Evaluation the Simple Spatial Interpolation Method to Identifying The Groundwater in Erbil Governorate
(PP 288 - 299)

ID No. 2223
https://doi.org/10.21271/zjhs.22.5.19

Paree khan Aabdulla Omer
Salahaddin University - College of Administration and Economics - Statistics department
Pareekhan.omer@su.edu.krd

Received: 22/05/2018
Accepted: 12/06/2018
Published: 01/11/2018

Abstract
In this research used Simple Kriging method as one of Geostatistics interpolation methods on the measured value of the specific part in Erbil. Geostatistics is a set of (tools and models) that are developed for statistical analysis of any continuous data that can be observed or measured at any location in the space. Verify three data features in statistical continuous data analysis: dependency, stationery and distribution. With these features you can proceed to the modeling of the Geostatistical data analysis like Kriging. Additionally, the goal of this work is to predict a new value at the unmeasured location by two models and compare the results of these two models based on the Simple Kriging method and understanding their spatial variability. The first step is modeling spatial dependency by semivariogram function. The Two types of the semivariogram are emphasized in this work (Exponential and Gaussian) model, and then different fitting models were taken to describe analyzing their influence over the interpolation results. The source of dataset is the observed values of the (550) wells that had been taken from known specific place called Qushtappa- in Erbil Governorate. Results of applying both models show that the predicted value by anisotropy semivariogram model is better than the isotropy semivariogram model depending on the value of the depth of groundwater. Additionally, the values of (RMSE, SME and SE) of each model are compared and the smaller values of them are the better interpolation as shown in analyzing to evaluate the precision of the prediction.

Keywords: Groundwater Surface Interpolation, Gaussian Random field, Assumptions, Semivariogram function, Covariance Function, Simple Kriging interpolation method.

1. Introduction
In general spatial statistics and Geographic information system (GIS) rely on each other's in many ways. Arc GIS is software can be used to create covariates for inclusion in all statistical models and to bring out the results from statistical models. The work in this study might be very important to evaluate the results from different models in simple Kriging interpolation approaches. This kind of comparison presents a relevant meaning for the variability of a physical model which used as reference to validate the interpolation results. GIS software package is a powerful tool to deal with analyzing groundwater and spatial distribution of the sample point data in this place. The (GIS) users can choose an optimal interpolation model to create surface according to their results. The performance of the interpolators were assessed and compared in terms of the root mean square errors (RMSE), standardized error of mean (SME) and standard error (SE) which obtained from the differences between the estimated values of two models and the information of the measured values.

2. Theoretical Background
2.1 Groundwater Surface Interpolation
Groundwater depth of the wells of any sample point data for generates surface need to be evaluated and preprocessed before interpolation. The locations and values of sample point data will impact the interpolation result. First, all the collected data should come from the
same type of wells in the same aquifer. The well information should be carefully evaluated to make sure the data reflect the dynamics of groundwater in the target aquifer, not other aquifers. Second, the spatial distribution of sample point data should be carefully considered. The clustered data and sparse data in one area will cause different interpolation results. After collecting suitable data for interpolation, raster calculator in GIS is used to calculate the elevation of groundwater table in each well by dividing the surface elevation (DEM) with the groundwater depth (Gou, 2010).

2.2 Gaussian Random Field

Used Gaussian Random Field \( Z(s) \) to identify the spatial correlation structure, spatial field is a set of random variables parameterized by some set \( D \subset \mathbb{R}^d \). The simplest stochastic process form is as follows:

\[
\{Z(s): s \in D \subset \mathbb{R}^d\} \quad \ldots (1)
\]

Where:

- \( Z(s) \): random spatial field.
- \( s \): spatial coordinates.
- \( D \): spatial domain.
- \( \mathbb{R}^d \): d-dimensional Euclidean space.

Any finite collection \( \{Z(s_1), Z(s_2), \ldots, Z(s_k)\} \) is multivariate normal:

\[
\begin{bmatrix}
Z(s_1) \\
\vdots \\
Z(s_k)
\end{bmatrix}
\sim N\left(
\begin{bmatrix}
\mu(s_1) \\
\vdots \\
\mu(s_k)
\end{bmatrix},
\begin{bmatrix}
\text{Cov}(Z(s_1), Z(s_1)) & \cdots & \text{Cov}(Z(s_1), Z(s_k)) \\
\vdots & \ddots & \vdots \\
\text{Cov}(Z(s_k), Z(s_1)) & \cdots & \text{Cov}(Z(s_k), Z(s_k))
\end{bmatrix}
\right)
\]

A random field is called second order stationary if the following assumptions hold (Webster and Olivier, 2007):

\[
\begin{align*}
E(Z(s)) &= \mu \quad \forall s \in D \quad \ldots (2) \\
\text{Cov}(Z(s_1), Z(s_2)) &= E[(Z(s_1) - \mu) (Z(s_2) - \mu)] = C(s_1 - s_2) \quad \forall s_1, s_2 \in D \quad \ldots (3)
\end{align*}
\]

Where:

- \( E \): The expected value. \( \text{Cov} \): Covariance function.
- \( h = s_2 - s_1 \): The vector distance between \( Z(s_1) \) and \( Z(s_2) \).

The covariance function can be expressed as follows:

\[
\text{Cov}(Z(s_1), Z(s_2)) = \text{Cov}[Z(s_1), Z(s_1 + h)] = C(h)
\]

\[
= E[(Z(s_1) - \mu)(Z(s_1 + h) - \mu)] = E[Z(s_1)Z(s_1 + h)] - \mu^2 \quad \ldots (4)
\]

When the expected value is equal to zero.

\[
\mu = E(Z(s)) = 0, \quad \forall s \in D \quad \ldots (5)
\]

Then,

\[
\text{Cov}(Z(s_1), Z(s_2)) = \text{Cov}[Z(s_1), Z(s_1 + h)] = C(h) = E[Z(s_1)Z(s_1 + h)] \quad \ldots (6)
\]

When the mean of a second order stationary of spatial random field is zero over the \( D \) and the covariance function not depend on location \( s_1 \) and \( s_2 \) but on the separation vector \( h \).

And if \( \text{Cov}(0) = \text{Cov}[Z(s), Z(s + h)] = V[Z(s)] \) for \( h = 0 \)
If the covariance function or variance does not exist, another hypothesis is introduced. The intrinsic hypothesis and the spatial random field are called intrinsic stationary if the following assumption holds:

\[ E(Z(s)) = \mu \quad \text{or} \quad E[Z(s_1) - Z(s_2)] = 0 \quad \forall s \in D \]
\[ V[Z(s_1) - Z(s_2)] = 2\gamma(s_1 - s_2) \quad \forall s_1, s_2 \in D \quad \ldots (7) \]

Where, 

\( V \): Variance,  \( \gamma \): semivariogram,  \( 2\gamma \): variogram

\[ 2\gamma = V[Z(s_1) - Z(s_2)] = V[Z(s_1) - Z(s_1 + h)] \]
\[ = E[(Z(s_1 + h) - Z(s_1))^2] - (E[Z(s_1 + h)] - E[Z(s_1)])^2 \quad \ldots (8) \]
\[ = E[(Z(s_1 + h) - Z(s_1))^2] \]

Additionally, if the covariance function \( C(s_1 - s_2) = C(h) \) or semivariogram \( \gamma(s_1 - s_2) = \gamma(h) \) depends only on separation distance between \( s_1 \) and \( s_2 \), \( h = \lVert s_1 - s_2 \rVert \), then the spatial random field is called isotropic (Marcin & Marek, 2010).

2.3 Assumptions
In some of interpolation methods especially Geostatistical methods, they have their own assumptions (Sluiter, 2008) as follow:

1. Stationarity.
2. Intrinsic hypothesis.
3. Isotropy is opposed anisotropy.
4. Unbiased.

In many cases, the assumption of isotropy is acceptable, but as we gradually refine our understanding of Earth structure and composition, the need to consider the effects of anisotropy becomes more important, because anisotropy can be visualized easily as an ellipsoid.

2.4 Semivariogram function and Covariance Function
The semivariogram is a structure function of intrinsically stationary spatial random field which describes a broader class of phenomena. In the case of the second order stationary spatial processes there is an equivalence between covariance function and semivariogram as follows:

\[ V[Z(s + h) - Z(s)] = V[Z(s)] + V[Z(s + h)] - 2Cov[Z(s), Z(s + h)] \]
\[ = 2V[Z(s)] - 2Cov[Z(s), Z(s + h)] \quad \ldots (9) \]
\[ = 2[C(0) - C(h)] = 2\gamma(h) \]
\[ \gamma(h) = C(0) - C(h) \quad \ldots (10) \]

The semivariogram is a measure of dissimilarity between pairs of observed value \( Z(s + h) \) and \( Z(s) \). There are three parameters of semivariogram for the second order stationary processes exist, nugget effect \((c_0)\), partial sill \((c)\) and range \((a)\) which are shown in Figure-1-. The sum \((c_0 + c)\) is called sill (Marcin & Marek, 2010), (Sluiter, 2009).
Figure-1:- The relation between the \( C(h) \) and \( \gamma(h) \) for second order stationary of spatial random field

2.4.1 The Exponential Semivariogram Model
The Exponential model, like the Spherical model, relies on two main parameters, the range and the sill. In addition, there may also be a nugget effect (Tesar, 2011). The equation for this model is given by:

\[
\gamma_z(h) = C_0 + C \left[ 1 - \exp\left( -\frac{h}{a_0} \right) \right], \quad h > 0 \quad \ldots (11)
\]

Figure-2:- The exponential semivariogram model with parameters.

2.4.2 The Gaussian Semivariogram model
The Gaussian model is commonly used to represent events with a small scale spatial structure (Clark & Harper, 2000), the equation for this model is similar to the Normal cumulative distribution function and it is given by:

\[
\gamma_z(h) = C_0 + C \left[ 1 - \exp\left( -\frac{h^2}{a_0^2} \right) \right], \quad h > 0 \quad \ldots (12)
\]

Figure-3:- the Gaussian semivariogram model with parameters

2.5 The Kriging
Kriging is a spatial interpolation Geostatistical method used for the first time in meteorology, geology, environmental sciences, agriculture, and others fields. This method is used to find the best estimator under the assumption of the second order stationarity. A geological process
may not be stationary in reality. In case where the process is non-stationary, we could use non-linear functions (Sarma, 2009). Here the classifying Geostatistical techniques as follows:

| Stationary            | Non-stationary       |
|-----------------------|----------------------|
| Linear                | Ordinary/simple      |
|                       | kriging              |
| Non-linear            | Disjunctive kriging   |
|                       | Simulation           |

Table-1: Classification of Geostatistical techniques

2.6 The Simple kriging interpolation method

Simple Kriging can deliver the value at any unmeasured location \((s_0)\) by using a linear estimator \(\lambda_i\) for the measured values at locations \((s_1,s_2,...,s_n)\):

\[
Z(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) , \quad \sum_{i=1}^{n} \lambda_i = 1 \quad (13)
\]

As the sum of all linear estimators equals one, the unbiased prediction is ensured. Using the Intrinsic Hypothesis, the estimation variance is calculated by the formula:

\[
\sigma^2(s_0) = \text{Var}(Z(s_0)) = E[(\hat{Z}(s_0) - Z(s_0))^2] \quad (14)
\]

\[
= 2\sum_{i=1}^{n} \lambda_i \gamma(s_i - s_0) - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(s_i - s_j)
\]

The semivariance \((s_j - s_0)\) and \((s_i - s_0)\) are taken from the variogram model. The goal is to minimize \(\sigma^2(s)\) under the unbiased conditions and to find the corresponding weight.

The weights are chosen to fulfil: \(E(\hat{Z}(s_0) - Z(s_0)) = 0\) accordring the assumption that \(E[Z(s_1) - Z(s_2)] = 0\) which called intrinsic stationary and to solve this the lagrange multiplier \((\mu)\) is introduced, not easy solving this model by any othre ways:

\[
Z(s_0) = \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_j) + \mu(s_0) = \gamma(s_j - s_0) \quad (15)
\]

The solustions to this linear equation system are the values of the linear estimators. This system is called kriging system (Disse and Amin, 2017). The kriging variance is determined by:

\[
\sigma^2_k = \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_0) + \mu(s_0) \quad (16)
\]

3. Application, Results and Discussion

3.1 The Study Area

Spatial variability of any regional variable is a result of complex processes which is working at the same time and over long periods of time. Variation of a regional variable has never been an easy task or work. Many regional random variables vary not only horizontally but also with depth such as wells of water, oil and gas. Additionally, they not only continuously but also abruptly. The study area is Qushtappa in Erbil.
3.2 Data sources
The source of dataset is the observed values of the (550) wells that had been taken from the known specific place which called Qushtappa-township in Erbil Governorate. These observations are collected by GPS by the ministry of agriculture. Figure-5 bellow expresses the location of each well and each point of dataset has its name and properties. Also each point has its goal for drilling the well, some of them for (drinking or Irrigation and Agriculture). Figure-5 shows the Geographical location of the data points of Qushtappa and all wells. They are shown as a surface which can be represented by the most probably prediction map and also by estimated prediction when we are applying two models in the simple kriging interpolation. The figure shows cycles that are many points closer together tend to be more alike than things that are farther apart (quantified here as spatial autocorrelation).

3.3 Results of Groundwater Surface Interpolation
These dataset can be described as continuous data and represented the random field. Figure - 6 shows histogram of the observed values and how they distributed in the region. It is Supposed to follow normal distribution (Gaussion process), but random process $Z(s)$ of data set is appeared non normal, therefore we applied (log) transformation to the data that can assist to make it approximately normal. It represents the curve of the observations values and the results after taken the suitable transformation that have small standard deviation equal to (0.61132), Skewness near to zero equal to (-0.43357) and Kurtosis near to (3) which is equal to (1.8977). These indicated that the distribution of levels of depth of Groundwater wells data were approximately normal.
3.4 Results of Simple kriging surface interpolation

3.4.1 The Exponential semivariogram Surfaces Interpolation

Spatial dependency can be detected in this dataset by using several tools which available in Geostatistical Analysis Exploratory Spatial Data Analysis (ESDA) and Geostatistical wizard in GIS software. In Geostatistics semivariogram called spatial modeling (called structural analysis or variography). In figures below two of the Exponential semivariograms are created, stationarity of autocorrelation can be examined and quantified. In the isotropy Exponential semivariogram will be have similarly in any direction, as it should for stationary data see figure (8, a.), but In the anisotropy Exponential semivariogram for the data set has change in continuity with direction which is described an ellipsoid see figure (8, b.)

In these two figures above we could describe the sample points in the area by focusing on the depth of wells. In the two dimensions figure (left-panel) there is only one well higher than other wells which its depth equal to (504)m, the purpose of drilling this higher well is because it was drilled for drinking.

If we look at the trend analysis provides a three dimension map which perspective of the data, it's very easy to observe that there is only one well among all wells deeper than the rest.
Figure-8: spatial dependency of the Exponential semivariogram model

Figure above of $\gamma(h)$ against $(h)$ called the semivariogram, this gives a quantitative description of variability about the regional variation and each small red circle represents a pair of points of random field. The important part of the variogram is the (rang) which describes the distance. Figure -8-(a) shows the isotropy of exponential semivariogram model for (554) wells of ground water of in Qushtappa, which all points have equal directional of variability as being north-south and east-west. Figure-8-(b) expresses anisotropy for the same data (554) where data point changes with direction which described an ellipsoid, and this ellipsoid specified by the length of two orthogonal axis (Major and Minor) with its orientation Angle $\theta$.

Table 2: Results of the Exponential semivariogram model

| Simple Kriging Exponential semivariogram | Isotropy | Anisotropy |
|------------------------------------------|----------|------------|
| Nugget ($C_0$)                           | 0        | 0          |
| Rang ($a$)                                | 0.453125 | 0.507617   |
| Major rang                               | 345.4818 | 345.4818   |
| Minor rang                               | 345.4818 | 170.9134   |
| Partial sill ($C$)                        | 0.394127 | 0.381967   |
| Lag size ($h$)                            | 28.79015 | 28.79015   |
| Number of lag                            | 12       | 12         |

Table (2) shows the height exponential semivariogram reaches the sill, which composed of two parts: nugget effect and partial sill, the three parameters of the exponential semivariogram model for second order stationary processes $Z(s_0)$ exist for both of isotropy and anisotropy with different values of parameters. These values of parameters show us the location of the line of Paris points in the specific area.
Figure 9: Surfaces of the Exponential semivariogram model
Figure (9) shows exponential semivariogram model for isotropy and anisotropy surfaces, these used to predict of random field (spatial field) of groundwater well $Z(s_0)$ unknown value in the same area, which depends only on maximum (5) neighbors value of the measured values. Table-3- contains all cross validation of isotropy and anisotropy semivariogram model and the prediction value of unknown value of the depth of well.

Table-3: Results of simple kriging Exponential semivariogram model

| Simple kriging | Isotropy | Anisotropy |
|----------------|----------|------------|
| **Exponential semivariogram** | | |
| Maximum neighbors | 5 | 5 |
| Minimum neighbors | 2 | 2 |
| Predicted value | | |
| X | 417328 | 417328 |
| Y | 397694 | 397694 |
| Depth (m) | 167.98 | 159.2723 |
| Prediction Errors | | |
| Samples | 554 | 554 |
| Mean | 1.640963 | 0.380092 |
| Root-mean-square error(RMSE) | 66.01085 | 65.59939 |
| Mean standardized error (SME) | 0.0008743 | 0.000121 |
| standard error (SE) | 91.11196 | 89.24593 |

Table (3) created to show all information about using Simple Kriging by Exponential semivariogram model to predict a new location of Groundwater in the same area for both of isotropy and anisotropy. Depending on the maximum value of neighbors (5) and at least only (2) neighbors value of the observed value from the area could predict new value of unknown measured of the depth of water well. This new predict value in the anisotropy exponential semivariogram model is equal to (159.2723) with longitude (Y) and latitudes (X) and its smaller than the depth of isotropy exponential semivariogram model which equal to (167.98) depends on all values of the three measurements in the table above.

3.4.2 The Gaussian semivariogram Surfaces Interpolation
Figure-10:- spatial dependency of the Gaussian semivariogram model

Figure 10 shows $\gamma(h)$ against $h$ another model called the Gaussian semivariogram. This plot gives a quantitative description of variability about the same regional variation and each small red circle represents a pair of points of random field. The important part of the variogram is the (rang) which describes the distance. Figure-10-(a) shows the isotropy of Gaussian semivariogram model for (554) data of ground water of wells in Qushtappa, which all points have equal directional of variability as being north-south and east-west. Figure -10-(b) expresses anisotropy for the same data (554) that changes with direction which described an ellipsoid, and this ellipsoid specified by the length of two orthogonal axis(Major and Minor) with its orientation Angle $\theta$.

Table-4:- Results of Gaussian semivariogram model

| Simple Kriging Gaussian semivariogram | Isotropy | Anisotropy |
|--------------------------------------|----------|------------|
| Nugget = $C_0$                       | 0        | 0          |
| Rang ($a$)                           | 1.367188 | 1.105273   |
| Major rang                           | 44.19868 | 82.30691   |
| Minor rang                           | 44.19868 | 38.38868   |
| Partial sill = $C$                   | 0.341554 | 0.341191   |
| Lag size                             | 26.71247 | 26.71247   |
| Number of lag                        | 12       | 12         |

Table-4- shows the height Gaussian semivariogram model reaches the sill, which also composed of two parts: nugget effect and partial sill, the three parameters of the Gaussian semivariogram model for second order stationary processes $Z(s_0)$ exist for both of isotropy and anisotropy with different values; it could be easy to observe it as it discussed in the table -2- section before.

Figure-11- : Surfaces of the Gaussian semivariogram model

Figure -11- shows Gaussian semvariogram model for isotropy and anisotropy surfaces, which also used to predict of random field (spatial field) of groundwater well of $Z(s_0)$unknown value in the same area, where it depends on only maximum (5) neighbors value of the measured value. Table -5- contains all results of both types (isotropy and anisotropy). The prediction value of unknown value of the new depth of well.

Table-5:- Results of Simple Kriging Gaussian semivariogram mode

| Simple kriging Gaussian semivariogram | Isotropy | Anisotropy |
|--------------------------------------|----------|------------|
| Maximum neighbors                    | 5        | 5          |
Table-5- shows all information using Simple Kriging by another model called Guassian semivariogram model to predict a new location of Groundwater in the same area for both of isotropy and anisotropy. The prediction value of unknown measured also depending on (5) maximum and (2) minimum value of neighbors from measured value for the depth of new well with longitude and latitude. in anisotropy exponential semivariogram model is equal to (156.559) and its little smaller than the depth of isotropy exponential semivariogram model which equal to (156.5875) depends on all values of the three measurements in table above.

4. Conclusions:

The major results of performing the analysis of two models of semivariogram (Exponential and Gaussian) by using Simple Kriging, the following main conclusions have been achieved:

1. The data follow approximately normal distribution after taken log transformation for variability and second order stationary to remove the trend.

2. The variability of the depth of Groundwater wells elevation in the region changes from north to south or from west-north to south-east, this is due to the spatial dependency in the area which is one of the reason of the depth of well .

3. For unmeasured value used only one spatial interpolation method- Simple Kriging with two semivariogram models (Exponential and Gaussian), which is in both models the predicted value by anisotropy semivariogram model is better than the isotropy semivarogram model depending on the value of the depth of groundwater and the values of (RMSE, SME and SE), which are regarded as good measures for comparing two or more than two models in Geostatistical methods to evaluate the precision of the prediction.

5. Recommendation

It is necessary to use these methods to charge a lot of losses in cost, time and efforts. Additionally, taken in to account the results of simple kriging for predicting the new depth of water well location and for other energy industry.
References:

1. Abdula, R. (2010): petroleum source Rock Analysis of the Jurassic Sargelu formation, Northern Iraq, thesis submitted to the Faculty and Board of Trustees of Colorado School, Master of Science (Geology).
2. Cressie, N. A. (1993): Statistics for Spatial Data, Second Edition, John Wiley and Sons., New York, USA.
3. Clark, I. & Harper, W. V. (2000): Practical geostatistics. Columbus, Ohio: Ecosse North America.
4. Disse, M. and Amin, K. (2017): Review of various methods for interpolation of rainfall and their applications in hydrology, thesis of Chair of Hydrology and River Basin Management, Technical University Munich.
5. Gou, Si (2010): Identifying Groundwater Dependent Ecosystems in the Edwards Aquifer Area, Zachery Department of Civil Engineering, and Texas A&M University.
6. Marcin, L. & Marek, K. (2010): Simple spatial prediction – least squares prediction, simple kriging and conditional expectation of normal vector, department of Geomatics AGH University of Science and Technology in Krakow, Poland.
7. Sarma, D. (2009): Geostatistics with Application in earth science, Second Edition, Springer, Capiatal publishing company.
8. Sluiter, R. (2009): interpolation methods for climate data; literature review, KNMI, R&D information and publishing company.
9. Disse, M. and Amin, K. (2017): Review of various methods for interpolation of rainfall and their applications in hydrology, thesis of Chair of Hydrology and River Basin Management, Technical University Munich.
10. Webster, R. and Olivier, M. (2007), Geostatistics Environmental Scientists, Second Edition, John Wiley and Sons, New York, USA.