



Comparison of different interpolation methods for DEM production

Emanuele Alcaras¹, Claudio Parente², Andrea Vallario³

¹Department of Sciences and Technologies, University of Naples “Parthenope”, Naples, Italy.

Email: emanule.alcaras@gmail.com

²Department of Sciences and Technologies, University of Naples “Parthenope”, Naples, Italy.

Email: claudio.parente@uniparthenope.it

³Department of Sciences and Technologies, University of Naples “Parthenope”, Naples, Italy.

Email: andrea.vallario@uniparthenope.it

ABSTRACT

Spatial interpolation, or the estimation of the variables at unobserved locations in geographic space based on the values at observed locations, is fundamental in all geophysical sciences, first of all for the construction of digital elevation model (DEM). Several methods are available in literature for spatial interpolation and the choice of the most suitable of them for building DEM, depends on many factors, particularly on the distribution of the sampled points, therefore, on the morphology of the area to be mapped.

This paper aims to choose the most appropriate interpolators for DEM production, by comparing different methods usually available in GIS software. For the purpose of developing the best performing model and comparing interpolators, a set of elevation data collected by digital vector map is used. The accuracy of interpolation methods is tested by analyzing 4 statistic parameters, which are achieved by cross-validation leave-one-out. Particularly, minimum, maximum, mean and root mean square error (RMSE) are calculated for each interpolation method considering the residual in each sampling point between measured and interpolated value.

Key words : digital elevation model, interpolation methods, cross validation, RMSE, GIS.

1. INTRODUCTION

Digital Elevation Model (DEM) can be defined as the digital representation of the land surface elevation with respect to any reference datum; it is frequently used to refer to any digital model of a topographic surface [1]. This term is generally adopted as synonymous of both Digital Terrain Model (DTM) and Digital Surface Model (DSM). The first indicates a three – dimensional representation of a terrain surface consisting of X, Y, Z coordinates stored in digital form [2]. The second represents the earth's surface and includes all objects on it, such as buildings, towers, trees, and other natural and manmade features [3].

In this work we consider DEM as grid representation of a

terrain surface displayable also in 3D, in other terms we consider bare terrain z -values. According to the U.S. Geological Survey, a grid DEM is the digital cartographic representation of the elevation of the land at regularly spaced intervals in x and y directions, using z -values referenced to a common vertical datum [4].

DEMs play an important role in all geophysical sciences and not only: they are used for hydrological, geomorphological, and biological applications [5], and support many studies and researches, such as evaluation of solar farm locations [6]-[7] and wind energy potential [8], production of high resolution satellite images [9], construction of 3D thematic models [10], intervisibility analysis [11], etc.

To preserve the quality of the resulting DEM, two aspects must be treated very carefully: the technology to acquire sampling points and the choice of an interpolation method to produce the surface that properly fills in the initial points [4]. About the first aspect, different techniques can be used, i.e., photogrammetry, lidar, land survey, etc. [12]. About the second aspect, several methods have been developed and made available in GIS software. They are reviewed in many works present in literature, e.g. [13], and a few of them are usually applied to the same dataset and compared, e.g. [14].

This paper aims to applied a considerable number of interpolation methods to the same dataset and compare them in order to select the most suitable for a good DEM of the study area. Furthermore, we intend to extend considerations about the choice of the interpolators for more general situations, in view of the findings of our research. Firstly, a brief description is given for the study area concerning Ischia island (Italy), and the dataset of sampling points extracted from digital maps is introduced. Next, the methodological approach based on interpolation methods available in the software ArcGIS 10.3 (ESRI) [15] is illustrated: it consists of cross validation leave-one-out applied to each method, calculation of statistical parameters of the residuals in sampling points between measured and interpolated values, comparison of the results. Finally, conclusions are reported in order to remark the importance of the work and suggest applications and extensions of it.

2. STUDY AREA AND DATASET

2.1 Study area

Ischia is a volcanic island in the Tyrrhenian Sea. It lies at the northern end of the Gulf of Naples, some 30 km WSW from Napoli, 35 km west from Vesuvio and about 8 km from the mainland coast at Capo Miseno [16].

It presents the largest area among the isles located in the Gulf of Naples and its location is mapped in figure 1. It extends about 7 km north to south and 10 km east to west, and presents a surface area of 46.3 square kilometers, with approximately 34 km of coastline.

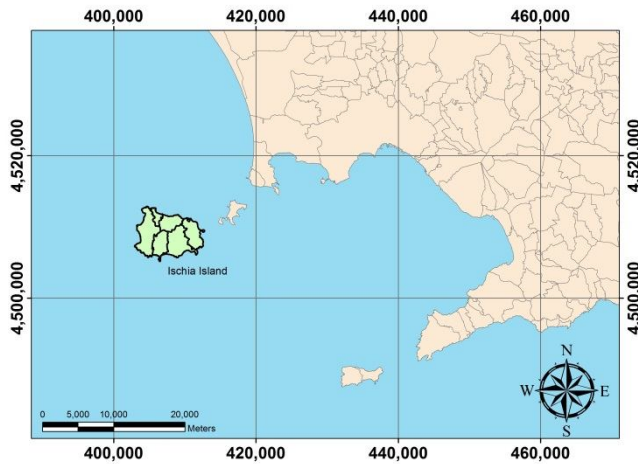


Figure 1: Ischia island in the Gulf of Naples

Ischia island represents the emerged portion of a wide volcanic field extending from the continental slope (facing westward on the Tyrrhenian batial plain) to the continental volcanic area of Campi Flegrei [17].

Its territory is mainly mountainous as it appears in the historical view in Figure 2 [18] and aerial photo in Figure 3; the highest peak is Mount Epomeo (788 meters), consisting of a Green Tuff ignimbrite deposit that was firstly submerged and then uplifted .



Figure 2: Historical view of Ischia Island published in 1776



Figure 3: Aerial photo of Ischia Island

As registered by ISTAT (the Italian Statistics Agency) on January 1, 2018, the island is very densely populated, including 60,000 residents (about 1,300 inhabitants per square km) [19]. The high value of population density represents a factor that contributes to increase the level of volcanic risk.

As can be seen also from the brief descriptions given above, Ischia Island presents historical, geological, volcanic, demographic characteristics that are so relevant and significant to require accurate DEM at large scale to support scientific studies and applications. For consequence, it represents a very interesting area to test different interpolation methods for DEM construction.

The part of Ischia island considered in this study is reported in Figure 4. It extends within the following UTM/WGS84 plane coordinates - 33T zone: $E_1 = 405,500$ m, $E_2 = 406,800$ m, $N_1 = 4,511,300$ m, $N_2 = 4,510,300$ m. Elevation values range between 29,090 m and 270 m.

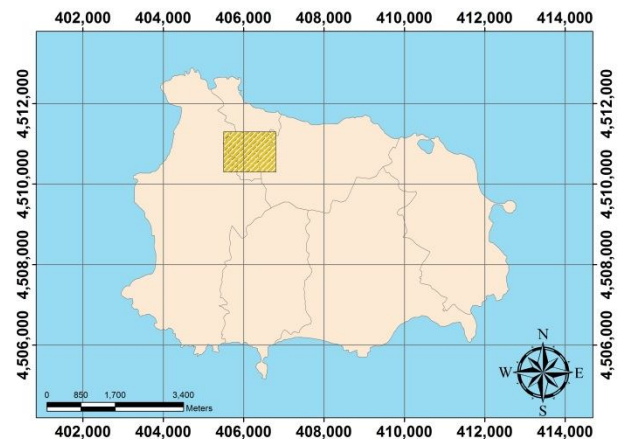


Figure 4: Map of Ischia island reporting the area considered in this study

2.2 Dataset

Elevation data are extracted from a topographic map in scale 1:5,000 produced by Campania Region and identified as CTR 430042. The original file is a vector file (shp) georeferred in UTM/WGS84 - 33T zone and contains all cartographic information (buildings, roads, etc.). Using a query in ArcGIS 10.3 (ESRI software), only spot heights and contour lines are extracted: vertices of contour lines and elevation points are grouped in one shape file used as dataset for the application of

12 interpolation methods available in Geostatistical Analyst, an extension included in ArcGIS software. In Figure 5, a particular of vector map showing vertices extracted from contour lines, is reported.

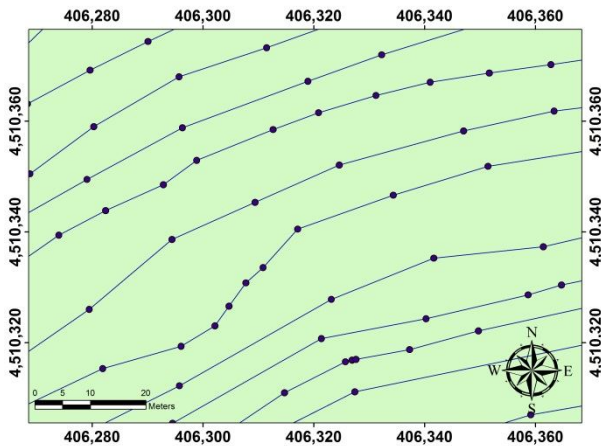


Figure 5: Particular of vector map showing vertices extracted from contour lines

In total, 4956 significant elevation points are used for interpolation processes.

3. METHODOLOGY

3.1 Interpolation methods

Interpolation permits to determine the value of a variable at any required location, based on the defined values at specified locations. Many interpolation methods are available in literature. In this study we consider several of them included in Geostatistical Analyst. Particularly, the following 12 methods are applied: Inverse Distance Weighting (IDW), Local Polynomial Functions (1st, 2nd, 3rd, 4th order), Radial Basis Functions (completely regularized spline, spline with tension thin-plate spline, multiquadric function, inverse multiquadric function), Ordinary kriging, Universal Kriging. However, also global polynomial functions are available in Geostatistical Analyst, but they are inappropriate for the considered area because its variable morphology that cannot be represented using one mathematic surface such as a plane (1st order), a surface with one bend (2nd order), a surface with two bends (3rd order), etc.

The adopted methods are largely discussed in literature, so the readers could refer to the numerous papers that describe their main characteristics, also in relation to their application for DEM construction, e.g. [4]-[13]-[20].

The application of each interpolation method needs to define some parameters. All methods considered in this study requires to define which points are involved in the estimation of elevation value in each prediction location.

Particularly, the user can assume that the correlation of the measured values with the prediction location decreases as locations get farther from the prediction location. For consequence, far points can be eliminated from the estimation of that particular prediction point by defining a search neighborhood. This purpose is achieved specifying the shape of the neighborhood. In our study, a circular isotropic model is used assuming that there are not defined directions and the

surrounding points have an equal influence on the central point that is being evaluated. Particularly, the radius equals 410 m for IDW, Local Polynomial Functions and Radial Basis Functions.

The choice of this distance defines the number of the neighbors. Additional parameters restrict the locations to be interpolated within the search neighborhood. The user can define the maximum and minimum number of neighbors to include and divide the neighborhood into sectors to ensure that values from all directions are included. Dividing the neighborhood into sectors, the specified maximum number and minimum number of neighbors are applied to each sector [21]. In our study, four sectors (without offset) are used, and maximum and minimum number of neighbors are set equal to 10 and 15 respectively.

Searching neighborhood step with the above mentioned parameters for IDW application is reported in figure 6.

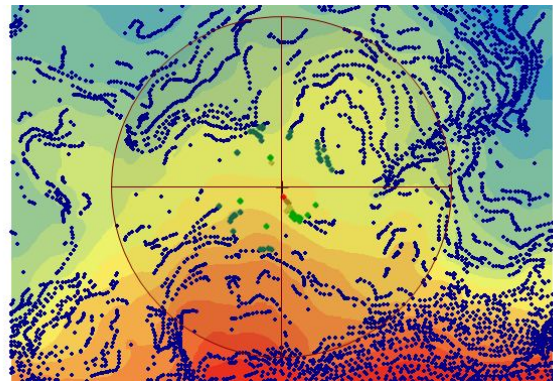


Figure 6: Searching neighborhood step for IDW application

For ordinary kriging and universal kriging, only minimum and maximum number of neighbors and sectors are defined (respectively 10, 15 and 4), while the radius is automatically derived by the adopted semi-variogram [22].

Some of the twelve models constructed in this research are shown in the next three figures in the order listed below. The model resulting from IDW application is shown in figure 7. The model resulting from Local Polynomial Function – 1st order application is shown in figure 8. The model resulting from Ordinary Kriging application is shown in figure 9.

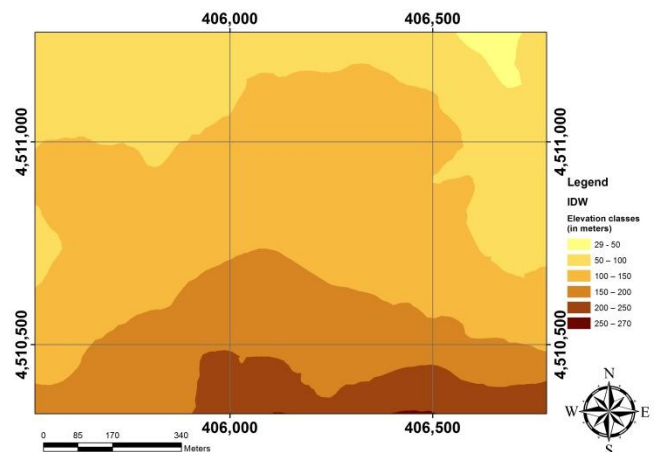


Figure 7: IDW model

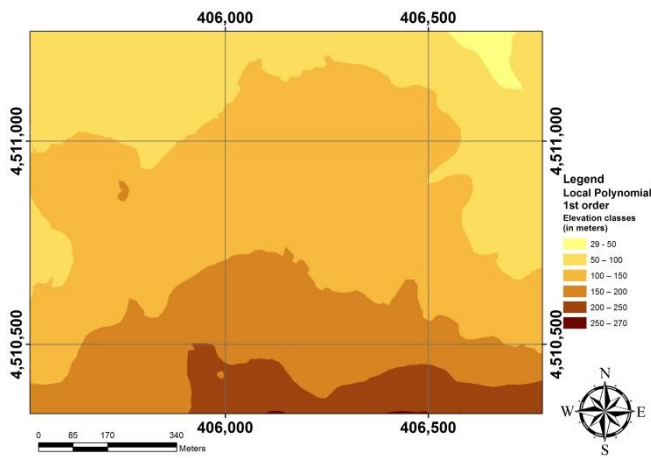


Figure 8: Local Polynomial Function – 1st order model

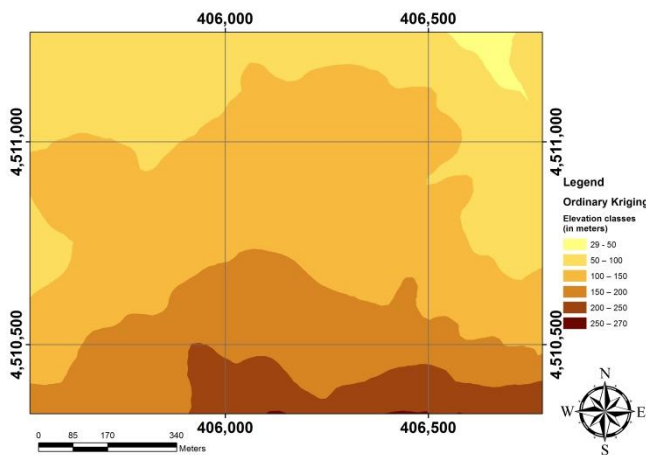


Figure 9: Ordinary Kriging model

3.2 Accuracy evaluation of each interpolation method

To compare different methods, the accuracy of each of them is to be evaluated. Cross validation is useful for this purpose: it permits to define the accuracy level of predictive values distinguishing between the training set and the validation set, the first used for model generation, the second for model evaluation.

Different approaches are usually adopted for cross validation. Leave-one-out method is based on the removal of a point from the data to be interpolated, the use of the other points to estimate a value at the location of the removed point, and the performance test by means of the removed data [23]. The difference between the known value and estimated value in each removed point, is calculated to evaluate the performance of the assumed interpolation method [24]. Statistical parameters of all residuals are calculated to facilitate the comparison of different approaches. Usually, root mean square error (RMSE) is the standalone parameter that better than another permits to evaluate the method performance. It is given by the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (h_i(x, y) - h'_i(x, y))^2}{N}} \quad (1)$$

Where:

N is the number of the elevation points;

$h_i(x, y)$ is the measured elevation at the location $i(x, y)$;

$h'_i(x, y)$ is the estimated elevation at the same location $i(x, y)$.

Leave-one-out method is applied in this study to each interpolation method.

4. RESULTS AND DISCUSSION

The residuals produced by cross validation applied to each interpolation methods are organized in form of tables that are easily managed using SQL [25]-[26]. The statistics parameters of those residuals calculated for each interpolation method are reported in Table 1.

The range of minimum values goes from -26.036 m obtained for Radial Basis Function - Multiquadric function, to -8.417 m resulting from Local Polynomial - 2nd order. The range of maximum values goes from 6.001 m obtained for Local Polynomial - 2nd order, to 14.086 m resulting from Radial Basis Function - Inverse multiquadric function. The range of mean values goes from -0.007 m obtained for Radial Basis Function - Multiquadric function, to 0.144 m resulting from Local Polynomial - 1st order. The range of RMSE goes from 2.226 m for IDW to 0.839 m resulting from Radial Basis Function - Spline with tension.

As resulting from RMSE value, Radial Basis function- Spline with tension and Radial Basis function - Thin-plate spline seem to be the most performing interpolation methods.

However, the range of residual values is too large for both. Ordinary Kriging and Universal Kriging are more suitable because they present small values of RMSE as well as shorter value ranges.

Usually, Radial Basis Functions tends to supply good performance in presence of uniform distribution of sampled points [27]. Because the morphology of the considered area and the cartographic origin of our elevation points, the uniform condition of point distribution is not respected everywhere, so the residuals in some locations are great.

For the same reason, also IDW performance is low.

Table 1: Parameters of the residuals resulting from cross validation application to each interpolation method

Method	min	max	mean	rmse
<i>All values are expressed in meters</i>				
Inverse Distance Weighting	-10.779	11.821	0.003	2.226
Local Polynomial 1 st order	-8.4862	8.520	0.144	1.714
Local Polynomial 2 nd order	-8.417	6.001	0.005	1.118
Local Polynomial 3 rd order	-8.677	13.464	0	1.052
Local Polynomial 4 th order	-14.259	13.121	0.008	1.021
RBF - Completely regularized spline	-15.575	11.087	-0.005	1.081
RBF - Spline with tension	-12.527	9.245	0.003	0.839
RBF - Thin-plate spline	-12.519	9.246	0.004	0.840
RBF - Multiquadric function	-26.036	10.863	-0.007	1.197
RBF - Inverse multiquadric function	-12.475	14.086	-0.005	1.086
Ordinary Kriging	-9.737	8.158	0.003	0.892
Universal Kriging	-9.735	8.158	0.003	0.893

Capability of kriging methods to adapt better than others to interpret and shape the variability of a territory is confirmed by our study.

5. CONCLUSION

Several interpolation methods are present in literature and most of them are available in GIS software and can easily be used for DEM construction. The choice of the most performing of them causes great difficulties because it depends on many factors, first of all the spatial distribution of sampled values and the morphology of the terrain to be modelled. For consequence, each situation is to be valued and the preferred approach is to estimate the accuracy of the model that the method is able to generate.

For the purpose, our research emphasizes the importance of the cross validation implemented by leave-one-out method. The calculation of the difference between the known value and estimated value in each removed point and statistical parameters of the residuals produced by each method in all locations of initial dataset, permit to compare different approaches. In other words, it supports the choice of the most performing method.

In our work, twelve methods are applied to 4956 elevation points contained in an area of 1.3 square kilometers. The experiments remark the differences between interpolation methods. On one site, they confirm the greater flexibility and adaptability of kriging approach to interpolate elevation points that are not equally distributed. On the other site, they remark the limit of some methods, i.e. IDW and Radial Basis Functions, for the same situation.

The results of the performed analysis and the conclusions that come out of them can be suitable for all use of interpolation methods, not only for DEM construction, but also for other applications.

Future developments of this work will be focused on the relationship between interpolation methods and terrain morphology as well as on the relationship between point density and model accuracy.

ACKNOWLEDGEMENT

We wish to thank the technical staff of our Department for their collaboration to this project. Special thanks should be given to Mr. Alberto Greco (Laboratory of Geomatics, Remote sensing and GIS), for his valuable technical support to this project.

CONTRIBUTIONS

C. P. conceived of the presented idea. C. P. and A. V. developed the theory and planned the experiments. E. A. selected and organized the dataset. E. A. and A. V. carried out experiments. C. P. supervised the experiments. C. P. and A. V. contributed to the interpretation of the results. All authors contributed to the writing of the article, provided critical feedback and helped shape the research, analysis and manuscript.

REFERENCES

1. A. Balasubramanian. **Digital elevation model (DEM) in GIS**, Technical report, University of Mysore, 2017. <https://doi.org/10.13140/rg.2.2.23976.47369>
2. C. L. Miller, and R. A. Laflamme. **The digital terrain model theory and application**, *Photogrammetric Engineering*, Vol. 24, no. 3, pp. 433-442, 1958.
3. Li Zhang. **Automatic digital surface model (DSM) generation from linear array images**, PhD thesis, ETH Zurich, 2005.
4. F. J. Aguilar, F. Agüera, M. A. Aguilar, and F. Carvajal. **Effects of terrain morphology, sampling density, and interpolation methods on grid DEM accuracy**, *Photogrammetric Engineering & Remote Sensing*, Vol. 71, no. 7, pp. 805-816, 2005.
5. I. D. Moore, R. B. Grayson, and A. R. Ladson. **Digital terrain modelling: a review of hydrological, geomorphological, and biological applications**, *Hydrological processes*, Vol. 5, no. 1, pp. 3-30, 1991.
6. M. Tahri, M. Hakdaoui, and M. Maanan. **The evaluation of solar farm locations applying Geographic Information System and Multi-Criteria Decision-Making methods: Case study in southern Morocco**, *Renewable and Sustainable Energy Reviews*, Vol. 51, pp. 1354-1362, 2015. <http://dx.doi.org/10.1016/j.rser.2015.07.054>
7. J. Brewer, D. P. Ames, D. Solan, R. Lee, and J. Carlisle. **Using GIS analytics and social preference data to evaluate utility-scale solar power site suitability**, *Renewable energy*, Vol. 81, pp. 825-836, 2015. <https://doi.org/10.23939/istecgcap2017.02.073>

8. B. Sliz-Szkliniarz, and J. Vogt. **GIS-based approach for the evaluation of wind energy potential: A case study for the Kujawsko–Pomorskie Voivodeship**, *Renewable and Sustainable Energy Reviews*, Vol. 15, no. 3, pp. 1696-1707, 2011.
<https://doi.org/10.1016/j.rser.2010.11.045>
9. O. Belfiore, and C. Parente. **Comparison of different algorithms to orthorectify WorldView-2 satellite imagery**, *Algorithms*, Vol. 9, no. 4, pp. 67, 2016.
<https://doi.org/10.3390/a9040067>
10. P. Maglione, C. Parente, R. Santamaria, and A. Vallario. **3D thematic models of land cover from DTM and high-resolution remote sensing images WorldView-2**, *Rendiconti Online della Società Geologica Italiana*, Vol. 30, pp. 33-40, 2014.
<https://doi.org/10.3301/ROL.2014.08>
11. A. Errico, F. Guastaferro, C. Oliviero, and R. Santamaria, **Intervisibility analysis of an offshore wind farm using GIS tools**. *WIT Transactions on Ecology and the Environment*, 149, pp. 87-95, 2011.
12. Z. Li, Q. Zhu, and C. Gold. **Digital terrain modeling: principles and methodology**, CRC Press, Boca Raton, FL, 2005.
13. N. S. N. Lam. **Spatial interpolation methods: a review**, *The American Cartographer*, Vol. 10, no. 2, pp. 129-150, 1983.
14. A. Šiljeg, S. Lozić, and S. Šiljeg. **The accuracy of deterministic models of interpolation in the process of generating a digital terrain model-the example of the Vrana Lake Nature Park**, *Tehnički vjesnik*, Vol. 22, no. 4, pp. 853-863, 2015.
<https://doi.org/10.17559/TV-20131010210942>
15. ESRI, **ArcGIS 10.3**, ESRI, Redlands, CA, USA (www.esri.com/software/arcgis)
16. Department of Geological and Mining Engineering and Sciences, **Ischia volcanic complex, Italy**, Michigan Technological University, 1996,
http://www.geo.mtu.edu/volcanoes/boris/mirror/mirrored_html/ISCHIA.html
17. A. Sbrana, P. Marianelli, and G. Pasquini. **Volcanology of Ischia (Italy)**. *Journal of Maps*, Vol 14, no. 2, pp. 494-503, 2018.
<https://doi.org/10.1080/17445647.2018.1498811>
18. W. Hamilton, and P. Fabris. **Campi Phlegraei. Observations on the Volcanos of the Two Sicilies**, Editor Mr. Peter Fabris, 1776.
19. ISTAT - Istituto Nazionale di Statistica, **Popolazioni e famiglie, Dati del censimento – Anno 2017**,
<https://www.istat.it/>
20. S. Erdogan. **A comparison of interpolation methods for producing digital elevation models at the field scale**, *Earth surface processes and landforms*, Vol. 34, no. 3, pp. 366-376, 2009.
21. ESRI, **Search neighborhoods**, in *ArcGIS 10.3- Help*, ESRI, Redlands, CA, USA
22. ESRI, **Understanding a semivariogram**, in *ArcGIS 10.3- Help*, ESRI, Redlands, CA, USA
23. U. Falchi, C. Parente, and G. Prezioso. **Global geoid adjustment on local area for GIS applications using GNSS permanent station coordinates**, *Geodesy and Cartography*, Vol. 44, no. 3, pp. 80-88, 2018.
<https://doi.org/10.3846/gac.2018.4356>
24. G. E. Fasshauer, and J. G. Zhang. **On choosing “optimal” shape parameters for RBF approximation**, *Numerical Algorithms*, Vol. 45 no. 1-4, pp. 345-368, 2007.
<https://dx.doi.org/10.1007/s11075-007-9072-8>
25. S. Chatvichienchai and Y. Kawasaki, **SpreadDB: Spreadsheet-Based User Interface for Querying and Updating**, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 7 no 2, pp. 6-10, 2018.
<https://doi.org/10.30534/ijatcse/2018/01722018>
26. B. Rama, K. S. Prasad and P. Sreeja, **Secure k-NN query on encrypted cloud data with multiple keys**, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 7 no 2, pp. 6-10, 2018.
<https://doi.org/10.30534/ijatcse/2019/82832019>
27. ESRI, **How radial basis functions work**, in *ArcGIS Pro help*, ESRI, Redlands, CA, USA,
<https://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/how-radial-basis-functions-work.htm>