Spatial Data Mining for Prediction of Unobserved Zinc Pollutants Using Various Kriging Methods

Durga pujitha. K*, Fathimabi. SK, JayaLakshmi. G, Suneetha.M

Department of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Kanuru, Vijayawada-520007, India

Journal of Advanced Environmental Research and Technology

Vol. 1, No.3
page 57-67 ,summer 2023
***
Received 20 December 2023
Accepted 13 January 2024

Abstract

Following years of contamination, rivers may experience significant levels of heavy metal pollution. Our research aims to pinpoint hazardous areas in these rivers. In our specific case, we focus on the floodplains of the Meuse River contaminated with zinc (Zn). Elevated zinc concentrations can lead to various health issues, including anemia, rashes, vomiting, and stomach cramping. However, due to limited sample data on zinc concentrations in the Meuse River, it becomes imperative to generate missing data in unidentified regions. This study employs universal Kriging in spatial data mining to investigate and predict unknown zinc pollutants. The semivariogram serves as a valuable tool for illustrating the variability pattern of zinc. To predict concentrations in unknown regions, the model captured is interpolated using the Kriging method. Employing regression with geographic weighting allows us to observe how stimulus-response relationships change spatially. Various semivariogram models, such as Matern, exponential, and linear, are utilized in our work. Additionally, we introduce Universal Kriging and geographically weighted regression. Experimental findings indicate that: (i) the Matern model, determined by calculating the minimum error sum of squares, is the most suitable theoretical semivariogram model; and (ii) the accuracy of predictions is visually demonstrated by projecting results onto a real map.

key words

spatial data mining

missing data

semivariogram

Universal Kriging

Geographically weighted Regression

*To whom correspondence should be addressed:
durgapujitha135@gmail.com

[DOI: 10.22034/jaert.1.3.57]

1. Introduction

"Spatial data mining" is the process of iden- tify- ing interesting and undiscovered patterns in spa- tial data. Spatial data mining [12], is the application of data mining techniques [5] to spatial data. Extracting meaningful and interesting patterns from spatial datasets is more challenging than extracting corresponding patterns from traditional numeric and categorical data due to their complexity. Research on spatial data min- ing has advanced significantly as a result of the blending of disciplines. [6] Geostatistics is a multidisciplinary field. [8] study that focuses on the spatial relationships between data and geology. It is applied in numerous disciplines, including geology, forestry, agriculture, and geography. [10,1] One of the main tools in geostatistics is Kriging, [4] an interpolation technique originally used to forecast mineral reserves. [11] The unobserved locations were filled in using the prediction results, and the gaps in particular areas were filled in by interpolating the available data. [2] Though its original application was in geostatistics, kriging is a general statistical interpolation technique that finds use in numerous other areas, including climatology [9] and education [12].

In 1962, Matheron introduced Kriging, also referred to as spatial Best Linear Unbiased Prediction (BLUP), as a tribute to D. G. Krige, a mining engi- neer from South Africa. Kriging, as it turns out, is an interpolation method that provides an objective linear estimate of the values of a point or block. With extensive use of kriging, the ongoing surface (i.e., estimation at each location in the study area) of related variables has been developed. There are different types of Kriging, depending on the sta- tionarity assumption and the stochastic properties of the random variables. Universal kriging (UK) is a spatial interpolation method that combines a deterministic model with a stochastic model. [7] It's a variant of ordinary kriging under non-stationary conditions. It is often used on data with a significant spatial trend, such as a sloping surface. It relaxes the assumption of stationarity by allowing the mean of the values to differ in a deterministic way in different locations for example Meuse River floodplain. Kriging can be easily applied in scenarios where obtaining a

spatial datum proves to be expensive because of the small sample size (n). Zinc (Zn) is one of the primary metals that contaminate the floodplain of the Meuse River. As such, identifying the location of the zinc-containing region is essential. However, the Meuse River's zinc concentration is only partially known, necessitating the generation of the missing data in unidentified regions. The Meuse River floodplain data needs to be applied with gstat and sp library in GStat-R to get a prediction index of pollutants in unobserved locations during the prediction computation using the Universal Kriging method. The pollutant prediction index of GStat-R has a minimum calculated average Kriging variance, which contributes to its accuracy. Additionally, it can show contours so that GStat-R can show the concentration and location of pollutants.

The remainder of this paper is organized as follows: The next section will show the works done by previous researchers. Section 3, will outline Kriging methods and discuss Universal Kriging and Geographically weighted regression in detail. In section 4, we implement Universal Kriging and GWR to zinc pollutants in the Meuse River dataset. Experiment results which show the results of measurement and visualize it on a meuse map are presented in this section. Finally, the last section presents the main conclusions of this work.

2. Literature Survey

Spatial analysis encompasses a variety of statis- tical and geographic information systems (GIS) methodologies, with Kriging standing out as a fundamental technique in this field. Kriging en- ables the prediction of spatial patterns, such as the distribution of zinc, a prevalent contaminant in the Meuse River floodplain. In a study by [1], ordinary point kriging coupled with Gaussian, Exponential, and Spherical semi-variograms was proposed to predict undiscovered zinc pollutants. Their approach aimed to interpolate and forecast the presence of zinc through spatial data analysis. Conversely, another kriging technique, Co-Kriging, was discussed in the work of [2]. This method predicts values at unobserved locations by considering spatially interconnected sample points and incorporating additional variables correlated with the primary variable. Co-Kriging is valuable not only for single variable predictions but also for simultaneous predictions involving multiple variables.

[3] proposed a machine learning-based approach utilizing the spatial features of coordinate information for spatial estimation. Their method, employing Random Forest (RF) among other machine

[DOI: 10.22034/jaert.1.3.57]

learning algorithms, exhibited superior performance compared to previous machine learning approaches and comparable results to Kriging. They emphasized the significance of re- trieved features with distinct spatial categorization properties, suggesting improved efficiency in spa- tial estimation. Furthermore,[4] delved into the interpretability of predictors in spatial data science, examining the conditions that lead to accurate statistics when modeling with such predictors. Their study also investigated the possibility of establishing an information horizon for scale and information content. In a related context, [5] introduced several spatial analysis techniques, encompassing Inverse Distance Weighting (IDW),Nearest Neighbor Inverse Distance Weighting (NNIDW), spline interpolation, and various types of Kriging. They applied these techniques to derive terrain measurements, emphasizing their significance in spatial component analysis.

In summary, these studies highlight diverse methodologies such as kriging variations, machine learning-based approaches, and the significance of different spatial analysis techniques for predicting and understanding spatial patterns. Whereas, we proposed different spatial analysis techniques to predict the unknown regions of zinc on a meuse map by using kriging methods and regression analysis (GWR).

3. Methods

To use Kriging or optimal prediction techniques, we must ascertain the spatial correlation's structure. This problem is known as the structural analysis problem in geostatistics, and it becomes important in the ensuing Kriging procedure. The accuracy of Kriging is determined by the functions that provide information about the found spatial correlation. Semivariograms must meet specific criteria to be classified as such. Typically, semivariograms are derived from observed datasets but may not inherently satisfy all the necessary criteria. Therefore, it becomes essential to fit them to one of the theoretical models that adhere to these criteria. Once a theoretical semivariogram is chosen, the next step involves employing Kriging techniques for spatial prediction. Additionally, we use a method known as geographically weighted regression shown in Fig. 1.

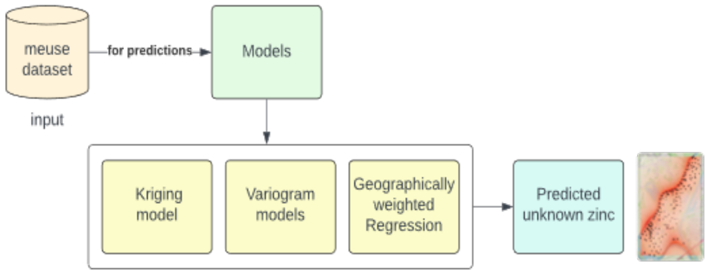


Fig. 1 Methodology

3.1 Theoretical Semivariogram

The semivariogram displays the spatial autocorrelation of the measured sample points. After the locations are plotted, a model is fitted through each pair of locations. These models are frequently defined in terms of a handful of particular characteristics. It quantifies how the variance between data points changes as a function of distance or lag between them.

The semivariogram at distance *h* h is defined as:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \qquad (1)$$

Where,

$\hat{\gamma}(h)$: semivariance at distance h

N(h) : number of pairs of points separated by distance h

$Z(x_i+h)$ and $Z(x_i)$: values of the variable of interest at points x_i+h and x_i,respectively

In general, an experimental semivariogram is not isotropic. Consistency across all orientations is known as isometropy. When generated from the observed dataset, it does not satisfy these requirements. The most well-known semivariograms with isotropic functions are exponential, linear, and matern models. These concepts Fitting the experiment's semivariogram to the models requires the application of three parameters, as indicated by the models: sill (c), range (a), and nugget (λ) in Fig. 2 below.

A thorough description of every semivariogram model with all required properties fully filed is

[DOI: 10.22034/jaert.1.3.57]

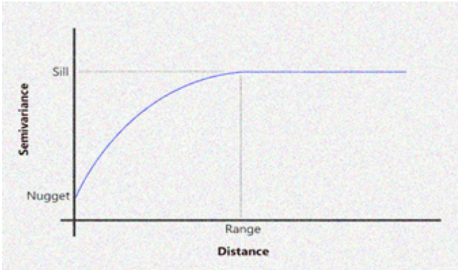


Fig. 2 Theoretical Semivariogram

given below :

3.1.1 Exponential Model

The spherical and exponential models share similarities in how spatial variability gradually approaches the sill. In both models, spatial dependence is marked by the semivariance increasing exponentially and asymptotically approaching the sill as distance increases. This behavior indicates that the models exhibit continuous but non-differentiable characteristics at the origin. The asymptotic approach to zero further characterizes these models, contributing to their representation of spatial dependence.

$$\gamma(h)=C.(1-\exp(-h/a))+\lambda \quad (2)$$

Where,

Sill (c) : represents the variance of the variable

range (a) : signifies the distance at which spatial correlation is significant

Nugget (λ) : represents the variance at very short distances or measurement error

3.1.2 Linear Model

The linear variogram model describes spatial dependence resulting in a linear increase in semivariance with distance. It's the simplest type of model without a plateau, meaning that the user has to arbitrarily select the sill and range.

$$\gamma(h)=C.h+\lambda \quad (3)$$

Note : c : sill, λ : nugget

3.1.3 Matern Model

The Matérn variogram model is a generalization of several theoretical variogram models. It incorporates a smoothness parameter and controls continuity with a shape parameter. The shape parameter must be larger than zero. The Matérn covariance function is named after the Swedish forestry statistician Bertil Matérn. It specifies the covariance between two measurements as a function of the distance between the points at which they are taken.

$$\gamma(h) = C.\left(1-\frac{1}{2^{v-1}}\cdot\frac{\Gamma(v)}{\Gamma\left(v+\frac{1}{2}\right)}\cdot\left(\frac{\sqrt{2v}h}{a}\right)^{v}.K_{v}\left(\frac{\sqrt{2v}h}{a}\right)\right)+\lambda \quad (4)$$

Note: c: sill, a : range, λ : nugget, v : Smoothness parameter, it dictates the smoothness of the transition between the nugget, partial sill, and the range.

3.2 Universal Kriging

Kriging stands out among various methods that utilize a limited sample of data points to estimate the value of a variable across a continuous spatial field. This approach is particularly useful when dealing with variables that exhibit spatial variation across a random field, such as the average monthly concentration of ozone over a city or the availability of healthy foods across neighborhoods. Unlike simpler methods like Gaussian decays, Linear Regression, and Inverse Distance Weighted Interpolation, Kriging leverages the spatial correlation between sampled points to interpolate values in the spatial field. What sets Kriging apart is its reliance on the spatial arrangement of empirical observations rather than an assumed model of spatial distribution. This approach allows for a more flexible and data-driven interpolation. Additionally, Kriging provides estimates of the uncertainty associated with each interpolated value, offering valuable insights into the reliability of the predictions made across the spatial field.

The different Kriging techniques have different levels of complexity and underlying assumptions. Universal kriging relaxes the assumption of stationarity by permitting the mean of the values to vary deterministically in various locations (for instance, as a result of a spatial trend). The only thing that stays constant across the field is the variance. This second-order stationarity (also known as "weak stationarity") is often a relevant assumption when taking environmental exposures into account. It involves incorporating a deterministic

[DOI: 10.22034/jaert.1.3.57]

trend or spatially varying mean into the Kriging prediction model, in addition to the spatial autocorrelation modeled through the semivariogram. For example, for zinc concentration data in the Meuse dataset, we can generate spatial predictions that integrate both the inherent spatial structure (modeled through the semivariogram) and any identifiable trend.

Under the assumptions: Universal Kriging can be expressed as a combination of the deterministic trend and the Kriging predictor. Z(u)=μ(u)+ε(u) Where Z(u) is the estimated value at the unsampled location u. The deterministic trend component μ(u) can take various functional forms and the kriging residual ε(u) is obtained by applying the kriging weights to the observed values and can be expressed as:

$$\varepsilon(u) = \sum_{i=1}^{n} \lambda_i(u).\,[Z(u_i) - u(u_i)] \qquad (5)$$

Where,

λ_i (u) : represents the kriging weights assigned to the sampled locations based on their spatial relationships with the prediction location u

$Z(u_i)$: denotes the observed value at location u_i

$u(u_i)$: value of the deterministic trend at location u_i

Universal Kriging involves estimating both the parameters of the deterministic trend and the kriging weights, ensuring that the prediction model accounts for both the systematic trend and the spatial autocorrelation in the dataset.

3.3 Geographically weighted regression

Geographically Weighted Regression (GWR) is an analytical technique designed for spatial point data, facilitating the interpolation of missing values within the dataset. This method recognizes that the direction and strength of the relationship between a dependent variable and its predictors may vary due to contextual factors. In essence, GWR produces individual Ordinary Least Squares (OLS) equations that incorporate the dependent and explanatory variables of locations within the bandwidth of each target location for every point in the dataset. The user has the flexibility to manually specify the bandwidth. By estimating regression parameters locally for different locations within a study area, GWR effectively captures the spatial heterogeneity in relationships. This allows for a more nuanced understanding of how the association between variables evolves across the spatial domain, acknowledging the impact of local context on the relationships under consideration.

$$Yi=\beta 0i+\beta 1iX1i+\beta 2iX2i+...+\beta piXpi+\varepsilon i \qquad (6)$$

Where,

Yi : dependent variable at location i

X1i,X2i,…,Xpi : independent variables at location i

The coefficients β0i,β1i,β2i,…,βpi are estimated locally at each location, capturing the spatially varying relationships between the dependent and independent variables and εi represents the error term or residual at location i . GWR allows for the examination of spatially varying relationships and provides insights into the spatial heterogeneity of the studied phenomena, making it valuable for spatial analysis, prediction, and understanding local variations in relationships between variables across a geographical area.

4. Implementation

4.1 Load Meuse dataset

Four heavy metals that were measured in the top soil of a flood plain along the Meuse River created the Meuse data set. According to the dispersed heavy metal distribution process, the contaminated sediment is deposited primarily in low-lying areas and along riverbanks, where it is carried by the

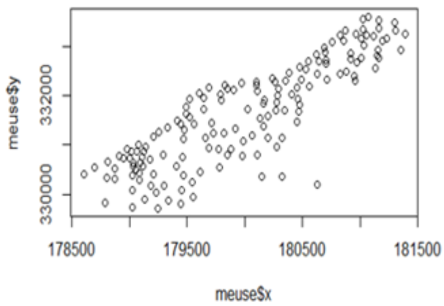


Fig. 3 Zinc concentrations on 155 samples in the flood plains near the Meuse River

[DOI: 10.22034/jaert.1.3.57]

river. The samples were taken near Stein village on the Meuse River's floodplain. In addition to several soil and landscape variables, the point data set includes 155 samples with the highest concentrations of soil heavy metals (ppm) shown in Fig. 3.

The R programming language was used to create the Gstat-R library, which is used for one-, two-, or three-dimensional geostatistical modeling, prediction, and simulation. Gstat-R provides a wide range of spatial prediction options, from simple to universal Kriging. It changed from being a standalone program to an R library package with more features for controlling the processing of geographic data for geostatistical applications with updates in 2004. Thus, in this work, we apply ordinary point Kriging using the GStat-R and sp package.

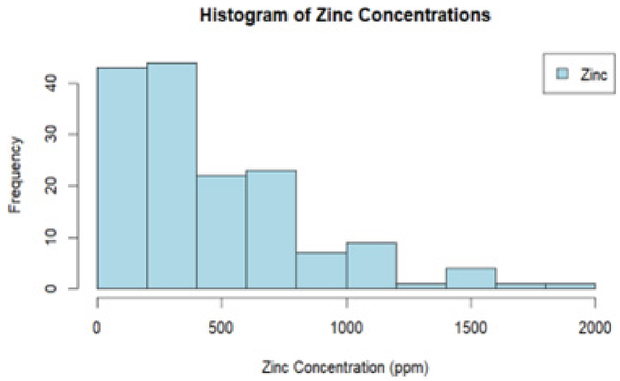


Fig. 4 Histogram of zinc concentration

4.2. Histogram Analysis

To see the distribution of each variable, create a histogram. Evaluate the distribution's shape to use universal kriging to predict the data. Fig. 4 shows a zinc concentration histogram plot.

4.3 Semivariogram models

The resulting theoretical semivariogram models depend on the selection of three parameters: sill (c), range (a), and nugget (λ). From the plot, we can infer that nugget=0, the range is between 300 and 700, and still is between 0.5 and 1. The best theoretical semivariogram models are then identified by fitting semivariogram models into experimental semivariograms based on these three parameters. In semivariogram model fitting, three models are used: linear, exponential, and matern. These fittings are shown in Figs. 5-7 below. Moreover, the best model can be determined using the smallest possible sum of squares.

The best semivariogram model with a minimum sum of square error (SSE) is the maternal model, highlighted in Table 1 below.

Using Universal Kriging and the Gstat-R package, we predict 3103 location points after choosing a theoretical semivariogram. Table 2 below shows the Universal Kriging output summary:

Fig. 8(a)-(b)) shows a contour map of the expected zinc concentration and the plot of the standard error of variance, due to the usage of the ggplot function in the R programming language.

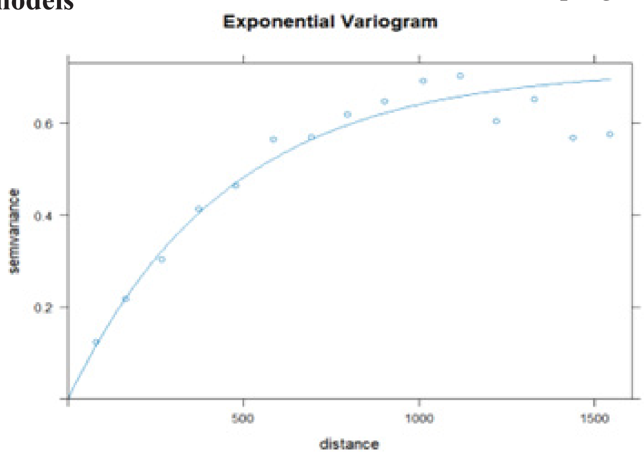


Fig 5. Exponential semivariogram model fitting

[DOI: 10.22034/jaert.1.3.57]

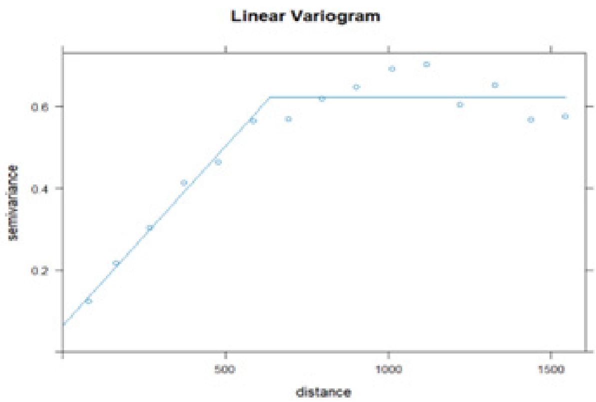


Fig. 6 Linear Semivariogram model fitting

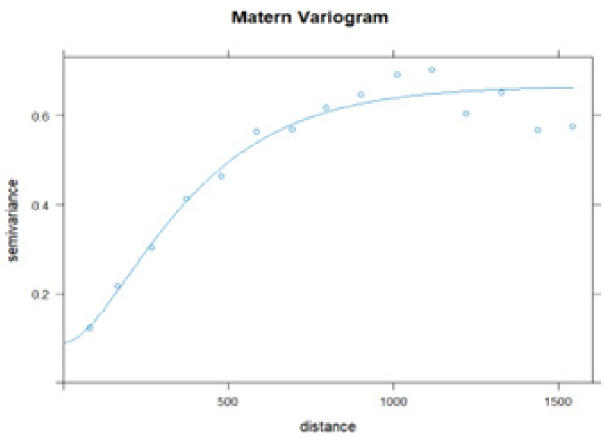


Fig. 7 Matern Semivariogram model fitting

Table 1 Minimum sum of squared error of the semivariogram models

Exponential	Linear	Matern
1.628328e-05	1.494981e-05	1.093181e-05

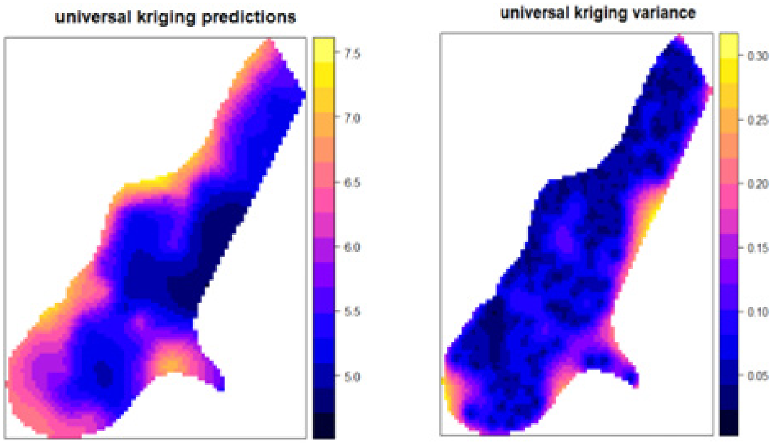
Table 2 Output Summary of Universal Kriging Predictions

S.No	Coordinates (x,y)	Predicted_values	Kriging variance
1	(181180,333740)	6.590687	0.18124505
2	(181140,333700)	6.688392	0.13251337
3	(181180,333700)	6.566027	0.14607826
4	(181220,333700)	6.443574	0.16155048
5	(181100,333660)	6.796248	0.08294027
: 3103	: :	: :	: :

Furthermore, examining the outcomes on the map as demonstrated in Fig. 9 enables one to confirm that accuracy of the Universal Kriging forecasts. Since there is no groundtruth in the verification data, this method is used.

4.4 Geographically weighted Regression analysis(GWR)

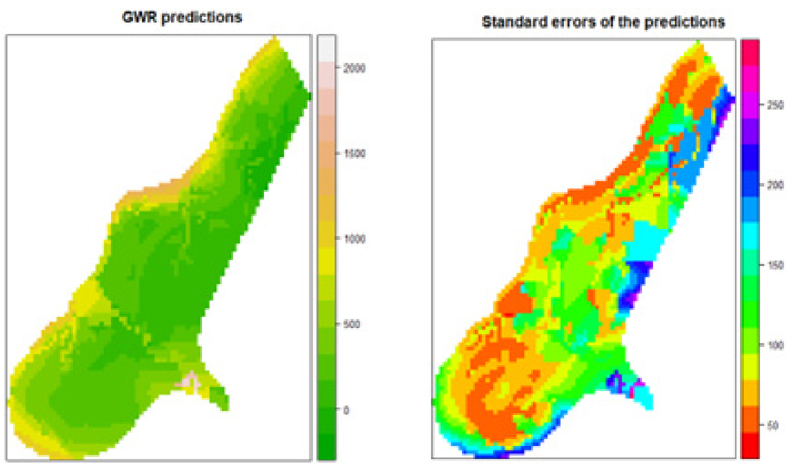
GWR can be performed with the spgwr package in R. Fig. 10 displays a contour map of the expected zinc concentration and a plot of the standard error of variance, which illustrates the degree of



(a) (b)

Fig. 8 Predictions of Universal Kriging (a) and Standard error of Universal Kriging (b)

Fig. 9 Map projection that uses colors to show areas with high and low zinc concentrations



(a) (b)

Fig. 10 The predictions of zinc by using GWR (a) and Standard errors of the GWR (b)

[DOI: 10.22034/jaert.1.3.57]

uncertainty and variability in the estimated zinc values across various study area locations as a result of using the ggplot function.

In a Geographically Weighted Regression (GWR) analysis, a comprehensive set of information is available, including a global model summary that encapsulates traditional regression results across the entire dataset. However, the distinctive feature of GWR lies in its local model statistics, which dissects the analysis into multiple localized models, each tailored to a specific geographic area.

to determine the Matern model, a theoretical semi-variogram model that closely aligns with the data according to experimental findings.

(ii) The results of the Kriging analysis can be projected onto a map, allowing for visual verification of the accuracy of predictions.

Moreover, the analysis demonstrated an accuracy rate of 86% using Geographically Weighted Regression (GWR). This success prompts consideration of the extent to which other fields, such as climatology, epidemiology, and education, should adopt the stationarity assumption inherent in Universal Kriging and similar techniques—an area that warrants further investigation.

```
Call:
gwr(formula = zinc ~ dist + soil + ffreq, data = meuse, bandwidth = 228,
    hatmatrix = TRUE)
Kernel function: gwr.Gauss
Fixed bandwidth: 228
Summary of GWR coefficient estimates at data points:
                  Min.    1st Qu.    Median   3rd Qu.      Max.   Global
X.Intercept.   490.972   754.490   859.657  1008.461  1731.987  853.460
dist         -4375.419 -1906.443 -1082.595  -574.289   184.974 -797.015
soil2         -516.161  -196.595   -40.672    89.968   777.370 -223.578
soil3         -391.662    21.245   128.138   617.601  2256.339   44.439
ffreq2       -1231.949  -361.349  -280.554  -205.268   -20.238 -275.986
ffreq3       -1731.295  -288.127  -195.038  -116.427   663.299 -296.978
Number of data points: 155
Effective number of parameters (residual: 2traceS - traceS'S): 59.32127
Effective degrees of freedom (residual: 2traceS - traceS'S): 95.67873
Sigma (residual: 2traceS - traceS'S): 171.8668
Effective number of parameters (model: traceS): 46.7098
Effective degrees of freedom (model: traceS): 108.2902
Sigma (model: traceS): 161.5494
Sigma (ML): 135.0311
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 2099.725
AIC (GWR p. 96, eq. 4.22): 2007.287
Residual sum of squares: 2826179
Quasi-global R2: 0.8638016
```

Fig. 11 Output summary of Geographically weighted regression

Furthermore, the GWR model for predicting zinc concentrations utilizes a Gaussian kernel function with a 228-unit bandwidth. This approach generates coefficient maps that visually depict the spatial variation of each variable's effect across the study area. The analysis includes local residuals and p-values, offering insights into the model's performance and significance at specific locations. Additionally, local parameter estimates provide a detailed view of the spatial variation in the regression coefficients, enhancing our understanding of how these coefficients change across different geographic regions. The model accuracy is shown as 86% of the R-squared value in Fig. 11 below.

5. Conclusion

Kriging proves to be a practical and effective method for filling gaps in missing spatial information, providing minimal errors and significant predictions for unobserved locations based on nearby observed locations. Two key points highlight the effectiveness of this approach:

(i) The minimal error sum of squares is employed

Data availability statement

The data can be provided upon reasonable request from the corresponding author.

Authors' Contribution

Suneetha. M conceived of the presented idea. Durga pujitha. K developed the theory and performed the computations, verified the analytical methods. Fathimabi. SK and JayaLakshmi. G encouraged Durga pujitha. K to investigate and supervised the findings of this work. Durga pujitha. K wrote the manuscript. All authors reviewed the results and approved the final version of the manuscript.

[DOI: 10.22034/jaert.1.3.57]

References

[1] Gunawan, A. A., Falah, A. N., Faruk, A., Lutero, D. S., Ruchjana, B. N., & Abdullah, A. S. (2016, October). "Spatial data mining for predicting of unobserved zinc pollutant using ordinary point Kriging". In 2016 International Workshop on Big Data and Information Security (IWBIS), IEEE, 83-88. https://doi.org/10.1109/IWBIS.2016.7872894.

[2] Falah, A. N., Hamid, N., Rusyaman, E., Abdullah, A. S., & Ruchjana, B. N. (2021). "Implementation of Ordinary Co-Kriging method for prediction of coal quality variable at unobserved locations". In Journal of Physics: Conference Series (Vol. 1722, No. 1, p. 012076). IOP Publishing.

[3] Ahn, S., Ryu, D. W., & Lee, S. (2020). "A machine learning-based approach for spatial estimation using the spatial features of coordinate information". ISPRS International Journal of Geo-Information, Mdpi, 9(10), 587. https://doi.org/10.3390/ijgi9100587

[4] Behrens, T., & Viscarra Rossel, R. A. (2020). "On the interpretability of predictors in spatial data science: The information horizon". Scientific Reports, 10(1), 16737. https://doi.org/10.1038/s41598-020-73773-y

[5] Paramasivam, C. R., & Venkatramanan, S. (2019). "An introduction to various spatial analysis techniques". GIS and geostatistical techniques for groundwater science, Elsevier, 23-30. https://doi.org/10.1016/B978-0-12-815413-7.00003-1

[6] Jiang, Z. (2018). "A survey on spatial prediction methods". IEEE Transactions on Knowledge and Data Engineering, 31(9), 1645-1664. https://doi.org/10.1109/tkde.2018.2866809

[7] Falah, A. N., Abdullah, A. S., Parmikanti, K., & Ruchjana, B. N. (2017, March). "Prediction of cadmium pollutant with ordinary point kriging method using GStat-R". In AIP Conference Proceedings (Vol. 1827, No. 1). AIP Publishing.

[8] Hussain, M. R. (2016). "An Overview of Geographic Information System (GIS)", Researchgate. https://doi.org/10.13140/RG.2.1.3569.5603

[9] Choudhury, N. H., Rahman, A., & Ferdousi, S. (2015). "Kriging infill of missing data and temporal analysis of rainfall in North Central region of Bangladesh. J. Climatol. Weather Forecast", Researchgate, vol. 3, 1-5. https://doi.org/10.4172/2332-2594.1000141

[10] Cressie, N. (2015). Statistics for spatial data. John Wiley & Sons.

[11] Montero, J. M., Fernández-Avilés, G., & Mateu, J. (2015). "Spatial and spatio-temporal geostatistical modeling and kriging". John Wiley & Sons.

[12] Setiawan, A., & Rosadi, R. (2011). "Spasial Data Mining menggunakan Model SAR-Kriging". IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 5(3), 52-61. https://doi.org/10.22146/ijccs.5213

[13] Shekhar, S., Evans, M. R., Kang, J. M., & Mohan, P. (2011). "Identifying patterns in spatial information: A survey of methods". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(3), 193-214. https://doi.org/10.1002/widm.25

[14] Goovaerts, P. (2009). "AUTO-IK: a 2D indicator kriging program for the automated non-parametric modeling of local uncertainty in earth sciences". Computers & geosciences, Elsevier, 35(6),1255-1270. https://doi.org/10.1016/j.cageo.2008.08.014

[15] Grimm, R., Behrens, T., Märker, M., & Elsenbeer, H. (2008). "Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis". Geoderma, 146(1-2), 102-113.

[16] Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). "The elements of statistical learning: data mining, inference, and prediction" (Vol. 2, pp. 1-758). New York: springer.

Oliver,M.A.,&Webster,R.(2007)."Kriging:a method of interpolation for geographical information systems". International Journal of Geographical Information System, 4(3), 313-332. https://doi.org/10.1080/02693799008941549

[17] Hengl, T., Heuvelink, G. B., & Stein, A. (2004). "A generic framework for spatial prediction of soil variables based on regression-kriging". Geoderma, 120(1-2), 75-93.

[18] Brivio, P. A., Colombo, R., Maggi, M., & Tomasoni, R. (2002). "Integration of remote sensing data and GIS for accurate mapping of

[DOI: 10.22034/jaert.1.3.57]

flooded areas". International Journal of Remote Sensing, 23(3), 429-441.

[19] Middelkoop, H. (2000). "Heavy-metal pollution of the river Rhine and Meuse floodplains in the Netherlands". Netherlands journal of geosciences, 79(4), 411-427. https://doi.org/10.1017/S0016774600021910.