Estimation of Unconfined Compressive Strength by Spatial Interpolation Using Non-Geostatistical Methods and Artificial Neural Networks

Thongchai Phothong, Boonsap Witchayangkoon

Department of Civil Engineering, Faculty of Engineering, Thammasat University, THAILAND.

Abstract

This study applies spatial interpolation to estimate soil engineering properties by using previous information in the neighborhood areas. This study focuses on soft clayey Bangkok soil data in the Bangkok Thailand. The non-geostatistic and artificial neural networks (ANN) methods are compared to estimate unconfined compressive strength of soil. The non-geostatistics are inverse distance weighted, triangulation, natural neighbor, b-spline approximation, cubic spline approximation, global thin plate spline, local thin plate spline and thin plate spline. For this study, ANN is the four layers feed forward neural networks with error back-propagation learning. From the computation with the testing data, the cubic spline approximation gives the lowest RMSE. ANN is also applicable with more input data.

1. Introduction

Any construction projects, it is important to know engineering soil properties in the construction area. For both feasibility and construction phases, knowing accurate engineering soil properties require field sample collections via borings which are somewhat costly and time consuming. By this, if one has estimated information of engineering soil properties, it would be of great benefit. Also, the estimated information can be used to crosscheck the real data. By having database engineering soil properties of previous multiple testing, this work therefore tries to apply and test various methods to see feasibility in estimation of engineering soil properties.
2. Literature Review

Suwanwiwattana et al. (2001) constructed geotechnical database system and subsoil model interpretation of Bangkok Clay, Thailand. The SPT soil engineering property was used to represent soil consistency that was interpolated by GRASS's spline module. The accuracy of the model was tested by manually interpreted and the comparison was satisfactory.

Soralump et al. (2010) developed a soil database system for infrastructure development. The soft Bangkok clay was used to be the case study same as Suwanwiwattana et al. (2001). Soralump et al. (2010) was emphasis on a support soil data system. It is not a soil property estimation system. But in the system has an evaluation algorithm for variation of soil properties with in 5x5 square kilometers.

For soil engineering estimation, Al-Ani et al. (2014) used 8 interpolation techniques, IDW by geostatistical analysis, Diffusion, Global Polynomial, Kernel, Ordinary Kriging, Universal Kriging, Spline and IDW to estimate standard penetration test value in Surfers Paradise, Australia. All methods were run and evaluated by observed data with the same position. The IDW by spatial analysis with parameters 2.719e-05 for output cell size, power by 2, search radius fixed and distance 0.25 km was outperformed other methods.

Gangopadhyay et al. (1999) illustrated a powerful performance of a combination tool of ANN and GIS. The ANN used to classify subsurface aquifer characteristics and GIS received that data to estimate depth-averaged aquifer parameters such as transmissivity, leakage factor and storage coefficient. The multilayer perceptron with the back-propagation algorithm was used. The input were location (x, y), depth, z, and extend of particular type material type, z-from and z-to. The output information was the aquifer material present for the input depth zone. The samples were divided to four strata by the variation of sand frequency, each strata was 50 meters.

Zhao et al. (2009) used the ANN to predict high resolution of soil texture map because though field survey is time consuming and expensive. The input of the ANN were coarse resolution and DEM data. The coarse composes of clay map, sand map. The DEM data is soil terrain factor map, soil drainage map, soil deliver ratio map and vertical position map. The relative overall accuracy was 88% for clay content and 81% for sand content.

Some main ideas from Jain et al. (1996), the ANN applications can apply to pattern classification, clustering/categorization, function approximation, prediction/forecasting, optimization, content-addressable memory and control. For nonlinear prediction problems, the
notable solving networks is feed forward network with two hidden layers combines to supervised learning paradigm and back propagation learning algorithm.

This study examines eight non-geostatistical methods and the artificial neural networks to simulate an engineering soil property (unconfined compressive strength) for Bangkok areas, Thailand. The target of the development is that for a certain situation that the researching results can be used to increase confidence of an engineer while decrease cost of soil investigation.

3. Methodology

Bangkok has 1,568.737 km² covering coordinates (1491347, 643245) and (1543301, 709475) (N, E) with about 6 million people (SED, 2012). This research used 74 sites with 155 soil boring data to build a database of soil engineering properties. The coordinates of each boring are pinpointed with Goggle Earth. The depths of the soil bore holes vary from 21 to 79.775 meters. This research consider only unconfined compressive strength (Su) because of testing consistency. The Su at depth 9.25m are selected to evaluate the interpolation methods because of number of data. The 86 samples at the selected depth are divided to two sets. The 64 samples for estimating and 22 for testing as shows in Figure 1. The non-geostatistics is shown by Li and Heap (2008) addressing that non-geostatistics is inverse distance weighted, triangulation, natural neighbor, spline approximation and thin plate spline (TPSP). The foundation equation of spatial interpolation at an unknown point by surrounding points is illustrated by weighted averages:

\[ \bar{z}(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i) \]

where \( \bar{z} \) is the estimated data value of the unknown point \( x_0 \), \( z \) is the observed data value at the known point \( x_i \), \( \lambda_i \) is the weight of known points. The inverse distance weighting (IDW) method uses weighted as an inverse function of the distance, \( d_i \), from unknown point to known points.

\[ \lambda_i = \frac{1/d_i^p}{\sum_{i=1}^{n} 1/d_i^p} \]

where \( P \) is a power parameter set to 2 and \( n \) is number of sampled points. The triangular irregular network (TRI) method estimates the data value of unobserved point from connected surrounding points that forms a series of triangles. The natural neighbors method (NN) combines the benefit of the nearest neighbors and TIN method. The NN creates a Delauney Triangulation of

*Corresponding authors (Boonsap Witchayangkoon) Tel +66-2-5643005 Ext 3101
the known points then weights their values by proportionate area. The splines are series of polynomials used to determine a curve and a surface that represent known points. The b-splines (BSP) are a basic of spline. The spline composes of knots and b-splines. The cubic spline interpolation (CSP) estimates unknown values by a polynomial that continuous though to the second derivative. The thin plate spline (TSP) is surface approximation by splines that try to form curve surfaces. The global thin plate spline (GTSP) is a technic to fit b-spline. The local thin plate spline (LTSP) method is an extension of the TSP. The LTSP uses not more than 10 closest observed points in the estimation processes. The System for Automated Geoscientific Analyses (SAGA) software is used in this study that provides eight non-geostatistics inverse distance weighted, triangulation, natural neighbor, b-spline approximation, cubic spline approximation, global thin plate spline, local thin plate spline and thin plate spline (TIN).

![Image](a) ![Image](b)

**Figure 1:** The ANN model (a) model 1 (b) model 2.

For The Artificial Neural Networks has two model both model are feed forward network. The first model has 50 unit for input layer, 50 unit for hidden layer one and two and one unit for output layer. The 50 input units composes of 16+16 difference in north and east direction between unknown point to known point, 16 units of Su value of known point and last 2 input units are north and east coordinate of unknown point. The output unit is Su of known point. The first model shows in Figure 2(a). The second model includes Su data from soil strata 9.00 meter, 47 samples and 10.5 meter, 43 samples. Both are using for training data set.
The input of the ANN model 2 has 194 units. The 16 samples from 9.00 meter layer each samples has 4 units, DN, DE, DZ and Su that means 54 units. Same as 9.00 meter layer, the input units from 9.25 and 10.5 meter layer are 54 units. Including 2 units, N and E from unknown point

*Corresponding authors (Boonsap Witchayangkoon) Tel +66-2-5643005 Ext 3101 drboonsap@gmail.com ©2013, American Transactions on Engineering & Applied Sciences. Volume 4 No.1 ISSN 2229-1652 eISSN 2229-1660 Online Available at http://TUENGR.COM/ATEAS/V04/049.pdf
The ANN model 2 shows on Figure 2(b). The number 16 samples from each model is the nearest points around the unknown point, Prasomphan and Mase (2013). The 64 samples for estimating will use to be training data and validating data set because of the limit of samples. The learning in ANN is error back-propagation and activation function is logistic. The different between the observed and calculated values of all estimations are assessed by the root mean squared error (RMSE).

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N}}
\]

(3)

where \( e \) is original value subtract estimated value. N is number of points to estimate.

4. Results and Discussion

Figure 2 shows results of non-geostatistic estimations. The b-spline and the local thin plate spline approximation cannot represent a good soil engineering property. The natural neighbor and the triangulation give the same triangle pattern of estimation. The inverse distance weighted, thin plate spline and global thin plate spline gave a smooth estimation but more concentrate on training data set.

The cubic spline and local thin plate spline estimation gave the good distribute and good
results. The graph of interpolation results is illustrated on Figure 3. The original testing data is shown on dash line graph. The different of original values and interpolation values are spike on points 1, 3, 7, 11 and 18. All spike points are around the middle of the area. The RMSE is in Table 1. The best is cubic spline spline estimation, 7.42 that relate to the Figure 3. The best ANN model is ANN model 2.

<table>
<thead>
<tr>
<th>Number</th>
<th>Method</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IDW</td>
<td>11.38</td>
</tr>
<tr>
<td>2</td>
<td>TRI</td>
<td>10.51</td>
</tr>
<tr>
<td>3</td>
<td>NN</td>
<td>10.86</td>
</tr>
<tr>
<td>4</td>
<td>BSP</td>
<td>12.86</td>
</tr>
<tr>
<td>5</td>
<td>CSP</td>
<td>7.42</td>
</tr>
<tr>
<td>6</td>
<td>GTSP</td>
<td>11.25</td>
</tr>
<tr>
<td>7</td>
<td>LTSP</td>
<td>13.81</td>
</tr>
<tr>
<td>8</td>
<td>TPSP</td>
<td>11.28</td>
</tr>
<tr>
<td>9</td>
<td>ANN1</td>
<td>11.46</td>
</tr>
<tr>
<td>10</td>
<td>ANN2</td>
<td>9.42</td>
</tr>
</tbody>
</table>

For the ANN model, the first model (ANN1) uses only data in the same layer. The RMSE is 11.46 it just similar most of non-geostatistic methods. To increase the number of samples to train the ANN model in this test is use data of different layer with significant improvement.

5. Conclusion and Recommendation

With spatial interpolation technique, this study can interpolation of soil unconfined compressive strength of clayey Bangkok soil data in Thailand. The non-geostatistic and artificial neural networks (ANN) methods are compared to estimate unconfined compressive strength of soil. The non-geostatistics are inverse distance weighted, triangulation, natural neighbor, b-spline approximation, cubic spline approximation, global thin plate spline, local thin plate spline and thin plate spline.

This study found possibility that ANN can be used to predict unconfined strength of engineering soil property. Training the ANN model with high number of data is importance. To increase the number of training data, the samples in different layer can be used. After multiple computational experiments with the testing data, the cubic spline approximation gives the best RMSE while ANN is the second best.
6. References

Al-Ani, H., Oh, E., Chai, G. and Al-Uzairy, B., "GIS-Interpolated Geotechnical Zonation Maps in Surfers Paradise, Australia", 2014


Jeff Heaton," Introduction to Neural Networks for Java", 2008


Prasomphan, S. and Mase, S., "Generating Prediction Map for Geostatistical Data Based on an Adaptive Neural Network Using only Nearest Neighbors", 2013


SED., 2012 STATISTICAL PROFILE OF BANGKOK METROPOLITAN ADMINISTRATION", Strategy and Evaluation Department of Bangkok Metropolitan Administration, 2012


Thongchai Phothong earned his Bachelor Degree (Civil Engineering) from King Mongkut's University of Technology Thonburi (KMUTT), Thailand, master degree in Geotechnical Engineering also from KMUTT. He is a PhD candidate at Thammasat University Thailand. He is a lecturer at KMUTT. He is interested in spatial technology and applications.

Dr. Boonsap Witchayangkoon is an Associate Professor of Department of Civil Engineering at Thammasat University. He received his B.Eng. from King Mongkut's University of Technology Thonburi with Honors in 1991. He earned his PhD from University of Maine, USA in Spatial Information Science & Engineering. Dr. Witchayangkoon current interests involve applications of emerging technologies to engineering.

Note: The original of this article has been submitted to 2nd International Workshop on Livable City 2014 (IWLC2014), a Joint Conference with International Conference on Engineering, Innovation, and Technology (EIT), held at Tabung Haji Hotel, Alor Star, Malaysia, during December 9-11, 2014. According to the IWLC2014 Conference Committee, this paper was given Technology Best Paper Award.