Journal of Integrity in Ecosystems and Environment
Volume: 3 Issue: 4 Year: 2025
ISSN: 3030-3729 | http://journals.proindex.uz

Using Kriging Method for Mapping Heavy Metal Pollution in Fallujah: Spatial Statistical Approach to Environmental Risk Assessment

Salih, Sufian Munther
Department of Management Economics Banking, College of Bussiness Economics, AlNahrain University, Baghdad, Iraq, sufian.m.salih@nahrainuniv.edu.iq

Ghufran Khalil Joad
University of Technology, Department of production and Metallurgy, ghufran.k.jwad@uotechnology.edu.iq

Abstract: Kriging is an important statistical method used in spatial studies within regional planning and predictive studies. It is employed to achieve optimal predictions, such as pollutant quantities or various environmental and health impacts. There are many types of pollutants, including:

1. Water Pollution: The leakage of chemicals from mining sites into groundwater and water bodies can affect water quality and vitality.
2. Ecosystem Destruction: Mining can destroy local plants and animals, affecting ecological balance and leading to biodiversity loss.
3. Soil Pollution: Mining can contaminate soil with harmful chemicals, affecting agriculture and plant growth.
4. Geological Changes: Mining can lead to changes in groundwater levels, as well as seismic impacts and land subsidence.

This research focused on one type of pollutant, such as copper dust, zinc dust, and iron dust. Each of these metals can have different environmental and health impacts.

In this study, Kriging was used to estimate the pollution caused by copper dust accumulated in specific areas within the city of Fallujah, based on data from 23 locations within the city. These locations were mapped on the city map.

The Kriging method was used to predict pollution levels in 6 sites based on their spatial distribution and not previously measured within the city. Errors in these predictions were calculated, along with the overall estimate of pollution and the confidence limits for the average pollution in the city. The mathematical methodology in this study relied on calculating the variogram function by selecting a central location in the city (CBD) and studying the distribution around it. After obtaining the variogram function, the covariance function was calculated assuming stationarity, which was discussed in this study. The covariance function was then used in the Kriging equations to obtain prediction values, and the results were very encouraging. All calculations were performed using Matlab. These results highlight the utility of spatial statistical methods in monitoring environmental pollution and provide valuable insights for local environmental policy and public health protection.

Keyword: Kriging, Statistical method, spatial studies, Regional planning, Predictivestudies, Optimal predictions, Pollutant quantities, Environmental impacts, Health impacts, Types of pollutants

1. Introduction

This research addresses the prediction of unmeasured random locations based on the measured regionalized variables using the Kriging method named after D.G. Krige. This method is represented by the variogram function or the autocovariance function of the spatial random variables.

Spatial statistics have evolved significantly through the work of scholars such as Matheron (1975), Journel and Huijbregts (1978), Stein (1987), and Cressie (1991).

The field of spatial statistics is continuously developing due to its importance. Spatial statistics involve mathematical statistical concepts used in regional planning studies to describe the correlation of randomly distributed spatial variables.

2. Basic Concepts

Consider D as a domain or region of the random variable z(x), x ∈ D ∈ R^p where 3 or 2 = P. This variable can be measured on a sample of size n of locations. These measurements are symbolized by the variable (x)z and its values are:

$z(x_1), z(x_2) \dots \dots \dots \dots z(x_n)$

Each z(x_i) represents the pollution value at location x_i

Because these measurements are distributed in space, z(x_i) is a regionalized variable in analytical geostatistics.

The location variable z (x) is a single observation and a single observation of the measurements is z(x_1),z(x_2)…………z(x_n)) for the random process {z(x),x ∈ D}.

One of the goals of analytical statistics is to estimate pollution at unmeasured locations in Region D.

For any unmeasured point x_0 in region D, the best estimator of the location variable z (x_0) at location x_0 is given by the conditional expectation.

z(x) at x=x_0 where x_0 is a new unobserved point. The best linear unbiased estimator. to z(x) is called the Kriging Estimator and has the following formula:

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i \, z(x_i) \dots \dots \dots \dots (1)$$

The weights $_i$ are estimated so that the mean square error (MSE) is as small as possible. Therefore, the best unbiased linear estimator of the parameter μ is:

$$\hat{\mu} = \frac{J' \Sigma^{-1} z}{J' \Sigma^{-1} zJ} \dots \dots \dots \dots (2)$$

Since J is a vector of dimension n*1 and all its elements are one and Σ is the variance-covariance matrix of dimensions n*n and must be a positive definite matrix.

2- When the mean μ is unknown:

To estimate z_0 by the best linear estimator which is in the form :

$$\hat{z}_0 = \lambda' z \dots \dots \dots \dots (3)$$

We need to meet two conditions:

$E(\hat{z}_0)=\mu$ 1)

2) $\sigma_k^2 = var(\hat{z}_0 - z_0)\ mi$

This variation is called Kriging variance and this variation is not

$\sigma^2_{\hat{z}_0} = var\ (\hat{z}_0)$

Which is called the variance of Crick's estimator. From equations (2) and (3) we can get Crick's estimator which is in the form:

$$\hat{z}_0 = \hat{\mu}\ + \sigma_0' \sum{}^{-1} (z - \hat{\mu}J) \dots \dots \dots \dots (4)$$

From the results of Equation (4), Crick's variance σ^2_k and Crick's estimated variance have the following two formulas :

$$\sigma^2_{\hat{k}} = \sigma_{00} - \sigma_0' \sum{}^{-1}\sigma + \left[\left(1 - J' \sum{}^{-1}\sigma_0\right)^2 / \left(J' \sum{}^{-1} J\right)\right] \dots \dots .. (5)$$

$$\sigma^{\ 2}_{\hat{z}_0} = \sigma_0' \sum{}^{-1}\sigma + \left[\left\{1 - \left(J' \sum{}^{-1}\sigma_0\right)^2\right\} / (J' \sum{}^{-1} J)\right] \dots \dots \dots (6)$$

The calculation of $\ ^2_k$ gives us an idea of the error included in the estimate, which represents a measure of the quality of the estimate and is also called the error variance. We note that σ^2 is the total variance, while $\sigma^2_{\hat{z}_0}$ is the variance of the estimator $\hat{z}_0$. See (1981) Ripley.

We can obtain from the estimates .i=1,2,…….M, $\hat{z}_i(x_0)$ measured by Crick's method the local estimation within the region D and as

$$\overline{z(x)} = \frac{1}{M}\sum_{i=1}^{M} \hat{z}_i(x_0) \dots \dots \dots .. (7)$$

The best linear unbiased estimate of the total amount within D is given by:

Total estimation=A $\overline{z(x)}$…………(8)

Where A represents the area of region D.

The local estimation error is:

$$\sigma^2_{LE} = var\left(\frac{1}{M}\sum_{i=1}^{M} \hat{z}_i(x_0) - z_i(x_0)\right) = \frac{1}{M^2}\sum_{i=1}^{M} var\left(\hat{z}_i(x_0) - z_i(x_0)\right) = \frac{1}{M^2}\sum_{i=1}^{M} \sigma^2_k(x_i)$$

............................(9)

Where M is the number of points at which the estimate was made.

As for the total variance of the error in the total estimate within the region D whose area is A, it is calculated from :

Total variance=var $\left(A * \frac{1}{M}\sum_{i=1}^{M} \hat{z}_i(x_0) - z_i(x_0)\right) = (A)^{\ 2} * \sigma^2_{LE} \dots (10)$

The confidence interval is found for the estimated quantity within D using the standard deviation, known as the standard error, for the total variance. For example, the standard confidence interval (95%) for the estimated quantity within D is given by:

Standard Error=$\sqrt{Total\ variance}$

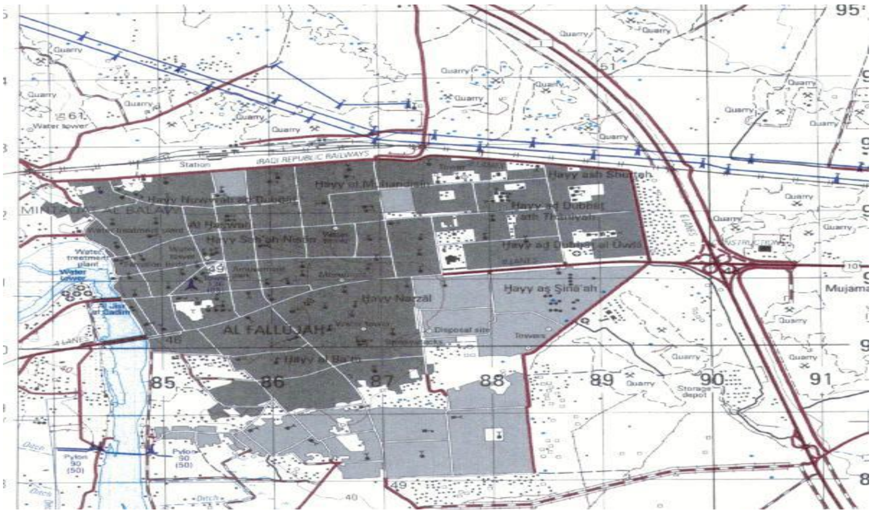
Total Estimation$\pm$1.96 Standard Error … … … … … … (11)

5-The applied aspect:

Since the nature of using Crick's prediction law in the prediction process requires data containing real site values, the data used in this study are real data on air and soil pollution in Fallujah city, which we obtained through the data of the Ministry of Planning. This is to know the pollution in rain, soil, and plant leaves resulting from air pollution with copper and zinc dust, with emphasis on copper due to the availability of information about it, and the extent of its effect on the surface of plant leaves and soil. See Al-Hayali (2000).

These data are shown in Table (1) and distributed on the map of Fallujah city, Figure (1). We note that it contains (23) sites with their calculated coordinates. For four seasons during the year 2020-2024, we randomly chose the spring season as a model for the study, as shown in Table (1), as each site in it represents an observation with its coordinates, which are (x)u representing north-south and (x)v representing east-west. Note that the pollution concentrations measured in this data are in micrograms/gram units. That is, equivalent to 10^{-6} grams. The data we have are in micrograms/gram units.

As we notice from Figure (1) it contains the areas that are required to be predicted, which are (Police District, Officers District, Yarmouk District, Nazzal District, and Golan), and the coordinates of these locations are shown in Table (1).



(Figure (1): Locations of neighborhoods from which dust samples were collected in the city of Fallujah and the locations that need to be predicted)

We note that the distribution of these data follows the normal distribution or is close to the normal distribution because the city of Fallujah is not a large industrial city, but rather an agricultural city much more than an industrial one, despite the presence of several factories such as the block and white cement factory, the foam concrete factory, and the high-ammonia refractory brick factory, as it does not contain factories that require copper and zinc as raw materials in their production. The copper dust present in this city is mainly produced from manufacturing and metal industries, especially in mining areas. For these reasons, we did not test the normal distribution of the data using one of the normal distribution testing methods such as the chi-square test or the Kolmogorov and Smirnov test. See Bernard (2018).

Table (1) Locations of lead dust concentrations in the city for the fall season with their coordinates.

u(x)	v(x)	value z(x)	z(x)	Areas
11.5	8.5	688	z_1	Al-Qadisiyah
10	10	90	z_2	Police
10.5	6.5	22.67	z_3	Officers
13	6.75	880	z_4	Teachers
9	8	200	z_5	Unity
14	9	251.67	z_6	Golan
9.5	11.5	54.33	z_7	Al-Moatasem
8	9.5	241.67	z_8	Republic
15	7	15	z_9	Al-Rasafi
8	10	50	z_{10}	Andalusia
13	5	461.67	z_{11}	message
13	12	1057.67	z_{12}	Nationalization

14.5	5.5	917.67	z_{13}	Green
7	7.5	507	z_{14}	Yarmouk
15	12	52.67	z_{15}	Al Mansour
11.5	3	364	z_{16}	peace
11.5	14	50	z_{17}	Al-Mamun
14	14	60	z_{18}	Secretary
16.5	12	55	z_{19}	Fight
12.5	2.5	90.33	z_{20}	Nationalization
13	14.5	77	z_{21}	Jerusalem
13.5	1.5	65	z_{22}	Staff
6	1	12.33	z_{23}	Victory

Table (2): Locations to be predicted

u(x)	v(x)	Areas
12	9	Officers' Quarter
12	4	Fight
9.5	6.5	Police district
8	8.5	Golan
10.5	10.5	Yarmouk neighborhood
15	6.5	Employees' Quarter

Using this data in Table (1), we calculated the experimental quasi-variogram function between pairs of observations that are offset from each other, and the results are shown in Table (3). See Cresses (1993). After converting the results in Table (1) to grams, we drew the relationship between the quasi-variogram function and the offset, as in Figure (2).

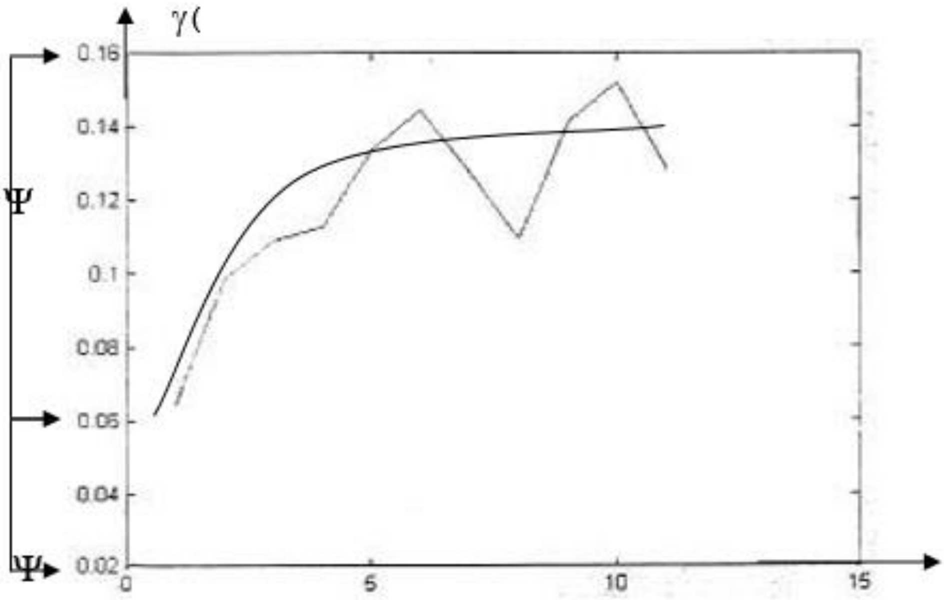


Figure (2) Graph of the semi-variogram function

Table (3): Results of the semi-variogram function

h	(h)
1	638220.9
2	95578.3
3	109010.7
4	112675.8
5	133647.3
6	144747.6
7	127564.6
8	109341.6
9	141346.8
10	151974.7
11	129152.9
12	35541.3
13	46508.5
14	68369.7
15	55240.14
16	77450.4
17	88621.6
18	110308.5
19	139002.6
20	62342.15
21	98540.4
22	228264.9

We notice from Figure (2) that the variogram function is an increasing function in terms of displacement and that it intersects the vertical axis at $\varphi_0 = 0.06$ and its height increases and almost stabilizes at h=0.14, meaning that the range is a=11 and the variogram function stabilizes at (h)$\gamma =$ 0.14, which is its highest height, meaning that $_2 = 0.2$ since $\sigma^2 = \varphi_0 + \psi$, and thus $\psi = 0.14$

The program function in Figure (2) is similar to the spherical variogram function, which has the formula

$$\gamma(h) = \begin{cases} \Psi_0 & h = 0 \\ \Psi_0 + \Psi\left[\frac{3h}{2a} - \frac{1}{2}(\frac{h}{a})^3\right] & 0 < h \le a \\ \Psi_0 & h > a \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

Cressie (1993), Istok and Cooper(1988)

To obtain the covariance function (h) c, we consider the following relationship:

Steven and Cresses (2019)

$$\gamma(h) = c(0) - c(h)$$

or

$$c(h) = c(0) - \gamma(h)$$

$$c(0) = (\sigma)^2 = \Psi 0 + \Psi$$

$$c(h) = \psi 0 + \psi - \psi 0 - \psi\left[\frac{3h}{2a} - \frac{1}{2}(\frac{h}{a})^3\right]$$

If the covariance function has the following form:

$$c(h) = \begin{cases} \Psi_0 + \Psi & h = 0 \\ \Psi\left[1 - \frac{3h}{2a} + \frac{1}{2}(\frac{h}{a})^3\right] & 0 < h \le a \\ 0 & h > a \end{cases} \quad \dots\dots\dots\dots (13)$$

After estimating the parameters a, φ, and φ_0 from the variogram function, the final formula for the covariance function is as follows:

$$c(h) = \begin{cases} 0.2\,, h = 0 \\ 0.14\left[1 - \frac{3h}{2(11)} + \frac{1}{2}(\frac{h}{11})^3\right], 0 < h \le a \\ 0\,, h > a \end{cases} \dots\dots\dots\dots (14)$$

We were able to project the coordinates of the observation sites onto the map, considering the scale of the map of Fallujah city, which is [1:50000] cm, meaning that after [50000 * 11] cm, the presence of pollution will decrease or disappear, which explains to us outside the study area. The second main reason that should be noted in choosing the range a = 11 is that we cannot exceed more than half the furthest length of the distance h because the covariance function is subject to estimation error when the pairs of points are distributed over large distances.

By taking the observed sites and their values in the study area and applying the random process (x) that has the spherical model, we found the Euclidean distance matrix between the measured areas according to:

$$h = \sqrt{(u(x_i) - u(x_j))^2 + (v(x_i) - v(x_j))^2} \dots\dots\dots\dots\dots (15)$$

Using the last matrix and after knowing the range and ψ and ψ, we found the matrix Σ according to formula (13) and its inverse Σ^{-1}. After that, we also calculated the distances between the regions to be predicted and the measured regions according to (15), through which we found the covariance function according to (13). After that, we found the sample weights λ_i from the formula:

$$\lambda = \left[\frac{1 - \sigma_0 \Sigma^{-1} J}{J' \Sigma^{-1} J}\right] \sum^{-1} J + \sum^{-1} J\sigma_0 \dots\dots\dots\dots\dots (16)$$

Al Bayati (2020

For each predicted area, which requires that an $\sum_{i=1}^{n} \lambda_i = 1$ see Table (4), then we calculated $\hat{z}(x_0)$ which represents the Crick estimator according to the relation (1).

As for the Crick variance and the Crick estimator variance ^2_k, we found them from (5), and (6) respectively.

The best-unbiased estimate of the total value was obtained from formula (8) and the total variance from (10) after determining the area of the area D which is equal to 11× 11=121km^2Then we addressed the local estimate and the error in the local estimate, as we found them from (7) and (9) respectively. Finally, we calculated the confidence interval (95%) for the estimated amount of pollution within D from the relation (11). The results of the study are shown in Table (4).

Table (4) The computational results of the prediction process carried out in the six locations of Fallujah city.

	$z(x_a)$	$z(x_b)$	$z(x_c)$	$z(x_d)$	$z(x_e)$	$z(x_f)$
λ_1	0.33	0.01	0.06	0.02	0.11	0.01
λ_2	0.11	0.00	0.02	0.05	0.31	0.00
λ_3	0.05	0.10	0.33	0.05	0.01	0.00
λ_4	0.09	0.06	0.04	0.00	0.01	0.09
λ_5	0.04	0.02	0.21	0.22	0.04	0.00
λ_6	0.16	0.00	0.00	0.00	0.04	0.08
λ_7	0.38	0.00	0.00	0.02	0.20	0.00

λ_8	0.00	0.00	0.03	0.22	0.04	0.00
λ_9	0.03	0.00	0.00	0.00	0.00	0.39
λ 10	0.00	0.00	0.01	0.14	0.05	0.00
λ 11	0.01	0.21	0.04	0.00	0.00	.006
λ 12	0.08	0.00	0.00	0.00	0.09	0.00
λ 13	0.00	0.05	0.00	0.00	0.00	0.25
λ 14	0.00	0.01	0.14	0.00	0.00	0.00
λ 15	0.02	0.00	0.00	0.23	0.01	0.01
λ 16	0.00	0.27	0.07	0.00	0.00	0.00
λ 17	0.00	0.00	0.00	0.00	0.05	0.00
λ 18	0.00	0.00	0.00	0.00	0.00	0.00
λ 19	0.00	0.00	0.10	0.00	0.00	0.04
λ 20	0.00	0.17	0.02	0.00	0.00	0.01
λ 21	0.00	0.00	0.00	0.00	0.01	0.00
λ 22	-0.01	0.06	0.00	0.00	0.00	0.03
λ23	-0.01	0.02	0.05	0.02	0.01	0.00
$\sum_{i=1}^{23} \lambda_i$	1.28	0.98	1.12	0.97	0.96	0.97
$\hat{z}(x_0)$	478.21	349.05	257.67	251.4	267.9	394.82
σ_k^2	0.21242	0.21089	0.20658	0.2155	0.21364	0.21481
$\sigma^2 \hat{z}_0(x)$	1.1952	1.1999	1.188	1.2217	1.2031	1.2299
$\overline{z(x)}$	333.175					
σ_{LE}^2	0.035					
Total estimator	40314.175					
Total variance	512.435					
Confidence interval	(40269.807, 40358.543)					

Conclusions and recommendation

Conclusions:

1. Crick's method can be used to predict pollution such as air pollution by gases and soil and water pollution. This method is better than classical methods because it takes into account the correlation between observations.
2. Geological and spatial statistics are well applied when the data are phased and normally distributed.

Recommendations

1. The application of the Kriging method must be on a specific area between the model values, otherwise it will give us wrong results.
2. **2-**The calculation of the area of Fallujah was approximate, but it requires calculating the pollution area accurately, and this is possible by using a surveyor to do so.

Sources:

1. Al-Bayati, Jaafar Musa Muhammad (2020): Spatial statistics in estimating environmental pollution. Master's thesis (unpublished), University of Mosul.
2. Al-Hayali, Afif Khalil Abdullah Najm (2019): Environmental study of some pollutants in Mosul. Master's thesis (unpublished), University of Mosul.
3. Bernard, W.L. (2020): Statistical Theory. Tenth editions. Collier Macmillan International. London.

4. Clark, I. (2021). Practical geostatistics 3th.Applied Science Publishers, London, U.K.
5. Cresses, N. (2017). Statistics for spatial Data. Wiley, New York.
6. Cresses, N. (2018). Statistics for spatial data, 2nd. Edition. New York, Wiley.
7. David, M. (2000). Geostatistical ore reserve estimation. Elsevier Scientific Publishing Company, New York.
8. Delfiner, P. (2016). Linear estimation of non-stationary spatial phenomena. Advanced geostatistics for the mining industry, M. Guarcascio, M. David, and C. Huijbregts, eds., Reidel publishing company, Dordrecht, Holland, 49-68.
9. Journel, A.G., and Huijbregts, C.I, (2021), mining geostatistics. Academic Press, New York.
10. Matheron, G. (2020): The Theory of Regionalized Variable and its applications. Les Cahiers du center de morphologie mathematique, No. 5, Fontainebleau, France.
11. Matheron, G. (2019). Random sets and Integral Geometry. New York, Wiley.
12. Mohammed, S.A. and Qassim, M.N.I. (2014). An Estimation of components of spatial covariance. J. Edu. Sc. Vol. (27), University of Mosul.
13. Olea, R.A. (2010). Optimum mapping techniques using regionalized variable theory. Kansas Geologic Survey series on spatial Analysis No.2 University of Kansas, Lawrence, Kans.
14. Ripley, B.D. (2007). Spatial statistics, New York, John Wiley.
15. Stein, M.L. (2009). Minimum Norm Quadratic Estimation of Spatial variograms, J. Am. Statist. Assoc. 982, p. 765- 772.
16. Steven, S.C. and Cressie, N. (2012). A comparison of geostatistical methodologies used to estimate snow water equivalent (Water resources bulletin). Journal of the American Water Resources Association Vol. 32, No. 2.