Chapter One

Introduction

1.1 General:

A set of data points needs to be acquired To model a piece of terrain surface. Indeed, data acquisition is perhaps the single most important stage in digital terrain modeling. Modeling refers to the selection of the location. The most Three important issues related to acquired DTM source (or raw) data are density, accuracy, and distribution. The accuracy is related to measurements. The optimum density and distribution are closely related to the characteristics of the terrain surface. For example, if a terrain is a plane, then only three points on any location will be sufficient. This is not a realistic assumption and, therefore, an analysis of the terrain surface precedes the discussion of modeling strategies.

If a full information about the geometry of a terrain surface is required, it is necessary to measure an infinite number of points. This means that it is impossible to obtain full information about the terrain surface. However, in practice, a point measured on a surface represents the height over an area of a certain size; therefore, it is possible to use a set of finite points to represent the surface. Indeed, in most cases, full or complete information about the terrain surface is not required for a specific DTM project, so it is necessary only to measure enough data points to represent the surface to the required degree of accuracy and fidelity. The main problem face a DTM specialist is that how to adequately represent the terrain surface using a limited number of elevation points, that is, what modeling process to use with a known surface. The fundamental modeling theorem that is being widely used in mathematics, statistics, engineering, and other related disciplines can be used as the theoretical basis.(Dr. Zhilin li, 2005)

1.2 Motivation:

Nowadays, a DTM projects are widely used in varies applications and purposes, and frequently used to make important decisions. In order to judge these decisions, DTM quality must be known. The quality of a DTM consists of several components such as precision, accuracy, and reliability. There are several methods used to assess the quality of a DTM.

1.3 Thesis objectives:

The main objective of this research is to assess the contribution of using the semivariogram and slope descriptors to improve DTM output.

1.4 Thesis outline:

The thesis is divided into five chapters including this Chapter provides an introduction, motivation and overview of the research objectives and thesis outline. Chapter Two presents a background of the digital surface modeling, overview of the geographic information system (GIS) and the modeling considerations (Similarity-Complexity). This chapter also introduces an important literature review about the modeling from different point of view such as geometry and statistics. Chapter Three introduces the data used in this research as well as the study area. It explains the methodology including data pre-processing, processing, and assessment. The chapter also describe the implementation techniques used for the different two methods (semivariogram-slope). Chapter Four shows the results of the digital surface modeling process, and also introduce an assessment of using this two methods of digital surface modeling. Chapter Five includes the conclusion and recommendations of the research.

Chapter Two
Background and literature review

2.1 Digital terrain model:

In representing the terrain surface, the digital terrain model (DTM) is one of the most important concept. A DTM is a topographic model of the bare earth terrain relief that can be processed by computer programs. The data files contain the spatial elevation data of the terrain in a digital format which usually is presented as a rectangular grid. A model may have a few specific purposes such as prediction and control, in which case, the model only needs to have just enough significant detail to satisfy these purposes.

Digital terrain models are used especially in civil engineering, geodesy and surveying, geophysics, and geography. The main applications are:(M. R. Gridan, 2012)

1. Visualization of the terrain.

2. Reduction (terrain correction) of gravity measurements (gravimetry, physical geodesy).

3. Terrain analyses in Cartography and Morphology.

4. Rectification of airborne or satellite photos.

5. Extraction of terrain parameters, model water flow or mass movement.

2.2 Terrain Features Overview:

All terrain features are derived from a complex landmass known as a mountain or ridgeline. The term ridgeline is not interchangeable with the term ridge. A ridgeline is a line of high points of ground, usually with changes in their elevations along its top and low ground on all sides from which a total of 10 natural or manmade terrain features are classified.

(A) Major Terrain Features:

1) Hill: a hill is an area of high ground. From a hilltop, the ground slopes down in all directions. A hill is shown on a map by contour lines forming concentric circles. The inside of the smallest closed circle is the hilltop.(Web1)

2) Saddle: a saddle is a dip or low point between two areas of higher ground. A saddle is not necessarily the lower ground between two hilltops; it may be simply a dip or break along a level ridge crest. If you are in a saddle, there is high ground in two opposite directions and lower ground in the other two directions. A saddle is normally represented as an hourglass.(Web1)

3) Valley: a valley is a stretched out groove in the land, usually formed by streams or rivers. A valley begins with high ground on three sides, and usually has a course of running water through it. Contour lines forming a valley are either U-shaped or V-shaped. To determine the direction of flowing, look at the contour lines. The closed end of contour line always points upstream or toward high ground.(Web1)

4) Ridge: a ridge is a sloping line of high ground. If you are standing on the centerline of a ridge, you will normally have low ground in three directions and high ground in one direction with varying degrees of slope. If you cross a ridge at right angles, you will climb steeply to the crest and then descend steeply to the base. When you move along the path of the ridge, depending on the geographic location, there may be either an almost unnoticeable slope or a very obvious incline. Contour lines forming a ridge tend to be U-shaped or V-shaped. The closed end of the contour line points away from high ground.(Web1)

5) Depression: a depression is a low point in the ground or a sinkhole. It could be described as an area of low ground surrounded by higher ground in all directions, or simply a hole in the ground. Usually only depressions that are equal to or greater than the contour interval will be shown. On maps, depressions are represented by closed contour lines that have tick marks pointing toward low ground.(Web1)

(B) Minor Terrain Features:

1) Draw: a draw is a less developed stream course than a valley. In a draw, there is essentially no level ground and, therefore, little or no maneuver room within its confines. If you are standing in a draw, the ground slopes upward in three directions and downward in the other direction. A draw could be considered as the initial formation of a valley. The contour lines depicting a draw are U-shaped or V-shaped, pointing toward high ground.(Web1)

2) Spur: a spur is a short, continuous sloping line of higher ground, normally jutting out from the side of a ridge. A spur is often formed by two roughly parallel streams cutting draws down the side of a ridge. The ground will slope down in three directions and up in one. Contour lines on a map depict a spur with the U or V pointing away from high ground.(Web1)

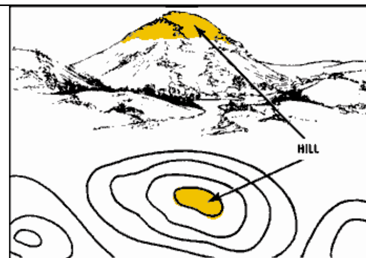
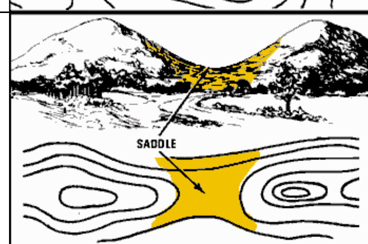
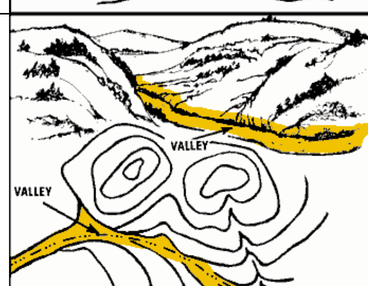
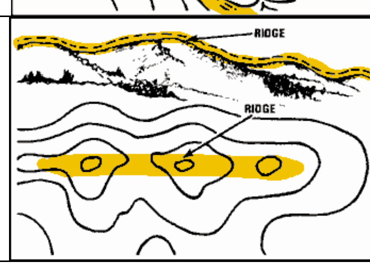
3) Cliff: a cliff is a vertical or near vertical feature; it is an abrupt change of the land. When a slope is so steep that the contour lines converge into one "carrying" contour of contours, this last contour line has tick marks pointing toward low ground. Cliffs are also shown by contour lines very close together and, in some instances, touching each other.(Web1)

(C) Supplementary Terrain Features:

1) Cut: a cut is a manmade feature resulting from cutting through raised ground, usually to form a level bed for a road or railroad track. Cuts are shown on a map when they are at least 10 feet high, and they are drawn with a contour line along the cut line. This contour line extends the length of the

cut and has tick marks that extend from the cut line to the roadbed, if the map scale permits this level of detail.(Web1)

2) Fill: a fill is a manmade feature resulting from filling a low area, usually to form a level bed for a road or railroad track. Fills are shown on a map when they are at least 10 feet high, and they are drawn with a contour line along the fill line. This contour line extends the length of the filled area and has tick marks that point toward lower ground. If the map scale permits, the length of the fill tick marks are drawn to scale and extend from the base line of the fill symbol.(Web1)

Major Terrain Features			
Hill			
Saddle			
Valley			
Ridge			

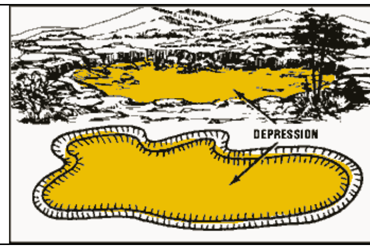
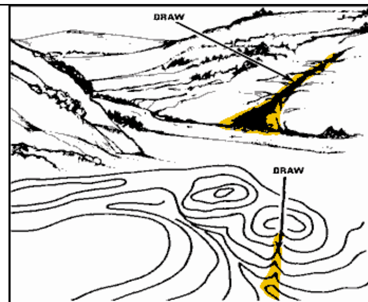
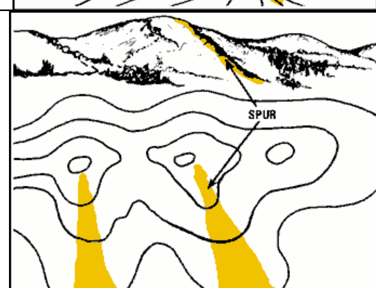
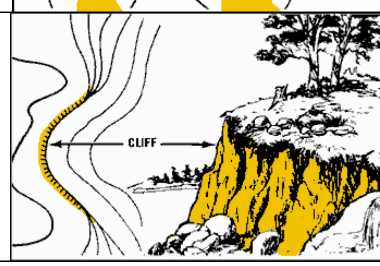
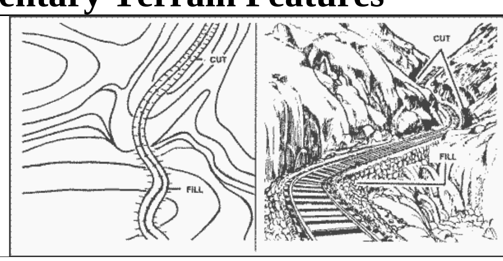
Depression		
Minor Terrain Features		
Draw		
Spur		
Cliff		
Supplementary Terrain Features		
Cut – Fill		

Figure (2-1) Major\ Minor\ Supplementary terrain features(Web1)

2.3 Digital surface modeling:

The nature of the studied phenomena is so complicated (urban areas-mountainous areas-flat areas etc.), that irregular modeling design is required. It means that certain modeling density is required for areas with low slopes, one for peaks, and another to depict linear objects. Even certain geodetic standards for number of surveying points exist, but so far, there is no generally applicable methodology on number of points, and its modeling design for specific categories of terrain descriptors. Thus, it would be very helpful from the methodological and practical point of view, to study modeling design relationship, by analyzing results with theoretically defined terrain descriptors.
Such study might consist of the following procedure: 1) Selection of the terrain properties to describe terrain. 2) Definition of terrain descriptors for certain terrain type. 3) Suggestion of modeling design based on previous two steps. Including minimal number of points required for interpolation fulfilling the criteria of accuracy.(Dr. Zhilin li, 2005)

2.3.1 Basic concepts of surface modeling:

Modeling is the process of selecting those points that have to be measured in certain positions. The operation can be characterized by two parameters, that is, distribution and density. Measurement is to determine the coordinates of a point whish concerned with accuracy. Modeling can take place before or after measurement. Modeling after measurement is to select points from a set of measured data points, usually with great density. Therefore, accuracy can also be included in the attribute set for the sampled data.(Dr. Zhilin li, 2005)
Distribution of sampled data is usually specified by the terms of location and pattern. The location is defined in terms of two positional coordinates, that is, longitude and latitude in a

geographical coordinate system or easting and northing in a grid coordinate system. Enter regarding pattern, a variety of them are available for selection, such as regular or rectangular grids. Regular 2-D data are produced by means of regular grid or progressive modeling. The resulting pattern could be a rectangular grid, a square grid, or a hierarchical structure of these two. The square grid is most commonly used. The hierarchical structured data, sampled by means of progressive modeling, can be decomposed into a normal square grid. Data that are regular in one dimension are produced by modeling with one dimension fixed (X, Y, or Z). That is, such a pattern is generated by using contouring or profiling. There are other special regular patterns, for instance, equilateral triangles and hexagons, etc. However, it seems that these structures are not as widely used as profiled or regular grid data. Enter irregular patterns may generally be classified into three groups, that is, random, cluster, and string data. Random data is mean that the measured points are located randomly (not in any specific form). Clustered data we mean that the measured points are clustered, which is often the case in geology. String data are not located in a regular pattern, yet they follow certain features (such as break lines). Enter the data sets that are sampled along rivers, break lines, or feature lines all belong to this pattern. For example, the pattern of the data resulting from composite modeling is usually a combination of string data with regular grid data.

Density can be specified by measures like the distance between two points, the number of points per unit area, and the cutoff frequency. The distance between two sampled points is usually referred to as the modeling process. If the modeling process varies with position, then an average value can be used. This measure is specified by a number with a unit. From another point of view, the required maximum frequency can also be used as a measure of data density because the modeling interval can also be obtained from it (the value of maximum frequency).(Dr. Zhilin li, 2005)

2.3.2 Modeling considerations:

In this section two of the most important considerations will be described and illustrated separately as well as the contribution of each consideration to this research.

2.3.2.1 Similarity:

For several applications it is interesting to find out how two or more points are alike, with (n) being the number of measured variables. Investigating variables individually disregards relationships between them and does not respect their multivariate characteristic. Having more than one distance and similarity indices can be calculated for all variables. The derived values describe relative interspaces between points by a single value and are a measurement for their alikeness. In larger data sets two or more naturally occurring groups can be distinguished, when the variables included in the analysis discriminate the objects in such a space. They are the basis for a couple of data ordination techniques to identify groups of points being similar or separate from others. It can be easily imagined that the distance between two points does not tell something about similarity. With increasing distances objects are more separated in their feature space. Thus, distance is a method to express an absolute value of dissimilarity. Dissimilarity itself is a relative value measuring the deviation between two points. To extend these to measures of spatial similarity, consider a scatter plot where the data pairs represent measurements of the same variable made some distance apart from each other. The separation distance is usually referred to as “lag”, as used in time series analysis. We’ll refer to the values plotted on the vertical axis as the lagged variable, although the decision as to which axis represents the lagged values is somewhat arbitrary.

The variogram is parameter used to describe the similarity of a DTM surface. At the heart of kriging is the semivariogram or structure function of the regionalized variables that you are trying to estimate. This amounts to the a priori information that you must supply to the software in order to make a regular grid out of your irregularly spaced data. Basically the idea is to have an estimate of the distance one would need to travel before data points separated by that much distance are uncorrelated. This information is usually presented in the form of the semivariogram.(Long, 1999)

$$\boldsymbol{\gamma}^{*} = \frac{\sum(\mathbf{y}(\mathbf{x}) - \mathbf{y}(\mathbf{x}+\mathbf{h}))^{2}}{\mathbf{2N}} \qquad (2.1)$$

Where. the $\boldsymbol{\gamma}^{*}$ ≡ an experimental variogram computed from the data.

(h) ≡ the lag distance between.

(N) ≡ data point pairs. There also are theoretical semivariograms which model the structure of the underlying correlation between data points, such as: spherical (Long, 1999)

$$\boldsymbol{\gamma}(\mathbf{h}) = \mathbf{C_0} + \mathbf{C}\left[\frac{\mathbf{3h}}{\mathbf{2a}} - \frac{\mathbf{1h^3}}{\mathbf{2a^3}}\right] \quad \text{For } \boldsymbol{h} \le \boldsymbol{a} \qquad (2.2)$$

$$\boldsymbol{\gamma}(\mathbf{h}) = \mathbf{C_0} + \mathbf{C} \qquad \text{For } \boldsymbol{h} > a \qquad (2.3)$$

model - which rises to the sill value more quickly than the exponential model, the general equation for it looks like:

Gaussian model - is a semivariogram model that displays parabolic behavior near the origin (unlike the previous models which display linear behavior near the origin). The formula that describes a gaussian model is:(Long, 1999)

$$\boldsymbol{\gamma}(\mathbf{h}) = \mathbf{C_0} + \mathbf{C}\left(\mathbf{1} - \mathbf{e}^{\frac{-\mathbf{h}^2}{\mathbf{a}^2}}\right) \qquad (2.4)$$

Where, (c_0) ≡ the nugget (Represents unresolved, sub-grid scale variation or measurement error and is seen on the variogram as the intercept of the variogram).

(c) ≡ the sill (The value of the semivariance as the lag (h) goes to infinity, it is equal to the total variance of the data set)

(a) ≡ the range (The scalar that controls the degree of correlation between data points, usually represented as a distance).(Long, 1999)

Practically, there are three tasks that we have to follow during the software process:

1) Computed sample variogram.

2) Detecting anisotropy.

3) Fitting sample variogram.

2.3.2.2 Complexity (roughness):

Complexity (roughness) cannot be completely defined by any single parameter, but must be a Complexity vector or a set of parameters. In this set of parameters, relief is used to describe the vertical dimension (or amplitude of the topography), while the terms grain and texture (the longest and shortest significant wavelengths) are used to describe the horizontal variations (in terms of the frequency of change). The parameters for these two dimensions are connected by slope. Thus, relief, wavelength, and slope are the roughness parameters. From both the theoretical and the practical points of view, relief, wavelength, and slope are the important parameters for terrain complexity (roughness). In geomorphology, a useful description of the landform at any point is given by altitude and the surface derivatives, slope and convexity (curvature), Slope is defined by a plane tangent to the surface at a given point and is completely specified by the two components: gradient (vertical component) and aspect (plane component), Gradient is essentially the first vertical derivative of the altitude surface while aspect is the first horizontal derivative. Further, land surface properties are specified by convexity (positive and negative convexity-concavity). These are the changes in gradient at a point (in profile) and the aspect (in the plane tangential to the contour passing through the point).(Dr. Zhilin li, 2005)

In other words, they are second derivatives. These attributes (altitude-gradient-aspect-profile, etc.) are the main elements used to describe terrain complexity. Among them, slope, comprising of both gradient and aspect, is the fundamental attribute. Gradient should be measured at the steepest direction. However, when taking the gradient of a profile or in a specific direction, it is actually the vector of the gradient and aspect that is obtained and used. Therefore, the term slope or slope angle is used in this context to refer to the gradient in any specific direction. The importance of slope has also been realized by others. Slope is the first derivative of altitude on the terrain surface. It shows the rate of change in height of the terrain over distance. From the practical point of view, using slope (and relief) as the main terrain descriptor for DTM purposes can be justified for the following reasons, 1) Traditionally, slope has been recognized as very important and used in surveying and mapping. 2) In the determination of vertical contour intervals for topographic maps.

Slope is perhaps the most important aspect of surface form, since surfaces are formed completely of slopes. Practically, there are two problems related to the estimation of its values need to be considered, that is availability and variability.(Mark, 2004)

By availability we mean that slope values should be available or estimated before modeling takes place, to assist in the determination of modeling intervals. If a DTM exists in an area, then the slope values for DTM points can be computed and the average can be used as the representative. Otherwise, slope may be estimated from a stereo model formed by a pair of aerial photographs with overlap or contour maps. The method of estimate the average slope of an area from the contour map is still widely used.(Dr. Zhilin li, 2005)

The average slope value (α) of a homogeneous are can be estimated as follows:

$$\boldsymbol{\alpha} = \mathbf{arctan}\left(\boldsymbol{\Delta}\mathbf{H} \times \frac{\sum \mathbf{L}}{\mathbf{A}}\right) \qquad (2.5)$$

Where, (Δ H) ≡ the contour interval.

(∑L) ≡ the total length of contours in the area.

(A) ≡ is the size of the area.

By variability we mean that slope values may vary from place to place so that the slope estimate that is representative for one area may not be suitable for another. In this case, average values may be used. If slope varies too greatly in an area, then the area should be divided into smaller parts for slope estimation. We need to determine not only the height of a hill, but the degree of the hill's slope as well. The rate of rise or fall of a terrain feature is known as its slope. The speed at which equipment or personnel can move is affected by the slope of the ground or terrain feature. This slope can be determined from the map by studying the contour lines and the closer the contour lines, the steeper the slope; the farther apart the contour lines, the gentler the slope. Four types of slopes that concern the military are as follows:(Web2)

1) Gentle: Contour lines showing a uniform, gentle slope will be evenly spaced and wide apart. Considering relief only, a uniform, gentle slope allows the defender to use grazing fire. The attacking force has to climb a slight incline.

2) Steep: Contour lines showing a uniform, steep slope on a map will be evenly spaced, but close together. Remember, the closer the contour lines, the steeper the slope. Considering relief only, a uniform, steep slope allows the defender to use grazing fire, and the attacking force has to negotiate a steep incline.

3) Concave: Contour lines showing a concave slope on a map will be closely spaced at the top of the terrain feature and widely spaced at the bottom. Considering relief only, the defender at the top of the slope can observe the entire slope and the terrain at the bottom, but he cannot use grazing fire. The attacker would have no cover from the defender's observation of fire, and his climb would become more difficult as he got farther up the slope.

4) Convex: Contour lines showing a convex slope on a map will be widely spaced at the top and closely spaced at the bottom. Considering relief only, the defender at the top of the convex slope can obtain a small distance of grazing fire, but he cannot observe most of the slope or the terrain at the

bottom. The attacker will have concealment on most of the slope and an easier climb as he nears the top.

2.3.3 Modeling from different point of view:

There are various ways to view points on the terrain surface from the differing viewpoints. Therefore, different modeling methods can be designed and evaluated according to the different of viewpoints as follow:

2.3.3.1 Geometry:

From the geometric point of view, a terrain surface can be represented by different geometric patterns, either regular or irregular in nature. The regular pattern can be subdivided into 1-D or 2-D patterns. If modeling is conducted with a regular pattern that is only regular in one dimension, then the corresponding method is referred to as profiling (or contouring). A 2-D regular pattern could be a square or a regular grid, or a series of contiguous equilateral triangles, hexagons, or other regularly shaped geometric figures.(Dr. Zhilin li, 2005)

2.3.3.2 Statistics:

From the statistical point of view, a terrain surface is a population (called a sample space) and the modeling can be carried out either randomly or systematically. The population can then be studied by the sampled data. In random modeling, any sampled point is selected by a chance mechanism with known chance of selection. The chance of selection may differ from point to point. If the chance is equal for all sampled points, it is referred to as simple random modeling. In systematic modeling, the points are selected in a specially designed way, each with a chance of 100 probability of being selected. Other possible modeling strategies are stratified modeling and cluster modeling. However, they are not suitable for terrain modeling and thus are omitted here.(Dr. Zhilin li, 2005)

2.4 Geographic Information Systems (GIS) overview:

Geographic Information Systems (GIS) are an integrated collection of software and data used to visualize and organize geographic data, conduct spatial analysis, and create maps and other geospatial information. Narrow definitions of GIS focus on the software and data, while broader definitions include hardware (where the data and software is stored), metadata (data that describes the data), and the people who are part of the system and interact with it as creators, curators, and users. Another definition: GIS is a visual system that organizes information around the concepts of place and location that can be used for geographic analysis, map making, database management, and geospatial statistics.(Donnelly, 2011) GIS can be (and has been) applied to virtually any discipline or endeavor. In a GIS, geographic features are represented as individual files or layers that can be added to a map. These features are not maps in and of themselves, but are the raw materials used for map making and analysis. For much of the 20th century cartographers drew geographic features on individual or acetate sheets and then layered those sheets over a paper base map to create maps. GIS uses the same principles of layering, with individual files consisting of features that can be layered on top of each other in GIS software. GIS software acts as an interface, or window, for viewing and manipulating GIS data. The ability to add different layers is quite powerful, as combining the layers allows for analysis that would be impossible if you were viewing single layers by themselves, Each GIS file is georeferenced, meaning that the file is actually tied and related to real locations on the earth. Just as paper maps were drawn based on map projections and coordinate systems, each GIS file has also been created based on a particular projection and coordinate system, which means that files that share the same reference systems can be laid on top of each other. Since projections and coordinate systems are highly standardized, GIS data can easily be shared.(Donnelly, 2011)

GIS files are stored in several formats, and each format comes in several different file types. Major formats and file include:

1) Raster: represent a continuous surface that is divided into grid cells of equal size. Each cell appears as a particular color based on some value (i.e. reflected light). Files in the raster format are similar to digital photos. Common raster objects include air photos, satellite imagery, and paper maps that have been scanned. Raster files can also consist of photos or imagery that have been generalized or have had value added to them to create a new layer, like a land use and land cover layer or a grid showing temperature. There are many different file formats, some common ones include Tiffs (.tif), JPEGs (.jpg), and SID (.sid). Unlike regular .tif or .jpg files, GIS raster files are georeferenced.(Donnelly, 2011)

2) Vector: consists of discrete coordinates and surfaces that are represented as individual points, lines, or polygons (areas). Vector files appear to be more "map-like", and are always abstractions rather than actual images (i.e. shapes to represent boundaries, points to represent cities). Common file formats are ESRI shapefiles (.shp)

ESRI coverages (.cov), Google KML files (.kml), and GRASS vector files.(Donnelly, 2011)

3) Tables: data tables that contain records for places can be converted to GIS files and mapped in several ways. If the data contains coordinates like latitude and longitude, the data can be plotted and converted to a vector file. If each data record contains unique ID codes for each place, those records can be joined to their corresponding features in a GIS file and mapped. Tables are commonly stored in text files like .txt or .csv, database files like .dbf, or in spreadsheets like Excel.(Donnelly, 2011)

4) Geodatabases: containers that can hold related raster, vector, and tabular data in one place. They are good for consolidating and organizing data. Geodatabases can be desktop (Microsoft Access .mdb, ESRI file geodatabases .gdb, Spatialite files .sqlite) or server based (PostGIS, ArcSDE).(Donnelly, 2011)

A standard interface for GIS software has evolved over time. Typically, GIS software has a data view that consists of a table of contents that lists files that have been added to a project, a data window that displays the GIS files, and a set of toolbars and menus for accessing various tools and launching various processes. Dragging the layers in the table of contents changes the drawing order of the layers, and right or left clicking on a layer in the table of contents will reveal individual properties for that particular feature. You can also access the attribute table of the feature and a symbol tab for changing how the features are depicted or classified. There are several tools for zooming in and out to examine different layers and to change the extent of the view. the way that coordinate systems and projections are handled is different for individual GIS software packages. In general, the options are: define the projection and coordinate system for the project before adding the files, or the project automatically takes the projection of the first file added. If you try to add GIS files that have different projections, some software may try to re-project the data on the fly, while others will simply fail to draw the new layers. Even if the software can correctly draw a layer without the user defining it, or even if it can re-project layers on the fly, users will run into problems later on when trying to manipulate the GIS files. You should always be sure to define the projection properly and make sure that all files share the same one - most GIS software will give you the ability to re-project data.

GIS software provides users with a variety of ways for querying geographic data, either by selecting records in the attribute table or shapes in the view, or by conducting searches where you build queries to high-light features that contain specific attributes, or that have some relationship with another geographic layer.
GIS software comes with a variety of editing tools that allow you to modify the geometry of GIS files. For example, you can merge features together, break them apart, or clip out or select certain areas to create new files. Collectively these processes are known as geoprocessing. You geoprocess layers in order to prepare raw data for analysis, to create new layers or data, or to simplify layers for cartographic or aesthetic purposes. GIS also provides the ability to edit files on a feature by feature basis. Most GIS programs have a separate map layout or print layout, where the user can create finished maps with standard map elements like titles, legends, scale bars, north arrows, and accompanying text. Finished maps can be exported out of the GIS as static files, such as pdfs or jpgs. Users can always save their GIS projects in a GIS project file. The scale and extent of the data view, symbolization and classification assigned to layers, map layouts, and links to GIS files used in the project are stored in the file. It's important to understand that the GIS files themselves are NOT stored inside the project file - the GIS data and the GIS project file exist independently. When adding data to a GIS, you are establishing a link from the GIS project to the GIS data, the GIS data is not stored within the project. Furthermore, changing the colors of the features or classifying them in a certain way has no effect on the actual GIS data files themselves. When you change

symbols, you are only changing how the GIS program views the data, you're not changing the data itself. This is an important concept to grasp. Essentially, the GIS software acts as a window for viewing and working with GIS data, which is stored outside the window. The GIS project file essentially stores the window dressing, of scale and symbolization. You never actually change the GIS data unless you go into an edit mode or conduct an operation that creates a new GIS file. This relationship is of crucial importance when it comes time to move or share files, if you move your project file or your data, the links between them will become broken, and you'll need to re-establish the location between the project and the data in order to repair your project file.(Donnelly, 2011)

Chapter Three

Data and methods

3.1 Study Area:

The study area is located on the east of Khartoum city- SUDAN, this area is meant to be a proposed location for the eastern Khartoum bridge, which will be cross the blue Nile river to connect the eastern Khartoum side. The location extended along the Nile from south: E 464984 N 1712397 up to north: E 464376 N 1714808. The total coverage of the study area is 3267 km², covered by 794 elevation points, frequents between 371.51 m (the minimum elevation) and 387.85 m (the maximum elevation).

Figure (3-1) Study area

3.2 Data acquisition:

For the data acquisition each observed point is designated by two elements, an alphanumeric name for its identification and a numerical code for its description. This codification method is used for the automatic recognition of any type of details, to help automatically construct objects of coverage from field data, and to support the use of symbols. Two lists of codes has been adopted for most details frequently met on the terrain, one for point features and another for linear features. In this procedure it has been respect the following rules, 1) Respect the codification method adopted for collected points of details. 2) Respect the course direction for linear details and contours, while going from the first point of the detail to the last one. The sample design for spatial data may or may not be on grid. 1) Uniform modeling: choosing locations uniformly throughout the region and independently. 2) Stratified random modeling: takes a uniform random sample. 3) Centric modeling: where the samples are taken on an equally spaced grid in the region. 4) Non aligned modeling: where the sample locations are of the nodes of a grid. Of these, Uniform modeling, Stratified random modeling, Non aligned modeling result in modeling that are not on a grid. Centric modeling is on a grid.

3.2.1 GPS data:

The electronic and computer technology have considerably changed the area of data acquisition and data processing. The use of the progress of spatial positioning methods, such as the GPS, made collection of data very rapid. Consequently, reserved time for data acquisition in the field is reduced, a great number of sources of errors are decreased, results are obtained rapidly, and data quality is improved. Once data are transferred in the computer, they are ASCII observation files or coordinate files, but in GPS formats. These files should be arranged in the ESRI Field data Format, can contain description data for occupied stations, observations and coordinates. Observations can be height, horizontal and vertical circle readings, distances, etc.

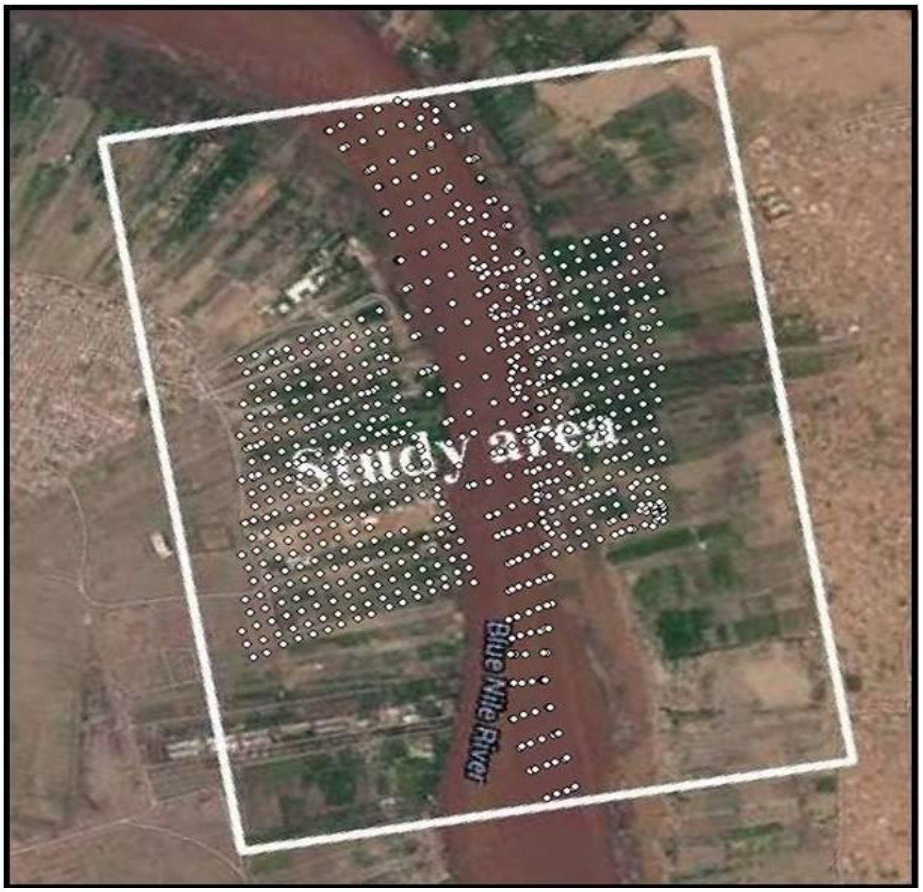


Figure (3-2) GPS data covered the study area

3.3 Modeling methodology:

Figure (3-3) describes the whole process of digital terrain modeling. It can be seen clearly that there are six different stages, in each of which one or more actions are needed to move to the next one. As it shows the research concern with the modeling which is between data source stage and raw data stage, also a specific DTM project may need only some of actions, such as feasibility study, project planning and design, contracting and shipment. But this research deals mainly with the theoretical and methodological aspects of digital terrain modeling.

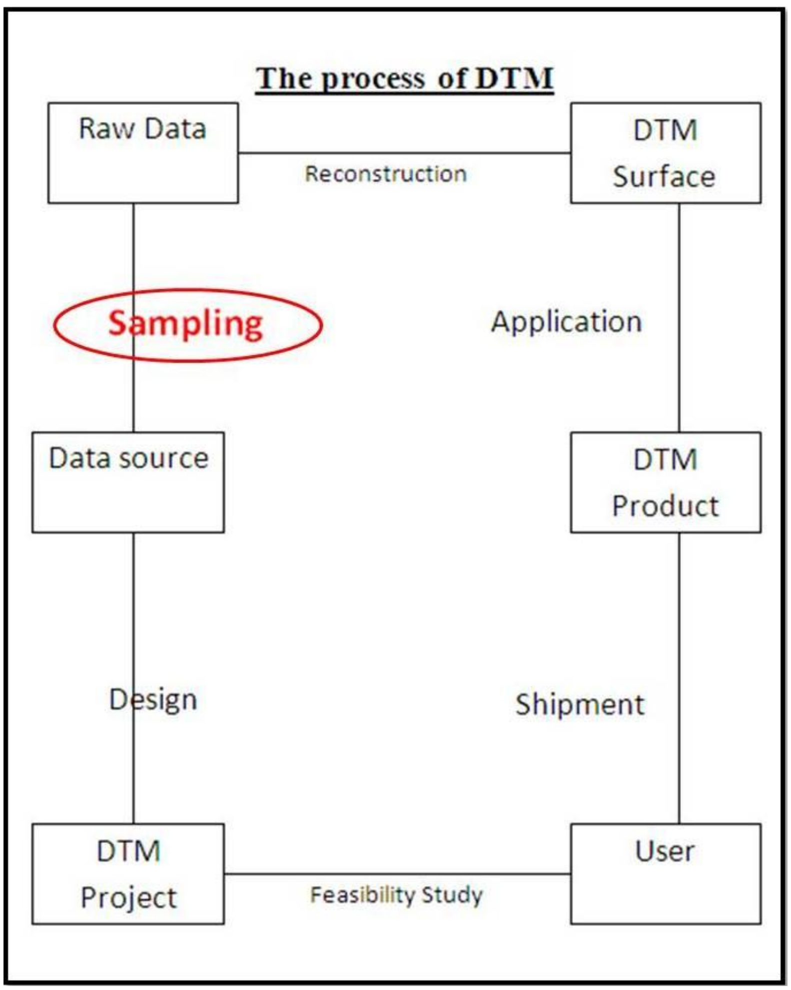


Figure (3-3) Stages of DTM project

3.3.1 Using semivariogram descriptor (similarity):

A semivariogram can be thought of loosely as an average of the various directional variograms. The semivariogram serves a useful starting point for establishing some of the parameters required for sample variogram calculations. It has been shown to be useful when the data are not on a grid to establish the lag tolerance. When the data are on a regular grid, they may be treated as row and columns of data. Assuming there are no row and column interactions, there are ways to fit, and the residuals become a new data set. In other method, anisotropy may still be an issue. With non gridded data as the data of the study area in this research, if there is a trend in the data, it must be removed. For the non-gridded data that we used (fig 3-4) the origin will represent a location in the design region. The azimuth is the measured angle from the y-axis, clockwise.

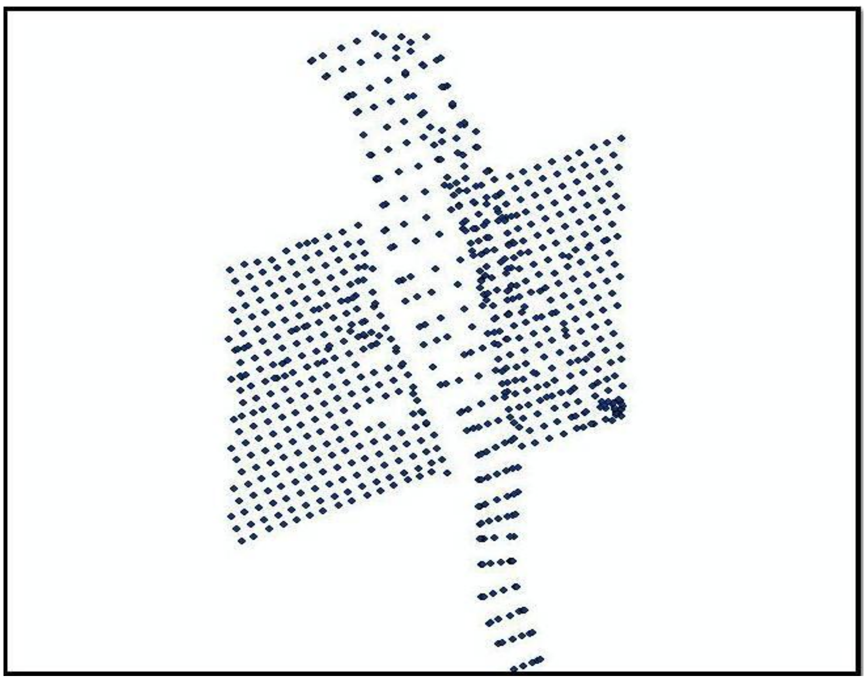


Figure (3-4) Elevation points

The number of lags should be determined to compute the sample variogram. in the sample region the lag size h(lag) is determined by dividing the maximum distance by the number of lags. (fig 3-5) shows the Two important tolerance intervals, the lag tolerance and azimuth tolerance. The lag tolerance is some amount from the current lag. Such that any location within h(lag) ± interval is considered to be at the current lag. Next is the azimuth tolerance, this amount such that any point within the interval azimuth ± is considered in the current azimuth.

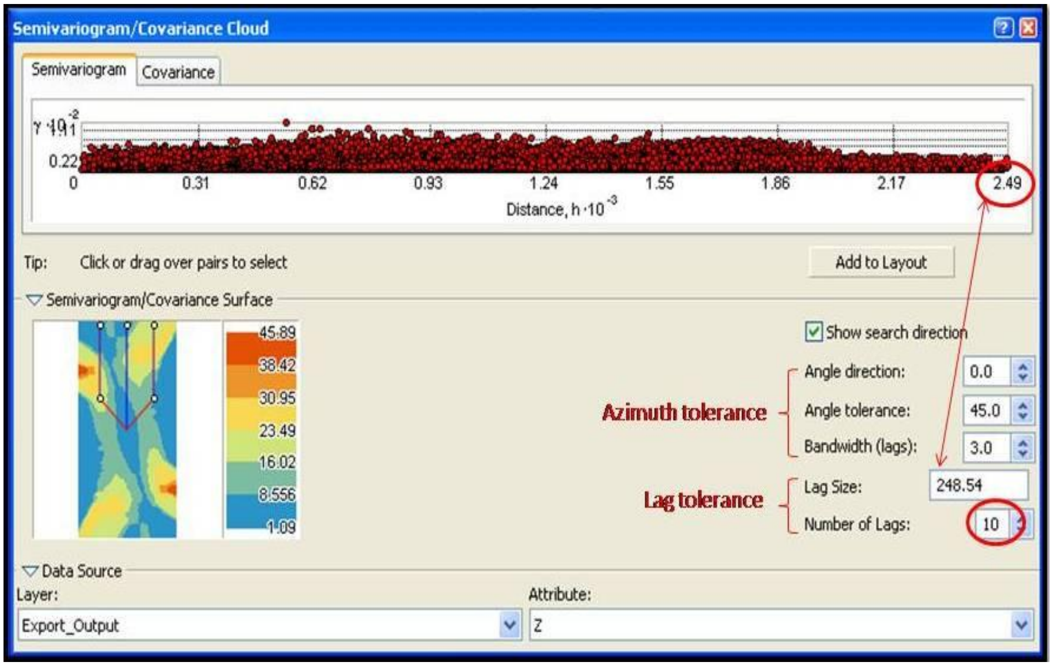


Figure (3-5) Data examination

They may be a number of reasons, such as outliner in the data or a trend on the data may result in violations in the stationary assumptions. The trend is estimated by a polynomial function and removed (fig 3-6). The residuals are obtained and used as new data set. Hence, an experimental variogram is computed on the residuals, and if the trend removal has been successful, the process yields an isotopic experimental variogram.

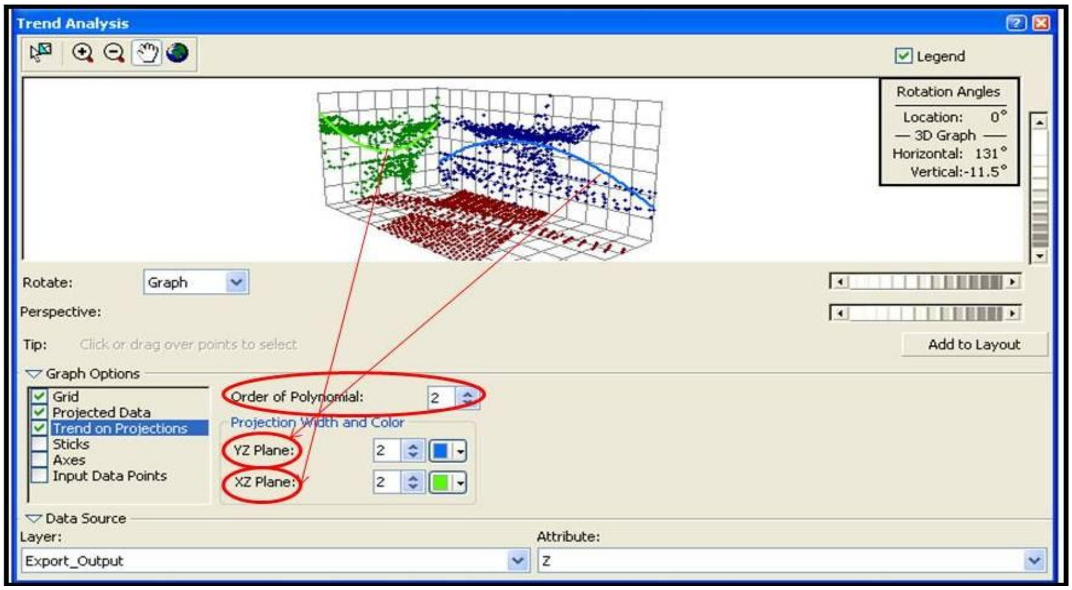


Figure (3-6) Trend analysis

As previously mentioned, outliners must be taken into consideration, and the trend should be remove (fig 3-7).

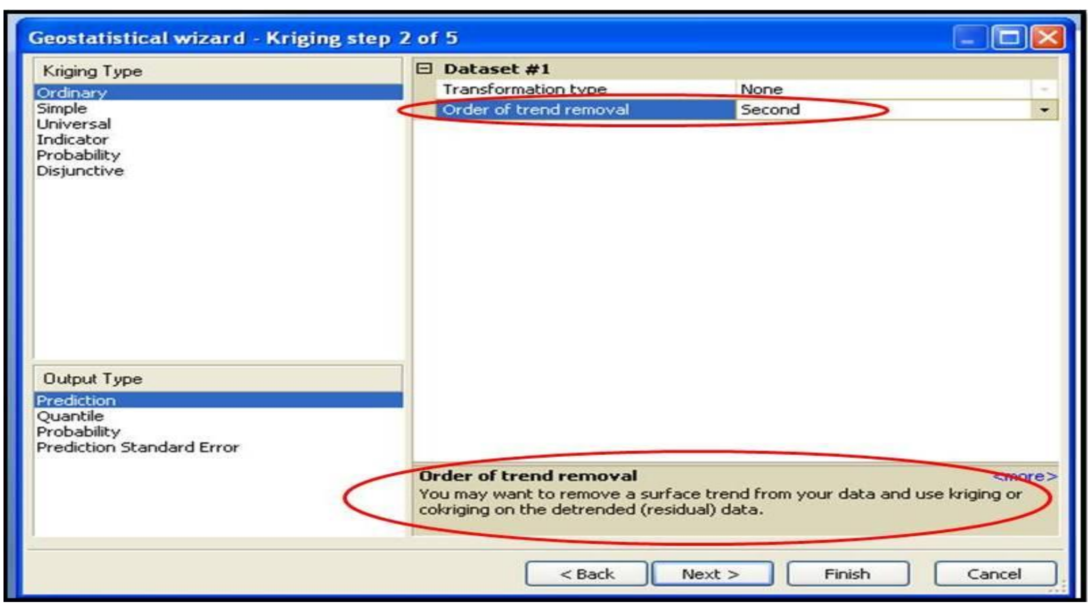


Figure (3-7) Trend removal

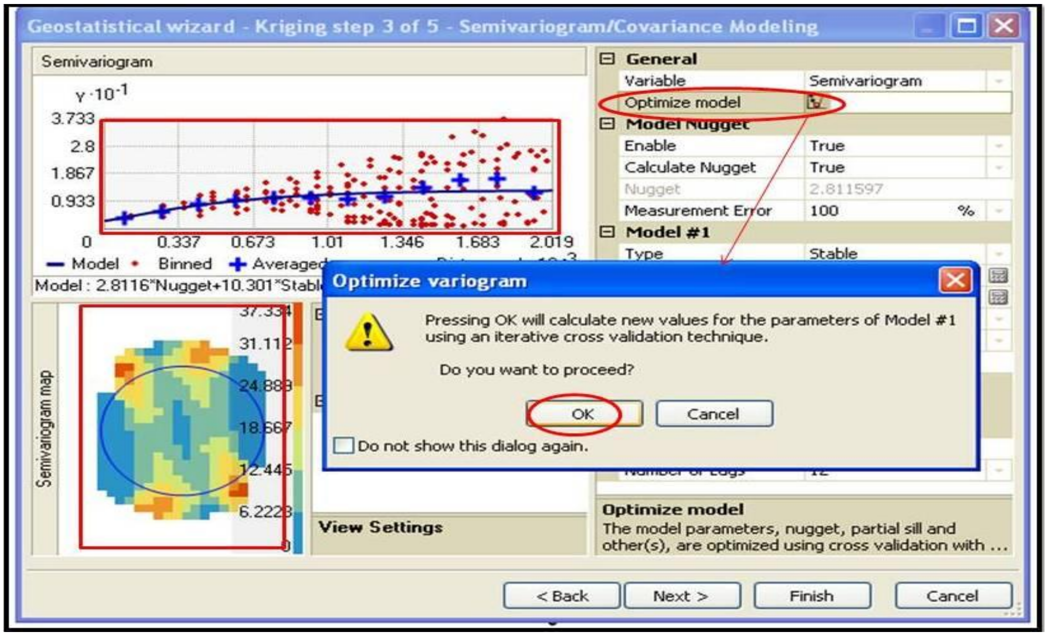


Figure (3-8) Semivariogram model carve and semivariogram surface

This iterative procedure is help to obtain the settings for the number of lags and the angular tolerance. The number of lags in the computed semivariogram has been reduced to produce the sample variogram in (fig 3-9). By reducing the number of lags, it increases the lag spacing, which in turn includes more pairs in estimating each point in the sample variogram. The result is a smoother sample variogram. When the experimental variogram is computed in various directions, and if they are different, the analyst must determine the reason. Another method uses linear smoothers. Once sample variogram has been computed and anisotropy has been detected, the next major task is to fit a parametric variogram function.

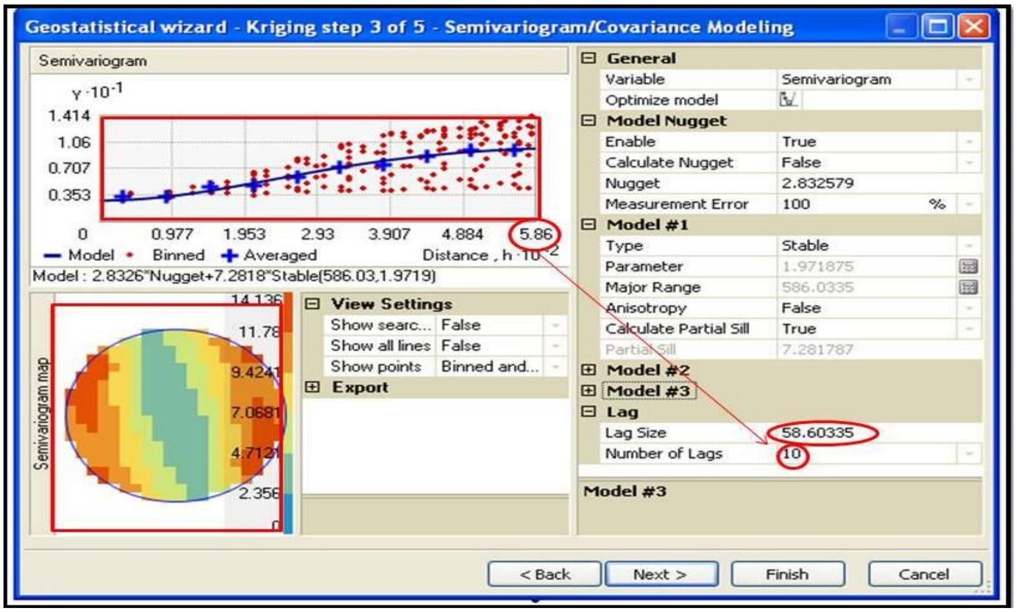


Figure (3-9) Fitting sample variogram

Once the lag tolerance is established, then directional sample variogram are computed, and the azimuth tolerance can be established. The data are displayed using various plots in (fig 3-8) semivariogram model carve and semivariogram surface. The default fit confirms a quadratic components. The default smooth is added to the residual plot to assess the fit of the model. The smooth shows no structure. Therefore, should be modeled the trend successfully, and the trend should be identified and removed. The next step is to use the residuals as new data set to compute a variogram.

3.3.2 Using slope descriptor (complexity):

First it is need to know the estimate value of the average slope (from the data figure- 1 and figure- 2) as follows:

$$\boldsymbol{\alpha = \arctan\left(\Delta H \times \frac{\sum L}{A}\right)} \qquad (3.1)$$

Where α is the slope

Δ H is the contour interval = 1.634 m

∑L is the total length of contour lines = 30288 m

A is the Area = 3267 km²

α = arctan (0.015) = 0.8593 (compare it with the mean in fig 3-10)

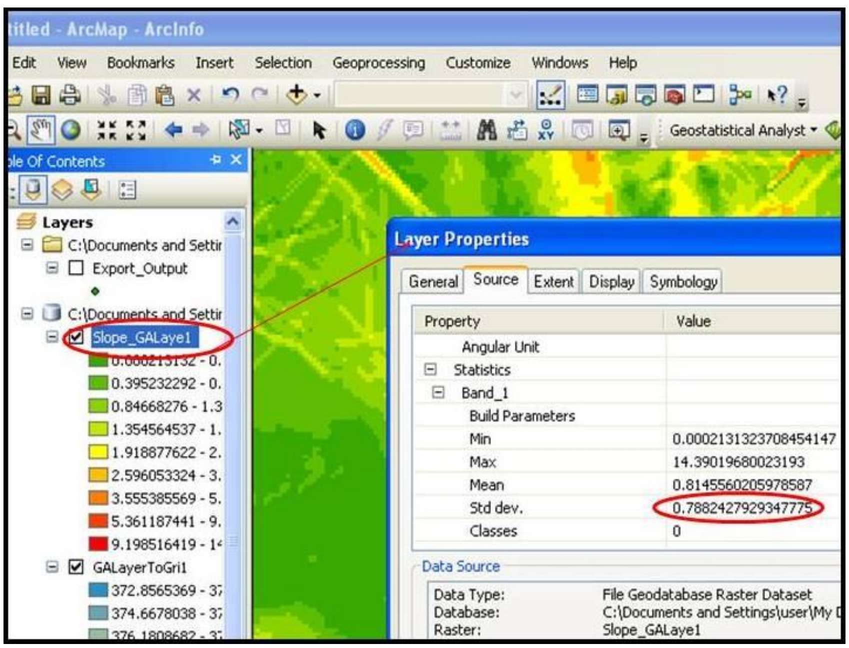


Figure (3-10) Slope statistics of the large area

Notice that it has a large standard deviation of the data used, so, the area should be divided into smaller parts for slope estimation, and that what is has done as shown (fig 3-12)

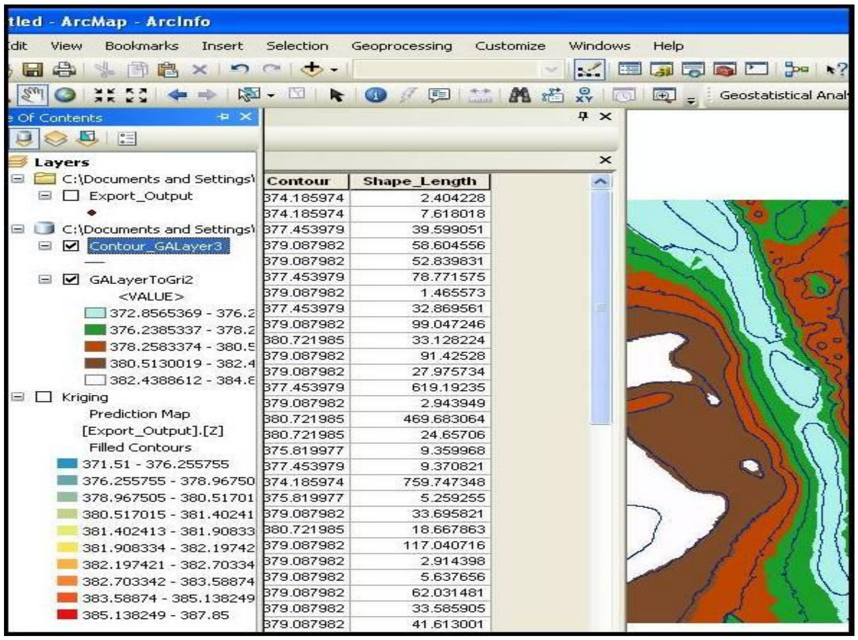


Figure (3-11) Calculate area, total length of contour lines

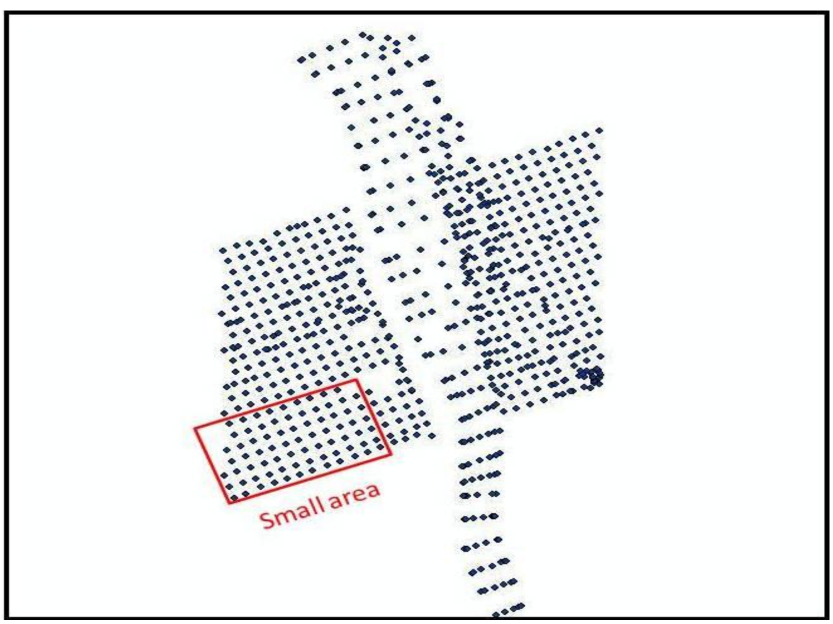


Figure (3-12) Divide large area to small study area

In (fig 3-13) after divided the area into small parts, the standard deviation is get smaller.

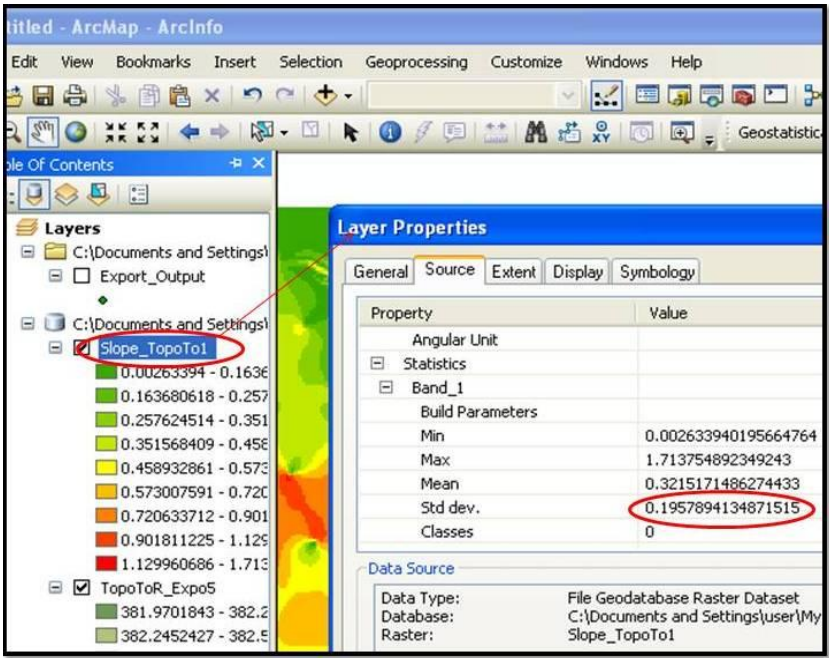


Figure (3-13) Slope statistics of the small area

The next step is to have a terrain of slopes as shows (fig 3-14).

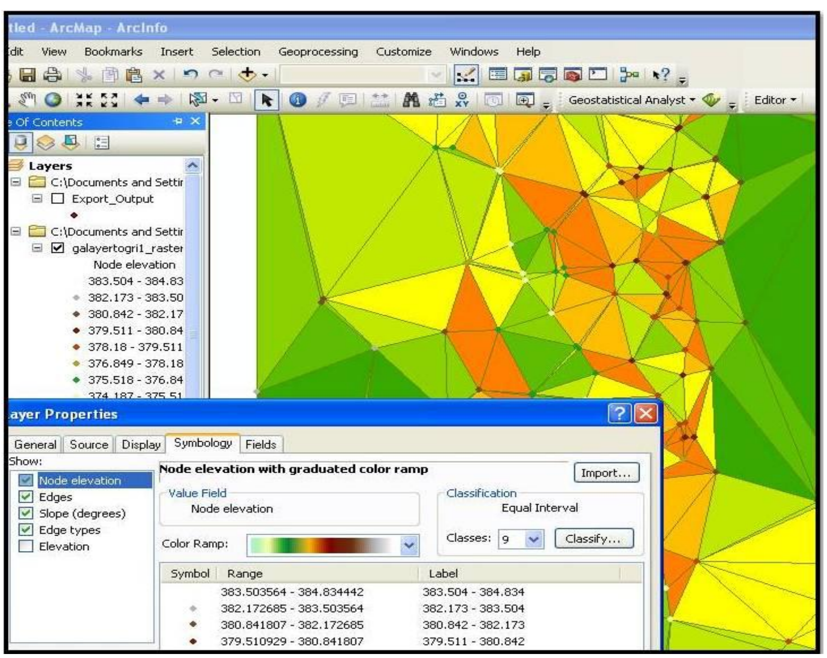


Figure (3-14) Terrain of slopes

Another way to represent the slope is by using ArcScene and it is create a 3D shape for visual assessment (fig 3-15).

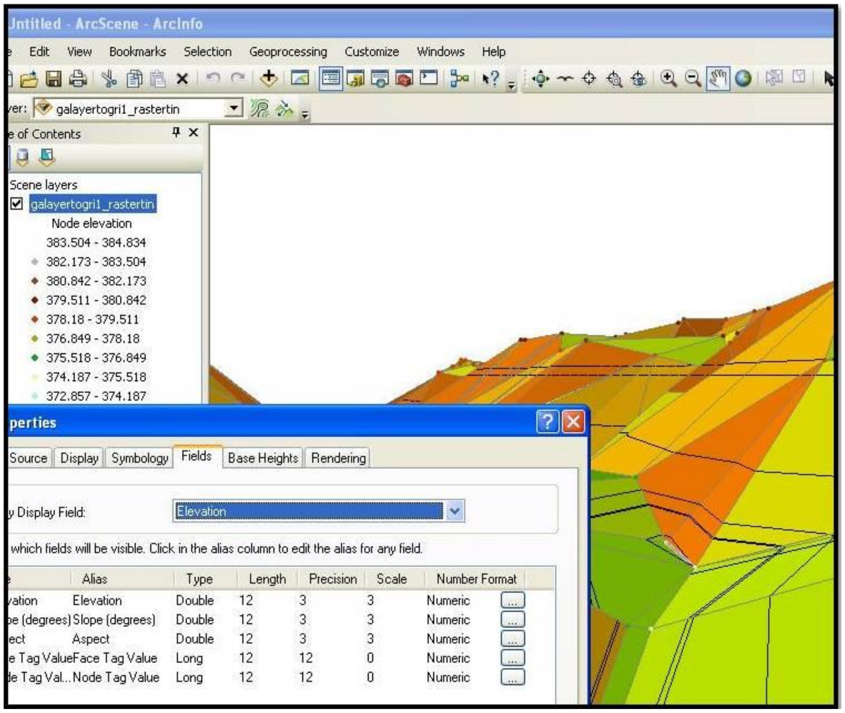


Figure (3-15) 3D terrain of slopes

Chapter Four

Result and discussion

4.1 Results:

4.1.1 Semivariogram descriptor usage:

The above analysis was an application of some of the geostatistical techniques used in practice. The data was plotted and trends were identified and removed. Throughout the process the iteration procedure started randomly with 36º, 130º, 220º and 310º as shows in the figures bellow.

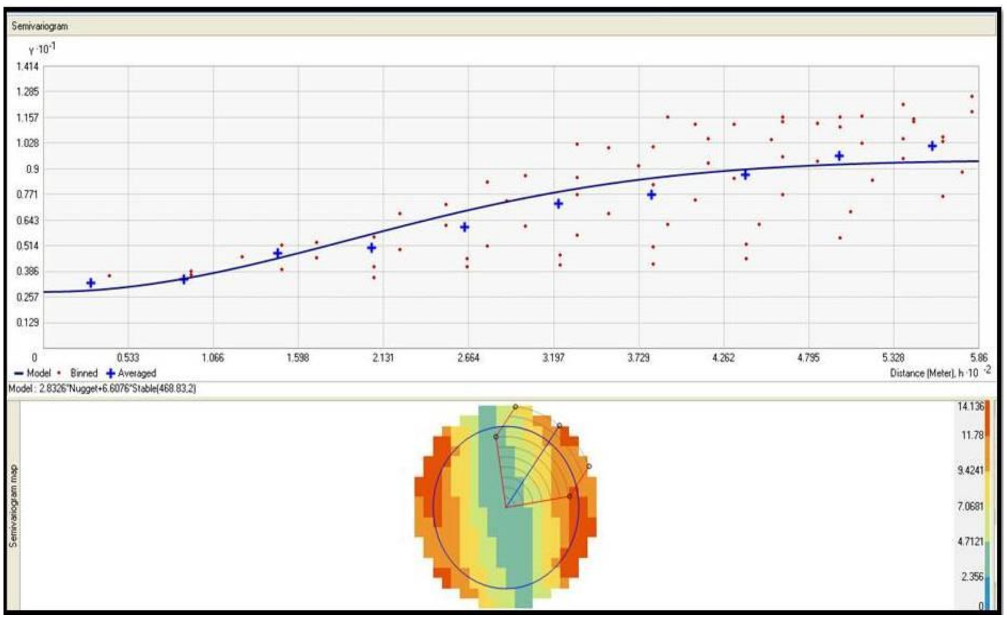


Figure (4-1) Semivariogram on 36º

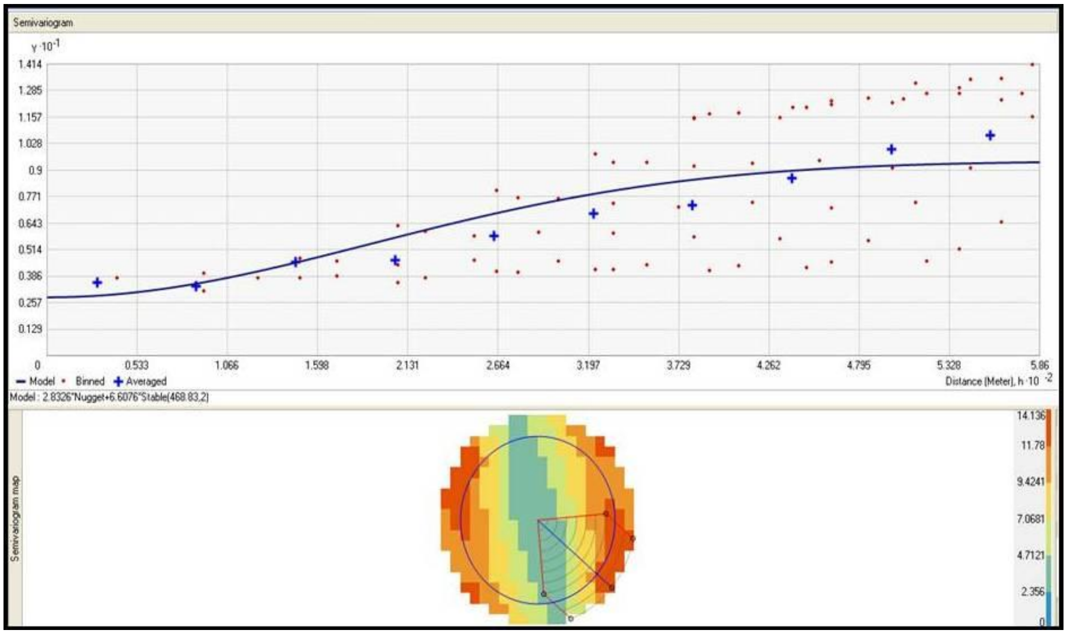


Figure (4-2) Semivariogram on 130º

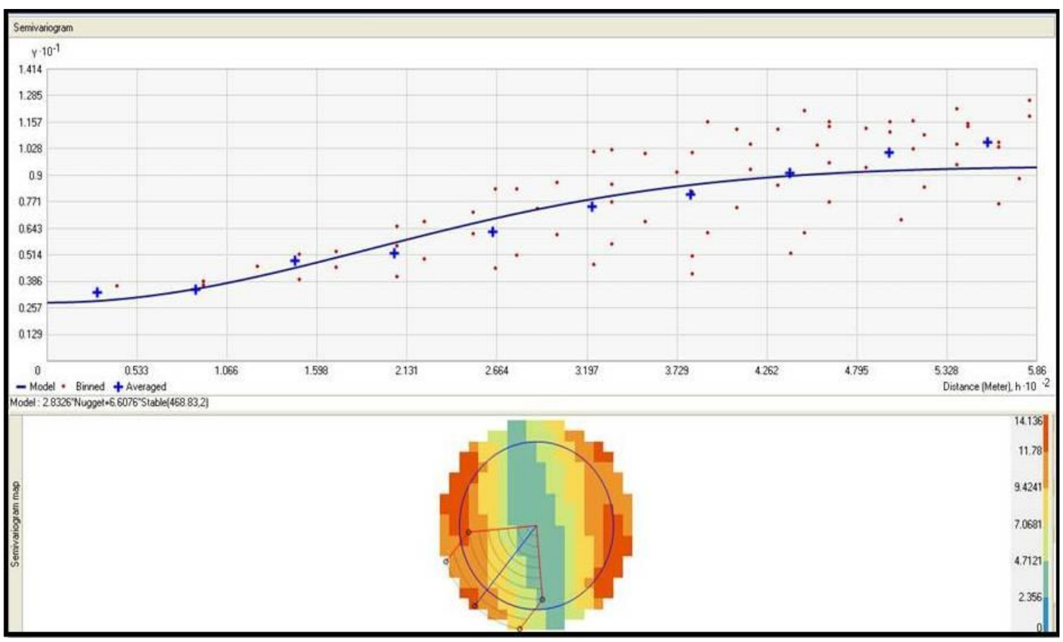


Figure (4-3) Semivariogram on 220º

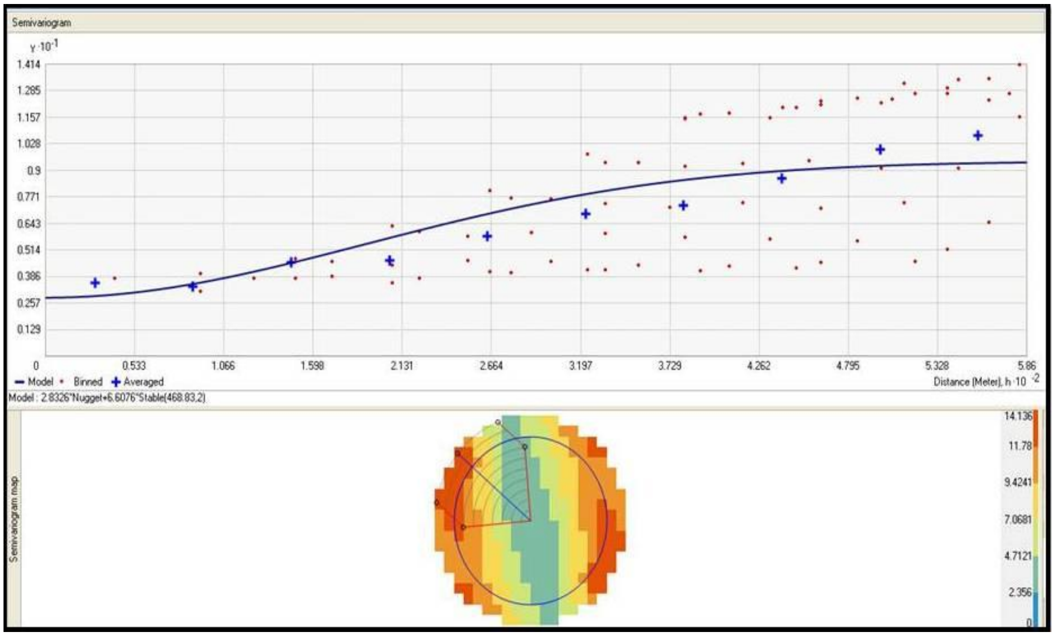


Figure (4-4) Semivariogram on 310º

The result show that in each direction, the sample variograms are similar, and the variogram surface is fairly circular. The simulation was performed 5 times with different directions and each time came to the same conclusion. The conclusion drawn from this analysis, is that it has an anomalous data set, and it needs to increase the angular tolerance to produce a reliable sample variogram.

The next step is to show the result map of using semivariogram method for modeling. As it shows in (fig 4-5) the best result is the continuous surface, and the other result map shows a discrete surface which is done without semivariogram process.

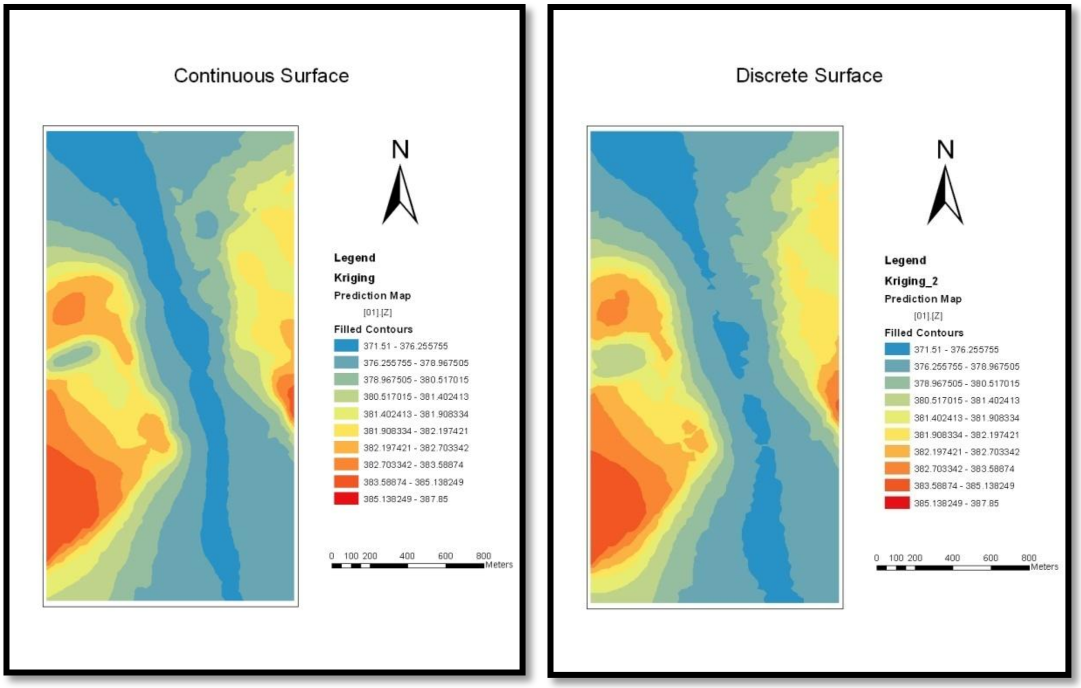
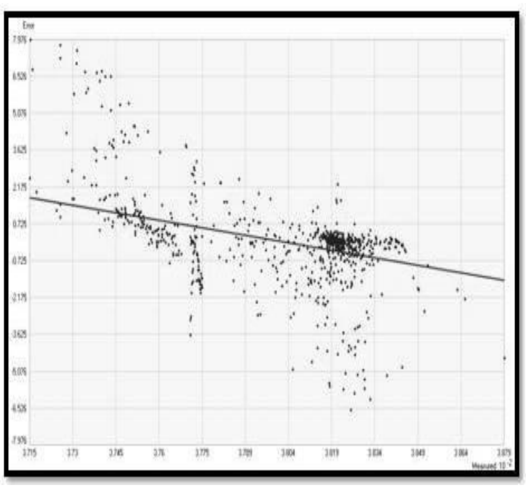
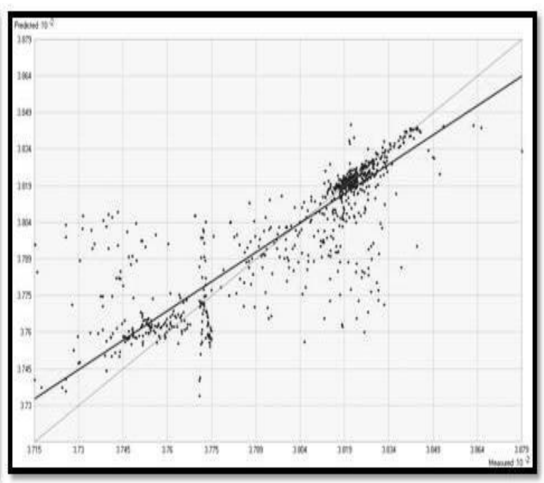


Figure (4-5) Result surfaces

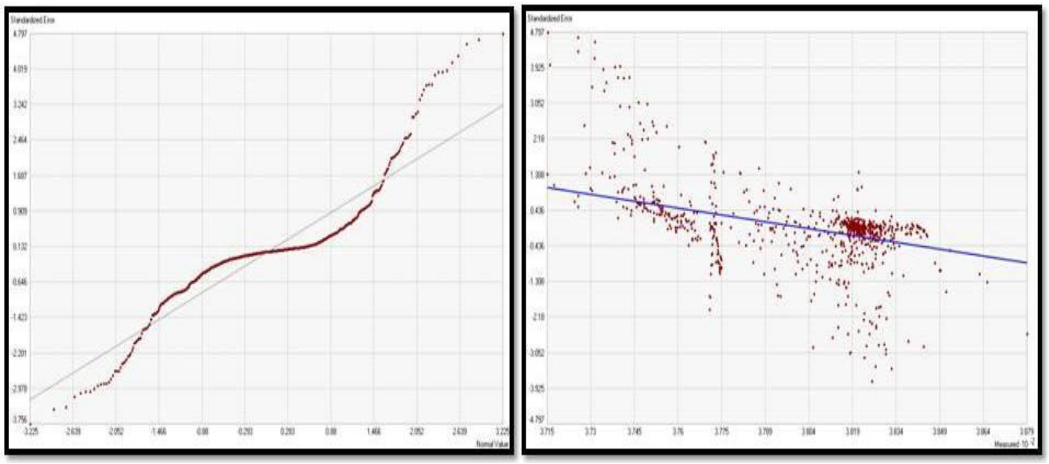
Also, needing of knowing the prediction errors of the semivariogram modeling.



Predicted



Error

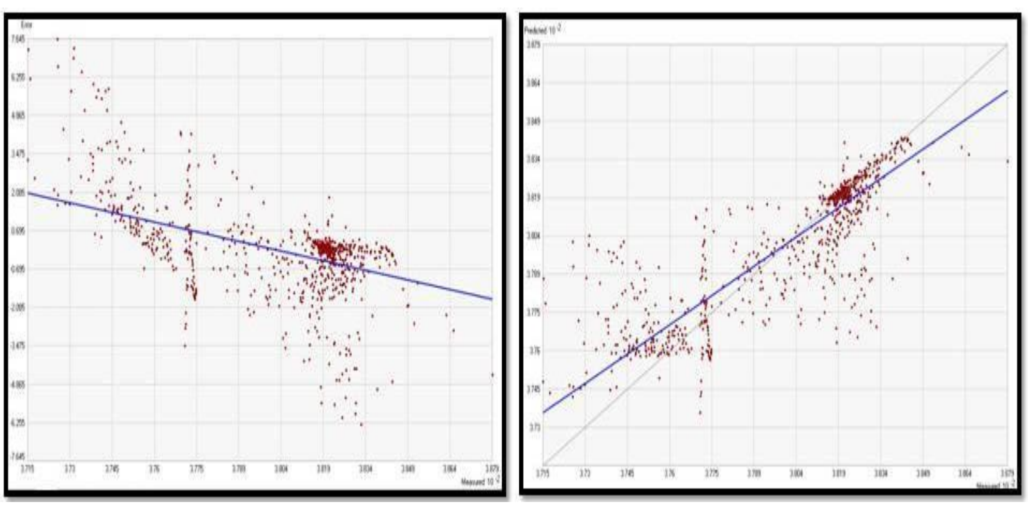


Standardized Error Normal QQ Plot

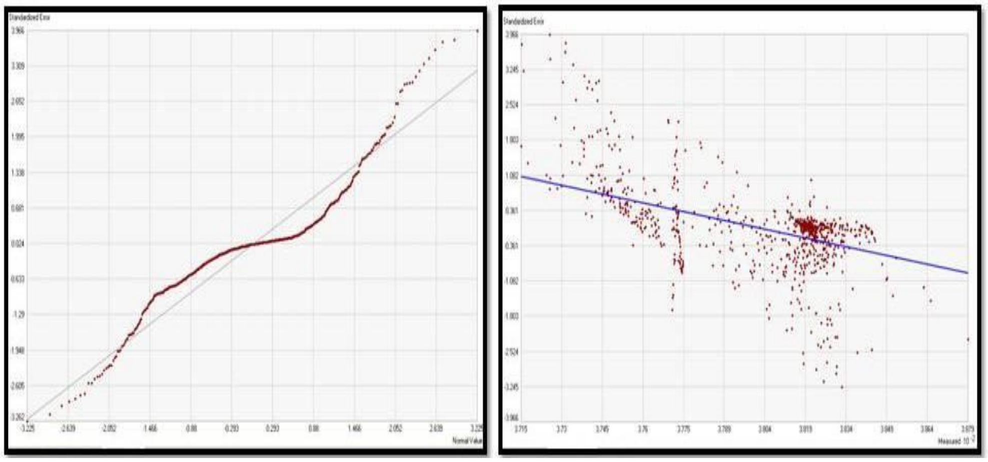
Figure (4-6) Cross validation for the continuous surface

Table (4-1) Semivariogram Prediction Errors for the continuous surface

Samples	794 of 794
Mean	0.005291349
Root-Mean-Square	1.79211
Mean Standardized	0.002480979
Root-Mean-Square Standardized	1.024862
Average Standard Error	1.781936



Predicted Error



Standardized Error Normal QQ Plot

Figure (4-7) Cross validation for the discrete surface

Table (4-2) Semivariogram Prediction Errors for the discrete surface

Samples	794 of 794
Mean	0.01982141
Root-Mean-Square	1.752548
Mean Standardized	0.01000494
Root-Mean-Square Standardized	0.900309
Average Standard Error	1.949704

The discussion of the result of prediction errors (fig 3-8) and (fig 3-9) has to through the best considering as follows:

1- The predictions should be unbiased, indicated by a mean prediction error as close to 0 as possible.
2- The standard error are accurate, indicated by a root-mean-square standardized prediction error close to 1.
3- The prediction don't deviate much from the measured values, indicated by root-mean-square error and average standard error that are as small as possible.

4.1.2 Slope descriptor usage:

It is true that any given input data carry an error value significant enough to change the resulting slope, even the high precision equipment. The results obtained inputs of the same resolution and acquired with other methods could be used for a better comparison and calculation in slope error estimation. Through the large area we found that low accuracy indicated by a large standard deviation of the data =0.788, and after we divided the area into small areas the standard deviation was getting small =0.196 The result of the slope usage shows in (fig 4-4)

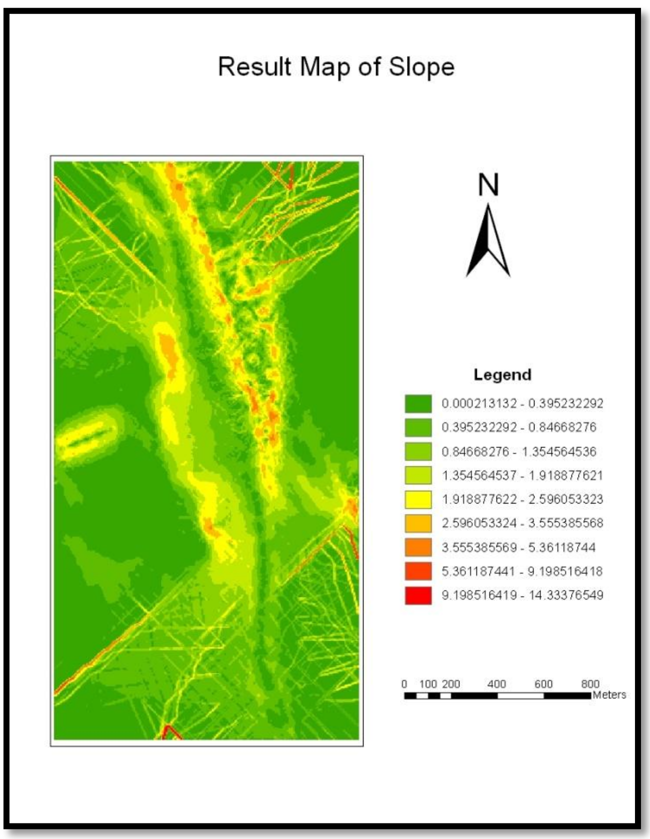


Figure (4-8) Result map of slope

As shown also in the tables bellow, so, it was increase the accuracy.

Table (4-3) Slop properties for model the large area

Minimum	0.0002131323708454147
Maximum	14.39019680023193
Mean	0.8145560205978587
Standard Deviation	0.7882427929347775

Table (4-4) Slop properties for model the small area

Minimum	0.002633940195664764
Maximum	1.713754892349243
Mean	0.3215171486274433
Standard Deviation	0.1957894134871515

Other software tools should be used to prove the simulated reality. Because of the time demanding computational process, less consuming processes should be investigated for the error pattern simulation. It is necessary to remember the main reason for dealing with the uncertainty, decreasing the risk that the outcome will be incorrect and will lead to wrong decisions. This study was made as an error propagation background. To know the uncertainty in the result is important in crisis management and other fields. Sometimes even one degree in slope can change the situation of the whole area.

4.2 Assessment:

4.2.1 Modeling with semivariogram descriptor:

By doing a comparison between the two semivariogram models we can find that the model (1) which has follow all the tasks is the best (fig 4-5).

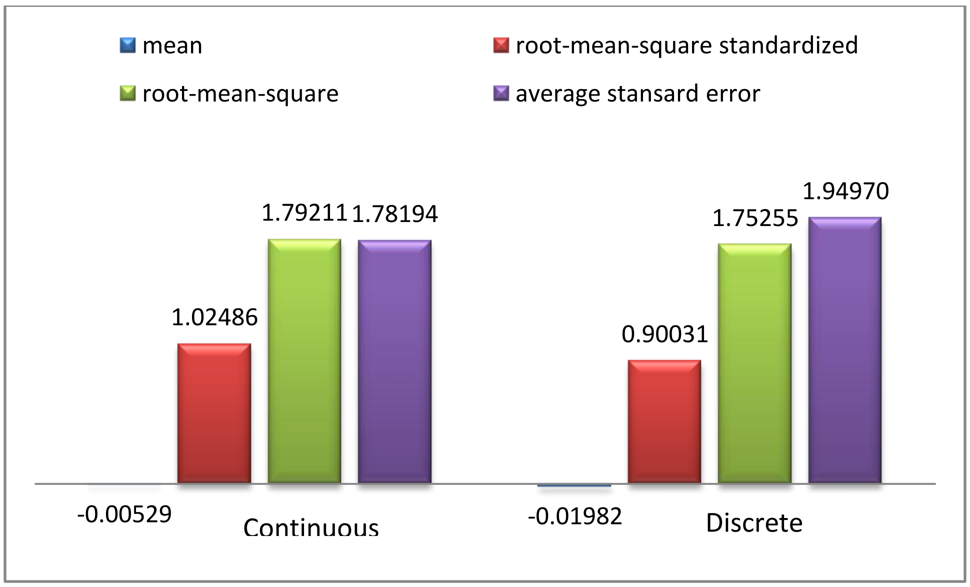
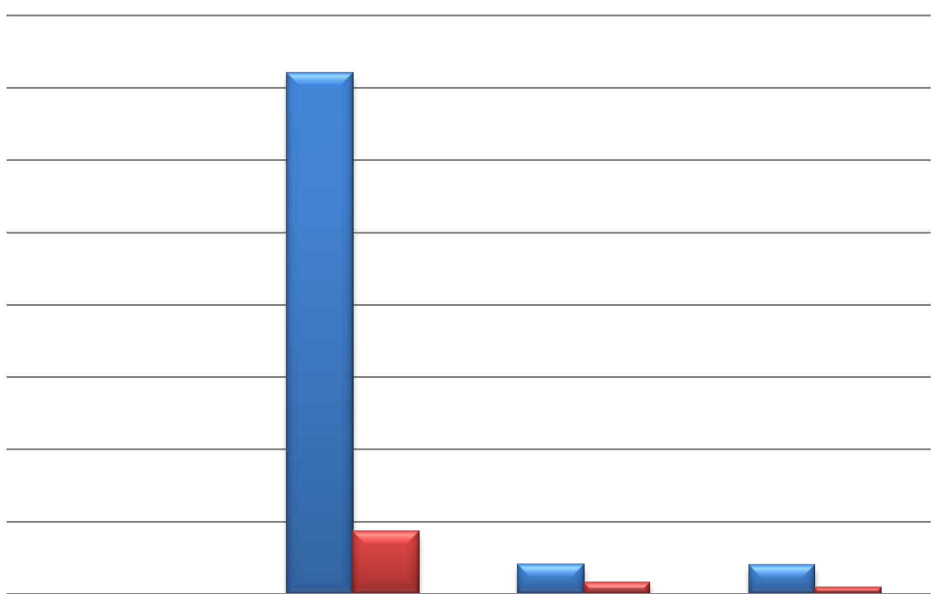


Figure (4-9) Comparison between the two surfaces

If the average standard errors are close to the root mean squared prediction errors, you are correctly assessing the variability in prediction. If the average standard errors are greater than the root mean squared prediction errors, you are overestimating the variability of your predictions; if the average standard errors are less than the root mean squared prediction errors, you are underestimating the variability in your predictions. Another way to look at this is to divide each prediction error by its estimated prediction standard error. They should be similar, on average, so the root mean squared standardized errors should be close to 1 if the prediction standard errors are valid. If the root mean squared standardized errors are greater than 1, you are underestimating the variability in your predictions; if the root mean squared standardized errors are less than 1, you are overestimating the variability in your predictions. The root-mean-squared prediction error may be smaller for a particular model. Therefore, you might conclude that it is the optimal model. However, when comparing to another model, the root-mean-squared prediction error may be closer to the average estimated prediction standard error. This is a more valid model, because when you predict at a point without data, you have only the estimated standard errors to assess your uncertainty of that prediction. When the average estimated prediction standard errors are close to the root-mean-squared prediction errors from cross-validation, you can be confident that the prediction standard errors are appropriate.

4.2.2 Modeling with slope descriptor:

This experimental study based on a unique actual data set has provide very interesting findings. These results are directly related to the topographical features observed in the study area dominated by a gently undulation relief. Using error propagation, the standard deviation in height may be estimated. This estimation can be used to predict the precision of derivatives of the DTM slope. As an alternative for the distinction between extrapolation and interpolation areas, a threshold for the predicted accuracy in height may be applied to classify insufficiently determined areas. As we have seen, the slope quality depends on the size of the study area. The slope error is reduced. Moreover, its influence is higher on slope than on the elevation. Thus, the choice of the slope parameter is of great importance and it constitutes one of many other production parameters that can influence the produced DTM. A small area size tends to be more accurate.



	Minimum	Maximum	Mean	Standard Deviation
Large Area	0.00021313237084541	14.3901968002319	0.814556020597858	0.788242792934777
Small Area	0.00263394019566476	1.71375489234923	0.321517148627445	0.195789413487152

Figure (4-10) Comparison between the large area usage and the small area usage

Chapter Five
Conclusion and recommendations

5.1 Conclusion:

- The purpose of this research was to assessment the contribution of using the semivariogram and slope descriptors to improve DTM output.
- two methods to assess the quality of a DTM are presented above, but the matter of their application, the consequences to take based on them have remained unanswered so far.
- In the majority of cases, DTM users require simply applicable information on exterior quality. The topic of accuracy is easy to communicate, but frequently the issue of reliability is not.
- A combination of a threshold for a minimum amount of reliability and an accuracy measure seems practical. Areas of the DTM that do not reach the threshold of reliability should be masked appropriately.
- During the creation of a DTM, deeper insights into data and model quality are needed in order to guarantee acceptable results. Subsequently, inconsistencies within different groups of data must be explored through difference models.

5.2 Recommendations:

- System calibration must further be enhanced.
- High accuracy DTM can be generated as a by-product from parallel guidance systems.
- The quality of the reference signal causes significant differences in the accuracy of the measured elevation, so Future research is needed for an automated workflow of high accuracy DTM generation and quality analysis.
- The underlying error pattern has to incorporate the outliers too, if there are any of them, then the sources of them must be found.
- The simulated error pattern has to be as closest as possible to the empirical one.
- This area is open for further research.

References:

1. Donnelly F 2011. AN OVERVIEW OF GISOn-line [URL] Year/Electronic Resource Number|.

2. Dr. Zhilin Li DQZ, 2005. DIGITAL TERRAIN MODELINGPrinciples and Methodology. In.

3. Long AE 1999. VARIOGRAM MODELLINGOn-line [URL] Year/Electronic Resource Number|.

4. M. R. Gridan CG 2012. "SURFER" SOLUTION, TOOL FOR TERRAIN MODELING On-line [URL] Year/Electronic Resource Number|.

5. Mark DM 2004. GEOMORPHOMETRIC PARAMETERS: A REVIW AND EVALUATIONOn-line [URL] Year/Electronic Resource Number|.

6. Web1. IDENTIFY MAJOR/ MINOR TERRAIN FEATURES. In.

7. Web2. ELEVATION AND RELIEF - TYPES OF SLOPES. In.