Comparison of Geostatistical Methods for Interpolation Groundwater Level  
(Case study: Lake Urmia Basin)  

Keivan Khalili  
Department of water engineering, Urmia University, Urmia, Iran  

ABSTRACT  
Interpolation is one of the most important techniques which have been used in zoning studies. In order to obtain the continuous and integrated maps and predict unknown values, the interpolation methods have been used. Since measuring groundwater levels is very difficult and costly, therefore water table maps are used. The main purpose of this research is to compare two geostatistical methods including Kriging and inverse distance weighted methods for interpolating ground water table in the west, south and south west of Lake Urmia basin. In this study, collected data in the 634 piezometric wells were used to interpolate the groundwater level through water year 2010. Stationary and independence of data were tasted using Wald and Wolfowitz tests. Groundwater level maps were drawn using Kriging and Inverse Distance Weighted (IDW) methods. Results showed that Kriging method with the higher correlation coefficient between measured and modeled data had better results than Inverse Distance Weighted method.  
KEYWORDS: Interpolation, Groundwater Level, Geostatistical, Kriging, Inverse Distance Weighted.  

INTRODUCTION  
Geostatistics is defined as application of all statistical methods such as classical and spatial statistics in the earth science studies. Geostatistics is a technology for estimating the local values of properties that vary in space (Oliver and Webster, 1991). The theory is based upon the concept of a random variable, which expresses a continuous variable depending on a location (Isaake and Srivastava, 1989). It is expected that observations close together in space will be more alike than those further apart (Stein et al., 1997). Geostatistical techniques are useful in providing estimates of sample attributes at locations with sparse information (Burrough and McDonnell, 1998). These methods work by defining the spatial structure of the phenomena by autocorrelation such as semi-variograms, then estimating values between measured points based on the degree of spatial autocorrelation or covariance found in the data (Robertson, 1987). Consequently, simpler alternatives to kriging, such as the Inverse Distance Weighting (IDW) have been used as interpolation methods. This technique is easier to implement due to the fact that the estimation of values does not require any measure of either spatial autocorrelation or spatial auto-covariance (Isaake and Srivastava, 1989). The Kriging techniques are called the best linear unbiased estimator since it tries to have a mean residual error equal to zero (Clark, 1979). It aims to minimizing the variance of the errors and hence is a strong advantage over other estimation methods like inverse distance weighting or moving average. Kriging forms weights from surrounding measured values to predict values at unmeasured locations. As with IDW interpolation, the closest measured values usually have the most influence. However, the kriging weights for the surrounding measured points are more sophisticated than those of IDW. IDW uses a simple algorithm based on distance, but kriging weights come from a semivariogram that was developed by viewing the spatial structure of the data. At first this method has been applied in the estimation of gold storage of South African mines 80 years ago (Nazeri-Tahruudi et al, 2013). Next studies for developing geostatistical methods done by Matheron (1960) and caused to invent a new branch of science called geostatistics. In the classical statistics, samples of population are mainly considered random and measured value could not present any information about the same value in the other sample. While in geostatistical methods, location of a sample is also considered in addition to the quantity of them. So the both quantity and location of samples can be analyzed together. The relationship between quantities of samples depends on the distance of samples from each other. This relation is the basis of geostatistics. Igúzquiza (1998) has been used Thiessen, ordinary Kriging, co-Kriging and Kriging along with trend methods to estimate average rainfall of Horse Pit River catchment in the northern of Spain and results showed that Kriging along with trend method gave better results than other methods. Prudhomme et al. (1999) demonstrated that the use of topographic data improves accuracy of predicting rainfall in ordinary Kriging method. Drogue et al. (2002) in order to produce spatial precipitation maps in the mountainous of northeastern of France using Pluvia simulation, compared multivariate linear relationship between rainfall parameters and the morphological parameters, multivariate geostatistical methods such as Kriging and co-Kriging. In their study 150 stations with 20 year data in the area of 30,000 km$^2$ used to producing and comparing rainfall maps. Diodato and Ceccarelli (2005) used multivariate geostatistical methods to prepare spatial precipitation maps in the Sanyo Mountain located in the northern of Italy. The aim of this study was to investigate the role of digital simulation to improve...
the interpolation process for preparing the sub basins mean annual rainfall and monthly rainfall maps from 40 years data in the region with an area of 1400 km². In that study, in addition to simple linear regression between height and precipitation, Kriging and inverse distance weighting methods was also applied. Evaluations showed that the method of inverse distance weighting and ordinary Kriging had more error than the Kriging and linear regression methods. Cheng et al. (2008) in order to estimate average regional rainfall and rainfall in the regions without station, used geostatistical methods. Variogram analysis showed that hourly rainfall indicates higher spatial variations than annual precipitation. Fathi-Marj et al. (2009) used the methods of interpolating underground water level concluded general Kriging, simple Kriging, SP line and inverse distance weighted, and selected an acceptable model in geographical information system for Fars Province. Latinopoulos (2006) interpolated the ground water table of Antimontas in Greece by Kriging method and CV criterion showed that this method was acceptable. Examples on the use of GIS and kriging techniques for groundwater quality assessment are found in the literature (Zhang et al., 1996; Van et al., 1999; Gogu et al., 2001; Liu et al., 2004; Babiker et al., 2004; Nas and Berktay, 2006; Babiker et al., 2007; Nas and Berktay, 2010). The main purpose of this study is to interpolate groundwater level and compare geostatistical methods using the measured data from 634 piezometric wells in different locations around the Lake Urmia basin.

MATERIALS AND METHODS

In this study groundwater level data measured at 634 piezometric stations have been used. These stations locate in the West, South and South West of Lake Urmia basin in the North West of Iran as shown in figure1. Also number and location of piezometers, presented in table 1.

![Study area](image)

Figure 1 - Study area

<table>
<thead>
<tr>
<th>Number of</th>
<th>Region</th>
<th>Number of</th>
<th>Region</th>
<th>Water year</th>
<th>Coordinate of</th>
<th>Province</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piezometers</td>
<td></td>
<td>Piezometers</td>
<td></td>
<td></td>
<td>UTM</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>Salmas</td>
<td>72</td>
<td>Urmia</td>
<td>2010-2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Saeen-Ghale</td>
<td>37</td>
<td>Oshnaviyeh</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Kahriz</td>
<td>36</td>
<td>Bokan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Mahabad</td>
<td>13</td>
<td>Rashkan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>Miyandoab</td>
<td>9</td>
<td>Ziveh</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>Nagadeh</td>
<td>12</td>
<td>Sero</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Geostatistical estimation consists of two stages. The first stage consists of identifying and modeling spatial structure of variable that can be investigated by semivariogram analysis. The second phase is estimation and interpolating of data that has been done in various ways, such as Kriging method. The main condition of using geostatistical methods is stationarity that can be tested using semivariogram and independence and stationarity tests. Also distribution of used data should be close to a normal distribution.

**Independence and stationarity test (WW Test)**

Independence and stationarity test is a non-parametric statistical test and proposed by Abraham Wald and Jacob Wolfowitz in 1943.

\[
R = \sum_{i=1}^{N} x_i x_{i+1} + x_i x_N
\]

\[
\hat{R} = \frac{R}{\left(\frac{s_i^2 - s^2}{N-1}\right)}
\]
\[ \text{var}(R) = \frac{(s_1^2 - s_2^2)}{N-1} - R^2 + \frac{(s_1^4 + 4s_1^2s_2 + 4s_1s_3 + s_2^2 - 2s_4)}{(N-1)(N-2)} \]  \hspace{1cm} (3)

\[ S_r = N \cdot m_r \]  \hspace{1cm} (4)

\[ U = \frac{(R - \bar{R})}{\sqrt{\text{var}(R)}} \]  \hspace{1cm} (5)

Where \( \bar{R} \) is the average test, \( U \) is test statistic and \( m_r \) is the rth of the sample about the origin. After calculating the value of \( U \), the value of this expression at significant levels 90, 95 and 99% will be tested and the acceptable range for this test is as follows:

\[-1.645 < U < 1.645 \quad -1.96 < U < 1.9 \quad -2.575 < U < 2.575 \]

Semivariogram, is as a fundamental tool in the field of geostatistics that describes the relationships of a variable. In other words, it is a vector quantity that shows spatial correlation between the square of subtracting two points with regard to the direction and distance and can be explained by three parameters: (Iguzuiza, 1998 and Nazeri-Tahrudi et al, 2013)

1- Random (\( C_0 \))
2- Threshold (\( C+C_0 \))
3- Impact range (\( A \))

![Figure 2 - Parameters of semivariogram](image)

- Spatial variance of variogram (\( C \))
- Amount of variogram increases with increasing in r-value. But after a specified distance value of the variogram tends to be constant. Some of the variograms reach a relatively constant value which after that the value of variogram does not change significantly (Nazeri-Tahrudi et al, 2013). Difference between minimum and constant value of variogram is called spatial variance.
- Random variance variogram (\( C_0 \))
- Theoretically, the variogram value for \( h = 0 \) should be reduced to zero, but in practice the actual variograms which are product of experiences almost do not follow such circumstances. The value of variogram for \( h=0 \) is called nugget effect. With this model, we can observe the best theoretical variogram model that is most correlated with the actual data, related parameters (\( C_0, C+C_0, r \)) and correlation coefficients (\( \text{RSS}, r_2, C/C_0 + C \)). Also in addition in this model with plotting isotropic variogram, also anisotropic variogram is observable (Nazeri-Tahrudi et al, 2013).

**Interpolation using Kriging method**

The most important method for estimating spatial statistics is named Kriging method. This method is presented by D.G Krig (1951), whom was a mining engineer in South Africa. Kriging estimation method is based on the logic of weighted moving average and is defined as follows:

\[ Z_n^* = \sum_{i=1}^{n} \lambda_i Z_{v_i} \]  \hspace{1cm} (6)

Where \( n \) is number of stations, \( Z_n^* \) is estimated parameter, \( \lambda_i \) is weight of \( i \), \( Z_{v_i} \) are known parameter (Hilaire et al, 2003. Chen et al, 2007 and Theodossiou and Latinopolos, 2006).

Kriging estimation is unbiased with minimum variance. Unbiased condition is assumed in other methods such as polygons and inverse square methods. But one the important properties of Kriging method is that in this method, coefficients are determined in such a way, so minimizes variance of estimation. Kriging method gives estimations with errors of each estimate and this subject help to calculate distribution of error beside to the average amount of error (variance estimation). Using this unique feature it can be revealed parts with more error that need more data. On the other hand the reduction in variance can be estimated and determined before digging observation well. So the best points for digging in water and petroleum industry can be recommended.

**Interpolating using Inverse Distance Weighted method (IDW)**

All interpolation methods have been developed based on the assumption that points which are closer together, are more similar and correlated than farther points. In this method it has been assumed that correlations and similarities between neighbors are proportional to the distance between them that can be defined as inverse function of the distance from any point to neighboring points (Nazeri-Tahrudi et al, 2013).


\[
Z \hat{\omega}_{t} = \frac{\hat{\omega}}{\hat{\sigma}} \left( \frac{Z(t_0)}{\hat{\sigma}} \right)^2 \frac{1}{i(h_{i0} - s)} \oplus i = 1, 2, ..., n
\]  

(7)

Where \( h_{i0} \) is distance between the two positions of \( t_0 \) and \( s \), \( t_i \) is smoothy factor and \( p \) is the weight factor that is between 1 and 5. Accuracy of plotted variograms can be observed by the Cross – validate method.

In this method, the value of each measured point is omitted and then estimated by variogram. Finally actual values of each point are placed on the chart and are compared with the estimated values (Nazeri-Tahrudi et al, 2013).

**RESULTS AND DISCUSSION**

Average annual water table measured in piezometric stations in the period of 2001 to 2011 presented in figure 3. As shown in this figure, groundwater table variations are obvious in the recent years.

![Figure 3 - Changes of groundwater level in the studied area during 2001 to 2011](image)

Independency and stationarity of daily and average monthly data recorded at different piezometric wells tested by Wald–Wolfowitz test. As results presented in table 2 all used data are accepted at 1% and 5% significant levels.

<table>
<thead>
<tr>
<th>Month</th>
<th>Annual</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.32</td>
<td>2.53</td>
<td>2.02</td>
<td>0.97</td>
<td>2.56</td>
<td>2.01</td>
<td>1.65</td>
<td>2.24</td>
<td>0.05</td>
<td>1.93</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.01</td>
<td>0.00</td>
<td>0.33</td>
<td>0.01</td>
<td>0.04</td>
<td>0.1</td>
<td>0.03</td>
<td>0.96</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Statistics

|        | 1.93   | 0.59  | 1.64  | 1.96|
|        | 1.64   | 0.11  | 0.05  | 0.96|
|        | 0.05   | 0.96  | 0.05  | 0.11|
|        | 1.64   | 0.05  | 0.11  | 0.05|
|        | 0.05   | 0.11  | 0.05  | 0.05|
| P-Value| 1.93   | 0.59  | 1.64  | 1.96|
|        | 1.64   | 0.11  | 0.05  | 0.96|
|        | 0.05   | 0.96  | 0.05  | 0.11|
|        | 1.64   | 0.05  | 0.11  | 0.05|
|        | 0.05   | 0.11  | 0.05  | 0.05|
| Significance Level | Statistics Value |

One of the main assumptions for estimating and interpolating is that used data should be follow normal distribution, so the groundwater level data of the studied area normalized using the function \( Y = \text{LN}(X) \) and plotted for some months as shown in Figures 4 and 5. Also variogram is plotted in figures 6 and 7 from different months.
To illustrate the spatial correlation of well data Piezometers variogram graphs were used to model the spatial correlation of two months from the data of Figures 3 and 4 were presented.

To check outliers in time series which will affect the variance of the model Claude variance and H-scatter diagrams have been used. H. Results presented in Figures 8 and 9.
Also correlation of actual and estimated average groundwater level by Kriging and IDW methods presented in figures 10 and 11.
Table 3 – Correlation coefficients of actual and estimated data using geostatistical methods

<table>
<thead>
<tr>
<th>Month</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kriging</td>
<td>0.53</td>
<td>0.52</td>
<td>0.52</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.5</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>IDW</td>
<td>0.49</td>
<td>0.49</td>
<td>0.48</td>
<td>0.47</td>
<td>0.48</td>
<td>0.43</td>
<td>0.48</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.48</td>
<td>0.49</td>
<td></td>
</tr>
</tbody>
</table>
Groundwater level of studied area has not been changed from period of 2001 to 2009 but it has been showed that in the period of 2009 to 2011 groundwater level decreased (Fig 3). In this study stationary normalized data used to interpolation groundwater level of Lake Urmia basin by two geostatistical methods include Kriging and IDW methods. Using these two methods, groundwater level data through water year 2010 for each month modeled and results compared. Number of 634 piezometric stations with the same depth have been used and caused to produce same distribution pattern of groundwater in the scale of monthly and total average in different methods. As showed in figures 10 and 11 and table 3, Kriging model with the higher correlation coefficient between actual and modeled data than IDW method selected to interpolate groundwater of Lake Urmia basin that is in accordance with the results of Theodossiou (2006), and Fathi Marj (2008). But in the study of Rngzan (2005) inverse distance weighting method preferred to the Kriging method. One of the important points of modeling and interpolating different parameters is that in order to plot fitted models normalized data should be returned to the actual data using normalization functions and then correlation between actual and fitted data determined. This subject causes to calculate actual correlation coefficient between observed and modeled data.

REFERENCES


