Comparison of Interpolation Techniques for Assessment of Spatial Variability of Soil Chemical Properties for Oil Palm Plantation Zonal Management

S Q Tan¹*, A S Mat Su¹, A Wayayok², and A S Sukor³

¹ Department of Agriculture Technology, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia.
² Department of Biological and Agricultural Engineering Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia.
³ Department of Land Management, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia.

*Email: tansq96@gmail.com

Abstract. Spatial variability map of the soil properties is highly dependent on availability of data either in detailed or semi-detailed resolution. Variation on the data resolution affects the interpolation technique, thus the zonal management zone. However, there are many interpolation techniques offered, thus raises the issues of the most suitable method in zone classification. The study aim is to evaluate the best interpolation method in assessing the spatial variability of the soil chemical properties from highly weathered tropical mineral inland 108.5 ha oil palm plantation. These soil samples were brought back for further lab analysis. The spatial variability of the soil chemical properties was done by using interpolation techniques were utilized for geospatial statistical and deterministic analysis. Due to a variety of algorithms embedded in the interpolation techniques, the strength of Pearson correlation coefficient (r), Nash–Sutcliffe efficiency (NSE), percent bias (PBIAS), and RMSE-observations standard deviation ratio (RSR) were applied for accurate quantification of interpolation method. On average, EBK and OK were the suitable interpolation methods with the least average model validation statistical value i.e. error. Therefore, the evaluation and comparison of r, NSE, PBIAS, and RSR were appropriate approaches in determining the accuracy of the interpolation method for the zonal management practices.

1. Introduction

Geostatistics has been widely applied in several studies on the spatial variability of soil physicochemical properties. Understanding of the soil spatial heterogeneity will provide better knowledge on soil pedon and soil type variation, such as from Rosemary et al. (2017) [1] who explored the spatial heterogeneity of Alfisol soil. In addition, Gaston et al. (2001) [2] studied spatial variability of soil physicochemical properties and the weed density. Besides that, soil properties of the paddy field had been reported by Aishah et al. (2010) [3] and using the zone delineated by Aimrun et al. (2011) [4] for nutrient recommendation management using geostatistics approach. The geostatistics methods were derived in several approaches and hence utilized in various type of software conventionally available in the market,
including ArcMap (ESRI, United States), GS+™ (Gamma Design Software, United States) and many more.

Site-specific management is the targeted aim for the stated research projects. Hence, gathering and acquisition of the site-specific information procedures are crucial [5]. To test the accuracy of the predictor model in geostatistics, several pieces of researches had been conducted. For instance, Mousavi et al. (2017) [5] had indicated that kriging (OK) perform better than inverse distance weighting (IDW) for soil properties (sand, silt, bulk density, total porosity, pH, electrical conductivity and calcium carbonate equivalent). For evaluating the performance of interpolators, the authors used mean bias error (MBE) and mean absolute error (MAE). Whilst, Bhunia et al. (2018) [6] had compared five interpolation techniques for the spatial distribution of soil organic carbon (SOC) in India. The authors found out OK is the best among IDW, local polynomial interpolation (LPI), radial basis function (RBF), and empirical Bayesian kriging (EBK). The authors had done cross-validation by evaluating four interpolation techniques. Thus, the best interpolation technique used to quantify the zoning for the site-specific management zone remains ambiguous especially dealing with high variability of the agriculture soil.

In this scenario, four interpolation methods were compared in this study, such as (1) empirical Bayesian kriging (EBK) and (2) ordinary kriging (OK) by using Esri’s ArcGIS Geostatistical Analyst, (3) Inverse distance weighting (IDW), and (4) spline (S) analysis by using Esri’s ArcGIS spatial analyst. The aim of this study is to evaluate the best interpolation method in assessing the spatial variability of the soil chemical properties for the highly weathered tropical mineral inland soil of oil palm plantation.

2. Methodology

2.1. Study site and soil sampling

The study site is located at an oil palm plantation in Bukit Senorang Estate, Temerloh District, Pahang, about 3 hours driving from Kuala Lumpur, with a total area of 108.5 ha. The study site consists of three (3) replanting regions which are Block 2018 A, B, and C, planted with mixed of one-year-old Yangambi-breed and DXP breed oil palm. The soil series is classified as Durian series, with the most dominant soil textures were silty clay and silty clay loam, situated between 12-24% slope. A total of 108 geo-referenced soil sampling points at two different depths of 0-30 cm and 30-60 cm were identified using a handheld GPS unit (Trimble Juno 3B, Trimble, California, USA). The samples then brought to the lab for further soil properties analysis and then were used to quantify the field spatial heterogeneity.

2.2. Soil chemical properties for laboratory analysis

Chemical properties analysis mainly involves soil exchangeable cation including potassium (Exc. K) calcium (Exc. Ca) and magnesium (Exc. Mg) were extracted by shaking method (ammonium acetate) which suggested by Sukor et al. (2017) [7]. Then, Bray and Kurt 2 extracting solution was used to extract soil available phosphorus (Avail. P), which suggested Standard Malaysia for Malaysian soil. Then, the leachates of Avail. P were analysed byAutoAnalyzer (AA), while, leachates of soil exchangeable cation was analysed by Atomic Absorption Spectrophotometer (AAS). Soil pH level (distilled water) was analysed by using pH Meter (HI 2211 pH/ ORP). Besides that, by referring to Jones Jr. (2001) [8] laboratory electrical conductivity (EC_{lab}) was measured by the same soil-to-water ratio as soil pH measurement, but need to make sure the soil samples in solution were settled down after 24 hours. In total 12 samples parameter (6 soil chemical properties x 2 depths) were used for the geostatistical evaluation.

2.3. Descriptive statistic and geostatistical analysis

Datasets were separated randomly into 70% of modelling subset for interpolation methods calibration and 30% of testing subset for estimation. Microsoft Excel software was used to pre-analyse descriptive statistic of the overall dataset for understanding the central tendency, data dispersion, data distribution, and skewness of data. Then, simulation map of the drilling subset was calculated by using EBK general computational model and explained by Krivoruchko and Gribov (2019) [9] as shown equation (1-4).
\[ z_i = y(s_i) + \epsilon_i, \quad i = 1 \ldots K, \]  

\[ \hat{y}(s) = \sum_{i=1}^{K_{sim}} (w_i \cdot E[y(s)|z, \Theta_i]) \]  

\[ \text{var}[y(s)|z] = \sum_{i=1}^{K_{sim}} (w_i \cdot ((\text{Var}[y(s)|z, \Theta_i]) + (E[y(s)|z, \Theta_i] - \hat{y}(s))^2)) \]  

\[ \gamma(h) = \text{nugget} - b|h|^{\alpha}, \quad \alpha \in (0.25, 1.75) \]  

where \( z_i \) is the measured vector value of observed location, \( S_i \). And, \( y(s_i) \) is the Gaussian process under study at the location \( s \). Whilst, \( \epsilon_i \) is the measurement error and \( K \) is the number of measurements. This equation was further described in equation (2-3). New spatial process \( \Theta_i \) which new values are unconditionally simulated at the location \( K_{sim} \), where \( i = 1 \ldots K_{sim} \). Weight \( w_i \) for each simulated model is calculated using Bayes’ rule. Predictions and prediction standard error are produced at specific locations using equation (2-3). Equation (4) is the power model which also default univariate EBK model was applied, where \( n \) is the nugget value, \( b \) is the slope of coefficient, \( h \) is the distance between the locations under study and \( \alpha \) is power parameter [9].

Then, second and fifth interpolator for assessment is OK which was expressed as equation (5-6). The variance of the differences of OK, usually denoted by \( \gamma \), is the semi-variance [10].

\[ \gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \]  

\[ \hat{z}(B) = \sum_{i=1}^{n} \lambda_i z(x_i) \]  

Semivariogram value \( \gamma(h) \) in equation (4) was calculated using differences of \( z \) value at position \( (x_i) \) and \( (x_i + h) \) \( N(h) \) is the pair number of observations which separated by lag, \( h \). Equation (6) is a prediction equation for both IDW and OK, where \( \hat{z}(B) \) is the estimate over a specific block of land. But, the weight \( \lambda_i \) of IDW only depends on the distance to predict the location. While, OK prediction also relies on the fitted model to the measured point, for instant, exponential model, spherical model, Gaussian model, and circular model. Parameter optimization is applied for OK which helps to fit semivariogram especially focused on the range as minimizing the mean square error for a specific semivariogram model [12].

Besides that, S was also tested for the soil chemical properties prediction and variability. Equation (7) is presenting an algorithm used for surface interpolation of S tool [14].

\[ S(x, y) = T(x, y) + \sum_{j=1}^{N} \lambda_j R(\eta) \]  

\[ T(x, y) = a_1 + a_2 x + a_3 y \]  

\[ R(\eta) = \frac{1}{2\pi} \left\{ \frac{r^2}{4} \left[ \ln \left( \frac{r}{2\tau} \right) + c - 1 \right] + \tau^2 \left[ K_0 \left( \frac{r}{\tau} \right) + c + \ln \left( \frac{r}{2\tau} \right) \right] \right\} \]  

Particularly, the equation consists of \( \lambda_j \) represent coefficient found by the system solution of linear equation and \( \eta \) is the distance from the point \( (x, y) \) to the \( f^{th} \) point. \( T(x, y) \) shown as equation (8) is the trend function depending on the selected option. Despite this, the default option (regularized option) of the linear equation is chosen for this study. For equation (9), \( r \) represents the distance between the point and the sample, \( \tau^2 \) represents the weight of the parameter and \( K_0 \) represents the modified Bessel function. In the equation, constant \( c \) has a value of 0.577215 [14].
2.4. Validation and selection of interpolation methods

The estimated testing subset \( (E_i) \) were further compared with observed testing subsets \( (O_i) \) for interpolation methods validation. Pearson’s correlation coefficient, \( r \) [11] was utilized for describing the degree of collinearity between \( E_i \) and \( O_i \). The value of \( r \) ranges from -1 to 1 which is the index of the degree of the linear relationship between \( E_i \) and \( O_i \). In formula (10 & 11), \( \bar{O} \) represents mean value of observed data. Whereas, dimensionless validation method includes Nash-Sutcliffe efficiency (NSE) [15], which expresses as following (equation 9).

\[
NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - E_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2}
\]  

(10)

NSE was applied to indicate how well the plot of \( E_i \) versus \( O_i \) fits. It ranges from \(-\infty\) to 1.0, but only values between 0.0 and 1.0 are generally viewed as acceptable levels of performance.

In addition, percent bias (PBIAS) and RMSE-observation deviation ratio (RSR) which also recommended by Moriasi et al. (2007) [15] as error-index validation methods. Equation (10) below is PBIAS which measures the average tendency of the modelling data to be larger or smaller than observed counterparts and expressed as a percentage. The optimal value of PBIAS is 0.0, which indicates accurate model simulation. Whilst, getting a positive value indicates overestimation bias and vice versa [13].

\[
PBIAS = \left[\frac{\sum_{i=1}^{n}(E_i - O_i) \cdot 100}{\sum_{i=1}^{n}(E_i)}\right]
\]  

(11)

RSR was also recommended by Moriasi et al. (2007) and Gupta et al. (1999) [15 & 17] because it incorporates the NSE which was applied to indicate how well the plot of \( E_i \) versus \( O_i \) fits. It ranges from \(-\infty\) to 1.0, but only values between 0.0 and 1.0 are generally viewed as acceptable levels of performance. In additional, percent bias (PBIAS) and RMSE-observation deviation ratio (RSR) which also recommended by Moriasi et al. (2007) [15] as error-index validation methods. Equation (11) below is PBIAS which measures the average tendency of the modelling data to be larger or smaller than observed counterparts and expressed as a percentage. The optimal value of PBIAS is 0.0, which indicates accurate model simulation. Whilst, getting a positive value indicates overestimation bias and vice versa [13] benefits of error-index and includes a scaling factor. The following equation (12) is representing its calculation.

\[
RSR = \frac{RMSE}{\text{Std Dev}_{test}} = \frac{\sqrt{\sum_{i=1}^{n}(O_i - E_i)^2}}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}}
\]  

(12)

Hence, the ‘best’ interpolation method was selected after going through three validation steps. After that, the maps of the study site were created to visualize the spatial variability of soil chemical properties. The summary of the data analysis procedure was illustrated in Figure 1.
3. Result and Discussion

The soil chemical properties vary from one sample point to another. This can be seen in Table 1, as most of the properties show relatively high in the coefficient of variation (CV) values. It is expected that this field to have quite a large spatial variability due to the landscape position, elevation, soil texture, and soil genesis or series, agronomic practices, and historical fertilization application in oil palm plantation. Most of the soil chemical results greater level of dispersion around the mean where CV range 50-120%, except soil pH. The Avail. P shows the highest variation (119.5%) especially soil depth of 30-60 cm, however soil pH show the lowest variation (4.88%) among soil chemical properties as similarly being reported in previous study including [18] and [19]. As mentioned in [20], the low variation of soil pH is due to log scale of proton (H+) concentration in soil solution. Thus, EC_{lab} tend to have low variability throughout the entire site.

| Parameter     | Soil pH | EC_{lab} | Avail. P | Exc. K | Exc. Mg | Exc. Ca |
|---------------|---------|----------|----------|--------|---------|---------|
| Depth (cm)    | 30 60   | 30 60    | 30 60    | 30 60  | 30 60   | 30 60   |
| Unit          | µS/cm   | µg/g     | cmol/kg  |        |         |         |
| Minimum value | 3.97 3.91 | 26.30 24.80 | 0.15 0.78 | 0.11 0.06 | 0.06 0.05 | 0.40 0.29 |
| Maximum value | 5.84 5.13 | 218.00 199.50 | 175.98 140.91 | 0.86 0.85 | 2.17 1.26 | 4.12 4.27 |
| Mean          | 4.45 4.43 | 82.10 82.18 | 30.47 24.28 | 0.35 0.33 | 0.29 0.23 | 1.17 1.04 |
| Std. Dev.     | 0.24 0.22 | 30.47 31.28 | 33.28 29.02 | 0.17 0.18 | 0.27 0.18 | 0.74 0.75 |
| CV (%)        | 5.31 4.88 | 37.12 38.06 | 109.23 119.50 | 48.60 52.61 | 94.49 76.55 | 62.78 72.28 |

*Note: Std. Dev. stand for standard deviation and CV stand for coefficient of variation

Most of the soil chemical properties show positive (+) Pearson’s correlation coefficient (r values), except for a few soil chemical properties on different interpolation techniques (Figure 2). There are significant relationships (p<0.05 and p<0.01) between predicted and observed data for Exc. K and Exc.
Mg by using most of the interpolation techniques. Among the interpolators, EBK tended to perform better for Exc. Mg by getting significant correlation ($r=0.52$ for depth 0-30cm, $p<0.01$ & $r=0.43$ for depth 30-60cm, $p<0.05$). Whilst, OK was well-performing on Exc. K interpolation by resulting significant correlation between predicted data and observed data ($r=0.52$ & 0.54 for both depths respectively, $p<0.01$). EBK able to predict most of the chemical properties in terms of Pearson’s correlation validation, except $EC_{lab}$, Avail. P and Exc. Ca for deeper soil sampling depth.

![Figure 2.](image)

Figure 2. Graph of Pearson’s correlation coefficient between the estimated versus observed soil chemical properties. Critical value of the validation are 0.349 ($\alpha=0.05$) and 0.449 ($\alpha=0.01$).

Normalized values using the NSE validation as shown in Figure 3(a), OK, IDW, and EBK interpolation tend to have a better performance in overall. Because, Moriasi et al. (2007) [15] mentioned values $\leq 0.0$ indicate that the mean observed value is better predictor than the estimated value from modelling data. By narrowing the selection, EBK results in fewer and lower strengths of negative values. Despite the good result, modelling value of Exc. K and Exc. Mg perform better in order to predict the testing subset.

From Figure 3(b), the result tells whether the estimated data when compared to observed data (testing subset) for an interpolation method is close to the optimal value of 0.0, which indicating accuracy model prediction [15]. By observing the graph trend, OK, IDW and EBK interpolation tend to have a similar, but slightly different in the strength of the tendency. For exploring RSR validation results, it is important to understand the lower the RSR same as RMSE which indicates that better prediction of an interpolator [15]. From Figure 3(c), EBK performed better by having the lowest RSR value (1.007) among tested interpolation methods. In the hierarchy order, EBK>OK>IDW>S with the respective values is presented in the sub-table below. It was worth to mention that, 25% of the sample parameters shows low RSR when predicted using OK than EBK, namely $EC_{lab}$ at 0-30 cm and 30-60 cm, and Ca at 30-60 cm.
Figure 3. Interpolation methods performance boxplots on soil chemical properties estimation; (a) for NSE, (b) for PBIAS, and (c) for RSR validation results. The box plots represent the minimum, first quartile, median, third quartile, and maximum of error index across the soil chemical properties.

On average, the EBK found to be the best interpolation method followed by OK with the least average model validation statistical value i.e. error. Therefore, the evaluation and comparison of r, NSE, PBIAS, and RSR were appropriate approaches in determining the accuracy of the interpolation method for the zonal management practices.

After that, the maps were produced for the zone delineation according to fertility status for the oil palm is suggested by [21]. The parameter which had been mentioned [21] are soil pH (Figure 4 a & b), Avail. P (Figure 4 c & f), Exc. K (Figure 4 g & h) and Exc. Mg (Figure 4 k & l). On the other hand, EC_{lab} and Exc. Ca were classified by the geometrical interval classification method.

The fertilizer recommendation is various according to fertilizer type used, palm age, and site-specific properties. Hence, it is important for us to understand more about the requirement of the palm, so, more accurate fertilizer program could be issued to that particular plantation site.
4. Conclusion
This study shows the importance of geostatistical analysis tools comparison hence more accurate spatial information could be contributed to the oil palm plantation. From the analysis conducted, EBK and OK were the suitable interpolation methods by producing the lowest error of prediction. In order to delineate zones of management (Figure 4), it is important for the authors to do a deeper exploration of other soil
physical and chemical properties. It worth also to highlight that the results may only apply for this study site, due to the unique soil characteristic from one location to another.

Acknowledgements
The authors would like to express special thanks of gratitude to United Malacca Berhad for funding the research project (Fund No.: UPM/UMB/2018/6300918). Besides that, authors are also thankful to laboratory staffs in Makmal Pengajaran 1, Faculty of Agriculture and Spatial Information System Laboratory, Faculty of Engineering in Universiti Putra Malaysia for their assistance and guidance in completing sample and data analysis.

References
[1] Rosemary F, Indraratne S P, Weerasooriya R, and Mishra U 2017 Exploring the spatial variability of soil properties in an Alfisol soil catena *Catena* 150 pp 53-61
[2] Gaston L A, Locke M A, Zablhotowicz R M and Reddy K N 2001 Spatial variability of soil properties and weed populations in the Mississippi Delta *Soil Science Society of America Journal* 65(2) pp 449-459
[3] Aishah A W, Zauyah S, Anuar A R, and Fauziah C I 2010 Spatial variability of selected chemical characteristics of paddy soils in Sawah Sempadan, Selangor, Malaysia *Mal. J. Soil Sci* vol. 14 pp 27-39
[4] Aimrun W, Amin M S M, Ezrin M H and Mastura M 2011 Paddy soil properties and yield characteristics based on apparent electrical conductivity zone delineation for a humid tropical rice farm. *African Journal of Agricultural Research* vol. 6(23) pp 5339-5350.
[5] Mousavi S R, Sarmadian F, Dehghani S, Sadikhani M R, and Taati A 2017 Evaluating inverse distance weighting and kriging methods in estimation of some physical and chemical properties of soil in Qazvin Plain *Eurasian Journal of Soil Science* vol. 6(4) pp 327-336
[6] Bhunia G S, Shit P K, and 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* vol. 17(2) pp 114-126
[7] Sukor A, Azira A Z A, and Husni M H A 2017 Determination of Cation Exchange Capacity of Natural Zeolite: A Revisit *Malaysian Journal of Soil Science* vol. 21 pp 105-112
[8] Jones Jr J B 2001 *Laboratory guide for conducting soil tests and plant analysis* CRC press.
[9] Krivoruchko K, and Gribov A 2019 Evaluation of empirical Bayesian kriging *Spatial Statistics*, vol. 32 100368
[10] Oliver M A, and Webster R 1990 Kriging: a method of interpolation for geographical information systems *International Journal of Geographical Information System* vol. 4(3) pp 313-332
[11] Pearson K 1895 Notes on regression and inheritance in the case of two parents *Proc. R. Soc. Lond.* 58 240–242.
[12] Parameter optimization Retrieved February 6 2020 from https://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/parameter-optimization.htm
[13] Zad M, Najja S, Zulkaflı Z, and Muharram F M 2018 Satellite rainfall (TRMM 3B42-V7) performance assessment and adjustment over Pahang River Basin, Malaysia. *Remote sensing*, vol. 10(3), no. 388.
[14] How Spline works Retrieved February 5 2020 from https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-spline-works.htm
[15] Moriasi D N, Arnold J G, Van Liew M W, Bingner R L, Harmel R D and Veith T L 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* vol. 50(3) pp 885-900
[16] Gupta H V, Sorooshian S and Yapo P O 1999 Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering* vol. 4(2) pp 135-143
[17] Singh J, Knapp H V and Demissie M 2004 Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT (ISWS CR 2004-08) Champaign, Ill.: Illinois State Water Survey
[18] Aimrun W, Amin M S M, Ahmad D, Hanafi M M and Chan C S 2007 Spatial variability of bulk soil electrical conductivity in a Malaysian paddy field: key to soil management *Paddy and Water Environment* vol. 5(2) pp 113-121
[19] Tsegaye T and Hill R L 1998 Intensive tillage effects on spatial variability of soil test, plant growth, and nutrient uptake measurements. *Soil Science* vol. 163(2) pp 155-165
[20] Sun B, Zhou S and Zhao Q 2003 Evaluation of spatial and temporal changes of soil quality based on geostatistical analysis in the hill region of subtropical China. *Geoderma* vol. 115(1-2) pp 85-99
[21] Goh K J and Po, S B 2005 March Fertilizer recommendation systems for oil palm: estimating the fertilizer rates In Proceedings of MOSTA (Best practices workshops-agronomy and crop management) Malaysian Oil Scientists' and Technologists' Association pp 1-37