MULTIPLE FACTORS AFFECTING THE SURFACE MODELING: INCORPORATE PATTERN ANALYSIS WITH INTERPOLATION TECHNIQUES

DR. AMR HANAFI AHMED ALI

Benha University
Faculty of Engineering-Shoubra
Surveying Engineering Department

تمثل عملية انتاج نماذج التضاريس الرقمية أهمية كبرى لمعظم التخصصات المختلفة، فالمهندس المدني يصمم ويشيد المباني علي سطح الأرض، والجيولوجي يقوم بدراسة ما بداخلها ؛ أما الجيوديسون يهتمون بشكل الأرض و كذلك الطبوغرافيين معنيين بقياس ووصف سطح الأرض وتمثيلها بطرق مختلفة فى شكل النموذج الرقمى للارتفاعات. وعلى الرغم من اختلاف وتنوع الاهتمام ، إلا انه تم الإجماع على أهمية إنتاج هذا النموذج بدقة عالية. وفى مجال الجيوديسيا، هناك العديد من العوامل التى تؤثر على النمذجة الرقمية الأرضية، حيث يظل أحد اهم هذه العوامل تقنيات الاستنتاج Interpolation المستخدمة فى إنتاج النموذج. وعلى الرغم من أهمية طرق الاستنتاج المستخدمة إلا ان الاعتماد عليها بمفردها قد يضلل المستخدمين نظرا لتعدد الطرق ولوجود عوامل أخرى قد تؤثر على النتيجة النهائية. لذلك فإن دراسة باقى العوامل المؤثرة على عملية تكوين النموذج الرقمى للارتفاعات يمكن يؤدى الى الحصول على تمثيل رقمى عالى الدقة. ويمثل دراسة نمط سلوك البيانات وتحليلها و التحقق من نظريات الإحصاءات المكانية و الإستنتاج أحد أهم العوامل المطلوب التعرف عليها، بجانب التقنيات المختلفة للإستنتاج.

ويهدف البحث اساسا الى اختبار مصداقية نتائج التمثيل ثلاثى الأبعاد للأسطح من خلال دراسة العوامل الرئيسية التي تؤثر على النتائج ، وهى تحليل نمط سلوك البيانات مع تقنيات الإستنتاج المختلفة. وعلى هذا، فلقد تم إقتراح منهجية الدراسة من خلال مرحلتين أساسيتين. أولا يتم دراسة سلوك البيانات من خلال نظريات تحليل النمط. وثانيا يتم تقييم بعض تقنيات الإستنتاج بإستخدام برمجيات نظم المعلومات الجغرافية والإحصاء ، والإحصاء الجغرافي لتطوير وتطبيق وتقييم المنهجيات المقترحة باستخدام دراسة حالات مختلفة.

ABSTRACT

People live on Earth and learn to cope with its terrain. Civil engineers design and construct buildings on it; geologists try to study its underlying construction; geomorphologists are interested in its shape and the processes by which the landscape was formed; and topographic scientists are concerned with measuring and describing its surface and presenting it in different ways. Digital terrain modeling is a process to obtain desirable models of the land surface. Despite these differences in emphasis and interest, these specialists have a common interest, that is, they wish the surface of the terrain to be represented conveniently and with confident accuracy.

Multiple factors affect the surface modeling. One of the major factors that affect the surface model is the applied interpolation technique. Most of the users are using the interpolation techniques regardless studying the pattern behaviors that mislead the user about the most proper interpolation technique for providing the surface. Therefore, theories of statistics and spatial interpolation utilized through 3D surface representation platforms by incorporating the spatial pattern analysis will be investigated.
The main objective of this research is to test the reliability of the results caused by the 3D surface representation platforms. Here, the main factors which affect on the results are examined, where the pattern analysis methods will be studied with different interpolation techniques. The applied methodology is divided into two main steps. Firstly, the pattern analysis will be studied. Secondly, different interpolation techniques will be evaluated. GIS, statistical, and geostatistical analysis software were used to develop, apply, and evaluate the proposed methodologies using different case studies.

KEYWORDS
DEM, DTM, Spatial Interpolation Techniques, Pattern analysis, and Geostatistics.

1. INTRODUCTION

Generally, spatial data is considered as an invaluable source of information to the society. It can be divided into planimetric maps (2D surface) that represents the natural and artificial features and the topographic maps, which represent also the relief or contours of the ground. The second form is the basis of the 3D representation.
By establishing a three-dimensional data grid using a set of points (X, Y, Z) at any common space, the surface modeling can be generated.

The pattern analysis and interpretation of any spatial datasets forms an important part of geostatistics where it is highly human dependent. For instance, it is well known that different individuals will take different approaches, yielding a large assortment of distinct solutions. It is often the case where judgment and experiences play a key role in selecting the proper spatial interpolation technique for each particular case. This is partly due to the variety of the available spatial interpolation methods, which range from simple intuitive predictions to more sophisticated and complex procedures [1].

The major factors that control the efficiency of the result are: - (1) The applied measuring technique, (2) The area of the site, (3) The topography, (4) The spatial interpolation technique, in other words the applied mathematical model to create the surface and (5) The Spatial Pattern analysis, which defines the quantitative methods, applied for describing and analyzing the distribution pattern of spatial data. Herein, the applied data points lies either in a regular or irregular (random) grid where the grid data values define the height (the third dimension Z-coordinate) above or below the X-Y plane. We are going to focus on the last three factors.

DEM is produced from different technologies such as Light Detection and Ranging (LiDAR) and Interferometric Synthetic Aperture Radar (IFSAR). These remotely-sensed methods provide users with high resolution DEM data that have vertical and horizontal accuracies in centimeters, making them more desirable, yet costly in both funding and processing requirements. For smaller study areas and limited budget, users can obtain DEMs by conducting field surveys using global positioning systems (GPS), and then interpolate DEM. No matter the source of gathering data, DEM products provide clear and detailed rendition of topography and terrain surfaces. These depictions can orient users into a false sense of results regarding the accuracy and precision of the data. Potential errors, and their effect on derived data and the based output on that data, are often far from user's consideration [2].

Herein, the attention is directed to the pattern analysis "density of points and their distribution" and the topography of the site with the different interpolation techniques in order to select the proper one that satisfy minimum uncertainty associated with DEM. Since locations of sample points are important for interpolation so, points should be located evenly over the area even if it is regularly or randomly spaced. The more the input points and the greater their distribution, the more reliable are the achieved results [3].

2. PATTERN ANALYSIS

Pattern analysis involves analyzing the arrangement of points in an area. It can reveal if the distribution pattern is random, dispersed, or clustered. Also, a pattern analysis can detect whether the distribution pattern contains clusters of high or low values. Analyzing and testing of pattern is performed via many methods, where the famous ones are Nearest Neighbor Analysis "NNA", Moran's I and G-Statistics [4].
Nearest Neighbour Analysis is a simple method that analyzes the spatial data distribution and depends only on the spatial data position. NNA uses the distance between each point and its closest neighbouring point in determining if the point pattern is random, regular or clustered [5]. NNA calculates the ratio of the observed average distance between nearest neighbours of a point distribution (d_{obs}) (i.e. that for each point, the shortest distance among all neighbors becomes the nearest distance, and then averaged using all points) to the expected average distance between nearest neighbors as determined by a theoretical pattern; the Poisson probability distribution (d_{exp}). It is denoted by the dimensionless statistic *R*, In its simplest form the nearest neighbor statistic, *R*, compares the observed, d_{obs}, with the expected, d_{exp} [6].

$$R = \frac{d_{obs}}{d_{\exp}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

If the (R) ratio is less than 1, the point pattern is considered more clustered than random where if it is greater than 1; the point pattern is biased to be more dispersed than random.

As Waldo Tobler notes, "I invoke the first law of geography: everything is related to everything else, but near things are more related than distant things" so, the correlation between the points (in 3-D) plays a major role of obtaining the proper results.

Moran's I is a method for analyzing point pattern, that is based on measuring spatial autocorrelation where, the analysis considers not only the points' position but also the variation of third dimension value where the spatial datasets are auto correlated [7].
Spatial autocorrelation therefore measures the relationship among values of variables according to the spatial arrangement of the values [8]. The relationship can be described as highly correlated if like values are spatially close to each other, and independent or random if no pattern is discerned from the arrangement of values. Moran's I is computed according to Equation (2).

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sigma^2 \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where x_i is the value at point i, x_j is the value at point i's neighbour j, n is the number of points, and σ^2 is the variance of x values with a mean of $\bar{x}$. The coefficient w_{ij} is the weight for measuring spatial autocorrelation. Typically, w_{ij} is defined as the inverse of the distance (d) between points i and j, or $1/d_{ij}$.The pattern is considered as random (i.e., not spatially correlated), if the index value I is 0, while positive I values indicate that adjacent points tend to have similar values (i.e. positive spatial autocorrelation) and negative ones indicate that adjacent points tend to have different values (negative spatial autocorrelation). Moran's I can only detect the presence of the clustering of similar values, but it cannot tell whether the correlation is strong or weak [5].
This has led to the use of the G-Statistics method, which can separate clusters according to their values. G-Statistics index which is based on a specified distance, d, can be defined as G(d) and is computed by using Equation(3)

$$G(d) = \frac{\sum \sum w_{ij}(d)_{x_i x_j}}{\sum \sum x_i x_j} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Where x_i is the value at position i, x_j is the value at position j that is laying within distance d of i, and $w_{ij}(d)$ is the spatial weight which is based on some weighted distance (e.g. inverse distance). G-Statistics is normally distributed method and can be standardized to facilitate its interpolation: the positive G values indicate cluster with high values "Strong autocorrelation", while negative values indicate clusters with low values "weak autocorrelation" [9].

3. SPATIAL INTERPOLATION TECHNIQUES

Spatial Interpolation is the process of using points with known values to estimate unknown values at other points. It is therefore a means of creating surface data from sample points where the surface data can be used for analysis and modelling. Spatial interpolation methods can be categorized in several ways. They can be grouped into global and local methods. A global interpolation method uses every available known point to estimate an unknown value. On the other hand, a local interpolation method uses a sample of known points to estimate an unknown value. Conceptually, a global interpolation method is designed to capture the general trend of the surface while a local interpolation method deals

with the local or short range variation. For many phenomena, it is more efficient to estimate the unknown value at a point using a local method than a global where the away points have little influence on the estimated value. Also, spatial methods can be grouped into exact and inexact interpolation. Exact interpolation predicts a value at the point location that is the same as its known value. So, exact interpolation generates a surface that passes through the control points. In contrast, inexact interpolation, predicts a value at the point location that differs from its known value. In case of elevations, the exact methods should be applied. Finally, the spatial interpolation methods can be deterministic or stochastic.
A deterministic interpolation method provides no assessment of errors with predicted value. On the other hand, a stochastic interpolation method, offers assessment of prediction errors with estimated variances. The assumption of a random process is normally required for a stochastic method [9].Table (1) shows a classification of spatial interpolation methods [5].

Table (1): classification of the major spatial interpolation techniques

Global		**Local**	
Deterministic	Stochastic	Deterministic	Stochastic
Trend surface (inexact)	Regression (inexact)	Density (inexact)	Kriging (exact)*
		Inverse distance weighted (exact)*	
		Splines (exact)*	

Since this research is focusing on the applied methods for creating DEM, the local exact methods should be used in interpolating elevation data where the contouring surfaces should pass the control points. The most applied ones are Splines, Inverse distance weighting and the Kriging techniques [10].

3.1. Splines

Splines is an interpolation method that estimates values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points. Therefore, this method is the best for generating gently varying surfaces such as elevation, water table heights, or pollution concentrations [4].
Splines interpolation consists of the approximation of a function by means of series of polynomials over adjacent intervals with continuous derivatives at the end-point of the intervals. Smoothing Splines interpolation enables to control the variance of the residuals over the dataset. The solution is estimated by an iterative process. It is also referred to the basic minimum curvature technique or thin plate interpolation as it possesses two main features: (a) the surface must pass exactly through the data points, and (b) the surface must have minimum curvature [1].
The approximation of thin-plate Splines is calculated according to Equation (4).

$$Q(x,y) = f_i = \sum_{i=1}^{n} A_i d_i^2 \log d_i + a + bx + cy \qquad \text{.............................} (4)$$

Where x and y are the coordinates of the point to be interpolated, x_i and y_i are the coordinates of control point i and $d_i^2 = (x-x_i)^2 + (y-y_i)^2$.
The component (a+ bx +cy) represents the local trend function. It has the same form as a linear or first-order trend surface and ($d_i^2 \log d_i$) represents a basis function, that is designed to obtain minimum curvature surfaces. The coefficients A_i, a, b, and c are determined by a

linear system of equations, where n is the number of control points, and f_i is the known value at control point i. Herein, the estimation of the coefficients requires n+3 simultaneous equations [5].

3.2. Inverse Distance Weighting (IDW)

Inverse Distance Weighting (IDW) is one of the simplest and most readily available methods. It is based on an assumption that the value at a predicted point can be approximated as a weighted average of values at points within a certain cut-off distance, or from a given number *m* of the closest points (typically 10 to 30) [11] where, weights are usually inversely proportional to a power of distance. The general equation for the IDW method is shown in Equation (5).

$$Z_O = \frac{\sum_{i=1}^{n} Z_i \frac{1}{d_i^k}}{\sum_{i=1}^{n} \frac{1}{d_i^k}} \qquad \text{.............................. (5)}$$

Where z_0 is the estimated value at point 0, z_i is the z value at known point i, d_i is the distance between point i and point 0, n is the number of known points used in estimation, and k is the specified power. The power k controls the degree of local influence. A power of 1.0 means a constant rate of change in value between points (linear interpolation), whiles the power of 2.0 or higher suggests that the rate of change in values is higher near a known point and the point levels are off away from it. The degree of local influence also depends on the number of known points used in estimation [12].

3.3. Kriging

Kriging is a geostatistical method for spatial interpolation that proves usefulness and popularity in many fields. Kriging is not deterministic but extends the proximity weighting approach of inverse distance to include random components where exact point location is not known by the function. Kriging depends on spatial and statistical relationships to calculate the surface [13]. It differs from the other interpolation methods as it can assess the quality of the prediction with estimated prediction errors. A unique aspect of geostatistics is the use of regionalized variables which are variables that neither totally random (stochastic) nor deterministic variables, where regionalized variables describe phenomena with geographical distribution (e.g. elevation of ground surface). In general, Kriging consists of three main terms: a drift that representing a trend; a spatially correlated term which representing the variation of the regionalized variable and a random error term. The interpretation of these three terms has led to different Kriging methods for spatial interpolation. Kriging uses the spatial autocorrelation or the semivariance $\gamma(h)$ which measures the spatially correlated component between known points, xi and xj, that separated by the distance h; and z is the third dimension value. $\gamma(h)$ is calculated by Equation (6)

$$\gamma(h) = \frac{1}{2}\left[z(x_i) - z(x_j)\right]^2 \qquad \text{.............................. (6)}$$

A semivariogram is a graph, which plots γ(h) against h for all pairs of known points in a data set. If spatial dependence does exit in a data set, known points that are close to each other expected to have small semivariances, and known points that are farther apart are expected to have large semivariances. To use semivariogram as interpolator in Kriging, it must fit with a mathematical model, where the fitted semivariogram can be used for estimating the semivariance at any given distance. A fitted semivariogram can be divided into three possible elements: nugget, range, and sill. Nugget defines the semivariance at the distance of 0, while range represents the distance at which semivariance starts to level off, and sill is the value of the semivariance at the end of the range. See figure (1) [14].

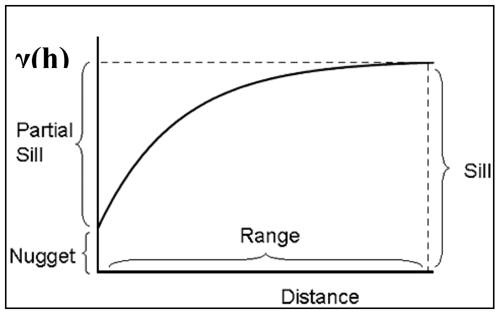

Figure (1): Nugget, Range, and Sill

The most applicable method is the Ordinary Kriging which assumed the absence of a drift. It focuses on the spatially correlated term and uses the fitted semivariogram directly for interpolation. Equation (7) is applied for estimating the z value at any point.

$$z_0 = \sum_{i=1}^{n} z_x w_x \qquad \text{.............................. (7)}$$

Where z_0 is the estimated value, z_x is the known value at point x, w_x is the weight associated with point x, and n is the number of sample points. The weight can be derived by solving a set of simultaneous Equation (8)

$$\sum_{i=1}^{n} \sum_{j=1}^{n} w_i \gamma(h_{ij}) + \lambda = \sum_{i=1}^{n} \gamma(h_{io}) \qquad \text{.............................. (8)}$$

Where $\gamma(h_{ij})$ is the semivariance between known points i, j, $\gamma(h_{i0})$ is the semivariance between the ith known points and the estimated point *0*. λ is a Lagrange multiplier that added to ensure the minimum possible estimation. Applying the above equation, the Kriging produces a variance measure for each estimated point to indicate the reliability of estimation where the variance estimation σ^2 is calculated using Equation (9) [5].

$$\sigma^2 = \sum_{i=1}^{n} w_i \gamma(h_{i0}) + \lambda \qquad \text{.............................. (9)}$$

To assess the accuracy of the above interpolation techniques, statistical procedures should be applied to confirm the consistency of each method.

4. VALIDATION TECHNIQUE

The interpolation accuracy can be measured by different methods, where the most straightforward one is to evaluate deviations between interpolated surface and the input points [15].

The major applied technique for this approach is the cross validation. Cross validation assess the accuracy of the interpolation methods by repeating the following procedure:

1. Remove a known point from the dataset.
2. Use the remaining points to estimate the value at the point previously removed.
3. Calculate the predicted error of the estimation by comparing the estimated with the known value.

After completing the procedure for each known point, one can calculate a diagnostic statistics to evaluate the accuracy of the interpolation method. This diagnostic statistics is the root mean square (RMS) as expressed in Equation (10)

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_{i,act} - z_{i,est})^2} \qquad \text{.............................. (10)}$$

Where n is the number of points, $z_{i,act}$ is the known value of point i, , $z_{i,est}$ is the estimated value of point i.

The interpolation of statistics is based on the rule “A better interpolation method should yield to a smaller RMS”. By extension, an optimal method should have the smallest RMS [16].

This particular form of the cross validation is especially suitable for relatively dense datasets, since removing points from already under-sampled areas can lead to misrepresentation of the surface to be interpolated. Despite the wide use of this technique for assessing interpolation schemes, one should be aware of its shortcomings. Specifically, cross validation will usually overestimate the interpolation error because the estimate is being computed at a position where data are actually available. In addition, the computed surface and hence the cross-validated estimate may be altered by the removal of the point being cross-validated. In practice these issues are unavoidable but with increasing number of input data points they have less impact [17].

Evaluating the predictive accuracy between points can be done by using a check dataset that contains un-used data in the interpolation. For each check point the deviation between actual and interpolated value is calculated and the overall accuracy is tested. The diagnostic statistic of RMS derived from the check dataset can then be used to assess the accuracy of the methods. However, in many applications and due to the limited number of input points, it becomes difficult to select independent evaluation dataset. Moreover, the accuracy information is available only for these independent points and they rarely cover the entire area of interest with a sufficient density [15].

5. DATA USED

For the computation process in this research, the following scenario has been followed. Sample sets of data that cover about 4 km^2 were prepared with variable topographic feature. The data derived from 4 raster DEMs with resolution 5 m, where these DEMs represent the ground truth (variability) of surface with high conformity. Each DEM is considered as a unique dataset that represents different topographic factor ranging from smooth topographic surface to complex topographic one. Each dataset is divided into two groups according to the distribution of points. A spatial resolution of the DEM determines its application properties and ability to represent the terrain features in a desired detail [18]. Therefore, each dataset samples were chosen from the original raster DEMs with resolution 25, 50 and 100 meters, respectively. They are divided as:

- Group (1) which represents a regular distribution region, where three datasets were arranged to cover different densities, the average spacing among points are (25, 50, and 100 meters respectively).
- Group (2) represents an irregular distribution region, where another three datasets were prepared to cover the same densities, the average spacing among points are (25, 50, and 100 meters respectively).

5.1. Dataset 1

The distinguished characteristics of this dataset that it is a flat surface with smooth slope where the difference between maximum and minimum elevations is about 1.5 m.

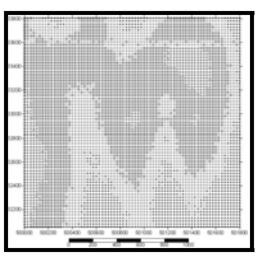

Fig (2): Dataset 1
Group (1) - Case (1)

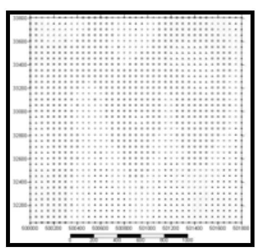

Fig (3): Dataset 1
Group (1) - Case (2)

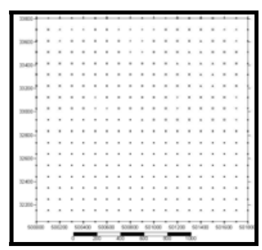

Fig (4): 1
Group (1) - Case (3)

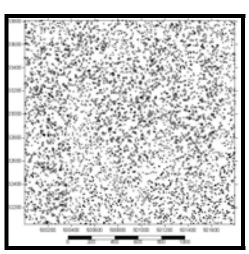

Fig (5): Dataset 1
Group (2) - Case (1)

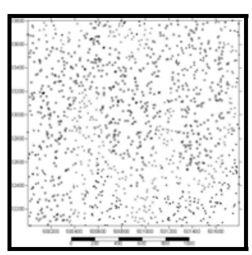

Fig (6): Dataset 1
Group (2) - Case (2)

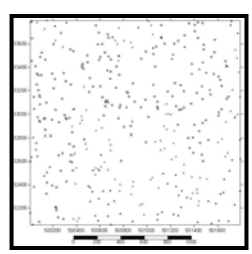

Fig (7): Dataset 1
Group (2) - Case (3)

5.2. Dataset 2

This dataset represents relatively flat area with gentle slopes where elevations differ from each other by about 48 m.

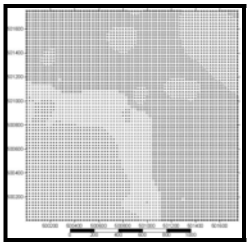

Fig (8): Dataset 2
Group (1) - Case (1)

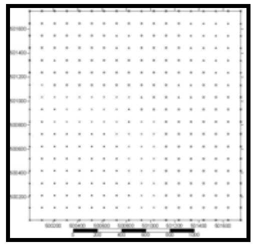

Fig (9): Dataset 2
Group (1) - Case (2)

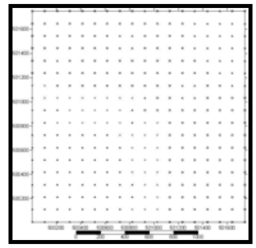

Fig (10): Dataset 2
Group (1) - Case (3)

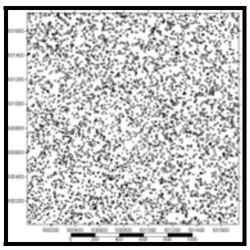

Fig (11): Dataset 2
Group (2) - Case (1)

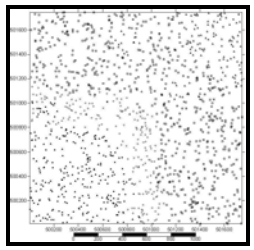

Fig (12): Dataset 2
Group (1) - Case (2)

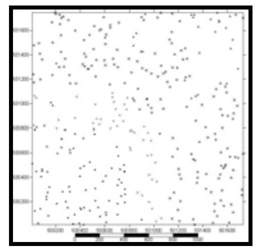

Fig (13): Dataset 2
Group (1) - Case (3)

5.3. Dataset 3

Herein, the topographic feature became much variable than before where elevations differ from each other by about 140 m.

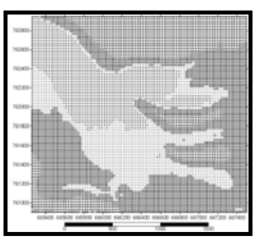

Fig (14): Dataset 3
Group (1) - Case (1)

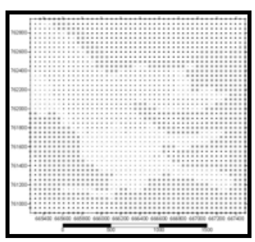

Fig (15): Dataset 3
Group (1) - Case (2)

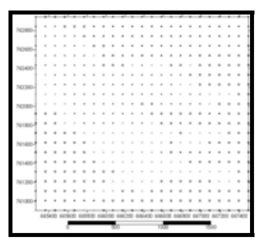

Fig (16): Dataset 3
Group (1) - Case (3)

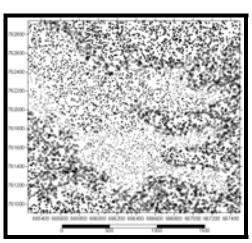

Fig (17): Dataset 3
Group (2) - Case (1)

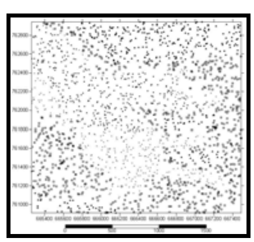

Fig (18): Dataset 3
Group (2) - Case (2)

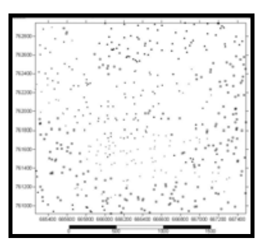

Fig (19): Dataset 3
Group (2) - Case (3)

5.4. Dataset 4

This dataset represents mountains area with steep slope where difference in elevations between minimum and maximum reaches 500 m.

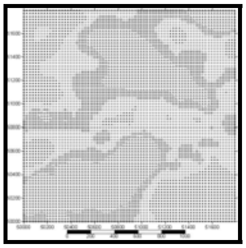

Fig (20): Dataset 4 Group (1) - Case (1)

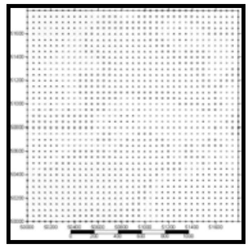

Fig (21): Dataset 4 Group (1) - Case (2)

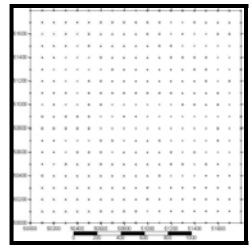

Fig (22): Dataset 4 Group (1) - Case (3)

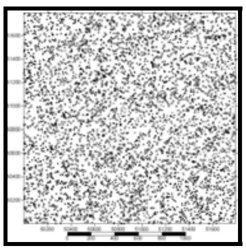

Fig (23): Dataset 4 Group (2) - Case (1)

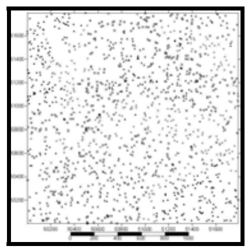

Fig (24): Dataset 4 Group (2) - Case (2)

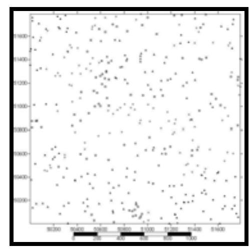

Fig (25): Dataset 4 Group (2) - Case (3)

6. METHODOLOGY

Since, the objective of this research is to study the major factors affecting the surface modeling; the pattern analysis methods will be studied with different interpolation techniques. The applied methodology is divided into two main steps.

- Firstly, the pattern analysis will be studied for each dataset. This will include the two groups with the different 3 cases of data, through the following methods, Nearest Neighbor analysis, Moran'I analysis and the G-Statistics analysis methods, where an indicator will be implemented to classify if the observed and random distribution is statistically significant or not (in our case study, a Z-score is calculated). If the score changes between -1.96 and +1.96 (confidence level of 95%), it indicates that there is no significant difference between the observed and random distribution statistically in spite of the result of indexes (R, I, and G) pattern indication. However, the greater amount of Z-score shows significant difference between observed and random distribution. As this Z-score expresses the divergence of the experimental result *Ind.* from the most probable result *E(Ind.)* as a number of standard deviations σ, so Z-score can be calculated according to equation (11) [19]

$$Z = \frac{Ind. - E(Ind.)}{\sigma} \qquad \text{……………..………… (11)}$$

- Secondly, different interpolation techniques will be examined for each dataset, where Splines, Inverse Distance Weighted (IDW) and Kriging methods are selected according to their suitability with elevation data as they represent the exact methods.
  Recalling that, RMS is the value which validates the interpolation methods with each dataset, where two methods were applied. Cross Validation that provides RMS for input data points and the other by using evaluation datasets as 100 points was prepared at each dataset to represent this evaluation dataset.

7. RESULTS AND ANALYSIS

7.1. First Step: Pattern Analysis

The statistics values of the four datasets are tabulated. The main items in the tables are the index value for each pattern analysis methods with their corresponding Z-score, and pattern description, where the value R will represent the index of nearest neighbor analysis method. If R is equal to 1, it indicates random pattern; if R is less than 1, the pattern exhibits clustering; if R is greater than 1, the trend is toward regular distribution pattern. Recalling that Z-Score confirmed this statistic index, where for Z- score values between -1.96 and +1.96 indicates random, if it is greater than 1.96, it confirmed regular distribution, and less than -1.96 indicates clustered pattern.
With respect to Moran's I method, the term I will represent the index; if I is equal 0, it means that there is no correlation between points; if I > 0, it directed to positive autocorrelation between points, while it implies negative autocorrelation, if the term I < 0. About the related Z-score, if it changes between -1.96 and 1.96, it indicates no correlation, if Z-score > 1.96, it indicates positive autocorrelation, and if Z-score < -1.96, it indicates negative autocorrelation.
Related to G-statistic method, G value represent the index associated also with Z-score, where for G is equal 0 and Z-score changes between -1.96 and +1.96, it indicates no correlation, if G >0 and Z-score > 1.96, it indicates strong correlation, and if G < 0 and Z-score < -1.96, it indicates weak correlation [19].

7.1.1. Dataset 1 Results

The results of dataset 1 with all cases (from case 1 to case 3) of its 2 groups are tabulated.
R, I, and G indexes as well as their Z-scores are computed and the pattern descriptions are summarized.

Table (2): Results of 3 pattern analysis methods for all cases of dataset 1

		NNA		Moran's I		G-Statistics		Pattern
		R	Z-Score	I	Z-Score	G	Z-Score	
Group (1)	Case (1)	2.03	141.61	0.20	589.69	0.0015	14.09	Regular of strong (+) correlation
	Case (2)	2.06	73.75	0.19	165.93	0.0015	7.50	Regular of strong (+) correlation
	Case (3)	2.08	39.14	0.17	47.22	0.0013	-1.90	Regular of weak (+) correlation

		NNA		Moran's I		G-Statistics		Pattern
		R	Z-Score	I	Z-Score	G	Z-Score	
Group (2)	Case (1)	1.01	1.64	0.23	425.51	0.0017	11.48	Random of strong (+) correlation
	Case (2)	0.98	1.39	0.23	101.80	0.0017	6.96	Random of strong (+) correlation
	Case (3)	0.99	-0.53	0.20	30.12	0.0017	-1.80	Random of weak (+) correlation

Within the first dataset, by applying nearest neighbor analysis method, the output values describe the pattern as regular one for group (1) and irregular for group (2). On the other hand, by using Moran's I analysis method, the points are described as positively correlated points. Finally, the G-Statistics method is performed to identify any strong or weak correlation. Within our test, this dataset is defined as strong correlated data points for case 1 and 2 and weak correlated data points for case 3, where this situation expected to be raised due to the low density of points

7.1.2. Dataset 2 Results

By applying the 3 methods for dataset 2 that is relatively flat area. The results of dataset 2 will be tabulated. R, I, and G indexes as well as their Z-scores are computed. Pattern description summarized.

Table (3): Results of 3 pattern analysis Methods for all cases of dataset 2

		NNA		Moran's I		G-Statistics		Pattern
		R	Z-Score	I	Z-Score	G	Z-Score	
Group (1)	Case (1)	2.00	136.38	0.37	1045.44	0.0016	5.77	Regular of strong (+) correlation
	Case (2)	2.04	71.44	0.36	303.22	0.0015	2.73	Regular of strong (+) correlation
	Case (3)	2.11	38.17	0.34	86.60	0.0013	1.21	Regular of (+) correlation , neither strong nor weak
Group (2)	Case (1)	1.01	1.42	0.38	691.77	0.0017	6.04	Random of strong (+) correlation
	Case (2)	1.03	1.94	0.40	183.40	0.0017	2.01	Random of strong (+) correlation
	Case (3)	1.03	0.87	0.41	45.84	0.0017	-2.90	Random of weak (+) correlation

The results show that group (1) described as regular according to nearest neighbor analysis method, positive correlation related to Moran's I method, and strong correlated data with respect to G-statistics except case (3) that described as random behavior in other words, "neither strong nor weak" . For group (2), the nearest neighbor analysis indicate random pattern, Moran's I point to positive correlation and G-statistics describe strong correlation except case (3) that described as weak correlation.

7.1.3. Dataset 3 Results

Executing the 3 methods for dataset 3 which is more complicated topographic surface, the results of dataset 3 will be tabulated. Indexes associated with their Z-scores are computed. As well as pattern description summarized.

Table (4): Results of 3 pattern analysis Methods for all cases of dataset 3

		NNA		Moran's I		G-Statistics		Pattern
		R	Z-Score	I	Z-Score	G	Z-Score	
Group (1)	Case (1)	2.02	167.83	0.21	829.17	0.0013	-24.95	Regular of weak (+) correlation
	Case (2)	2.05	87.90	0.20	236.58	0.0012	-12.78	Regular of weak (+) correlation
	Case (3)	2.09	45.90	0.18	64.60	0.0011	-6.28	Regular of (+)weak correlation
Group (2)	Case (1)	1.01	1.11	0.22	533.95	0.0014	-19.52	Random of (+)weak correlation
	Case (2)	1.01	0.83	0.22	139.29	0.0014	-10.50	Random of weak (+) correlation
	Case (3)	1.05	2.08	0.20	46.50	0.0014	-6.71	Random of weak (+) correlation

Nearest neighbor analysis indicate regular pattern for group (1), random one for group (2), Moran's I notify positive correlation for the two groups, and finally G-Statistics imply weak correlation for the whole cases.

7.1.4. Dataset 4 Results

Again, applying the 3 methods for dataset 4 that represents the complex topography, the results of dataset 4 will be tabulated where indexes associated with their Z-scores are computed. As well as pattern description summarized.

Table (5): Results of 3 pattern analysis Methods for all cases of dataset 4

		NNA		Moran's I		G-Statistics		Pattern
		R	Z-Score	I	Z-Score	G	Z-Score	
Group (1)	Case (1)	2.01	140.66	0.19	576.42	0.0015	29.27	Regular of strong (+) correlation
	Case (2)	2.04	73.46	0.18	160.06	0.0015	14.52	Regular of strong (+) correlation
	Case (3)	2.10	39.94	0.15	43.68	0.0013	-1.95	Regular of weak (+) correlation
Group (2)	Case (1)	1.02	2.19	0.20	404.07	0.0017	30.26	Random of strong (+) correlation
	Case (2)	1.01	0.55	0.21	108.74	0.0017	14.04	Random of strong (+) correlation
	Case (3)	1.02	0.88	0.22	35.34	0.0013	-1.80	Random of weak (+) correlation

For this dataset the same indication recorded as dataset 1.
To conclude all the previous datasets, figures (26) and (27) show the R index value and the Z-score respectively of the nearest neighbor analysis method for the 4 datasets with their 3 cases of group (1). Figures (28) and (29) plot the Moran's I index value and the Z-score correspondingly for the same datasets "group (1)". Figures (30) and (31) show the index of G-statistics and Z-score for group (1) of the 4 datasets. Figures (32- 37) illustrate the equivalent figures for group (2).

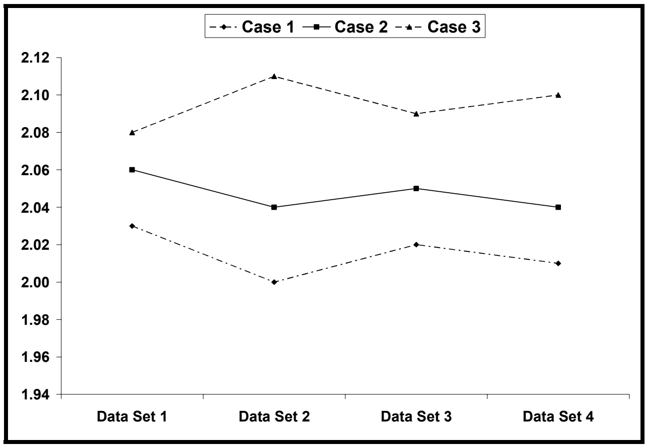

Fig. (26): Group (1)
NNA "Ratio Value"

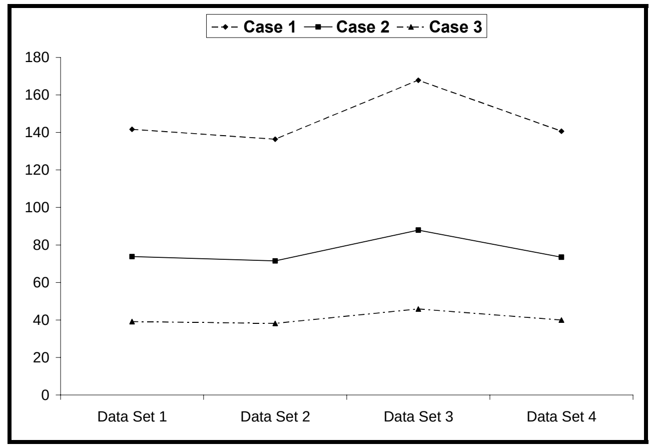

Fig. (27): Group (1)
NNA "Z-Score Value"

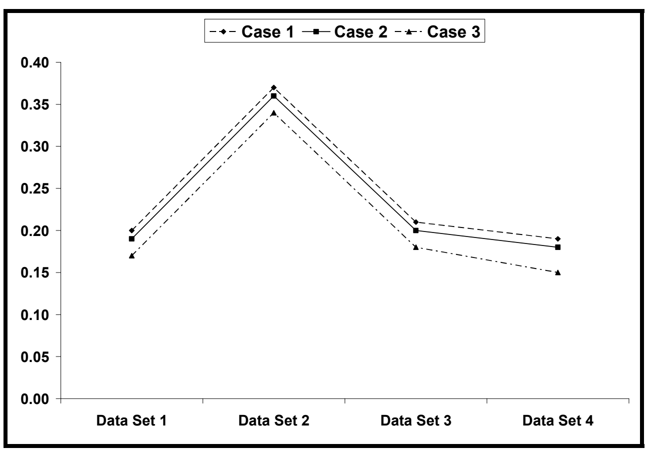

Fig. (28): Group (1)
Moran's I "Ratio Value"

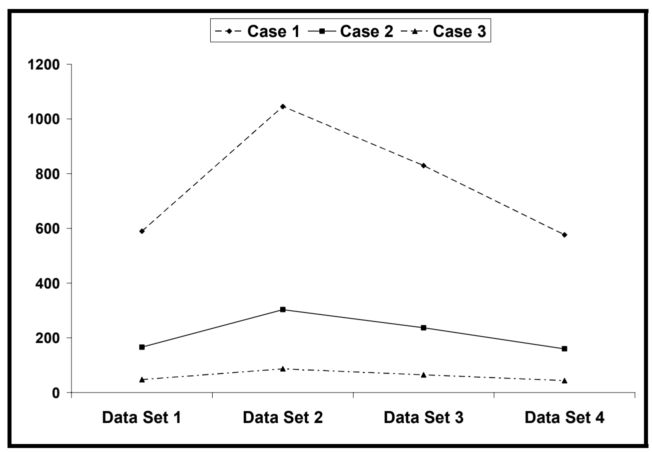

Fig. (29): Group (1)
Moran's I "Z-Score Value"

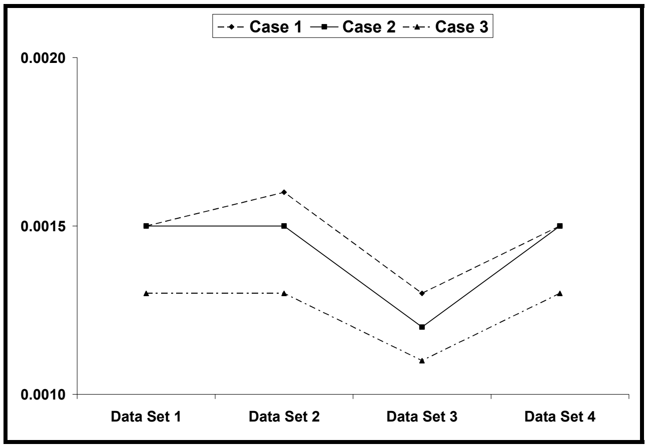

Fig. (30): Group (1)
G-Statistics “Ratio Value”

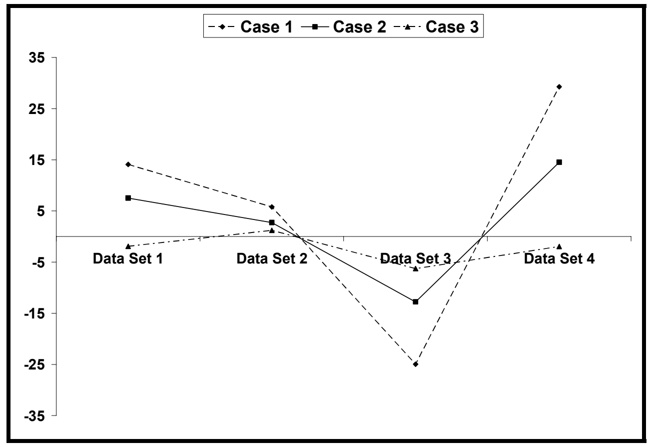

Fig. (31): Group (1)
G-Statistics “Z-Score Value”

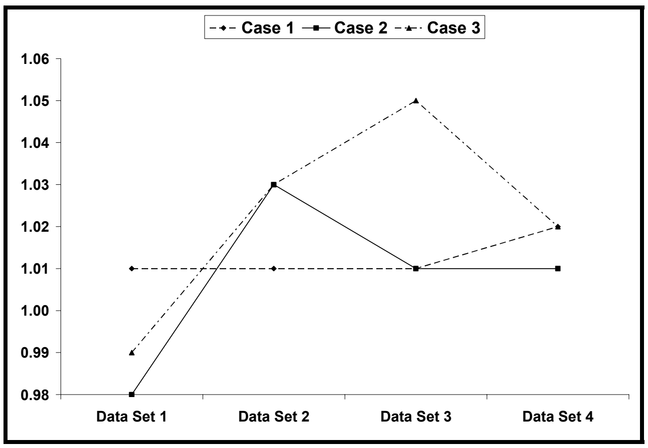

Fig. (32): Group (2)
NNA “Ratio Value”

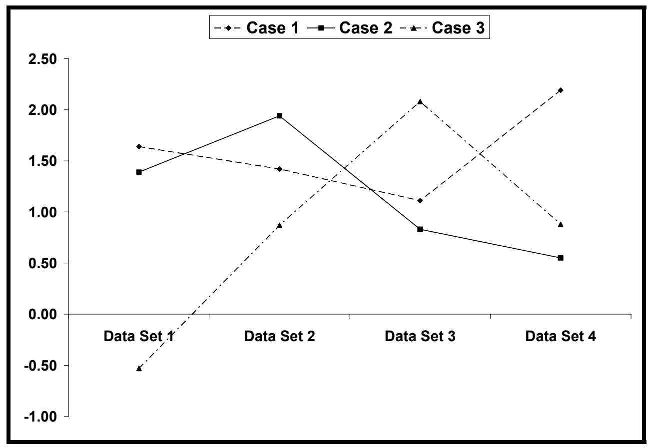

Fig. (33): Group (2)
NNA “Z-Score Value”

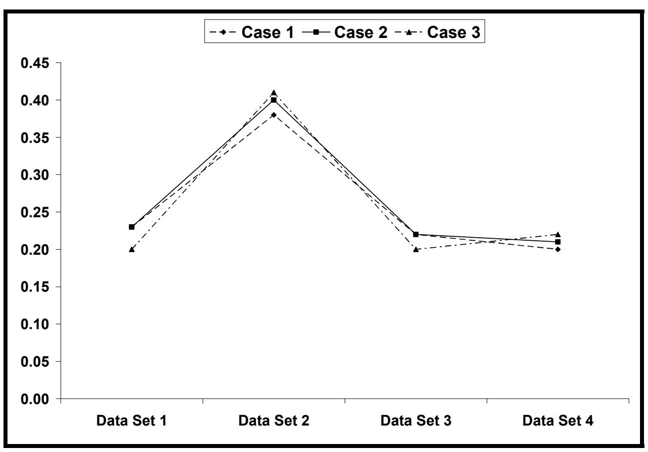

Fig. (34): Group (2)
Moran’s I “Ratio Value”

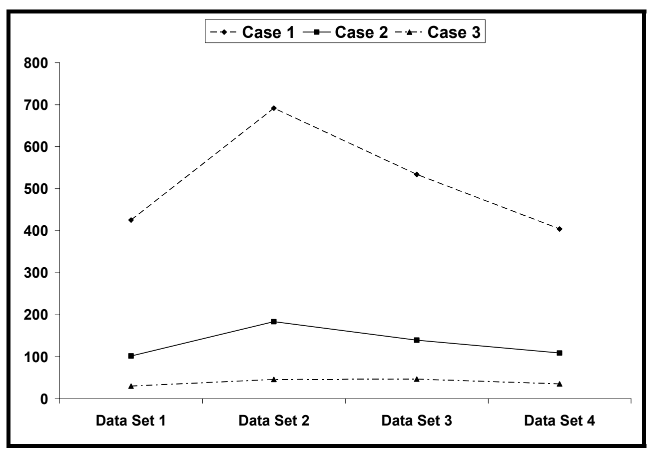

Fig. (35): Group (2)
Moran’s I “Z-Score Value”

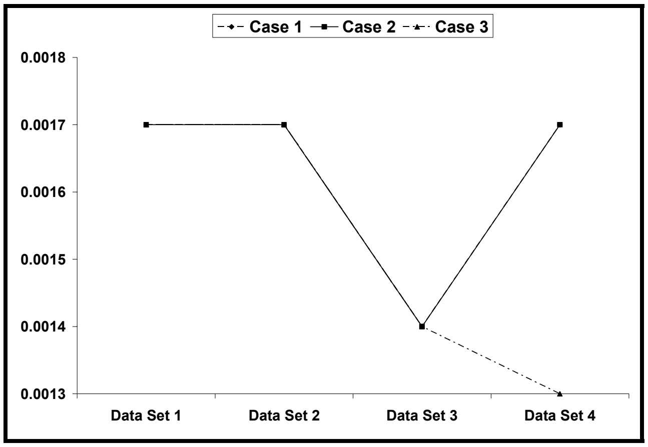

Fig. (36): Group (2)
G-Statistics "Ratio Value"

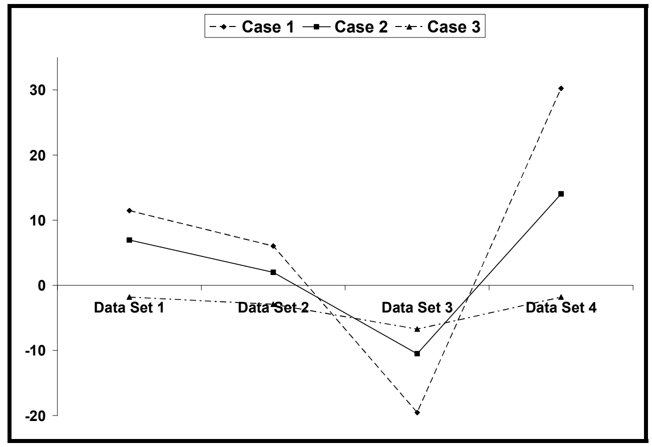

Fig. (37): Group (2)
G-Statistics "Z-Score Value"

7.2. Second Step: Comparing Interpolation Techniques

Locations of sample points are important for interpolation. Ideally, for mapping, points should be located evenly over the area. However, samples can be regularly or randomly spaced. More the input points and greater their distribution, more reliable of the results can be achieved [4]. To assess the interpolation techniques suitability, the samples at different factors of topography, density and distribution are taken and evaluated. The RMS value calculated for each case using input data points and check points, where 100 check points are generated for each dataset. The overall results are compared in order to find the minimum statistical errors.

7.2.1. Dataset 1 Results

For the first dataset with the regular distribution samples, the results of the three interpolation techniques are almost the same. All the 25 m, 50 m, and 100 m samplings fit the surface trend well by applying the Splines, IDW, Kriging interpolation techniques, where RMS values is very small and close to each other. The similar results were obtained from analyzing the random samples. This reflects that for flat surface with smooth slope, the three interpolation techniques can fit the surface well regardless the number and distribution of sample points. However Kriging is relatively better in both weak and strong correlated cases. Table (6) summarizes the results.

To confirm the results, the RMS values for the 100 check points were calculated. They provide results as good as the recorded input data, where each of the three interpolation techniques fits the surface well. Also, Kriging is the best in all cases. Table (7) reviews the results of the dataset-1 check points.

Table (6): Results of 3 interpolation techniques for all cases of dataset 1 "input data points"

	RMS in m					
	Group (1)			Group (2)		
	Case (1)	Case (2)	Case (3)	Case (1)	Case (2)	Case (3)
Splines	0.036	0.042	0.069	0.015	0.031	0.042
IDW	0.039	0.046	0.073	0.020	0.029	0.056
Kriging	0.035	0.041	0.052	0.013	0.028	0.042
Pattern description	Regular High density + Strong correlation ΔH=1.5 m	Regular Medium density + Strong correlation ΔH=1.5 m	Regular Low density + Weak correlation ΔH=1.5 m	Irregular High density + Strong correlation ΔH=1.5 m	Irregular Medium density + Strong correlation ΔH=1.5 m	Irregular Low density + Weak correlation ΔH=1.5 m

Table (7): Results of 3 interpolation techniques for all cases of dataset 1 "check points"

	RMS in m					
	Group (1)			Group (2)		
	Case (1)	Case (2)	Case (3)	Case (1)	Case (2)	Case (3)
Splines	0.124	0.484	1.352	0.012	0.035	0.041
IDW	0.389	1.010	2.189	0.021	0.028	0.054
Kriging	0.164	0.551	1.641	0.012	0.024	0.039
Pattern description	Regular High density + Strong correlation ΔH=1.5 m	Regular Medium density + Strong correlation ΔH=1.5 m	Regular Low density + Weak correlation ΔH=1.5 m	Irregular High density + Strong correlation ΔH=1.5 m	Irregular Medium density + Strong correlation ΔH=1.5 m	Irregular Low density + Weak correlation ΔH=1.5 m

7.2.2. Dataset 2 Results

For the second dataset, it is observed that the RMS values of case (3) represent relatively higher values. The RMS values of Splines and Kriging techniques are close to each other, where Splines technique gave relatively better results using input data points. Results tabulated at table (8). While table (9) represents the obtainable RMS for the check points. The RMS values of Splines and Kriging techniques are close to each other, where Kriging technique gave relatively better results for regular distribution, strong correlated cases. However Splines gives the best result for weak correlated case. For irregular distribution, Splines is a slightly better than the other techniques in both strong and weak correlated cases. Table (9) summarizes the results.

Table (8): Results of 3 interpolation techniques for all cases of dataset 2 "input data points"

	RMS in m					
	Group (1)			Group (2)		
	Case (1)	Case (2)	Case (3)	Case (1)	Case (2)	Case (3)
Splines	0.124	0.484	1.352	0.071	0.273	1.166
IDW	0.389	1.010	2.189	0.325	0.799	1.869
Kriging	0.164	0.551	1.641	0.123	0.421	1.386

	RMS in m					
	Group (1)			Group (2)		
	Case (1)	Case (2)	Case (3)	Case (1)	Case (2)	Case (3)
Pattern description	Regular High density + Strong correlation ΔH=48 m	Regular Medium density + Strong correlation ΔH= 48 m	Regular Low density + Random correlation ΔH= 48 m	Irregular High density + Strong correlation ΔH= 48 m	Irregular Medium density + Strong correlation ΔH= 48 m	Irregular Low density + Weak correlation ΔH= 48 m

Table (9): Results of 3 interpolation techniques for all cases of dataset 2 “check points”

	RMS in m					
	Group (1)			Group (2)		
	Case (1)	Case (2)	Case (3)	Case (1)	Case (2)	Case (3)
Splines	0.035	0.163	0.502	0.080	0.325	0.830
IDW	0.253	0.293	1.461	0.339	0.634	1.516
Kriging	0.018	0.110	0.839	0.156	0.414	0.926
Pattern description	Regular High density + Strong correlation ΔH=48 m	Regular Medium density + Strong correlation ΔH= 48 m	Regular Low density + Random correlation ΔH= 48 m	Irregular High density + Strong correlation ΔH= 48 m	Irregular Medium density + Strong correlation ΔH= 48 m	Irregular Low density + Weak correlation ΔH= 48 m

7.2.3. Dataset 3 Results

For the third dataset, the RMS values of Splines technique gave relatively better results in all cases. Results are tabulated in table (10). This outcome is confirmed when using the check points. The results have the same behavior. The irregular RMS values “Group 2” is less than the regular one “Group 1”. For both regular and irregular distribution, the best results are obtained from Splines technique. Table (11) summarizes the results.

Table (10): Results of 3 interpolation techniques for all cases of dataset 3 “input data points”

	RMS in m					
	Group (1)			Group (2)		
	Case (1)	Case (2)	Case (3)	Case (1)	Case (2)	Case (3)
Spline	0.866	2.641	6.356	0.253	0.685	2.064
IDW	2.320	5.195	9.323	0.642	1.565	4.101
Kriging	1.717	4.094	8.042	0.658	1.462	3.055
Pattern description	Regular High density + Weak correlation ΔH=140 m	Regular Medium density + Weak correlation ΔH= 140 m	Regular Low density + Weak correlation ΔH= 140 m	Irregular High density + Weak correlation ΔH= 140 m	Irregular Medium density + Weak correlation ΔH= 140 m	Irregular Low density + Weak correlation ΔH= 140 m

Table (11): Results of 3 interpolation techniques for all cases of dataset 3 “check points”

	RMS in m					
	Group (1)			Group (2)		
	Case (1)	Case (2)	Case (3)	Case (1)	Case (2)	Case (3)
Splines	1.251	2.015	4.704	0.541	1.566	3.155
IDW	2.030	3.851	7.002	1.266	2.352	5.120
Kriging	1.822	2.991	5.868	0.864	1.922	4.075
Pattern description	Regular High density + Weak correlation ΔH=140 m	Regular Medium density + Weak correlation ΔH= 140 m	Regular Low density + Weak correlation ΔH= 140 m	Irregular High density + Weak correlation ΔH= 140 m	Irregular Medium density + Weak correlation ΔH= 140 m	Irregular Low density + Weak correlation ΔH= 140 m

7.2.4. Dataset 4 Results

For the fourth dataset, the Splines technique is the best interpolation techniques in all cases. Group 2 which represent the irregular distribution gave smaller RMS values than Group 1 which define the regular distribution. The results tabulated in table (12).
Within the check points procedure, almost the same behavior were recorded, where Splines considers the best interpolation technique in all cases. Herein, the RMS of regular distribution cases “group 1” is smaller than the RMS of irregular distribution cases “group 2”. As a final conclusion Splines is the best techniques in both cases at input data points and at check points. The results summarizes at table (13).

Table (12): Results of 3 interpolation techniques for all cases of dataset 4 “input data points”

	RMS in m					
	Group (1)			Group (2)		
	Case (1)	Case (2)	Case (3)	Case (1)	Case (2)	Case (3)
Splines	1.29	4.64	15.22	0.77	3.80	11.24
IDW	4.45	12.31	29.35	4.27	10.43	22.53
Kriging	4.22	10.27	24.39	3.31	8.86	15.03
Pattern description	Regular High density + Strong correlation ΔH=500 m	Regular Medium density + Strong correlation ΔH=500 m	Regular Low density + Weak correlation ΔH=500 m	Irregular High density + Strong correlation ΔH=500 m	Irregular Medium density + Strong correlation ΔH=500 m	Irregular Low density + Weak correlation ΔH=500 m

Table (13): Results of 3 interpolation techniques for all cases of dataset 4 “check points”

	RMS in m					
	Group (1)			Group (2)		
	Case (1)	Case (2)	Case (3)	Case (1)	Case (2)	Case (3)
Splines	0.28	1.31	6.16	0.79	3.18	10.07
IDW	2.01	4.70	14.19	4.37	9.56	22.51
Kriging	2.20	5.61	14.02	2.53	7.09	13.84
Pattern description	Regular High density + Strong correlation ΔH=500 m	Regular Medium density + Strong correlation ΔH=500 m	Regular Low density + Strong correlation ΔH=500 m	Irregular High density + Strong correlation ΔH=500 m	Irregular Medium density + Strong correlation ΔH=500 m	Irregular Low density + Strong correlation ΔH=500 m

The following table summarizes the above results and defines the proper interpolation technique for different alternatives.

Table (14): Summary of the Results

	Regular		Irregular	
	Strong pattern	Weak pattern	Strong pattern	Weak pattern
Smooth (ΔH=0 - 5 m)	Kriging	Kriging	Kriging	Kriging
More Complicated (ΔH=5 - 150 m)	Kriging	Spline	Spline	Spline
Complex (ΔH=> 150 m)	Spline	Spline	Spline	Spline

8. CONCLUSIONS

This research is prepared to test the major factors that control the efficiency of surface modeling in order to select the proper interpolation technique for representing the elevation data. Herein, some of these factors are examined, which comprised of two main parts. The first part is the pattern analysis that evaluated the distribution and density of data points, where three major methods are performed for fulfillment. They are NNA; Moran’s I, and G-statistics. The second one is the applied interpolation techniques to create the surface, where three well-known interpolation techniques are examined; Splines, IDW, and Kriging. These two stages applied for four different datasets, which represents various topographic factors ranging from smooth surface to complex topographic one. An experimental approach was followed which allowed the relative performances of the three interpolators to be compared under different conditions of data density, distribution and topographic surface. Recalling that the assessment of the results depends on RMS statistics value computed from check points rather than input points.

Based on RMS statistical values, the following can be concluded:

- The results were consistent with expectations of the knowledge of sampling theory in that the accuracy of interpolation generally improved with sample point’s density; regular distributions of sample points usually captured more accurately the trend in strong positive correlated data sets; and irregular distributions produced slightly more accurate results when the sample data were spatially positive correlated with random or weak correlation.

- In case of smooth topographic surface, the 3 interpolation techniques can fit the surface well regardless the number and distribution of sample points. However the best results in both strong and weak correlated cases come from Kriging method.
- For more complex surface, Splines and Kriging gave relatively better results, however, Kriging acts well with regular point's distribution in case of strong correlated points in spite Splines gave relatively better results in case of weak correlation. For irregular distribution in both strong and weak correlated cases Splines represents the best method.
- For the mountain area with steep slopes, Splines method produced smoothest surface, which fits well in all cases.

REFERENCES

1. Naoum S.and Tsanis I.K.:"Ranking Spatial Interpolation Techniques using a GIS-based DSS", The Global Network for Environmental Science and Technology: the International Journal Volume 6 No.1, pp 1-20, Copyright© 2004 Global NEST, Printed in Greece.
2. S. P. Wechsler,"Uncertainties associated with digital elevation models for hydrologic applications: a review", Hydrology and Earth System Sciences, 11, 1481–1500, 2007. Copernicus Publications on behalf of the European Geosciences Union, www.hydrol-earth-syst-sci.net/11/1481/2007/
3. Taher Buyong, "Spatial Data Analysis for Geographical Information System", Universiti Teknologi Malaysia Press, 2007.
4. Kulapramote Prathumchai and Lal Samarakoon, "Elevations Surface Interpolation of Point Data Using Different Techniques – A GIS Approach" Asian Conference on Remote Sensing (ACRS), Ulaanbaatar, Mongolia, 2006
5. Kang-tsung Chang, "Introduction to Geographic Information System", third edition, University of Idaho, Copyright © 2006, 2004, 2002 by the McGraw-Hill Companies, Inc.
6. Bishop, M. A., "Comparative Point Pattern Analysis of Hyperboreae Undae, Mars, and the Rub' al Khali Sand Sea, Earth", Planetary Dunes Workshop: A Record of Climate Change, held April 29-May 2, 2008 in Alamogordo, New Mexico. LPI Contribution No. 1403, p.13-14.
7. Suleyman Demirci, ABD, and I-Shian, Suen "Spatial Pattern Analysis of PKK-KONGRA GEL Terror Incidents in Turkey: 2003-2004" In Ozgur Nikbay, Suleyman Hancerli, "Understanding and responding to the terrorism phenomenon", IOS Press, 2007.
8. Francis Harvey, "A Primer of GIS: Fundamental Geographic and Cartographic Concepts", The Guilford Press, New York London, 2008.

9. Marie-Josée Fortin, Mark R. T. Dale, “Spatial Analysis: a Guide for Ecologist”, Cambridge University Press, UK, 2005.
10. Spencer Chainey and Neil Stuart, “Stochastic simulation: an alternative interpolation technique for digital geographic information”. In Steve Carver, “Innovations in GIS 5: Selected Papers from the Fifth National Conference on GIS Research UK (GISRUK)” , Taylor & Francis e-Library, 2005.
11. Mitas, L., Mitasova, H., Spatial Interpolation. In: P.Longley, M.F. Goodchild, D.J. Maguire, D.W.Rhind (Eds.), “Geographical Information Systems: Principles, Techniques, Management and Applications”, Wiley, 1999.
12. Lubos Mitas and Helena Mitasova , “MultiDimensional Spatial Interpolation” , University of Illinois at Urbana-Champaign, 1998, hmitaso@unity.ncsu.edu.
13. Saad Bakkali and Mahacine Amrani, “About the use of spatial interpolation methods to denoising Moroccan resistivity data phosphate “Disturbances” map”, Acta Montanistica Slovaca - International Scientific Journal, Volume 13, Number 2, 216 -222, 2008.
14. Giles M. Foody and Peter M. Atkinson, “Uncertainty in Remote Sensing and GIS”, Copyright© John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2002.
15. Jaroslav Hofierka, Tomáš Cebecauer, and Marcel Šúri, “Optimisation of Interpolation Parameters Using a Cross-validation”. In Robert J. Peckham and Gyozo Jordan, “Digital Terrain Modelling: Development and Applications in a Policy Support Environment”, Copyright © Springer; illustrated edition (30 Jul 2007).
16. ESRI White Paper, “ArcGIS TM Geostatistical Analyst: Statistical Tools for Data Exploration, Modeling, and Advanced Surface Generation”, August 2001.
17. S. Je_rey, J. Carter, K. Moodie, and A. Beswick. “Using spatial interpolation to construct a comprehensive archive of australian climate data”, Environmental Modelling & Software, Vol. 16, No. 4:309-330, June 2001.
18. Weibel, R. and Heller, M. 1991. Digital Terrain modelling. In: Maguire D J, Goodchild M F, Rhind D W. Geographical Information Systems: Principles and Applications. Longman, Essex: 269-297.
19. Yousef ErfaniFard, Jahangir Feghhi, Mahmoud Zobeiri and Manouchehr Namiranian, Comparison of Two Distance Methods for Forest Spatial Pattern Analysis (Case Study: Zagros Forests of Iran), Journal of Applied Sciences Vol.8 No. 1- 152-157, © 2008 Asian Network for Scientific Information, 2008