Prediction of Suspended Sediment Concentration in Kinta River Using Soft Computing Techniques

By

Ahmad Safwan Bin Abu Bakar

Dissertation submitted in partial fulfillment of

the requirement for the

Bachelor of Engineering (Hons)

(Civil Engineering)

SEPTEMBER 2013

Universiti Teknologi PETRONAS

Bandar Seri Iskandar

31750 Tronoh

Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Prediction of Suspended Sediment Concentration in Kinta River Using Soft Computing Technique

By

Ahmad Safwan bin Abu Bakar

A project dissertation submitted to the
Civil Engineering Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF ENGINEERING (Hons)
(CIVIL ENGINEERING)

Approved by,

____
(Dr. Muhammad Raza Ul Mustafa)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

September 2013

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.

____

AHMAD SAFWAN BIN ABU BAKAR

ABSTRACT

The prediction of suspended sediment concentration in hyperconcentrated rivers is crucial in modeling and designing hydraulic structures such as dams and water intake inlets. In this study, suspended sediment concentration in Kinta River is predicted using soft computing technique, specifically radial basis function. Suspended sediment concentration and stream discharge from the year of 1992 to 1995 and data from the year of 2009 are used as input. The data are divided into three sections, namely training, testing and validation. 824 data are allocated for training, 313 data are allocated for testing purpose and 342 data are allocated for validation purpose. All data are normalized to reduce error. The determination of input neuron is based on correlation analysis. The number of hidden neurons is determined by the application of trial and error method. As for the output, only one output neuron is required which is the predicted value of suspended sediment concentration. The results obtained from the radial basis function model are evaluated to identify the performance of radial basis function model. Performance of the prediction is measured using statistical parameters namely root mean square error (RMSE), mean square error (MSE), Coefficient of efficiency (CE) and coefficient of determination (R^2). Radial basis function model performed well producing the value of R^2 (0.9856 & 0.9884) for training and testing stages, respectively. However the performance of RBF model in the prediction of suspended sediment concentration for the year 2009 is poor, with the value of R^2 of 0.6934. Recommendations to improve the prediction accuracy are by incorporating a wider data span and by including other hydrology parameters that may impact the changes in the value of suspended sediment concentration

ACKNOWLEDGEMENT

The author would like to express his sincere gratitude to the supervisor, Dr.Muhammad Raza Ul Mustafa; lecturer of Civil Engineering Department, Universiti Teknologi PETRONAS for his guidance throughout this project. Thank you for the continuous supports and motivation.

To the author research partner; Munira Aira binti Malim who has been conducting the study together with the author, the author would like to express appreciation and gratitude for helping along the whole duration of the study.

A heartfelt appreciation is extended to the Department of Irrigation and Drainage Perak, for the assistance given in providing research data for the study.

An earnest appreciation to Universiti Teknologi PETRONAS for the opportunity given to conduct a research study for the author final year project. The author had received an opportune chance in meeting many wonderful people along the way of completing this project.

To my family, thank you for your prayers and moral support. And last but not least, millions of thanks to the lecturers, technicians and friends who have been contributing for this project.

Thank you.

AHMAD SAFWAN BIN ABU BAKAR

Civil Engineering Department

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

River is an important element to humankind. Throughout the history of mankind, rivers are the city of London, United Kingdom. Sedimentation occurs in all rivers as a natural phenomenon and the rate of sedimentation vary with time. The obstruction of the flow of sedimentation could occur due to blockage by natural and man- made structures such as dam. Consequently, sediments are known as the source of provisions and transportation, which led to human establishments. 70% of establishments are located nearby rivers, such as Cairo for the river of Nile, Bangkok of the Chao Phraya River, Baghdad of the Tigris River, Belgrade of the Danube River, Ho Chi Minh City of the Saigon River, Rome of the Tiber River, Moscow of the Moskva River and the famous Thames River that flows through trapped upstream and affected the downstream, especially where the human establishments are. Anthropogenic activities such as agriculture are deeply affected by the blockage of sediments as it carries lots of nutrients for the crops.

In practice, quantification of sediment is difficult and costly. The difference in inflow of sediment and outflow of sediment could assist in estimating the sediments trapped. The quantification of sediment helps the maintenance and operation of civil purposes such as hydropower and irrigation.

A reservoir’s life span relies on the sedimentation rate of the river. The higher the sediments quantity, the water storage capacity of a reservoir would be lesser. Therefore, it is important to estimate the quantity of sediments accurately for proper management of water related projects (Reddy & Ghimire, 2009).

1.1 Problem Statement

Estimation of suspended sediments is important to the field of civil engineering as it would determine the design of a structure (Aytek & Kişi, 2008). Suspended sediments also induce pollution in the river. The common method of estimation is by establishing sediment rating curves to describe the relationship between discharge and sediment concentration ("Dimensionless Bedload and Suspended Sediment Rating Curves," 2012). However, sediment rating curves method is found to be less accurate and often displays errors due to the non-linear behavior of suspended sediment. Therefore, an alternative method is required in order to solve non-linear problem hence accurately estimate and predict the suspended sediment concentration of a river. In this research, the prediction of suspended sediment concentration would be based on radial basis function neural network modeling.

1.2 Significance of the project

It is important to quantify the suspended sediment in order to mitigate any problems related to sediments, such as pollution and structural concerns. However, manual quantification is costly as constant monitoring is required to ensure all data are well measured. It would cover the maintenance of the equipment as well as manpower to monitor the equipment on a regular basis. Failure of constant-monitoring will cause loopholes in the data, for instance inadequate data. To overcome the issue of manual quantification, researchers are adapting to the method of forecasting the data based on the input data. Prediction is done using soft computing technique. The soft computing technique for this study is specified on radial basis function model.

1.3 Objective

The main objective of this study is to predict suspended sediments in Kinta River with the following specific objectives:

1) To develop a radial basis function model for the prediction of suspended sediment concentration in Kinta River
2) To evaluate the performance of radial basis function model using statistical parameters

1.4 Scope of Study

In this research, the scope of study would encompass the following elements:

1) Understanding the mechanism of soft computing in predicting suspended sediments concentration
2) Developing soft computing model using MATLAB software for the purpose of suspended sediments concentration's prediction
3) Assessing the performance of soft computing models in predicting suspended sediments concentration and validating the accuracy of the results

1.5 Feasibility of the project within the scope and time frame

To predict suspended sediments of any river, vital information such as daily mean flow, daily mean rainfall and daily mean sediments are required. Fortunately, these data can be obtained from the Department of Irrigation and Drainage (DID) of Malaysia, which is a convenient move as no fieldwork is required hence reducing the time spent for data collection which allowing more allocation of time for the purpose of data analysis and construction of soft computing model. Kinta River was chosen as the study area because of its nature of being heavily laden with sediments, which makes it a hyperconcentrated river. Hyperconcentrated flow can be described as having high level of suspended sediment concentration and some fine sediment. Being the main river that runs through several major towns of Perak such as Batu Gajah, Pusing, Ipoh and Pasir Putih, the contribution of Kinta River is significant. Infrastructures such as bridges need to consider sediment loading to avoid deposition of suspended sediment that may cause Kinta River to become shallow, hence inducing disaster such as flash floods. Kinta River was chosen as the study subject due to its property of high concentration of suspended sediments. A proper study on Kinta River could assist in managing these water- related structures with an improvement in efficiency.

CHAPTER 2

LITERATURE REVIEW

2.1 Suspended Sediment Prediction Using Soft Computing Techniques

The application of soft computing techniques to estimate suspended sediments concentration is common among researchers nowadays. Soft computing techniques are replacing sediment rating curve and other methods in predicting suspended sediment concentration in rivers. Examples of soft computing techniques are artificial neural network (ANN), gene expression programming (GEP), support vector machine (SVM), adaptive neuro fuzzy inference system (ANFIS) and fuzzy logic (FL). However, the focus of this study is a subsection of ANN, which is radial basis function (RBF). In the prediction of suspended sediment concentration using soft computing techniques, researchers have done numerous studies to evaluate the performance of each model.

Kisi et al. (2012) used GEP to predict the suspended sediment concentration and compares the results with ANN, SVM and ANFIS. GEP showed a better performance compared to ANN, SVM and ANFIS. Determination of the best input combination of GEP model is using statistical analysis. Among 8 input combinations, it was found that the best input combination for GEP model are current stream discharge, one antecedent stream discharge and one antecedent suspended sediment concentration ($Q_t, Q_{t-1}$and S_{t-1}). The best input combination shows highest correlation value (R^2) between suspended sediment concentration and stream discharge and the lowest mean absolute error (MAE) value. However, GEP has several disadvantages. The program size of GEP may increase but not the fitness of the model. Consequently, the program size will cease to stop growing hence causing the interpretation of the program to be difficult.

2.2 Suspended Sediment Prediction Using Artificial Neural Network (ANN)

The inspiration to develop ANN came from biological neural system, or the brain. The neural system has a vast amount of neurons that are able to execute tasks better, in comparison with modern day high speed computer. Real neurons transmit signals to other neurons. These signals are transmitted over biased or weighted connection. The similarity of function between real neuron and artificial neuron is significant.

The evolution of ANN into other algorithm is based on three key elements. The key elements are as follow:

- Arrangement of neuron
- Selection of training paradigm
- Connections

The evolution has created many other derivation of ANN. The most important are multilayer perceptron (MLP) and radial basis function (RBF). The basic architecture of ANN is as figure 2.1.

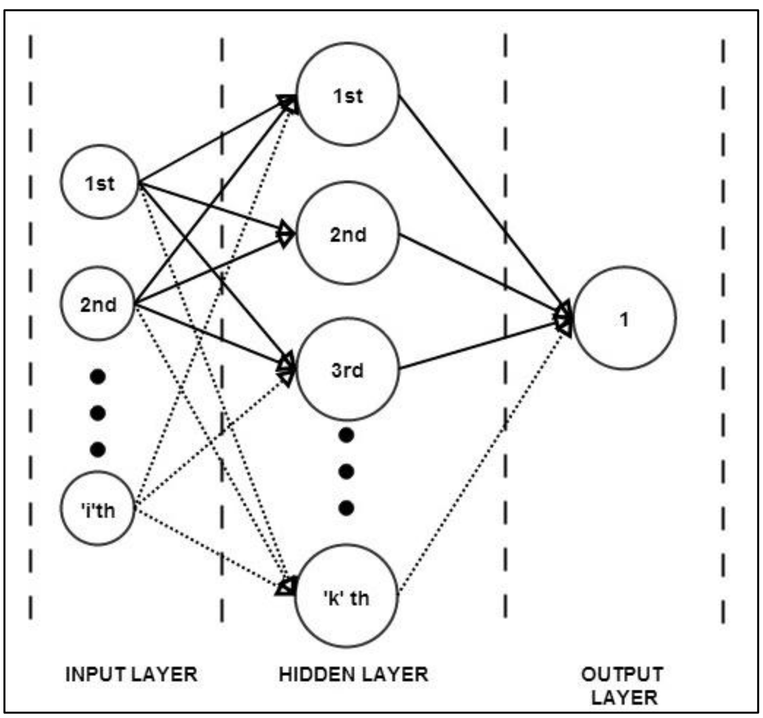

Figure 2.1: Basic architecture of ANN.

ANN basically has 3 main components, which are input layer, hidden layer and output layer. The input layer has i^{th} neuron, the hidden layer has k^{th} neuron and the output layer, in the context of this study is 1 output, which is the suspended sediment concentration. The hidden layer of ANN is not limited to one layer only, but could expand more. The most important derivations of ANN algorithm are radial basis function and multilayer perceptron. However, ANNs don't have exact formulae or equation to simulate the prediction process, unlike conventional method that use formulae and equation with multiple parameters. Due to this reason, ANNs are considered as black box models, where the input determines the output produced (Kisi et al., 2012).

Similar to MLP and RBF, the computation of weight to produce the expected outcome from the given input is done by neuron in hidden layer. The computation is done by activation functions. Among the common activation functions are tangent sigmoid, Gaussian and polynomial.

2.3 Suspended Sediment Prediction Using Multilayer Perceptron (MLP)

An MLP has one or more hidden layers in its architecture. Hidden neurons are also known as computation nodes have the purpose of arbitrating the external inputs and the output of the network. MLP is able to execute complex situation by adding hidden layer. The application of MLP in civil related issues, particularly water and hydrological aspect is quite a focus nowadays. In the prediction of suspended sediment concentration, MLP is proven to be accurate. The basic architecture of MLP is similar to ANN, with the same allowance of hidden layer expansion. The expansion of layer, or multilayer, allows the MLP model to work on more complicated task. The ability to solve highly complicated task comes with a drawback which is longer computational time, compared to single hidden layer.

Feyzolahpour et al. (2012) uses MLP in predicting suspended sediment concentration. The study area is Givichay River that is located in Iran. Due to the high sediment yield rate, Givichay River is an example of hyperconcentrated rivers. Mustafa et al. (2012) also applied MLP in predicting the suspended sediment concentration of Pari River, which is located in Silibin, Perak, Malaysia. MLP also being applied by Khalilabad et al. (2009) in forecasting river suspended sediment yield in Bar River, Neyshaboor, Iran. Bar River is selected as study area due to the location of the river which is in arid and semi- arid basin. The estimation of suspended sediment concentration in semi- arid and arid basin is important due to the complicated erosion and sedimentation problem.

It is important to choose the input data. Input data could affect the complexity during training session (Mustafa et al., 2012). Feyzolahpour et al. (2012) uses 3 input of MLP network which are Q_t, Q_{t-1} and S_t. Q is stream discharge, S is suspended sediment concentration and t is the expected day. These parameters may affect the yield of suspended sediment concentration. Based on the correlation coefficient between suspended sediment concentration and stream discharge, Mustafa et al. (2012) determined the input layer to have 3 neuron which are $Q_t, Q_{t-1}$and Q_{t-2}. Khalilabad et al. (2009) uses trial and error method and determined the input to be 3 neuron.

The hidden layer is the most important part of MLP architecture. Khalilabad et al. (2012) has 5 neurons in the hidden layer, with the arrangement of the neuron as 3-1-1. Where there are 3 neurons in the first hidden layer, 1 neuron in second hidden layer and 1 neuron in third hidden layer. Trial and error method is used to determine the 3 neuron in hidden layer for the MLP model (Mustafa et al., 2012). Feyzolahpour et al. (2012) uses trial and error method to justify the selection of three neuron of the hidden layer.

The similarity between Khalilabad et al. (2009), Feyzolahpour et al. (2012) and Mustafa et al. (2012) is the output layer. The output layer for all 3 papers is the predicted suspended sediment concentration, S_t.

To analyze the performance of the develop MLP model, statistical parameter analysis should be conducted. Khalilabad et al. (2009) and Feyzolahpour et al. (2012) employed root mean square error (RMSE) and correlation coefficient (R^2). From their analysis, MLP in general is able to predict suspended sediment concentration. The predicted and observed data show high value of coefficient of determination ($R^2 = 0.88$) (Khalilabad et al., 2009). Feyzolahpour et al. (2012) found that the best result for their MLP prediction model is R^2 of 0.90003 and root mean square error (RMSE) of 330 mg/L, which is relatively low compared to the data range. Mustafa et al. (2012) compared the performance of difference MLP training algorithms. The best training algorithm in predicting suspended sediment concentration using MLP is Levenberg- Marquadt (LM). The RMSE is considerably the lowest for both training and testing stage, with RMSE of 47 as the best performing algorithm.

2.4 Suspended Sediment Prediction Using Radial Basis Function (RBF)

Radial basis function is another derivation of ANN, apart from MLP. Unlike MLP, RBF only has a single layer of hidden neuron. The advantage of having a single hidden layer is fast converging time, which could be translated as less computational time and produce results faster.

In predicting suspended sediment concentration, RBF has been widely used among researchers and new derivations have been made to improve accuracy. Feyzolahpour et al. (2012) develops RBF model to be compared with neural differential evolution (NDE) and MLP. The input is 3, which are Q_t, Q_{t-1} and S_t. Using the trial and error method, the best number of neuron for hidden layer is 17. The spread of the RBF model is 0.39 and determined by trial and error method as well. The output is 1 which is the predicted value of suspended sediment concentration, S_t. The performance of RBF is judged by RMSE and R^2. The RBF model performs better than MLP with RMSE of 318 mg/L and R^2 of 0.9114.

Trial and error method is a common method in determining the best number of neuron in hidden layer. In comparing the performance of RBF and MLP in predicting suspended sediment concentration, both RBF and MLP employs the method (Memarian & Balasundram, 2012). The number of neuron of hidden layer for RBF is 20 and 30 for MLP. The first layer of hidden layer has 20 neurons while the remaining 10 neurons is in the second layer of MLP's hidden layer. The dataset of the study are stream discharge and suspended sediment concentration. Based on the result analysis, MLP performs better than RBF as MLP is more capable of following the changing pattern of daily suspended sediment concentration. However, RBF has a faster convergence time compared to MLP hence RBF is able to produce prediction faster. The mean square error (MSE) for MLP and RBF is 274089.00 and 281938.38, respectively.

Aydin & Eker, (2012) compared two learning rule of RBF which are Quickprop (QP) and Delta-bar-Delta (DBD). The learning rules were paired with transfer functions which are linear-tangent-hyperbolic-axon (litanhaxon) and tangent-hyperbolic-axon (tanhaxon). The number of neuron in hidden layer is kept as a constant with the value of 1. After the learning rules with transfer functions being analyzed, it was found that the difference in learning rule and transfer function is insignificant. Both of the learning rule with different transfer function yield relatively similar result. Table 2.1 displays the summary of literature reviews that are being used as reference in this study.

Table 2.1: Summary of literature review

No	Title	Author	Year	Methodology/ Findings	Results
1	Daily suspended sediment load prediction using ANN and SVM	E. K. Lafdani, A. M. Nia & A. Ahmadi	2013	-Data used: streamflow, suspended sediment, rainfall -Employs ANN with 3 inputs, determined using correlation analysis.	ANN has higher accuracy in prediction than SVM
2	Modeling the daily suspended sediment concentration in hyperconcentrated river on the Loess Plateau, China using Wavelet-ANN (WANN) approach	Q. J. Liu, Z. H. Shi, N. F. Fang, H. D. Zhu & L. Ai	2013	-Data used: streamflow, suspended sediment concentration, rainfall. -partitioning of data based on dry and rainy season -Input data: 3, determined using correlation analysis -6 neurons in hidden layer determined using trial & error method, with activation function of tangent sigmoid.	WANN model produces predictions with high accuracy. Further research could be done incorporating precipitation and vegetation.
3	Investigation and evaluation of ANN in Babolroud River suspended load estimation	E. Kia, A. R. Emadi & R. Fazlola	2013	-ANN input determined by trial and error method, MLP input determined by correlation analysis. -MLP has 1 hidden layer with 6 neuron & determined by trial & error method	-SRC underestimates SSC, MLP predictions are closer to the observed value -MLP predicts better than SRC & RBF.
4	Sediment estimation study using ANN for Karaj Dam reservoir in Iran	M. Salimi, Y. Hassanzadeh, R. Daneshfaraz & M. Salimi	2013	-Employs ANN model with 3 input, determined using correlation analysis. -3 neuron in hidden layer, determined using trial & error method. Activation function of hidden layer is tangent sigmoid.	ANN is suitable in predicting large data.
5	Comparison between MLP and RBF networks for sediment load estimation in a tropical watershed	H. Memarian & S. K. Balasundara m	2012	-Employs MLP with 30 hidden neuron. -Employs RBF with 20 neuron in hidden layer. -Kernel Function: Gaussian function -Activation function of RBF is Hyperbolic Tangent and Logistic for MLP's 1^{st} and 2^{nd} hidden layer.	MLP performs better than RBF

6	Prediction of daily suspended sediment load using radial basis function neural network	A. Aydin & R. Eker	2012	-Hidden neuron is constant with the value of 1. -Kernel Function: Gaussian function	All combination perform well with an insignificant difference between them
7	Estimating suspended sediment concentration using NDE, MLP and RBF	M. Feyzolahpour, Rajabi & R Shahram.	2012	-Employs MLP with 3 inputs, 3 neuron in 1 hidden layer and 1 output. -Employs RBF with 3 input, 17 neuron in hidden layer and 1 output., spread of 0.39 -Kernel Function: Gaussian function -Determination of hidden neuron for both MLP and RBF by using trial and error method.	RBF performs better than MLP
8	Estimating river suspended sediment yield using MLP neural network in arid and semi- arid basins	H. M. Khalilabad, S. Feiznia & K. Zakikhani	2009	-Employs MLP with 3 input neuron, 5 hidden neuron with the arrangement of 3 neuron in the first hidden layer and 1 neuron for the remaining second and third layer. The output layer is 1 neuron.	MLP model able to predict value and obtained sediment rating curve
9	River suspended sediment prediction using various multilayer perceptron neural network training algorithm	M. R. Mustafa, R. B. Rezaur, S. Saiedi & M. H. Isa	2012	-Employs MLP with 3 input, 3 hidden neuron and 1 output. -Hidden neuron determination is using trial and error method. -Compares the performance of different training algorithm	Levenberg-Marquadt is the best training algorithm for MLP.
10	Suspended sediment modeling using genetic programming and soft computing techniques	O. Kisi, A. H. Dailr, M. Cimen & J. Shiri	2012	-ANN model uses 3 input, 1 hidden layer and 1 output. Hidden layer activation function is logarithm sigmoid.	-ANN estimations has negative values, unlike GEP that has all positive predictions. -GEP predicts better than ANN
11	Modeling of suspended sediment concentration at Kasol in India using ANN, fuzzy logic and decision tree algorithms.	A.R.S. Kumar, C.S. P. Ojha, M. K. Goyal, R. D. Singh & P. K. Swamee	2012	-ANN input data selection using correlation analysis. Best ANN structure is 6 neuron in input layer, 4 neuron in hidden layer and 1 output. -Optimum structure for RBF is 6 input, 2 neuron in hidden layer and 1 output.	-RBF predicts better with more hidden neurons. -ANN predicts better than fuzzy logic, RBF and decision tree.

12	Suspended sediment load prediction of river systems: An ANN approach	A. M. Melesse, S. Ahmad, M. E. McClain, X. Wang & Y. H. Lim	2011	-Employs MLP. Best input determined using trial and error method. -The study compares training length, 3 years data for training and 2 years data for testing, and 2 years data for training and 3 years data for testing.	-3 years training data & 2 years testing data produces better estimation than 2 years training data & 3 years testing data.
13	Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models	T. Rajaee, S. A. Mirbagheri, M. Z. kermani & V. Nourani	2009	-Employs ANN with 3 input, 4 neuron in hidden layer and 1 output. Determination of input and hidden neuron is using trial & error method. -Hidden layer activation function is tangent sigmoid function.	-Neuro Fuzzy model performs better than ANN, SRC, multilinear regression and non-multilinear regression
14	Event-based sediment yield modeling using ANN	R. K. Rai & B. S. Mathur	2008	-Input data selection using correlation analysis. -Hidden layer uses tangent sigmoid activation function	ANN is able to model event-based sediment with no computational difficulty
15	Estimation and forecasting of daily suspended sediment data using wavelet-neural networks.	T. Partay & H. K. Cigizoglu	2008	-Input of ANN model determined by wavelet analysis -First application of wavelet analysis in predicting suspended sediment concentration	Wavelet-ANN predicts better than ANN and SRC

Based on the summary above, RBF and MLP prediction model is accurate. Most of the architecture of RBF and MLP consist of 3 input parameters. The performance of the prediction models are measured in statistical analysis. The common parameters being used are RMSE, MSE, MAE, CE and R^2.Trial and error method is proven to be an effective method in determining the hidden neuron and RBF spread.

The formula for each statistical analysis parameters is as below.

Mean Square Error (MSE) = $\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{N}$

Root Mean Square Error (MSE) = $\sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{N}}$

Coefficient of Determination (R^2)= $\frac{[\sum_{i=1}^{n}(P_i - \overline{P})(O_i - \overline{O})]^2}{\sum_{i=1}^{n}(P_i - \overline{P})^2 . \sum_{i=1}^{n}(O_i - \overline{O})^2}$

Nash- Sutcliffe Coefficient of Efficiency (CE) = 1-$\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2}$

Mean Absolute Error (MAE) = $\frac{1}{n}\sum_{i=1}^{n}|O - P|$

The notation of observed value, predicted value, number of data, mean of observed data and mean of predicted data are O, P, n, $\overline{O}$ and $\overline{P}$, respectively.

CHAPTER 3
METHODOLOGY

3.1 Study Area and Data Source

The research data are obtained from the Department of Irrigation and Drainage (DID) for the river of Sungai Kinta. Sungai Kinta is located in the state of Perak. Figure 3.1 shows the location of Kinta river catchment which is nearby the main river of Perak River.

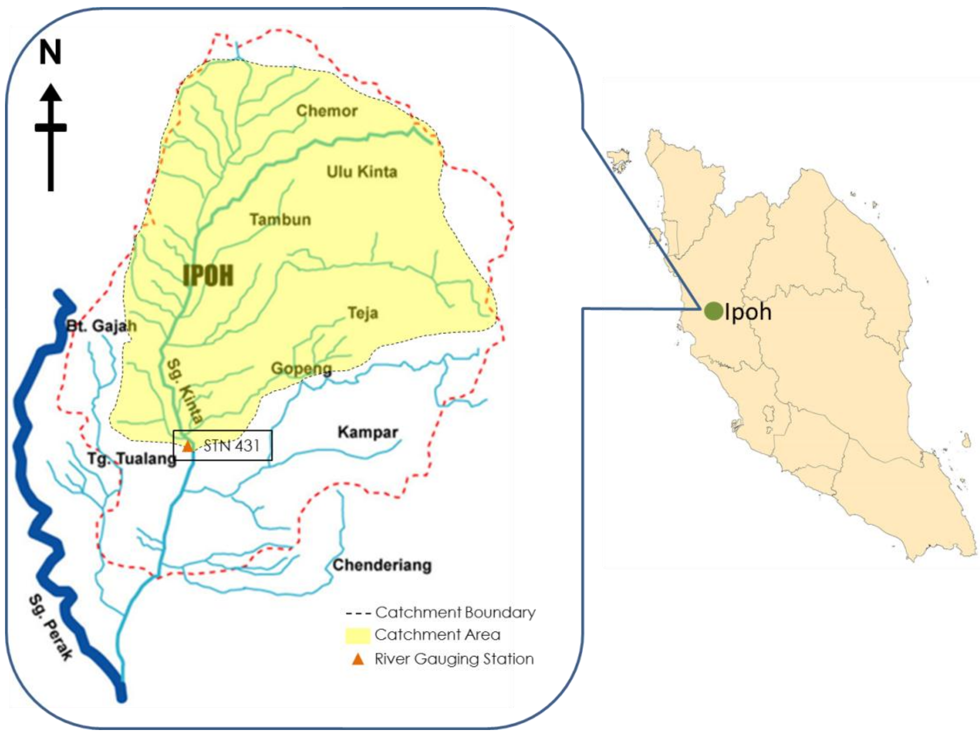

Figure 3.1: Kinta River catchment area

Kinta River has the catchment area of 2540 $km^2$and the major contributors for Kinta River are Pari River, Raia River and Kampar River. The measuring station, Station 431, is located nearby Tanjung Tualang and indicated on the figure 3.1. The geography of the Kinta River catchment in the north and east side comprises high degree of steepness mountains that are covered with forest while in the southern part of the Ipoh lies Kinta Valley, where the usage of land is concentrated for human activities. Formerly, Kinta Valley was famous for its tin mining activities which resulted to formation of ponds and lakes around the area (Ghani, A. A., 2007). Throughout the year 1992 to 2009, many developments were carried out on areas nearby Kinta River.

3.2 Development of RBF Model

Based on the general structure of RBF, there are 3 main layers to be considered namely input layer, hidden layer and output layer. Apart from that, the spread coefficient of the RBF model needs to be determined as well. Prior to the development of RBF model, the data need to be properly selected, partitioned and normalized to reduce the complexity of the learning process of RBF model, hence providing prediction with high accuracy.

3.2.1 Selection of Data

Data of suspended sediment concentration and stream discharge of Kinta River are from the year of 1992, 1993, 1994, 1995 and 2009. These years were chosen because of their recentness and the most completed data available from the DID. Having a complete data is essential in establishing a pattern to be identified by RBF model, in order to produce a highly accurate estimation. Loopholes or missing data will resulted to high skewness and scattered data, which consequently increasing the complexity of the learning process of RBF model.

3.2.2 Partitioning of Data

The partitioning of data for training and testing was based on the data trend. Based on figure 3.2, there are gaps between the years 1992 – 1993 and 1994-1995. However, the gap between 1994- 1995 is larger than 1992- 1993. Hence it was decided that data prior to 1995 were used for training and data for 1995 were used for testing purpose. Data in between 1st January 1992 until 7th September 1994 for both stream discharge and suspended sediment concentration were used for training. Data in between 1st January 1995 until 30th December 1995 were used for testing. Data for the year 2009 were used to examine the efficiency of RBF model in predicting suspended sediment concentration for a long gap.

Hence, data in between 1st January 2009 and 18th December 2009 were used for validation stage.

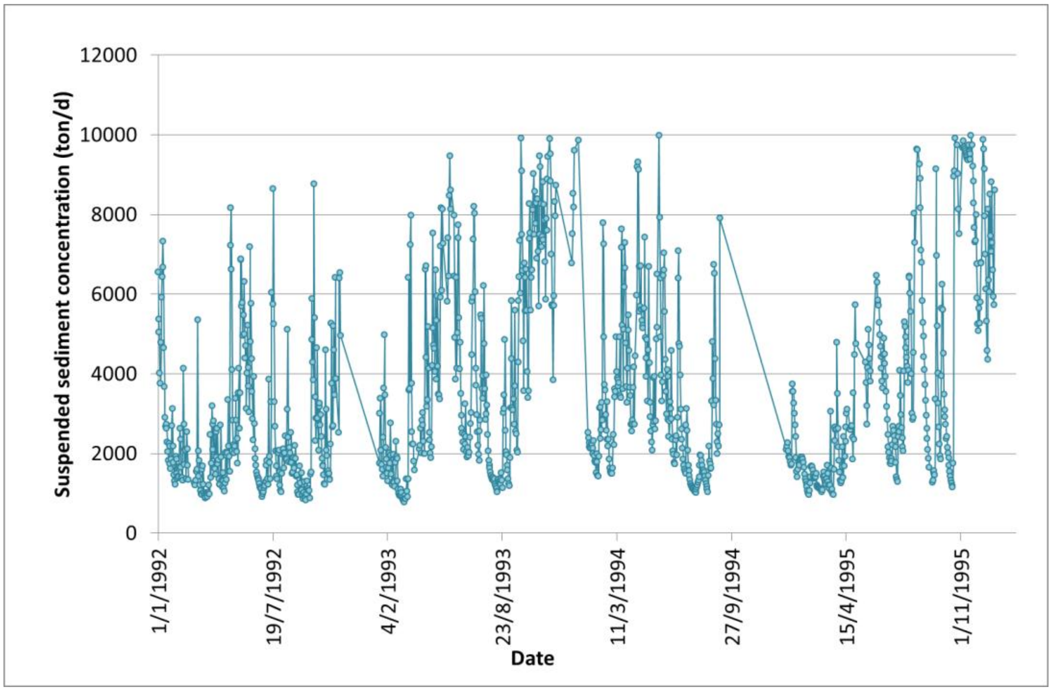

Figure 3.2: Time series of the whole data

Total available data are 1137 and 824 of them were used for the purpose of training and the remaining 313 will be used for testing purpose, for both stream discharge and suspended sediment concentration. 348 data from the year 2009 were used as validation data. The time series of daily suspended sediment concentration for training and testing data is displayed in figure 3.3 and 3.4.

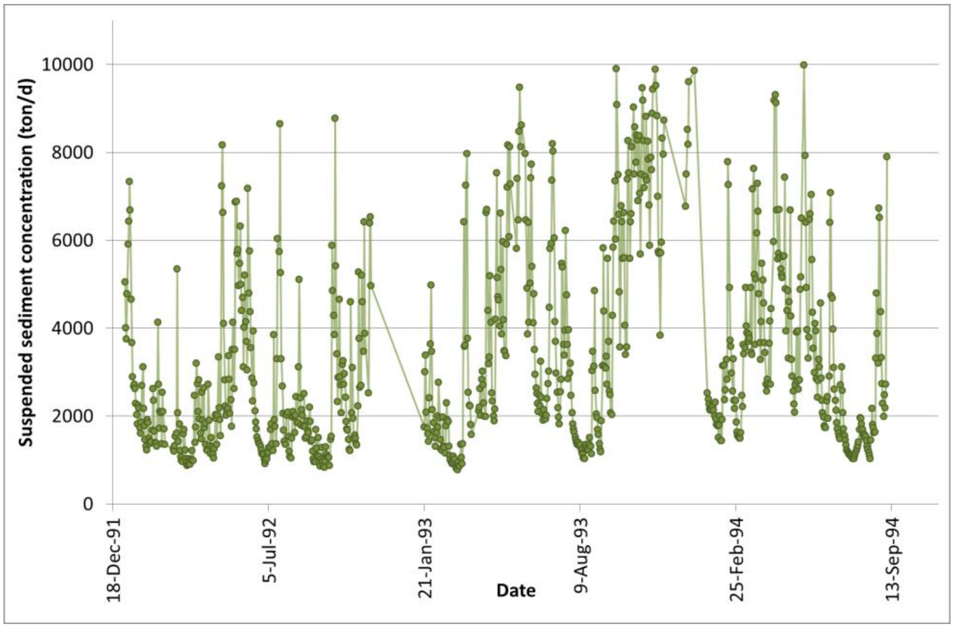

Figure 3.3: Time series of training data after partitioning

It can be seen from figure 3.3 that the suspended sediment concentration of Kinta River from the year 1992 to 1994 mostly distributed below 4000 ton/day. However, Data for the year 1995 in figure 3.4 were used as testing data. Figure 3.4 show a fluctuating pattern with small scale reading as the beginning. During the fourth quarter of the year 1995, the magnitudes of suspended sediment concentration are mostly larger than 70% of the whole data.

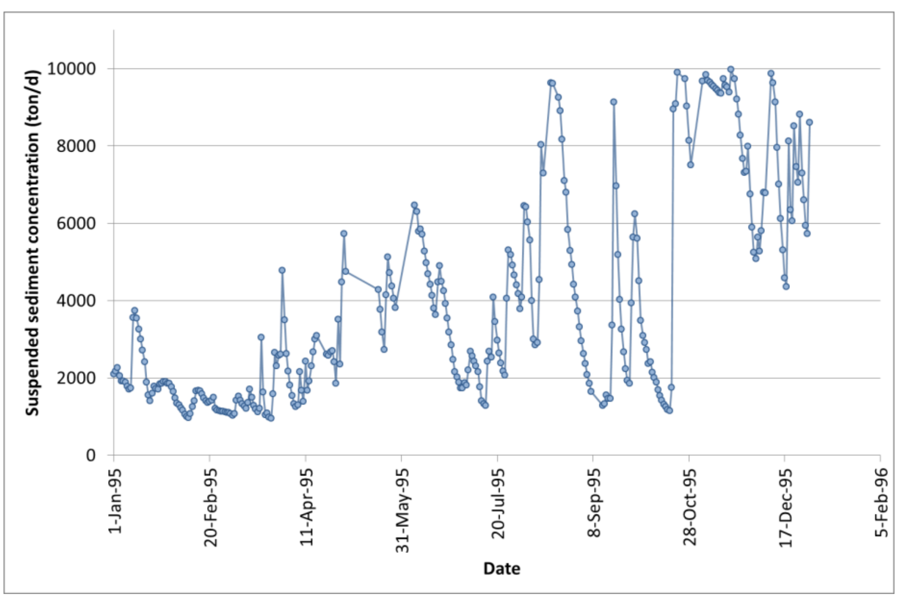

Figure 3.4: Time series of testing data after partitioning

3.2.3 Data Analysis

Table 3.1: Statistical parameters of the applied data set

Data Set	Unit	Xmax	Xmin	Xmean	Xsd	Csx
Training	Q (m3/s)	183.41	35.29	78.67	32.68	1.019
	SSC (ton/d)	9986.00	780.00	3322.77	2211.18	1.066
Testing	Q (m3/s)	183.21	39.77	86.96	39.91	0.977
	SSC(ton/d)	9974.80	959.80	3860.76	2663.47	0.933
Validation	Q (m3/s)	708.91	32.76	138.73	94.29	2.268
	SSC(ton/d)	45634.86	234.77	7349.91	6330.94	2.268

Statistical parameters included in the data analysis are maximum and minimum value, mean, standard deviation ($\mathrm{s_d}$) and coefficient of skewness ($\mathrm{C_{sx}}$).

Data analysis is important as to foresee any challenges or underlying factors that could affect the accuracy of RBFNN model. Table 3.1 above shows the statistical parameters of training and testing data. For the daily stream discharge, the maximum value of training and testing data is relatively close, with a slight difference of $0.2m^3/s$. However, for the minimum value, the difference between training and testing data is relatively larger than the difference between maximum values which is $4.48m^3/s$. This implies the maximum capacity, in term of streamflow, which Kinta River can hold during wet season is around $183m^3/s$. The river keeps flowing during dry season even with low discharge. Apart from that, the minimum discharge of Kinta River for the testing data has increased. This could be interpreted as Kinta River having modification to allow for larger capacity. Considering that areas around Kinta River are surrounded by human establishment, having larger capacity of discharge could reduce the possibility of flooding.

The difference between the mean of training and testing data is $8.29m^3/s$, which is relatively low. Low mean difference signifies that both training and testing data have a relatively constant stream discharge with low fluctuation. The standard deviation of training and testing data is quite large. Large standard deviation means the stream discharge data for both training and testing are not concentrated around the mean, which is $78.67m^3/s$ for training and $86.96m^3/s$ for testing. Large standard deviation also signifies the distribution of data is scattered. Apart from that, having large standard deviation is an indicator that the dataset may contain outliers.

The skewness for both training and testing data is quite large; this implies that the discharge data are mostly above the average or mean value for both training and testing data. Similar to standard deviation, large skewness indicate a scattered distribution of data (Fazlola et. al., 2013).

As for suspended sediment concentration, the difference between maximum value of training and testing data is very small, which is 11.20 ton/day compared to the maximum value of 9986.00 ton/day. The minimum value for training and testing data however, shows a slightly large difference with the value of 179.80 ton/day. This corresponds to the significant difference in minimum daily stream discharge of Kinta River. Similar to stream discharge, the mean difference between training and testing for suspended sediment concentration is small with the value of 8.29 ton/day. The standard deviation however, for both training and testing data is large. The skewness for both testing and training is small, close to normal distribution. For the data of training and testing for both suspended sediment concentration and stream discharge, the positive value of skewness indicating the datasets are skewed to the right-hand-side. This is due to the mode of the datasets are consist of high value of stream discharge and suspended sediment concentration. In predicting suspended sediment concentration using radial basis function, low skewness is important as high value may affect the performance of the radial basis function negatively (Liu, et al., 2013). This is due to the increased complexity of the model's learning process as no significant pattern can be identified.

3.2.4 Normalization of Data

From the aspect of computer science, data normalization is important to represent the data in their unique form or commonly known as standard form. The formula that is used in this research to normalize the data is as equation below.

$$V_p = \frac{X_p - X_{min}}{X_{max} - X_{min}}$$

The current normalized data is denoted as V_p, X_p is the current original data, X_{min} denotes the minimum value of the whole data and X_{max} denotes the maximum value of the whole data. The data in the context of this study are the suspended sediment concentration and stream discharge value. Normalization of data is important in ensuring a fast learning process of RBF model, hence producing estimation in a short time. In this study, the data were normalized between -1 and 1.

3.2.5 Input Layer

The determination of input layer of RBF model depends on the number of input and the type of input variables. There are 3 inputs variables for this study and they are current stream discharge, 1–antecedent stream discharge and 2–antecedent stream discharge. The notation for each variable is Q_t for current stream discharge, Q_{t-1} for 1–antecedent stream discharge and Q_{t-2} for 2–antecedent stream discharge. The determination of these input variables was based on the recommendation of previous research papers. The method used in determining the number of neuron in input layer is using correlation coefficient analysis.

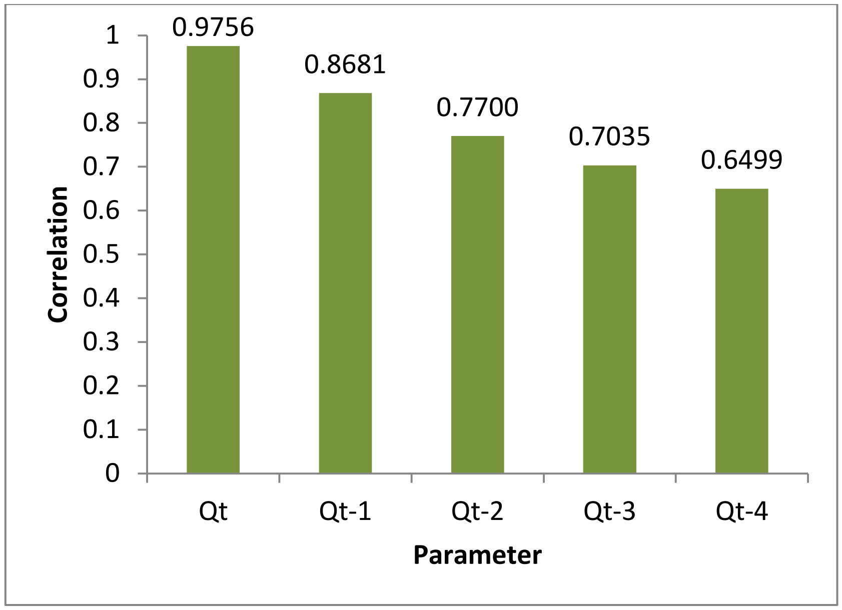

Figure 3.5: Correlation analysis of input variables

Based on figure 3.5, parameter that has the highest correlation with suspended sediment output is Q_t with the value of 0.9756, followed by Q_{t-1} and Q_{t-2} with the value of 0.8681 and 0.7700, respectively. Only three parameters were considered

3.2.6 Kernel

For this study, thin plate spline function has been chosen as the kernel of RBF model.

3.2.7 Spread Coefficient

The spread of RBF model was determined by using the default equation in the MATLAB software. In this study, the calculated spread, σ, is 0.8077.

3.2.8 Hidden Layer

The method to determine the number of neuron in the hidden layer is trial and error method. Table 3.2 shows the details of the trial and error method in determining the number of neuron in hidden layer.

Table 3.2: Determination of number neuron in hidden layer using trial and error method

No. of trial	No. of neuron in hidden layer	MSE	
		Training	Testing
1	4	0.064498	0.091086
2	5	0.059653	0.142820
3	6	0.039682	0.076834
4	7	0.023830	0.031443
5	8	0.050598	0.089422
6	9	0.008477	0.015718
7	10	0.003533	0.009489
8	11	0.007717	0.021315
9	12	0.005804	0.008810
10	13	0.026173	0.003802
11	14	0.002874	0.003991
12	15	0.002655	0.004497
13	16	0.000875	0.001654
14	17	0.001762	0.003744
15	18	0.000847	0.001201
16	19	0.001254	0.003097
17	20	0.001211	0.005003
18	50	0.000064	0.000347
19	100	0.000013	0.000593

MSE denotes statistical parameter of mean square error. For the trial and error method, mean square error is chosen as the criteria to justify the best number of neuron in the hidden layer. The lowest value of MSE provides the best choice. The analysis summary of the trial and error method is shown in table 3.3.

Table 3.3: Analysis of trial and error method

	MSE	
	Training	Testing
Lowest value	0.000013	0.000347
No. of layer in hidden neuron	100	50
Highest value	0.064498	0.142820
No. of layer in hidden neuron	4	5

Based on the trial and error method, the number of neurons that yield the best result, which is the lowest MSE, is 100 neurons for training stage and 50 neurons for testing stage. By right, the number of neuron in hidden layer should be chosen from these two options. However, during the testing stage of both 50 and 100 neurons in hidden layer, both shows a higher MSE compared to training stage. This could be due to the phenomenon of “over fitting”. Over fitting problem is a common issue of neural network modeling because of high number of hidden neuron. The neural network models, for example RBF, have the tendency to smoothen the curve of the time series plot.

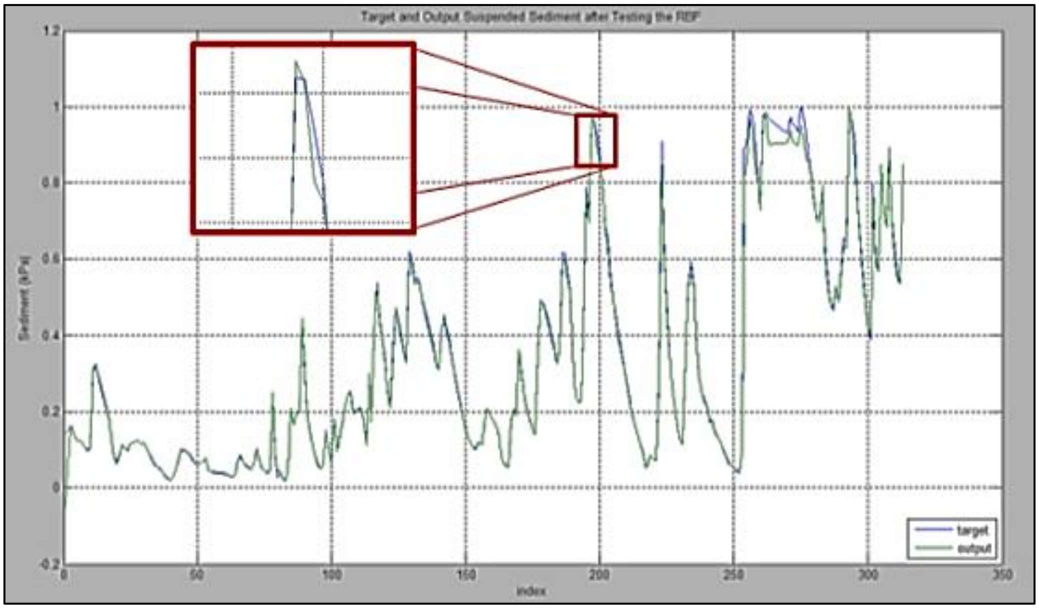

Figure 3.6: Time series of testing of RBF model with a close up of overfitting

Figure 3.6 above displays the time series of RBF model for testing of 100 neuron. Figure 3.7 is the snapshot of the result of trial and error method which was done using MATLAB software. The vertical axis is the suspended sediment concentration and the horizontal axis. The blue chart indicates the observed values while the green chart is the predicted values of suspended sediment concentration. The RBF model smoothen the curve and fails to follow the pattern of the observed values. In between 4 to 20 number of neuron, 18 neuron shows the lowest MSE value for both training and testing phase. Hence the number of hidden neuron is 18.

3.2.9 Output Layer

There is only one output layer for RBF model. The output is the current predicted suspended sediment concentration, which is S_t. S denotes suspended sediment concentration and t denotes the current time. Summary of the RBF model is as follows:

- Spread, σ = 0.80777
- Kernel Function = Thin plate spline
- Input Variables = 3 (Q_t, Q_{t-1} and Q_{t-2})
- Hidden Layer = 18 neurons
- Output Neuron = 1(S_t)

The architecture of RBF model is as figure 3.7.

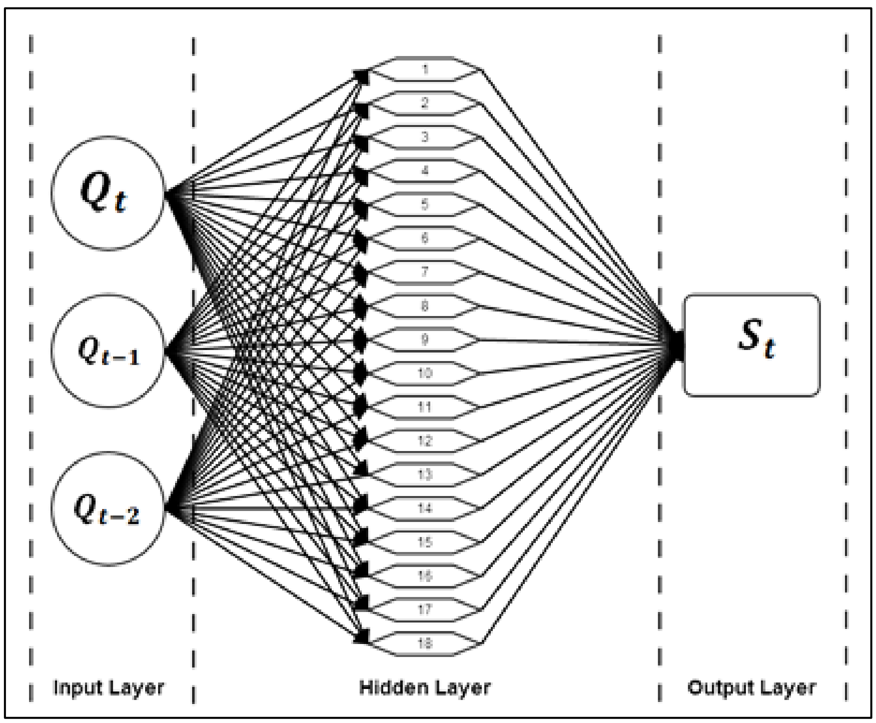

Figure 3.7: Architecture of RBF model

The three input variables, Q_t, Q_{t-1} and Q_{t-2}, are connected to each of the neurons in hidden layer, therefore producing only one output, which is the predicted suspended sediment concentration at present, S_t. The results obtained from the output were analyzed and presented in the form of table and chart.

3.3 Project Activities Flow

Below are the steps for the project throughout the FYP 1 and 2 until completion.

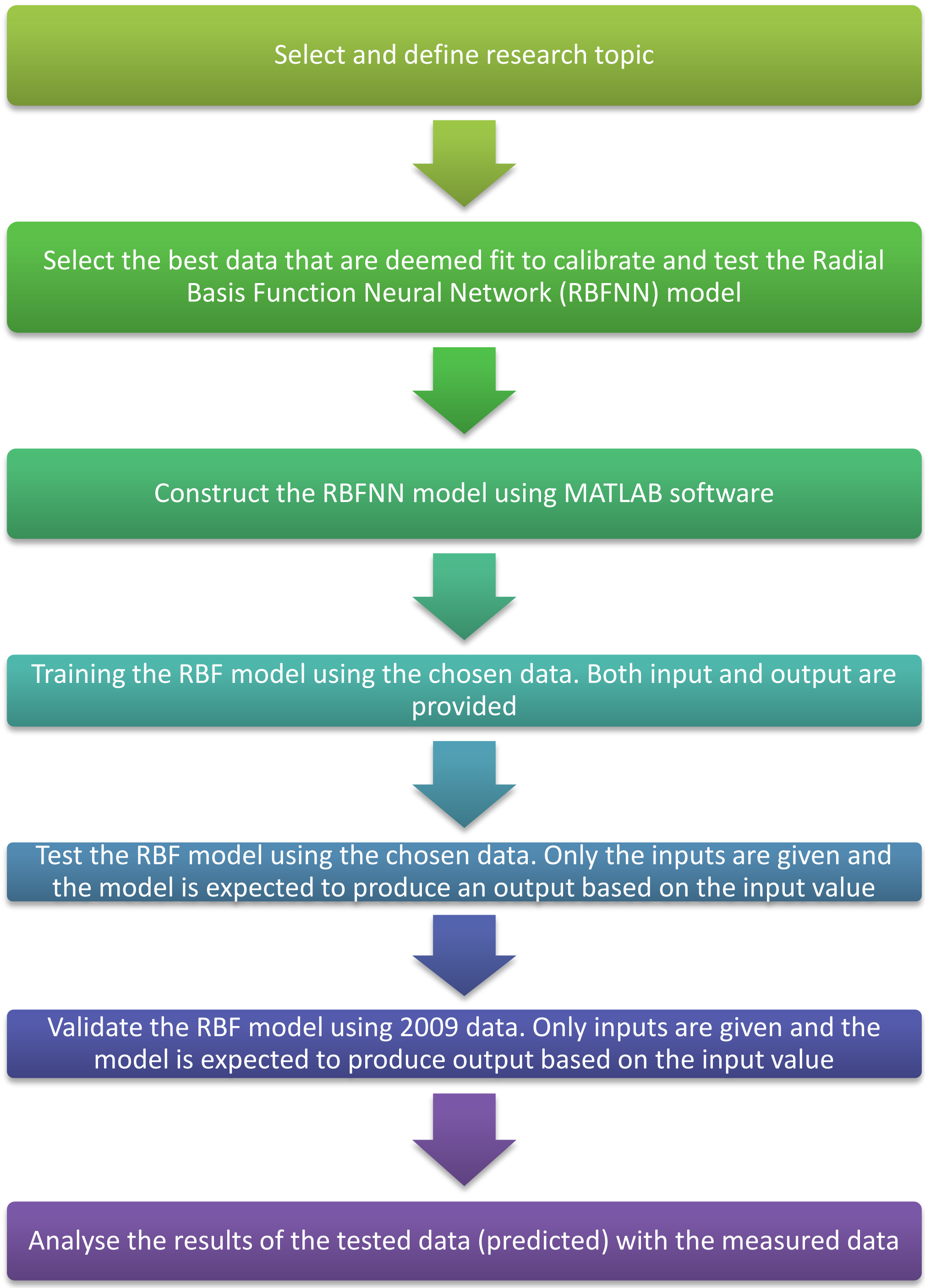

Figure 3.8: Flow of activities

3.4 Key Milestone

Final Year Project (FYP) is divided into 2 sections, namely FYP 1 and FYP 2. Below are the milestones for both FYP 1 and 2.

- Semester 1 (FYP 1)

Table 3.4: Key milestone for FYP 1

Milestone	Week
Project Proposal	Week 1
Extended Proposal (10%)	Week 7
Proposal Defense (40%)	Week 9
Interim Report (50%)	Week 14

- Semester 2 (FYP 2)

Table 3.5: Key milestone for FYP 2

Milestone	Week
Progress Report (10%)	Week 8
Pre-SEDEX (10%)	Week 11
Technical Report (10%)	Week 13
VIVA presentation (30%)	Week 14
Dissertation (40%)	Week 15

3.5 Gantt Chart

Below is the Gantt chart for the whole course of FYP. For FYP 1, the total time allocated is 14 weeks while 15 weeks for FYP 2. FYP 1 would focus on the execution of research and FYP 2 would focus on complete documentation of the research.

Table 3.6: Gantt chart for FYP 1

No	Detail	Week														
		1	2	3	4	5	6	7	Mid sem Break	8	9	10	11	12	13	14
1	Selection of project topic	■	■													
2	Preliminary research work		■	■	■	■										
3	Submission of extended proposal defence						■	■								
4	Proposal defence									■	■					
5	Continuation of project work											■	■	■		
6	Submission of interim draft report														■	
7	Submission of interim report															■

Table 3.7: Gantt chart for FYP 2

No	Detail	Week															
		1	2	3	4	5	6	7	Mid sem Break	8	9	10	11	12	13	14	15
1	Continuation of project work	■	■	■	■	■	■	■									
2	Submission of progress report									■							
3	Continuation of project work									■	■	■	■	■			
4	Pre-SEDEX												■				
5	Submission of draft report													■			
6	Submission of dissertation(soft bound)														■		
7	Submission of technical paper														■		
8	Oral presentation															■	
9	Submission of project dissertation (hard bound)																■

3.6 Tools and Software

No complicated tools and software are required to achieve the objective of the research. The main software that is being used is MATLAB. MATLAB is a programming language and software that is being used for many purposes such as math and computations, algorithm development, data acquisition, modeling, simulation and prototyping, data analysis, exploration and visualization, scientific and engineering graphics and application development, including graphical user interface building. For this project, MATLAB is used for the purpose of developing the soft computing model. Figure 3.9 is the snapshot of MATLAB with an example of algorithm.

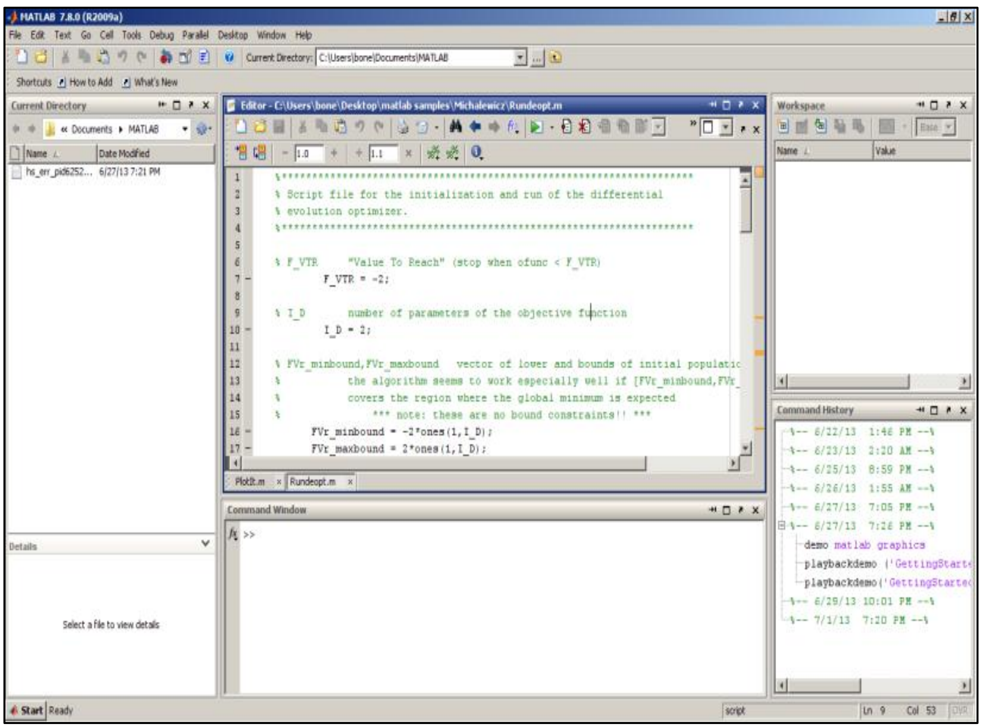

Figure 3.9: Interface of MATLAB software

Apart from MATLAB, the common software such as Microsoft Excel, Microsoft Word and Notepad will be used throughout the research.

CHAPTER 4

RESULT AND DISCUSSION

The analysis of model performance is done in the form of statistical parameters. The statistical parameters involved are mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE).

Table 3.8 below displays the statistical analysis of RBF model performance in predicting suspended sediment concentration in Kinta River.

Table 3.8: Statistical analysis of model's performance

Data Set	MSE	RMSE	MAE	CE	R^2
Training	71592.83	267.57	182.43	0.9852	0.9856
Testing	101758.12	319.00	201.41	0.9845	0.9884
Validation	8323623.38	2885.07	1050.33	0.1900	0.6934

Based on table 3.8, the MSE, RMSE and MAE values are quite high. This is due to the large range of data, specifically suspended sediment concentration data which are within the highest limit of 10000 ton/day to 100 ton/day as the lowest limit. MSE values are increasing from training, testing to validation data. The MSE for training is 71592.83 ton/day, followed by testing with 101758.12 ton/day and validation data with MSE value of 8323623.38 ton/day.

The same increasing pattern is found in RMSE and MAE. Training data produced lowest RMSE value of 267.57 ton/day, followed by testing data with RMSE value of 319.00 ton/day and validation data with RMSE of 2885.07 ton/day. The lowest MSE value is produced by training data, followed by testing and validation data. The MAE value for training, testing and validation is 182.43 ton/day, 201.41 ton/day and 1050.33 ton/day, respectively.

The most important parameter is Nash- Sutcliffe efficiency coefficient, CE. Predictive performance of hydrological models can be gauge by using CE (Nash & Sutcliffe, 1970).Both datasets of training and testing have significantly high value of efficiency, which are close to 1. There is a decreasing pattern in CE value. Training data has the highest CE, 0.9852, followed by testing data, 0.9845, and the lowest CE value is validation data, 0.1900. Based on the value of CE, it can be deduced that prediction of suspended sediment concentration of validation data is inefficient.Inefficiency in prediction is affected by the presence of large magnitude of stream discharge data. Therefore CE is an indication that the RBF model developed is able to perform prediction with high efficiency. This could be explained by referring to table 3.9.

Table 3.9: Analysis of suspended sediment concentration data

Data	Increment of standard deviation from training data for discharge (%)	maximum value (m^3/s)	minimum value (m^3/s)	Mean (m^3/s)
training	0.00	183.41	35.29	78.67
testing	22.31	183.21	39.77	86.96
validation	188.53	708.91	32.76	138.73

Based on table 3.9, the standard deviation of validation data is 188.53% higher than standard deviation of training data, whereas standard deviation for testing data is only 22.31% higher than training data. The distribution of testing data is still within the limit training data, which enabled RBF model to predict efficiently. This corresponds to the maximum and minimum value of each data. The maximum value of training data for discharge is 183.41 m^3/s and maximum value for testing data is 183.21 m^3/s. Note that the maximum value of training data is higher than testing data. The same pattern also found in the minimum value. The minimum value of training and testing data for discharge is 35.29 m^3/s and 39.77 m^3/s , respectively. The minimum value of training data is much lower than testing data. The maximum and minimum values of testing data indicate that the testing data are still within the limit of training data. However, validation data shows a significantly larger maximum value than training data, which is 708.91m^3/s. Validation data has the lowest minimum value of discharge, with 32.76 m^3/s.

The mean for the dataset increases from training with 78.67 m^3/s, followed by testing data with the mean of 86.96 m^3/s. Validation data has the highest mean with 138.73 m^3/s. The mean of validation data almost reach the maximum value of training data indicating the range of data is too high for training data. For the prediction to be made accurately, the discharge of validation should fall within the limit of training data, not exceeding it. Clear example of successful prediction is during the testing stage. The data range of testing stage is within the maximum and minimum limit of training data. The developed RBF model could produce highly accurate prediction data during testing stage because all of the testing data is within the limit of training stage. Validation data shows an extreme condition for the learning process of RBF model because most of the data are exceeding the maximum range of training data. To improve the prediction of testing stage, the training data span should be increased.

Another important parameter to measure the prediction ability is coefficient of correlation, R^2. The value of R^2 is in between 0 to 1 with the latter as the best option. From table 3.8, testing data has the highest correlation which is 0.9884. Training data produced a slightly lower value of correlation, which is 0.9856. Validation data has the lowest correlation, which is 0.69334. Nevertheless, both training and testing stage display high correlation between the predicted value and the observed value, compared to validation stage. Testing stage has the highest correlation because the developed RBF model is able to predict suspended sediment concentration close to the observed or measured value. Because of the presence of high value of discharge in validation data, it is difficult for the developed RBF model to establish a pattern to predict higher range of data. This could be improved by training the RBF model with higher range of data.

Plotting of observed suspended sediment concentration against the corresponding observed value is as figure 3.10 and 3.11.

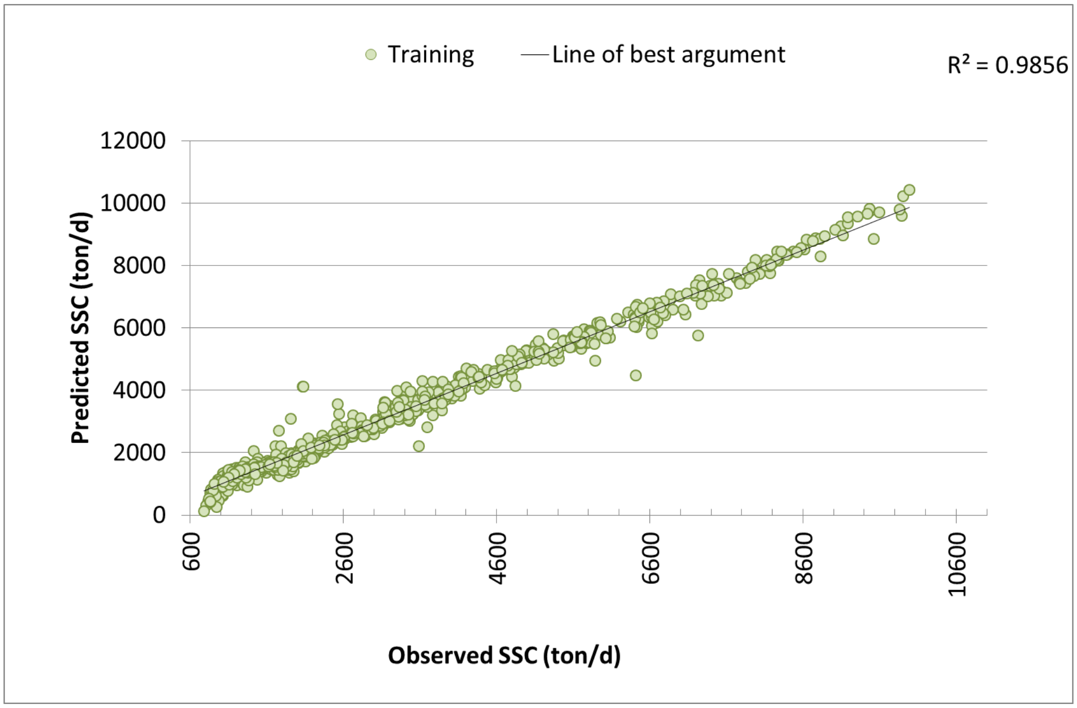

Figure 3.10: Plotting of predicted suspended sediment concentration and observed suspended sediment concentration for training data.

Based on figure 3.10, it can be seen that the outcome, which is the predicted suspended sediment concentration, is highly correlated with the observed value. This could imply that the RBF model developed is able to learn pattern and produce accurate prediction. The presence of outliers in the plot is not that significant, as the deviation of outliers from the best line of agreement is not large. The comparison between the observed and predicted value can be seen clearly in the form of time series. The R^2 value of training data is 0.9856. Time series for training data is as figure 3.12.

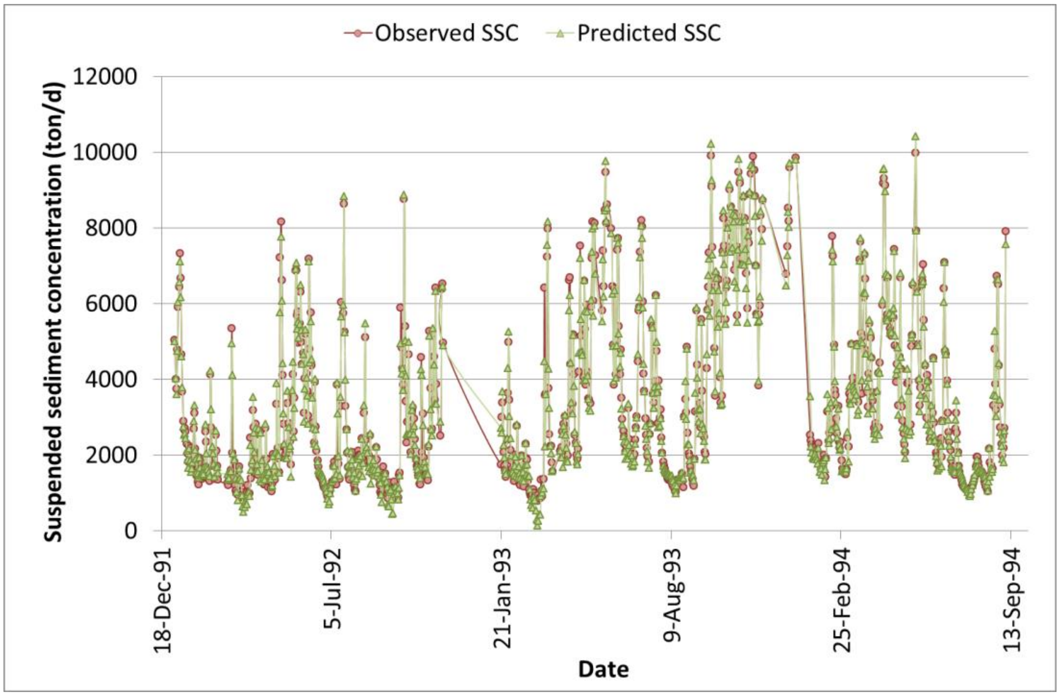

Figure 3.12: Time series of training data of the RBF model.

Note that the predicted values are following the same pattern as the observed values. The high correlation between the observed and predicted data allows the prediction to be made close to the actual and following the pattern of observed data. All spikes and troughs are predicted with high precision as none of the spikes and troughs are predicted as the opposite; pike become trough and vice versa (Hicks, 2011). Increased in spike values of suspended sediment concentration is insignificant throughout the year 1992 to 1995, implying that the geographical elements of Kinta River catchment area remained relatively constant throughout the period.

Testing stage performs better than training stage. Figure 3.11 displays the plotting of observed and predicted suspended sediment concentration for testing stage. This may be due to the large size of data being trained during the training stage. Therefore it provides a better learning for the testing data as the data size is much smaller. The likely outcome should testing data size is larger than training data size would be lower correlation coefficient value as the learning of model is limited by small data size. R^2 value of testing data is slightly higher than training data, which is 0.9884.

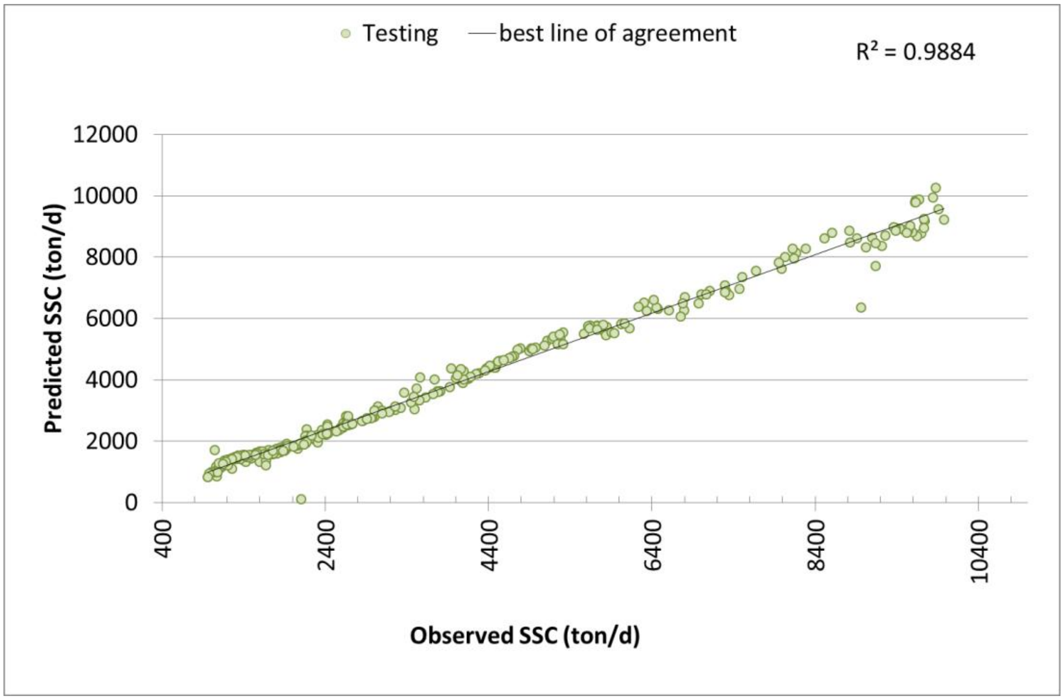

Figure 3.12: Plotting of predicted suspended sediment concentration and observed suspended sediment concentration for testing data.

There are few outliers that can be identified. Most of the outliers exist at the peak of the time series. Figure 3.13 shows the time series of predicted and observed suspended sediment concentration of testing stage.

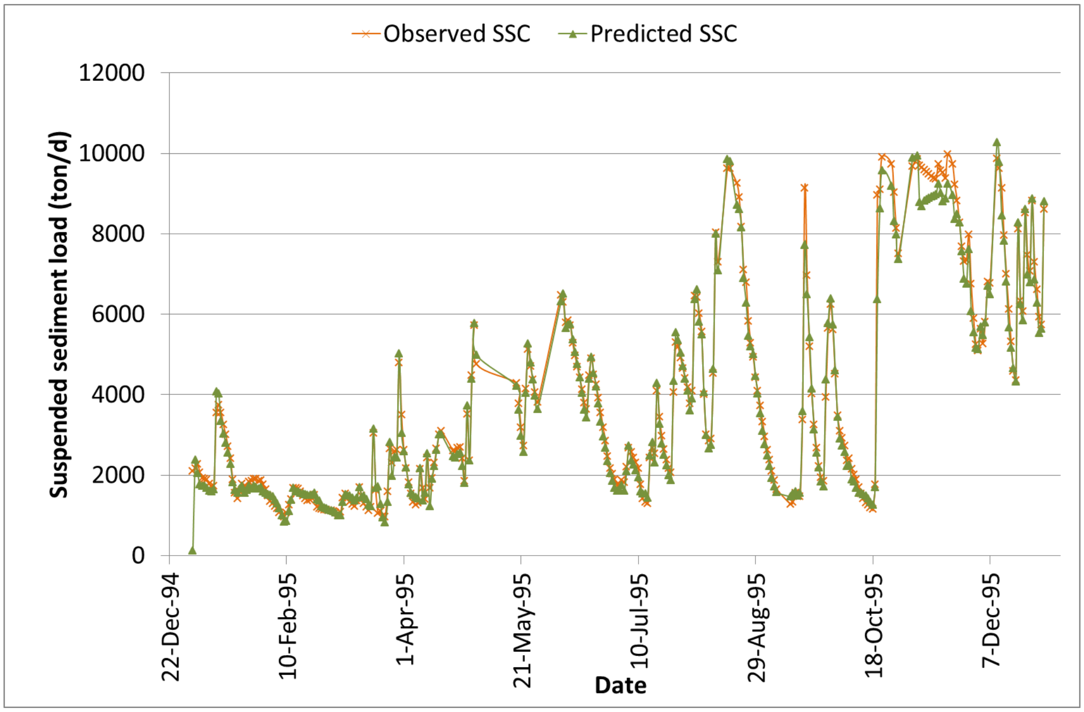

Figure 3.13: Comparison of time series of observed and predicted suspended sediment concentration during testing stage.

Based on the time series, it can be identified that most of the predicted values, specifically at the peaks of graph, did not follow closely the pattern of observed values. This is the part where outliers in the plotting of predicted-observed suspended sediment concentration for testing data occur. Outliers should not be neglected as they could reduce the accuracy of prediction. The impact of outliers in the prediction of suspended sediment concentration can be seen in the plotting of observed and predicted data for the year 2009, which are the validation data. Figure 3.14 displays the plotting of observed and predicted suspended sediment concentration for validation data.

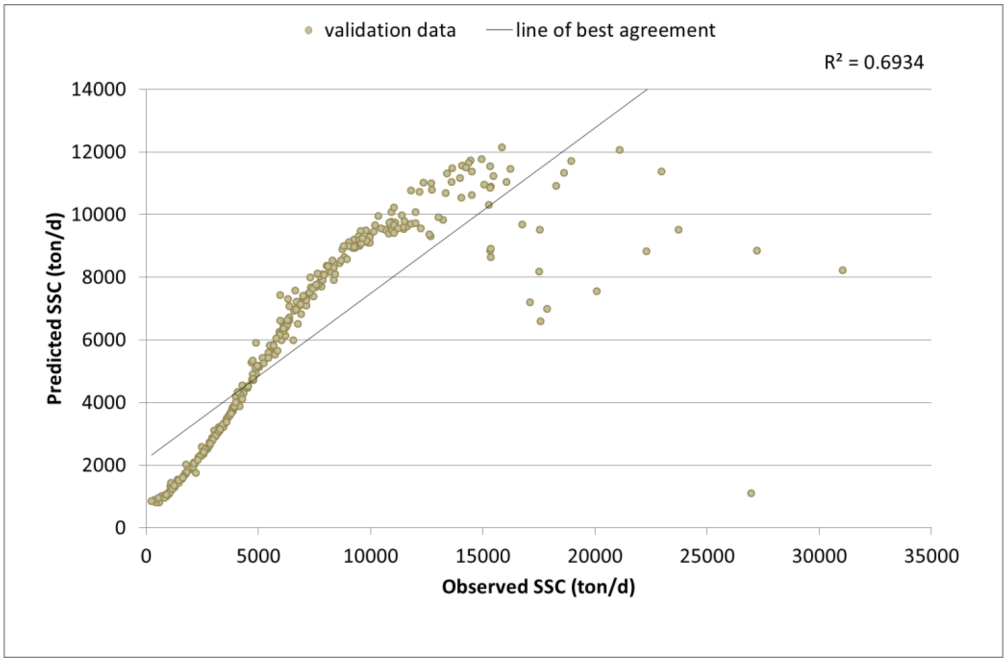

Figure 3.14: Plotting of observed and predicted suspended sediment concentration for validation data

As can be seen on figure 3.14, many outliers can be identified. The deviation of outliers from the mean is very large. The outliers have negatively affected whole distribution of data, hence producing low correlation between predicted and observed value. Presence of outliers started within the range of 10000 to 15000 ton/day of the observed value, which coincides with the maximum range of training data. The range of suspended sediment concentration of validation data is too high for the RBF model to predict, hence producing underestimation value of suspended sediment concentration. In order to have a better comparison between observed and predicted suspended sediment of validation data, a time series was plotted. Figure 3.15 shows the time series for observed and predicted suspended sediment concentration of validation data.

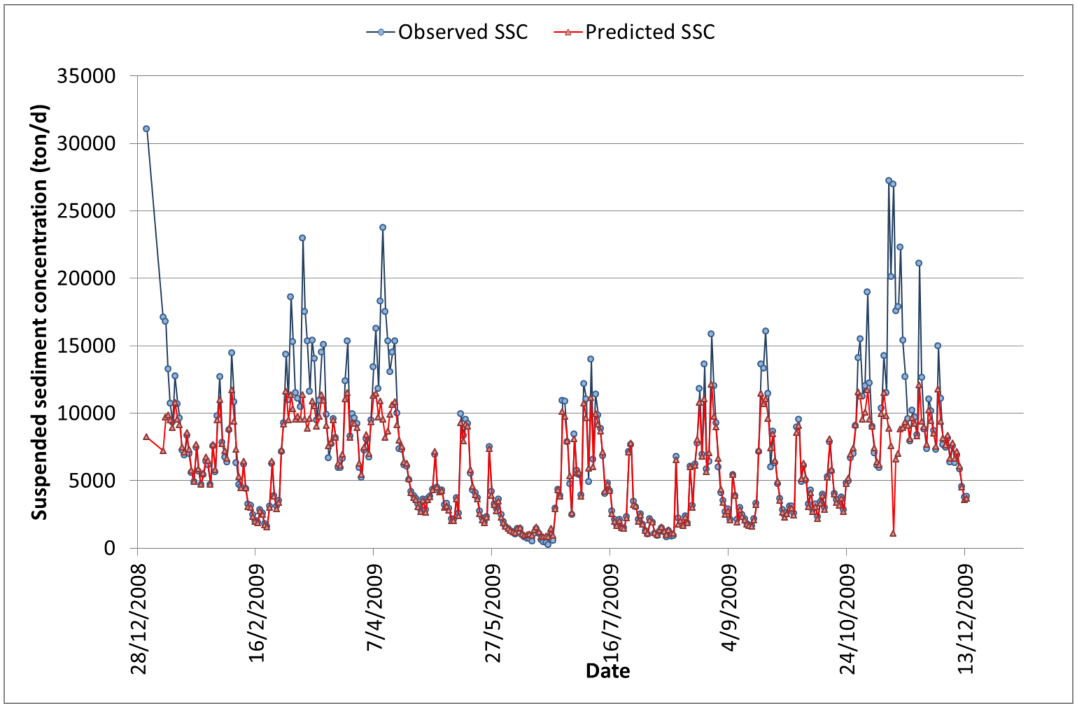

Figure 3.15: Time series for observed and predicted suspended sediment concentration of validation data

Red time series indicates the predicted values while the blue time series shows the observed value. Prediction of suspended sediment concentration at the trough is not as difficult at the spikes. The predicted data fails to follow the spikes of observed data. Extreme spikes produced greatest error as compared to the majority of the data. There is a noticeable upper limit of the predicted value, which is in around 12000 ton/day. This corroborates the plotting of observed and predicted suspended sediment concentration of validation data, where the presence of outliers start between the value of 10000 ton/day and 15000 ton/day of observed suspended sediment concentration. If the value of observed suspended sediment concentration is high, the margin of prediction error is high as well. To accommodate large observed data, the training stage should incorporate large amount of high magnitude data. This is possible through the expansion of training data to more than 5 years. Validation data shows higher suspended sediment concentration as compared to training and testing data because of the recentness. Training and testing data were taken from the year of

1992 to 1994 and 1995, respectively, whereas validation data were taken from the year 2009. The gap between training and validation is 15 years. Within 15 years, Kinta River catchment areas had undergone numerous developments. The developments of catchment area may affect the behavior of suspended sediment transportation in Kinta River. Possible addition of sediments comes from surface runoffs. Development of Kinta River catchment area causes the water penetration into soil decreases as most of the land surfaces are covered with pavements and buildings.

CHAPTER 4

CONCLUSION AND RECOMMENDATION

It is important to ensure that the progress of the research is aligned with the objective. RBFNN is one of the soft computing techniques available and the prediction of suspended sediment concentration will be done using RBF model. The summary of the developed RBF model is as below:

- 3 input variables : Q_t, Q_{t-1} and Q_{t-2}
- Spread, σ : 0.80777
- Kernel : Thin plate spline
- Hidden layer : 18 neuron
- 1 output : S_t

RBF model developed was able to predict suspended sediment concentration well. The important parameter that was used in measuring the performance of RBF model is correlation coefficient, R^2. Testing of RBF model yielded higher R^2 value than training, which is 0.9884 and 0.9856, respectively. The highly accurate predictions also supported by good data input, as presented in the analysis that both suspended sediment concentration and stream discharge are highly correlated to each other for both training and testing. This allows the learning process of RBF to be simplified and increase the accuracy of predictions. Outliers were identified from the results and it was found that outliers may affect prediction of suspended sediment negatively.

It is recommended for further research to include the outliers and use necessary method to eliminate outliers without affecting the good data in order to produce a reliable prediction model and accurate predictions. Apart from that, for a validation of prediction models using raw data collected from Kinta River is highly recommended. The models will be used to predict suspended sediment concentration using recent suspended sediment concentration and stream discharge. The training of RBF model should include a wide range of data. Therefore, the increasing value of stream discharge and suspended sediment concentration could be anticipated by the model hence producing prediction with high accuracy.

REFERENCES

Reddy, M. J. and Ghimire, B. N. S. (2009).“Use of Model Tree and Gene Expression Programming”. Journal of Intelligent Systems **18**(3).

Aytek, A. and Kişi Ö. (2008). "A genetic programming approach to suspended sediment modelling." Journal of Hydrology **351**(3): 288-298.

Walling, D. E. "Limitation of the rating curve technique for estimating suspended sediments loads, with particular reference to British rivers."

2012. < http://water.epa.gov/scitech/datait/tools/warsss/dimless.cfm.>

Khalilabad, H. M., Feiznia, S. & Zakikhani, K. (2009) “Estimating River Suspended Sediment Yield using MLP Neural Network in Arid and semi-arid basins; Case Study: Bar River, Neyshaboor, Iran.”

Aydin, A. and Eker, R. 2012. “Prediction of Daily Suspended Sediment Load Using Radial Basis Function Neural Network.”

Mustafa, M. R., Rezaur, R. B., Saiedi, S. & Isa, M. H. (2012). “River Suspended Sediment Prediction Using Various Multilayer Perceptron Neural Network Training Algorithms- A Case Study in Malaysia.” *Water Resource Management* **26**: 1879- 1897.

Memarian, H. and Balasundram, S. K. (2012). “Comparison between Multi-Layer Perceptron and Radial Basis Function Networks for Sediment Load Estimation in Tropical Watershed.” *Journal of Water Resource and Protection* **4:** 870- 876.

Feyzolahpour, M., Rajabi, M. & Roostaei, S. (2012). “Estimating Suspended Sediment Concentration using Neural Differential Evolution, Multi layer Perceptron and Radial Basis Function Models”. *Journal of Physical Sciences* **7** (29): 5106- 5117

Jain, A. K. and Mao, J. (1996). “ Artificial Neural Networks: A Tutorial”

Nash, J.E. and Sutcliffe, J.V. “River flow forecasting through conceptual models part I — A discussion of principles,” *Journal of Hydrology*, Volume 10, Issue 3, April 1970, Pages 282-290, ISSN 0022-1694, http://dx.doi.org/10.1016/0022-1694(70)90255-6.

Hicks, D. M., 2011. *Review of Turbidity and Suspended Sediment Monitoring at Horizon Regional Council,* New Zealand.

Lafdani, E. K., Nia, A. M. and Ahmadi, A. 2013, "Daily suspended sediment load prediction using artificial neural networks and support vector machine," *Journal of Hydrology* **478** (2013): 50-62

Liu, Q. J., Shi, Z. H., Fang, F.N., Zhu, H. D. and Ai. L. 2013, "Modeling the daily suspended sediment concentration in a hyperconcentrated river on the Loess Plateau, China, using the Wavelet-ANN approach" *Geomorphology* **186** (2013): 181-190

Kia, E., Emadi, A. R. and Fazlola, R. 2013, "Investigation and evaluation of artificial neural networks in Baboulroud River suspended load estimation*" Journal of Civil Engineering and Urbanism* **3(4)**: 183-190

Salimi, M, Hassanzadeh, Y., Daneshfaraz, R. and Salimi, M. 2013,"Sedimentation estimation study using artificial neural network for Karaj Dam Reservoir in Iran," *Journal of Basic and Applied Scientific Research* **3 (8)**: 185-193

Kisi. O., Dailr, A. H., Cimen, M. and Shiri, J. 2012 "Suspended sediment modeling using genetic programming and soft computing techniques" *Journal of Hydrology* **450-451** (2012): 45-58

Kumar, A.R.S., Ojha, C.S.P., Goyal, M.K., Singh, R.D. and Swamee, P.K. 2012, "Modeling of suspended sediment concentration at Kasol in India using ANN, fuzzy logic, and decision tree algorithms" *Journal of Hydrologic Engineering* **17**: 394-404

Melesse, A.M, Ahmad, S., McClain, M.E., Wang, X. and Lim, Y.H., 2011 "Suspended sediment load prediction of river systems: an artificial neural network approach" *Agricultural Water Management* **98(2011)**: 855-866

Rajaee, T., Mirbagheri, S.A., Kermani, M. Z. and Nourani, V. 2009," Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models" *Science of the Total Environment* **407 (2009)**: 4916-4927

Rai. R.K. and Mathur, B.S. 2008, “ Event based sediment yield modeling using artificial neural network” *Water Resource Management* **(22)**: 423-441

Partal, T. and Cigizoglu, H.K. 2008, “Estimation and forecasting of daily suspended sediment data using wavelet-neural networks” *Journal of Hydrology* **358**:317-331