To Compare Two Interpolation Methods: IDW, KRIGING for Providing Properties (Area) Surface Interpolation Map Land Price.
District 5, Municipality of Tehran area 1.

Yousefali Ziary, Hormoz Safari, ISLAMIC REPUBLIC OF IRAN

Key words: zoning, IDW, kriging

SUMMARY

To provide a uniform and continuous map from required place needs the interpolation methods in order to anticipate unknown values. In this regard, there are different interpolation methods and geo statistical estimations. The important factors such as number, the local distribution of sampling points and the capability of interpolation model have an effective role in the accuracy of providing the zoning map, several methods such as the methods of Tisen polygons and weighting based on distance photograph (IDW) which are among customary methods, have been presented for the interpolation of a variable, AT any rate, because of not considering the correlation between datum and not optimizing the size and from of used neighborhood, these methods often haven’t enough precision. Geostatistical theory is a branch of applied statistics that plays an effective role in the description and analysis of ground observation. In recent years, the development of Geostatistical theories and techniques has made many improvements in analyzing local datum in GIS environment. Using analysis based on Geostatistical theories decreases very much corresponding defects and limitations to classical statistical which have been developed by the random distribution theory of processes and variables, Datum interpolation is important for local analyzing by GIS, because many used maps for GIS operation are made by interpolation. In fact, it is possible to make continuous and smooth models through the time and local distribution of datum which are studied by interpolation. In research, two methods, kriging, to IDW have been examined for land price zoning and their advantages and disadvantages were considered. the evaluation and test of error which two interpolation models, have been studied by willmott way (1989) that shows kirging methods has a higher precision than IDW method.

1. INTRODUCTION

GIS is a set of a soft ware and hardware system which is able to link descriptive and local information. GIS has different ports such as datum input, datum management (to save, recover, control and provide data base ...), the analysis of datum and out put as hard copy and soft copy. The zoning is one of the local analysis methods that make a continuous surface by local datum as equivalent point and lines, which shows the changes of environmental value. Used functions of zoning are as two statistical and mathematical states. It shows datum demography in statistical state and numerical expression and data calculations in mathematical states. As the definition of anticipation and the calculation of points values and
unknown surfaces by using known points, to zoning the zoning map by interpolation method, this anticipation is determined based on the local correlation itself. Current methods of interpolation are as deterministic and indeterminate (geostatistical). In deterministic method, mathematical functions will be used in terms of the distance or closeness of sample points than neighbours (IDW) and or clear or flatness (spline) degrees. In indeterminate method, further using mathematical functions, statistics is used to anticipate unknown points. In this method, the rate of confidence level and correctness of anticipating points is calculated there are definitions for this method, for example, it studies a phenomenon that is changed in space and time (Deutsch, 2002) and Geostatistics is a description of the local continuity of natural phenomena and is in accordance with the techniques of classical statistics (Isaaks and et al, 1989). It is important to notice that local distribution aspect, a method is never proposed as an optimum model for different states. (Rangzan and others, 2005) because different precisions of a special model are obtained in different conditions. The dispersion of points and their local distribution, the compression and density of points and the capability of finding–mediator method define the accuracy and precision of interpolation.

1.1. IDW (inverse distance weight)

All interpolation methods have been developed based on this theory that closer points to each other have more correlations and similarities than farther points. In IDW method, it is assumed substantially that the rate of correlations and similarities between neighbors is proportional to the distance between them that can be defined it as a distance reverse function of every point from neighboring points. It is necessary to remember that the definition of neighboring radius and the related power to the distance reverse function are considered as important problems in this method. This method will be used by a state in which there are enough sample points with a suitable dispersion in local scale levels. One of the advantages of this method is to be suitable for showing Barriers discontinuous lines such as fractures, quasiling, faults levees and rivers which make fracture and discontinuity on the surface.

\[ Z_i = \frac{\sum_{j=1}^{N} z_j \cdot d_j^{-n}}{\sum_{j=1}^{N} d_j^{-n}} \]

\( Z_i \) = the estimation value of variable z in point i.
\( z_i \) = sample value in point i.
\( D_j \) = the distance of sample point to estimated point.
\( N \) = A coefficient that determines weigh based on a distance.

1.2. Geostatistics

Geostatistical methods are calculated based on statistical models that include correlation itself (the statistical relation of points has been measured). To done geostatistical methods, used datum must have two features: one the spatial normal distribution of datum and the
other, there aren’t meaningful change for average and variance in the place (Stationary). These two features are observable in the histogram of normal distribution and variance. The mathematical expression of geostatistical method is in the following relation:

\[ Z(u) = \mu(u) + \varepsilon(u) \]

\( Z(u) \) = A district variable value.
\( \mu(u) \) = A certain trend and structural component \( Z(u) \).
\( \varepsilon(u) \) = Error of sample than \( \mu(u) \) or a random component.
\( \varepsilon(u) \) = A random component structure is called as spatial structure that is a function of distance and coordinate independent, some how if the values of sample are close to each other and the difference of value is increased by the increment of points distances , the continuity of a district variable will be shown.

In terms of environmental features, the calculation of formula \( \mu(u) \) is as following polynomial function.

\[ \mu(u) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy \]

Two stages must be done for the geostatistical estimation of unknown values in the required place. In this stage, first the recognition and modeling of samples spatial structure and their homogenous or heterogenous state are studied in a diagram as called variogram. In the second stage, values are estimated by kriging method.

In the first stage, samples local structures are examined in variogram curves. The features of this carve are the local variance value of samples toward each other on axis \( y \) and the value of distance in terms of step number (lag) on axis \( x \) that been represented as \( h \) in formula (diagram 1).

First, variogram value is increased by the increment of \( h \) but in a point, it reaches saturation limit, it is called sill which is equal to \( z(x + h) \).

Nugget as a random variance and without structure is not dependent on direction and distance. The formula of calculation variogram is as follows:

\[ \lambda(h) = \frac{1}{2(N-h)} \sum_{i=1}^{N-h} [z(x+h) - z(x)]^2 \]
There are two kinds of variogram theory models:
One is some variograms in which $\lambda(h)$ value is increased by the increment of $h$ and they don’t tend to approach a fixed limit. Second, some variograms in which first, $\lambda(h)$ value is increased by the increment of $h$ and then they reach a fixed limit that is a same limit. In roofed (limited) model, the limit is equal to the variance of variable value in the sampled society.

Diagram 2: The variogram curve of samples in this research.

The second stage of estimating unknown values by using interpolation in which kriging method has been used there are different methods for estimation. One by classical statistics method and the latter geo statistical method which is based on available spatial structure in environment and on weighted movable average logic. One of the features of this method can
be calculated per related error estimation to it, to exert this method from first estimation, variable Z must have normal distribution, other wise we must use non linear kriging or convert datum into normal datum and we use logarithm for their transformation.

\[ Z_v = \sum_{i=1}^{N} \lambda_i Z_{vi} \]

Formula of interpolation calculation by kriging method:

\[ Z_{vi} \text{ is the value of estimation, } \lambda_i \text{ is, dependent quantity weight on sample and } Z_{vi} \text{. Is the sample initial value, } I. \]

The estimation of value \( \lambda_i \) is obtained from following equation:

\[
\begin{bmatrix}
C_{11} & C_{14} & \ldots & C_{1n} \\
C_{41} & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots \\
C_{n1} & \ldots & \ldots & C_{nn}
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{bmatrix}
= 
\begin{bmatrix}
C_{01} \\
\vdots \\
C_{0n}
\end{bmatrix}
\]

\( C_{ij} = \) Covariogram average for a state in which vector h is changed in 2 sample I o j.

\( C_{oi} = \) Covariogram average for a state in which vector h will be in sample I and anticipation sample o.

\( \lambda_i = \) kriging coefficients.

One of the most important features of this method is to calculate the estimation variance for every sample in all the levels and this helps as so that samples in which the estimation variance is high, show remote points at required place.

Its other feature is to lesson changes (reducing fluctuations) I. e. the changes will decreased between anticipation value and exact value (Hosni pak, 1999).

2. METHODOLOGY

The place under study is at district 5, Tehran municipality of area 1 where is at north of Tehran and includes Aghdasieh, Kashanak, Darabad, Saheb Gharineh and Ajodanieh (map 1). The area price was obtained from 480 geo samplings to provide price widening map that in some points, because of military position, these samples hadn’t good dispersions (map 2).

Among 3763 parcel in district 5 of area, 480 samples were selected about % 12.75 of all parcels of area. The criterion of selecting sample is based on the form of block, the number of parcels in block, the number of intersections around block and dead locks. Some how, the current parcels of intersection top, the end of dead lock and the center of block were selected as sample parcels (figure 1). In this distinct, the value of area price is between 666 to 3333 $ and the value average of samples price is 1485 $.
Figure 1: the way of selecting sample parcels in block.

Map 1: the situation of research local in country.
Map 2: the dispersion of sample points in district 5

Arcgis V. 9 software, spatial Analyst & Geostatistical Analyst program used to done interpolation methods and provide zoning map. After geosampling, the sample points are putted on the parcels of district 5 on map.

To calculate the interpolation of IDW method, first the value of coefficient n which has the least error RMS for interpolation was determined.

Table 1: to calculate optimum coefficient in IDW method:

<table>
<thead>
<tr>
<th>N</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS</td>
<td>198.33</td>
<td>185.88</td>
<td>182.77</td>
<td>184.22</td>
<td>186.88</td>
<td>189.55</td>
</tr>
</tbody>
</table>

In accordance with calculated error rate, RMS, n = 3 coefficient was suitable for the interpolation by IDW method.

To provide the widening map by kriging method, first sample were studied from statistical distribution, they had abnormal statistical distribution and to solve this problem, normalization was used in a logarithmic manner In kriging manner spherical variogram model was selected by abnormal sample be cause of having the least estimation error (k. krivoruchk and et al, 2004)
Than other models for linear estimators.
To determine area price for every parcel, first the widening map was converted from Restry structure into vector structure and then area price map of any parcel was extracted from the overlapping of area parcel maps on the vector widening map. The prepared map showed the higher price of area in Saheb Gharanieh amounting about 333 $ and also the lowest price in Kashanak and Darabad locales amounting to 111-1000 $ (map 5).
3. EVALUATION AND CONCLUSION

Willmott five indices method (1984) is used to evaluate accuracy and surrey the error rate of above mentioned interpolation methods.

1. Mean Absolute error (MAE)

\[
MAE = \frac{\sum_{i=1}^{n} | P_i - O_i |}{n}
\]

2. Root mean square (RMS)

\[
RMSE = \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{N}^{0.5}
\]

3. Systematic Root mean square (RMAEs)

Map 5: the area price map of parcels in district 5.
4. Un systematic Root mean square (RMSeu)

\[
RMSe_u = \frac{\sum_{i=1}^{N} [(P^n - P^2)^2]^{0.5}}{N}
\]

\[P^n = a + b o_i\]

5. Acceptability index (d)

\[d = 1 - \frac{N^* \text{RMSE}^2}{PE}, \quad PE = \sum_{i=1}^{N} (|P_i - o| + |o_i - o^-|)^2\]

\(O_i = \) taken sample

\(P_i = \) the estimation value of samples.

\(N = \) total number of samples.

\(A, b = \) linear regression between o and p

Diagram No 4: linear regression between observed values and anticipation in IDW model.
Diagram No 5: the linear regression between the observed values and anticipation in KRIGING model

The table No 2: the factory of regression and coefficient between the observed values and anticipation in two models.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>a</th>
<th>b</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDW</td>
<td>0.936</td>
<td>78.376</td>
<td>0.958</td>
</tr>
<tr>
<td>KRIGING</td>
<td>0.942</td>
<td>95.79</td>
<td>0.96</td>
</tr>
</tbody>
</table>

The table No3: the evaluation of the error rote in two methods of interpolation

<table>
<thead>
<tr>
<th>MODEL</th>
<th>MAE</th>
<th>RMSE</th>
<th>RMSE_s</th>
<th>RMSE_u</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDW</td>
<td>123.143</td>
<td>123.145</td>
<td>33.565</td>
<td>117.914</td>
<td>0.99045</td>
</tr>
<tr>
<td>KRIGING</td>
<td>122.938</td>
<td>122.941</td>
<td>33.85</td>
<td>118.29</td>
<td>0.9905</td>
</tr>
</tbody>
</table>

4. THE OBTAINED RESULTS ARE AS FOLLOWING

1. It is necessary to account the optimum factory of “n” when we evaluate the interpolation as IDW method, because there are the different values of RMSE for the different amounts of “n”. In this research, the amount of optimum factor of “n” that has the lowest error of RMSE, evaluated as 3.
2. According to the table No 2 , the value of R , the coefficient between the observed value and anticipation in two methods IDW and KRIGING is shown that the KRIGING method than the IDW is a little high .
3. According to the table No 3 , the evaluation of accounted error by the five above mentioned methods is shown that in two investigation methods MAE , RMSE and the
accept index of d, the method of KRIGING has the more accuracy than IDW method.

4. By normalizing and the elimination of remote samples by using the relation “\((Q_3 - Q_1) - \mu \leq X_i \leq (Q_3 - Q_1) + \mu\)" that \(Q_1\) is the first quarter, \(Q_3\) is the third quarter and \(\mu\) is for the mean, it is obtained a bout 68 sample of non sense from a total of 480 samples, the area price of it in this district is between 666 $ to 2429.46 $ and the mean value of the samples price is 1273.732 $ in the table No4, the investigation of error in two methods MAE and of the remote samples against their true amounts is shown. We can find the more accuracy in the method of interpolation KRIGING by investigating the result of the table No4. Also the effect radius decreased by the reduction of the numbers and the normal distribution of samples and this makes the table.

No4: the evaluation of error rate in two methods of interpolation with the nonsense samples.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDW</td>
<td>182.347</td>
<td>182.35</td>
</tr>
<tr>
<td>KRIGING</td>
<td>168.562</td>
<td>168.568</td>
</tr>
</tbody>
</table>

At last, it is not recommended an optimum method for interpolation, but it is related to the local conditions, the method of sampling, the local distribution of samples and their compression.

It is recommended to investigate the accuracy of the interpolation models by using the evaluation of the error and a model which has the lowest error selected.

BIBLIOGRAPHY

CONTACT

Safari Hormoz
Iran-Tehran-aqdassieh st-n 40-TGIC com.
Tehran
Iran (Islamic Republic of)
Tel.: +98-021-22296969
Fax: +98-021-22295898
Email: safarihormoz@yahoo.com