Proc. of the 6th Intl. Symposium on Reliability Engineering and Risk Management (6ISRERM)
31 May – 1 June 2018, Singapore
Editor(s) Xudong Qian, Sze Dai Pang, Ghim Ping Raymond Ong, Kok-Kwang Phoon

3D GEOLOGIC MODELING WITH BOREHOLE DATA BY GENERAL REGRESSION NEURAL NETWORK

Zhou W. H.[1*], Zhao L. S.[1], Chen G. M.[1], and Yuen K. V.[1]

[1]*Department of Civil and Environmental Engineering, Faculty of Science and Technology, University of Macau, Macau, China.*
**E-mail: hannahzhou@umac.mo*

It is a challenged task to examine the spatial variation of soil with sparse measured data. In this study, the general regression neural network (GRNN) method was presented to predict the 3D geologic soil profile of a region based on the data of seven boreholes. A probability vector was introduced to represent the soil type at the associated location. By ensuring the difference between the predicted soil type and the measured one less than 1% at these measured boreholes, the smoothing parameter σ in the GRNN method was determined. With the selected smoothing parameter, the regression model was developed using the borehole data and a 3D geologic model was established to present the spatial distribution of soil type. It indicates that the GRNN method can realize a simple and intuitive geologic modeling using only the spatial coordinates.

Keywords: general regression neural network, vector variable of probability, parameter optimization, 3D geologic modeling.

1 Introduction

In geotechnical engineering, the heterogeneity and spatial variability of soil properties usually cause geological uncertainty, which makes a great effect on the performance of geotechnical constructions (e.g., slope, subgrade, and foundation excavation). The spatial variation of soil has attracted many researchers' attention (Tang et al. 1989; Li et al. 2004; Qi et al. 2016). One reason is that it is very important to determine the soil distribution of a certain zone before analyzing the geological uncertainty. The other one is that the understanding of the spatial distribution of soil variability is vital to soil management and environmental research (Kite and Kauwen 1992). Li et al. (2004) carried out a two-dimensional simulation of spatial distribution of soil types using the coupled Markov chain model. Qi et al. (2016) investigated the soil distribution characteristics of the cross sections of a selected area using a proposed practical method based on two-dimensional coupled Markov chain and pointed out that the predicted soil distribution is related to the sampling interval.

In contrast to the two-dimensional modeling of soil distribution, 3D geologic modeling can specifically and entirely reflect the spatial distribution of soil properties. However, the sparse survey data at some regions make it difficult to develop a 3D geologic model. Among the existing prediction methods, the neural network technique is widely used to solve complex nonlinear problems (Hubick, 1992). Especially, as a one-pass learning algorithm with a highly parallel structure (Specht 1991), the general regression neural network (GRNN) method has been applied to deal with some issues related to regression, prediction and classification in geotechnical engineering due to its several advantages, such as high accuracy, no requirement of backpropagation, and handling noises in inputs. Additionally, GRNN method can respond much better to many types of problems (Salem and Zahaby 2010). Pradeep et al. (2006) built the GRNN model using CPT data to predict the soil composition and determine soil type. The feasibility of using GRNN method was validated to estimate the soil type.

In this study, based on the data of seven boreholes in a region, a regression model was developed using the GRNN method to describe the relationship between the soil type and the three-dimensional coordinate at the associated point. The soil types at unknown zone in this region were predicted using the obtained regression model. Finally, a 3D geologic model was developed based on the predicted results to reveal the spatial distribution of soil type in the selected region.

2 Working Principle of GRNN

As one type of memory-based network, GRNN is a one-pass learning algorithm with a highly parallel structure, and it only needs a fraction of training data. GRNN involves with Gaussian functions, which enables it to have high accuracy. Each training sample, X_i, is used as the mean of a Gaussian distribution. The estimated $\hat{Y}(X)$ can be visualized as a weighted average of all of the observed values, Y_i, where each observed value is weighted exponentially according to its Euclidean distance from X (Specht 1991):

$$\hat{Y}(X)=\frac{\sum_{i=1}^{n} Y_i \exp(\frac{-D_i^2}{2\sigma^2})}{\sum_{i=1}^{n} \exp(\frac{-D_i^2}{2\sigma^2})} \tag{1}$$

where $D_i^2=(X-X_i)^T(X-X_i)$. Y and $\hat{Y}(X)$ can be vector variables or scalars. σ is the standard deviation or the smoothing parameter as shown in Specht (1991).

3 Probability Combined with GRNN Model

In this study, the input variable, X, represents a series of three-dimensional coordinates at the boreholes, i.e., $[(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_n, y_n, z_n)]$, where n is the number of these coordinates. Y is denoted as the soil type with respect to the corresponding coordinate at the boreholes. Because three types of soils (sand, silt, and clay) were observed at these boreholes, each component of Y is written into a 1x3 vector, $Y_i = [Y_{i1}\ Y_{i2}\ Y_{i3}]$, to consider the possibility of each of these soil types, where $i = 1, 2, \ldots, n$. Herein, each element of Y_i is denoted as Y_{im} ($m = 1, 2, 3$). The value of Y_{im} means the probability of a certain soil type. For example, $Y_i = [0.6, 0.2, 0.2]$ means that the probabilities of soil type at the corresponding location for sand, silt, and clay are 0.6, 0.2, and 0.2, respectively. The predicted soil type, $\hat{Y}_i$, after using GRNN method can be written as follows:

$$Y_i=\begin{bmatrix}Y_{i1} & Y_{i2} & Y_{i3}\end{bmatrix}\xrightarrow{GRNN}\hat{Y}_i=\begin{bmatrix}\hat{Y}_{i1} & \hat{Y}_{i2} & \hat{Y}_{i3}\end{bmatrix} \tag{2}$$

It should be noted that the summation of three components for an arbitrary Y_i is equal to 1. The summation of that for $\hat{Y}_i$ after using GRNN method is shown below:

$$\begin{aligned}\sum_{m=1}^{3}\hat{Y}_{im} &= \frac{\sum_{i=1}^{n} Y_{i1}\cdot w_i}{\sum_{i=1}^{n} w_i}+\frac{\sum_{i=1}^{n} Y_{i2}\cdot w_i}{\sum_{i=1}^{n} w_i}+\frac{\sum_{i=1}^{n} Y_{i3}\cdot w_i}{\sum_{i=1}^{n} w_i}\\ &= \frac{\sum_{i=1}^{n}\left(Y_{i1}+Y_{i2}+Y_{i3}\right)\cdot w_i}{\sum_{i=1}^{n} w_i}=1\end{aligned} \tag{3}$$

where $w_i = \exp\left(-D_i^2 / 2\sigma\right)$. It means that GRNN method does not change the summation of the probabilities for the three soil types. In this study, the maximum probability in the updated $\hat{Y}_i$ determines the soil type. For instance, Y_i = [0.7, 0.1, 0.2] means that the soil type should be sand. For convenience in the following interpretation, this maximum probability is defined as decision probability.

4 3D Geologic Modeling

4.1 *Borehole Data*

The borehole data adopted in the following case study comes from Australian Geomechanics Society and the Institution of Engineers, Australia (Qi et al. 2016). Seven boreholes are scattered within a 72 m × 40 m area, as shown in Figure 1(a). The boreholes from left to right (boreholes with original number of D15, D16, D14, D74, D73, D18, and D17) were re-labeled as borehole 1, 2, 3, 4, 5, 6, and 7 respectively. Figure 1(b) demonstrates the stratigraphy of the boreholes by projecting these boreholes on the *x-z* plane. And three types of soils, i.e., sand, silt, and clay were observed at this region. In this study, the coordinates of the sampling points were developed by dividing a continuous borehole data into a series of discrete points. To balance the accuracy and the computational efficiency, the depth interval for these boreholes was selected as 0.4 m. The coordinate values were normalized using their mean and standard deviation in each direction for diminishing the variance of the coordinate value among the three different directions.

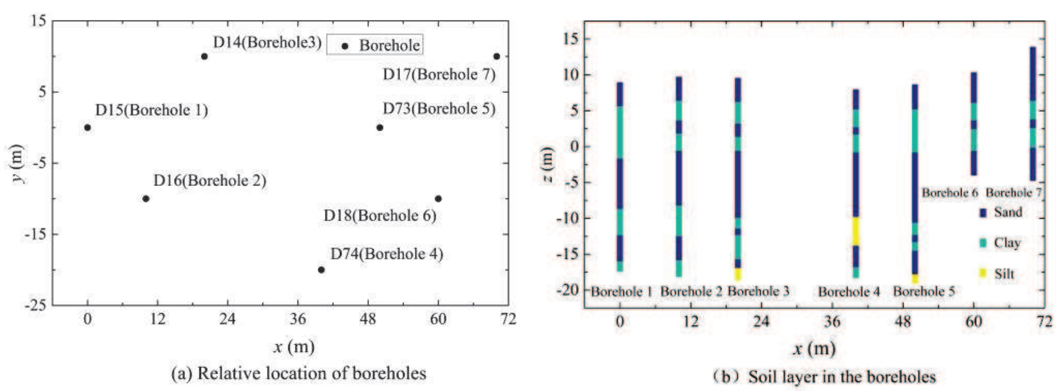

Figure 1. The relative location and stratigraphy of the boreholes in Perth city, Australia

4.2 *Determining smoothing parameter σ*

According to the GRNN method, a larger smoothing parameter results in a smoother regression surface while a smaller smoothing parameter produces a better fitness to the sampling points. It is expected that the predictions at these boreholes have no much difference from the measured data as well as a smoother regression surface. Hence, the smoothing parameter was examined starting from a relatively small value. Then, it expanded gradually until the difference between the predicted soil type and the measured one is less than 1%. It was determined that σ is 0.11.

Based on the optimal value of σ, the predicted soil types at the boreholes are presented in Figure 2 using GRNN model. In this prediction, the soil types of the target borehole were set as testing data while that of the other six boreholes were set as training data. The differences between the predicted soil types and the measured ones at borehole 1, 2, 3, 4, 5, 6, 7 are 15.6%, 23.5%, 30.8%, 26.2%, 29.4%, 20.6%, and 4.4%, respectively. The average of these differences is 21.5%. The testing reveals that the GRNN method can get an acceptable prediction based on the existing data. It should be noted that more boreholes need to be drilled around the borehole 3 and 5 because of the great spational variation of soil. In the present study, only 7 borehole data are available. It can present the soil type distribution roughly with some degree of accuracy. The balance between the accuracy and the cost (borehole numbers) should be considered in real practice.

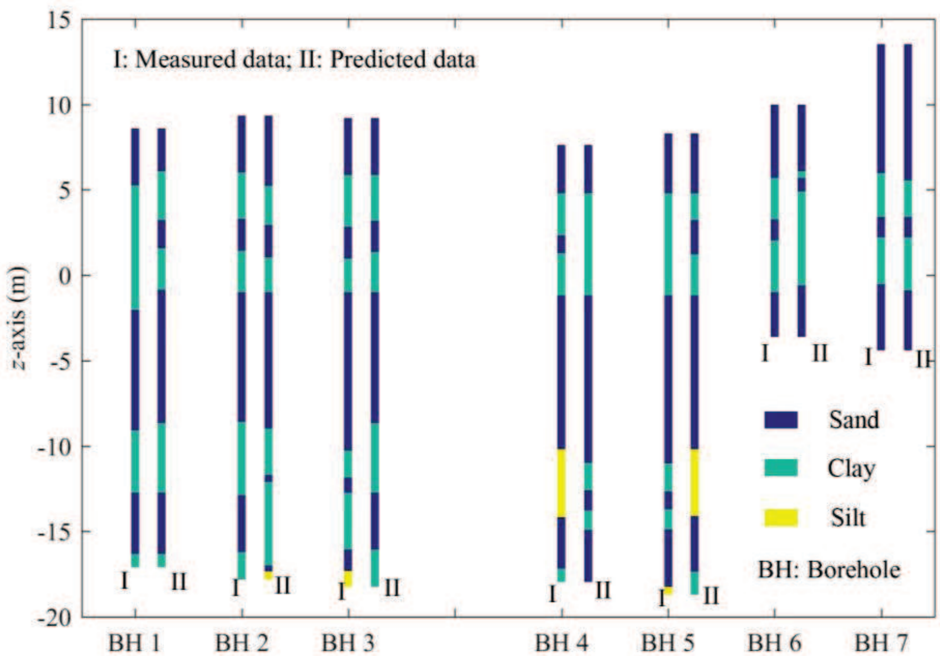

Figure 2. Predicted soil types at different borehole locations

4.3 *3D Geologic Model*

Based on the determined smoothing parameter, σ, a regression model was generated using GRNN method associated with the data of seven boreholes. The three-dimensional coordinates of the whole region were obtained through meshing the domain in x direction of (0~70) m, in y direction of (-25~15) m, and in z direction of (-20~15) m with the meshing interval of 0.4 m. The soil type at the corresponding location can be obtained using the trained regression model based on the definition of decision probability. After that, a 3D stratification was presented based on the predicted soil types at different locations as shown in Figure 3. This prediction can reveal the spatial distribution of the three soil types based on the information of these seven boreholes. However, due to the uniform smoothing parameter adopted in all the three dimensions, the sudden transition at the boundaries between different soil types occurs as shown in Figure 3.

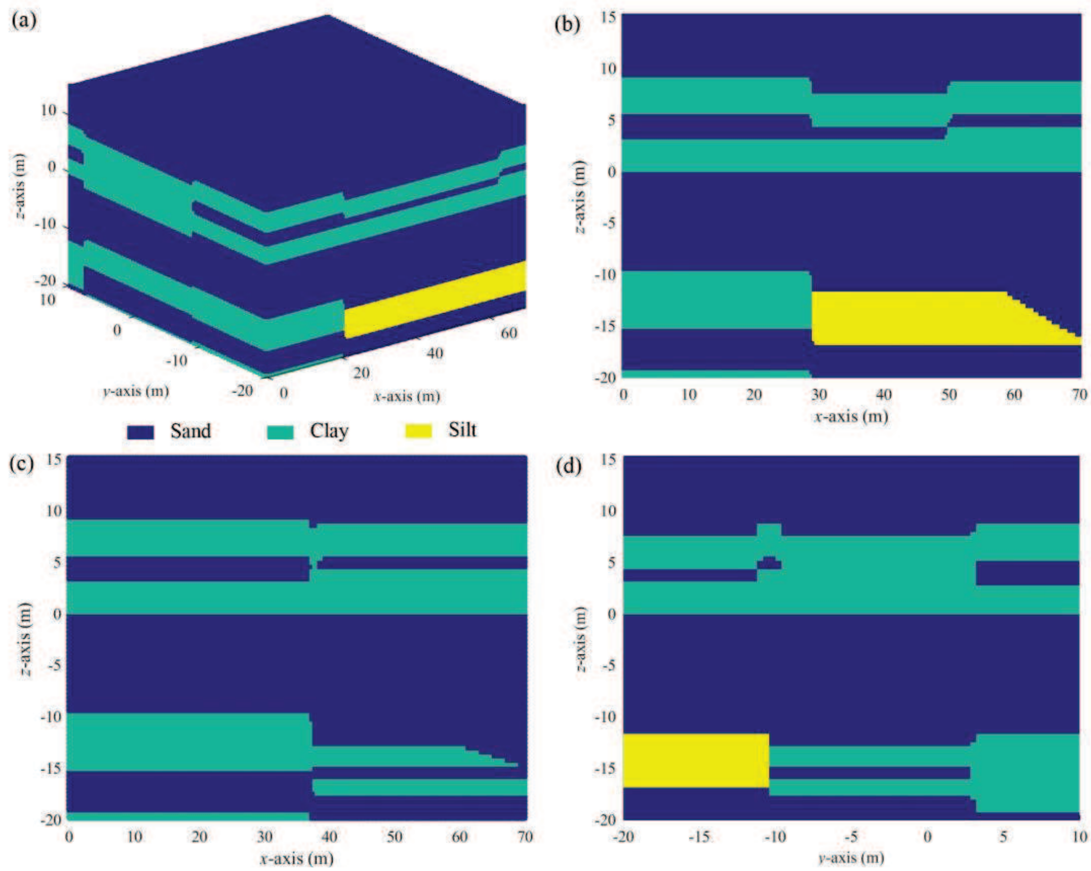

Figure 3. Geologic model: (a) 3D view; (b) x-z plane with y = -15 m; (c) x-z plane with y = -10 m; (d) x-z plane with x = 40 m

5 Conclusions

In this study, the GRNN method was adopted to predict the 3D stratification of a region. The three-dimensional coordinates at different locations were designed as the inputs while the probability vectors with respect to the soil types were taken as the outputs. The smoothing parameter σ in GRNN method was determined as 0.11 based on the meshing interval of 0.4 m. The GRNN method was verified to be reasonable to predict the spatial distribution of soil type. A regression model was developed using the measured seven borehole data at the region. Then, the soil types at the region were predicted using the proposed regression model. Finally, a 3D geologic model was built to reflect the overall soil distribution of the region. The prediction can present the spatial variation of soil at this region.

Acknowledgments

The authors gratefully acknowledge the financial supports from the Macau Science and Technology Development Fund (FDCT) (125/2014/A3) and University of Macau Research Fund (MYRG2015-00112-FST and MYRG2017-00196-FST).

References

Hubick, K. T., Artificial Neural Networks in Australia, Department of Industry, *Technology and Commerce*, Commonwealth of Australia, Canberra, 1992.

Kite, G. W. and Kauwen, N., Watershed Modeling Using Land Classification, *Water Resour. Res.,* 28, 3193–3200, 1992.

Li, W., Zhang, C., Burt, J. E., Zhu, A. X. and Feyen, J., Two-Dimensional Markov Chain Simulation of Soil Type Spatial Distribution, *Soil Sci. Soc. Am. J.*, 68 (5), 1479–1490, 2004.

Pradeep, U., Kurup, P. E. and Griffin, E. P., Prediction of Soil Composition from CPT Data Using General Regression Neural Network, *J. Comput. Civ. Eng.*, 20(4), 281-289, 2006.

Qi, X. H., Li, D. Q., Phoon, K. K., Cao, Z. J. and Tang, X. S., Simulation of Geologic Uncertainty Using Coupled Markov Chain, *Eng. Geol.*, 207, 129–140, 2016.

Specht, D. F., A General Regression Neural Network Neural Networks, *IEEE Trans*, 2, 568–576, DOI: 10.1109/72.97934, 1991.

Salem, S. S. and Zahaby, K. E., Application of General Regression Neural Networks (GRNNs) in Assessing Liquefaction Susceptibility, Proceedings of the 5th International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics, San Diego, California, May 24–29, 2010.

Tang, W. H., Sidi I. and Gilbert R. B., Average Property in Random Two-State Medium, *J. Eng. Mech.*, 115(1), 131–44, 1989.