Indonesian Geotechnical Journal, April 2022,, Vol 1 (1). Page 15-24
DOI: https://doi.org/10.56144/igj.v1i1.5
E-ISSN 2829-7520
https://indonesian-geotechnical-journal.org/index.php/JGI/issue/archive

Interpolation Method for Spatial Distribution of Clay Content within Residual Soil

Alfrendo Satyanaga[1,*], Aswin Lim[2], Nurly Gofar[3]

[1]Department of Civil and Environmental Engineering, Nazarbayev University, Nur-Sultan, Kazakhstan, 010000; alfrendo.satyanaga@nu.edu.kz

[2]Department of Civil Engineering, Engineering Faculty, Parahyangan Catholic University, Bandung, Indonesia, 40141; aswinlim@unpar.ac.id

[3]Department of Civil Engineering, Postgraduate Program, Universitas Bina Darma, Palembang, Indonesia 30251; nurly_gofar@binadarma.ac.id

***Correspondence: alfrendo.satyanaga@nu.edu.kz**

SUBMITTED 7 April 2022 **REVISED** 17 April 2022 **ACCEPTED** 19 April 2022

ABSTRACT Residual soil is covering some parts of the tropical country such as Singapore. It is commonly known that residual soil is associated with variability either horizontally or vertically. As a result, it is necessary to have a method to determine the distribution of soil properties to minimize the need to have an excessive amount of boreholes before the commencement of the construction project. This paper presents the suitable interpolation method for the spatial distribution of clay content across Singapore island. The common term related to the spatial distribution of soil properties is called digital soil map. The selection of the best method to generate this map depends on different factors. Therefore, each case requires a thorough evaluation and comparison based on the interpolation results. Singapore's digital elevation model and its boundary were utilized in this study. A digital soil map was developed using ArcGIS software based on two interpolation methods such as ordinary kriging, and the inverse distance weighted method. These two methods were cross-validated and compared based on the regression analysis of the analyses results. Cross-validation was performed to verify the correctness of the mentioned interpolation methods. In order to validate the interpolation results, the mean error and the root mean square error were used. The results suggest that ordinary kriging could be used as a suitable method to generate the spatial distribution of clay content obtained from boreholes in Singapore.

KEYWORDS ordinary kriging, Inverse Distance Weighted, digital soil map, spatial distribution

1 INTRODUCTION

The size of the soil grains and their characterization are essential in every step of construction engineering and valuable information to engineers and researchers. However, soil samples collected from the field could be of limited number due to many constraints such as budget, climate, and accessibility of particular locations. Much environmental and social simulation analysis needs soil parameters to calculate and predict variations, and disasters in our future living conditions. However, the availability of soil data is limited on both national scales. Using tools such as GIS software and interpolations methods in the software, researchers could predict the value of a parameter of soil at an unobserved location through ‘Digital soil mapping (DSM)’. In more detail, DSM is the process of generating a geographically referenced soil map based on quantitative parameters between spatial data and measurements made in the field and laboratory (Minasny and Mcbratney, 2015). Most researchers use a special function in ArcGIS environment called Spatial Analyst Tools for soil mapping. The software could be exploited to estimate parameters (elevation, precipitation, concentration, or other numerical data) for any geographic point on the map, provided that the input into the software is also numerical data (Daly et al., 2008).

ArcGIS has been exploited by many researchers including Nunes et al. (2016) who used the software to generate a soil vulnerability map based on numerical data (soil characteristics) obtained during soil sampling. The authors were able to continuously monitor the impact of irrigation on soil degradation. Therefore it would be interesting to have further studies on the accuracy of this technique for a broader area.

Further studies by Nistor et al. (2019) obtained a suitability map for the determination of soil sampling locations in Singapore by incorporating a digital elevation model (DEM), slope angle, and the soil sampling locations from previous research. This generated map in ArcGIS using the Spatial Analyst Tool could help engineers and further research to determine where to conduct soil sampling.

Recent research by (Ip et al., 2020) identified the locations prone to slope failure and developed a map. In their study, they estimated unsaturated and saturated soil properties using ordinary kriging interpolation. Similar research using GIS mapping was the study by Nistor et al. (2019), who analyzed the effect of climate on the spatial and temporal distribution of 5-day antecedent rainfall in Singapore. Kriging interpolation method in ArcGIS software generated a map of rainfall variation in Singapore. Interpolation was based on the historical data in 22 stations. However, in all this research, the sufficiency of the input data, accuracy of interpolation method, and obtained map have not been discussed thoroughly.

On the other hand, research conducted earlier by Mueller et al. (2004) analyzed the relative performance of Ordinary Kriging and IDW. However, evaluation was conducted using regression and prediction efficiency which is not the case in this research. Ip et al. (2021) attempted to evaluate the effectiveness of using some soil properties in predicting the spatial variation of shear strength properties: effective cohesion (c') and effective friction angle (ϕ'). The different interpolation methods in predicting c' and ϕ' of residual soils in Singapore were analyzed and discussed. The performance of the interpolation method was evaluated using the root mean square error (RMSE) and the mean prediction error (ME). The same approach will be used in the current study.

The main objective of the current research is to analyze the interpolation methods and sufficiency of data that should be input to obtain accurately and representative interpolated map. For this purpose, Singapore will be chosen as a study area due to 300 available borehole data throughout the city. Soil mapping will be developed by ArcGIS software for further analysis. This will be obtained by two interpolation methods such as Ordinary kriging (OK), and Inverse distance method (IDW).

2 RESEARCH METHODOLOGY

In the research, the author will use the soil database from different locations within 4 zones in Singapore. All data points used in this study were collected from boreholes within 2–6 m (Rahardjo et al., 2019). Subsequently, consolidated undrained triaxial tests and in-situ SPT were conducted. There was also index properties test such as grain-size distribution, Atterberg limit, specific gravity and moisture content tests.

ArcGIS 10.4 software will be used to make soil mapping through interpolation. For now, only ordinary kriging and inverse distance method (IDW) methods will be used for interpolation. Both methods are based on the strategy that nearby coordinates have more correlations with the particular point than distant ones and are usually used in environmental and engineering studies. Ordinary kriging interpolates the point based on an average weight of already known neighboring points within a specific area (Li and Heap, 2011).

"Arc Toolbox" window can help to convert data in an excel file into a readable format by ArcGIS 10.4 software. The 'Excel To Table' function was used from 'Conversions Tools'. The Attribute table could be converted into an Excel file using the 'Table To Excel' function later. The 'Geostatistical Analyst Tools' in ArcGIS performed the statistical computations using 'Ordinary Kriging' interpolation and Inverse Distance Weighting interpolation methods.

Ordinary kriging is a common interpolation method and easy form kriging that has a feature called B.L.U.E. (best linear unbiased estimator). It is simple to use and needs only soil data locations and values of the properties. However, it is a univariate method and when the number of data points is limited, it may result in poor predictions. Namely, OK can result in higher root mean square error (RMSE) and mean error (ME), due to its sensitivity to short-range variations (Laslett & McBratney, 1990). The distance between surrounding data points and interpolation point, direction are the main influencing factors to the variation of variables (Goovaerts, 1997).

The kriging is a linear weighted average of surrounding observations (Ip et al., 2021). The relevant equations for ordinary kriging interpolation are presented in Equations 1-3.

$$\hat{Z}_{OK}(x_0) = \sum_{i=1}^{N} \lambda_i^{OK} z(x_i) \tag{1}$$

$$\sum_{i=1}^{N} \lambda_i^{OK} \gamma(x_i, x_j) + \boldsymbol{\psi}(x_0) = \gamma(x_j, x_0) \tag{2}$$

$$\sum_{i=1}^{N} \lambda_i^{OK} = 1 \tag{3}$$

where $\hat{Z}_{OK}(x_0)$ - estimated soil property by ordinary kriging; N- Number of observations; $z(x_i)$ - the observed soil parameter; λ_i^{OK}- the ordinary kriging weights; $\boldsymbol{\psi}(x_0)$- the Lagrange multiplier; $\gamma(x_i, x_j)$- the semivariance between estimation locations; $\gamma(x_j, x_0)$- the semivariance between observation locations.

The inverse distance weighted method assumes that data points that are nearby are more similar to each other than objects that are distant from each other. The measured values closest to the predicted location have a greater influence on the predicted value than the ones farther from it. A relevant equation for interpolation of data based on the inverse distance weighted method can be seen in Equation 4.

$$\hat{Z} = \frac{\sum_1^n z_i / d_i^k}{\sum_1^n 1/d_i^k} \tag{4}$$

where, $\hat{Z}$ −estimated value at an unsampled point; n −number of control points used to estimate a grid point; k −power to which distance is raised; d −distances from each control point to unsampled points.

The interpolation results can be checked for accuracy using different methods but the simplest one is to evaluate using the root mean square error (RMSE) and the mean prediction error (ME). In the best case, ME should be as close as possible to zero, while RMSE should be small and equal to standard error. The relevant equations are shown in Equations 5 and 6 (Webster & Oliver, 2007).

$$ME = \frac{1}{l} \sum_{j=1}^{l} [\hat{z}(x_j) - z^*(x_j)] \tag{5}$$

$$RMSE = \sqrt{\frac{1}{l} \sum_{j=1}^{l} [\hat{z}(x_j) - z^*(x_j)]^2} \tag{6}$$

where, l −the number of validation point; $\hat{z}(x_j)$ −estimated values; $z^*(x_j)$ −actual observations;

3 STUDY AREA

The georeferencing is applied on the raster files (e.g. tiff, jpg, png) that need to become a truemap with real geographical coordinates. A reference map (Figure X) could be used for georeferencing. The 'Georeferencing Toolbar' should be activated and the raster file that needs to be georeferenced should be open into ArcGIS. The fit of the raster file with a real map is necessary to be completed. The fitting with a real map is called the georeferencing by shape with a reference map (Figure 1).

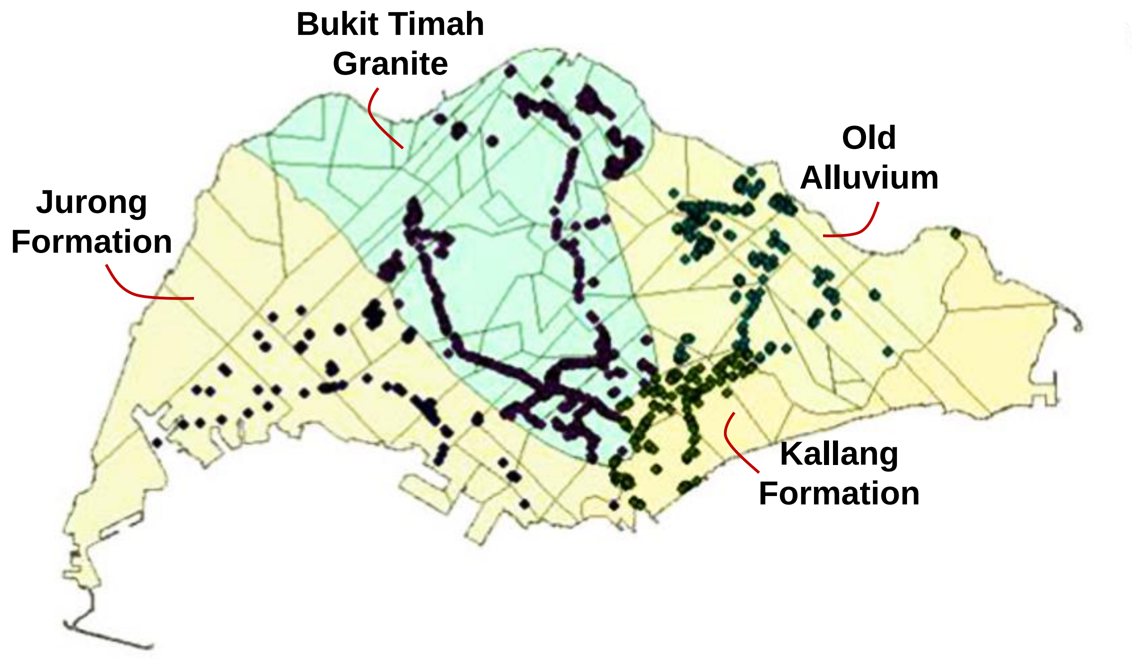
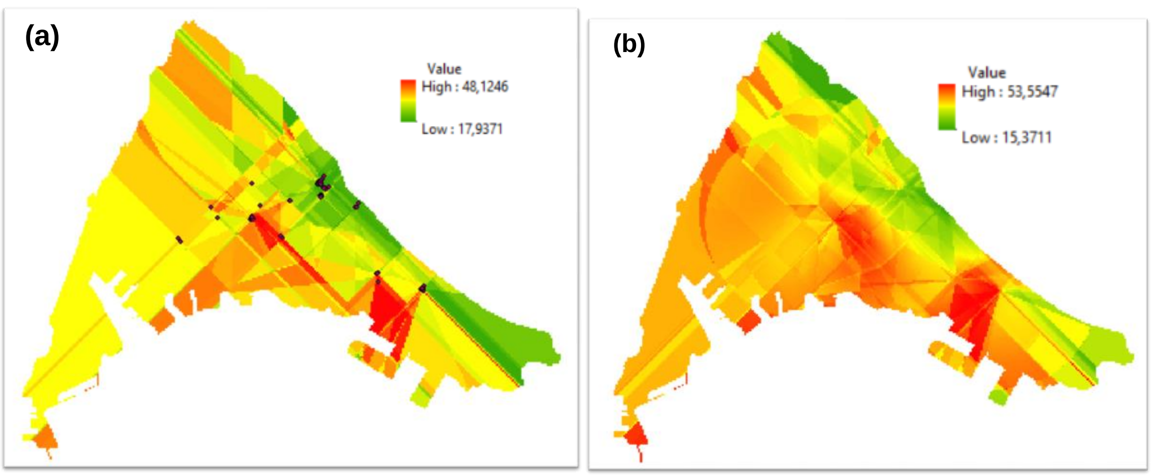


Figure 1. Locations of boreholes within residual soils from different formation in Singapore

The study area is located in Singapore. The main soil formations of Singapore were also given in Figure 1. The file of the geological formation of Singapore was obtained by manual vectorization using analogical maps. Singapore's geological landscape comprises four main formations, namely: BTG zone (Bukit Timah Granite), JF zone (Jurong Formation), KF zone (Kallang formation), OA zone (Old Alluvium). This study focuses on Jurong Formation (JF) zone, which is one of the four main geological formations. JF (sedimentary rocks) is located in the west and this sedimentary soil mainly consists of clayey with silt, sand material.

4 RESULTS AND DISCUSSION

To find the optimum number of data points, which gives an accurate interpolation, the analysis was carried out on subgroups from data points, which contains only 50, 100, 150, 200, 250, and finally all 300 data points. The spatial distributions of data based on the Kriging method and IDW method are presented in Figure 2 and Figure 3, respectively.



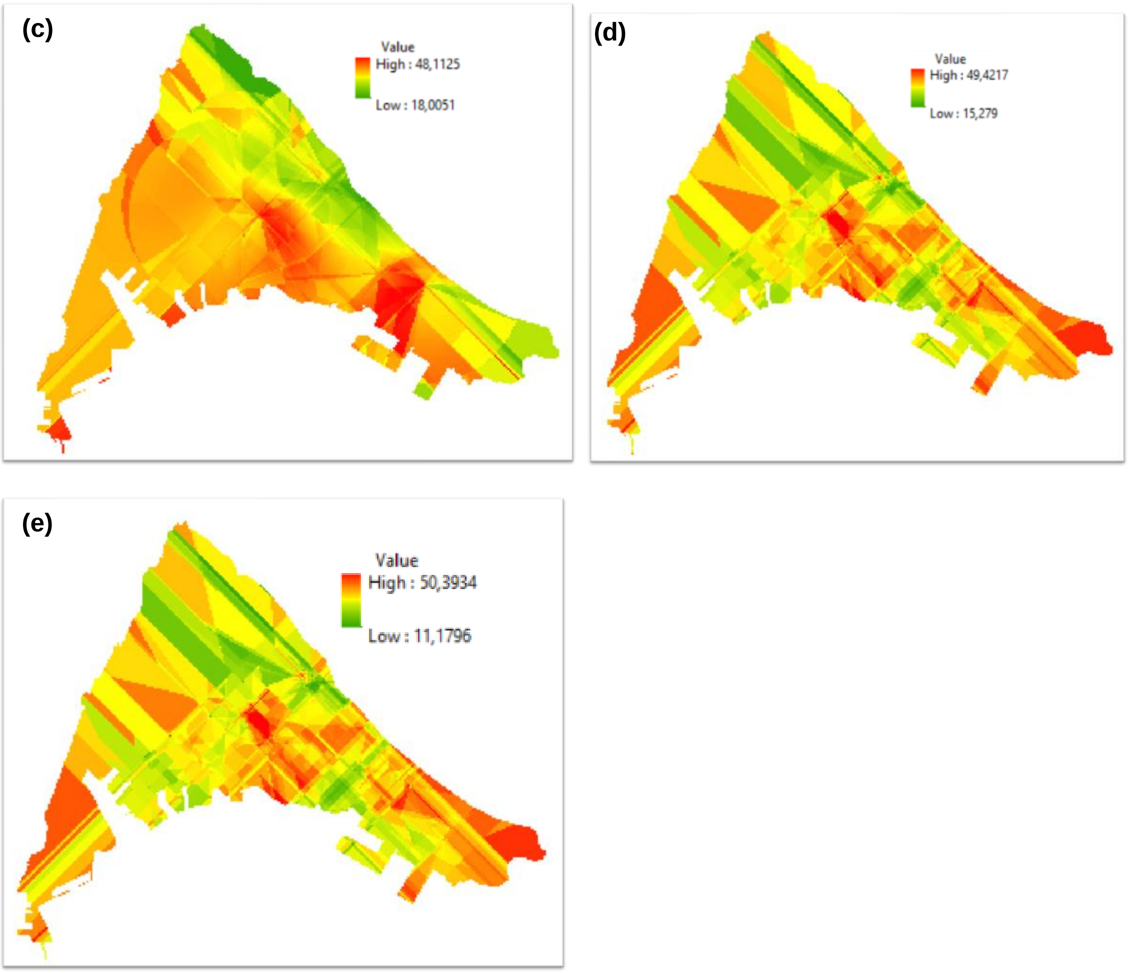
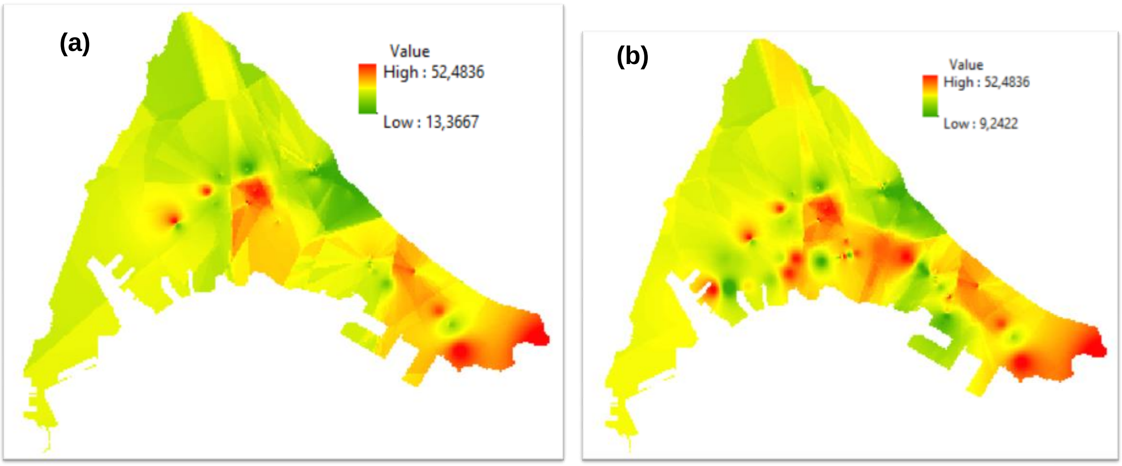


Figure 2. Spatial analyses of clay percentage distribution after Ordinary Kriging based on (a) 100 samples; (b) 150 samples; (c) 200 samples; (d) 250 samples and (e) 300 samples.



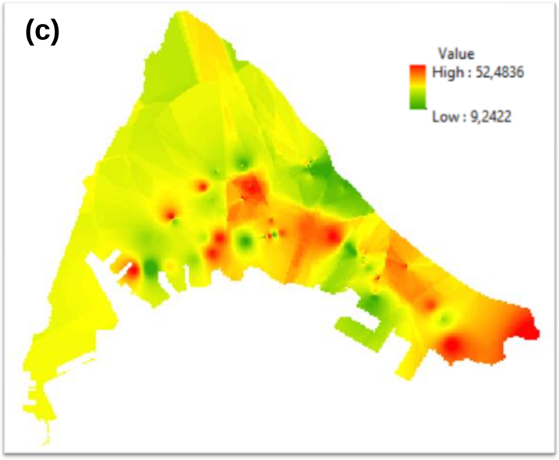


Figure 3. Spatial analyses of clay percentage distribution after IDW based on (a) 200 samples; (b) 250 samples and (c) 300 samples.

In many cases, it is necessary to assess the quality of the interpolation results obtained through different interpolation methods and choose the accurate one. This method analysis can be done by comparing the estimated values with actual known values at the data points. This is called method validation. The problem is that in many cases there is no data available for independent validation and we only have the data based on which the model was built. Cross-validation allows you to analyze the quality of a model using only such data.

In our case, validation of the spatial distributions of soil properties was carried out by cross-validation method. In more detail, this method removes one point from the data, then interpolates the value at the appropriate location based on the remaining data. This tool is used primarily for comparing an interpolated value with a measured value to obtain important information about some parameters of the model.

Cross-validation results of Ordinary Kriging for Cases 1-5 are presented in

Figure 4. The results of the cross-variation of Ordinary Kriging are summarized in Table 1. Cross-validation results of Inverse distance weighing for Cases 1-5 are presented in

Figure 5. The results of the cross-variation of Ordinary Kriging are summarized in Table 2.

Overall, the number of data points influenced the accuracy of both interpolation methods. When Mean Error is taken as the main criterion for evaluation Case 5 shows the least error for both OK and IDW. However, ME should be used when RMSE are equal in value which is not the case in this research (Ding et al. 2018). Therefore, the RMSE parameter was used as the main comparative indicator. The lower RMS error means the quality of the prediction.

For better understanding, the RMSE that were obtained from both OK and IDW interpolations were shown in Figure 6. When the sample size was around 100, IDW method was more sensible to sample size. Afterward, the number of data points similarly affected both OK and IDW. Regression analysis was done to check the optimum number of data required to generate a minimum error. The value of x-absis where the curve starts to bend is indicating the optimum sample sizes. Based on the result of IDW and OK interpolations, the error starts to be constant when the number of data points is more than approximately 168 for IDW and 155 for OK. At the optimum sample size, the RMSE is around 12.306 for OK and 12,904 for IDW which means OK showed more accurate results.

Nevertheless, the RMSE value for both interpolation methods is relatively high. It can be due to the high special variation of input values of clay distribution. To avoid such variation, it is necessary to get more samples of the soil on-site for accurate measurements. When looking at the results, it's obvious that OK provides better spatial visualization of the data according to its values from cross-validation. The obtained visualization and cross-validation values suggest that ordinary kriging could be used in the spatial presentation of soil parameters obtained from boreholes.

Similar research conducted in 2004 by Mueller et al. came to opposite results, IDW outperforming OK method. Also, the authors note that cross-validation should not be the sole criterion for deciding whether to use one interpolation procedure. Therefore, one limitation of this research and suggestion for further development would be to use other validation methods.

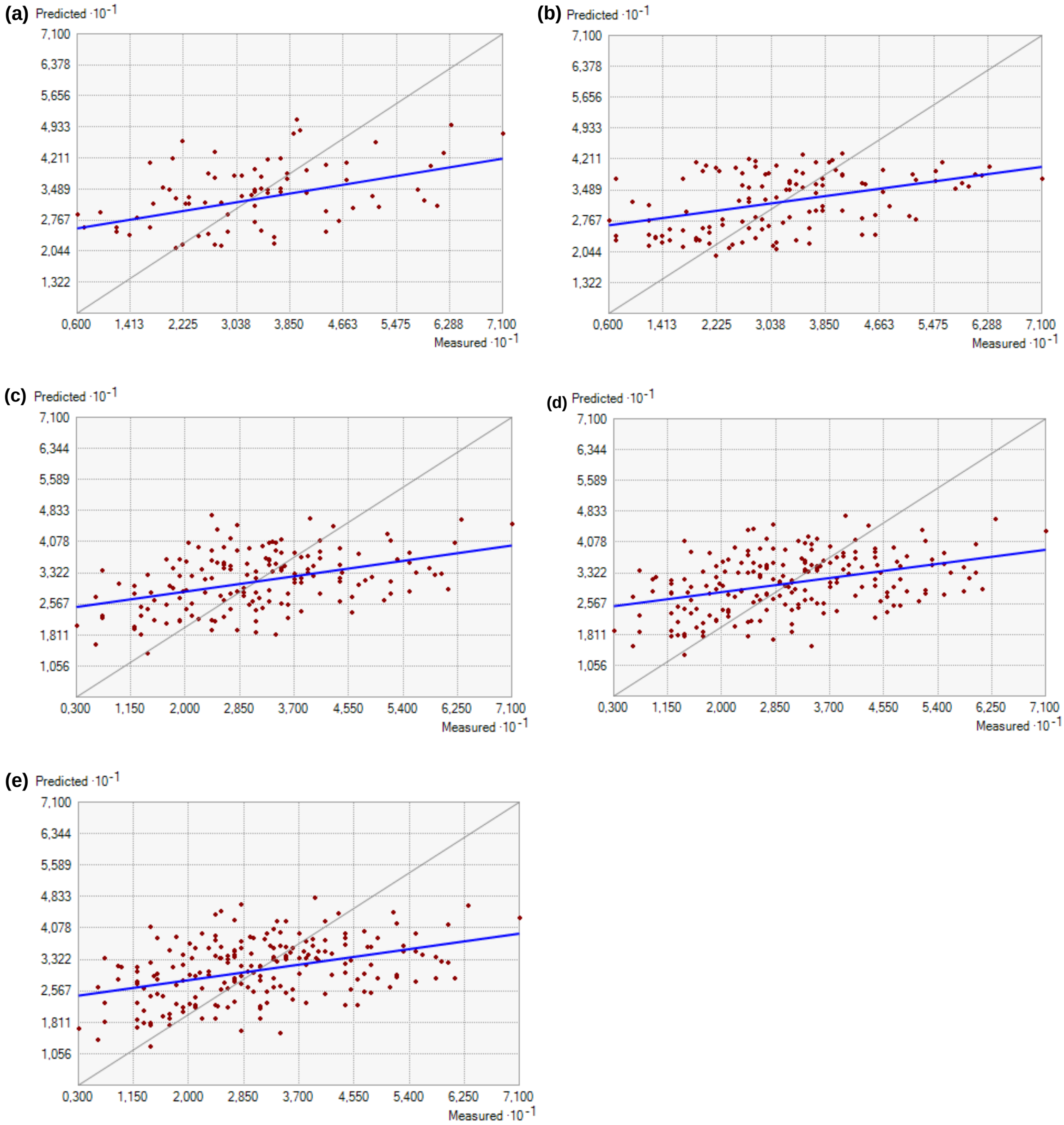


Figure 4. Cross validation of the Ordinary Kriging interpolation from a) Case 1 b) Case 2 c) Case 3 d) Case 4 e) Case 5

Table 1. Prediction errors from Ordinary Kriging analyses

Interpolation	Case 1	Case 2	Case 3	Case 4	Case 5
Samples	1-100	1-150	1-200	1-250	1-300
Mean	0,741	0,359	0,233	0,149	0,127
Root-Mean-Square	12,776	12,434	12,212	12,48	12,434
Mean Standardized	0,0427	0,029	0,0186	0,0122	0,010
Root-Mean-Square Standardized	0,8653	0,9865	0,9906	1,018	0,997
Average Standard Error	15,052	12,581	12,3452	12,259	12,462

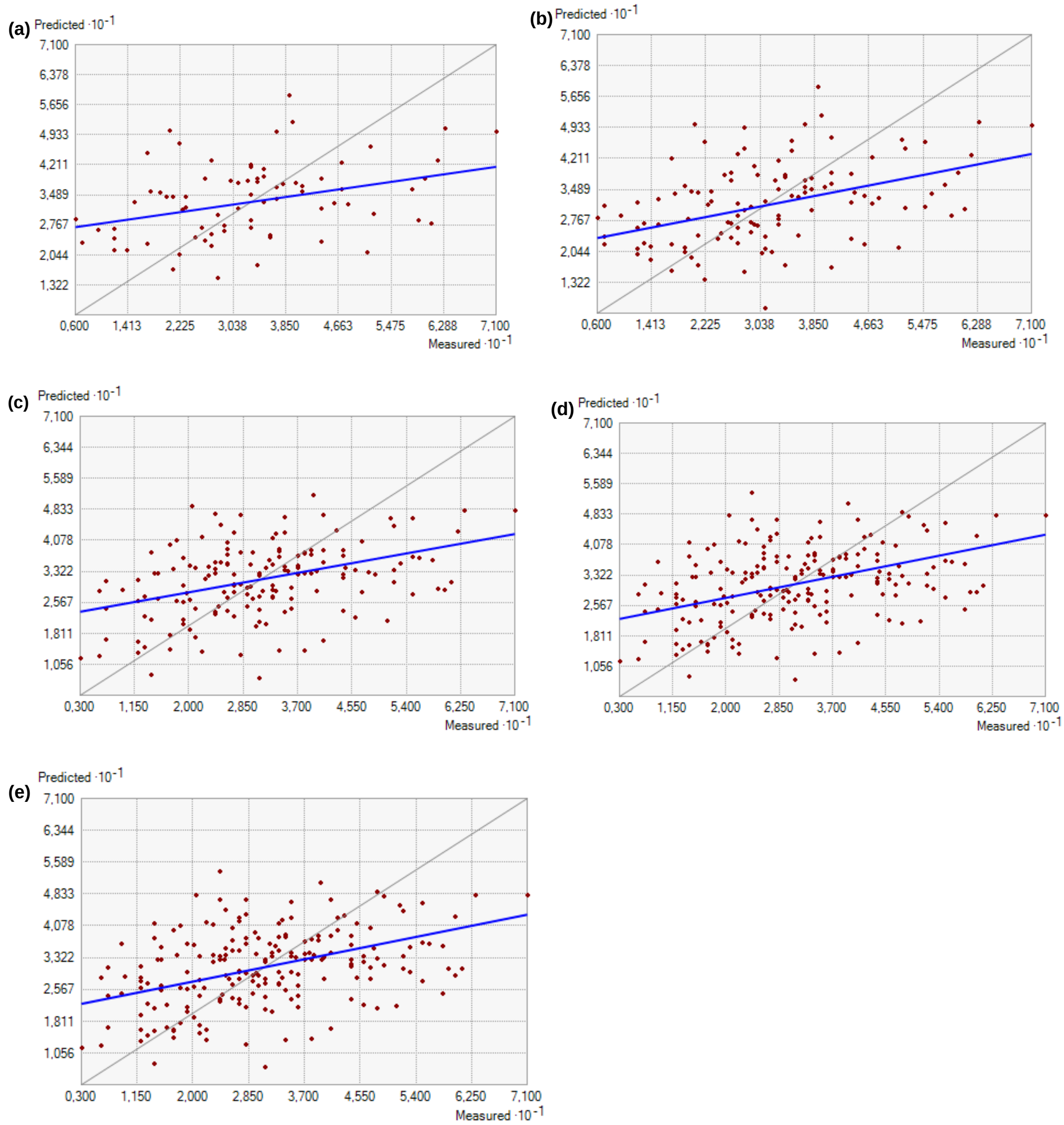


Figure 5. Cross validation of the IDW interpolation from a) Case 1 b) Case 2 c) Case 3 d) Case 4 e) Case 5

Table 2. Prediction errors from IDW analyses

Interpolation	Case 1	Case 2	Case 3	Case 4	Case 5
Samples	1-100	1-150	1-200	1-250	1-300
Mean	0,6765	0,1918	-0,1857	0,0767	0,0844
Root Mean Square	13,758	13,234	12,802	13,081	13,017

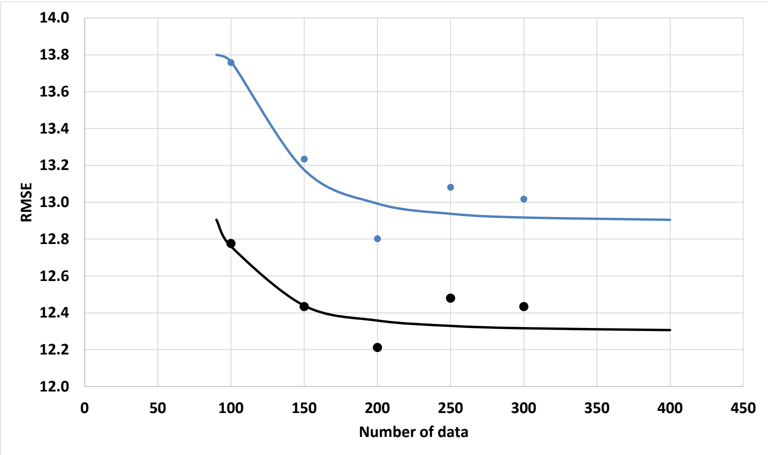


Figure 6. RMSE from analyses using Ordinary Kriging and IDW

5 CONCLUSION

To find a suitable interpolation method and identify a sufficient amount of data that should be input to obtain an accurate and representative interpolated map, Singapore was chosen as a study area. Soil mapping has been developed by ArcGIS software for further analysis. This was obtained by two interpolation methods such as Ordinary Kriging (OK), and Inverse distance method (IDW). The two methods were cross-validated and compared after developing regression analysis. Each method showed different spatial visualization. When estimating the performance of both methods, the average Root Mean Square error (RMSE) was higher for IDW, showing a max RMSE of 13.758 while max RMSE for OK was 12.776. Nevertheless, RMSE values for both interpolation methods were relatively high. The optimum sample size is approximately 168 for IDW and 155 for OK, and this was found by developing regression analysis. The value of x-absis where the curve starts to bend indicated the optimum sample sizes. At the optimum sample size, the RMSE is around 12.306 for OK and 12,904 for IDW, which means OK showed more accurate results. The obtained visualization and cross-validation values suggest that ordinary kriging could be used in the spatial presentation of soil parameters obtained from boreholes. One limitation of this research and suggestion for further development would be to use other validation methods.

DISCLAIMER

The authors declare no conflict of interest.

AVAILABILITY OF DATA AND MATERIALS

All data are available from the author.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support from Building Construction Authority and the sharing of the data from Singapore Land Authority. First Author was supported by the Nazarbayev University Research Fund under Grants 11022021CRP1512. The authors are grateful for this support. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Nazarbayev University.

REFERENCES

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, and J. Curtis, 2008. ‘Physiographically Sensitive Mapping of Climatological Temperature and Precipitation Across the Conterminous United States.’ International Journal of Climatology, vol. 28, pp. 2031–2064.

Ding, Q., Wang, Y. & Zhuang, D. 2018. ‘Comparison of the commo spatial interpolation methods used to analyze potentially toxic elements surrounding mining regions’. Journal of Environmental Management, vol. 212, pp. 23-31.

Ip, S. C., Rahardjo, H., & Satyanaga, A., 2020. ‘Three-dimensional slope stability analysis incorporating unsaturated soil properties in Singapore’. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, pp. 1-15.

Ip, Sabrina CY, Alfrendo Satyanaga, and Harianto Rahardjo. 2021. ‘Spatial variation of shear strength properties incorporating auxiliary variables.’ Catena, 200: 105196.

Li, Jin & Heap, Andrew. (2011). ‘A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors’. Ecological Informatics. Vol.6, pp. 228-241.

Minasny, Budiman & Mcbratney, Alex., 2015. ‘Digital soil mapping: A brief history and some lessons. Geoderma’. 264.

Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., Barnhisel, R. I., & Shearer, S. A. 2004. ‘Map Quality for Ordinary Kriging and Inverse Distance Weighted Interpolation.’ Soil Science Society of America Journal, Vol.68, no.6, p. 2042.

Nistor, M., Rahardjo, H., Satyanaga, A., 2019. ‘Development assessment of the Singapore land: a GIS spatial-temporal approach based on land cover analysis’, Geographia Technica, [online] vol.14, no. 2, pp. 60 – 73.

Nistor, M., Rahardjo, H., Satyanaga, A., Hao, K., Wai Lun, S. A., And Leong, C., 2019. 'GIS-based approach to identify the suitable locations for soil sampling in Singapore', Geographia Technica, vol. 14, no. 1, pp. 103-117.

Nunes, J. R., Loures, L., Lopez-Piñeiro, A., Loures, A., & Vaz, E., 2016. 'Using GIS towards the characterization and soil mapping of the caia irrigation perimeter', Sustainability. vol. 8, no. 4, p. 368.

Scull, P.; Franklin, J.; Chadwick, O. A.; McArthur, D., 2003. ‘Predictive soil mapping: a review’. Progress in Physical Geography, vol. 27, no. 2, pp. 171–197.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists, 2nd. ed. John Wiley & Sons, England, p. 2007.