Methods of Spatial Interpolation in The Representation of Soil Properties/Nineveh Governorate Soil Case Study

Suhaib H. Khadir Taha\textsuperscript{1a}\textsuperscript{*} and Osama M. Najim\textsuperscript{1b}

\textsuperscript{1}Geography Department, Human Education College, University of Mosul, Mosul, Iraq.
\textsuperscript{a}Sahlah80@gmail.com, \textsuperscript{b}smslmn408@gmail.com

Abstract: Spatial interpolation is a process used to guess cell values at points or locations where real models are not available. The principle of interpolation is based on Spatial autocorrelation and Spatial Dependence, which measures the degree of relationship or dependence between points or objects near and far and by using statistical techniques to address the effects of data values and the relationship between the various interrelated. The study relied on more than one method of spatial interpolation (Kriging - Inverse Distance Weighted-Radial Basis Functions-Global Polynomial-Local Polynomial). To represent data for one of the most important elements of the natural environment and the soil and within the spatial space of the governorate of Nineveh, In the present study, some of the physical properties of the soil are expressed (soil texture, pH, bulk density, and chemical components such as (organic matter, total Bases, cation exchange capacity, exchangeable sodium percentage, calcium carbonate, available water capacity, and electrical conductivity based on (92) points distributed according to the results of the spatial interpolation test based on the mean bias error values and root mean square error, which was used as a criterion for the differentiation between the methods of completion used in the study. The study also showed that the Kriging method is suitable for the parts of spatial completion of soil properties. It was possible to reach (27) maps to represent some of the physical and chemical properties of the soil of the study area.

Keywords: Geostatistical analysis, GIS, soil Characterization, spatial Interpolation, maps.

1. Introduction

Spatial statistics technologies have proven to be highly effective in studying many elements of the natural environment with a wide area span that makes it difficult to survey and cover as well as "the material cost in the case of its study, but Interpolation Methods have provided solutions to such cases and have become used In many studies, especially after the results matched with the results of the prediction of traditional methods, Spatial Interpolation is a smart guessing process in which the researcher or user attempts to make a reasonable estimate of a value in a continuous field and in natural places where the value has not been measured, It is an important way to reach the aim of estimate the values of unmolded interfaces by applying point views. Where statistical techniques are used to address the effects of data values and the relationship of the different incoming among them, and the self-correlation analyzes provide quantitative estimates of the spatial relationship between each of the two data taking into account the distances between them, so the effect on the converged sites is more than the sites that are far from each other The other one.

The research aims to use spatial prediction methods to study some of the physical and chemical soil properties so that it is possible to estimate the field data not studied and express the spatial change in the values along the surface of the study. Nineveh governorate was chosen as a study area to know the spatial distribution of (soil tissue, organic matter (OM), (pH), bulk density (BD), total exchangeable...
bases (TEB), cation exchange capacity (CEC), exchangeable sodium percentage (ESP), calcium carbonate (CaCO3), gypsum, available water capacity (AWC), and electrical conductivity (EC) based on (92) randomly distributed points. While the importance of the research is summarized in finding the best representation for the distribution of some physical and chemical soil properties in the study area, and by using spatial interpolation methods (spatial statistics) in the geographic information systems environment where they save time, effort and reduce costs. The study problem lies in the ability to benefit from the accuracy of methods Spatial completion and its effectiveness in representing some physical and chemical properties of the soil, which are available within the programs of geographic information systems, and conducting various statistical Solutions to choose a method of interpolation that is appropriate to distribute the values of soil properties in Nineveh Governorate. To achieve the aim of the study, a quantitative analytical research approach was followed by providing the required data on the phenomenon studied with a view to interpolation it spatially, as well as making use of the geographic information systems technology, programs, and spatial analytical tools.

2. Data Sources and Programs Used in The Study

The sources of data used in the current study are:
1) Land Sat 8 ETM+
2) Digital Elevation Modal
3) HWSD (Harmonized World Soil Database (version 1.2.2009) (FAO) Food and Agriculture Organization of the United Nations, (IIASA) International Institute for Applied Systems Analysis, ISRIC-World Soil Information, (ISSCAS) Institute of Soil Science–Chinese Academy of Sciences, (JRC) Joint Research Centre of the European Commission).
4) Soil data for some areas of Nineveh Governorate from (Ministry of Water Resources -National Center for Water Resources Management - Department of Environmental Studies).

The programs used in the study are:

2.1. Global Mapper v.18.
It is considered a GIS application in data processing and is characterized by its ease of use, with access to an unparalleled variety of spatial data, and was used to integrate and process Images, various maps and the digital height model.

2.2. Arc GIS v10.4
It is an integrated set of applications of geographic information systems, and provides tools for spatial analysis and is used to build information layers, conduct spatial analysis and producing maps. The area of Nineveh Governorate is about (37863) km². The study area represented by Nineveh Governorate is located in the northern and northwestern part of Iraq, bordered to the north by Dohuk Governorate and Lake Mosul Dam, while it is bordered to the east by the governorates of Erbil and Kirkuk, and south by the provinces of Salahuddin and Anbar and to the west of the Syrian Arab Republic, and accordingly is limited to The study area between the two latitudes (34 55) - (30 37) north and between longitudes (41 25) - (44 25). Figure 1 shows the study area and locations of boreholes.

3. Natural Characteristics of The Study Area

3.1. Geology.
As the type of soil prevailing in a region is determined by the type of parent rock (the original material), which depends to a large extent on the prevailing rock and its characteristics and development over time [1]. There is a group of rock formations and sediments in Nineveh Governorate that reflect different sedimentary environments, extends the chronological age from the Cretaceous period from the second geological time to the Holocene era from the fourth geological time, and the characteristics of these formations and sediments of different rocks between sandy, mud, alluvial, limestone, gypsum and halite [2]. The properties of these rocks with other environmental variables in the study area were reflected in determining many of their physical and chemical properties.
3.2. Surface
In terms of terrain and slope, the study area is divided into each of the flat sector and its lands have few and medium slopes (0-7.9%) according to the (Zink) classification and constitute an appropriate environment for soil development and stability and of great thickness and is the most widespread form of sloping surfaces in Nineveh Governorate with an area (35.875450) km², 94.75%, Compared to the other sector (the folds) whose lands have high slopes (8-30)% as these surfaces do not allow the soil to develop as their moisture content is low, so we find soil with shallow rocky or naked soils in steep slopes [3].

3.3. Climate
The climate is the active factor among the factors of soil formation and is essential in all biochemical reactions in which it occurs. The erosion factors depend entirely on the elements of rain, temperature, and wind in weathering and transporting materials [4]. The study area is characterized by its transitional location between three types of climatic regions according to the classification (Koppen) Mediterranean climate (Cs) and prevails in the northern and northeastern regions and the Sinjar Heights, and the semi-arid climate (Bs) and dry (Bw) in most parts of the study area [5]. In general, the climate of Nineveh Governorate is hot dry summer, and rainy in winter The temperature rates are varied, due to the effects of location, height, and different atmospheric pressure, and accordingly, the roles of each heat factor in mechanical and chemical weathering (rock debris and rocky crumbs) that the various soil formation factors in the region depend on, and the role of falling rain Some elements of the melt this rock fragment mixed with water to carbon dioxide, such as calcium carbonate, sodium, magnesium, and calcium sulfate. To keep gravel, sand, clay, clay, iron oxides, and lime.

3.4. Natural Vegetation Cover
The vegetation cover its importance lies in that it works to stabilize the soil and protect it from erosion, not to mention the direct contribution to increasing the concentration of organic matter inside the soil after its death and its decomposition, and that the natural plant in the study area consists mainly of herbaceous plants, where it covers the largest part of the conservative lands, while The trees form a limited part, in addition to other plant species [6]. The contribution of the vegetation to the composition of the soil of the study area varied according to its density and durability.
4. Spatial Interpolation Methods
To compare the spatial interpolation methods, the five representation methods (Kriging, IDW, Radial Basis Functions, Global Polynomial, and Local Polynomial) in this study were compared to choose the best method for all variable data and present the results according to statistical parameters [7].

1) **Mean Bias Error (ME):** Equation of errors in the results is:

\[
\text{Mean Bias} = \frac{1}{N} \sum_{i=1}^{N} (M - O)
\]

Where
N represents the number of data or points
M The estimated value of the studied property
O represents the measured value of the studied property.

2) **Root Mean Square Error (RMSE):** It is used to evaluate the model's susceptibility to graphical representation [17].

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_p - X_m)^2}
\]

Where
N represents the number of data or points
X_p is the estimated value represents studied models
X_m is the measured value represents the studied models.

To find out the efficiency of spatial interpolation methods, compare the efficiency of the method in integrating data through error results (mean error and the square root of the error square resulting from using the method) [10].

a) When the value of the mean error or the bias error is equal to zero or close to zero, then this means that the method fits the data, that is, the larger the value it means less representation and greater bias, incompatibility [11].

b) The small value of the square root of the error square indicates that the estimate of the results is accurate and appropriate.

4.1. Data Distribution Patterns
The process of determining the quality and distribution of data helps in the appropriate selection of statistical methods and methods used with geographic information systems based on the results. Most data follow the symmetric natural distribution, in which all measures of central tendency apply to each other and their values are equal. The natural distribution is not achieved in most data for the natural properties of soils that may contain anomalous or extreme values or large values within a small set of values or vice versa when the graph or distribution is twisted Iterative towards outliers. In this case, it should not be sufficient to describe the distribution in measures of central tendency, but rather it needs another measure that expresses the extent of the distribution in the distribution from the natural or the degree of (Skewness) and (Kurtosis) [12]. The frequency distribution is crooked, positive if the extreme data are large values, the long end of the curve extends to the right and the measures of central tendency become mean then the mode, and negative, if the extreme data are small values, then the long end of the curve extends to the left and the central tendency measures become the mode and then the median. Then the mean. Figure 2 below shows the shapes of the three distributions curves for the positions of the arithmetic mean for the mean, median, and mode [13,14].
5. Data Analysis

Geographic information systems meet with geostatistic at the common point of interest in conducting analytical operations on quantitative data, calculating statistical measures, and knowing the movement of contrasting models within one region. The next stage is choosing the best programs to produce the results and display. The study of the region begins by collecting data on its elements followed by arrangement and coordination to facilitate its statistical analysis to obtain the information. The statistical approach uses two types of statistical applications for data [8].

1) **Descriptive analysis** and is concerned with the digital description of a particular society (study models) so do not withdraw its results to another environment (the rest of the study area).

2) **The inferential analysis** includes the sampling process any small sample (study models) representing the large community (study area) provided that the final results are approximate and within the limits of a statistically calculated error.

Sample consistency the problem of extracting reliable information from the sample is concentrated in determining its conformity with the study community (study area) and although it is impossible to confirm conclusively from that, we can come up with appropriate estimates for calculating the standard error. The process of calculating the standard error includes identifying the amount of difference or fluctuation in the measurements that the sample provides us with the averages for the different samples (central tendency measures) and calculating the standard deviation mathematically from the averages and determining the pattern of data distribution for each variable, whether natural or abnormal (dispersion measures) and then Determine the value and direction of deviations (data distributions) [14].

5.1. Steps and Modus Operandi

The steps of the work in the study were to set up the basic data approved in the study, which is the soil data of the study area which was obtained from the Harmonized World Soil Database (HWSA), and the soil data obtained from the Ministry of Water Resources-National Center for Water Resources Management - Department Environmental studies (paper data) for some regions of the study. And then exporting that data in the format (XLS.) To the Arc Map v10.4, and projecting it in the form of points bearing each of them (XY) values and drawing a shapefile for each point projected with the entry of data for each point separately carrying each point the physical and chemical properties, to select the ideal method for the spatial interpolation of the data, statistical values (central tendency measures and dispersion measures) have been found for each of the soil properties to know how to distribute data and make the necessary transformations for them, and choose an appropriate method for spatial completion based on the values of the mean error and the square root of the mean error and apply that to methods Spatial interpolation available in the (Geostatistical analysis) window within the Geographic Information Systems program, and choosing an appropriate data method with high accuracy and less.
The spatial analyzer ArcGIS v10.4 provides several tools or spatial interpolation methods for creating a surface cell network of raster data. There are two interpolation techniques:

- **Deterministic** to design surfaces based on measured points or mathematical equations, a method based on the extension of cell symmetry or similarity, such as the inverse of the weighted distance.
- **Geostatistics** depending on Statistical processors and is used in the design of advanced models of surfaces, and is characterized by including accuracy calculations for points estimated as in the kriging method.

It is evident from the above that the spatial interpolation methods present in the program's spatial analyzer follow different methods to reach the desired map of the data, all of which are correct when they meet the conditions on which they were established and the map representing the study area is closer to reality. It contains only a very small percentage of Mistakes. The result that can be reached from all methods gives us maps that vary from one to the other and may share small details or move away from them a lot. To choose the ideal way to represent the soil data for the study area and represent it more closely to the truth, rely on the statistical values and the square root of the error square, which were used to compare spatial interpolation methods.

### Table 1. Values of the statistical measures of the soil of the study area.

| Properties (Depth)(m) | Statistical measures |
|-----------------------|----------------------|
|                       | Count | Min | Max | Mean | Median | Std. Dev. | Skewness | Kurtosis |
| Top Sand% (0-30 cm)   | 92    | 16  | 43  | 32.6 | 35     | 7.878     | -1.258    | 3.251    |
| Sub Sand (%) (30-100 cm) | 92  | 0   | 40  | 30.6 | 35.5   | 9.662     | -1.547    | 4.466    |
| Top Silt% (0-30 cm)   | 92    | 29  | 47  | 39.39| 40     | 5.994     | -0.5252   | 1.985    |
| Sub Silt% (30-100 cm) | 92  | 0   | 45  | 35.16| 37     | 8.0688    | -2.7218   | 12.407   |
| Top Clay% (0-30 cm)   | 92    | 18  | 55  | 28.141| 23     | 12.629    | 1.391     | 3.238    |
| Sub Clay% (30-100 cm) | 92  | 0   | 57  | 30.65| 27     | 13.148    | 0.708     | 3.488    |
| AWC (mm/m) (0-100 cm) | 92  | 15  | 150 | 110.38| 125    | 45.563    | -0.563    | 1.610    |
| Top Bulk Density (0-30 cm) | 92 | 1.25 | 1.65| 1.427| 1.39  | 0.1177    | 0.550     | 2.1495   |
| Sub Bulk Density (30-100 cm) | 92 | 0   | 1.76| 1.394| 1.41  | 0.2927    | -3.321    | 17.47    |
| Top OM (0-30 cm)      | 92    | 0.37| 0.96| 0.5405| 0.47   | 0.1582    | 1.226     | 3.730    |
| Sub OM (30-100 cm)    | 92    | 0   | 0.53| 0.298| 0.22   | 0.122     | 0.1014    | 2.0874   |
| Top pH (0-30 cm)      | 92    | 7.6 | 8.3 | 7.9022| 7.9    | 0.1176    | 0.5277    | 5.003    |
| Sub_pH (30-100 cm)    | 92    | 0   | 8.2 | 7.7054| 8      | 1.428     | -5.195    | 28.239   |
| Top CEC_{Soil} (0-30 cm) | 92 | 10  | 44  | 19.72| 16.5   | 11.424    | 1.284     | 3.249    |
| Sub CEC_{Soil} (30-100 cm) | 92 | 0   | 43  | 17.77| 15     | 11.842    | 1.11      | 3.257    |
| Top TEB_{Soil} (0-30 cm) | 92 | 10.4| 50.8| 25.182| 24.65  | 12.245    | 0.239     | 1.913    |
| Sub TEB_{Soil}(30-100 cm) | 92 | 0   | 58.3| 30.649| 31.45  | 17.672    | -0.0859   | 1.323    |
| Top ESP (0-30 cm)     | 92    | 1   | 41  | 4.641| 3      | 7.355     | 3.788     | 16.65    |
| Sub ESP (30-100 cm)   | 92    | 0   | 43  | 15.837| 17     | 12.228    | 0.3017    | 1.623    |
| Top CaCO_{3} (0-30 cm) | 92 | 0.2 | 31.6| 14.09| 9      | 12.645    | 0.501     | 1.449    |
| Sub CaCO_{3} (30-100 cm) | 92 | 0   | 35  | 12.183| 9      | 10.05     | 0.641     | 2.187    |
| Top CaSO_{4} (0-30 cm) | 92 | 0   | 15.1| 4.859| 1.1    | 6.652     | 0.890     | 1.816    |
| Sub CaSO_{4} (30-100 cm) | 92 | 0   | 23  | 7.87 | 2.5    | 9.915     | 0.836     | 1.791    |
| Top EcE (0-30 cm)     | 92    | 0.1 | 22.3| 2.44 | 1      | 4.335     | 3.98      | 18.26    |
| Sub EcE (30-100 cm)   | 92    | 0   | 14.8| 3.504| 1.5    | 3.66      | 1.29      | 4.33     |
| Top USDA Tex (0-30 cm) | 92 | 1   | 3   | 1.79 | 2      | 0.433     | -1.034    | 3.210    |
| Sub_USDA Tex (30-100 cm) | 92 | 0   | 3   | 2.33 | 3      | 0.952     | -1.022    | 2.583    |
6. Discussion of Results
The Kriging method was chosen to produce Spatial Interpolation maps as appropriate to the natural properties of the soil according to the results of the Spatial Interpolation test as in Table (2) as well as its ability to deal with data whether it is of natural or abnormal distribution (logarithmic distribution or BOX distribution - COX) as Kriging method is classified as a spatial and statistical method and depends on the principle of correlation between the variables in the sample values and the spatial direction of the distribution of these points, and takes into account both the distance, degree, and direction of the change between the known points, It also has several mathematical models (Spherical, Circular, Ordinary, Universal, Exponential,...) that give it the ability and flexibility to deal with the data, which gave the best results and all the studied soil variables and proved their high effectiveness according to the results in their application to soil properties data of study area.

| Interpolation Methods | Soil properties | ME     | RMSE  |
|-----------------------|----------------|--------|-------|
|                       | Top Sand, %    | 0.12   | 0.44  |
|                       | Sub Sand, %    | 0.09   | 0.670 |
|                       | Top Silt, %    | 0.05   | 0.41  |
|                       | Sub Silt, %    | 0.01   | 0.45  |
|                       | Top Clay, %    | 0.006  | 0.40  |
|                       | Sub Clay, %    | 1.1    | 0.30  |
|                       | AWC, mm/m      | 0.688  | -0.033|
|                       | Top Bulk Density| -0.003 | 0.11  |
|                       | Sub Bulk Density| -0.0009 | 0.2   |
|                       | Top OM, %      | -0.0002 | 0.15 |
|                       | Sub OM, %      | -0.0007 | 0.16 |
|                       | Top pH         | -0.003  | 0.4   |
|                       | Sub pH         | 0.07    | 1.20  |
| Kriging/CO_kriging    | TOP_CEC_Soil   | -0.003  | 0.38  |
|                       | SUB_CEC_Soil   | -0.010  | 0.39  |
|                       | TOP-TEB_Soil   | -0.013  | 0.48  |
|                       | SUB-TEB_SOil   | -0.03   | 0.44  |
|                       | TOP-ESP        | 0.007   | 0.32  |
|                       | SUB-ESP        | 0.033   | 0.41  |
|                       | TOP_CaCo3      | 0.02    | 0.5   |
|                       | SUB-CaCo3      | -0.002  | 0.38  |
|                       | TOP-CaSo4      | 0.001   | 0.26  |
|                       | SUB-CaSo4      | -0.001  | 0.43  |
|                       | TOP-ECe        | -0.01   | 0.29  |
|                       | SUB-ECe        | 0.0009  | 0.3   |
|                       | T_USDA_TEX     | 0.01    | 0.41  |
|                       | S_USDA_TEX     | 0.01    | 0.84  |

As a result, according to the final on the previous stages were obtained maps of the spatial interpolation shown in Fig. 3 which represents some physical and chemical properties in the study area and on two levels (surface (0-30 cm) and subsurface (30-100 cm)).
Figure 3. Some physical and chemical properties of soil in the study area.
7. Conclusions and Proposals

- The distribution of the data studied in the study area was not a natural distribution of most of the physical and chemical soil properties except for the (pH), (silt) and (Bulk density) properties, so logarithmic distribution was used in the data distribution process, to make it appropriate in Spatial Interpolation processes.

- The possibility of preparing Spatial Interpolation maps to represent soil properties using geological statistics and according to the results of the mean error and the square root of the mean error of the data, and the process took place within the GIS software with high accuracy.
The kriging method was used in the Spatial Interpolation process, as it uses more than one mathematical and statistical formula to reach the desired result.

The study showed, through the values of the statistical measures and iterative scales, as well as in the results of the Spatial Interpolation values and the final maps represented, the variation of the characteristics of the studied properties of the soil between the surface and subsurface, due to the different influencing conditions.

The study showed that there is a discrepancy between the parts of the study area in its physical and chemical properties of the soils in the maps representing the study area and they are identical to the prevailing natural data in the study area between the diversity of geology, geomorphological, climatic and natural layers of vegetation.

The study also showed that the values of most of the studied soil properties in the study area are acceptable and within certain specifications except for the gypsum characteristic.

Recommendations

- The study recommends applying this type of study at larger levels and more models so that its results are closer to simulating reality and comparing the results obtained in the field.
- The study recommends increasing attention to this field of scientific and modern technology and its introduction in various fields and other environmental studies and linking it to other branches of science as it creates a database that can be used by decision-makers.

References

[1] M.M.S. Al-Mesmari, Wadi Al-Slayeb Basin - Study in Geography of Soil, Libyan International Journal, College of Education Al-Marj, University of Benghazi, (2016).
[2] T. Buday, The Regional of Iraq Stratigraphy and Palecography, Dar Al- Kutib Publishing house, University of Mosul, Iraq (1980).
[3] S. Marin, GIS solution in, Natural Resource Management, Ed. Renewable Natural Resource Foundation and, Natural Academy of Sciences Natural Resource, Council, Washington, (1999).
[4] A.H. Al-Shalash, Geography of Soil, Baghdad University Press, (1985).
[5] S.B. Al-Najmawi, Nineveh Governorate Climate, Master Thesis (unpublished), Department of Geography, College of Education, University of Mosul, (2001) p. 45
[6] K.S. Al-Juburi, The effect of climate change on the development of vegetation in Nineveh Governorate, Ph.D. thesis, Department of Geography, College of Education, University of Mosul, (2015).
[7] A. Hani and S. A.H. Aburi, Determination of cd, Zn, K, pH, TNV, Organic material and Electrical Conductivity (CE), Distribution in Agricultural Soils using Geostatistics and GIS (case study: western of Natanz –Iran). World Academy of Science, Engineering and Technology, No.60 22-24.
[8] R. Webster and M.A. Oliver, Geostatistics for environmental scientists. John Wiley & Sons, 2007.
[9] E. Verfaillie, V. Van Lancker, and M. Van Meirvenne, M., Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas, Continental Shelf Research, 26(19) 2006 2454-2468.
[10] NWillmott, C. J. and K.Matsuura (2005). Advantages of the mean absolute error (MAE) OVER the root mean square error (RMSE) in assessing average model performance. climate Research, vol.30, pp.79-82.
[11] S.A. Moasheri, A. S. Goshki and A. Parsaie, “SAR” Qualities parameter persistence by a compound method of geostatic and artificial neural network (Case study of Jiroft plain), Int. J. Agri. Crop Sci., 6(3) 2013 157-166
[12] K. Johnston, J.M. Ver Hoef, K. Krivoruchko, and N. Lucas, Using ArcGIS geostatistical analyst. Vol. 380. Redlands: Esri, 2001.
[13] Kroulik, M., A Slejska, D, Kokoskova, and V.Venclova(2005). Soil Electrical Conductivity as one Possible Tool for Predicting of Cirsium Arvense Infestation Occurrence. Herbicides, Theory and Applications WWW.intechopen.com.

[14] L. Brodský, V. Vaněk, L. Borůvka, and J. Száková, Consistency of spatial dependence of soil chemical properties in two fields: a geostatistical study." Plant Soil Environ, 50(11) 2004 507-512.