15th International Congress of the Geological Society of Greece
Athens, 22-24 May, 2019 | Harokopio University of Athens, Greece
Bulletin of the Geological Society of Greece, Sp. Pub. **7** Ext. Abs. GSG2019-242

Leave-one-out cross validation method in hydrogeochemistry using GIS: A case study of Schinos village, NE Peloponnese, Greece

P. Papazotos[1], I. Polychroni[2], E. Vasileiou[1], M. Perraki[1]

(1) School of Mining and Metallurgical Engineering, Division of Geo-sciences, National Technical University of Athens, 9 Heroon Polytechniou St., 15773 Zografou, Greece, papazotos@metal.ntua.gr
(2) Department of Geology and Geoenvironment, Laboratory of Climatology and Atmospheric Environment, National and Kapodistrian University of Athens, Panepistimiopolis 15784, Zografou, Greece

Interpolation methods categorized either as deterministic or as geostatistical, are based on known values in order to estimate unknown values for any data. They are mostly used for generating spatial distribution maps. Deterministic interpolation methods are based on mathematical equations. The most common deterministic methods are inverse distance weighting (IDW) and radial basis function (RBF) (Varouchakis and Hristopoulos, 2013; Bhunia *et al.*, 2018). Geostatistical interpolation methods are based on statistics of the measured points by means of semivariogram. They quantify the distance among neighboring points. Some of the most important and widely used geostatistical methods in the scientific field of geology are kriging techniques (Krige, 1951; Matheron, 1963). The most frequently used sub-types of kriging are ordinary kriging (OK), simple kriging (SK), universal kriging (UK) and empirical Bayesian kriging (EBK) (Varouchakis and Hristopoulos, 2013; Bhunia *et al.*, 2018).

The aim of the present study is to propose the optimum interpolation method by comparing the main deterministic (IDW and RBF) and the geostatistical (OK, SK, UK and EBK) methods on the basis of the parameter of electrical conductivity (EC) in the coastal aquifer of Schinos village, NE Peloponnese, Greece. This study will contribute to reducing the uncertainties in the spatial distribution maps of the hydrogeochemical parameters.

The Schinos village is located in NE Peloponnese, Greece and belongs to the Corinth prefecture. The main aquifer system in the area is a granular unconfined alluvial coastal aquifer consisting of sands, gravels and conglomerates of ultramafic and carbonate origin; it is characterized by deterioration of quality due to the seawater intrusion and high concentrations of nitrate (NO_3^-) and hexavalent chromium (Cr^{6+}) (Papadopoulos and Lappas, 2014; Pyrgaki *et al.*, 2016).

A total of twenty (20) groundwater samples were collected from irrigation wells in the alluvial aquifer of Schinos village during the period of November to December 2016. The evaluation of the best interpolated method was achieved by performing the cross validation method, which uses the total number of samples and estimates the value of a known position by means of leave-one-out sample, so there is one measured and one estimated value. This procedure was repeated for the twenty (20) groundwater samples and for all the methods examined herein (IDW, RBF, OK, SK, UK and EBK). Geostatistical methods were performed with semivariogram using Gaussian-model-function, since it produced the most favorable outcome. The root-mean-square error (RMSE) and the mean absolute error (MAE) are two approaches to quantify the predictive performance of different interpolation methods and were calculated using ArcGIS 10.3. The equations for the estimation of RMSE and MAE are:

$$\mathrm{RMSE} = \sqrt{\frac{\sum (Z_i - Z)^2}{n}}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} | Z_i - Z |$$

Where Z_i is the predicted value, Z is the measured value, and n is the total number of observations.

EC constitutes one of the most representative parameter for groundwater quality, since it is a significant indicator of groundwater quality and presents strong to very strong correlation coefficients with the majority of major ions and the total dissolved solids (TDS) (R^2=0.90). Furthermore, elevated values of EC, which are directly associated with increased concentrations of Cl^-, usually reveal a seawater intrusion regime into the aquifer making explicit the qualitative degradation of the groundwater in the study area (Papadopoulos and Lappas, 2013). EC values in the Schinos village range from 972 to 3943 μS cm^{-1} with a mean of 2032.35 μS cm^{-1} and a median of 2017.5 μS cm^{-1}. 95% of the samples have EC > 1000 μS cm^{-1}. The EC has set a maximum guideline value of 2500 μS cm^{-1} in drinking water according to World Health Organization (WHO, 2011).

The physical parameter of EC was interpolated using deterministic (IDW and RBF) and geostatistical methods (OK, SK, UK and EBK). The IDW method was estimated for three (3) different weighting power (1,2,3), since these weights could affect significantly the results. Spatial distribution maps of EC present different spatial patterns for each interpolation method. According to RMSE, the most appropriate deterministic and geostatistical methods are in decreasing order IDW1>IDW2>IDW3>RBF, and UK>SK>OK>EBK, respectively. Lower value of RMSE indicates the better interpolation method. According to MAE, the most appropriate deterministic and geostatistical methods are in decreasing order IDW2>IDW3>IDW1>RBF and SK>UK>OK>EBK, respectively. Lower value of MAE suggests more precise results. Data treatment and cross validation method revealed, that the best interpolation methods for generating spatial distribution maps of EC are RBF and EBK for deterministic and geostatistical methods, respectively. The method with the lowest values of RMSE and MAE was the EBK suggesting that this is the most suitable interpolation method (Figure 1), as also stated by Hussain *et al.*, 2014.

The optimum interpolation methods for estimating the spatial variability of the values of EC in the coastal aquifer of Schinos village are geostatistical methods using GIS techniques. Selecting the most appropriate interpolation method is a complex process due to the variability of the examined parameters and the high number of manually-adjusted parameters, which are dependent on the user's experience. The most crucial factors are: the magnitude and the spatial distribution of the dataset (dependence of the distance), the interpolated qualitative or quantitative parameter since some methods are more suitable for datasets with extreme values (e.g. NO_3^-, Cr^{6+}), while others are proper for datasets without outliers (e.g. EC, TDS and the majority of major ions).

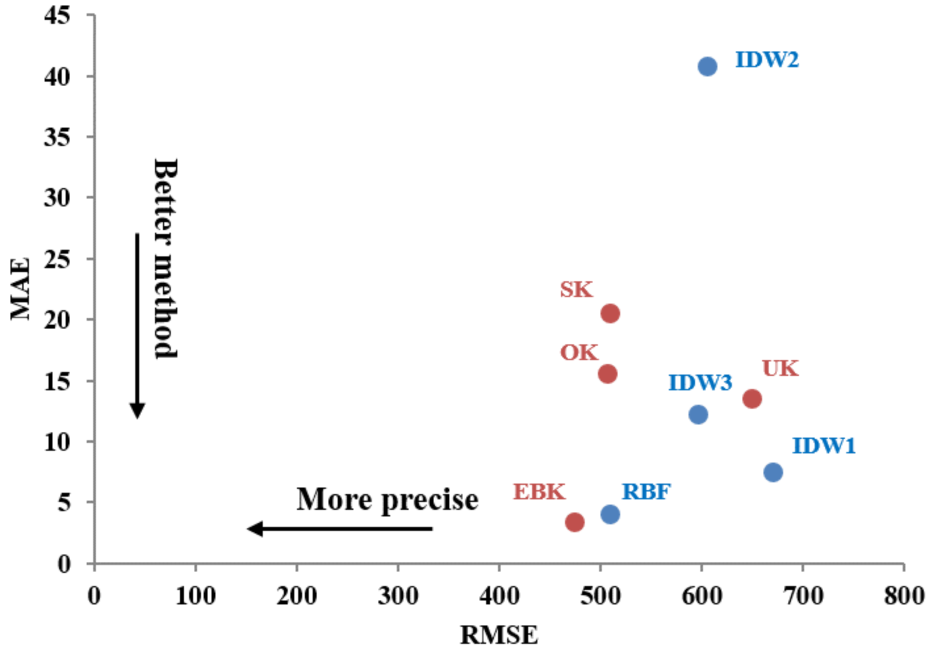

Figure 1. RMSE and MAE for deterministic (blue circles) and geostatistical (red circles) interpolation methods for the groundwater samples in the Schinos village.

References

Bhunia, G.S., Shit, P.K., Maiti, R., 2018. Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon(SOC). Journal of the Saudi Society of Agricultural Sciences, 17 (2), 114–126.

Hussain, I., Shakeel, M., Faisal, M., Soomro, Z.A., Hussain, T., 2014. Distribution of total dissolved solids in drinking water by means of Bayesian kriging and gaussian spatial predictive process water quality. Exposure Health, 6 (4), 177–185.

Krige, D., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa, 52, 119–139.

Matheron, G., 1963. Principles of geostatistics. Economic Geology, 58, 1246–1266.

Papadopoulos, K., Lappas, I., 2014. Groundwater quality degradation due to Cr6+ presence in Schinos area, prefecture of Corinth, Central Greece. 10th International Hydrogeological Congress of Greece, Thessaloniki.

Pyrgaki, K., Argyraki, A., Kelepertzis, E., Paraskevopoulou, V., Botsou, F., Dassenakis, E., Mitsis, I., Skourtsos, E., 2016. Occurrence of hexavalent chromium in the ophiolite related aquifers of Loutraki and Schinos Areas. Bulletin of the Geological Society of Greece, 50 (4), 2261-2270.

Varouchakis, E.A., Hristopulos, D.T., 2013. Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. Environmental Monitoring Assessment, 185, 1–19.

World Health Organization (WHO), 2011. Guidelines for drinking water quality, World Health Organization Geneva, 4th Edition.