CERTIFICATION OF APPROVAL

River Discharge Prediction at Kinta River using Multi Quadric Radial Basis Function

By

NURUL NEESA BT IDRIS

16704

A project dissertation submitted to the

Civil Engineering Programme

Universiti Teknologi PETRONAS

in partial fulfilment of the requirements for the

BACHELOR DEGREE OF ENGINEERING (Hons)

(CIVIL)

Approved by,

____

(Dr Muhammad Raza Ul Mustafa)

Universiti Teknologi PETRONAS

Tronoh, Perak

May 2015

RIVER DISCHARGE PEDICTION AT KINTA RIVER USING MULTI QUADRIC RADIAL BASIS FUNCTION

By

NURUL NEESA BT IDRIS

16704

Dissertation submitted in partial fulfilment of

the requirements for the

Bachelor Degree of Engineering (Hons)

(Civil Engineering)

MAY 2015

Universiti Teknologi PETRONAS

Bandar Seri Iskandar,

31750 Tronoh,

Perak Darul Ridzuan.

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 undertaken or done by unspecified sources or persons.

NURUL NEESA BT IDRIS

ABSTRACT

In recent years, data evaluation approaches such as artificial neural network (ANN) techniques are being increasingly used for river flow forecasting. For efficient management of water resources, accurate and reliable flow prediction is extremely important. Additionally, it is also true for effective flood risk management. In general, streamflow prediction models when incorporated within flood forecasting systems serve as tools for early warning systems so as to reduce flood damages on one hand and may also result in considerable economic and social benefits. The specific objective of this study is to develop Multi-Quadric Basis function Neural Network model for the prediction of river discharge at Kinta River and to evaluate the performance of the Multi-Quadric basis function model using different statistical performances measures. The ANNs model for this study is developed in MATLAB software. To measure the performance of the model, four criteria performances, including a coefficient of determination (R2), the sum squared error (RSE), the mean square error (MSE), and the root mean square error (RMSE) are used. The results of this study could be used to help local and national government plan for the future and develop appropriate to the local environmental conditions new infrastructure to protect the lives and property of the people of Perak.

ACKNOWLEDGEMENT

In the name of Allah, the Most Beneficent, the Most Merciful, I would like to express my gratitude to who contributes great experience and learning opportunities for the past 8 weeks to complete my final year project.

First and foremost, sincerely thanks to my supervisor, Dr. Muhammad Raza Ul Mustafa for his guidance and supervision along the way I undertaken final year project. Under his supervision, I gained a lot of experiences and knowledge as an undergraduate student.

Next, I would like to thanks Dr. Kamaruzzaman for being my internal examiner in Proposal Defense, Pre-Sedex and Viva. Comments and evaluations given are accepted with an open heart as an advice. Not to forget, my external examiner, Ir. Abdul Isham b. Abd. Karim.

I also want to dedicate a special thanks to all my beloved family members and friends that always give support to me and help me to build up my confident in completing my dissertation.

Without the commitment and support from all parties that involved directly or indirectly it is impossible for me to successfully complete my final year project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1:

INTRODUCTION

1.1 Background

Hydrologic forecasting is significant for effective operation of a water resources planning or flood mitigation system and or to plan for future expansion or reduction. Flow forecasting also provides information about the sediment amount carried by the river to the reservoirs (Kişi, 2007). Water experts, with a reliable river flow forecast, can allocate water supplies for water users such as hydropower generation, agricultural, domestic and for the maintenance of environmental flows. Therefore, the study of river flow forecasting of Kinta River which are located at Perak River catchment is crucial as to maintain its function and to overcome the flooding issues happen again in future due to inconsistencies of river water levels.

Fundamentally, there is two techniques for river flow forecasting which is conventional method and soft computing technique. Conventional method tends to be inaccurate because it is linear and is measured using complex conceptual model such as curve fitting and regression model. When compared to soft computing technique for instance artificial neural network (ANN) they are able to predict nonlinear function such as river discharge. Many studies have compared ANN with linear regression approaches and verified that ANN can perform statistical technique. (Yu, Qin, Larsen, & Chua, 2013).

Artificial neural networks are flexible mathematical structures that are able of identifying complex non-linear relationships between input and output data sets (Huo et al., 2012). A neural network comprises of a large number of simple processing elements that are variously called neurons, units, cells, or nodes. Each neuron is connected to other neurons by means of direct communication links. The network usually has two or more layers of processing units where each processing unit in each layer is connected to all processing units in the adjacent layers. (Mustafa, Rezaur, Saiedi, Rahardjo, & Isa, 2012)

ANN technique will be using in this study at Kinta River, which focus on the application using Multi-Quadric. Artificial neural networks are chosen for various reasons. One of the reason is it do not underestimate a detailed understanding of a river's physical characteristics and require extensive data pre-processing. This is the advantages of ANNs because it can manage incomplete and ambiguous data (Dawson et al., 2002). Between two of the most popular neural network multilayer perceptron (MLP) and RBF, RBF is chosen because recent study, Dawson et al. (2002) stated in their paper that RBF predicts river flow accurately than the MLP.

1.2 Problem Statement

Nowadays, there are many technique have been used to predict river flow. However, the main problems figured was regarding the appropriateness of the technique conducted to measure and plotted the river flow data in previous study. They were using the conventional flow rating curve to determine the river flow discharge prediction. Those technique required the data obtained to be plotted into a graphical form before a linear function is apply. Nevertheless the application of this linear technique tend to produce less accurate result. As a matter of fact, the stage and discharge data is a non-linear form in its nature due to the variability of water level and time measurement. Therefore, Artificial Neural Networks is the alternative method to approximate nonlinear functions and data thus Multi Quadric Radial Basis Function is used in this study since it has been identify that this function has never been performed previously in Kinta River.

1.3 Objectives

The objectives of this research study are listed as follows:

1) To develop Multi-Quadric Basis function Neural Network model for the prediction of river discharge at Kinta River.
2) To evaluate the performance of the Multi-Quadric basis function model using different statistical performances measures.

1.4 Scope of Study

In this study is described about prediction of river discharge by using Multi-Quadric radial basis function (RBF). Prediction of river discharge in Kinta River is performed using hydrological data such as discharge, and river water level. The scope of study can be described as below:

1) The scope of study area is limited towards prediction of river discharge in Kinta River by developing Radial Basis Function (RBF). Few types of function listed inside the RBF namely Multi Quadric (MQ), Gaussian, and Thin Plate Spline (TPS) and logarithmic have been known to perform their own specific algorithm and function. Though, Multi Quadric radial basis function is chosen in this research study to be applied in developing the selected basis function model using MATLAB computing software.

2) To evaluate the performance of the Multi Quadric radial basis function by using different statistical measures that are root mean square error (RMSE), coefficient of determination (R^2), and the mean square error (MSE).

CHAPTER 2:

CRITICAL LITERATURE REVIEW

Accurate and reliable flow prediction is extremely important for efficient management of water resources. Moreover, it is also useful for flood risk management. In general, streamflow prediction models when incorporated within flood forecasting systems serve as tools for early warning systems to reduce flood damages (Shamseldin, 2010) and may also result in substantial economic and social benefits. In recent times, Artificial Neural Networks (ANNs) have been used for flow predictions, flow simulation, parameter identification and to model nonlinear input and output time series. Generally, an ANN is a network that relates the inputs and outputs of a system. The enormous success with which ANNs have been used to model the nonlinear system behavior in a wide range of areas indicates that this approach can be useful in river flow prediction also. Instead of its complexity structures, ANN is aimed to meet several purposes objective in solving hydrological flow predicting problems. Zhou and Han (1993) claimed that the principle of the existence of ANN is to discourse the problems of flooding. The process could be applied through evaluating the algorithm of the neural networks using the load of past input data, neural cells and noise containing data without required to design mathematical prototypes (Brion & Lingireddy, 2003). However, it is challenging to describe the variable using others network such as Linear regression analysis function during flood condition.

Artificial Neural Networks (ANNs)

ANNs are mathematical models of human insight that can be trained for performing a specific task based on accessible empirical data. When the relationships between data are unknown, it can become a greatest tool for modeling. (Masoud et. al., 2011). Moreover, ANN are capable of identifying complex non-linear relationships between input and output data that consist of data processing units called nodes or neurons arranged in layers. It is supported by Jain and Chalisgaonkar (2000) and (Supharatid, 2003), they stated in their research paper an ANN is a network of parallel, distributed information processing systems that relate an input vector to an output vector which consist of neurons organized in layers.

Furthermore, according to Jain and Chalisgaonkar (2000), an ANN was created in a very special way to try to be like the function of human intelligence which consist of billions of interconnections. In the ANNs, it consists of a number of information processing elements called neurons or nodes, which are grouped in layers (Jain and Chalisgaonkar, 2000). There are three layers in the neurons, the first layer is known as input layer or processing elements that receive the input vector and transmit the values to the next layer across connections where this process is continued and the last layer is known as output layer, whereas layers in between are known as hidden layer.

This classification of network, where data flow one way or forward, is known as a feedforward network. As mentioned before, a feedforward ANN has three layers. Each of the neurons in a layer is connected to all the neurons of the next layer, and the neurons in one layer are linked only to the neurons of the immediate next layer. The strength of the indication passing from one neuron to the other depends on the weight of the interconnections (Jain and Chalisgaonkar, 2000). Dawson et al., 2002 also stated the same fact that is the neurons in a layer are interconnected with neurons in adjacent layers by connection weights.

Since the 1980s, study in artificial neural networks has enhanced and today neural networks are utilized in many diverse applications using different network types, training algorithms and structures. (Dawson et al., 2002). Besides that, one of the important parts in the ANN system is the determination of the hidden layer numbers. The number of nodes in the hidden layer was determined using the application of Kolmogorov's theorem whereby the least number of nodes should follow the formula of 2n+1 (where n represent the number of nodes in the input layer) (Feng and Lu, 2010). This is because the hidden layers enhance the network's ability to model complex functions. A three-layer feedforward ANN along with a typical processing element is shown in Figure 1. The data passing through the connections from one neuron to another are influenced by weights that control the strength of a passing signal. When these weights are adjusted, the data transferred through the network change and the network output alters.

Moreover, Jain and Chalisgaonkar (2000) claimed the neurons in a layer share the same input and output connections, but do not communicate among themselves. All the nodes within a layer act synchronously. Therefore, at any point of time, they will be at the similar stage of processing. The activation levels of the hidden nodes are transmitted across connections with the nodes in the output layer. The level of activity generated at the output nodes is the network's solution to the problem presented at the input nodes.

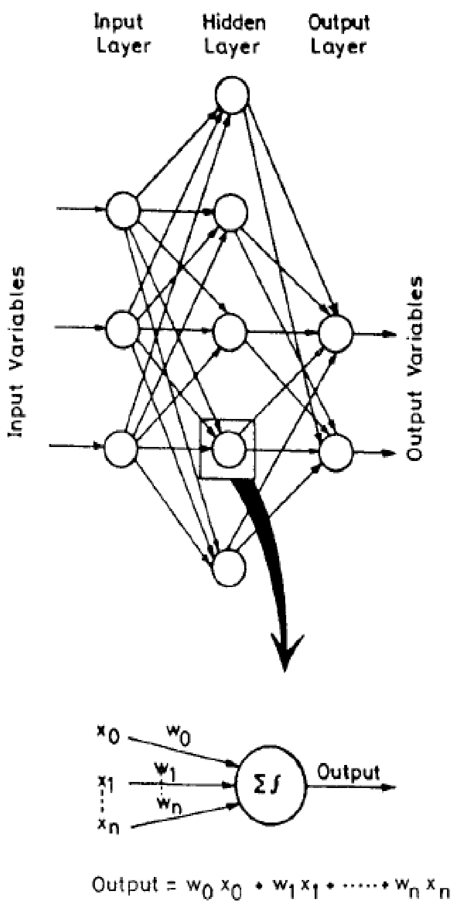


Figure 2.1: Typical Three-Layer Feedforward Artificial Neural Network

Nevertheless, there were boundaries in the work scope of ANN, a detailed review made by ASCE, 2000 found that even though there were general application of ANN in the hydrological engineering, ANN cannot be treated as a replacement for the other hydrological modelling technique because the physics of the basic or foundation process in the system was confidentially stored in the optimal weight and threshold value and never been visible to the user even after the end of training stage.

Consequently, thorough studies regarding the application of ANN must be done in order to ensure that this system will able to meet the objective designed.

In addition, a comparison between model performances was made by Hsu et al. (1995) using daily steps as stated in their paper. They prove that ANN could better stimulate the rainfall-runoff relationship on a river basin in Mississippi, USA, when compared to a conceptual model. According to Wang et al., 2009; Lohani et al., 2011 and Lin et al. 2006 ANNs have been compared to other methods including Genetic Algorithm (GA), Support Vector Machine (SVM), Fuzzy Logic (FL), and linear transfer function for river flow simulation and obtained better performance. Furthermore, the recent decade has seen a tremendous growth in the interest of application of ANNs in streamflow modeling

Artificial neural networks is chosen as the functional technique in most of the research study for various reasons. One of them is, it do not underestimate a detailed understanding of a river's physical characteristics, or require extensive data pre-processing (Dawson et al., 2001). This is because ANNs can handle incomplete and ambiguous data. Additionally artificial neural networks are simpler to implement than physically-based hydrological models. ANNs are also well-matched to dynamic problems and are parsimonious in terms of information storage within the trained model.

Therefore, ANN is found to be the best alternative to solve those problems since artificial neural network was able to complete the data in the network from end to end relations between the neural cells. Furthermore, the system also needs special learning process to enable the process of mapping the variables to be possible to produce accurate result (Feng and Lu, 2010).

Radial Basis Function Neural Networks (RBFNs)

RBF neural networks (RBFNs) are a class of feedforward neural networks that are used for classification problems, function approximation, noisy interpolation, and regularization. They have gradually attracted curiosity for engineering applications due to their advantages over traditional multilayer perceptron, namely faster convergence, smaller extrapolation errors, and higher reliability (Moradkhani et al., 2003). Moreover, the RBF technique offers good generalization ability with a minimum number of nodes to avoid unnecessarily lengthy calculations, in comparison with multilayer perceptron networks, which showed that RBFs are highly promising for multivariable interpolation given irregularly positioned data points.

The objective of any RBFN design process is to determine centers, widths and the linear output weights connecting the RBFs to the output neuron layer. The most traditional learning procedure has two stages first, learning of centers and widths, and then, training of output weight. Girosi and Poggio (1990) and Moradkhani et al. (2003) presented that RBFNs have the best approximation property, which is not for multi-layer perceptron type of neural networks. Their use in neural networks has found applications in solution of classification problems, function approximation, noisy interpolation, and regularization in various engineering fields due to their advantages over traditional multilayer perceptron, such as smaller extrapolation errors, and higher reliability (Girosi and Pogio, 1990).

The architecture of radial basis function neural networks in Figure 2 consists of an input layer, one hidden layer and one output layer. Each node in the hidden layer evaluates a radial basis function on the incoming input. Differing a general type of a MFN network, the connections between the input and hidden layer are not weighted A distinct advantage of RBFNs is the possibility of choosing appropriate parameters for the transfer functions at the hidden nodes, by estimation in advance without having to accomplish a full nonlinear optimization of the network. As stated by Moradkhani et al., (2003) the RBF technique provides good generalization ability with a minimum number of nodes to avoid unnecessarily lengthy calculations, in comparison with multilayer perceptron networks.

Kasiviswanathan and Agarwal (2012) mentioned that the function node in the RBFNN is different compared to the one applied in the Back Propagation Neural Network. It does not implement the same mechanism of multiply and add of the weighted summation, it computes a respective field from the individual function overlaps. In addition, the function nodes is not a problem dependent function since it rely heavily on the network designer on how to set up the function based on the model performances (Kasiviswanathan & Agarwal, 2012).

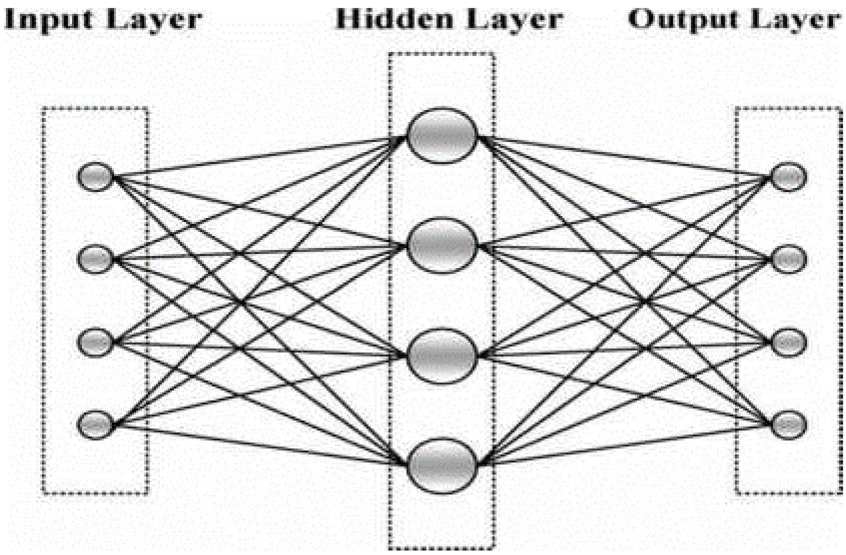


Figure 2.2: Typical Radial Basis Function

Nevertheless, in the RBFNN the main uniqueness lying in the structure of the hidden layer and the output layer. The hidden layer comprised of non-linear function which has its own specific function shape. While, the output layer is normally comprise of only one node. In point of fact the numbers of nodes in the output layer in RBFNN depend only on the variables fixed. On the other hand, it is known that RBFNN has a greater reliability, faster convergence and analysis and produce very minor error compared to the conventional multilayer perceptron.

Besides that, according to Mustafa. et.al, 2014, the use of trial and error method to classify the number of neurons in the hidden layer has been found to produce a better result as compared to the existing conventional regression analysis method. In fact this method is vital to ensure that throughout the training, the configuration set which gave the maximum Efficiency Index (EI) and minimum Root Mean Squared Error (RMSE) and Mean Absolute Deviation (MAD) is selected and this must be done with reference to the minimum allowable number of hidden nodes (Shamseldin, 2010).

	MSRE	*RMSE (cumecs)*	*CE (%)*	*AIC*	*R-squared*
ARMA	0.0202	4237	75.98	1674	0.91
MLP	0.0073	2068	94.24	1609	0.97
SWMLR	0.0107	2998	87.90	<u>1607</u>	0.97
ZOF	*0.0362*	*6590*	*41.50*	*1759*	*0.49*
RBF (MQ)	<u>0.0070</u>	<u>1933</u>	<u>94.97</u>	1611	<u>0.98</u>

Figure 2.3: Experimental Result of RBF

Based on the experiment done by Dawson et al., (2002) result obtained shows that RBF obtained the lowest value for mean square root error, (MSRE) when comparing with other neural networks such as ARMA, MLP, SWMLR and ZOF.

Multi Quadric Basis Function

Between two of the most popular neural network multilayer perceptron (MLP) and RBF, RBF is chosen because of recent study, Dawson et al. (2002) stated that RBF predicts river flow accurately than the MLP. According to Mustafa et al. (2012), the model architecture's performance are measured by using error basis measurement such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of efficiency (E), Mean Squared Relative Error (MSRE) and coefficient of determination (R2) to indicate the overall performance of the selected network. In Figure 4 shows that Multi Quadric obtained minimum value for MSRE compared to other function that was made by Dawson et al., (2002).

	MSRE	*RMSE (cumecs)*	*CE (%)*	*AIC*	*R-squared*
Cubic	0.0261	*5398*	*60.76*	*2096*	0.93
Gaussian	0.0148	3647	82.14	1976	0.94
IMQ	0.0140	3240	85.91	2079	0.95
Linear	0.0119	2606	90.86	2049	0.96
MQ	0.0070	1934	94.97	1611	0.98
TPS	*0.0322*	4545	72.18	1720	*0.85*

Figure 2.4: Experimental Result of MQ

Based on the literatures reviews above, Multi-Quadric Radial Basis Function used has been proven to show improvement in the water flow forecasting techniques in comparison to the other function. This technology is very important in river flow calculation process since the stage, discharge and other non-linear hydrological variables play significant roles in determining the correct discharge value from the inserted stage data. The application and the development of radial basis function seem to bring more advantages in producing the accurate outcome result for the betterment of hydrological research study.

CHAPTER 3:

METHODOLOGY

3.1 Data Source and Study Area

The research data are obtained from Department of Irrigation and Drainage (DID) Daerah Kinta, Perak. The data used in this study consist of two variables of hydrological resources which are Water Level (WL, m) and Discharge (DC, m3/s). For this paper, the study area chosen is Kinta River located at Perak, Malaysia. Based on the data given, the records consist of three variables of hydrological resources which are Water Level (WL), and Discharge (DC). Each of these data comprises of their own specific value and unit (m and m^3/s respectively) which was tabulated into group form according to subsequent years onwards starting from year 1990 until 2013. The daily data were tabulated according to the months from January until December for each and every years. Out of these 23 years historical data merely 3 recent data starting from 2008, 2009 and 2010 were chosen to be presented into graph and table form due to the recentness and relevancy factors.

Figure 5 below shows location of Kinta River and its stream flow. Kinta River is a sub-catchment of Perak River which is drainage area approximately 2540 km^2 and the stream length is 100km long.

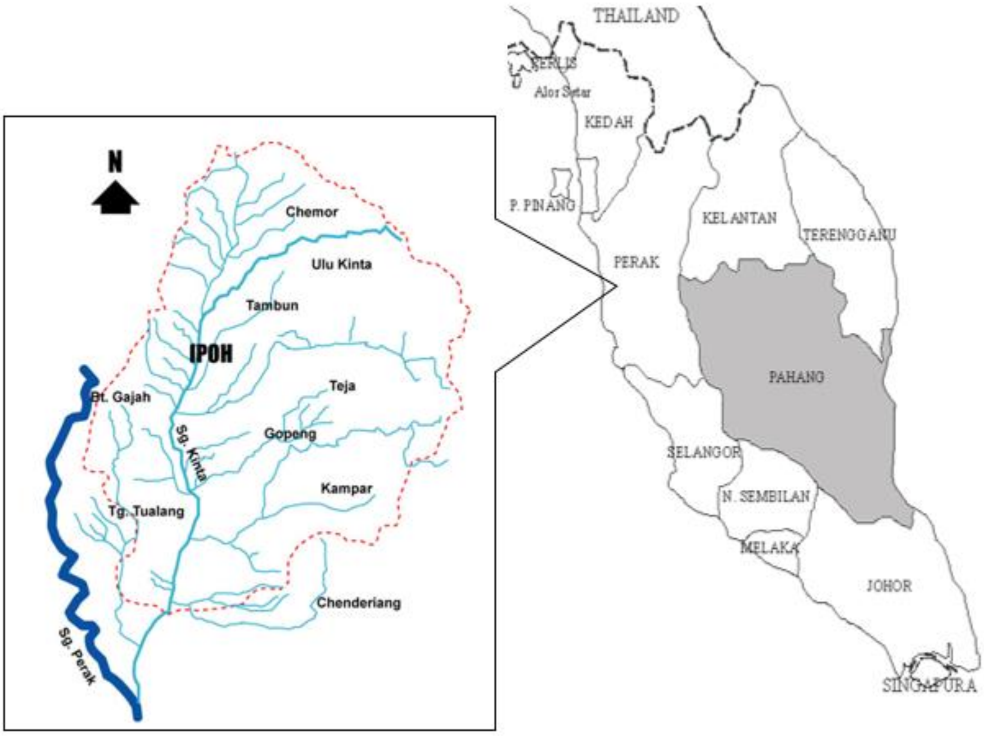


Figure 3.1: Location Map of Study Area, Kinta River (Figure adopted from Fahkaruden, 2014)

3.2 Development of Radial Basis Function (RBF) Model

The RBF model development included input data selection, statistical data analysis and normalization data.

3.2.1 Data Selection

Data collected for water level and discharge at Kinta River was from year 1990 until 2013. For this study, the data used was from year 2008 until 2010 because of the recentness and completeness of the data to prevent skew and scattered profile. In fact, it is essential for the selected data to have a consistent data set since it will affect the accuracy of the end result obtained. In order to get accurate estimation, the data must be adequate and specific for the modelling task. Other than that, input data must be as limited as possible to reduce the training time and possibility of over fitting.

Datasets were divided into two training datasets and testing dataset. For each one of the input variables, the time series was divided in two different subsets. One subset for training the neural network (1January 2008 – 14 December 2009) and one for model testing (1 January 2010- 31 December 2010). Total available data are 1079 and 714 of them was used for training purpose meanwhile 365 used for testing purpose. The figure below is the time series of daily river discharge and water level for training and testing. Data is divided into training and testing by follows condition that all data must be available and consistent and data for training is more than data for testing. Training set is used to adjust the weights on the neural network meanwhile testing set is used only for testing the final solution in order confirm the actual predictive power of the network. After that, training and testing data will analyzed by Matlab software.

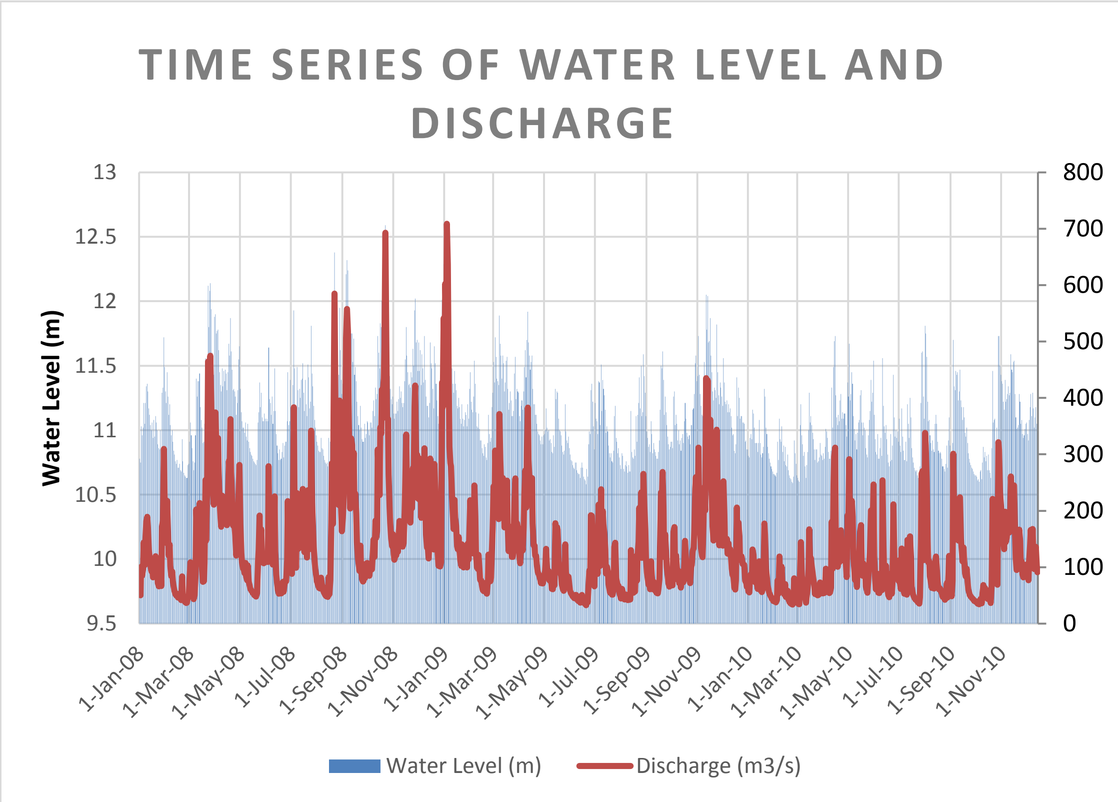


Figure 3.2: Time Series of Water Level and Discharge

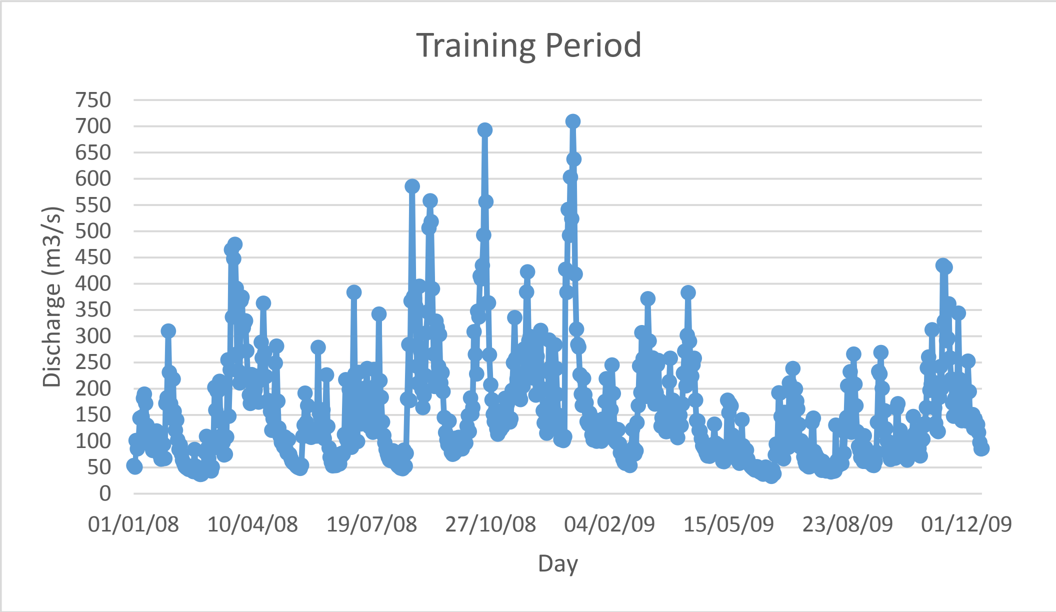


Figure 3.3: Daily Hydrograph of Discharge vs Date for Training

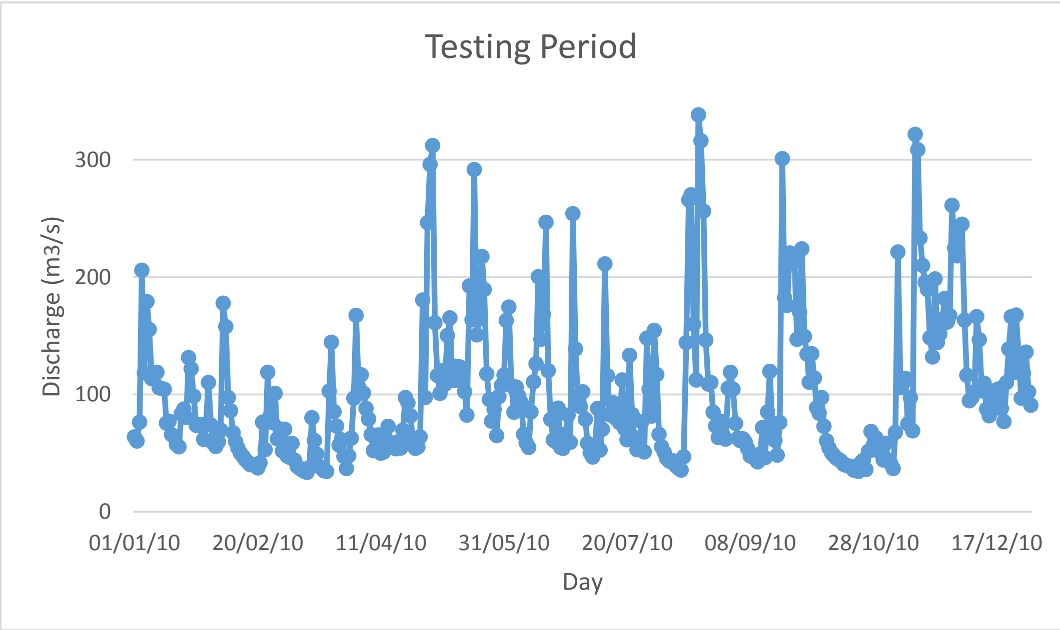


Figure 3.4: Daily Hydrograph of Discharge vs Date for Testing

Based on the daily discharge for training period of the Perak River shown it was found, from the months of July 2008 until February 2009 the discharge has recorded for a vigorous fluctuation trend which is mainly due to the inconsistencies of the rainfall event. However, the discharge value increase drastically from months of May 2009 until August

2009 thus recorded for the minimum discharge value at 32.76 m3/s on the day of 45 before fluctuating again. The value of discharge then record for a gradual decrease along the days after. The hypothesis show that, a smaller marginal difference between the maximum and minimum discharge value will tend to produce more consistent water flow prediction.

3.2.2 Statistical Data Analysis

Parameters	Training		Testing	
	Water Level	Discharge	Water Level	Discharge
Mean	11.2	155.9893	10.99	100.69717
Variance	0.1259	11020.62703	0.0654	3661.47
Standard Deviation	0.35486	104.979	0.2559	60.51
Minimum	10.58	32.76	10.59	33.62
Maximum	12.62	708.91	11.81	338.17

Table 3.1: Summary of Statistical Data Analysis

Statistical parameters involved are Mean, Variance, Standard Deviation (SD), Minimum and Maximum value. Statistical analysis is prepared to determine the complexity of the data, to determine maximum and to compare testing and training data.

The table above shows the summary of the statistical data analysis performed for training and testing data. Mean value for discharge of testing data is lower than training data with difference of 55.31 m^3/s. It means that training has higher river discharge. When compared mean value of water level of training and testing, water level for training is higher than training data with difference of 0.21 m.

Value for mean differences relatively low means that both training and testing data have a relatively constant stream discharge with low fluctuation. As shown in the table above, the mean of the water level for testing in year 2010 is lower compared to the training value. This might happened due to the weather changes and the lower frequency of rainfall during those period. In addition, the value of the maximum water level and discharge for both training and testing was found to be proportionally increased with the increase of the water level value due to the natural phenomenon reaction.

Standard deviation difference for discharge of training data is higher than testing data by 44.46 m^3/s means that the difference quite high. Low standard deviation signifies the distribution of data is converged. Apart from that, having large standard deviation is an indicator that the data may contain no outliers. Whereas, the large difference in maximum value implies the maximum capacity in term of the stream flow in which the area in Kinta River can hold during wet season. Thus, it shows that this river can bear the worst flooding impact due to the heavy rain condition.

3.2.3 Normalization of Data

Normalization of data is vital to ensure for minimization of global error during the network training as mentioned by Rojas (1996). On the other words, Mustafa et al. (2012) claimed that it is a process in which the data set is scaled with the intention of optimize the accurateness of the numerical calculation by reducing redundancy hence minimizes the simulation failure. The formula that is commonly been used to normalize the subsequent data is shown in the equation below.

$$\upsilon_\rho = 2 \times \frac{\chi_\rho - \chi_{min}}{\chi_{max} - \chi_{min}} - 1$$

Where υ_ρ = normalized or transformed data set
χ_ρ = Original data set such that $1 \leq \rho \leq P$ and P = number of data
$\chi_{max},\ \chi_{min}$ = minimum and maximum value of the original data set respectively

The current υ_p symbol represents for the normalized or transformed data series whereas the x_p is the raw data series such that $1 \leq p \leq p$ in which p is the number of data and x_m and x_{max} are the minimum and the maximum value of the original data series respectively which is in this case the data referred to the water level and discharge data series (Mustafa et al., 2012)

3.3 ANN Model Architecture Selection

In this research paper, radial basis function is used as the design model. Thus, there are three layers which are input, hidden and output layer. The layers consist of specific number of neurons that should to be decided in this stage. Maier et al., (2010) stated that selection of suitable figure of neuron in the input, hidden and output has a great consequence on the accuracy of the model structure established.

3.3.1 Input Layer Selection

Identification of the input layer is based on the number of input and the type of input Variables. In this study, there are three input variables and they are current water level,1-antecendent water level, and 2-antecedent water level. The notation for each type of variable is W_t for current water level,W_{t-1} for 1-antecendent water level and W_{t-2} for 2-antecedent water level. The method to carry out the selection is subjected to recommendation from previous research papers.

3.3.2 Kernel

For this study, Multi-Quadric function has been chosen as the kernel of the model.

3.3.3 Spread Coefficient

Default equation in the MATLAB software defined the spread of RBF model. In this study, the calculated spread is 0.89601. The spread values were evaluated through numbers of trial. In this study, the number of hidden layer and spread which created the lowest mean square error (MSE) value was selected as the best optimum criterion for the model architecture.

3.3.4 Hidden Layer Selection

Trial and error is the method used for the process of determination of the hidden layer. This is because this method obtained an effective result for the selection of the optimal number of hidden layer. Figure 9 below shows the correct method on how the hidden neuron is execute. In this trial and error process, layer number is computed by using Microsoft Excel and MATLAB software.

The selected data were input into the Excel sheet as part of the process to enable the selection process. The simulation will run automatically until the basic load graph appear. By entering the fixed value of testing and training data at 365 and 714 data respectively, the desired value of the hidden neuron will be requested. In this research paper, the number of hidden neuron is started with 4 and will increase by one neuron for the subsequent trials. This is mainly because, the hidden value of 4 is the optimal minimum number of hidden neuron to be inserted before the spread value could be identified.

No. of Trial	No. of Neuron in Hidden Layer	MSE	
		Training	Testing
1	4	0.68	12.15
2	5	0.528	10.489
3	6	0.762	14.88
4	7	0.872	16.327
5	8	0.642	12.228
6	9	0.591	9.31
7	10	0.712	8.668
8	11	0.521	11.87
9	12	0.405	8.466
10	13	0.672	13.478
11	14	0.743	10.351
12	15	0.822	15.95
13	16	0.819	7.71
14	17	0.336	4.368
15	18	0.806	8.927
16	19	0.513	13.42
17	20	0.462	8.402
18	50	0.881	12.859
19	100	0.921	18.21

Figure 3.5: Determination of the optimal of neurons in hidden layer using the trial and error approach

Parameter	MSE	
	Lowest Error Value	
Number of Hidden Layer	17	
Stages	Training	Testing
Value	0.336	4.368

Table 3.1: Summary of data obtained for trial and error method

From the table above it was found, the lowest value of Mean Square Error (MSE) produced during the training is 0.336 and 4.368 for testing. Since the number of MSE produced during the testing using 17 number of hidden layer are the lowest among the others, this layer was found to be the best layer for optimum hidden neuron selection to be used inside the radial basis function architecture.

3.3.5 Performances Evaluation Measures

The most commonly employed error measured were the root mean square error (RMSE), the mean square relative error (MSRE), the coefficient of efficiency (CE) and the coefficient of determination (r^2) (Dawson and Wilby, 1999). They claimed that a reliable measure of goodness of fit at the high flows can be produced by square error despite the fact that relative errors are partial towards moderate flows. Based on formula shown below, z_n the observed discharged value and y_n the predicted value for discharged and z bar is the mean of the observed discharged value and N is the total number of observation for the computed error.

$$RMSE = \left[\frac{1}{N} \sum_{n=1}^{N} (z_n - y_n)^2 \right]^{1/2}$$

$$MAE = \frac{1}{N} \sum_{n=1}^{N} (z_n - y_n)$$

$$CE = 1 - \frac{\sum_{n=1}^{N} (z_n - y_n)^2}{\sum_{n=1}^{N} (z_n - \bar{z})^2}$$

Figure 3.6: Example of performance measurement

3.3.6 Output Layer Selection

There is only one output layer for this study of radial basis function using multi quadric function. The output is discharge value with respect to forecast water level. Summary of the RBF model is as follows:

- Spread, = 0.89601
- Kernel function = Multi Quadric function
- Input variables = 3
- Hidden layer = 17 neurons
- Output neuron = 1

As per summarized in the line above, the final result of the model architecture were construct based on the description listed above in order to get the full picture of the network :

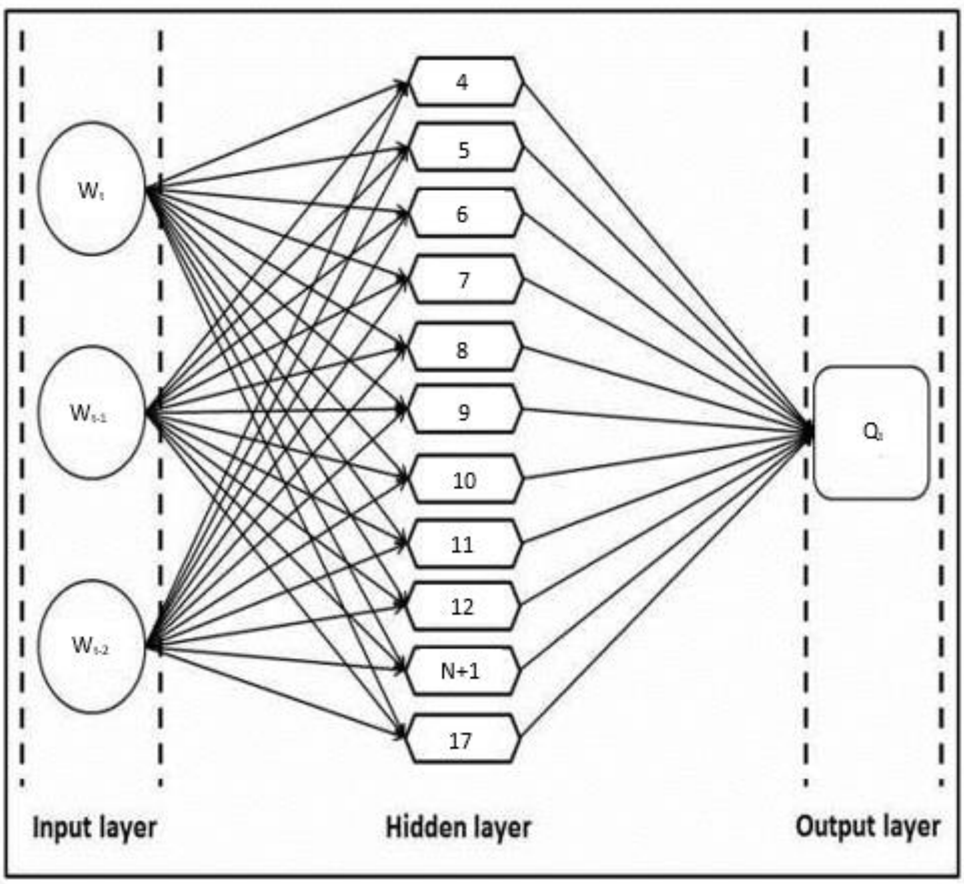


Figure 3.7: Final model of Multi Quadric Radial Basis Function

3.4 Project Flow Activities

Phase 1

Define the topic proposed by supervisor.

Phase 2

Analysed data, choose the consistent and latest data available.

Phase 3

Test the MQ-RBF model.

Phase 4

Design the model using the MATLAB software.

Phase 5

Load the chosen data as input and test the RBF model for training process.

Phase 6

Validate the model using MATLAB software for testing.

Phase 7

Analyse the result and make comparison between the predicted and observed data series.

3.5 Project Key Milestone

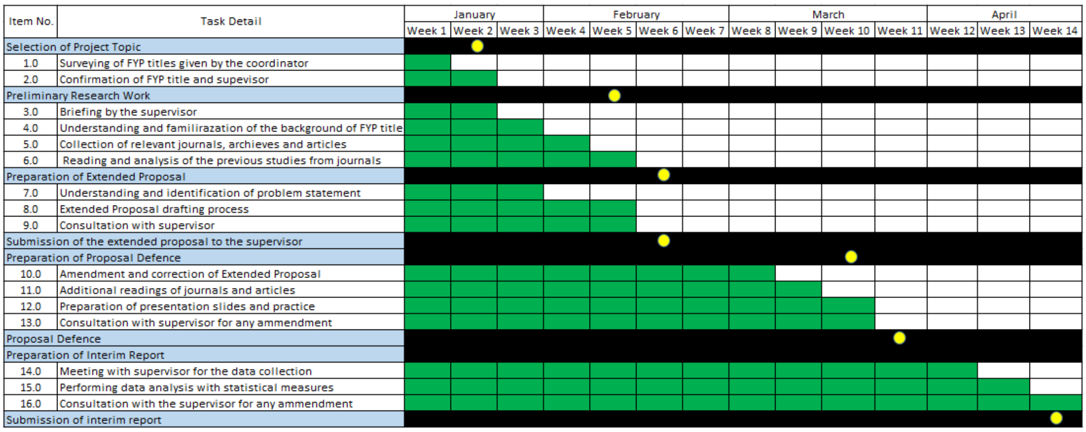


Figure 3.8: FYP 1 Key Milestone

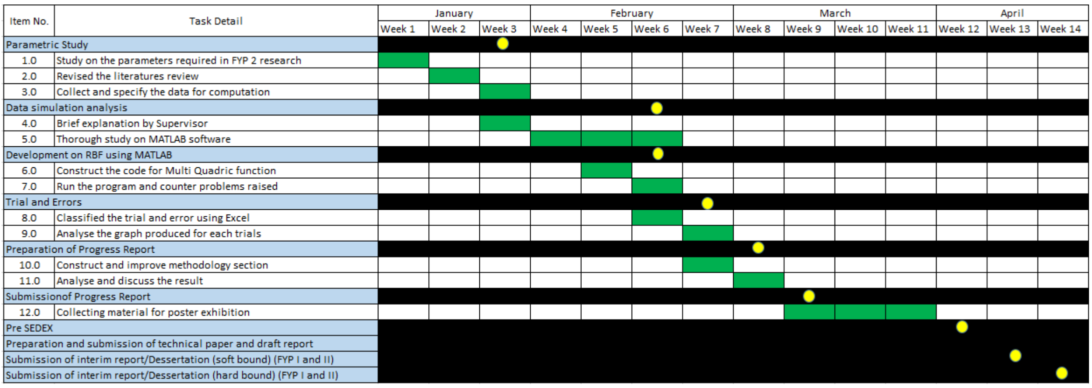


Figure 3.9: FYP 2 Key Milestone

3.6 Project Timeline Gantt chart

No	Detail/Week	WEEK NO.													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Selection of the Project Topic														
2	Preliminary Research Work														
3	Identify and Understanding the Problem Statement														
4	Familiriaze to the Existing ANN Technique and Framework														
5	Submission of the Extended Proposal														
6	Proposal Defense														
7	Developed the Framework and Model using MATLAB Software														
8	Submission of interim draft report fo improvement														
9	Submission of Interim Report														
10	Collect Data	FYP 2													
11	Validate Data														
12	Developed Structured Framework for Troubleshooting														

Figure 3.10: FYP 1 Gantt Chart

No	Detail/Week	WEEK NO.													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Continuation of Project Work														
2	Submission of Progress Report														
3	Continuation and Improvement on Project														
4	Pre-SEDEX														
5	Submission of Draft Report														
6	Submission of Dissertation (Soft Bound)														
7	Submision of Technical Paper														
8	Oral Presentation														
9	Submission of Project Dissertation (Hard Bound)														

Figure 3.11: FYP 2 Gantt Chart

CHAPTER 4:

RESULT AND DISCUSSION

4.1 Statistical Model Analysis

The comparison of the predictive and the observed data between the training and testing is shown in the Figure 16 and 17. R^2 was calculated from the formula stated in Figure 10 by using Excel Spreadsheet. Through the calculation that have been done, the value of coefficient of determination, R^2 of the training data set is 0.981 which is higher than the value of testing data set, 0.941 From this view, it shows that that during training, the model basis function analyzed with a higher precision to the targeted result value since there is less variation to the existing perfect line of agreement. This is mostly because of the system has gained an adequate learning process due to the high numbers of loaded input data and plenty learning time. As found in the graph above, there were some data points which was far from the best fit line. It is because of the high marginal difference between the predicted and observed value thus resulted in lower accuracy of predictive performance for testing and training model.

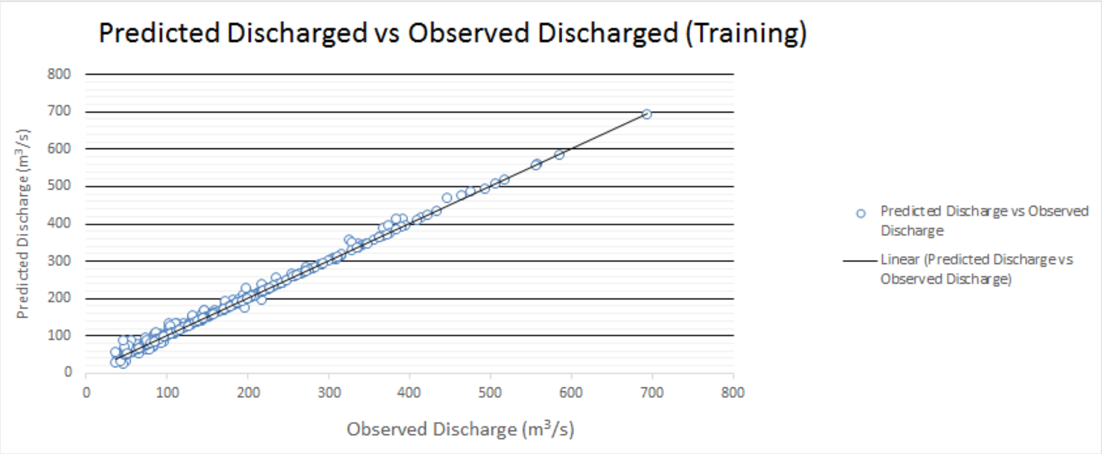


Figure 4.1: Graph of predicted vs observed discharge value for training

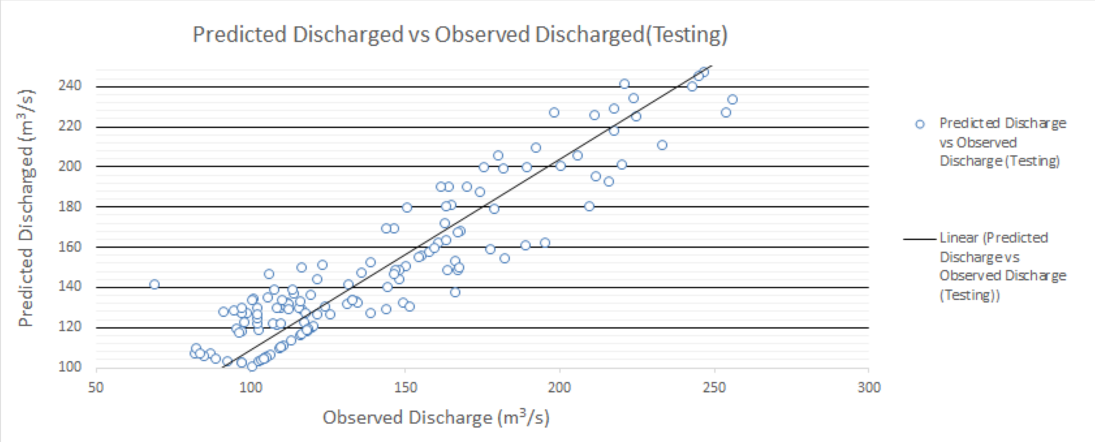


Figure 4.2: Graph of predicted vs observed discharge value for testing

Moreover, as we can see in Figure 17, certain point in testing recorded a value of 247 m3/s for predicted discharged, which is quite high compared to the other values in the data set. Therefore, multi quadric algorithm is found to encounter with a problem to learn with a large magnitude value and thus result in inconsistency of the data along the line of agreement. The same condition also happen to one particular point picked at the observed discharged value at 253.52 m3/s for testing data set. This circumstances might happened due to large marginal difference between the observed and predicted value and also inconsistency in maximum and minimum value in data set which attributed to low accuracy of the model predictive performance later.

However, with the slight difference between the coefficient determination, R^2 between the training and testing model it can be concluded that the RBF model architecture using the multi-quadric algorithm has shown a good agreement with line of perfect agreement and able to forecast the data as close as possible to the observed data.

Training data set in graph below shows crowded data set compared to the testing. This is because of the huge numbers of loaded input data which has been selected for the learning process at 714 data instead of 365 for testing. This is deliberately been done in order to promote a sufficient learning process for the algorithm before the testing could be performed.

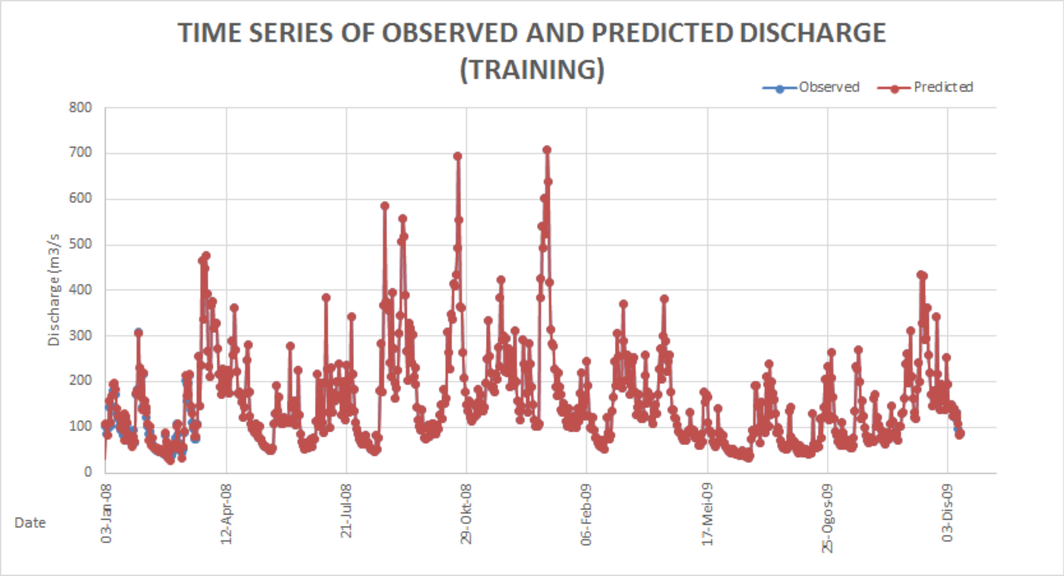


Figure 4.3: Time Series of Observed and Predicted Discharge for Training

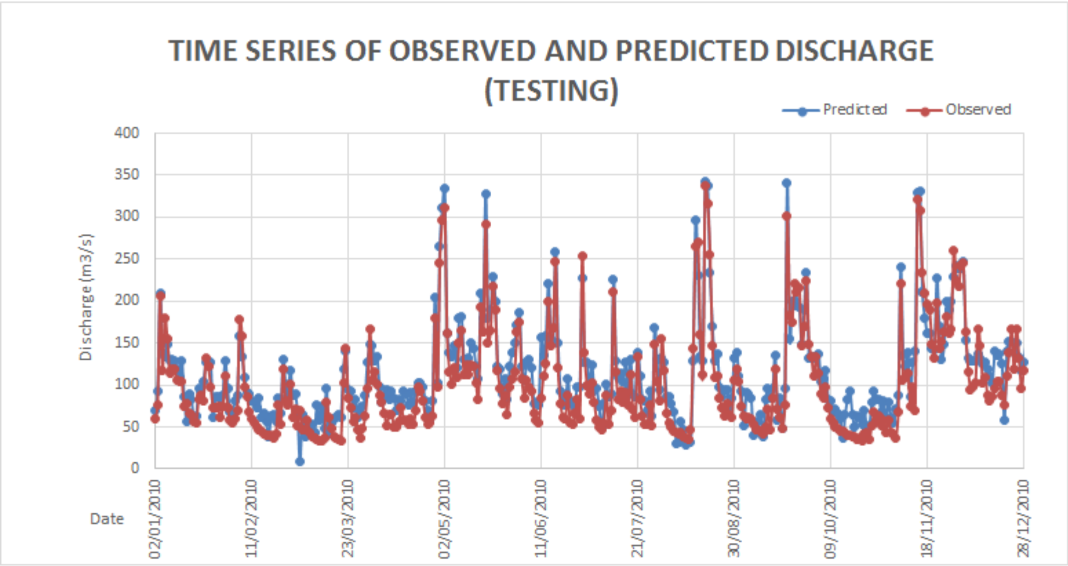


Figure 4.4: Time Series of Observed and Predicted Discharge for Testing

Even though there are tons of data loaded during the training in the network system, the trend shows a very systematic growth and decrement of linear line shape by closely follow the shape of the line in the observed discharged data. Therefore, this recommend that the network system has learned the pattern of water level variation in response to discharged very well during the training process. Apart from this, the application of multi quadric algorithm during testing did performed well which actually showed a good correlation between the observed and predicted value pattern. Henceforth, it show that the network of Radial Basis Function using the multi quadric basis function could generalize at its best function when subjected to different surrounding.

4.1 Statistical Performance Measure Analysis

Table 4.1 : Statistical analysis of the model performance

Data Set	RMSE	R^2
Training	2.810	0.981
Testing	2.108	0.941

The table above demonstrated the simplified form of the result obtained for each parameters involved for both testing and training data set. The analysis of the model performance is completed by measuring the basis of error. The error involve in statistical performance measure includes Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) and Coefficient of Efficiency (CE). Indeed, each of these parameters is a very dominant indicator towards the predictive of the overall performance of the developed model. The formula stated in Figure 11 was used for calculating each of these parameters before the result could be interpreted.

Meanwhile result obtained from testing have lower error value compared to testing value thus RBF architecture have standards to be a perfect model in forecasting the discharge flow. Because of well-trained learning development undergo during the training it was found that the values of error in each parameter for both training and testing did not differ much from each other. Nevertheless, a detail analysis regarding the result should be determined first.

From the excel program, it was found that the value for MSE, RMSE and MAE for testing did produce a very well and satisfactory result in predicting the flow discharge of the Perak River. However for the simplification purpose, only RMSE and Coefficient of Determination, R^2 were chose to be presented in the result part. Indeed, Root Mean Square Error is the Root factor to the actual MSE thus this parameters is adequate to evaluate and analyze the performance of the model.

As in the table, the high value of RMSE for the training data is much higher compared to the testing data set due to the size of the error which correlate the predicted with the observed discharged value in the system. Therefore as a result, the squared error basis such as RMSE shows a higher tendency of being dominated by the high magnitude error during the training process. From the table, the value of RMSE recorded for training is much higher at 2.810 compared to the testing at 2.108. Therefore, the training show that the cluster of input inserted into the system is far from the actual mean value obtained thus result of high error magnitude.

In contrast, for the testing it disclosed that the model were certainly forecast the observed data set with a great predictive accurateness due to good correlation between the water level and discharge data used in testing as a consequence stimulate for a minor magnitude of error value as compared to the training. As an alternative, the important of low error measurement for MAE value in testing would demonstrate that there is less absolute error of difference between the predictive and targeted output.

Number of load input value which higher compared to testing resulted in coefficient of determination, R^2 for the training is greater than testing as shown in table. Indeed, as more input is loaded the higher the improvement and the performance of R^2 value due to satisfactory learning process.

CHAPTER 5:

CONCLUSION AND RECOMMENDATION

Throughout this study of prediction of river discharge at Kinta River, it was discovered that multi quadric radial basis function produced acceptable result. The model architecture were generate to accomplish with three input layer namely water level, water level antecedent 1, water level antecedent 2 and 17 number of neuron in the hidden layer with one output neuron of discharge value. Apart from that, the performance of multi quadric basis function was evaluated by using various statistical measures such root mean square error (RMSE), mean absolute error (MAE), coefficient of efficiency (CE) and coefficient of determination (R^2). The result achieved from the two stages of training and testing showed a very remarkable and significant accuracy of predictive performance for testing at 0.981 and 0.941 for training. In a nutshell, the model able to produce a very good relationship between the predicted and the observed discharge value. As a result it can be decided that objective to predict river discharge using multi quadric radial basis function at Kinta River has been achieved. Finally, it is recommended to use multi quadric radial basis function in future to predict for the other hydrological data in the related hydrological field hence provide a precise and consistent data sources for the application in the industry.

REFERENCES

Dawson, C. W., Harpham, C., Wilby, R. L., & Chen, Y. (2002). Evaluation of artificial neural network techniques for flow forecasting in the river yangtze, china. Hydrology & Earth System Sciences, 6(4), 619-626.

Feng, L. H., & Lu, J. (2010). The practical research on flood forecasting based on Artificial neural networks. *Science Direct*, 37, 2974-2977.

Fernando, D. A., & Shamseldin, A. Y. (2009). Investigation of internal functioning of the radial-basis-function neural network river flow forecasting models. Journal of Hydrologic Engineering, 14(3), 286-292

Huo, Z., Feng, S., Kang, S., Huang, G., Wang, F., & Guo, P. (2012). Integrated neural networks for monthly river flow estimation in arid inland basin of Northwest China. *Journal of Hydrology, 420–421*(0), 159-170. doi: http://dx.doi.org/10.1016/j.jhydrol.2011.11.054

Jain, S. K., & Chalinsgaonkar, D. (2000). Setting Up Stage-Discharge Relations Using ANN. *Journal of Hydrologic Engineering*, 5(4), 428-433

Kasiviswanathan, K. S., & Agarwal, A. (2012). Radial Basis Function Artificial Neural Network: Spread Selection. International Journal of Advanced Computer Science, 2(11).

Kişi, Ö. (2007). Streamflow Forecasting Using Different Artificial Neural Network Algorithms. Journal of Hydrologic Engineering, 12(5), 532-539. doi: 10.1061/(ASCE)1084-0699(2007)12:5(532)

Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environmental Modelling & Software, 25(8), 891-909.

Moharrampour, M., Eskandari, M. R., Rahimi, H., Ghafouri, S. R., & Abad, M. R. A. A. (2012). Predicted Daily Runoff Using Radial Basic Function Neural Network RBF. *Advances in Environmental Biology*, *6*(2), 722-725.

Moradkhani, H., Hsu. K. L., Gupta, H. V., & Sorooshian, S. (2003). Improved streamflow forecasting using self organizing radial basis function artificial neural network. *Journal of Hydrology*, 295, 246-262.

Mustafa, M., Rezaur, R., Saiedi, S., Rahardjo, H., & Isa, M. (2012). Evaluation of MLP-ANN Training Algorithms for Modeling Soil Pore-Water Pressure Responses to Rainfall. Journal of Hydrologic Engineering, 18(1), 50-57. doi: 10.1061/(ASCE)HE.1943-5584.0000599

Ruslan, F. A., Samad, A. M., Zain, Z. M., & Adnan, R. (2013, November). Modelling flood prediction using Radial Basis Function Neural Network (RBFNN) and inverse model: A comparative study. In *Control System, Computing and Engineering (ICCSCE), 2013 IEEE International Conference on* (pp. 577-581). IEEE.

Supharatid, S. (2003). Application of neural network model in establishing a stage-discharge relationship for a tidal river. Hydrological Process, 17, 3085- 3099.

Shamseldin, A. Y. (2010). Artificial neural network model for river flow forecasting in developing country. *Journal of Hydrinformatics*.

Yin, J. C., Zou, Z. J., & Xu, F. (2013). Sequential learning radial basis function network for real-time tidal level predictions. *Ocean Engineering*, *57*, 49-55.

Yu, J., Qin, X., Larsen, O., & Chua, L. (2013). Comparison between Response Surface Models and Artificial Neural Networks in Hydrologic Forecasting. Journal of Hydrologic Engineering, 19(3), 473-481. doi: 10.1061/(ASCE)HE.1943-5584.0000827

APPENDICES

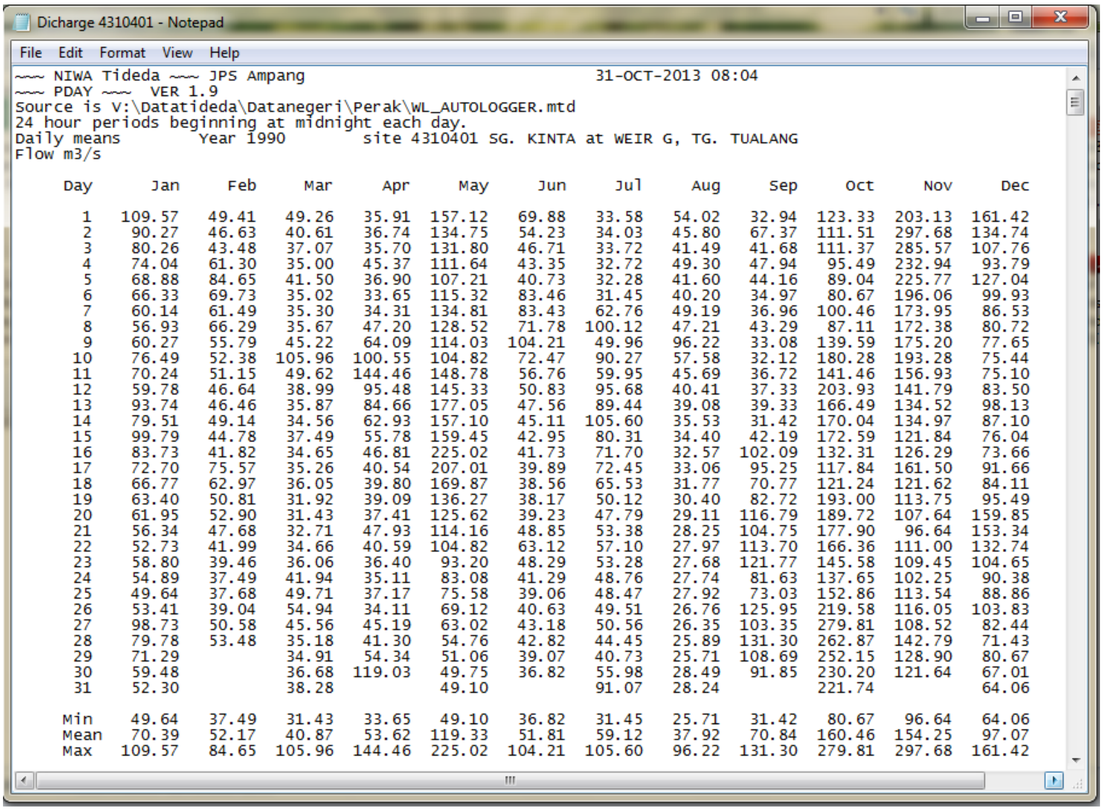
Appendix 1

Appendix 1: Kinta River

Appendix 2

Appendix 2: Residential Area near to Kinta River

Appendix 3



Dicharge 4310401 - Notepad

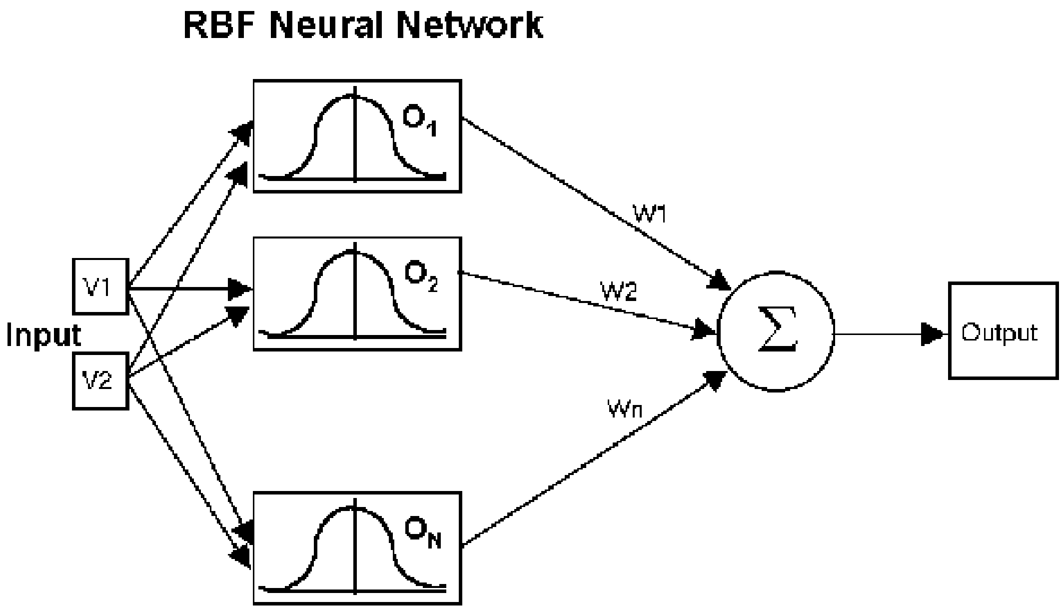
File Edit Format View Help

```
~~~ NIWA Tideda ~~~ JPS Ampang                              31-OCT-2013 08:04
~~~ PDAY ~~~  VER 1.9
Source is V:\Datatideda\Datanegeri\Perak\WL_AUTOLOGGER.mtd
24 hour periods beginning at midnight each day.
Daily means         Year 1990        site 4310401 SG. KINTA at WEIR G, TG. TUALANG
Flow m3/s
```

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	109.57	49.41	49.26	35.91	157.12	69.88	33.58	54.02	32.94	123.33	203.13	161.42
2	90.27	46.63	40.61	36.74	134.75	54.23	34.03	45.80	67.37	111.51	297.68	134.74
3	80.26	43.48	37.07	35.70	131.80	46.71	33.72	41.49	41.68	111.37	285.57	107.76
4	74.04	61.30	35.00	45.37	111.64	43.35	32.72	49.30	47.94	95.49	232.94	93.79
5	68.88	84.65	41.50	36.90	107.21	40.73	32.28	41.60	44.16	89.04	225.77	127.04
6	66.33	69.73	35.02	33.65	115.32	83.46	31.45	40.20	34.97	80.67	196.06	99.93
7	60.14	61.49	35.30	34.31	134.81	83.43	62.76	49.19	36.96	100.46	173.95	86.53
8	56.93	66.29	35.67	47.20	128.52	71.78	100.12	47.21	43.29	87.11	172.38	80.72
9	60.27	55.79	45.22	64.09	114.03	104.21	49.96	96.22	33.08	139.59	175.20	77.65
10	76.49	52.38	105.96	100.55	104.82	72.47	90.27	57.58	32.12	180.28	193.28	75.44
11	70.24	51.15	49.62	144.46	148.78	56.76	59.95	45.69	36.72	141.46	156.93	75.10
12	59.78	46.64	38.99	95.48	145.33	50.83	95.68	40.41	37.33	203.93	141.79	83.50
13	93.74	46.46	35.87	84.66	177.05	47.56	89.44	39.08	39.33	166.49	134.52	98.13
14	79.51	49.14	34.56	62.93	157.10	45.11	105.60	35.53	31.42	170.04	134.97	87.10
15	99.79	44.78	37.49	55.78	159.45	42.95	80.31	34.40	42.19	172.59	121.84	76.04
16	83.73	41.82	34.65	46.81	225.02	41.73	71.70	32.57	102.09	132.31	126.29	73.66
17	72.70	75.57	35.26	40.54	207.01	39.89	72.45	33.06	95.25	117.84	161.50	91.66
18	66.77	62.97	36.05	39.80	169.87	38.56	65.53	31.77	70.77	121.24	121.62	84.11
19	63.40	50.81	31.92	39.09	136.27	38.17	50.12	30.40	82.72	193.00	113.75	95.49
20	61.95	52.90	31.43	37.41	125.62	39.23	47.79	29.11	116.79	189.72	107.64	159.85
21	56.34	47.68	32.71	47.93	114.16	48.85	53.38	28.25	104.75	177.90	96.64	153.34
22	52.73	41.99	34.66	40.59	104.82	63.12	57.10	27.97	113.70	166.36	111.00	132.74
23	58.80	39.46	36.06	36.40	93.20	48.29	53.28	27.68	121.77	145.58	109.45	104.65
24	54.89	37.49	41.94	35.11	83.08	41.29	48.76	27.74	81.63	137.65	102.25	90.38
25	49.64	37.68	49.71	37.17	75.58	39.06	48.47	27.92	73.03	152.86	113.54	88.86
26	53.41	39.04	54.94	34.11	69.12	40.63	49.51	26.76	125.95	219.58	116.05	103.83
27	98.73	50.58	45.56	45.19	63.02	43.18	50.56	26.35	103.35	279.81	108.52	82.44
28	79.78	53.48	35.18	41.30	54.76	42.82	44.45	25.89	131.30	262.87	142.79	71.43
29	71.29		34.91	54.34	51.06	39.07	40.73	25.71	108.69	252.15	128.90	80.67
30	59.48		36.68	119.03	49.75	36.82	55.98	28.49	91.85	230.20	121.64	67.01
31	52.30		38.28		49.10		91.07	28.24		221.74		64.06
Min	49.64	37.49	31.43	33.65	49.10	36.82	31.45	25.71	31.42	80.67	96.64	64.06
Mean	70.39	52.17	40.87	53.62	119.33	51.81	59.12	37.92	70.84	160.46	154.25	97.07
Max	109.57	84.65	105.96	144.46	225.02	104.21	105.60	96.22	131.30	279.81	297.68	161.42

Appendix 3: Raw data obtained from Drainage and Irrigation Department

Appendix 4



Appendix 4: Configuration of Radial Basis Function’s Neural Network