Comparation of Spatial Interpolation Methods on Slowly Available Potassium in Soils

Lingling Liu¹, Yu Jin¹, Jie Wang¹, Qi Hong¹, Zhenggao Pan² and Jinling Zhao¹

¹National Engineering Research Center for Analysis and Application of Agro-Ecological Big Data, Anhui University, Hefei 230601, China
²Information Engineering Institute, Suzhou University, Suzhou 234000, China
aling0123@163.com

Abstract. Soil nutrients are important parameters for maintaining the plant growth. Conventional in-situ methods are usually labor-intensive, low efficiency and time-consuming. Conversely, spatial interpolation is an extremely efficient method to derive the spatial distribution trend from discrete ground truth data. In this study, a total of 402 sample data of soil nutrients was collected in the wheat fields of Tongzhou District, Beijing, China. Slowly available potassium (SAK) was used as the interpolation factor. Four methods were comparatively employed to evaluate the availability and accuracy, specifically including inverse distance weighted (IDW), ordinary kriging (OK), spline function (Spline) and Trend interpolation. There are three models of OK, spherical model, exponential model and Gaussian model. Cross-validation was performed to compare the interpolation accuracy using the mean error (ME) and root mean square error (RMSE). The results show that OK is superior to the other methods, in which the exponential model is the best. In comparison with spherical function and Gaussian function, exponential model has the best performance with the ME and RMSE of 176.7084 and -1.1283, respectively.

1. Introduction
Soil nutrients are the essential nutrient elements for plant growth provided by the soil [1-3]. Understanding the spatial distribution of soil nutrients is the basis of performing the soil testing and formulated fertilization and carrying out the precision agriculture. Due to the complexity of soil nutrient distribution and the limitation of sample data, it is difficult to learn about the distribution of nutrients in the whole study area depending on traditional in-situ measurements. Conversely, spatial interpolation techniques have greatly facilitated the operation of evaluating soil nutrients at a larger scale [4].

Spatial interpolation analysis is a commonly used spatial analysis method in GIS, and it is also one of the important characteristics of GIS system different from other information systems [5]. The closer the spatial position is to the known observation point, the more likely it is to have similar characteristic values. The further away from the known observation point, the less likely it is that its characteristic values are similar, which is the most basic theoretical hypothesis of spatial interpolation technology [6]. On this basis, a variety of interpolation methods are proposed such as the inverse distance weighted, Kriging, spline function interpolation, trend surface interpolation and so on.

In general, different interpolation methods need to be selected according to the distribution characteristics of sample data. In our study, as one of the important elements of soil nutrients, slowly available potassium (SAK) was used as the case study to compare the availability of different methods in ArcGIS 10.2 software. The primary objective is aimed at finding out the optimal interpolation method for SAK in soils.
2. Materials and Methods

2.1. Collection and Preprocessing of Soil Nutrient Data

In this study, the soil nutrient data were collected in the rural areas of wheat fields in Tongzhou District, Beijing, China. The grid single point sampling method was used and the sub-meter high-resolution GPS receiver was simultaneously used to accurately locate the sampled plots. A total of 402 sample points was collected. The GPS-measured sampling points with coordinate records are converted into spatial points with spatial coordinates by using GIS software ArcGIS, and perform a projection transformation to generate a sample distribution map with slowly available potassium nutrient content information.

2.2. Inverse Distance Weighted

Inverse distance weighted (IDW) interpolation uses a linear combination of weights from a set of sampling points to determine the pixel value [7]. The weight is an inverse distance function. Corresponding formula are shown in Eqs. (1) and (2).

\[ v_e = \sum_{i=1}^{N} \lambda_i v_i \]  
\[ \lambda_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^{N} d_{i0}^{-p}} \]  
\[ \sum_{i=1}^{N} \lambda_i = 1 \]  

where \( v_e \) is the variable value of \((x_e, y_e)\), \( \lambda_i \) is the weight coefficient and \( p \) is the parameter. The best value can be determined by finding the minimum value of the prediction error of the root mean square. The surface to be interpolated should be a surface with local dependent variables. This method assumes that the mapped variable is reduced due to the influence of the distance between the sample location and it is a global interpolation method, that is, all sample points are involved in the estimation of \( Z \) value of a certain unknown point, which is suitable for sample data sets that are uniformly distributed and dense enough to reflect local differences. It is a precise interpolation method, that is, predicted sample point value in the surface generated by interpolation and the measured sample point value is exactly equal.

2.3. Ordinary Kriging

Kriging is a high-level statistical process of generating estimated surfaces through a set of dispersive points with \( Z \) values [8]. To be different from other interpolation methods, the best estimation method used to generate the output surface should be used to conduct a comprehensive study on the spatial behavior of phenomena represented by \( Z \) value, which assumes that the distance or direction between sampling points can reflect the spatial correlation that can be used to illustrate the surface changes. The kriging tool can fit mathematical functions with a specified number of points or all points within a specified radius to determine the output value for each position. The commonly used formula consists of the weighted sum of the data, as shown in formula (3).

\[ \hat{Z}(s0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \]  

where the \( Z(S_i) \) is the measured value of position \( i \), \( \lambda_i \) is the unknown weight of position \( i \), \( s0 \) is an unmeasured position and \( N \) is the measuring index. The weight is not only determined by the distance between the measured points, but also by the predicted position. It also depends on the overall spatial arrangement based on the measured points. The method is usually used in soil science and geology.

2.4. Spline Method

The interpolation method used by the spline function method tool estimates values using mathematical functions that minimize overall surface curvature to produce smooth surfaces that just pass through the
input point [9]. There are two methods of spline function: regular spline function and tension spline function. The rule spline function method uses values that may be outside the scope of the sample data to create a gradual smooth surface. The tension spline function method controls the surface hardness according to the characteristics of the modelling phenomenon. It uses values that are more tightly constrained by the scope of the sample data to create a less smooth surface. The surface interpolation formula is as shown in Eq. (4).

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

where \( N \) is the number of points, \( \lambda_j \) is a coefficient obtained by solving a system of linear equations, and \( r_j \) is the distance from the point \((x, y)\) to the \( j \) point. Depending on the option selected, \( T(x, y) \) and \( R(x, y) \) are different.

For REGULARIZED option, \( T(x, y) \) and \( R(x, y) \) are shown in Eqs. (5) and (6).

\[
T(x, y) = a_1 + a_2x + a_3y
\]

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

where \( r \) is the distance between a point and a sample, \( \tau^2 \) is a weight parameter, \( K_o \) is a modified Bessel function, and \( c \) is a constant of 0.577215. The weight parameter specifies the weight to be attached to the third derivative term during the minimization, and increasing this value will get a smoother surface. Values between 0 and 0.5 are more appropriate. Using the REGULARIZED option ensures a smooth surface and a smooth first derivative surface. This method is useful if you need to compute the second derivative of the interpolation surface.

For TENSION option, \( T(x, y) \) and \( R(x, y) \) are shown in Eqs. (7) and (8).

\[
T(x, y) = a_i
\]

\[
R(r) = \frac{1}{2\pi \varphi^2} \left[ \ln\left( \frac{r\varphi}{2} \right) + c + K_o \left( r\varphi \right) \right]
\]

where \( r \) is the distance between the point and the sample, \( \varphi^2 \) is the weight parameter, \( K_o \) is modified Bessel function, and \( c \) is a constant of 0.577215. The weight parameter specifies the weight that is attached to the first derivative term during the minimized period. When the weight is zero, it will become the interpolation method of the basic sheet spline function interpolation.

2.5. Trend Interpolation

Trend interpolation is a global polynomial interpolation method which can fit the smooth surface defined by the mathematical function (polynomial) and the input sample points [10]. The trend surface will gradually change and capture the coarse-scale model in the data. The model function is shown in Eq. (9).

\[
Z_i = \hat{Z}(x_i, y_i) + \varepsilon_i
\]

where \( Z_i(x_i, y_i) \) is the actual observation data, \( \hat{Z}(x_i, y_i) \) is the trend fitting value, and \( \varepsilon_i \) is the residual value. Polynomial regression analysis is generally used to minimize the sum of squared residuals.

2.6. Test of Interpolation Results

Before generating the final surface, you should know how accurate the model is in predicting the value of the unknown location. In this paper, cross-validation is adopted to determine the quality of the model [11-13]. The objective OK is to have an average error (ME) close to 0, a smaller root mean
square error (RMSE). Finally, the RMSE is taken as the indicator for accuracy evaluation of the four interpolation methods. The formula is shown in Eq. (10).

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (Z_i - \tilde{Z}_i)^2}
\]  

(10)

where, \(Z_i\) is the true value of the test sample point, \(\tilde{Z}_i\) is the predicted value of testing sample points, and \(n\) is the number of samples for testing sample points. The smaller the RMSE value, the higher the accuracy of spatial interpolation prediction, and the smaller the error of spatial interpolation prediction of sample points.

3. Results and Analysis of Spatial Interpolation

3.1. Comparison of Interpolation Results

3.1.1. Comparison of Kriging Interpolation Models. Ordinary kriging interpolation is the most widely used kriging interpolation, it includes many semivariogram types, three models are selected for interpolation in this paper, spherical function, exponential function and Gaussian function, the interpolation results are shown in Fig. 1. The prediction error is shown in Table 1. It can be concluded from the table that the exponential function model has the best interpolation effect, because the RMSE is smallest.

![Figure 1. Comparison of Kriging interpolation models.](image)

(a) Spherical function  (b) Exponential function  (c) Gaussian function

| Model  | ME  | RMSE  |
|--------|-----|-------|
| Fig. 1a | -1.1994 | 178.1152 |
| Fig. 1b | -1.1283 | 176.7084 |
| Fig. 1c | -2.5456 | 181.1867 |
3.1.2. **Comparison of Spline Function Interpolation Models** Regular spline function method and tension spline function method are used for Spline interpolation. Figure 2 shows that tension spline function interpolation is more detailed. The prediction error is shown in Table 2. Based on the root mean square error of the two interpolation methods, the RMSE of tension spline is smallest. Consequently, the interpolation method of tension spline functions is more appropriate.

![Comparison of Spline function interpolation.](image)

**Figure 2.** Comparison of Spline function interpolation.

| Model  | RMSE    |
|--------|---------|
| Fig. 2a | 179.1171 |
| Fig. 2b | 178.8853 |

**Table 2.** Comparison of spline interpolation model

3.2. **Analysis of Interpolation Accuracy**

The final results of the four interpolation methods are shown in Fig. 3. RMSE is selected as the comparison index for the interpolation accuracy. As we know, it is better that the closer the absolute value of the smaller the root mean square error. Table 3 is the cross-validation results of various interpolation methods. From the table, it can be concluded that the root-mean-square error of kriging interpolation method is the minimum, among which the exponential function method interpolation is optimal.

![Interpolation Results](image)

**Figure 3.** Interpolation Results.

| Method  | RMSE    |
|---------|---------|
| IDW     |         |
| Kriging |         |
| Spline  |         |
| Trend   |         |
4. Conclusion
The results of interpolation of slowly available potassium in soil nutrients in Tongzhou district of Beijing were compared by different interpolation methods, ordinary kriging method has higher interpolation precision than the other three methods. Therefore, considering the accuracy of interpolation and the complexity of calculation process, the ordinary Kriging is the best choice for slowly available potassium interpolation. The choice of interpolation method is a balance between data type and computational efficiency, and any method is not absolutely unique. The results generated by different interpolation may be quite different, which is affected by the density of sampling points, the value or variation range of site data, and the complexity of ground liquid. In this paper, only four interpolation methods are selected for comparison, and the selection of other interpolation methods needs further study.

5. Acknowledgements
This work was supported by the Natural Science Research Project of Anhui Provincial Education Department (KJ2018A0009), Anhui Provincial Major Scientific and Technological Special Project (17030701062), Application Research of Anhui Provincial Public Welfare Technology on Linkage Projects (1704f0704059), and the Scientific and Technological Project of Suzhou City (SZ2017GG39).

6. References
[1] Tsegaye, T., Hill, R.L.: Intensive tillage effects on spatial variability of soil test, plant growth, and nutrient uptake measurements. Soil Sci. 163, 155–165 (1998)
[2] Calvaruso, C., Kirchen, G., Saint-André, L., Redon, P.O., Turpault, M.P.: Relationship between soil nutritive resources and the growth and mineral nutrition of a beech (Fagus sylvatica) stand along a soil sequence. Catena. 155, 156–169 (2017)
[3] Steffan, J.J., Brevik, E.C., Burgess, L.C., Cerdà, A.: The effect of soil on human health: an overview. Eur. J. Soil Sci. 69, 159–171 (2018)
[4] Ylöstalo, P., Seppälä, J., Kaitala, S., Maunula, P., Simis, S.: Loadings of dissolved organic matter and nutrients from the Neva River into the Gulf of Finland–Biogeochemical composition and spatial distribution within the salinity gradient. Mar. Chem. 186, 58–71 (2016)
[5] Oliver, M.A., Webster, R.: Kriging: a method of interpolation for geographical information systems. Int. J. Geogr. Inform. Syst. 4, 313–332 (1990)
[6] Luo, W., Taylor, M.C., Parker, S.R.: A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. Int. J. Climatol. 28, 947–959 (2008)
[7] Zhang, G., Rui, X., Fun, Y.: Critical review of methods to estimate PM$_{2.5}$ concentrations within specified research region. ISPRS Int. J. Geo-Inf. 7, 368 (2018).
[8] Arcidiacono, G., Berni, R., Cantone, L., Placidoli, P.: Kriging models for payload distribution optimisation of freight trains. Int. J. Prod. Res. 55, 4878–4890 (2017)
[9] Ajaj, Q.M., Pradhan, B., Noori, A.M., Jebur, M.N.: Spatial monitoring of desertification extent in Western Iraq using Landsat images and GIS. Land Degrad. Dev. 28, 2418–2431 (2017)
[10] Rodríguez-Amigo, M.D.C., Díez-Mediavilla, M., González-Peña, D., Pérez-Burgos, A., Alonso-Tristán, C.: Mathematical interpolation methods for spatial estimation of global horizontal irradiation in Castilla-León, Spain: A case study. Sol. Energy. 151, 14–21 (2017)

[11] Shi, B., Wang, P., Jiang, J., Liu, R.: Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies. Sci. Total Environ. 610, 1390–1399 (2018)

[12] Cheng, X., Xiao, X., Chou, K.C.: pLoc-mPlant: predict subcellular localization of multi-location plant proteins by incorporating the optimal GO information into general PseAAC. Mol. Biosyst. 13, 1722–1727 (2017)

[13] Qiu, W.R., Sun, B.Q., Xiao, X., Xu, Z.C., Jia, J.H., Chou, K.C.: iKcr-PseEns: Identify lysine crotonylation sites in histone proteins with pseudo components and ensemble classifier. Genomics. 110, 239–246 (2018)