Geodesy and Cartography
ISSN 2029-6991 / eISSN 2029-7009
2023 Volume 49 Issue 1: 60–65
https://doi.org/10.3846/gac.2023.16591

UDC 528.932.3

THE EFFECT OF THE NUMBER OF INPUTS ON THE SPATIAL INTERPOLATION OF ELEVATION DATA USING IDW AND ANNS

Sara RESPATI*, Totok SULISTYO

Department of Civil Engineering, Balikpapan State Polytechnic, Balikpapan, Indonesia

Received 24 February 2022; accepted 17 February 2023

Abstract. Spatial interpolation is a required method to generate a continuous surface such as Digital Elevation Model (DEM) because field investigation for most of the surface's part is time-consuming with a high demand in both human resources and monetory cost. One of the most used deterministic interpolation models is Inverse Distance Weighting (IDW) model. The model takes several neighbors' information, and the weights are constructed based on the distance between the interpolated point and the neighbors' points. From the machine learning model, Artificial Neural Networks (ANNs) model has also been used for spatial interpolation. The input of ANNs model is also one of the parameters that need to be defined when building the model. This paper evaluated the effect of the number of inputs (neighbors) on the elevation interpolation accuracy. We applied IDW and ANNs to interpolate the elevation of Balikpapan City, Indonesia. The results show that the accuracy increases significantly when the number of inputs is between one and three. However, after three inputs, additional input would not change the accuracy significantly. ANNs performed better than IDW. For three or more inputs, the MAE of ANNs and IDW interpolations are below 1.1 and around 2 meters, respectively.

Keywords: artificial neural network, digital elevation model, elevation interpolation, interpolation, inverse distance weighting, spatial interpolation.

Introduction

The spatial surveys are generally conducted at some number of points in the field. Spatial interpolation methods help to generate a continuous surface from scattered sample points. The continuous surface can be used for various spatial data applications such as urban planning, preliminary mineral exploration, geologic mapping, etc. One of the most popular used statistical-based models for spatial interpolation of surface elevation is the inverse distance weighting (IDW) model (Ajvazi & Czimber, 2019; Bartier & Keller, 1996; Ikechukwu et al., 2017; Li et al., 2004; Shiode & Shiode, 2011). The concept behind IDW is simple. Attributes of points in a geographic area are deemed to correlate with each other, however, the correlations are inversely related to the distances between the points. The distance is specified using Euclidean distance i.e. the distance between two points is measured using a straight line (Peterson & Pearse, 2017). Li et al. (2004) tested different interpolation methods and reported that IDW is suitable for hilly and flat areas. IDW has been also applied for spatial interpolation of other variables: Tomczak (1998) applied IDW for interpolation of rainfall magnitude; Robinson and Metternicht (2006) used IDW for soil properties interpolation; Nusret and Dug (2012) applied IDW for spatial interpolation of annual precipitation. IDW is also available as GIS package making it easy to be applied and widely applied (Jumaah et al., 2019; Keskin et al., 2015; Noori et al., 2014).

From the machine learning-based model, artificial neural networks (ANNs) have been applied for non-linear regression. One of the applications is a spatial interpolation. Rigol et al. (2001) employed feed-forward back-propagation ANNs for spatial interpolation of daily minimum air temperature. The result shows that the ANNs model can account for the non-linear relationship between the data. Merwin et al. (2002) examined the performance of ANNs in interpolating DEM. The study investigated the interpolation accuracy with regards to the effect of different input sizes (six and sixteen neighbors) on the low and high reliefs interpolation accuracy. Liu et al. (2009) proposed generalized regression neural network residual kriging (GRNNRK) to interpolate terrain surfaces. The result shows that the model achieved better accuracy performance compared to that of kriging. Sivapragasam et al. (2010) proposed ANNs for hydrological variable interpolation, such as rainfall and groundwater level. The result shows that ANNs outperformed Kriging for spatial interpolation.

*Corresponding author. E-mail: *sara.wibawaning@poltekba.ac.id*

Both IDW and ANNs are affected by the number of their neighbors or input. In IDW, the neighbors can be defined based on the number of neighbor points and the technique will take the exact number of nearest points. It also can be defined using a search radius where all the points that lie within the radius will be considered in the interpolation. To build an ANNs model, a fixed number of inputs is required. Therefore, a search radius technique may not be suitable for a simple ANNs model application. Besides, by setting a fixed number of inputs, it allows further points to be taken into account in the interpolation. This is of benefit to the application of interpolation with low-resolution survey data. A study to investigate the effect of points considered in spatial interpolation would give an insight into the required survey activities.

Therefore, in this study, we investigate the effect of the number of neighbors in the interpolation using IDW and ANNs. IDW and ANNs represent the conventional spatial interpolation model, and the machine learning-based model, respectively. The remainder of this paper is as follows. A brief explanation of IDW and ANNs, also their structures used in this study are presented in the 1 section. The data is presented in the 2 section. The result and conclusion are presented in 3 section.

1. Methodology

1.1. IDW

IDW is a deterministic spatial interpolation technique proposed by Shepard (1968). It estimates unsampled values from a set of weighted sample points with measurement values. The interpolated value is defined using the following equation:

$$u(x) = \sum_{i=0}^{N} \frac{w_i(x) u_i}{\sum_{i=0}^{N} w_i(x)}, \quad (1)$$

where $u(x)$ is the unknown value u at a given point x. $w_i(x)$ and u_i are the weight of sample point i with respect to the given point x, and the value of sample point i, respectively. N is the number of sample points. The weight is determined for each sample point as a function of the distance between the interpolated point and the sample location, d, and power parameter, p, which is a positive real number. A greater value of p assigns a higher weight to the sample points that are closer to the interpolated point. Equation (2) shows that each sample point affects the interpolated value where the influence diminishes with distance.

$$w_i(x) = \frac{1}{d(x, x_i)^p}. \quad (2)$$

Figure 1 illustrates how IDW interpolation works. For example, the IDW takes three closest samples. A value at position x is determined based on values of sampling point 1, 2, and 3 with the distances to x, $d(x, x_i)$, are d_1x, d_2x, and d_3x. The respective weights are calculated using Equation (2), and the value at x can be found using Equation (1).

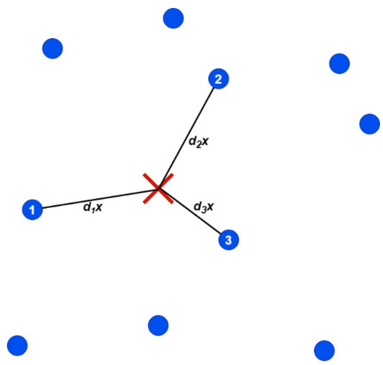

Figure 1. Illustration of IDW

The sample points used in interpolation can be specified based on variable and fixed search radius. The former uses the fixed number of samples required. The latter considers all the samples within the search radius. In this study, we considered variable search radius and evaluated the effect of the number of samples incorporated in the interpolation. The number of samples ranges between one to 10 samples.

1.2. ANNs

ANNs structure is comprised of a node layer, one or more hidden layers, and an output layer (as illustrated in Figure 2). The model determines the mathematical model between the input layer and the output layer as shown in Equation (3). Each node connects to another and has an associated weight that is defined during training. The output y is defined by the bias, b, and each input, x_i, and its respective weight, w_i.

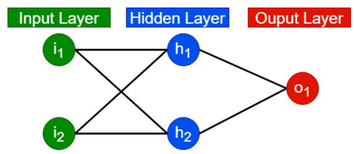

Figure 2. Illustration of ANNs

$$y = b + \sum_{i=1}^{n} x_i w_i. \quad (3)$$

Here, to be conservative, we used one hidden layer. We first selected the training function using a fixed number of inputs. Levenberg–Marquardt and Bayesian Regularization were evaluated as the training function. The training algorithm minimizes the linear combination of squared errors and weights. It also modifies the linear combination, so in the end, the network has good generalization. Then, a various number of inputs were tested ranging from one to 10 inputs. ANNs with one input means that it only considers the closest point to the interpolated location in the interpolation process.

The distinct characteristic of IDW and ANNs is that IDW considers the inverse of distance as the sample weight and ANNs assign weights to its inputs during the training process that the closer sample may not always obtain higher weights. IDW can be applied without training data, but ANNs model requires training data to build the model.

2. Site and data

In this study, we used DEM data of Balikpapan city, East Borneo, Indonesia, shown in Figure 3. The data was collected from Sentinel Hub EO Browser (Sentinel Hub, 2021). The source of DEM dataset is the Mapzen DEM that is based on the SRTM30 which is less accurate in elevation estimates compared to DEM W3D30 (Chymyrov, 2021). Figure 4 shows the histogram of the study area elevation. It can be seen that the elevation is between five to around 84 meters. As the focus of this paper is to evaluate the performance of interpolation models, we deem the data is the ground truth elevation data. The raster shown in Figure 3 covers the area of 20 km^2 and was converted to points. The distance between the points is 35 meters. The DEM resulted in around 17.000 points of which the 70% and 30% of the points were used as samples (known) and interpolated (unknown) points, respectively. The closest sample distance to the interpolated points varies from 35 to 105 meters. To build ANNs model, 70% of the unknown points were randomly selected from the data as training data; and the rest of the 30% data were used as testing data. The training and testing interpolated points were around 5000 and 1500 points, respectively. The IDW model was applied for the testing data.

3. Result

The first step in interpolating using ANN was to select the training function. Using the training data with five inputs and 10 hidden layers, the Bayesian Regularization

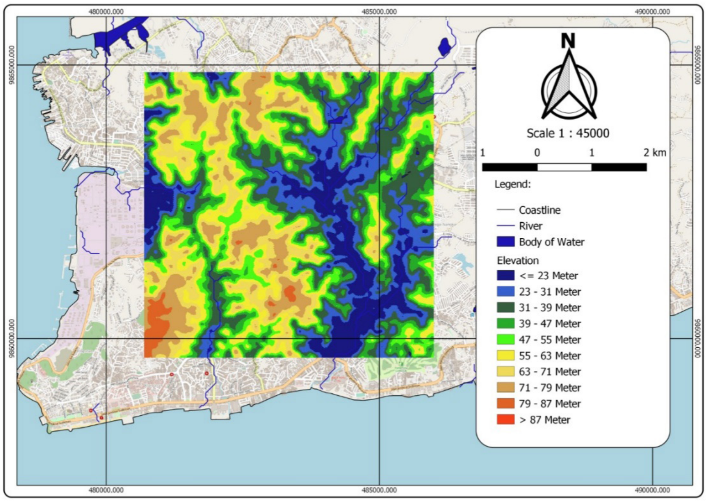

Figure 3. Study site: Balikpapan city, East Borneo

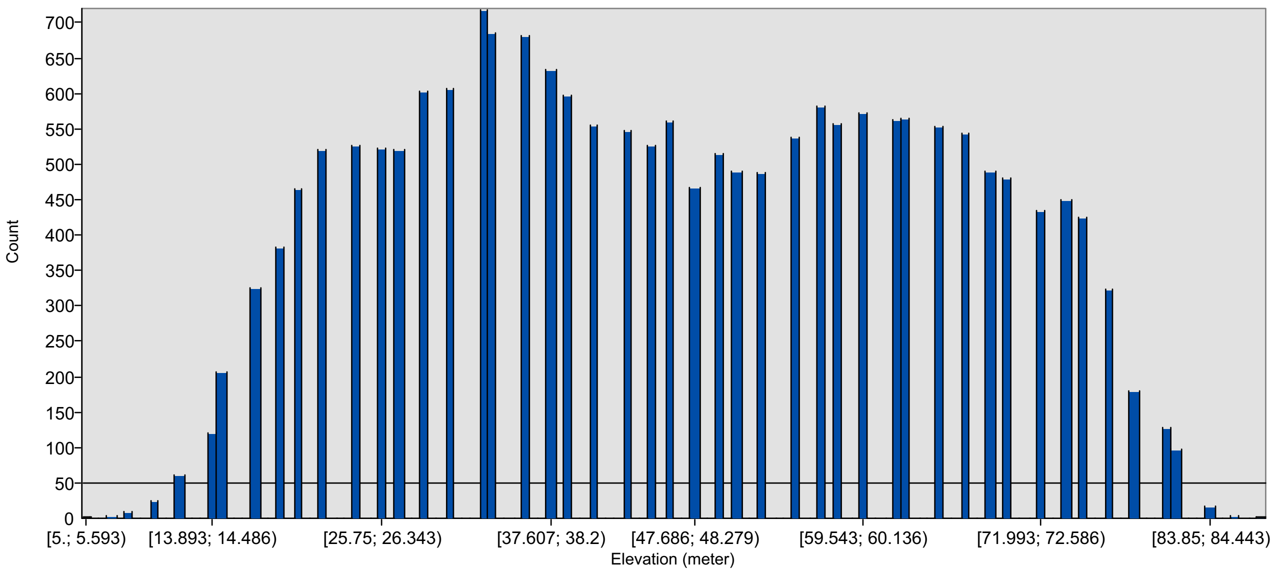

Figure 4. The histogram of the study area elevation

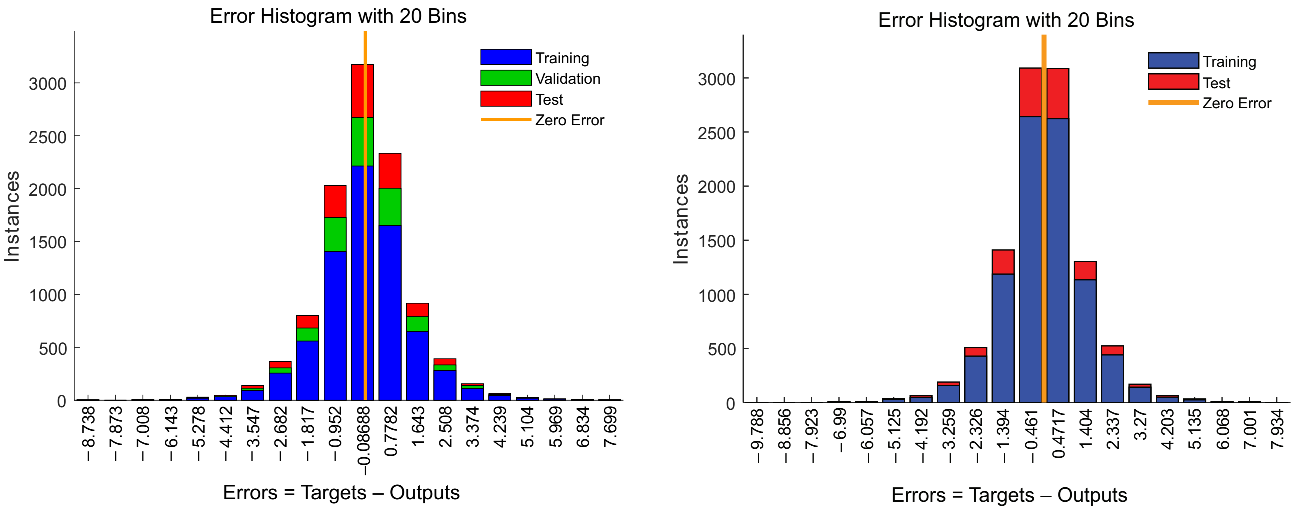

Figure 5. Training error with Levenberg–Marquardt (left) and Bayesian Regularization (right)

gave slightly better interpolation accuracy than the Levenberg–Marquardt training function. The training errors are presented in Figure 5. The models were tested on the testing data, and the mean absolute errors (MAE) are 1.2 m and 1.3 m for Bayesian Regularization and Levenberg–Marquardt training function, respectively. However, the training duration using Levenberg–Marquardt function was only around a quarter of training using Bayesian Regularization function. Therefore, we selected Levenberg–Marquardt as the training function.

Using the selected training function, we evaluated several number of hidden layers ranging from three to ten layers. Figure 6 shows the MAE of the various number of hidden layers. Between three to five hidden layers, as the number of hidden layers increases, the MAE decreases. However, after five hidden layers, the MAE increases significantly. Therefore, this study selected five hidden layers for the neural network's structure.

After both training function and number of hidden layers were selected, we evaluated a various number of inputs to the ANNs model starting from one to ten inputs. The input selection was based on the closest distance to the interpolated points. From Figure 7, it can be seen that the MAE decrease significantly when we added inputs up to three inputs. After that, the additional inputs keep the model performance stable. It can be argued that with training data, accurate interpolation of elevation can be obtained even the sample data is small.

Interpolation using IDW was set using variable search radius so that the effect of the number of neighbors on the interpolation accuracy can be evaluated. IDW has the parameter of p that needs to be defined. We first used the number of neighbors of five to evaluate the effect of p-value. Figure 9 shows the MAE for different p-value, ranging from one to three, with five neighbors. The MAE values for different p are similar at around two meters. We selected p equal to one as the value resulted in the lowest MAE. With the fixed p-value, different numbers of

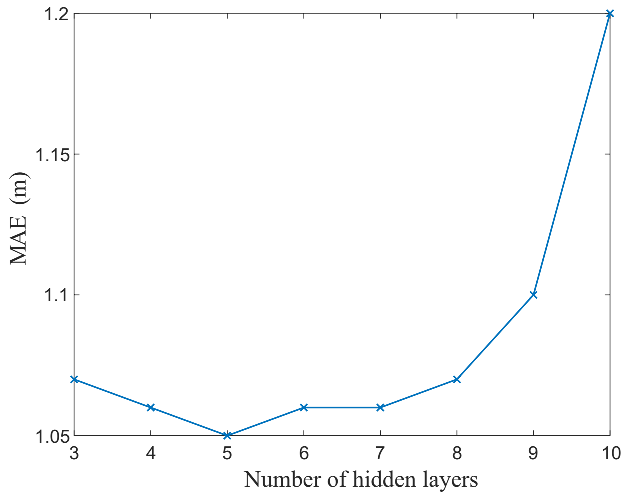

Figure 6. MAE of interpolation using ANNs for different numbers of hidden layers

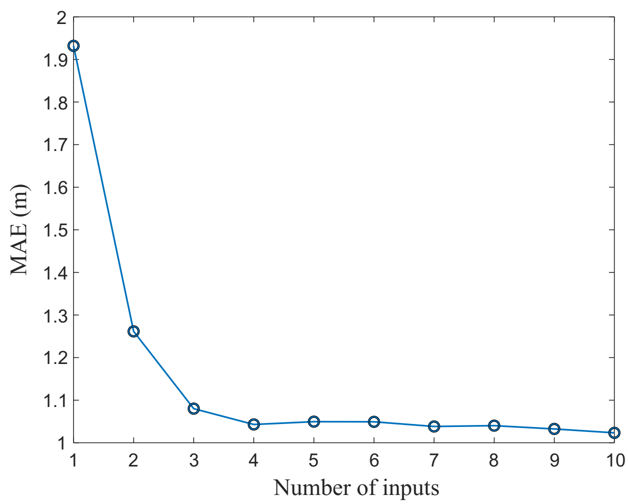

Figure 7. MAE of interpolation using ANNs for different numbers of inputs

neighbors ranging from one to ten were tested to interpolate the sample data points. The MAE is presented in Figure 8. It can be seen between one to three neighbors; the MAE decrease significantly when we added neighbors. However, starting from three to ten neighbors, the MAE becomes stable.

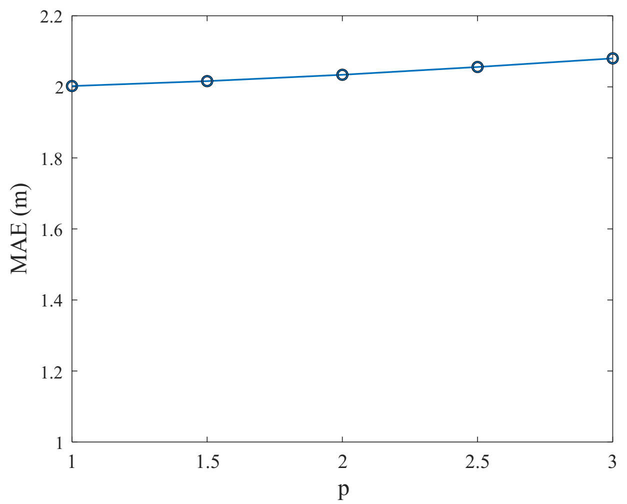

Figure 8. MAE of interpolation using IDW for different p with five number of neighbors

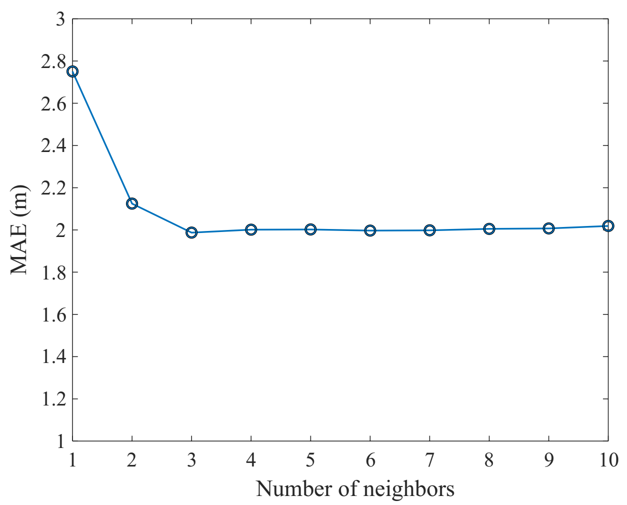

Figure 9. MAE of interpolation using IDW for different number of neighbors

In this study, both ANNs and IDW show the same trend with regards to the number of inputs considered in the interpolation. After three inputs, additional input would not increase the interpolation accuracy significantly. However, ANNs interpolation model gives better interpolation accuracy with MAE below 1.1 m for three or more inputs, compared to IDW with MAE around 2 m for three or more neighbors.

Conclusions

In this paper, the effects of the number of inputs in the elevation spatial interpolation using IDW and ANNs were evaluated. Using trial and error, the ANNs model structure that yielded the most accurate interpolation consists of five hidden layers. Also, Levenberg–Marquardt was chosen as the training function. The *p*-value in IDW was also evaluated and the result shows that *p* equal to one gave the best accuracy. Both models was tested to interpolate the elevation data using one to ten inputs. The result shows that both models give a similar trend in interpolation accuracy. From one to three inputs, additional input would give significant accuracy enhancement. However, after three inputs, additional input would give a stable performance of the interpolation. However, ANNs yield better accuracy which is under 1.1 meters of error for three or more inputs; compared to that of IDW which is around two meters of error for three or more inputs.

References

Ajvazi, B., & Czimber, K. (2019). A comparative analysis of different DEM interpolation methods in GIS: Case study of Rahovec, Kosovo. *Geodesy and Cartography*, *45*(1), 43–48. https://doi.org/10.3846/gac.2019.7921

Bartier, P. M., & Keller, C. P. (1996). Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). *Computers & Geosciences*, *22*(7), 795–799. https://doi.org/10.1016/0098-3004(96)00021-0

Chymyrov, A. (2021). Comparison of different DEMs for hydrological studies in the mountainous areas. *The Egyptian Journal of Remote Sensing and Space Science*, *24*(3), 587–594. https://doi.org/10.1016/j.ejrs.2021.08.001

Ikechukwu, M. N., Ebinne, E., Idorenyin, U., & Raphael, N. I. (2017). Accuracy assessment and comparative analysis of IDW, spline and kriging in spatial interpolation of landform (topography): An experimental study. *Journal of Geographic Information System*, *9*(3), 354–371. https://doi.org/10.4236/jgis.2017.93022

Jumaah, H. J., Ameen, M. H., Kalantar, B., Rizeei, H. M., & Jumaah, S. J. (2019). Air quality index prediction using IDW geostatistical technique and OLS-based GIS technique in Kuala Lumpur, Malaysia. *Geomatics, Natural Hazards and Risk*, *10*(1), 2185–2199. https://doi.org/10.1080/19475705.2019.1683084

Keskin, M., Dogru, A. O., Balcik, F. B., Goksel, C., Ulugtekin, N., & Sozen, S. (2015). Comparing spatial interpolation methods for mapping meteorological data in Turkey. In *Energy systems and management* (pp. 33–42). Springer. https://doi.org/10.1007/978-3-319-16024-5_3

Li, Z., Zhu, C., & Gold, C. (2004). *Digital terrain modeling: Principles and methodology*. CRC Press. https://doi.org/10.1201/9780203357132

Liu, F., He, X., & Zhou, L. (2009). Application of generalized regression neural network residual kriging for terrain surface interpolation. In *Proceedings SPIE: International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining* (Vol. 7492). https://doi.org/10.1117/12.837425

Merwin, D. A., Cromley, R. G., & Civco, D. L. (2002). Artificial neural networks as a method of spatial interpolation for digital elevation models. *Cartography and Geographic Information Science*, *29*(2), 99–110. https://doi.org/10.1559/152304002782053323

Noori, M. J., Hassan, H. H., & Mustafa, Y. T. (2014). Spatial estimation of rainfall distribution and its classification in Duhok governorate using GIS. *Journal of Water Resource and Protection*, *6*, 75–82. https://doi.org/10.4236/jwarp.2014.62012

Nusret, D., & Dug, S. (2012). Applying the inverse distance weighting and kriging methods of the spatial interpolation on the mapping the annual precipitation in Bosnia and Herzegovina. In *6th International Congress on Environmental Modelling and Software* (pp. 1–8). https://scholarsarchive.byu.edu/iemssconference/2012/Stream-B/229

Peterson, E. E., & Pearse, A. R. (2017). IDW-Plus: An Arc GIS toolset for calculating spatially explicit watershed attributes for survey sites. *JAWRA Journal of the American Water Resources Association*, *53*(5), 1241–1249. https://doi.org/10.1111/1752-1688.12558

Rigol, J. P., Jarvis, C. H., & Stuart, N. (2001). Artificial neural networks as a tool for spatial interpolation. *International Journal of Geographical Information Science*, *15*(4), 323–343. https://doi.org/10.1080/13658810110038951

Robinson, T., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture*, *50*(2), 97–108. https://doi.org/10.1016/j.compag.2005.07.003

Sentinel Hub. (2021). *EO Browser*. https://www.sentinel-hub.com/

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM National Conference* (pp. 517–524). https://doi.org/10.1145/800186.810616

Shiode, N., & Shiode, S. (2011). Street-level spatial interpolation using network-based IDW and ordinary kriging. *Transactions in GIS*, *15*(4), 457–477. https://doi.org/10.1111/j.1467-9671.2011.01278.x

Sivapragasam, C., Arun, V., & Giridhar, D. (2010). A simple approach for improving spatial interpolation of rainfall using ANN. *Meteorology and Atmospheric Physics*, *109*(1), 1–7. https://doi.org/10.1007/s00703-010-0090-z

Tomczak, M. (1998). Spatial interpolation and its uncertainty using automated anisotropic inverse distance weighting (IDW)-cross-validation/jackknife approach. *Journal of Geographic Information and Decision Analysis*, *2*(2), 18–30.