Data-driven modelling for pore water pressure variation responses to rainfall

M. R. Mustafa[1], M. H. Isa[1], R. B. Rezaur[2] & H. Rahardjo[3]
[1]Department of Civil Engineering,
Universiti Teknologi PETRONAS, Malaysia
[2]Golder Associates Ltd., Calgary, Canada
[3]School of Civil and Environmental Engineering,
Nanyang Technological University, Singapore

Abstract

Measurement of soil pore water pressure is always a tedious, time consuming and expensive exercise. Moreover, unavailability of any physical based or mathematical relationship to get information of pore water pressure leads researchers to perform data-driven modelling. This study presents a data-driven modelling approach to predict soil pore water pressure variations in a slope. Point measurements based time series data of soil pore water pressure variations and corresponding rainfall was used to develop the data-driven model. The model was developed using radial basis function neural network with Multi-quadric basis function. The inputs of the model consist of 5 antecedent pore water pressure, two antecedent rainfall and one current rainfall values. Trial and error procedure was adopted to obtain the appropriate number of neurons in the hidden layer. Normalization method was used to determine the spread of the basis function. Mean absolute error (MAE) and coefficient of determination (R^2) as statistical measures were used to evaluate the performance of the model. The results revealed that the data-driven model predicted the pore water pressure values close to the observed values. The minimum value of MAE during test stage was observed as 0.327 with a coefficient of determination $R^2 = 0.975$. Multi-quadric basis function was found to be suitable for the prediction of soil pore water pressure variations.
Keywords: neural network, pore-water pressure, prediction, radial basis function, rainfall, slope failure.

1   Introduction

Tropical climate regions like Singapore and Malaysia are always exposed to heavy rainfalls and humid environment due to their topographical location in the globe. Consequently, increase of pore water pressure may reach positive values that are highly undesirable in residual soil slopes. Therefore, due to these positive or high pore water pressure values, the regions are always prone to slope failures. Hence, it is of prime importance to get knowledge of pore water pressure fluctuations for any slope stability analysis. However, the measurement of soil pore water pressure requires extensive resources in terms of space, time and cost. Additionally, due to unavailability of any physical based or mathematical relationship between pore water pressure and the related parameters, there is a demand to establish data-driven models to collect information of pore water pressure variations. Since the soil pore water pressure depends on several climatic factors (i.e. rainfall, humidity, temperature, evaporation etc.) as well as soil characteristics (i.e. moisture content, hydraulic conductivity, porosity etc.), it is challenging to incorporate all these factors in any data-driven model. However, few studies have been conducted to predict soil pore water pressure variations in a slope using some data-driven techniques [1–3].

Variations of pore water pressure in a residual soil slope using radial basis function neural network were explored by Mustafa *et al.* [1]. They investigated the input structure of the model by an auto-correlation and cross correlation analysis between the rainfall and pore water pressure data. Gaussian basis function was used in the hidden layer to capture the non-linear behaviour of pore water pressure. They used several statistical measures to evaluate the model performance and concluded that the application of Gaussian basis function is suitable for prediction of pore water pressure variations. Mustafa *et al.* [2] investigated the effect of antecedent pore water pressure values for prediction of pore water pressure variations in a slope. The data-driven model was developed using scaled conjugate gradient training algorithm in multilayer perceptron neural net-work. They suggested that the addition of antecedent pore water pressure values in the input structure of data-driven model is highly important and have a significant contribution for accurate predictions. A comparative analysis between different training algorithms in multilayer perceptron neural networks for prediction of pore water pressure was performed by Mustafa *et al.* [3]. Performance of four different training algorithms was evaluated to identify an appropriate multilayer perceptron training algorithm for modelling pore water pressure variations in a residual soil slope. They suggested that Levenberg Marquardt training algorithm could be preferred over other algorithms due to its fast convergence, self-selection ability of learning rate and accuracy in predicted values.

Although a few studies have been conducted for prediction of pore water pressure variations in a soil slope, it appears that application of Multi-quadric basis function has not been explored yet. Application of Multi-quadric basis function neural network for prediction of pore water pressure variations may explore an alternative method to the existing techniques. Therefore, the objective of this study is to develop a data-driven model for prediction of soil pore water pressure

variations in a residual soil slope using Multi-quadric basis function neural network.

2   Site description and data statistics

Thirteen month time series data of rainfall and pore water pressure was acquired at a soil slope in Yishun, Singapore. The field investigation was mainly conducted for understanding the infiltration effects on residual soil slope stability analysis and slope failure phenomenon due to excess rainfall [4, 5]. However, these data have been partly used in this study to conduct data-driven modelling for pore water pressure variations in a slope. The climate at the site is hot and humid throughout the year with mean relative humidity of about 84% and an average temperature of 26.6°C. The slope area is plain, and 100% protected with vegetation cover. The length of slope is 16.5 m, height 7 m and slope angle of 23° [6]. The pore water pressure data used in this study was recorded at soil depth of 0.5 m. The minimum and maximum recorded values of rainfall are 0 and 44.75 mm, whereas, those of pore water pressure are -55.90 and 5.54 kPa respectively.

3   Data-driven modelling

Data-driven modelling is one of the emerging techniques which provides a functional representation between the variables comprising complex and non-linear relationships. Data-driven approach analyse the data without assuming any underlying mechanism between the variables. In recent years, data-driven techniques have achieved significant importance due to their learning ability during training stage and have been successfully applied in multiple disciplines including water resources engineering [1, 7–11]. Perhaps, artificial neural network is one of the most popular data-driven techniques which has been broadly used in modelling hydrological variables. Radial basis function is one of the commonly used neural network types. Thorough literature on radial basis functions and its different types including Multi-quadric basis function was discussed by Bishop, [12]. Development of data-driven model using Multi-quadric basis function neural networks has been explained in the subsequent sections.

3.1 Model structure

General architecture of radial basis function neural networks consists of three layers named as input, hidden and output layers. The input layer consists of the number of inputs required for the model construction that may be equal to the number of input variables or includes some antecedent conditions. It is important to include some antecedent rainfall and pore water pressure values in the input structure of a data-driven model developed for the prediction of pore water pressure variations [2]. Therefore, in this study the input structure contains five antecedent pore water pressure values, two antecedent rainfalls and one current rainfall value. Previously some studies used similar input structure and suggested

it to be appropriate for data-driven modelling [1–3]. Therefore, the input layer of the model consists of eight numbers of input neurons.

The hidden layer is equipped with basis functions, and the number of hidden neurons is equal to the number of basis functions. In this study, Multi-quadric basis function has been used in the hidden layer. Appropriate number of hidden neurons and spread of the basis function was determined by trial and error procedure. Multiple options were examined by fixing either the spread or the number of the basis during the investigation of appropriate values. The appropriate number of hidden neurons was found to be 10 with best spread value of 3.0. The centers of the basis functions were randomly selected from the input data structure. Since the output expected from the model was only current pore water pressure value, therefore, the number of output neuron was limited to one. Thus, the structure of the model may be defined as 8-10-1 (eight input neurons – ten hidden neurons – one output neuron).

3.2 Data division for training and testing

Since the data-driven approach mainly consists of two stages *viz.*, training and testing stage, it is foremost important to select an appropriate data set for efficient training of the model. To avoid any discrepancy in the result during the testing stage, the minimum and maximum values of the whole data set should be included during the training stage. Furthermore, the training data set should be a true representative of the complete variations involved in the time series. Based on these recommendations, the total data was divided into two sets i.e. 75% of the data was selected for the training stage to accommodate the maximum variations involved in the time series and the remaining 25% was used for the testing stage of the model.

3.3 Pre-processing of the data

In data-driven modelling, particularly while using radial basis function neural network technique, it is recommended to normalize the data according to the range of the basis function to ensure fast convergence of the model [1, 3, 13]. Pre-processing or data normalization is advantageous to minimize the global error of the network. Since the time series data of pore water pressure contains negative as well as positive values, the data was scaled from -1 to 1.

3.4 Model evaluation

Coefficient of determination (R^2) is one of the broadly used performance evaluation measures in data-driven modelling techniques. Plot of observed versus predicted data with the perfect line of agreement provides a true comparison and clearly describes the outliers of the predicted data. Therefore, the performance of the model was evaluated based on the value of the coefficient of determination and on plotting the data against the perfect line of agreement. Additionally, to provide digital explanation about the model error, the model was also evaluated using mean absolute error (MAE).

4   Results and discussion

Fluctuations of soil pore water pressure have been modelled successfully by a data-driven technique called radial basis function neural network using Multi-quadric basis function. It was observed that the data-driven model with network configuration 8-10-1 was able to capture the pore water pressure behaviour. A comparison between the time series of observed and predicted soil pore water pressure data during the testing stage of the modelling is shown in Figure 1. The pore water pressure increases due to rainfall events and decreases in the drought period. A comparison of observed and predicted values showed that the predicted values followed the pattern of the observed data. The precision between observed and predicted data during testing stage verified that the model had learned the nonlinear behaviour between the inputs and targets during the training stage. The model successfully predicted the data during both dry and wet periods. Two main drought periods (15th to 26th July and 30th August to 18th September) were observed in the testing dataset. During these droughts, the pore water pressure decreased significantly and reached the extreme negative values. The model was able to capture both drought events.
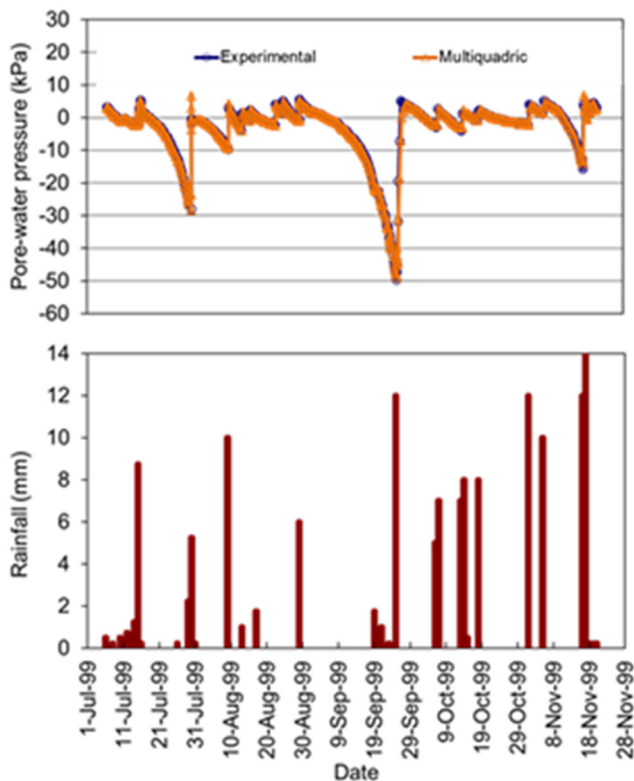

Figure 1:   Time series of rainfall, predicted and observed pore water pressure.

Comparison of pore water pressure data (observed and predicted) with perfect line of agreement at training and testing stages of the model are shown in Figure 2. It is shown that the observed and predicted data showed nearly perfect agreement and produced coefficient of determination close to one ($R^2 = 0.993$) during training stage of the network (Figure 2(a)). Almost all the data lies on the perfect line of agreement. This depicts the effective learning of the model during training stage. A similar trend was observed when the testing data set was plotted with the perfect line of agreement. The predicted values showed a good agreement with the observed pore water pressure data. The coefficient of determination is also close to one ($R^2 = 0.975$ as shown in Figure 2(b)). However, there are few outliers which are very common to be expected in any data-driven modelling. Overall, the nearly perfect agreement and well-synchronized behaviour between the observed and predicted data suggest the suitability of Multi-quadric basis function for modelling pore water pressure variations.
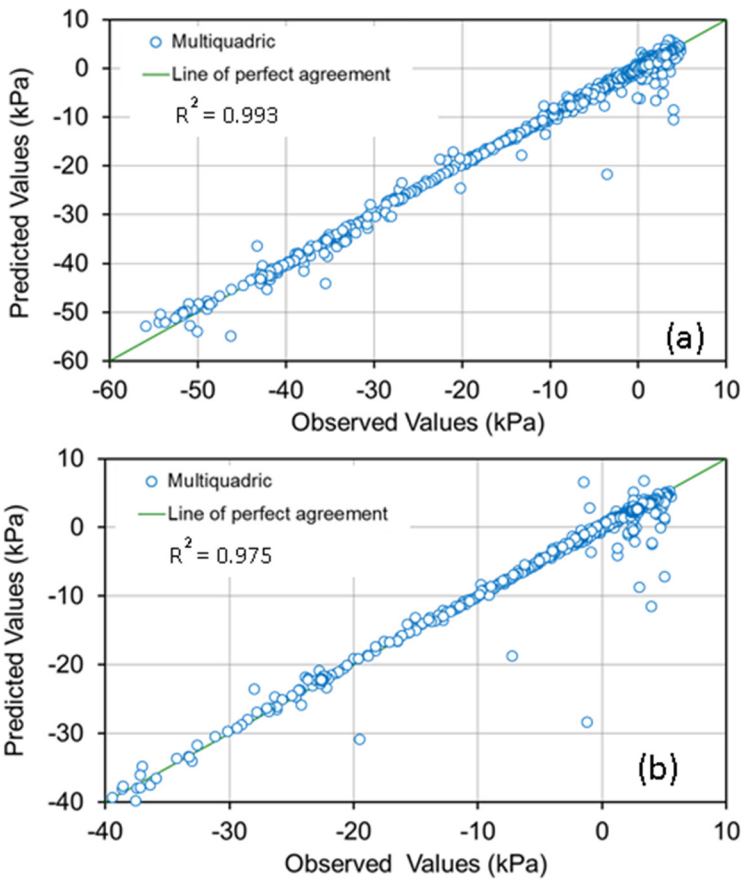

Figure 2: Comparison of predicted and observed pore water pressure with perfect line of agreement. (a) Training; (b) Testing.

A summary of the statistical performance analysis performed for the model evaluation is presented in Table 1. The mean absolute error during training stage (MAE = 0.3270) is comparatively less than the testing stage (MAE = 0.4459). Similar trend was observed in previous studies related to data-driven prediction modelling [1, 2, 11, 14]. Since during training stage both input and targets are presented to the network but in test stage only the inputs are provided to the network, training performance could be expected to be better than the testing stage. Additionally, the amount of data used in the training stage is larger than the testing stage.

Table 1:   Summary of performance statistics of the model.

Criteria	Training	Testing
MAE	0.3270	0.4459
R^2	0.9930	0.9750

Only two types of neural networks were observed to in literature for successful prediction of pore water pressure. Mustafa *et al.* [2, 3] developed prediction models using Multilayer Perceptron (MLP) neural network using different training algorithms. While, Mustafa *et al.* [1, 15] predicted pore water pressure using Radial Basis Function (RBF) with Thin Plate Spline and Gaussian basis functions respectively. A comparison between the current model with previous studies on different data-driven models showed that Multi-quadric basis function model produced nearly similar performance in terms of R^2 (Tables 1 and 2).

Table 2:   Performance statistics of different data-driven models for prediction of pore water pressure.

Reference	Year	Type/Function	Testing R^2
Mustafa *et al.* [15]	2014	Thin Plate Spline	0.978
Mustafa *et al.* [3]	2013	Levenberg-Marquardt	0.983
Mustafa *et al.* [1]	2012b	Gaussian	0.971
Mustafa *et al.* [2]	2012d	Scaled Conjugate Gradient	0.973

Obviously, all the results indicated the suitability of Multi-quadric basis function for modelling pore water pressure fluctuations in a soil slope due to rainfall events. The comparison of the results with previously employed data-driven techniques recommended the applicability of the Multi-quadric basis function to predict pore water pressure behaviour. Particularly, the wide range of pore water pressure variations of the data set which covered both dry and wet period was nicely captured by the model. Apart from the few outliers, all the predicted values had excellent agreement with the observed time series data.

5   Conclusions

This study presents a successful application of Multi-quadric basis function neural network for prediction of time series of soil pore water pressure variations in a residual soil slope. The network configuration with some antecedent rainfall and pore water pressure with structure of 8-10-1 (input-hidden-output neurons) was appropriate to recognise the nonlinear pattern of pore water pressure variations. The results showed that the predicted time series of pore water pressure were very close to the observed ones. Furthermore, the application of Multi-quadric basis function for prediction of time series of pore water pressure is suitable and competitive with other data-driven techniques. Therefore, Multi-quadric basis function can be recommended as an alternative to the previously developed pore water pressure data-driven models.

Acknowledgement

References

[1]   Mustafa, M.R., Rezaur, R.B., Rahardjo, H., Isa, M.H., 2012b. Prediction of pore-water pressure using radial basis function neural network. Engineering Geology 135-136, pp. 40-47.

[2]   Mustafa, M.R.U., Bhuiyan, R.R., Isa, M.H., Saiedi, S., Rahardjo, H., 2012d. Effect of Antecedent Conditions on Prediction of Pore-Water Pressure Using Artificial Neural Networks. Modern Applied Science 6, pp. 6-15.

[3]   Mustafa, M.R., Rezaur, R.B., Saiedi, S., Rahardjo, H., Isa, M.H., 2013. Evaluation of MLP-ANN Training Algorithms for Modeling Soil Pore-Water Pressure Responses to Rainfall. Journal of Hydrologic Engineering 18, pp. 50-57.

[4]   Rahardjo, H., 2000. Rainfall-induced Slope Failures. Research Report-NSTB 17/6/16, Nanyang Technological University, Singapore, 407.

[5]   Rezaur, R.B., Rahardjo, H., Leong, E.C., 2002. Spatial and temporal variability of pore-water pressures in residual soil slopes in a tropical climate. Earth Surface Processes and Landforms 27, pp. 317-338.

[6]   Rahardjo, H., Leong, E.C., Rezaur, R.B., 2008. Effect of antecedent rainfall on pore-water pressure distribution characteristics in residual soil slopes under tropical rainfall. Hydrological Processes 22, pp. 506-523.

[7]   Nastos, P.T., Paliatsos, A.G., Koukouletsos, K.V., Larissi, I.K., Moustris, K.P., 2014. Artificial neural networks modeling for forecasting the maximum daily total precipitation at Athens, Greece. Atmospheric Research 144, pp. 141-150.

[8]   He, Zhibin, Wen, Xiaohu, Liu, Hu, Du, Jun. 2014. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support

vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology 509, pp. 379-386.

[9]   Lohani, Anil Kumar, Goel, N.K., Bhatia, K.K.S., 2014. Improving real time flood forecasting using fuzzy inference system. Journal of Hydrology 509, pp. 25-41.

[10] Liu, Qian-Jin, Shi, Zhi-Hua, Fang, Nu-Fang, Zhu, Hua-De, Ai, Lei, 2013. Modeling the daily suspended sediment concentration in a hyper-concentrated river on the Loess Plateau, China, using the Wavelet–ANN approach. Geomorphology 186, pp. 181-190.

[11] Mustafa, M.R., Rezaur, R.B., Saiedi, S., Isa, M.H., 2012c. River Suspended Sediment Prediction Using Various Multilayer Perceptron Neural Network Training Algorithms – A Case Study in Malaysia. Water Resources Management 26, pp. 1879-1897.

[12] Bishop, Christopher M. 1995. Neural Networks for Pattern Recognition. Oxford University Press, 482.

[13] Rojas, R., 1996. Neural Networks: A Systematic Introduction. Springer Verlag, Berlin, pp. 151-184.

[14] Kisi, O., 2011. Wavelet Regression Model as an Alternative to Neural Networks for River Stage Forecasting. Water Resources Management 25, pp. 579-600.

[15] Mustafa, M.R., Rezaur, R.B., Isa, M.H., Rahardjo, H., 2014. Estimation of soil pore-water pressure variations using a thin plate spline basis function. High performance and optimum design of structures and materials, WIT press 137, pp. 615-623.