Evaluation of the Population Distribution Using GIS-Based Geostatistical Analysis in Mosul City

Sabah Hussein Ali¹ · Faten Azeez Mustafa¹ ²†

Abstract: The purpose of this work was to apply geographical information system (GIS) for geostatistical analyzing by selecting a semi-variogram model to quantify the spatial correlation of the population distribution with residential neighborhoods in the both sides of Mosul city. Two hundred and sixty-eight sample sites in 240 km² are adopted. After determining the population distribution with respect to neighborhoods, data were inserted into ArcGIS10.3 software. Afterward, the datasets was subjected to the semi-variogram model using ordinary kriging interpolation. The results obtained from interpolation method showed that among the various models, Spherical model gives best fit of the data by cross-validation. The kriging prediction map obtained by this study, shows a particular spatial dependence of the population distribution with the neighborhoods. The results obtained from interpolation method also indicates an unbalanced population distribution, as there is no balance between the size of the population neighborhoods and their share of the size of the population, where the results showed that the right side is more densely populated because of the small area of residential homes which occupied by more than one family, as well as the right side is concentrated in economic and social activities.

Key Words: GIS, Geostatistical Analyst, Population, Semi-variogram, Nineveh

1. Introduction

Recent developments in remote sensing and GIS make it possible to detect changes and devise strategies based on these changes. GIS-based geostatistical tools for urban planning have been increasingly powerful and versatile. GIS can be used to map and analyze the geographical distribution of population for the purpose of making future decisions and increase planning for agriculture, economic, education and health development (Dionissios et al., 2013). Geostatistics is a methodology for integrating spatial and temporal coordinates of observations into data processing (Simon, 2000). The precision of various geostatistical approaches to spatial interpolation of spatial distribution patterns has been tested by many scientists, (Goovaerts, 2006) provides...
a generalization of Poisson kriging, according to which the size and shape of administrative units and the population density (Xiao and Alexis, 2019). shows the effectiveness of parcel and decennial census data in highly developed urban cores for areal geostatistical interpolation, and calls for future research on population characteristics (Ferry et al., 2016). reviewed and contrasted three commonly used spatial interpolation techniques: ordinary kriging, inverse distance weighting and triangular irregular networks, and developed a new distribution-based distance weighting (DDW) spatial interpolation method to be used with temporal dimensions for large-scale spatiotemporal interpolation in environmental modeling. The geographical distribution of the population of any region reflects natural, human and financial circumstances, because these variables have a direct effect on the population and thus on the population concentration in a specific region or on the random allocation in another region. Natural factors may be the direct impact on the spatial distribution of the population in a given region, but economic and social factors may also make the area densely populated despite the lack of relevance in terms of natural human habitat (El Nahry, 2007).

Generally, by knowing certain patterns and understanding which factors have the greatest effect on population density and overall population, it’s possible to forecast future growth or possible decrease in population at any area of the world, as well as spatially locate them all over the world. It will then be possible to create policies that can safeguard the environment, enable the planet’s sustainability and enable the worldwide population to continue the alter (Minmin et al., 2018).

The present study is focus on applying GIS for the purpose of geostatistical analyzing and exploring data variability and spatial relationships of the geographical distribution pattern with the size of population according to the area in the residential neighborhoods of in the both sides of Mosul city. It’s attempted to find the relationship between the statistical variables and the spatially population size which will be provide a new thinking for the future urbanization of Mosul city.

2. Study Area

1) Location

Mosul city is the center of Nineveh governorate which is located in the northern part of Iraq. The geographical location the city is shown in Fig. 1 and can be noted to be found at longitudes 43°02′59.65″E and 43°13′57.89″E. and at latitudes 36°17′23.86″N and 36°25′45.05″N. Mosul is located on both sides

Fig. 1. Map of the study area.
of the Tigris river, dividing it into two parts locally known as left and right sides. The morphological characteristics of both sides of the city differ. The right side is characterized by its height as compared to the left side, where the heights exceed (289 m) in some places. Due to the existence of the Tigris flood plain, which represents the largest percentage of the left side and reaches the height of (220 m), the left side, it appears flatter than the right side.

2) Climate

According to the Koppen classification, Mosul city has a hot semi-arid climate (BSh), verging on the Mediterranean climate (Csa), with extremely hot dry summers and moderately wet, relatively cool winters (Deliang and Hans, 2013).

3) Population Size

Population distribution refers to the distribution of a population at a given time in a certain geographical area or spatial form of a population (Chen, 2015), that’s mean, a certain time connected with the total summation must be noted when speaking about the population and their distribution, because there is no a fixed population group but it’s always subject to change by influencing variables (births rate, deaths rate, all kinds of migration and administrative changes). As the process of urbanization accelerates, the population is increasingly concentrated in urban areas, as is the case in Mosul city according to the census carried out. Table 1 lists the total population of Mosul city according to population censuses (Special Development Plan-Nineveh (2010-2020)).

From Table 1, it concludes very important data to enable us to draw the results of the population increases of the Mosul city over time periods, of course these population increases led to the establishment of modern residential neighborhoods within the city. However, the increase in population expansion on the left side is more than the right side of the city due to the modern urban planning and the increase of commercial and economic centers as well as the presence of most government departments in this side, which led to the movement of many families from the right side and even from the rural areas in the districts of the governorate to the left side of city.

3. Theoretical Consideration

Semi-variogram is a graphical demonstration for presenting an image of spatial correlation of data. A GIS geostatistical analyst was used to determine Mosul’s spatial distribution of population variable through Kriging approach. Kriging uses the covariance and semi-variogram to make optimal prediction (Omer et al., 2012) Semi-variogram is a graphical illustration of the spatial correlation of data and a high value indicates a high level of co-movement. Thus, it takes two sample location and find the distance between them as it showed in equation (1) (Johnston et al., 2003):

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

Where n(h) is the number of pairs of data that are approximately separated by lag (h) within a given distance and direction class. If the values at z (xi) and z (xi + h) are automatically correlated, the result will be small relative to an uncorrelated pair. The semi-variogram modeling also can determine the best fit for the model that passes through semi-variogram points by calculating the variance between the input pairs.
variables. The Cross-validation was used to assess performance of predictions. Possible and the root mean standardized square error should be near 1 (Bilgehan and Ali, 2010). The correlation between the variables is an indication of the changes in the variance of the variables and is denoted as $\gamma(h)$ using the formula in equation (2) (Kazem et al., 2015):

$$2\gamma(h) = 1/n \sum_{i=1}^{n} [Z(x_i + h) - Z(x_i)]$$  (2)

The distance is indicated by $h$, $Z(x_i + h)$ and $Z(x_i)$ are indicated by point $x_i + h$ and $x_i$. Three parameters; nugget, sill and range are important in the interpretation of the semi-variogram as shown in Fig. 2. The sill corresponds to the overall variance in the dataset and the range is the maximum distance of spatial auto correlation. The nugget variance is the positive intercept of the semi-variogram and can be caused by measurement errors or spatial sources of variation at distances smaller than the sampling interval or both (Mauricio et al., 2014). Finally, the model’s efficiency can be evaluated using the root mean square error (RMSE) that is a $Z^*(x_i)$ feature and can use the following expression:

$$\text{RMSE} = \sum_{i=1}^{n} (Z^*(x_i) - Z(x_i))^2$$  (3)

Where, $Z(x_i)$ and $Z^*(x_i)$ are the observed and predicted values, respectively.

### 4. Methodology

#### 1) Data

The data adopted for this research were taken from Nineveh Planning Directorate, its included spatial and descriptive data about population censuses of Mosul city for the year 2013. The dataset also included a Shapefiles for the Mosul district and its neighborhoods as well as a shapefile for the Tigris River. The dataset also included a Panchromatic image satellite (Worldview 2) with a spatial resolution of (46 cm) covered the study area (https://www.geoimage.com.au/satellite/worldview-2). The sensor bands and resolution are listed in Table 2.

#### 2) Method

Geostatistics is a combination of techniques and accidental models that analyze the characteristics of spatial data. This technique has the capacity and

<table>
<thead>
<tr>
<th>Sensor bands</th>
<th>Panchromatic</th>
<th>Panchromatic: 450-800 nm 8 450-800 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal</td>
<td>400-450 nm</td>
<td>450-510 nm</td>
</tr>
<tr>
<td>Blue</td>
<td>510-580 nm</td>
<td>585-625 nm</td>
</tr>
<tr>
<td>Green</td>
<td>630-690 nm</td>
<td>605-745 nm</td>
</tr>
<tr>
<td>Yellow</td>
<td>770-895 nm</td>
<td>860-1040 nm</td>
</tr>
<tr>
<td>Red</td>
<td></td>
<td>Multispectral: 1.85 m GSD at nadir, 2.07 m GSD at 20° off-nadir</td>
</tr>
<tr>
<td>Red Edge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIR1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIR2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Typical semi-variogram and its parameters.
power of estimation and can also express an error in estimation. The estimation of spatial structure and distribution variable values and the evaluation of the error associated with this estimate is called kriging geostatistical approach (Andreas, 2013).

In current study, the ArcGIS10.3 / ordinary kriging approach was used to predict the relationship between the area and population attribute data fields provide by the Nineveh Planning Directorate (NPD) for the year (2013) for both sides of the Mosul city as shown in Fig. 3. Mosul city is divided into nine sectors (five in the left side and four in the right side) for the purpose of facilitating the management of municipal services in the city (Mosul Municipality Directorate, 2019), all these sectors are including one hundred forty-one neighborhoods and covering an area of about 240 km². Based on the population size and area attribute field, the total sample points taken by the study are 268 samples (including those who transgressed on agricultural land).

The spatial distributions of the samples is shown in the Fig. 4 Below for the years (2013), while Fig. 5
shows the classification map of the neighborhoods with respect to population size of the study area. Fig. 4 and Fig. 5 gives an important indicator of population distribution on both sides of the city of Mosul, as we note from the two figures, that the density of population distribution in areas with proper municipal planning is greater than the areas spread across the borders of the city of Mosul, which have been distributed by cooperative societies and are mainly agricultural land and are not subject to municipal planning. Also, the high-density areas shown in Fig. 4 are characterized by offering health and municipal services as well as the available of schools, water and electricity.

The samples dataset were used for prediction the spatial distribution of the population size with the area of the neighborhoods by using Ordinary Kriging interpolation. Before the actual interpolation was done, a semi-variogram analysis was conducted to determine the spatial similarity of the data samples using the ArcGIS10.3 geostatistical analysis tool. After defining and analyzing the spatial pattern distribution of the samples using the semi-variogram function, cross-validation have been used to decrease the errors of sample points distribution by implementing the regression function. The prediction map was calculated to estimate the unsampled points.

5. Results and Discussion

Ordinary kriging is the most widely used method of kriging interpolation. This helps to calculate a value at a point in a region known to be a semi-variogram by using estimation location data in the neighborhood. Ordinary kriging implicitly tests the mean in a moving neighborhood with local second-order stationarity. Many models of semi-variogram are checked to get the one that fits the dataset of samples, Spherical model was chosen as a best one to illustrate the spatial autocorrelation (covariance scales of input data) of the studied samples dataset and gives the best interpolation map.

Fig. 6 shows the semi-variogram model for the data of population samples in both sides of Mosul city. The semi-variogram displays the variances within groups of observations (red dots) plotted as a function of distance between the observations to display the tendency for nearby observations to be more alike than distant observations.

From Fig. 6, it is noted that the semi-variogram model is flattened at a range of (2351.7) in the x-axis which’s corresponding to sill of ($7 \times 10^7$) at the y-axis, then all the sample locations separated by distances closer to the range are spatially autocorrelated. That is to say, the covariance function decreased between...
points and obtain similar function, whereas locations farther apart than the range are not autocorrelated (increase the distance between points).

Table 3 lists the resulted values of the semi-variogram model, the sill was calculated by adding the partial sill to the nugget. To evaluate the accuracy and precision of the resulted spatial interpolation, the cross-validation graph must be determined.

Fig. 7 shows the resulted cross-validation graph which illustrates the relation between the measured and predicted error values using spherical model.

According to the semi-variogram model and the cross-validation approach, the anticipated error values are presented in Table 4.

After analysis of spatial dependence, we evaluated sites that had not been sampled in order to produce surface maps by ordinary kriging. This has been done by defining a circle with four sectors to enclose the

<table>
<thead>
<tr>
<th>Semi-variogram</th>
<th>Statistical Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.35*10^7</td>
<td>Partial Sill</td>
</tr>
<tr>
<td>5.23*10^7</td>
<td>Nugget</td>
</tr>
<tr>
<td>85.59</td>
<td>Lagsize</td>
</tr>
<tr>
<td>1015.107</td>
<td>Range</td>
</tr>
<tr>
<td>No</td>
<td>Anisotropy</td>
</tr>
<tr>
<td>12</td>
<td>No. of Lags.</td>
</tr>
</tbody>
</table>
dataset that are used to predict a value at an unmeasured location as shown in Fig. 8.

The final spatial prediction map by ordinary kriging method are presented in Fig. 9 for the population size at the year of 2013 in the both sides of the Mosul city. The prediction map of the population size shown in Fig. 9 Confirms the validity of the results shown in Fig. 6 and Fig. 8 that prediction standard error of that location near sample points generally have a lower error and spatially autocorrelation unlike the far sample points. Fig. 5 and Fig. 9 gives the impression of an irregular distribution of the population size in relation to the area of the neighborhood. There are small areas densely populated, as in many neighborhoods on the

<table>
<thead>
<tr>
<th>Value</th>
<th>Type of error</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.3897</td>
<td>Mean</td>
</tr>
<tr>
<td>0.0043</td>
<td>Mean Standardized</td>
</tr>
<tr>
<td>1.014</td>
<td>Standardized Mean Square Error</td>
</tr>
<tr>
<td>546.523</td>
<td>Root Mean Square</td>
</tr>
</tbody>
</table>

Table 4. Performance of the implemented criteria
right side (spatially in old Mosul city), while there are large areas with little population, and this is evident in the new neighborhoods in the city of Mosul, whether on the right side or the left side of them, noting that there is a large urban expansion on the left side.

6. Conclusion

This study has attempted to predict the spatial distribution of the population size in the Mosul city by using 268 sample dataset using GIS geostatistical analyst / ordinary kriging interpolation method. Generally, results showed that, the semi-variogram/spherical model was selected as the best to illustrate the spatial autocorrelation of the studied samples dataset with standardized mean square error of (1.014) and addresses spatial autocorrelation and cross-autocorrelation related to the distribution of people in urban areas. Rather than ignoring spatial dependency, it models the spatial autocorrelation of population and the impermeable surface fraction by means of variograms and applies them in population interpolation. The resulted prediction map illustrates that the population density is concentrated in the old neighborhoods on the right side and then the left side, while the distribution is random in many neighborhoods that are not subject to city planning. The limitations of the present study were the effort involving of obtaining data from several official authorities in the city, after which it was re-revised and corrected before carrying out the geostatistical analysis and preparing the required maps for the study.

Future work should be done on implementation of GIS classification model to identified the land use of the Mosul city and then find the relation between the population spatial distribution and the land use maps to obtain newmapf or classifying population distribution according to land use of the both side. Another study can be done to study the distribution of health centers and the services they offer according to the spatial distribution of the population in every neighborhoods of the city.

The last point is that estimates of obtained population spatial distribution are essential for urban planning applications. For example, the population density in residential neighborhoods is the important indicator in sustainability studies of self-dependent regions and in future population census process.

Acknowledgements

The authors thank the Nineveh Planning Directorate for the facilities provided by them in preparing the data adopted in the study. We also thank the University of Mosul for the continuous scientific support in carrying out scientific research that serves the community.

References


methods to estimate forest volume in the municipal forest of the Greek island Skyros, *Geo-Spatial Information Science*, 16(2): 100-112.


