

UDK 528.344

A COMPARATIVE ANALYSIS OF DIFFERENT DEM INTERPOLATION METHODS IN GIS: CASE STUDY OF RAHOVEC, KOSOVO

Besim AJVAZI¹, Kornél CZIMBER^{2*}¹*Faculty of Civil Engineering and Architecture, University of Prishtina “Hasan Prishtina”, Prishtine, Kosovo*²*Institute of Geomatics, Forest Exploration and Water Management, University of Sopron, Sopron, Hungary*

Received 21 January 2019; accepted 01 March 2019

Abstract. Geographic Information System (GIS) uses geospatial databases as a model of the real world. Since we are speaking of the real world this entails that in many cases the information about the Earth's surface is highly important. Therefore, the generation of a surface model is significant. Basically, the quality of the Digital Elevation Model (DEM) depends on the source data or techniques used to obtain them. However, different spatial interpolation methods used for the same data may provide different results. This paper compares the accuracy of different spatial interpolation methods such as IDW, Kriging, Natural Neighbor and Spline. Since interpolation is essential in DEM generation, then is important to do a comparative analysis of such methods to find out which one provides more accurate results. The DEM data set used is from an aero photogrammetric surveying. According to this data set, three scenarios are performed for each of the methods. Selected random control points are derived from the base data set. The first example includes 10% of randomly selected control points, the second example includes 20%, and the third example includes 30%. The Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) are calculated. We find out that results do not have much difference; however, the most accurate results are derived from the Spline and Kriging interpolation methods.

Keywords: GIS, DEM, Spatial Interpolation, IDW, Kriging, Natural Neighbor, Spline.

Introduction

The usage of GIS tools facilitates a more efficient decision-making process because of their spatial analysis capabilities. The utility of the decision-making process, in operational terms, will be significantly improved when surface models are considered. GIS software packages offer several methods to create reliable surface models. Once we have obtained sample points from certain techniques, by interpolation methods we can create a data structure for the entire area. Hence, interpolation is a crucial operation in GIS. It can help the visualization process and a better understanding of the dataset for the respective areas. The main objective of this paper is to assess the impact of different interpolation methods on the accuracy of DEM generation. In some cases, the quality of an interpolation method varies in relation with the data sample size and landform types, but in this case the research was carried out considering the study area as a whole.

In the context of being an essential tool for estimating spatial continuous data, interpolation methods have been

applied to various disciplines concerned with the Earth's surface (Tan & Xu, 2014). One of the most common applications of interpolation methods is in the creation of a DEM for the purpose of agricultural land analysis. Surface modeling is one of the prime approaches that can be used to assess the spatial variability of agricultural fields and their surrounding ecosystems (Samsuzana, 2008). In addition, spatial interpolation methods are essential for mapping soil properties (Robinson & Metternicht, 2005). Therefore, it is important to analyze, compare and find the appropriate method for spatial interpolation regarding study area and available data in advance. The selected study area is in one of the municipalities with the biggest area of agricultural land in Kosovo. Although many studies have been conducted regarding the accuracy of various interpolation methods for generation of the DEM, they do not offer consistent results among each other. So, it is still difficult to select which is the most appropriate method for spatial interpolation regarding Digital Elevation Model.

*Corresponding author. E-mail: czimber.kornel@uni-sopron.hu

1. Study area

Municipality of Rahovec is located in the south-western part of Kosovo with a total area of 276 km² and 76,577 inhabitants. The territory of this municipality largely has a good geographic position, which covers the central part of Dukagjini region. It is positioned between 42° 30' and 42° 50' north latitude and 20° 21' and 20° 55' longitude. Regarding the topography, settlements are divided into three categories: valley-plains, hills and mountains. Most of the area is lying in an altitude which ranges from 350 to 750 m. This municipality has a continental climate with Mediterranean influence. Considering statistical data regarding climate and soil, this municipality provides a suitable location for agriculture, respectively viticulture. The annual rainfalls reach 765 mm and the average air temperature is 11.8 °C. Another advantage of this municipality regarding agriculture is high quality of soil. For the purposes of this research, the study area of 7.38 km² was selected (Figure 1).

2. Data and methodology

The dataset used in this research is generated based on aerial photogrammetric surveying in Kosovo in 2009. Regarding the study area, 15583 points with known values (x, y, z) are selected. They do cover the whole study area. In order to analyze the accuracy of the DEM by different

interpolation methods, the test data were divided into three scenarios. The first one included 10% of the base points or 1536 points in total. The second scenario included 20% and the third one 30%, or 3117 points respectively 4675 points in total (Figure 2). Three groups of the test points are randomly selected and are uniformly distributed throughout the study area. Within one set of data, base points were used to model the DEM, while the test data was used to compare them with the predicted values from the DEM. By this approach we are able to analyze the accuracy of each interpolation method, depending on different set of control points. But the main aim is to compare the interpolation methods and not the quality of different data sources. This section will provide an overview of the theory behind the spatial interpolation methods.

2.1. Interpolation methods

The data regarding the Earth's surface can be visualized by different data formats, but in this case the grid format is taken into consideration. A grid representation of a surface is considered to be a functional surface because for any given x, y location, it stores only a single Z value as opposed to multiple Z values (Childs, 2004). By using GIS tools, it is possible to interpolate the height for each cell which will derive from the sample points. There are different approaches for interpolation. In this study, the four (IDW, Kriging, Natural Neighbor and Spline) commonly

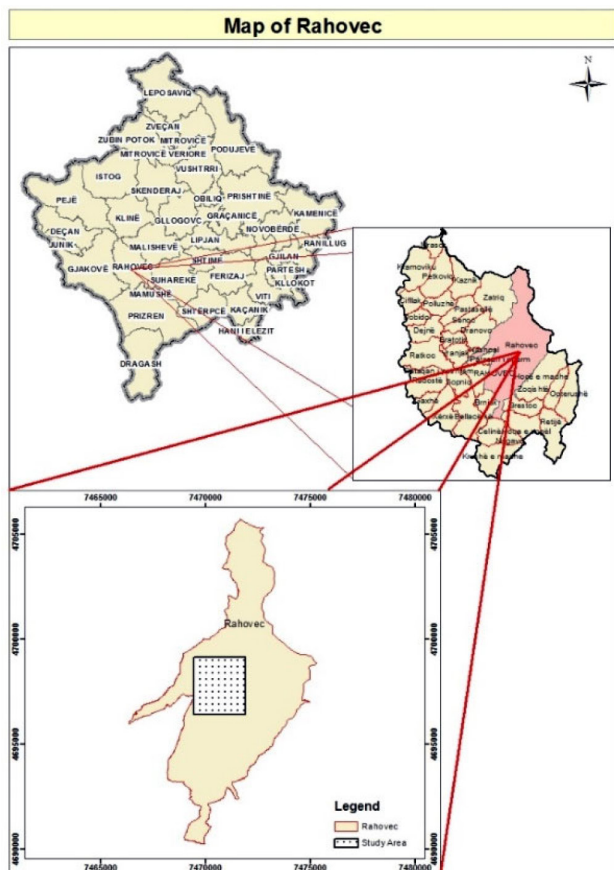


Figure 1. Study area (Rahovec)

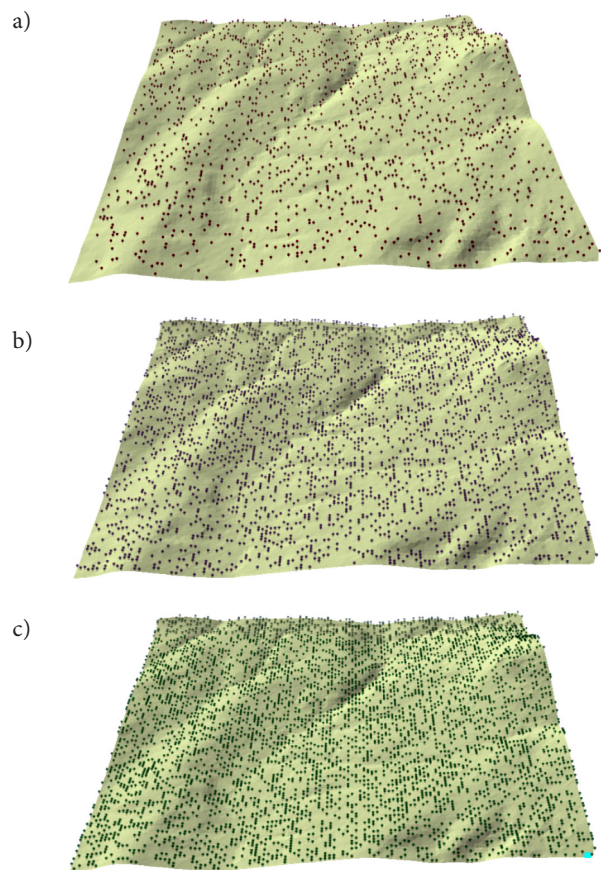


Figure 2. Control points: a) 10% of base points, b) 20% and c) 30%

used spatial interpolation methods are applied. Generally, interpolation methods are classified into two groups: deterministic and geostatistical analysis. The IDW, Natural Neighbor and Spline method belong to the deterministic group while Kriging belongs to the geostatistical analysis. Deterministic methods use mathematical functions to predict unknown values, while geostatistical analysis use statistics and mathematical functions as well (Johnston, VerHoef, Krivoruchko, & Neil, 2001). However, deterministic methods can be divided into two groups as well, namely: global and local. Local methods predict the value of an unknown point based on the values of neighboring pixels. Compared to the global methods, local methods usually yield less smooth surfaces. Local methods are not as sensitive to outliers, because their effects will not influence the entire interpolated surface, but only local regions of the interpolated surface (Tan & Xu, 2014). Global interpolation methods such as polynomial interpolation functions use all the available sample points to generate predictions for a particular point (Arun, 2013). In the following sections an overview for each method is introduced and accuracy analyses are performed.

2.1.1. Inverse distance weight (IDW)

The IDW approach is a local deterministic interpolation technique that calculates the value as a distance-weighted average of sampled points in a defined neighborhood (Burrough & McDonnell, 1998). Sample points are points with a known value. Search neighborhood of the sample points as well as the number of points can be defined by the user. In addition, by setting the power value p we can modify the influence of nearby data points in a search neighborhood. A higher power value results in less influence from distant points. The IDW approach can be used mainly when dense sample points exists and when they are distributed better throughout the area and are not clustered (Johnston, VerHoef, Krivoruchko, & Neil, 2001).

The main equation is (Johnston et al., 2001):

$$\hat{Z}(s_o) = \sum_{i=1}^N \lambda_i Z(s_i),$$

where: $\hat{Z}(s_o)$ is the value we are trying to predict for location s_o ; N is the number of measured sample points surrounding the prediction location that will be used in the prediction; λ_i are the weight assigned to each measured point that we are going to use. These weights will decrease with distance.

$\hat{Z}(s_i)$ is the observed value at the location s_i .

The formula to determine the weights is the following:

$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^N d_{i0}^{-p};$$

$$\sum_{i=1}^N \lambda_i = 1.$$

As the distance becomes larger, the weight is reduced

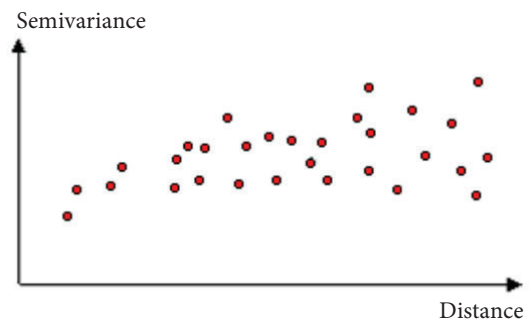


Figure 3. Semivariogram graph

by a factor of p . The quantity d_{i0} is the distance between the predicted location, s_o and each of the measured locations, s_i .

2.1.2. Kriging

Kriging refers to a geostatistical group of spatial interpolation. Like other local methods, Kriging works based on a principle that things closer together are more alike than things further away. The main difference between Kriging and deterministic methods is in the usage of the statistical model, which includes autocorrelation. This relationship between the values of data points and the distance between them is known as spatial autocorrelation. Kriging is the most appropriate when you know there is a spatially correlated distance or directional bias in the data. In order to model the surface while using Kriging method, initially a semivariogram based on known data points will be created (Figure 3). Firstly, it is necessary to calculate the distance and squared differences between each pair of locations. The empirical semivariance is 0.5 times the difference squared. Since a dataset may have a large number of points, it means that we may have many pairs of points. While calculating the pairs in this way is difficult, grouping them into lag bins and taking the average semivariance of all pairs of points is a better solution. After that, the next step is fitting a model to the semivariogram. The following models are possible: circular, spherical, exponential, Gaussian, and linear. Based on the fitting model it is possible to predict the unknown values by Kriging.

This method is often used in the mining and petroleum industries, geochemistry, geology, soil science and ecology, where its statistical properties are of great value and on the contrary it has been less successful for applications where local geometry and smoothness are the key issues and other methods prove to be competitive or even better (Mitas & Mitasova, 1999).

2.1.3. Natural neighbor

Based on the natural neighbor coordinates, Robin Sibson developed a weighted average interpolation technique that he named natural neighbor interpolation (Ledoux & Gold, 2005). The basic equation used in this method is identical with IDW. It uses a weighted average of neighboring sample points. The natural neighbors of any point are those associated with the neighboring Voronoi polygons. Initially,

a Voronoi diagram is constructed from the sample points and a new Voronoi polygon is then created around the interpolation point. The proportion of overlap between this new polygon and the initial polygons are then used as weights. The surface created by this approach does not exceed the minimum or maximum values in the sample data set (Garnero & Godone, 2013).

2.1.4. Spline

Spline is a deterministic interpolation method where the predicted values are estimated using a function minimizing the total curvature of the surface. The result is a smooth surface that passes through all known points. There are two types of spline interpolation: regularized and tension. Usage of the regularized type results in smoother surface as compared to the tension type. The impact on smoother value is through the weight parameter. By selecting the regularized type and setting a high value of the weight parameter the surface will be smoother. On the other hand, by selecting the tension type and setting a high value of the weight parameter, in the surface will be coarser. The Number of points identifies the number of points used in the calculation of each interpolated cell. The greater the value of the number of points the smoother the surface of the output raster.

Referring to Johnston et al. (2001), the algorithm for spline approach uses a linear combination of n functions, one for each known point.

$$\hat{Z}(s_0) = \sum_{i=1}^n \omega_i \phi(\|s_i - s_0\|) + \omega_{n+1},$$

where $\phi(r)$ represents the interpolation function, $\|s_i - s_0\|$

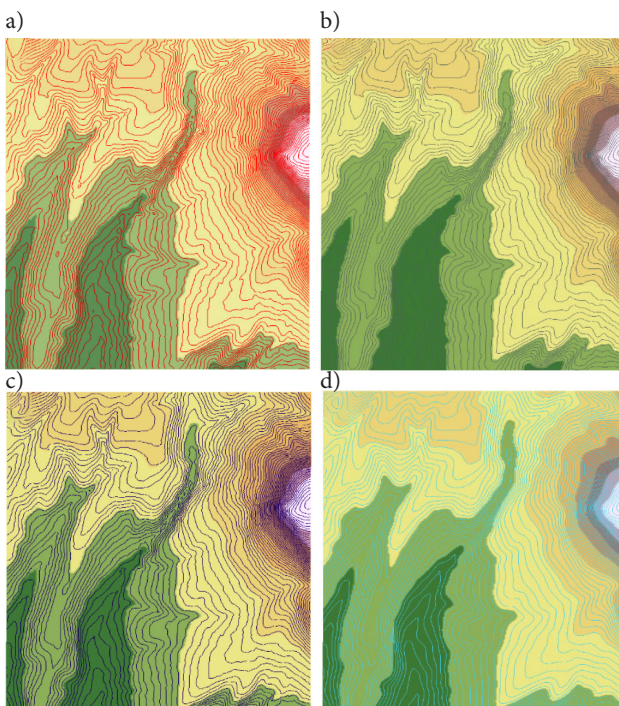


Figure 4. DEM generation with contours based on 14047 points: a) Spline, b) Natural Neighbor, c) IDW, d) Kriging

the Euclidean distance r between an unknown point s_0 and an observed one s_i , while ω_i , with $i = 1, 2, \dots, n+1$, are weights.

Weights are assigned according to the distance of known points, under the constraint that, in their locations, the function must necessarily give the measured value. These conditions lead to the computation of a system of N equations with N unknown points with a unique solution. Splines produce good results with gently varying surfaces, and thus are often not appropriate when there are large changes in the surface values within a short horizontal distance (Robinson & Metternicht, 2005).

To compare different interpolation methods, we examined the difference between the known data and the predicted data using the mean absolute error and root mean squared error, equation as below (Johnston et al., 2001; Webster & Oliver, 2001; Kravchenko & Bullock, 1999).

$$MAE = \frac{1}{N} \sum_{i=N}^N \{z(x_i) - \hat{z}(x_i)\};$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=N}^N \{z(x_i) - \hat{z}(x_i)\}^2},$$

where $\hat{z}(x_i)$ is the predicted value, and $z(x_i)$ is the observed (known) value.

3. Results and discussions

Four interpolation methods as introduced theoretically in section three are applied for DEM generation in four examples as well. Dataset used in these examples is from the same source. This study intends to find the most appropriate spatial interpolation method for the selected study area. In order to analyze the accuracy, three scenarios are supposed. This is done to verify whether the results of accuracy assessment will differ depending on the number of control points.

3.1. The first scenario

In the first case of the study, 10% of the sample points were taken as control points. It means that the height values of such points are compared to the predicted values as generated from the DEM. A total of 1536 points were used as control points toward 14047 sample points, which were used to model the surface.

Referring Figure 4 which shows the generated DEM by the four interpolation methods, it is not easy to note the differences between them even if some differences practically exist. This figure, as mentioned above presents the visibility comparisons on the generated DEM for the four interpolation methods. Furthermore, some error analyses are performed (see Table 1 and Figure 5). Based on the mean absolute error (MAE), the best performing interpolation method was Spline, while the worst one was Inverse Distance Weight. Also, according to results of (RMSE), the situation is the same.

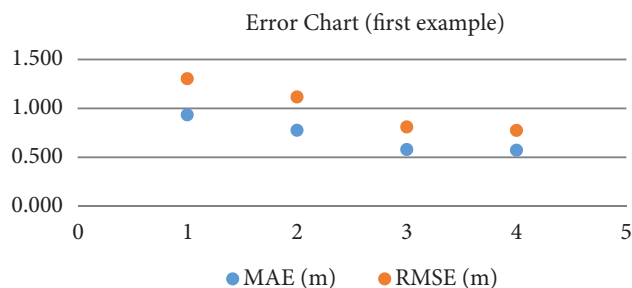


Figure 5. Error chart of the four interpolation methods for 14047 sample points

Ledoux and Gold (2005) in his article mentioned that spline interpolation method performs better with irregularly distributed data and is valid in any dimensions. The initial data used in the study are regularly distributed, however, after randomly selected the 10% of the points, the most of the points will be irregularly distributed.

Table 1. Results on MAE and RMSE analysis

Method	MAE (m)	RMSE (m)
IDW	0.933	1.302
Kriging	0.776	1.115
Natural Neighbor	0.579	0.808
Spline	0.572	0.774

3.2. The second scenario

In the second scenario of the paper, 20% of the sample points were taken as control points. A total of 3117 points were used as control points toward 12466 sample points, which were used to model the surface. Figure 6 shows the differences on the DEM generation based on the four interpolation methods.

According to the results on error analysis, some minor differences are visible compared to the first example. Referring to MAE, Kriging provided the best results while Spline is only 2 cm less accurate than Kriging (Table 2 and Figure 7). IDW has highest value. A similar situation is observed with RMSE results.

Table 2. Results on MAE and RMSE analysis

Method	MAE (m)	RMSE (m)
IDW	1.017	1.429
Kriging	0.568	0.804
Natural Neighbor	0.630	0.894
Spline	0.587	0.820

3.3. The third scenario

In the third scenario, 30% of the sample points were taken as control points. A total of 4675 points were used as control points toward 10908 sample points, which were used to model the surface. Figure 8 shows the differences on the

DEM generation based on the four interpolation methods.

Based on the error analysis results, smaller differences appear in this case as compared to the second example. In this case, the best results are from Spline, however, they are almost equal with Kriging. IDW again provided the worst results (Table 3 and Figure 9).

Table 3. Results on MAE and RMSE analysis

Method	MAE (m)	RMSE (m)
IDW	1.079	1.476
Kriging	0.602	0.838
Natural Neighbor	0.655	0.906
Spline	0.592	0.815

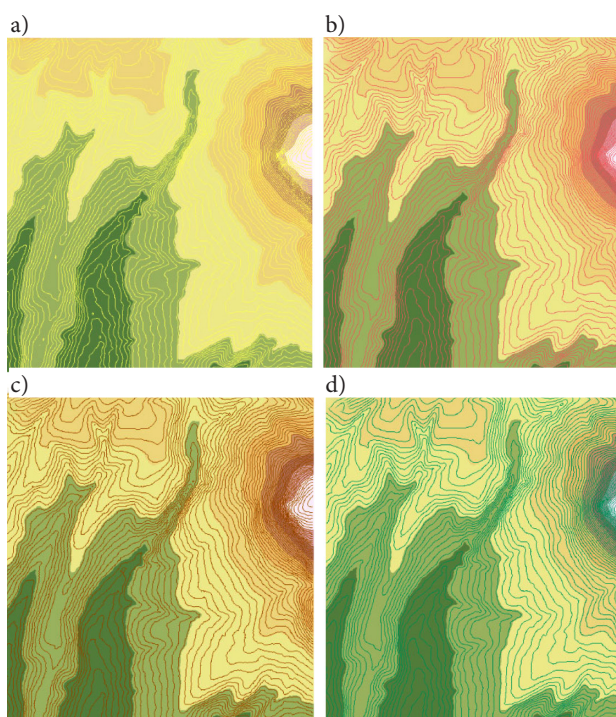


Figure 6. DEM generation with contours based on 12466 points: a) Spline, b) Natural Neighbor, c) IDW, d) Kriging

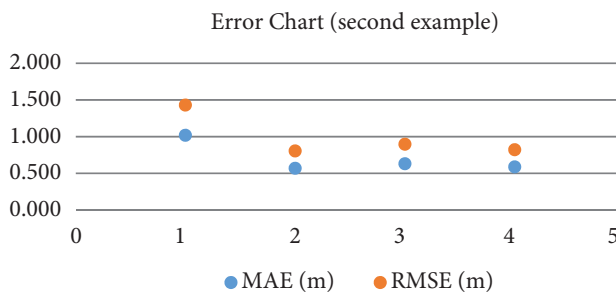


Figure 7. Error chart of the four interpolation methods for 12466 sample points

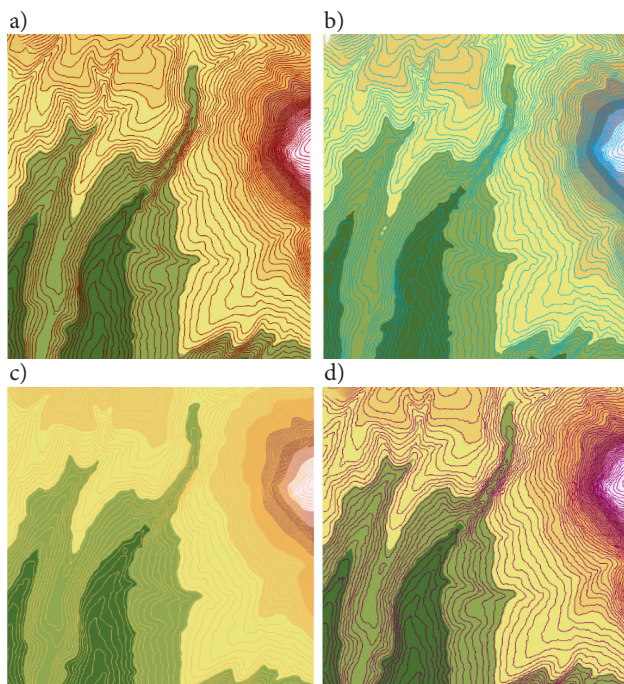


Figure 8. DEM generation with contours based on 10908 points: a) Spline, b) Natural Neighbor, c) IDW, d) Kriging

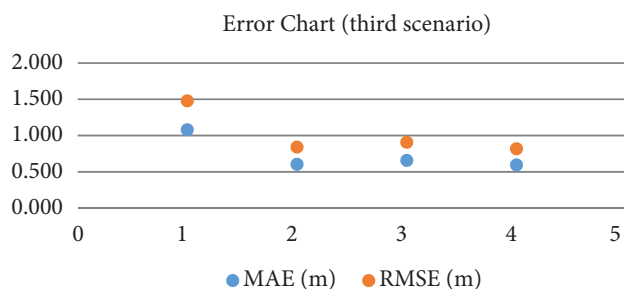


Figure 9. Error chart of the four interpolation methods for 10908 sample points

Conclusions

As introduced in this paper, there are several spatial interpolation methods that can be used by applying GIS software. It is important to analyze them by comparing the results they provide based on the available sample points. In many cases, DEM is essential while doing spatial analyses, especially for agricultural land management. This paper intended to analyze and compare which one of the interpolated methods provides a better quality of the DEM based on the dataset and study area as well. The following conclusions for each of the four interpolation methods are achieved: IDW can be used mainly when dense sample points exist and when they are distributed better throughout the area and are not clustered, as they were in the study; Kriging is most appropriate when there is a

spatially correlated distance or directional bias in the data; Natural neighbor is particularly useful if we have break lines or an irregularly shaped data area; Spline showed to be more useful when the range of the sample values may not include the extremes of the phenomenon being interpolated. Considering the three examples that were demonstrated in the study as well as the quality control analysis, it can be concluded that Spline and Kriging are the methods which provided a better quality of the DEM in a more efficient manner.

References

- Arun, P. V. (2013). A comparative analysis of different DEM interpolation methods. *The Egyptian Journal of Remote Sensing and Space Sciences*, 16, 133-139. <https://doi.org/10.1016/j.ejrs.2013.09.001>
- Burrough, P. A., & McDonnell, R. A. (1998). *Principles of geographic information systems* (pp. 333-335). New York: Oxford University Press.
- Childs, C. (2004). Interpolating surfaces in ArcGIS spatial analyst. In *ArcUser, ESRI* (pp. 32-35). California.
- Garnero, G., & Godone, D. (2013). Comparisons between different interpolation techniques. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W3, 2013, 139-144.
- Johnston, K., VerHoef, J., Krivoruchko, K., & Neil, L. (2001). *Using ArcGIS™ Geostatistical Analyst*. ESRI™, US.
- Kravchenko, A. N., & Bullock, D. G. (1999). A comparative study of interpolation methods for mapping soil properties. *Agronomy Journal*, 91, 393-400. <https://doi.org/10.2134/agronj1999.00021962009100030007x>
- Ledoux, H., & Gold, C. (2005). An efficient natural neighbour interpolation algorithm for geoscientific modelling. In P. F. Fisher (Ed.), *Developments in spatial data handling* (pp. 97-108). Springer, Heidelberg. https://doi.org/10.1007/3-540-26772-7_8
- Ledoux, H., & Gold, C. M. (2004). An efficient natural neighbour interpolation algorithm for geoscientific modelling. In P. Fisher (Ed.), *Developments in spatial data handling – 11th International Symposium on Spatial Data Handling* (pp. 97-108). Springer.
- Mitas, L., & Mitasova, H. (1999). Spatial interpolation. *Geographical information systems: principles, techniques, management and application* (pp. 481-492). John Wiley & Sons.
- Robinson, T. P., & Metternicht, G. (2005). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture*, 50, 97-108. <https://doi.org/10.1016/j.compag.2005.07.003>
- Samsuzana, A. A. (2008). *Development of digital elevation models (DEMs) for agricultural applications*. Graduate Theses and Dissertations. Paper 11017.
- Tan, Q., & Xu, X. (2014). Comparative analysis of spatial interpolation methods: an experimental study. *Sensors & Transducers*, 165(2), 155-163.
- Webster, R., & Oliver, M. A. (2001). *Geostatistics for environmental scientists*. John Wiley and Sons, Brisbane, Australia.