VOL. 19, NO. 9, MAY 2024 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

PREDICTION OF GRAIN SIZE DISTRIBUTION USING ORDINARY KRIGING AND COMPOSITIONAL KRIGING METHODS

Ahram Kim, Bashir Busahmin and Stephen Tyson
Department of Petroleum and Chemical Engineering, Universiti Teknologi Brunei, Brunei Darussalam
E-Mail: ahram.kim0805@gmail.com

ABSTRACT

Grain size analysis plays a crucial role in understanding the geological characteristics of the coastal environments that influence the optimizations for oil and gas production operations. This paper aims to explore a sophisticated geostatistical approach using the ordinary kriging and compositional kriging techniques, to forecast the grain size fluctuations of sediments in the Long Island region located in the United States. In addition, utilizing a comprehensive dataset collected from the same region about an integrated seventeen compositional components for investigation using the spatial model of the grain size distribution. Moreover, a variogram and the scatter plot predicted a distinctive spatial dependency was achieved. The compositional kriging method used to predict the grain size distribution in the coastal areas presented an accurate result based on the shape of the histogram, Root Mean Square Error (RMSE), and the Mean Squared Error (MSE). In conclusion, the geo-statistics assisted in the integration of the sedimentological analysis in the coastal settings and showed an effective configuration for the decision-making in the oil and gas industry business.

Keywords: spatial Interpolation, compositional kriging, continuous data, grain size, sediment data.

Manuscript Received 15 March 2024; Revised 3 June 2024; Published 7 July 2024

1. INTRODUCTION

The vitality of sedimentary settings is vital in petroleum systems that involve emphasizing the essential part of the sedimentary deposit, where geological properties carried in the development, movement, and accumulation of the hydrocarbons where [1] stated that the sedimentary settings are important in petroleum systems because it starts with the origin of source rocks, the development of reservoir rocks, and the regulation of migratory patterns. In addition, the geological properties of sediments, as well as the processes that shaped them, are critical for forecasting the presence quality, and the distribution of the hydrocarbons in the underground oil reservoirs [2]. Sedimentary rocks are so significant in defining the reservoir properties that include porosity and permeability, however, the reservoir performance is to store and transport fluids influenced by the dimensions, arrangement of sediments, compaction, and grain cementations, [3] in addition, lithology and the effect of diagenesis are essential for the success of petroleum exploration and production [4]. Moreover, by identifying the key studies that successfully links the surficial sediment data to reservoir quality is instrumental in advancing the understanding of the subsurface dynamics governing hydrocarbon reservoir behaviour [5], which is focused on integrating the high-resolution surficial sediment data with the reservoir characteristics in a fluvial-dominated deltaic system, where many researchers such as [6] meticulously correlated the grain size variations in the surficial sediments using the petrophysical properties of the reservoir rocks, and therefore revealing a nuanced relationship between the surface and the subsurface of the sedimentary processes and the study exemplifies how a comprehensive analysis of the surficial sediment data can enhance the predictions of reservoir quality by capturing the dynamic interactions between the depositional environment and the subsurface rock properties. Furthermore, the research stood out for the application of the advanced sediment logical and geochemical analyses to link the surficial sediment properties with the reservoir quality in a coastal depositional setting, moreover, the authors went beyond the traditional grain size analyses and examined the influence of the sediment composition, organic content, and the mineralogy on the reservoir quality, whereas the researchers successfully established robust correlations between the surficial sediment attributed and the reservoir petrophysical parameters, showcasing the potential of multidisciplinary approaches in refining the predictions of the subsurface reservoir behaviour based on surface sediment data and the key studies collectively highlighted the significance of integrating the surficial sediment information into the reservoir characterization efforts for more comprehensive understanding of the hydrocarbon reservoir quality. Nevertheless, examining sediment samples along the entire coastline is impracticable, resulting in a paucity of information about the distribution of the coastal sediments therefore, the grain size of the sediments predicted at all unsampled sites as a result of geostatistical approaches for delineating the spatial patterns are also advisable for mapping and forecasting of the sediment characteristics, as well as contributing to the creation of maps depicting the coastal and seabed sediment features [7]. [8] explained the geostatistical interpolation methods such as the Ordinary Kriging (OK) are useful for anticipating the spatial change of the geological data, however, the approach assumes the spatial stationarity, which means that the spatial variability of the sediment grain size is assumed to be constant throughout the region. Additionally, in most papers, the choice of the variogram models affected the Ordinary Kriging (OK), and choosing the incorrect values may lead to a biased

ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

forecast. The main purpose of this research paper is to compare the predicted performance of the two spatial interpolation algorithms for the coastal sediment distribution by seventeen groups of grain size, as this paper studied the comparison between ordinary kriging and compositional kriging.

2. METHODOLOGY

2.1 Study Area

A total of 931 samples of coastal sediments were collected along the coastline, which were taken from the top 0 to 55 m at each sample position on the Long Island surface located in the United States. Figure-1 is created by Python programming to detect the location of data. Table-1 presents the sediment data used for the classification. PHIM5 which is the weight percentage of the sample in 5 phi fractions (nominal diameter of the particles that are greater than or equal to 32 mm, but less than 64 mm); very coarse pebbles to PHI11 which is the weight percentage of the sample in 11 phi fractions (nominal diameter of that particles which are greater than or equal to 0.5 mm, but less than 0.001 mm particle size). In addition, Table-1 presents the Group which is numbered in order of the size to assist the investigation and explains the data.

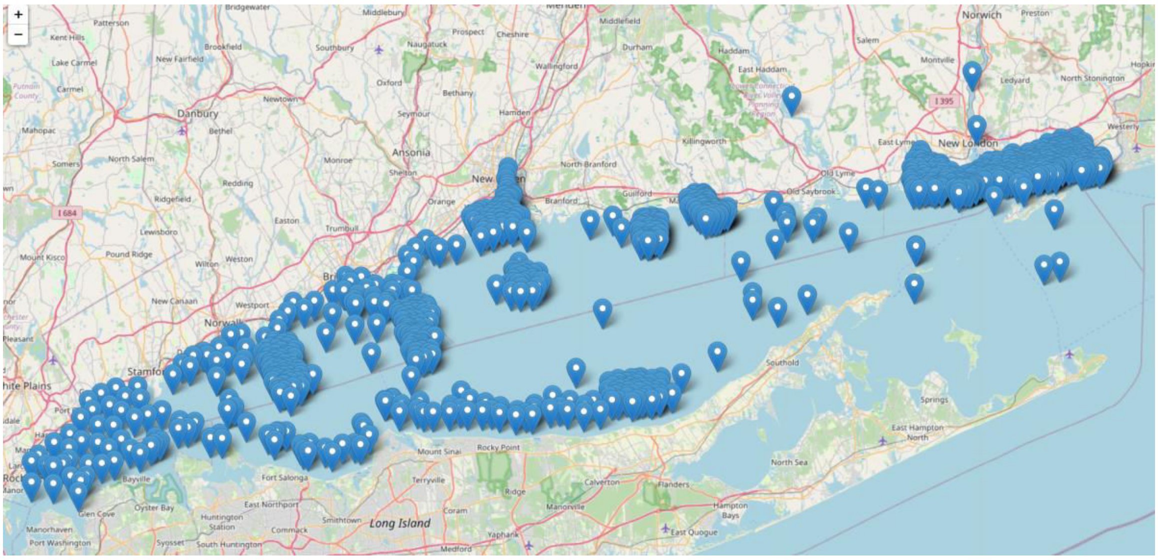

Figure-1. Plotting long island in US sediment data on the map.

ARPN Journal of Engineering and Applied Sciences

Table-1. Group number, attribute label, and grain size of long island sediment data in the US**.**

Group	Attribute Label	Size
1	PHIM5	Nominal diameter greater than or equal to 32 mm, but less than 64 mm
2	PHIM4	Nominal diameter of particles greater than or equal to 16 mm, but less than 32 mm
3	PHIM3	Nominal diameter of particles greater than or equal to 8 mm, but less than 16 mm
4	PHIM2	Nominal diameter of particles greater than or equal to 4 5mm, but less than 8 mm
5	PHIM1	Nominal diameter of particles greater than or equal to 2 mm, but less than 4 mm
6	PHI0	Nominal diameter of particles greater than or equal to 1 mm, but less than 2 mm
7	PHI1	Nominal diameter of particles greater than or equal to 0.5 mm, but less than 1 mm
8	PHI2	Nominal diameter of particles greater than or equal to 0.25 mm, but less than 0.5 mm
9	PHI3	Nominal diameter of particles greater than or equal to 0.125 mm, but less than 0.25 mm
10	PHI4	Nominal diameter of particles greater than or equal to 0.0625 mm, but less than 0.125 mm
11	PHI5	Nominal diameter of particles greater than or equal to 0.031mm, but less than 0.0625 mm
12	PHI6	Nominal diameter of particles greater than or equal to 0.016mm, but less than 0.0625 mm
13	PHI7	Nominal diameter of particles greater than or equal to 0.008mm, but less than 0.016 mm
14	PHI8	Nominal diameter of particles greater than or equal to 0.004mm, but less than 0.008 mm
15	PHI9	Nominal diameter of particles greater than or equal to 0.002 mm, but less than 0.004 mm
16	PHI10	Nominal diameter of particles greater than or equal to 0.001 mm, but less than 0.002 mm
17	PHI11	Nominal diameter of particles greater than or equal to 0.5 mm, but less than 0.001 mm

2.2 Spatial Interpolation Analysis Using Kriging Method

Two geostatistical techniques which are the Ordinary kriging (OK) and the Compositional kriging (CK) were implemented to interpolate the grain size and compare it with the geostatistical analyst tool using Python programming. Deterministic approaches, such as Inverse Distance Weighting (IDW) typically measure the distance

VOL. 19, NO. 9, MAY 2024 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

between the points. However, the challenges of the interpolation method are sensitive to the outliers [9]. [10] discussed the statistical characteristics of the data and the semi-variogram which serves as the foundation for the geostatistical models. [11] Emphasized the interpolation method, by considering the statistical properties of the measured data using the Ordinary Kriging (OK) technique as shown in equations (1,2) which is straightforward of Kriging methods.

$$Z_{OK} = \sum_{i=1}^{N} \lambda_i Z_i \tag{1}$$

$$\sum_{i=1}^{N} \lambda_i = 1 \tag{2}$$

$Where$: Z_{OK} (x_0) is the interpolated value for point x_0, $Z\ (x_i)$ $is\ the\ known\ value$, λ_i is the Ordinary kriging (OK) weightage for the $Z\ (x_i)$ value, moreover, [12] stated that statistical approach for estimating the values at the unsampled places uses the spatial correlation information, as well as examined the required Ordinary Kriging (OK) method and the variogram, which measures the degree of the spatial dependency between the two pairs of the locations and characterized the geographical connection between the sample points. therefore, the variogram is important to pose the details on how the variability data varies along the distance and the direction. However, when the data collection contains a substantial outlier and exhibits non-stationary behaviour, it may produce skewed results, and therefore inaccurate spatial interpolations might occur [13]. Compositional kriging is a geostatistical technique that combines the examination of the spatial distribution of several compositional components to provide a complete framework for forecasting and charting the variability of a compositional parameter over a given area. Compositional kriging provides a significant advantage in estimating the surficial sediment particle size because it takes into account the interaction of many compositional components.

3. FINDINGS

Although Kriging may produce a decent interpolation result in non-normal data, however, the best result is obtained when the data is normal or close to normal. As a result, The Long Island sample data set is log-transformed to generate the distributions that are as close to normal as possible. After the log transformation, the sampling data set is normally distributed as shown in Figure-2.

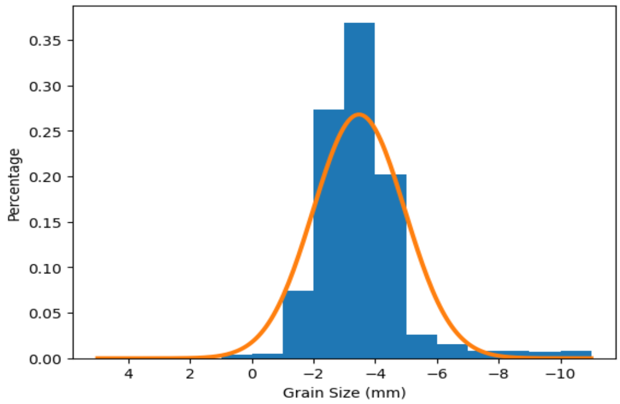

Figure-2. Normal distribution of log transformation for long island in US sediment data.

3.1 Dependency

Figure-3 presents the scatter plot between the mean and the variance of the Long Island in United States of America sediment data, however, are not strongly correlated which is aligned with [14] and revealed using the non-gaussian symmetrical distributions, mean and the variance parameters but it cannot be quantified to be an independent random estimator.

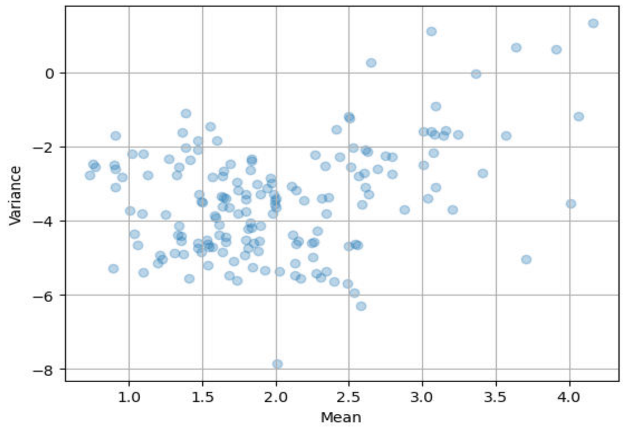

Figure-3. Scatter plot between mean and variance of long island in US sediment data.

Figure-4 presents the scatter plot between PHI_9 and PHI_10 which is strongly correlated with the correlation coefficient of (r = 0.962). Similarly, Figure 5 shows a strong correlation between PHI_7 and PHI_8 with a correlation coefficient of (r = 0.960), Furthermore, the mean and the variance are not that much related, however, each variable may be correlated. Since they are dependent it is possible to examine and predict the semi variogram.

 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

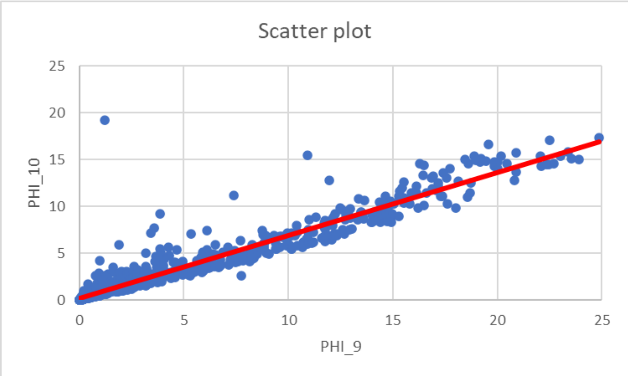

Figure-4. Scatter plot between PHI_9 and PHI_10.

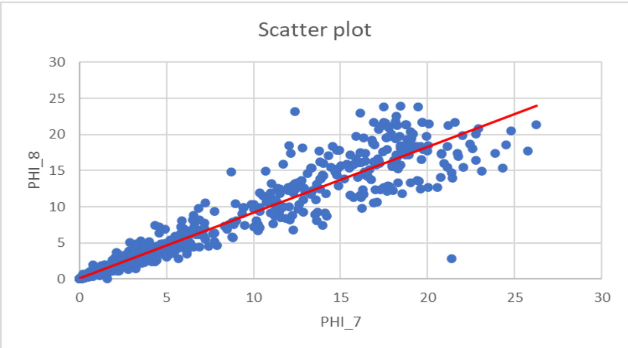

Figure-5. Scatter plot between PHI_7 and PHI_8.

3.2 The Semi-Variogram

Depending on the data inputs for each geostatistical interpolation method, the semi-variogram parameters nugget, sill, and the range were found as shown in Figure-6, the variogram has a range of 0.025 km (25m) where the data points behaved as an independent beyond the range distance and more likely the data are comparable when are the closer to each other. As a result, it shows 1.6 sill and a little nugget that demonstrates a small amount of randomness in the sites which are remarkably close to each other.

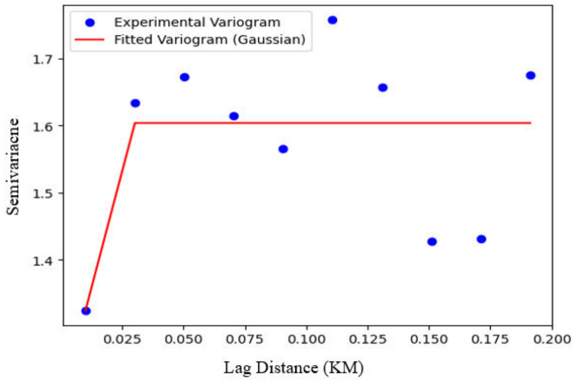

Figure-6. A semi variogram for long island in US sediment data.

3.3 A Comparison between the Distribution Histogram of Two Kriging Methods

After investigating the spatial dependency, the distribution of the data is examined the shape of the distribution between the original data and the kriging data for the ordinary kriging and the compositional kriging. For instance, data number 22 showed the original data distribution using the Compositional Kriging data distribution. It is noticeable that the highest value of original data is Group 8 but the interpolation value of Ordinary Kriging is Group 10, as shown in Figures 7 and 8, the Group 8 the interpolation value in compositional kriging and the original data showed a similar shape of the distribution.

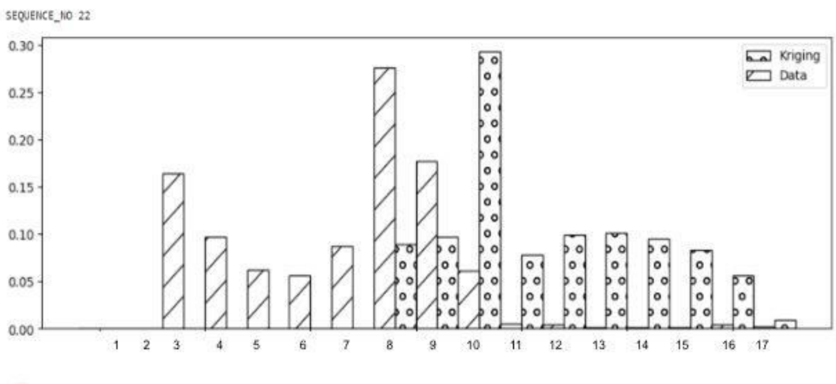

Figure-7. The distribution for ordinary kriging and original data.

VOL. 19, NO. 9, MAY 2024 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

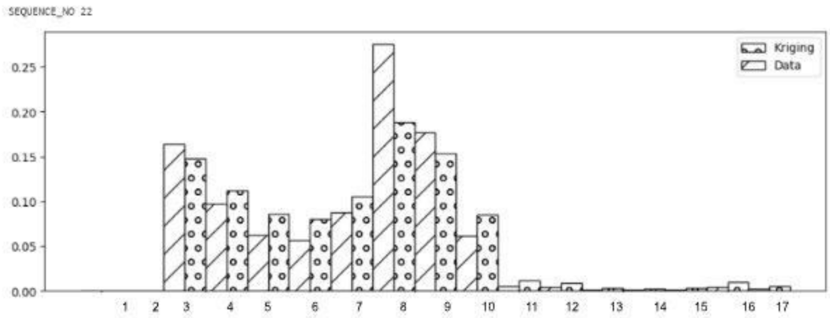

Figure-8. The distribution for compositional kriging and original data.

3.4 Cross Validation

Cross-validation is a crucial strategy when it comes to evaluating the statistics models. This method of statistics separates the dataset into subsets systematically, allowing for a thorough evaluation of the model's efficacy and adaptability. In this section, the Mean Squared Error (MSE) and the Root Mean Square Error (RMSE) are considered to quantitatively assess the predicted values as depicted in Table-2, this amount of 0.000176 is less than 0.012924, which means that in this situation, the associated model has a lower average squared error. Since a lower Mean Squared Error (MSE) indicates a greater model performance in terms of accuracy, therefore, the Compositional Kriging is a better perception to predict the grain size in the coastal sediment data. Similarly, the Root Mean Square Error (RMSE) calculated the average size of the errors between the expected and the actual values by taking the square root of the mean squared difference and that is 0.079414 is less than 0.093389, where it indicates that the associated model has less average magnitude errors. Thus, 0.079414 indicates a lower average magnitude of the error than 0.093389, suggesting a potentially improved model performance in terms of the Root Mean Square Error (RMSE).

Table-2. Cross-validation results of compositional kriging and ordinary kriging.

	Compositional Kriging	Ordinary Kriging
Mean Squared Error (MSE)	0.000176	0.012924
Root Mean Square Error (RMSE)	0.079414	0.093389

4. CONCLUSIONS

This study investigates the performance of several spatial interpolation algorithms for predicting the sediment distribution histogram. Two geostatistical methods named Compositional Kriging (CK) and Ordinary Kriging (OK) were used in the coastal sediments of Long Island, United States, for interpolation of the grain size and to select the most suitable method for the specific location. It is concluded that the data is dependent spatially as shown by the variogram and the scatter plots; it is required to compare the two interpolation methods using the shape of the distribution by using the histogram along the original data. Compositional Kriging showed a similar distribution in the shape with the original data. Similarly, the Root Mean Square Error (RMSE) and the mean Square Error (MSE) were calculated, and the Compositional Kriging presented more accuracy than the ordinary kriging. In addition, Compositional Kriging enabled a more accurate depiction of the spatial distribution of sediment features, which means a better prediction that takes into consideration the intricate interactions between the diverse sediment constituents.

REFERENCES

[1] Z. Yang, J. Chen, S. Wang, X. Li, W. Cheng, and S. Chen. 2022. Indication of formation water geochemistry for hydrocarbon preservation: New applications of machine learning in tight sandstone gas reservoirs. ACS Omega, vol. 7, no. 41, pp. 36263–36276, 2022..

[2] Y. Huang, I. A. Kane, and Y. Zhao. 2020. Effects of sedimentary processes and diagenesis on reservoir quality of submarine lobes of the Huangliu Formation in the Yinggehai Basin, China. Mar. Pet. Geol., vol. 120, no. 104526, p. 104526, 2020.

[3] Q. U. Z. Dar, R. Pu, C. Baiyegunhi, G. Shabeer, R. I. Ali, and U. Ashraf and Mehmood M. 2022. The impact of diagenesis on the reservoir quality of the early Cretaceous Lower Goru sandstones in the Lower Indus Basin. Pakistan. Journal of Petroleum Exploration and Production Technology, pp. 1–16.

[4] Nichols T. E., Worden R. H., Houghton J. E., Duller R. A., Griffiths J. and Utley J. E. P. 2023. Sediment texture and geochemistry as predictors of the sub-depositional environment in a modern estuary using machine learning: A framework for investigating clay-coated sand grains. Sedimentary Geology. *458*, 106530.

[5] T. E. Nichols, R. H. Worden, J. E. Houghton, R. A. Duller, J. Griffiths, and J. E. P. Utley. 2023. Sediment texture and geochemistry as predictors of sub-depositional environment in a modern estuary using machine learning: A framework for investigating clay-coated sand grains. Sediment. Geol., vol. 458, no. 106530, p. 106530, 2023.

[6] O. E. Aro, S. J. Jones, N. S. Meadows, J. Gluyas, and D. Charlaftis. 2023. The importance of facies, grain size, and clay content in controlling fluvial reservoir quality example from the Triassic Skagerrak Formation, Central North Sea, UK. Central North Sea, UK. Petroleum Geoscience, vol. 29, no. 2, 2023.

[7] R. M. Lark, D. Dove, S. L. Green, A. E. Richardson, H. Stewart, and A. Stevenson. 2012. Spatial prediction of seabed sediment texture classes by cokriging from a legacy database of point observations. Sediment. Geol., vol. 281, pp. 35–49, 2012.

[8] V. Boumpoulis, M. Michalopoulou, and N. Depountis. 2023. Comparison between different spatial interpolation methods for the development of sediment distribution maps in coastal areas. Earth Science Informatics, pp. 1–19, 2023.

[9] U. Bronowicka-Mielniczuk, J. Mielniczuk, R. Obroślak, and W. Przystupa. 2019. A comparison of some interpolation techniques for determining spatial distribution of nitrogen compounds in groundwater. Int. J. Environ. Res., vol. 13, no. 4, pp. 679–687, 2019.

[10] J. Aalto, P. Pirinen, J. Heikkinen, and A. Venäläinen. 2013. Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models. Theor. Appl. Climatol., vol. 112, no. 1–2, pp. 99–111, 2013.

[11] S. K. Al-Mamoori, L. A. Al-Maliki, A. H. Al-Sulttani, K. El-Tawil, and N. Al-Ansari. 2021. Statistical analysis of the best GIS interpolation method for bearing capacity estimation in An-Najaf City, Iraq. Environ. Earth Sci., vol. 80, no. 20, 2021.

[12] G. Pellicone, T. Caloiero, G. Modica, and I. Guagliardi. 2018. Application of several spatial interpolation techniques to monthly rainfall data in the Calabria region (southern Italy). Int. J. Climatol., vol. 38, no. 9, pp. 3651–3666, 2018.

[13] S. J. Hadi and M. Tombul. 2018. Comparison of spatial interpolation methods of precipitation and temperature using multiple integration periods. J. Ind. Soc. Remote Sens., vol. 46, no. 7, pp. 1187–1199, 2018.

[14] A. Vlad and B. Badea. 2008. Analysing the dependence between variance and mean estimated values: A theoretical and experimental approach. In Computational Science and Its Applications – ICCSA 2008, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 752–767.