Cross-Validation of Elevation Data Interpolation

Expert Dr. Hussain Zaydan Ali

Ministry of Science and Technology

hussainzayali53@yahoo.com

Abstract: *Most geographical spatial analysis requires a continuous data set and this study is designed to create such a surface. Digital model of landscape is an important part within creation of geo-information systems. It is an important tool in applications, which model an Earth's surface like geomorphology, hydrology, geology, cartography, ecology, etc. Many software products offer different interpolation methods for creation of digital model of landscape. Its accuracy and quality is impacted by selection of an interpolation method and precision input data. Several studies have demonstrated that various spatial interpolation techniques perform differently depending on the type of attribute, geometrical configuration of the samples, spatial resolution, world region, etc. Hence, selecting the best interpolation technique for each particular situation is a key factor. The major objective of this paper is to assess the spatial variability of elevation* *data in Iraq by comparing different interpolation procedures. The elevation data were interpolated using a deterministic method (Inverse square distance) and geostatistical methods in ArcGIS. Cross-validation is a sample reuse algorithm for quantitative comparison of experimental performance of alternative interpolation methods. Cross validation can help make an informed decision as to which method provides the best results. Two diagnostic statistics are mainly considered in this paper from the results mean error, and root mean square error.*

Keywords: Digital elevation model, Interpolation techniques, Validation.

1. Introduction

Digital Elevation Model (DEM) is an important component of GIS applications in many socio-economic areas. Especially, DEM has a very important role in monitoring and managing natural resources, preventing natural hazards, and supporting spatial decision making. Digital elevation model (DEM) is an important part of the spatial data infrastructure [1], [4]. Usually, DEM is built by interpolation from a limited set of sample points. Thus, the accuracy of the DEM is depended on the used interpolation method. The inverse distance weighted (IDW) interpolation determines the elevation of a specific point using a linearly weighted combination of the elevations of nearby located sample (known) points [6].

The kriging interpolation assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain the variation in the surface [7]. Kriging fits a mathematical function to a specified number of points, or all points within a specified radius, to determine the output value for each location. It is a multi step process including: exploratory statistical analysis of the data, variogram modeling, creating the surface. Kriging is most appropriate when there is a spatially correlated distance or directional bias in the data. Kriging is similar to IDW in that it weights the surrounding measured values to derive a prediction for an unmeasured location.

However, in kriging, the weights are based not only on the distance between the measured points and the prediction location but also on the overall spatial arrangement of the measured points. To use the spatial arrangement in the weights, the spatial autocorrelation must be quantified through empirical semivariograms. The semivariogram can have one of the following models: circular, spherical, exponential, Gaussian, and linear [7].

Geostatistical methods can provide reliable estimates at unsampled locations provided that the sampling interval resolves the variation at the level of interest [3]. Spatial prediction techniques, also known as spatial interpolation techniques, differ from classical modeling approaches in that they incorporate information on the geographic position of the sample data points [1].

All interpolation methods have been developed based on the theory that points closer to each other have more correlations and similarities than those farther. Ordinary kriging is one of the most basic of kriging methods. It provides an estimate at an unobserved location of variable z, based on the weighted average of adjacent observed sites within a given area. Also in Inverse Distance Weighting (IDW) method, it is assumed substantially that the rate of correlations and similarities between neighbors is proportional to the distance between them that can be defined as a distance reverse function of every point from neighboring points.

It is necessary to remember that the definition of neighboring radius and the related power to the distance reverse function are considered as important problems in this method. The main factor affecting the accuracy of inverse distance interpolator is the value of the power parameter [6]. In this study, we compared different method including Ordinary Kriging and Inverse Distance

Weighting. Important step within the geostatistic analysis is adjustment of the experimental model by appropriate type of the variogram model.

Each model has its own basic structure, which is the function of the distance among data and we express her by the parameters sill (upper limit of the variogram model) and range (boundary distance between selection locality at which the model reaching of the maximum value or „sill"). The example of a variogram with typical components is shown in figure (1).

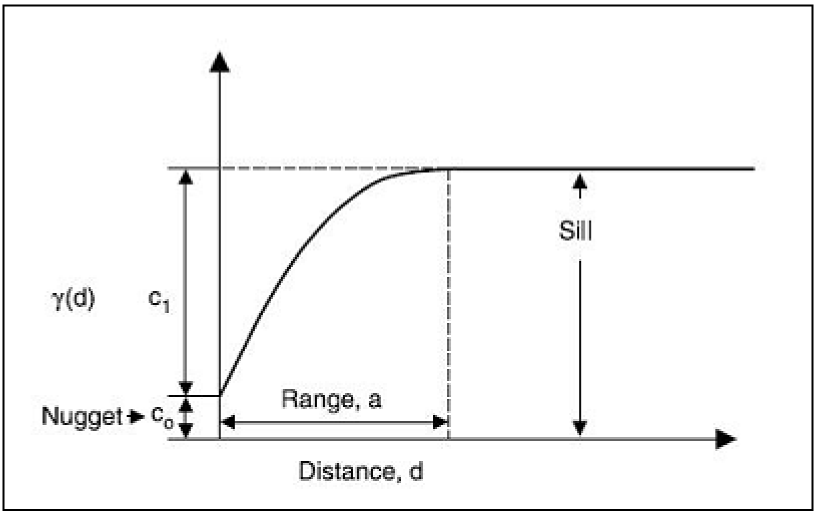

Figure (1): Typical Parameters of a semivariogram.

2. Modeling semivariogram and covariance

Kriging is divided into two distinct tasks: quantifying the spatial structure of the data and producing a prediction. Quantifying the structure, known as variography, is where the fit of a spatial-dependence model to the data. Making a prediction for an unknown value for a specific location, kriging will use the fitted model from variography, the spatial data configuration, and the values of the measured sample points around the prediction location.Variography is the process of estimating the theoretical Semivariogram. It begins with exploratory data analysis, then computing the empirical semivariogram, binning, fitting a semivariogram model, and using diagnostics to assess the fitted model. The semivariogram and covariance function quantify the assumption that things nearby tend to be more similar than things that are farther apart. They both measure the strength of statistical correlation as a function of distance [3].

3. Validation:

Cross-validation is a sample reuse algorithm for quantitative comparison of experimental performance of alternative interpolation methods. The cross-validation procedure ignores an observation in the data set and uses the remaining observations to estimate the ignored observation using a particular interpolation technique. The process is repeated for each observation in the data set to obtain a complete set of interpolated values by each technique. Each set of interpolated. Validation should be carried out before producing the final surface, where it helps in making an informed decision as to which model provides the best predictions should have[5].

4. Study Area:

Figure(2) represent the study area located west of Al Razzaza lake, digital elevation model of the region is shown in figure(3).

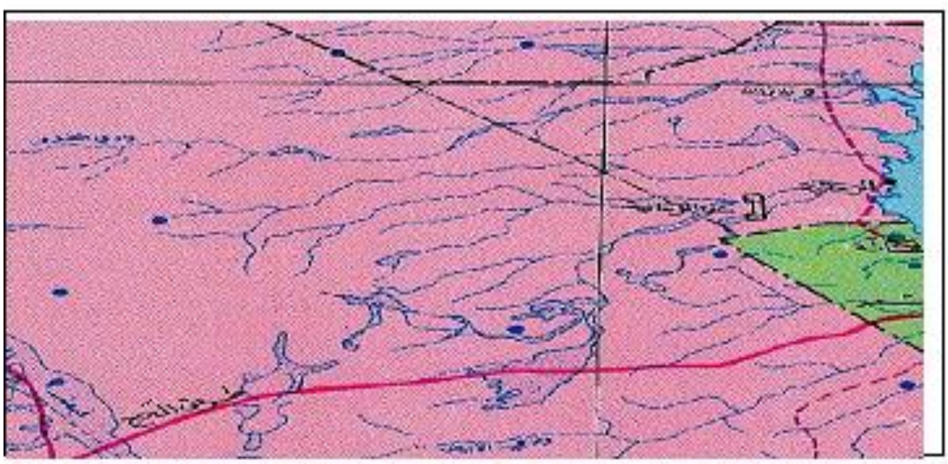

Figure (2): The Study Area.

Figure (3): The produced DEM of Study Area.

5. Results and conclusions:

Figure (4) represent the triangular irregular networks (TIN).

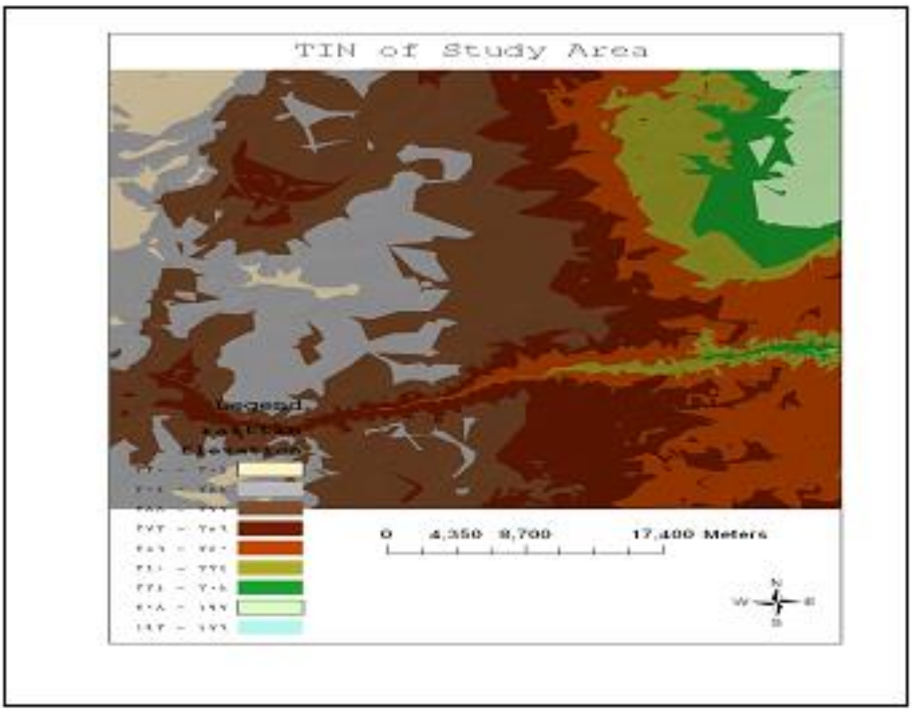

Figure (4): Triangular Irregular Network (TIN).

A network of non-overlapping triangles that follow Delaunay criterion used in digital elevation models.

Interpolation technique plays an important role in achieving a high accuracy of DEM. The influence of interpolation technique on the DEM accuracy depends on the type of topography, and the distribution of sample points.

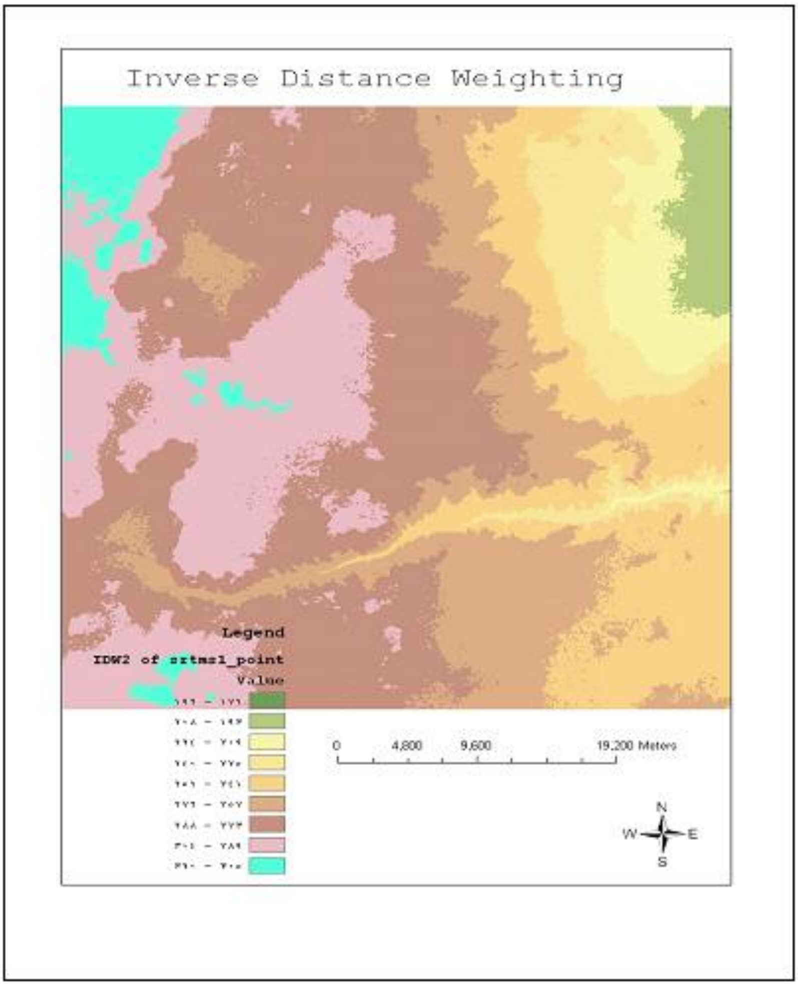

Figure (5): Surface produced using IDW method.

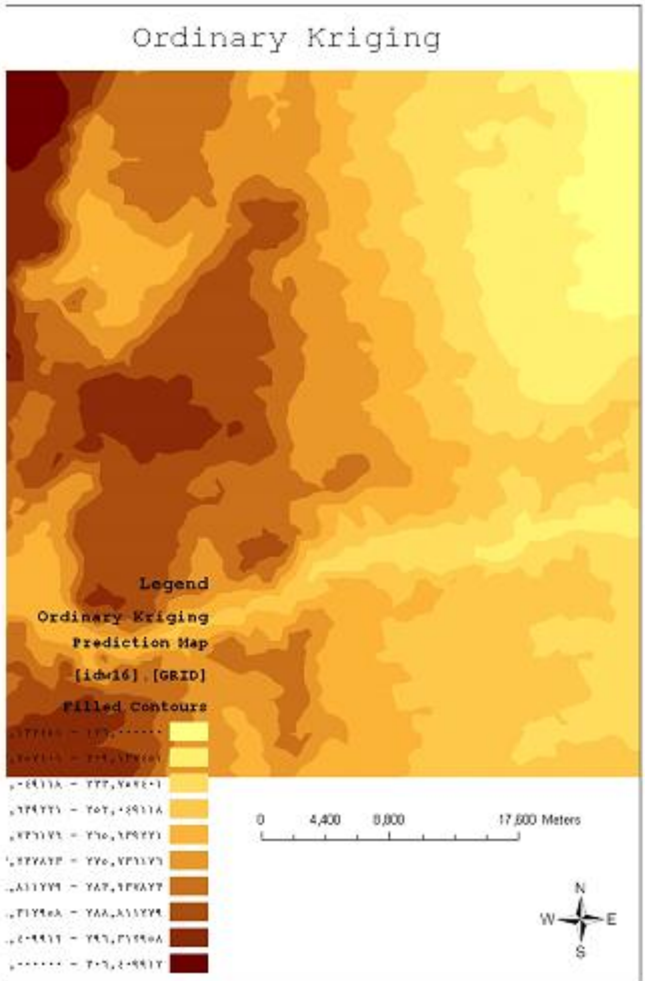

Figure (6): Surface produced using kriging method.

This research has examined two interpolation techniques IDW and ordinary kriging as shown in figures (5), and (6). The influence of interpolation technique on the DEM accuracy depends on the type of topography, and the distribution of sample points, what is directly related to the surveying method. The relation between predicted and measured points is shown in figure (7).

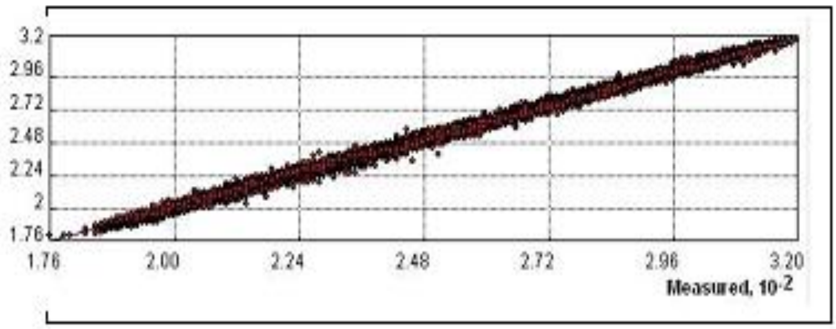

Figure (7): The relation between measured and predicted elevation data.

The following represent the results obtained from cross validation, which serve as diagnostics that indicate whether the model and/or its associated parameter values are reasonable. If the mean prediction error is near zero, predictions are centered on the measurement values. The closer the predictions are to their true values the smaller the root-mean-square prediction errors. The average root-mean-square prediction errors are computed as the square root of the average of the squared difference between observed and predicted values. For a model that provides accurate predictions, the root-mean-squared prediction error should be as small as possible. The average standard error and the mean standardized prediction error should be as small as possible. Also the root-mean-squared standardized prediction error should be close to one.

Prediction Errors

Mean: -0.000023
Root-Mean-square: 1.086
Average Standard Error: 1.442
Mean Standardized: -0.000013
Root-Mean- Square- standardized: 0.753.

References:

[1] Burrough, P. A. &McDonnell, R. A. 1998 Principles of Geographical Information Systems. Oxford University Press, Oxford.

[2] V. Chaplot et al., Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density, Geomorphology 77 (2006) 126.

[3] Isaaks, H. E. & Srivastava, R. M. 1989 an Introduction to Applied Geostatistics. Oxford University Press, Oxford.

[4] Z.L. Li, Q. Zhu, C. Gold, *Digital terrain modeling: principles and methodology*, CRC Press, Boca Raton, 2005.

[5] J. McCoy, K. Johnston, Using ArcGIS Spatial Analyst, ESRI Press, Redland, CA, USA, 2001.

[6] L. Mitas, and H. Mitasova, General variational approach to the interpolation problem, Computer and Mathematic Application 16 (1988) 983.

[7] M.A. Oliver, Kriging: a method of interpolation for geographical information systems, International Journal of Geographic Information Systems 4 (1990) 313.

التدقيق المتقاطع لاستنباط بيانات الارتفاع

د.حسين زيدان علي
خبير معالجة الصور الفضائية
وزارة العلوم والتكنولوجيا

المستخلص:

تتطلب معظم التحاليل المكانية الجغرافية بيانات مستمرة وهذة الدراسة مصصمة لانتاج مثل هذا السطح. يعتبر الموديل الرقمي للارض جزء مهم ضمن انتاج انظمة المعلومات المرتبطة بالارض، ويعتبر اداة مهمة في التطبيقات التي تستخدم لنمذجة سطح الارض مثل الجيومورفولوجي ،الهايدرولوجي ، الجيولوجي ، علم الخرائط ، علم البيئة ، ....الخ. تستطيع العديد من البرامجيات توفير طرق مختلفة للاستنباط لانتاج الموديل الرقمي للارض . تتأثر دقتها ونوعيتها بأختيار طريقة الاستنباط ودقة البيانات .لقد بينت العديد من الدراسات بأن تقنيات الاستنباط المكانية المختلفة تسلك سلوكا مختلف اعتمادا على نوع المعلومة الوصفية ، الترتيب الهندسي للنماذج ، قابلية التحليل المكانية ، المنطقة الجغرافية ، ....الخ . هنا ، فأن اختيار افضل تقنية للاستنباط تعتبر عامل حاكم . ان الهدف الرئيسي من البحث هو تقيم التغاير المكاني لبيانات الارتفاع في العراق بمقارنة طرق استنباط مختلفة . لقد تم استنباط بيانات الارتفاع باستخدام الطرق الحتمية و الطرق الاحصائية الارضية في نظم المعلومات الجغرافية. يعتبر التدقيق المتقاطع كخوارزمية لاعادة استخدام النموذج لهدف المقارنة الكمية للاداء التجريبي لطرق الاستنباط المختلفة . يستطع التدقيق المتقاطع المساعدة في تسهيل مهمة اتخاذ القرار عن اي الطرق التي تعطينا افضل النتائج . تم اعتبار طريقتين احصائية للتشخيص بشكل رئيسي في هذا البحث من النتائج وتشمل معدل الخطأ ، و خطأ جذرمربع المعدل.

الكلمات المفتاحية: موديل الارتفاع الرقمي ، تقنيات الاستنباط ، التدقيق.