is located at (x = 1, y = 1), and thus the center of W should be $O(x_0 = a, y_0 = a)$. According to the definition of the ellipse, two of the elliptical focuses are located at $F_1\left(x_1 = a + \sin(\theta)\sqrt{a^2 - (a*r)^2}, y_1 = a - \mathrm{con}(\theta)\sqrt{a^2 - (a*r)^2}\right)$ and F_2 $(x_2 = a - \sin(\theta)\sqrt{a^2 - (a*r)^2}, y_2 = a + \mathrm{con}(\theta)\sqrt{a^2 - (a*r)^2})$. The summed distances between a point and the two focus points can be expressed as

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions

$l_{ij} = \sqrt{(i-x_1)^2 + (j-y_1)^2} + \sqrt{(i-x_2)^2 + (j-y_2)^2}$, where i and j take values from 1 to $2a-1$. According to the elliptical focus formula, we can decide whether a grid in R is located in W. For any grid in R, if the sum of the distances between the two focal points and a grid center is greater than $2a$, then the grid is located in W, vice versa. For the grid points outside of W, the weight is assigned as 0, and the equivalent distances should be calculated for the grid points within W. As mentioned above, the parameters for the ellipse can be determined using the kriging method. In the ellipse W where the semi-major axis is a, we keep r and θ as constants, so we can obtain countless ellipses centered at the center of W, and the equivalent distance is the same on the same elliptical orbit. Thus, the equivalent distance template can be obtained for the elliptic local window. Figure 2 shows the equivalent distance templates under the conditions that a = 11 grid points, r = 0.5, and the azimuths for the semi-major axis are 0°, 45°, 90°, and 135°, where the weight template can also be calculated based on Fig. 2.

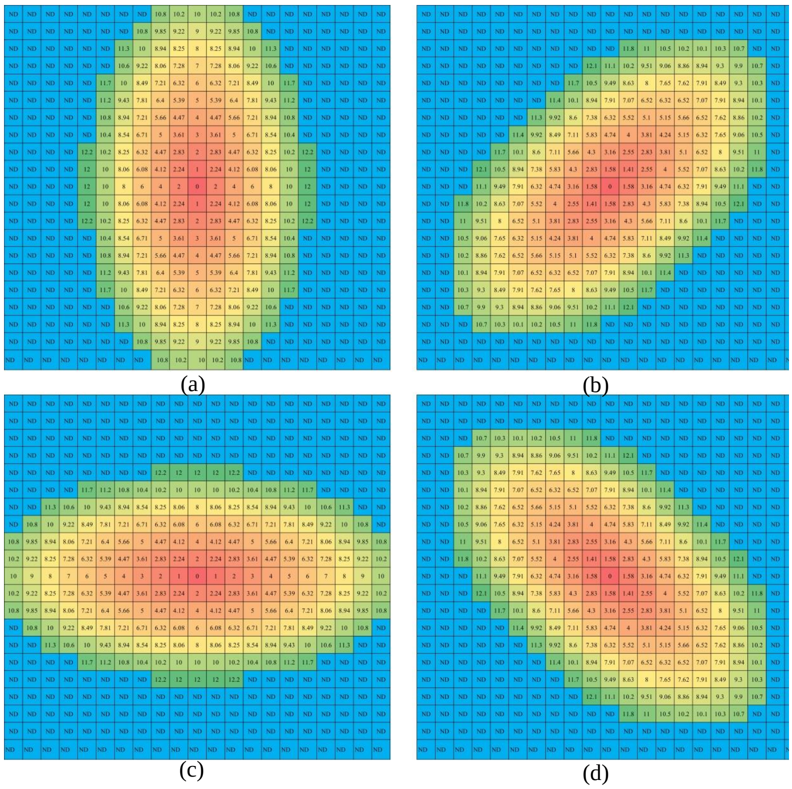


Fig. 2 Construction of the distance template based on a elliptic local window: *a* = 11 grid points, *r* = 0.5, and the azimuths for the semi-major axis are 0 °(a), 45 °(b), 90 °(c), and 135 °(d).

3.2 Pseudocode for ILRBSWT

The ILRBSWT method focuses mainly on two problems, i.e., spatial non-stationarity and missing data. We use the moving window technique to establish a local model, which can overcome the spatial non-stationarity better compared with the global model. The spatial *t*-value employed in the WofE method is used to binarize spatial variables based on the local window, which is quite different from binarization based on the global range, where the missing data can be handled well because positive and negative weights are used instead of the original "1" and "0" values, and the missing data can then be represented well as "0." Both the isotropy and anisotropy window types are possible in our new proposed model. The geographical weights and the window size can be determined by the users themselves. If the

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

geographic weights are equal and there are no missing data, then ILRBSWT will yield the same posterior probabilities as logistic regression; hence, the later can be treated as a special case of the former. The core ILRBSWT algorithm is as follows.

Step 1. Establish a loop for all of the grid points in the study area according to both the columns and rows. Determine a basic local window with a size of $r_{\min}$ based on a variation function or other method. In addition, the maximum local window with a size of $r_{\max}$ is set, with an interval of ΔR. If we suppose that a geographical weight model has already been given in the form of a Gaussian curve determined by variations in the geostatistics, i.e., $W(\mathrm{g}) = \mathrm{e}^{-\lambda d^2}$, where d is the distance and λ is the attenuation coefficient, then we can calculate the geographical weight for any grid in the current local window. The equivalent radius should be used in the anisotropic situation. When other types of weights are considered, e.g., the degree of exploration or research, it is also necessary to synthesize the geographical weights and other weights (see equation 10).

Step 2. Establish a loop for all of the independent variables. In a circular (elliptical) window with a radius (equivalent radius) of $r_{\min}$, apply the WofE (Agterberg, 1992) model according to the grid weight determined in step 1, thereby obtaining a statistical table containing the parameters of W_{ij}^{+}, W_{ij}^{-} , and t_{ij}, where *i* is the *i*-th independent variable and *j* denotes the *j*-th binarization.

Step 2.1. If a maximum t_{ij} exists and it is greater than or equal to the standard *t*-value (e.g., 1.96), record the values of $W_{i-\max\_t}^{+}$, $W_{i-\max\_t}^{-}$, and $B_{i-\max\_\mathrm{t}}$, which denote the positive weight, negative weight, and corresponding binarization, respectively, under the condition where *t* takes the maximum value. Go to step 2 and apply the WofE model to the other independent variables.

Step 2.2. If a maximum t_{ij} does not exist or it is smaller than the standard *t*-value, go to step 3.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Step 3. In a circular (elliptical) window with a radius (equivalent radius) of $r_{\max}$, increase the current local window based on $r_{\min}$ according to the algorithm in step 1.

Step 3.1. If all of the independent variables have already been processed, go to step 4.

Step 3.2. If the size of the current local window exceeds the size of $r_{\max}$, then disregard the current independent variable and go to step 2 to consider the remaining independent variables.

Step 3.3. Apply the WofE model according to the grid weight determined in step 1 in the current local window, which has increased. If a maximum t_{ij} exists and it is greater than or equal to the standard *t*-value, record the values of $W_{i-\max\_t}^{+}$, $W_{i-\max\_t}^{-}$, $B_{i-\max\_\mathrm{t}}$, and r_{curent}, which represents the radius (equivalent radius) for the current local window.

Step 3.4. If a maximum t_{ij} does not exist or it is smaller than the standard *t*-value, go to step 3.

Step 4. Suppose that n_s independent variables are remaining.

Step 4.1. If $n_s \le 1$, then calculate the mean value for the dependent variable in the current local window with a radius size of $r_{\max}$ and retain it as the posterior probability in the current location. In addition, set the regression coefficients for all of the independent variables as missing data. Go to step 6.

Step 4.2. If $n_s \ge 1$, then find the independent variable with the largest local window and apply the WofE model to all the other independent variables, before recording the values of $W_{i-\max\_t}^{+}$, $W_{i-\max\_t}^{-}$, and $B_{i-\max\_t}$ for this time, and then go to step 5.

Step 5. Apply the logistic regression model based on geographic weights and for each independent variable: (1) use $W_{i-\max\_t}^{+}$ to replace all of the values that are less than or equal to $B_{i-\max\_t}$; (2) use $W_{i-\max\_t}^{-}$ to replace all of the values that are greater than $B_{i-\max\_t}$; and (3) use 0 to replace no data ("-9999"). The posterior probability and regression coefficients can then be obtained for all of the independent variables at the current location, and go to step

6.

Step 6. Take the next grid as the current location and repeat steps 2–5.

4 Interface Design

In addition to the improved GWLR, we developed other modeling processes, where all of the visualization and mapping procedures are performed using the ArcGIS 10.2 platform and GeoDAS 4.0 software. The maps are stored in grid format, which are transformed into ASCII files based on tools included in the Arc toolbox before the improved GLWR is performed.

As shown in Fig. 3, the main interface for the improved GLWR comprises four parts. The upper left part is for the layer input settings, where independent variable layers, dependent variable layers, and global weight layers should be assigned if they exist. Layer information is shown at the upper right corner, including the row numbers, column numbers, grid size, ordinate origin, and missing data. The local window can be defined in the middle. Using the drop-down list, we can prepare a circle or ellipse to represent various isotropic and anisotropic spatial conditions, respectively. The corresponding window parameters should be set for each window type. For the ellipse, it is necessary to set parameters comprising the initial length of the equivalent radius (initial major radius), the final length of the equivalent radius (largest major radius), the increase in the length of the equivalent radius (growth rate), the threshold of the spatial t-value used to determine the need to enlarge the window, the length ratio of the major and minor axes, the orientation of the ellipse's major axis, and the compensation coefficient for the sill. Next, it is necessary to define the attenuation function and a variety of kernel functions, such as exponential model, logarithmic model, Gaussian model, or spherical model, via the drop-down menu. More parameters can be set when a certain model is selected. The output file settings are defined at the bottom and the execution button is at the lower right corner.

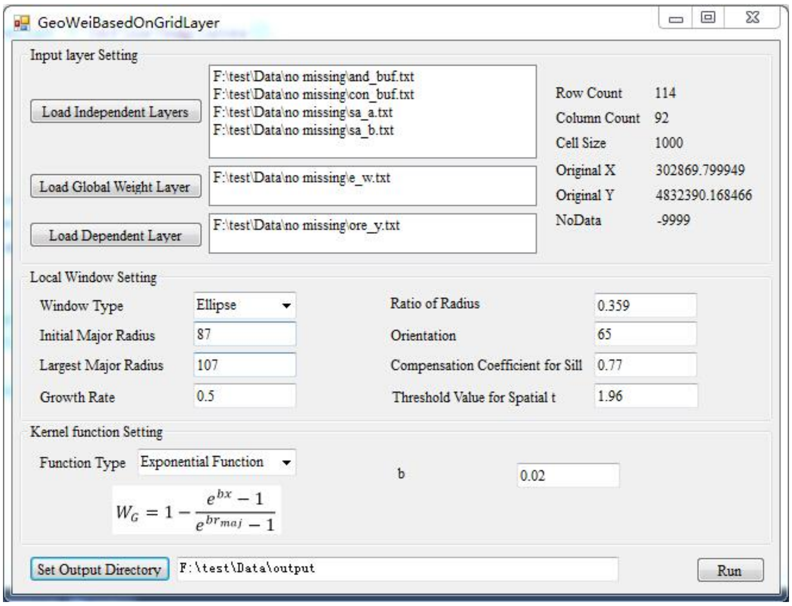


Fig. 3 User interface design.

5 Real Data Testing

5.1 Data source and preprocessing

The test data used in this study were obtained from the case study reported by Cheng (2008). The study area (≈7780 km^2) was located in western Meguma Terrain, Nova Scotia, Canada. Four independent variables were used in the WofE model for gold mineral potential mapping by Cheng (2008), i.e., buffer of anticline axes, buffer for the contact of Goldenville–Halifax Formation, and background and anomaly separated with the S-A filtering method based on the loadings of the ore elements of the first component. More information about the data set can be found in Cheng (2008).

Four independent variables mentioned above were also used for ILRBSWT modeling in this study. In order to demonstrate the advantages of the new method when processing missing data, we designed a situation where the geochemical data were missing for the northern part of the study area, as shown in Fig. 4. In that case, grids in region A own values

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

at all of the four independent variables; however, grids in region B only own values at two independent variables, and they have no values in the two geochemical variables.

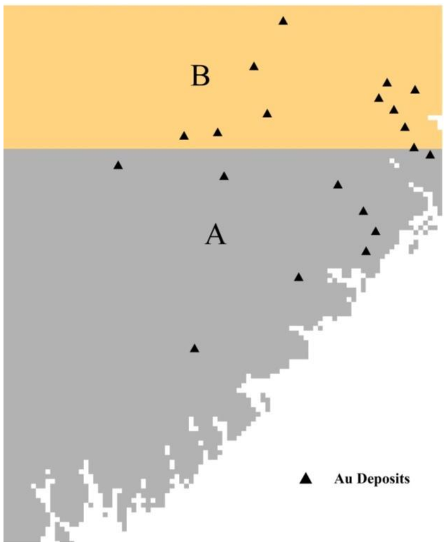


Fig. 4 Study area (A and B) and the scope with missing geochemical data (B).

5.2 Mapping weights for the exploration level

These types of weights can be determined based on prior knowledge according to differences in the exploration data, e.g., different scales may exist throughout the whole study area. They can also be obtained quantitatively. The density of known deposits is a good index for the exploration level, where the degree of research is higher when more deposits are discovered. The exploration level weights for the mapped study area obtained using the kernel density tool provided by the ArcToolbox in ArcGIS 10.2 are shown in Fig. 5.

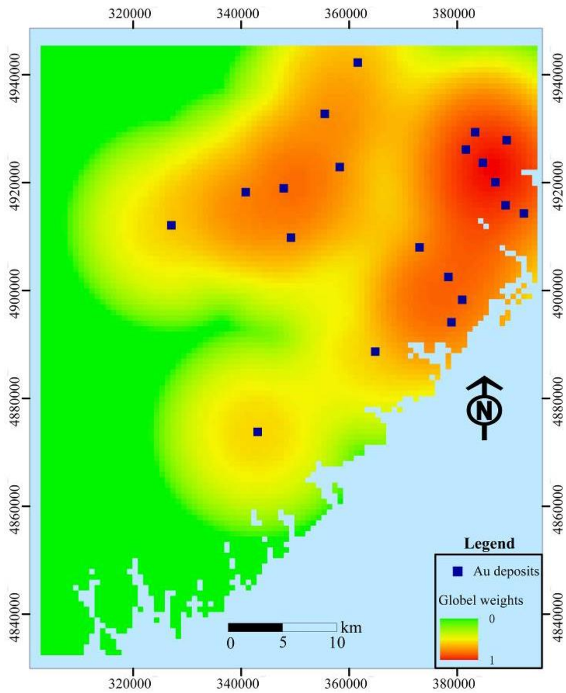


Fig. 5 Exploration level weights.

5.3 Assignment of local window parameters and geographical weights

Empirical and quantitative methods can be used to determine the local window parameters and the attenuation function for geographical weights. The variation function in geostatistics is an effective method for describing the structures and trends of spatial variables, so it was used in this study. In order to calculate the variation function for a dependent variable, it is necessary to first map the posterior probability using the global logistic regression method, before establishing the variation function to determine the local window type and parameters. Variation functions are established in four directions in order to detect anisotropic changes in space. If there are no significant differences among the various directions, a circular local window can be used for ILRBSWT, as shown in Fig. 1; otherwise, an elliptic local window should be used, as shown in Fig. 2. The specific parameters for the local window in the study area were obtained as shown in Fig. 6, and the final local window and geographical weight attenuation were determined as indicated in Fig. 7 (a) and 7(b), respectively.

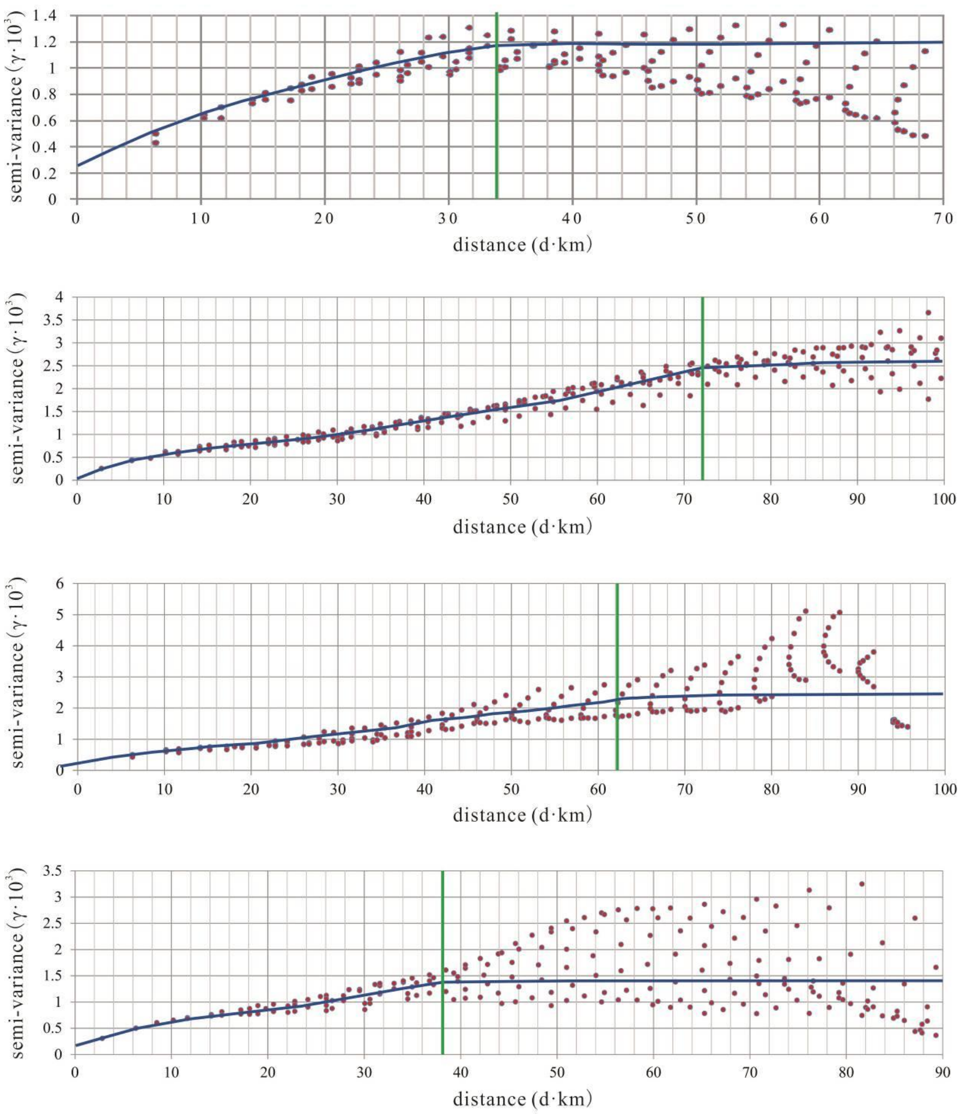
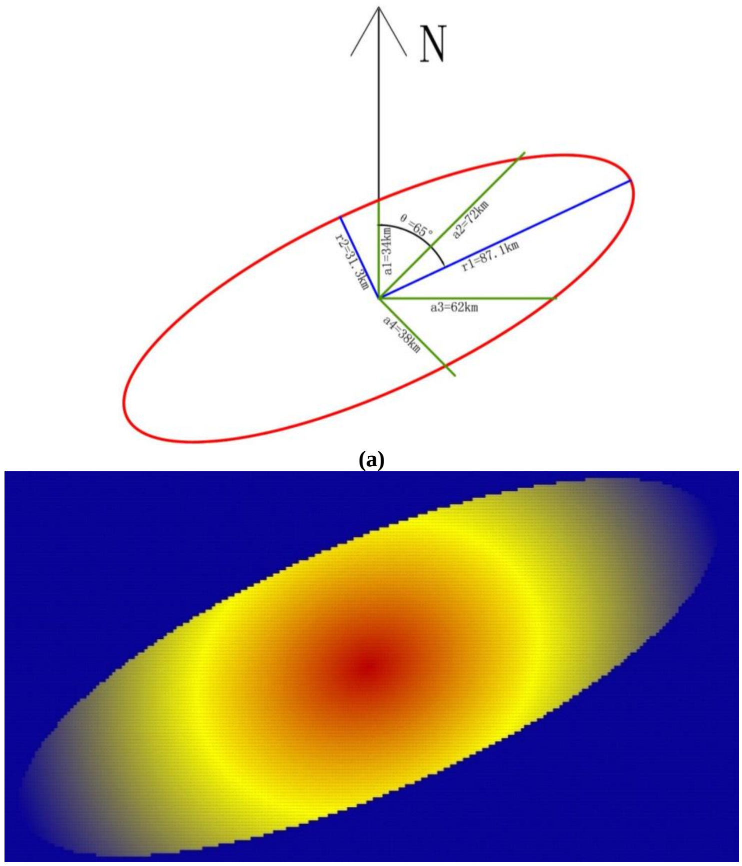


Fig. 6 Experimental variogram fitting in different directions, where the green lines denote the variable ranges determined for azimuths of (a) 0 °, (b) 45 °, (c) 90 °, and (d) 135 °.



(b)

Fig. 7 Nested spherical model for different directions. The green lines in (a) correspond to those in Fig. 6, and (b) shows the geographical weight template determined based on (a).

5.4 Data integration

Using the algorithm described in section 3.2, ILRBSWT was performed for the study area according to the settings in Fig. 3. The estimated probability map obtained for intermediate and felsic igneous rocks by ILRBSWT is shown in Fig. 8 (b), while Fig. 8 (a) presents the results obtained by logistic regression. It can be seen from Fig. 8 that ILRBSWT can better weak the effect of missing data than logistic regression, since the Au deposits in the north part of the study area (where missing data exist) are well felled into the region with relatively higher posterior probability in Fig. 8 (b) than in Fig. 8 (a).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

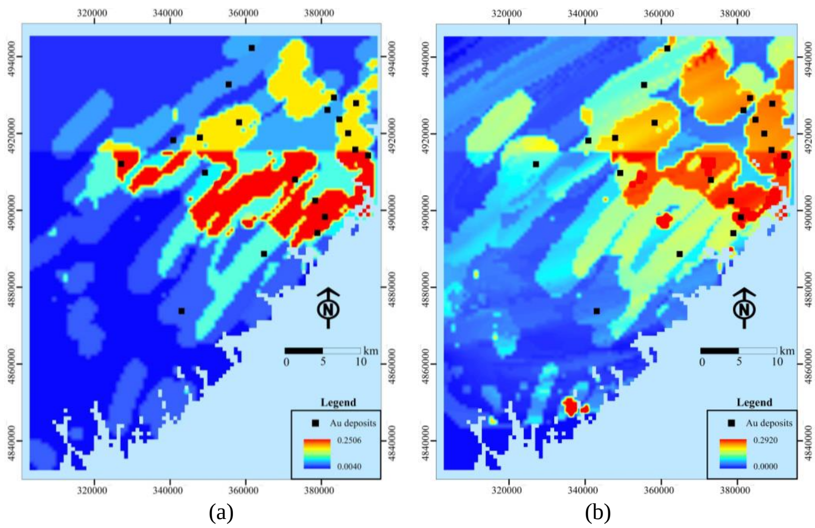


Fig. 8 Posterior probability maps obtained for an Au deposit by (a) logistic regression and (b) ILRBSWT.

5.5 Comparison of the mapping results

In order to evaluate the predictive capacity of the newly developed method and the traditional method, the posterior probability maps obtained by logistic regression and ILRBSWT shown in Fig. 8(b) and 8(a), respectively, were divided into 20 classes by the quantile method and the *t*-values were then calculated using WofE modeling (Fig. 9). Clearly, ILRBSWT performed better because higher *t*-values were obtained, especially when a smaller area was delineated as the target area, which is much more realistic. In the northern part of the study area, the known deposits fitted better to the high posterior probability area shown in Fig. 8(b) than that in Fig. 8(a), which indicates that ILRBSWT can deal with missing data better than logistic regression.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

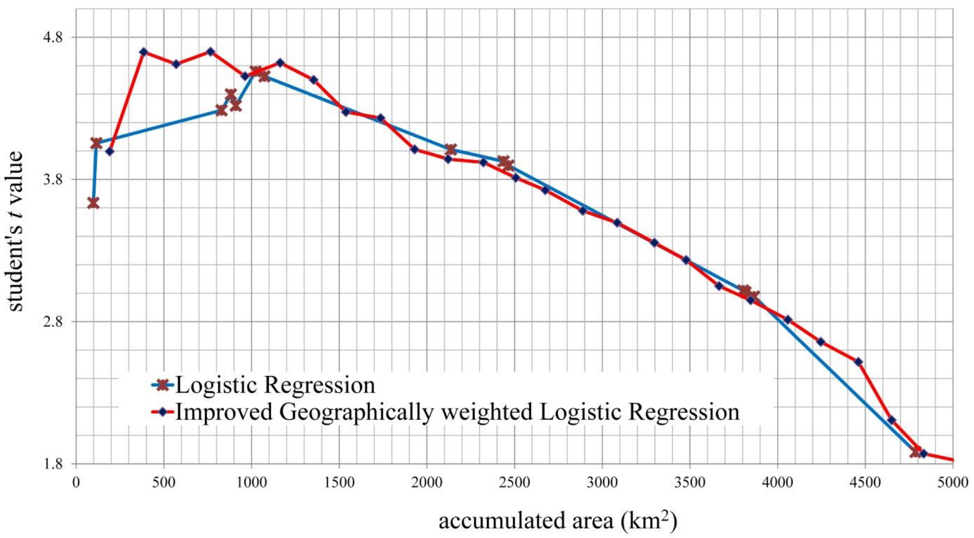


Fig. 9 Student's *t*-values calculated for the spatial correlation between the known Au deposit layer and the predicted posterior probability layers obtained by logistic regression and ILRBSWT at different threshold levels.

6 Conclusions

In this study, we developed an improved GWLR model ILRBSWT based on logistic regression, WofE, and the current GWR model. Furthermore, a software module was developed for ILRBSWT and a case study demonstrated its capacities and advantages. Following objectives were achieved:

(1) A moving window technique is employed for spatial variable–parameter logistic regression, which can overcome or weaken the effect of spatial non-stationarity in MPM and improve the accuracy of mineral prediction.

(2) The variogram model in geostatistics is used to determine the spatial anisotropic parameters and geographical weight attenuation model, which makes the local window parameter design more objective and tenable.

(3) The spatial *t*-statistics method based on WofE is introduced to perform

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions

binarization/discretization for the independent variables in each local window, and the new model can better handle missing data.

(4) The global weight layer in ILRBSWT can reflect differences in the data quality or exploration level well.

Code availability

The software tool ILRBSWT v1.0 in this research is developed by using C#, and the main codes and key functions are prepared in file "Codes & Key Functions". The executable program files are placed in the folder "Executable Programs for ILRBSWT". Please find them in gmd-2017-278-supplement.zip.

Data availability

The data used in this research is sourced from the demo data of GeoDAS software (http://www.yorku.ca/yul/gazette/past/archive/2002/030602/current.htm), and this data is also used by Cheng (2008). All spatial layers used in this work is included in the folder "Original Data" in the format of ASCII file, which can be also found in gmd-2017-278-supplement.zip.

Acknowledgments

This study benefited from joint financial support by the Programs of National Natural Science Foundation of China (Nos. 41602336 and 71503200), China Postdoctoral Science Foundation (Nos. 2016M592840 and 2017T100773), Shaanxi Provincial Natural Science Foundation (No. 2017JQ7010), and the Fundamental Research Funds for the Central Universities (No. 2017RWYB08). The first author thanks former supervisor Drs. Qiuming Cheng and Frits Agterberg for discussions about spatial weights and for providing constructive suggestions.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions

References

Agterberg, F.P., & Cabilio, P., 1969. Two-stage least-squares model for the relationship between mappable geological variables. Journal of the International Association for Mathematical Geology, 1(2), 137-153.

Agterberg, F.P., & Kelly, A.M., 1971. Geomathematical methods for use in prospecting. Canadian Mining Journal, 92(5), 61-72.

Agterberg, F.P., 1964. Methods of trend surface analysis. Colorado School Mines Quart, 59(4), 111-130.

Agterberg, F.P., 1970. Multivariate prediction equations in geology. Journal of the International Association for Mathematical Geology, 1970 (02), 319-324.

Agterberg, F.P., 1971. A probability index for detecting favourable geological environments. Canadian Institute of Mining and Metallurgy, 10, 82-91.

Agterberg, F.P., 1989. Computer Programs for Mineral Exploration. Science, 245, 76 – 81.

Agterberg, F.P., 1992. Combining indicator patterns in weights of evidence modeling for resource evaluation. Nonrenewal Resources, 1(1), 35–50.

Agterberg, F.P., 2011. A Modified WofE Method for Regional Mineral Resource Estimation. Natural Resources Research, 20(2), 95-101.

Agterberg, F.P., Bonham-Carter, G.F., & Wright, D.F., 1990. Statistical Pattern Integration for Mineral Exploration. in Gaál, G., Merriam, D. F., eds. Computer Applications in Resource Estimation Prediction and Assessment of Metals and Petroleum. New York: Pergamon Press: 1-12.

Agterberg, F.P., Bonham-Carter, G.F., Cheng, Q., & Wright, D.F., 1993. Weights of evidence modeling and weighted logistic regression for mineral potential mapping. Computers in geology, 25, 13-32.

Blyth, C.R., 1972. On Simpson's paradox and the sure-thing principle.Journal of the American Statistical Association, 67(338), 364-366.

Bonham-Carter, G.F., Agterberg, F.P., & Wright, D.F., 1988. Integration of Geological Datasets for Gold Exploration in Nova Scotia. Photogrammetric Engineering & Remote Sensing, 54(11), 1585-1592.

Bonham-Carter, G.F., Agterberg, F.P., & Wright, D.F., 1989. Weights of Evidence Modelling: A New Approach to Mapping Mineral Potential. In Agterberg F P and Bonham-Carter G F, eds. Statistical Applications in the Earth Sciences, 171-183.

Brunsdon, C., Fotheringham, A.S., & Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis, 28(4), 281-298.

Casetti, E., 1972. Generating models by the expansion method: applications to geographic research. Geographical Analysis, 4, 81-91.

Cheng, Q., 1997. Fractal/multifractal modeling and spatial analysis, keynote lecture in proceedings of the international mathematical geology association conference, 1, 57-72.

Cheng, Q., 1999. Multifractality and spatial statistics. Computers & Geosciences, 25, 949–961.

Cheng, Q., 2008. Non-Linear Theory and Power-Law Models for Information Integration and Mineral Resources Quantitative Assessments. Mathematical Geosciences, 40(5), 503-532.

Fotheringham, A.S., Brunsdon, C., & Charlton, M.E., 1996. The geography of parameter space: an investigation of spatial non-stationarity. International Journal of Geographical Information Systems, 10, 605-627.

Fotheringham, A.S., Brunsdon, C., & Charlton, M.E., 2002. Geographically Weighted Regression: the analysis of spatially varying relationships, Chichester: Wiley.

Fotheringham, A.S., Charlton, M.E., & Brunsdon, C., 1997. Two techniques for exploring nonstationarity in geographical data. Geographical Systems, 4, 59-82.

Hosmer, D.W., Lemeshow, S, & Sturdivant, R.X., 2013. Applied logistic regression, 3rd edn. Wiley, New York

Nakaya, T., 2016. GWR4.09 user manual. WWW Document. Available online: https://raw.githubusercontent.com/gwrtools/gwr4/master/GWR4manual_409.pdf (accessed on 16 February 2017).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-278
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 16 January 2018

Geoscientific Model Development Discussions
Open Access
EGU

Pearson, K., Lee, A., & Bramley-Moore, L., 1899. Mathematical contributions to the theory of evolution. VI. Genetic (reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred racehorses.Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 192, 257-330.

Reddy, R.K.T., Agterberg, F.P., & Bonham-Carter, G.F., 1991. Application of GIS-based logistic models to base-metal potential mapping in Snow Lake area, Manitoba. Proceedings of the Canadian Conference on GIS, 18-22.

Simpson, E.H., 1951. The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society. Series B (Methodological), 238-241.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography, 46(2), 234-24.

Xiao, F., Chen, J., Hou, W., Wang, Z., Zhou, Y., & Erten, O., 2017. A spatially weighted singularity mapping method applied to identify epithermal Ag and Pb-Zn polymetallic mineralization associated geochemical anomaly in Northwest Zhejiang, China. Journal of Geochemical Exploration.

Yule, G.U., 1903. Notes on the theory of association of attributes in statistics. Biometrika, 2(2), 121-134.

Zhang, D., 2015. Spatially Weighted technology for Logistic regression and its Application in Mineral Prospectivity Mapping (Dissertation). China University of Geosciences, Wuhan (in Chinese with English abstract).

Zhang, D., Cheng, Q., & Agterberg, F.P., 2017. Application of spatially weighted technology for mapping intermediate and felsic igneous rocks in fujian province, china. Journal of Geochemical Exploration, 178, 55-66.

Zhang, D., Cheng, Q., Agterberg, F.P., & Chen, Z., 2016. An improved solution of local window parameters setting for local singularity analysis based on excel vba batch processing technology. Computers & Geosciences, 88(C), 54-66.

Zuo, R., Carranza, E.J.M., & Wang, J., 2016. Spatial analysis and visualization of exploration geochemical data. Earth-Science Reviews,158, 9-18.