Prediction of Soil Pore Water Pressure Using Polynomial Kernel Support Vector Machine

by

Muhamad Alif Bin Muhd Hasbi

18903

Dissertation submitted in partial fulfilment of
the requirement for the
Bachelor of Engineering (Hons)
Civil Engineering

September 2017

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
32610 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Prediction of Soil Pore Water Pressure Using Polynomial Kernel Support Vector Machine

by

Muhammad Alif Bin Muhd Hasbi
18903

A project dissertation submitted to the
Civil Engineering programme
Universiti Teknologi PETRONAS
In partial fulfilment of the requirement for
BACHELOR OF ENGINEERING (HONS.)
(CIVIL)

Approved by,

____

(DR MUHAMMAD RAZA UL MUSTAFA)

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
32610, Tronoh
Perak Darul Ridzuan
September 2017

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

____

(MUHAMMAD ALIF BIN MUHD HASBI)

ABSTRACT

Knowledge in soil pore water pressure is important in slope stability analysis. The fluctuation of soil pore water pressure is mainly affect by the rainfall intensity. The pore water pressure can cause effect on the soil strength. A few problems has been identified in obtaining the pore water pressure reading which are time consuming and intensive labor work that are from field instrumentation. However, the development of soft-computing in nowadays has become focus as alternative technology to monitor the changes of pore water pressure in soil, and scope in this study is to use support vector machine. Then, a study on developing a model to predict the soil pore water using polynomial kernel support vector machine and to evaluate the model performance has been conducted. A considerable good correlation between observed and predicted pore water pressure has been found by the end of the studies with performance evaluation of R^2 and RMSE of 0.93 and 0.60 respectively A slope in Universiti Teknologi PETRONAS, Perak has been chosen as study scope to predict the soil pore water pressure.

ACKNOWLEDGEMENT

Bismillahhirahmanirrahim,

First of all a great gratitude to The Almighty, upon His blessing, because I am able to complete this dissertation after two semester of conducting Final Year Project. I also would like to express my gratitude towards Universiti Teknologi PETRONAS for the giving opportunities for me to have exposure in doing the research. Moreover, I considered as a lucky individual to have unlimited moral supports and advices from parents and family members. Their encouragement and motivation has been one of the key that keep strengthen my spirit to endure this period of completing the project. Words are inadequate to express my deepest gratitude to my project supervisor, Dr Muhammad Raza Ul Mustafa, for endless support in ensuring this project runs smoothly, for both technical and non-technical sector. Lastly, sincere thanks are presented to all parties that are directly or indirectly involved in this research project.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

UTP – Universiti Teknologi PETRONAS

PWP – Pore Water Pressure

RMSE – Root Mean Square Error

R^2 – Coefficient of determination

RBF – Radial basis function

SM – Sample mean

SV – Sample variance

SD – Sample deviation

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Study of pore water pressure of soil is one of key factor in slope stability analysis apart of climatic changes, imposed loadings and others. Pore water pressure is the pressure which inhibit inside soil acting either suction or exerting outward of soil. Rainfall infiltration is the main factor which will cause the fluctuation of the pore water pressure. The pore water pressure is usually can be obtained from slope instrumentation which create problems of both time consuming and intensive labour work. Thus, alternative method to counter the problems is desired and the ability to predict the upcoming occurrence due to pore water pressure due to rainfall is vital process so a prevention step can be introduced.

It is found that application of soft computing to predict the pore water pressure is a promising alternative. Ibrahim D, (2016) has claimed that soft computing is tolerant of imprecision, uncertainty, partial truth, and approximations. It is based on techniques such as fuzzy logic, genetic algorithms, artificial neural networks, machine learning, and expert systems. Since prediction of pore water pressure is required the machine learning based, then support vector machine (SVM) has been chosen due to its ability to perform regression and classification function. SVM role is to minimize the generalization error in predicting the pore water pressure. Aided with specialty SVM to utilize the kernel trick, it has become main advantage of SVM since kernel trick is able to minimize model prediction complexity and prediction error (Sujay & Paresh, 2013)

The study is conducted on a slope within Universiti Teknologi PETRONAS, more specific is in front of Block 5, Chemical Engineering department block. The field is chosen because of the existing field is already instrumented with tensiometer and rain gauge which will be used to collect the pore water pressure (PWP) and gauge rainfall intensity respectively.

TABLE 1. Slope Characteristics

Slope Properties	
Area (m^2)	550
Length (m)	20
Height (m)	11
Angle (^{o})	33
Cover	Turfed

1.2 PROBLEM STATEMENT

Knowledge of pore water pressure is significant in slope stability study. It has become more concern during the rainfall period since the rainfall infiltration is affecting the pore water pressure to rise. This is very undesirable condition in slope stability study. However, the only method to obtain the data of pore water pressure is by setting up the instrumentation on the slope. This is difficult to be done since setting up the instrumentation is time consuming and intensive labour work (Mustafa et al, 2012). It is a tedious work.

It is desirable to seek a more efficient alternative method to obtain the pore water pressure. Thus, process of predicting the pore water pressure can provided ability to counter the problem of instrumenting the slope. Provided with the present data of pore water pressure and rainfall intensity, a model of support vector machine has the capability to perform the prediction of pore water pressure. The development of the model will require the antecedent data of pore water pressure and rainfall.

1.3 OBJECTIVE AND SCOPE OF STUDY

1.3.1 Objective:

To predict soil pore water pressure responses to rainfall events using polynomial kernel support vector machine (SVM) model. The development of SVM model will using the antecedent data of pore water pressure and rainfall. The observed data is taken from the data collected from instrumented slope date 1 January 2015 to 31 March 2015 (3 months).

To evaluate the performance of the polynomial kernel support vector machine model in predicting the pore water pressure responses to rainfall events. The three months data obtained are divided into training period and testing period. The machine learning of SVM will provided training data for training of data. Then, the data will be tested using the data of testing period. Thus, performance of model will be measured using standard statistical measure.

1.3.2 Scope of Study

The project is bounded within a scope thus to perform a focused study;

1.3.2.1 Location of study is a slope within area of Universiti Teknologi PETRONAS (UTP), Perak. The slope is opposite of Block 5, Chemical Engineering Department block. The reason of the slope is chosen is due to the feasibility of instrumenting the set-up of tensiometer and rain gauge which purposely to collect the data of soil pore water pressure and rainfall.

1.3.2.2 Polynomial kernel Support Vector Machine will be used to develop the model. Presently, there are few well known or basic kernel which are available other than the polynomial kernel. Those are Radial Basis Function (RBF) kernel, linear kernel, sigmoid kernel and a few more. The intention of using the polynomial kernel is because of its able to perform the process much faster compared to other available kernels.

1.3.2.3 The study is to predict the soil pore water pressure due to rainfall using data collected from 1 January 2015 to 31 March 2015. The period of collected data is three (3) months.

CHAPTER 2

LITERATURE REVIEW AND THEORY

2.1 SOIL PORE WATER PRESSURE

Pore water pressure can be defined as the pressure exerted by water which exhibit within soil pore or void. Mustafa et al. (2016) urge that, sometime, it can be referred as unsaturated soil mechanics or matric suction (negative PWP) because it can caused section effect. The pressure is fluctuated depending on current wet or dry soil condition. During the rainfall, the water will seep into the soil thus filling in the present void within soil and cause the pressure to escalate. Inversely, without the present of water within the void, the surrounding soil will give pressure and cause cohesive between soil particles. Thus it can be deduced that rainfall intensity will cause the shear strength of the soil and act as vital role in affecting the slope stability (Schneellmann et al., 2010)

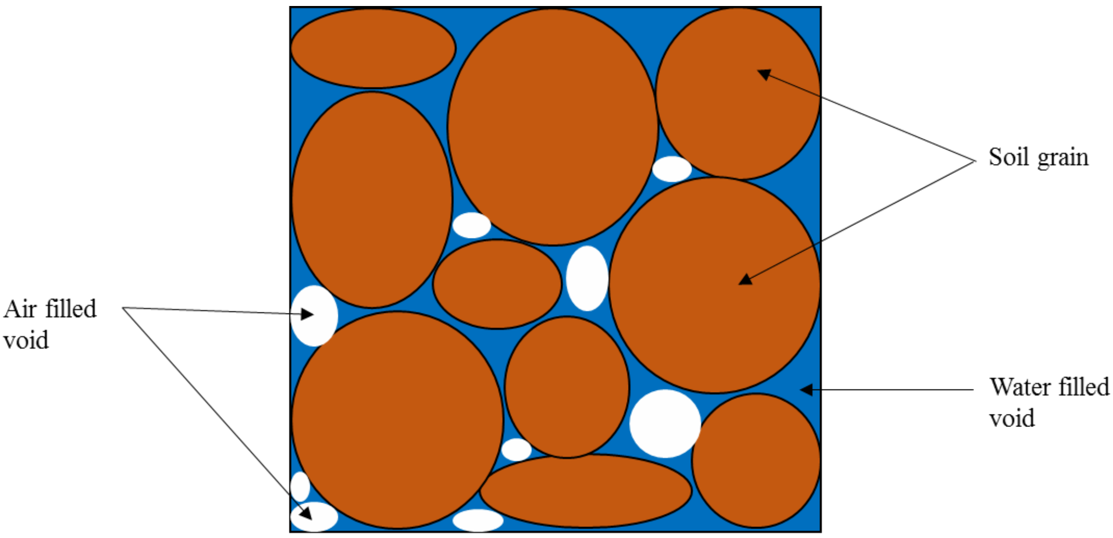

FIGURE 1. Saturated soil section component

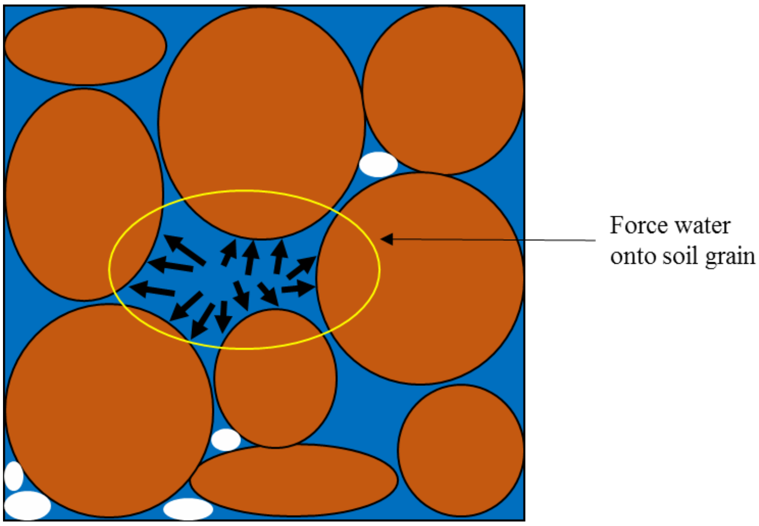

FIGURE 2. Pressure action in wet soil

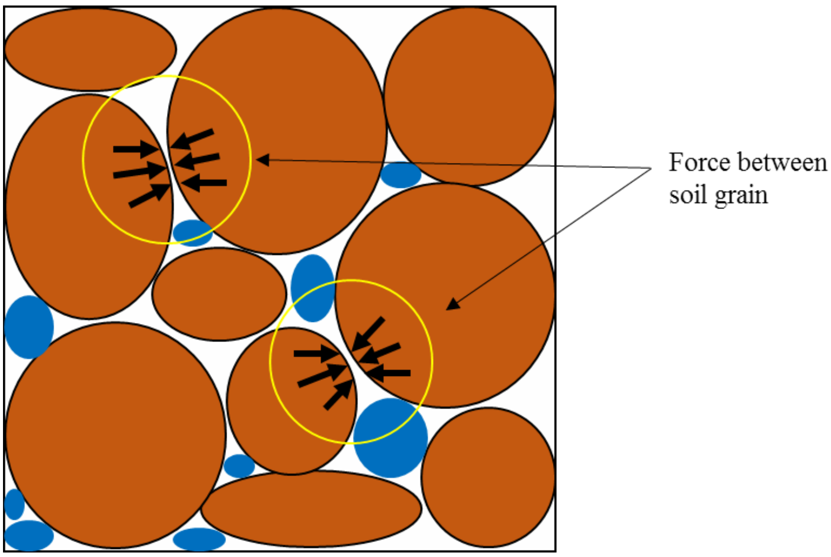

FIGURE 3. Pressure action in dry soil

2.2 SUPPORT VETOR MACHINE

Support Vector Machine (SVM) was introduced by Vapnik and others in the early 1990s. SVM is a machine learning systems that utilize a hypothesis space of linear functions in a high dimensional feature space, trained with optimization algorithms that implements a learning bias derived from statistical learning theory. Hong and Wan (2011) has claimed that fluctuation of rainfall data collected at instrumented field is complex, thus for identifying and modelling the system to predict the behaviour of pore water pressure, which much affected by rainfall intensity, developing the parametric equation is extremely difficult.

To encounter with such problem, Mustafa et al. (2016) has initiated that data-driven models on artificial intelligence or soft computing is the way out with a claim that Artificial Neural Network (ANN) is the most notable system as to modelling a complex relationship and widely accepted. ANN has proven as well performer in many field of study including engineering. It was then followed with development of SVM which claimed as most successful method of classification in machine learning, and has been demonstrated in various studies to be much more robust in many classification and recognition fields than the next best method. Burgess (1998) has established that SVM is more robust and more promising method in both classification and regression while Zhao et al. (2013) has added it can be performed even with fewer training datasets.

2.3 APPLICATION OF SOFT COMPUTING IN PREDICTING PWP

TABLE 2. Application of Prediction of PWP

Items	**Radial Basis Function Neural Network**	**Support Vector Regression**	**Low Degree Polynomial Kernel SVM**
Author	- M.R. Mustafa - R.B. Rezaur - H. Rahardjo - M.H. Isa	- N.M. Babangida - M.R. Mustafa - K.W. Yusuf - M.H. Isa	- N.M. Babangida - M.R. Mustafa - K.W. Yusuf - M.H. Isa - I. Baig
Year Published	2012	2016	2016
Reason	RBFNN is suitable for mapping the non-linear, complex behavior of porewater pressure responses to rainfall	SVR model is effective in providing an accurate and quick means of obtaining pore-water pressure response, which may be vital in systems where response information is urgently needed	It generally does not give better accuracy. However, polynomial kernels are found to perform better in natural language predictions

CHAPTER 3

METHODOLOGY/ PROJECT WORK

3.1 RESEARCH METHODOLOGY

General methodology has been listed out to ensure the smoothness and effectiveness of the whole study process.

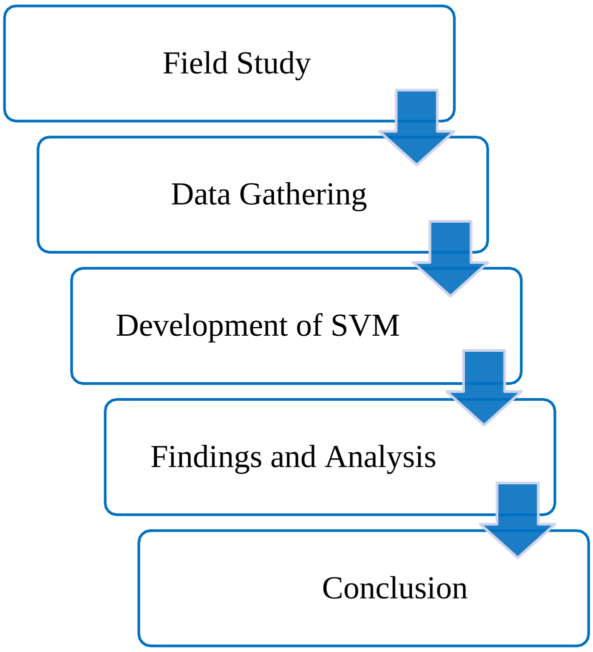

FIGURE 4. Research Methodology

3.1.1 Field Study

To implement the prediction of PWP using SVM, a suitable slope was selected and furnished with the necessary instruments to collect rainfall and corresponding PWP response. Prior to this project, an instrumented slope at Universiti Teknologi PETRONAS (UTP) in Malaysia is selected, specifically is a slope in opposite of Block 5, Chemical Engineering Department block.

FIGURE 5. Slope for Field Study

3.1.2 Data Gathering

The slope is instrumented with tensiometer which fitted with transducer and data logger to gather data of soil pore water pressure and rain gauge to obtain data of rainfall.

3.1.3 Development of SVM

The model development will cover on data partitioning, input data determination and model structure, data normalization, model generation, and determination of model performance evaluation criteria.

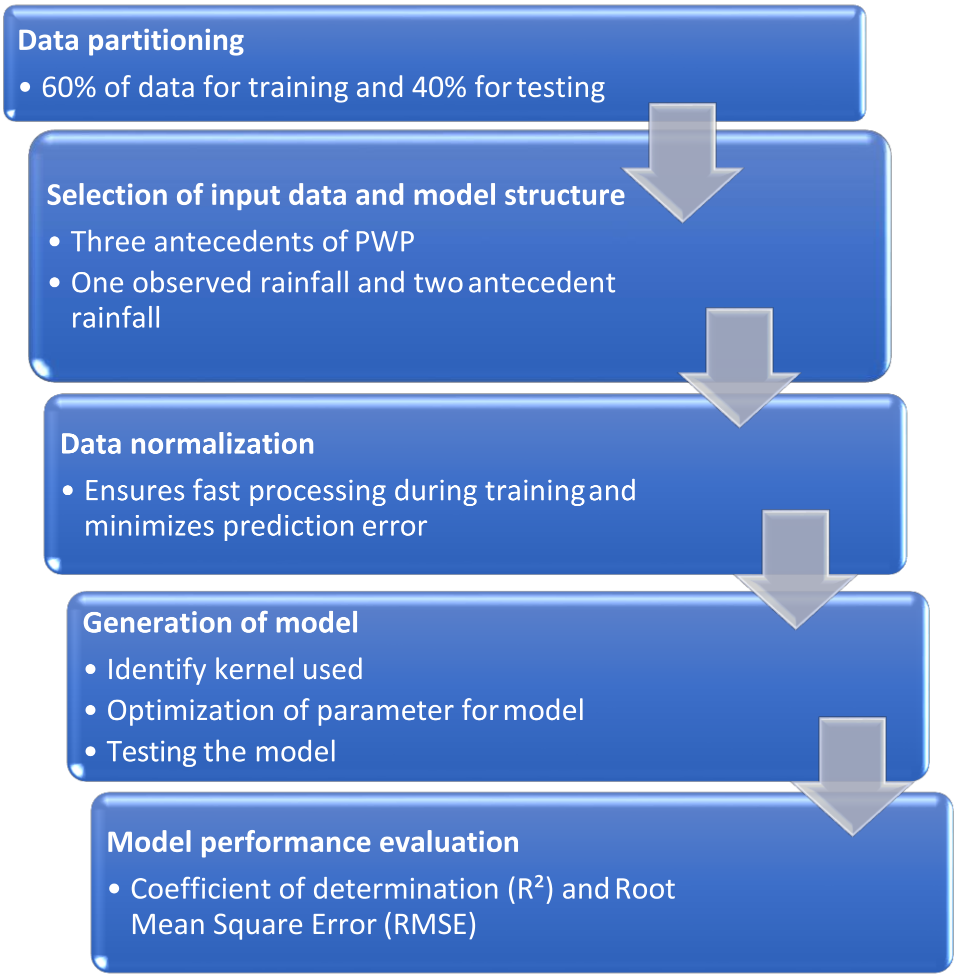

FIGURE 6. Development Stage of Model

3.1.3.1 *Data Partitioning;* accumulated of 4312 number of data are to be analysed in this project. Thus, the data has been divided into two category which is for training and testing purpose. 60% of data (2588 number of data) is for the training purpose while the rest of 40% (1724 number of data) is for the testing purpose. The reason of percentage of training data is higher than the testing data is because of the model need to be trained more so that it can perform better in testing of the model.

3.1.3.2 *Selection of input data and model structure* ; selection of input data is a selection of training data which represent non-linear mapped by model during training are highly affecting the accomplishment in mapping of complex non-linear behaviour (Rojas, 1996). The structure of the model are represented in form algorithm. Whereas, selection of model algorithm is important to get the fastest and most appropriate training to solve given problem. Meanwhile, performance of algorithm training depend on the problem complexity and type to be modelled, data used for training the network and network architecture (Rafik et. al, 2008). Babangida N.M et al., (2016) has considered five antecedent of PWP , present and five antecedent rainfall for analysis and determination of required feature which make up to eleven input feature even though the relevancy of input feature may be restricted to its subsets only. However, this project will be chosen three antecedent of PWP and present and two antecedent rainfall as structure of the model.

$$U_t = f\ (u_{t-1}, u_{t-2}, u_{t-3}, r_t, r_{t-1}, r_{t-2} \qquad \textbf{(1)}$$

Based on the Equation 1;

$\boldsymbol{U_t}$ = the model structure, $\boldsymbol{u_t}$ = the pore water pressure, $\boldsymbol{r_t}$ = the rainfall

3.1.3.3 *Data Normalization*; Rojas (1996), has claimed that the normalization of data is purposely to have better processing in term of reducing computational timing as well to reduce the error in predicting. Normalization or scaling down the data has been applied to all training and testing datasets and has been set to the range of 0 to 1 only. Equation 2 is the formula;

$$v_p = \frac{(x_p - x_{min})}{(x_{max} - x_{min})} \qquad \textbf{(2)}$$

Based on equation 2;

$\boldsymbol{v_p}$ = the normalization equation, $\boldsymbol{x_p}$ = current data value, $\boldsymbol{x_{min}}$ = minimum dataset value, x_{max} = maximum dataset value

3.1.3.4 *Generation of model ;* three main stages are implemented in this step starting with choosing the kernel of the model, optimizing the parameter that will be used and do testing on the developed model. First, kernel purpose is to provide a simple avenue that will transform the non-linearity to linearity of algorithm as well to map the input space into higher dimensional feature space

(Babangida et al, 2016 & Babangida et al, 2016). There are a few common kernel that widely known which are Radial Basis Function (RBF) kernel, linear kernel, sigmoid kernel and few others. The polynomial is chosen because of its capable to perform the process much faster compared to other available kernels as well its advantage in performing better in natural language prediction (Babangida et al, 2016)

$$k\,(\boldsymbol{x},\boldsymbol{y}) = (-\gamma\boldsymbol{x},\boldsymbol{y} + r)^d \qquad \textbf{(3)}$$

Based on Equation 3;

x and **y** are input vectors, ***γ***, ***r*** and ***d*** are parameters that need to be tuned for the polynomial kernel.

3.1.3.5 *Model Performance Evaluation*; the performance parameter is using the standard statistical measure which to evaluate the capability of the developed model to predict the pore water pressure in response to rainfall. There are few model for using of performance evaluation. In this project, the performance parameter that been used is Root mean square error (RMSE) and Coefficient of determination (R^2). RMSE has purpose to measure the difference between values of the observed PWP with the predicted PWP. Better value for RMSE is when the value is closer to 0. As for the R^2, can be defines as predictable of the proportion of the variance in the dependent variable using of the independent variable. The closer of the R^2 to value of 1 is favourable in this project scope since it will indicate the developed model has performed well in predicting the PWP.

$$RMSE = \left[\frac{1}{N}\sum_{k=1}^{N} (\hat{u}_k - u_k)^2\right]^{1/2} \qquad \textbf{(4)}$$

$$\boldsymbol{R^2} = \frac{\sum(\boldsymbol{u_i} - \bar{u}_i)^2 - \sum(\hat{u}_i - \boldsymbol{u_i})^2}{\sum(\boldsymbol{u_i} - \bar{u}_i)^2} \qquad \textbf{(5)}$$

Based on Equation 4 & 5;

$\boldsymbol{u}$ and $\hat{\mathbf{u}}$ represent observed and predicted values of PWP respectively, $\boldsymbol{u}$ = mean of observations, $\boldsymbol{n}$ = number of observations, $\boldsymbol{k}$ = number of model parameters.

3.1.4 Findings and Analysis

The stage which the results obtained will be presented and in depth evaluating the data using proper analytical and logical reasoning.

3.1.5 Conclusion

The last stage which to conclude the overall project based on the findings and analysis that has been conducted in previous stage.

3.2 PROJECT ACTIVITIES

3.2.1 Final Year Project 1

TABLE 3. FYP 1 Project Milestone

PROJECT FLOW	WEEK													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Selection of Project Topic														
Preliminary Research Work														
Submission of Extended Proposal						●								
Field Study														
Proposal Defence									●					
Gathering Data														
Submission of Interim Draft Report													●	
Submission of Interim Report														●

LEGEND :

Gantt Chart

● Key Milestone

3.2.2 Final Year Project 2

TABLE 4. FYP 2 Project Milestone

PROJECT FLOW	WEEK														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Model Development															
Findings and Analysis															
Submission of Progress Report							●								
Study Conclusion															
Pre-SEDEX										●					
Submission of Final Draft Report											●				
Submission of Dissertation												●			
Submission of Technical Paper												●			
Viva													●		●
Submission of Project Dissertation															

LEGEND :

Gantt Chart

● Key Milestone

CHAPTER 4

RESULT AND DISCUSSION

4.1 Comparison between Training Datasets and Testing Datasets

Based on data obtained from the project site, initial data analysis has been made, which is to acquire statistical data using simple statistics formula.

TABLE 5. Statistics Formula

No	Name	Formula
1	Sample mean,	$x' = \frac{Total\ value,\ \sum x}{Total\ number, N}$
2	Sample variance, σ^2	$\sigma^2 = \sqrt{\frac{\sum(x - x')^2}{(N-1)}}$
3	Standard deviation, σ	$\sigma = \sqrt{(\sigma^2)}$

The data acquired will be separated into two (2) stage which are the *Training* and *Testing*. This separation is purposely to evaluate the validity and consistency of the training and testing datasets.

A statistical data analysis has been done on both data of PWP and rainfall, and the statistics can be observed below;

TABLE 6. Statistics Data of PWP & Rainfall

	Training (1/1/15-24/2/15)		Testing (24/2/15-31/3/15)	
	PWP	Rainfall	PWP	Rainfall
N	2588	2588	1724	1724
Min	-34.1 kPa	0.0 mm	-19.4 kPa	0.0 mm
Max	-3.5 kPa	46.0 mm	-3.4 kPa	26.5 mm
Total value	-27784.8 kPa	306.5 mm	-17754.4 kPa	224.5 mm
SM	-10.74	0.12	-10.30	0.13
SV	16.97	1.57	4.84	1.89
SD	4.12	1.25	2.20	1.38

Legend :

N=No of data, Min=minimum value of data, Max=maximum value of data, SM=sample mean, SV=sample variance, SD=standard deviation

4.2 Model Performance Evaluation

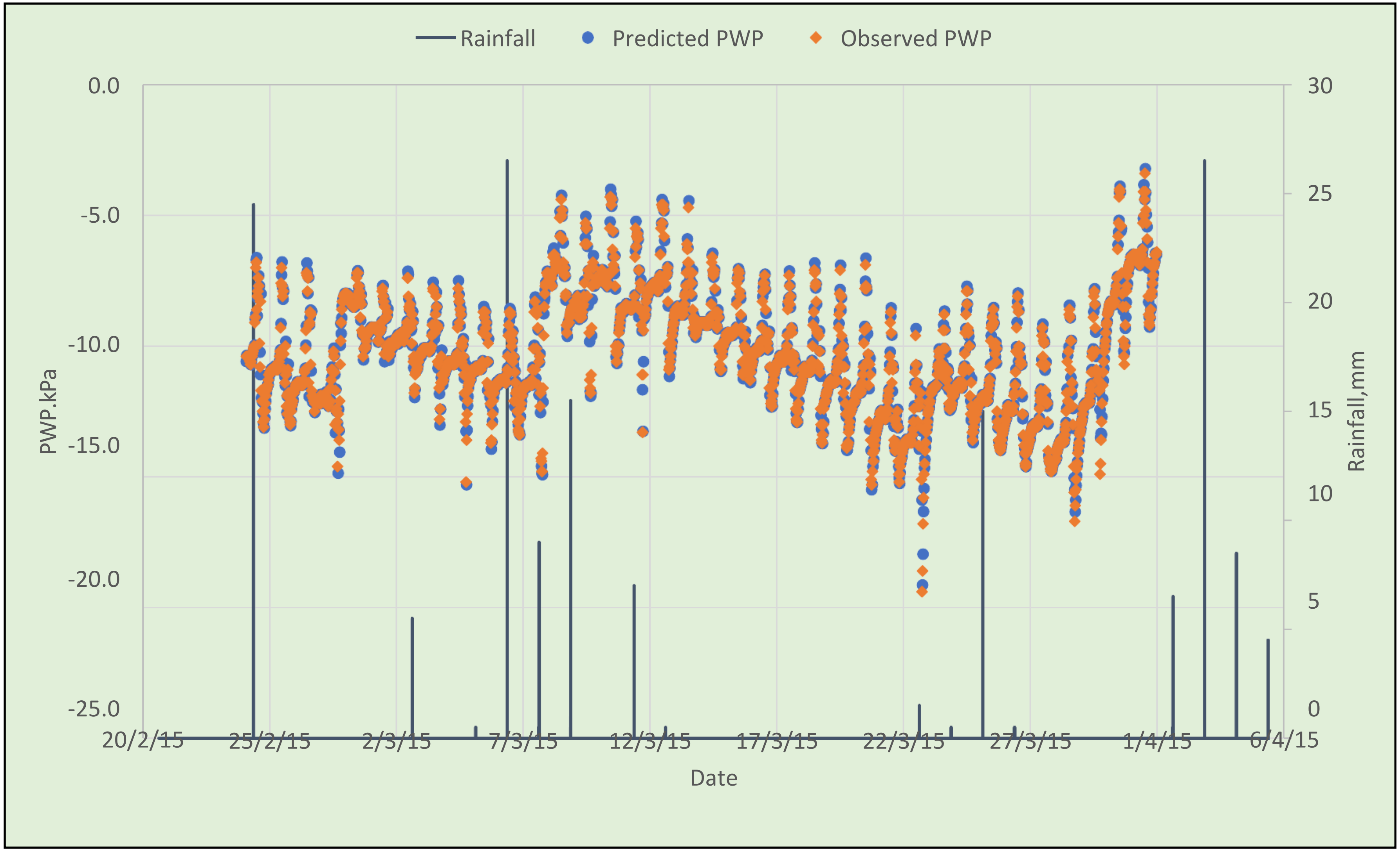

FIGURE 7. Comparison of PWP between observed PWP and predicted PWP responses to rainfall

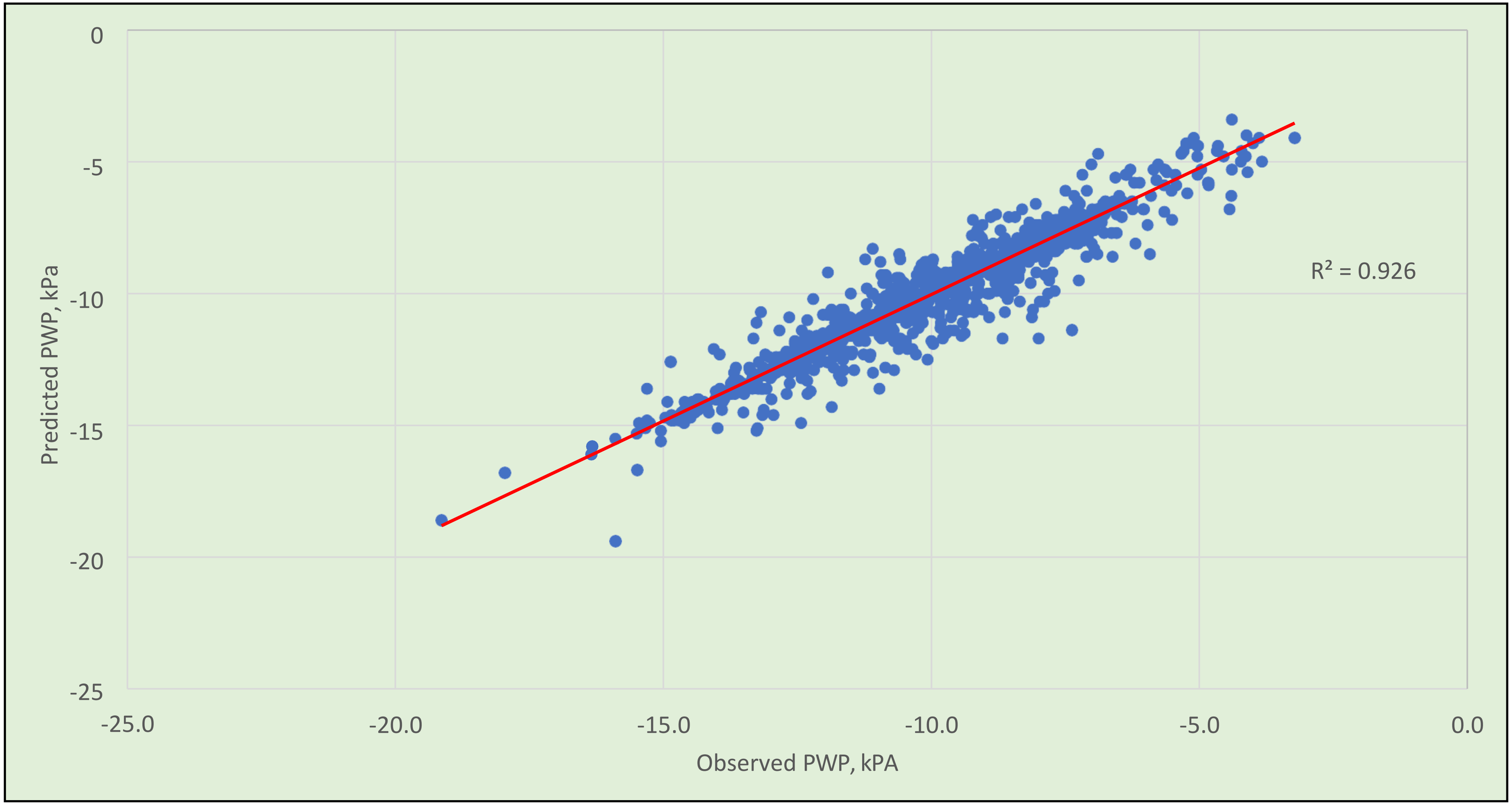

FIGURE 8. Scatter plot between observed PWP and predicted PWP

TABLE 7. Performance of Model

Performance parameter	Value
R^2	0.93
RMSE	0.60

The Figure 7 has shown the comparison of developed model in predicting the pore water pressure responses to rainfall. Blue color plot represent the predicted PWP as the orange color plot represent the observed PWP. Meanwhile, the line graph is to indicate the rainfall intensity in conjunction with the observed PWP and predicted PWP.

Based on the graph plot, it has shown that the predicted PWP has mimicked the observed PWP since the pattern of the plotting is identical to each other. This has shown the capability of the model to produce the prediction that can give almost similar reading compared to the observed data.

It is also can be observed that, the pattern of both observed and predicted PWP is fluctuating. The reason of this occurrence may due to its comparison in conjunction with the rainfall event. It has shown that the pore water pressure inside the soil will increase after a high rainfall intensity. Whereas, the pore water pressure inside the soil will decrease after a few time of low rainfall intensity.

The fluctuation pattern of pore water pressure has given the proofing to claim that the rainfall event can cause the variability to the pore water pressure. To explain more, this occurrence could possibly due to infiltration of rainfall into the soil that may cause the water to fill in the void then exerting the pressure outward to the soil. Thus, causing the pore water pressure to increase.

Based on the Figure 8, it has shown the scatter plot between observed PWP and predicted PWP. It also featured the line fitness which produced from the plotting between the observed PWP and predicted PWP. The purpose of line of fitness or linear line is to show the linearity of the plotted data and how well to indicate the comparability between the observed PWP and predicted PWP.

Referring to the graph of Figure 8, it also has able to evaluate the performance of the prediction model in predicting pore water pressure responses to rainfall intensity. R^2 or coefficient of determination is one of performance evaluation parameter that can be obtained from the graph. The value of obtained R^2 is 0.93 which mean that predicted PWP obtained has almost similar to the observed PWP. This is because the closer of R^2 to value of 1 is favorable. Meanwhile, the value of RMSE is 0.6. It is better get the closer value of RMSE to value of 0 because it can indicate that the model can produce lower differences compared to the observed PWP. Thus, it has shown that a good result in R^2 is not necessarily will give a good RMSE value.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

The choice of polynomial kernel has given a much faster time in processing the model structure. Using three antecedent of PWP and current and two antecedent rainfall are chosen in this project as input feature. It is a simple model yet able to predict the PWP with a good performance of 0.93 and 0.60 of R^2 and RMSE respectively. Although, based on the literature that been reviewed, there are a better performance that can be achieved, prominently is by using radial basis function (RBF) kernel. However, considering to use polynomial kernel support vector machine has given a good correlation in predicting the pore water pressure in responses to rainfall intensity. Provided, further manipulating the parameter of $\boldsymbol{\gamma}$, $\boldsymbol{r}$ and $\boldsymbol{d}$ in the polynomial function, input feature and model algorithm may give a better correlation between observed and predicted PWP.

In the nutshell, the ability to predict the upcoming soil pore water pressure can give advantage in applying the model in the slope management studies. Maybe from the predicted PWP fluctuation pattern can used to improvise the slope failure warning system or revising the factor of safety to be imposed in slope stability studies. Thus, all the objective has been achieved successfully, which are to develop a model to predict the soil pore water pressure using polynomial kernel SVM and to evaluate performance of SVM in predicting the pore water pressure responses to rainfall intensity.

REFERENCE

Babangida N.M., Yusof K.W., Mustafa M.R., and Isa M.H (2016). Comparison of Support Vector Machine Kernel Functions for Pore-Water Pressure Modelling. *Taylor & Francis Group, London* 294-297

Burges CJC (1998). A Tutorial On Support Vector Machines For Pattern Recognition. *Data Min Knowl Disc* 2:121–167

Hong YM, Wan S (2011) Information-Based System Identification For Predicting The Groundwater-Level Fluctuations Of Hillslopes. *Hydrogeol J* 19:1135–1149

Mustafa M. R., Rezaur, R. B., Rahardjo, H., and Isa, M. H. (2011). Prediction Of Pore-Water Pressure Using Radial Basis Function Neural Network. *Eng. Geol*., 135–136, 40–47

Mustafa MR, Rezaur RB, Saiedi S., Rahardjo H., and Isa M.H (2013) Evaluation of MLP-ANN Training Algorithms For Modeling Soil Pore-Water Pressure Responses to Rainfall. *J Hydrol Eng* 18:50-57

Mustafa MR., Nuraddeen MB, Khamaruzaman WY., Isa MH. (2016). Prediction of Pore- Water Pressure Response To Rainfall Using Support Vector Regression. *Springer Hydrogeol J*

Rafik, Z., Ridha, B., and Daniel, R. (2008). Levenberg-Marqurdt Learning Neural Network For Adaptive Pre-Distortion For Time Varying Hpa With Memory In Ofdm Systems. *16th European Signal Processing Conference (EUSIPCO 2008),* Lausanne, Switzerland.

Rojas, R., 1996. Neural Networks: A Systematic Introduction. *Springer Verlag,* Berlin.

Schnellmann, R., Busslinger, M., Schneider, H.R., Rahardjo, H., 2010. Effect Of Rising Water Table In An Unsaturated Slope. *Engineering Geology* 114, 71–83

ZhaoW, Tao T, Zio E (2013) Parameters Tuning In Support Vector Regression For Reliability Forecasting. *ChemEng Trans* 33:523–528.