

**NITROGEN PREDICTION MODEL THROUGH NEW HYBRID
MODEL USING ANT COLONY OPTIMIZATION AND ELMAN
NEURAL NETWORK**

PAVITRA KUMAR

**FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2021

**NITROGEN PREDICTION MODEL THROUGH NEW
HYBRID MODEL USING ANT COLONY
OPTIMIZATION AND ELMAN NEURAL NETWORK**

PAVITRA KUMAR

**THESIS SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY**

**FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2021

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **PAVITRA KUMAR**

Registration/Matric No: **KVA180055**

Name of Degree: **DOCTOR OF PHILOSOPHY**

Title of Thesis: **NITROGEN PREDICTION MODEL THROUGH NEW HYBRID MODEL
USING ANT COLONY OPTIMIZATION AND ELMAN NEURAL NETWORK**

Field of Study: **WATER RESOURCES ENGINEERING**

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature:

Date: 22-Jul-21

Subscribed and solemnly declared before,

Witness's Signature:

Date: 22-Jul-21

Name:

Designation:

NITROGEN PREDICTION MODEL THROUGH NEW HYBRID MODEL USING ANT COLONY OPTIMIZATION AND ELMAN NEURAL NETWORK

ABSTRACT

The prediction of nitrogen not only assists in monitoring the nitrogen concentration in streams but also helps in optimizing the usage of fertilizers in agricultural fields. A precise prediction model guarantees the delivery of better-quality water for human use, as the operations of various water treatment plants depend on the concentration of nitrogen in streams. Considering the stochastic nature and the various hydrological variables upon which nitrogen concentration depends, a predictive model should be efficient enough to account for all the complexities of nature in the prediction of nitrogen concentration. For two decades, artificial neural networks (ANNs) and other models (such as autoregressive integrated moving average (ARIMA) model, hybrid model, etc.), used for predicting different complex hydrological parameters, have proved efficient and accurate up to a certain extent. The current research work presents the methodology of development of artificial neural network models for the prediction of nitrate-nitrogen and ammonia-nitrogen in streams. A new training method is proposed which paves the way of selection of models among the number of models having different combinations of internal parameters providing different levels of prediction accuracy. Using the new training approach, four optimum models have been selected for prediction of both the nitrogen compounds at two different measuring stations i.e. LUI and KAJANG, in Langat River basin in Malaysia. These models provided considerable accuracy of prediction. However, for real time prediction the accuracy needed to be enhanced.

To enhance the prediction of the developed artificial neural network models, a new hybrid model was developed. Ant colony optimization and Elman neural network was integrated to form a new ACO-ENN hybrid model. ACO performs a decision-making task and ENN

performs learning task. ACO selects a best combination of internal parameters and provides it to ENN for an enhanced starting point, leading to a better result. The same four models were developed based on the same historical data using hybrid modelling technique. The prediction results obtained from the hybrid models had higher accuracy than that obtained from standalone artificial neural network models.

Keywords: Monthly Nitrate-Nitrogen, Monthly Ammonia-Nitrogen, Ant Colony Optimization, Elman Neural Network, Multilayer Neural Network

Universiti Malaya

RAMALAN NITROGEN MELALUI MODEL HYBRID BARU MENGUNAKAN OPTIMASI KOLONI SEMUT DAN RANGKAIAN NEURAL ELMAN

ABSTRAK

Ramalan nitrogen bukan sahaja membantu dalam pemantauan kepekatan nitrogen di dalam aliran sungai tetapi juga membantu dalam mengoptimumkan penggunaan baja di ladang pertanian. Model ramalan yang tepat menjamin penyampaian air berkualiti lebih baik untuk kegunaan manusia, kerana operasi pelbagai loji rawatan air bergantung pada kepekatan nitrogen di dalam aliran sungai. Mempertimbangkan sifat stokastik dan pelbagai pembolehubah hidrologi yang mana kepekatan nitrogen bergantung, model ramalan seharusnya cukup cekap untuk mempertimbangkan semua kerumitan alam dalam ramalan kepekatan nitrogen. Selama dua dekad, rangkaian saraf tiruan (ANN) dan model lain (seperti model purata bergerak terpadu autoregresif (ARIMA), model hibrid, dan lain-lain), yang digunakan untuk meramalkan pelbagai parameter hidrologi yang kompleks, telah membuktikan kecekapan dan ketepatannya sehingga tahap tertentu. Kerja penyelidikan ini membentangkan metodologi pengembangan model rangkaian saraf tiruan untuk ramalan nitrate-nitrogen dan ammonia-nitrogen dalam aliran sungai. Satu kaedah latihan baru telah dicadangkan untuk pemilihan model di antara sejumlah model yang mempunyai kombinasi parameter dalaman yang berbeza, dan memberikan tahap ketepatan ramalan yang berbeza. Dengan menggunakan pendekatan latihan baru, empat model optimum telah dipilih untuk meramalkan kedua-dua sebatian nitrogen di dua stesen pengukuran yang berbeza iaitu Lui dan Kajang, di lengangan Sungai Langat di Malaysia. Model-model ini memberikan ketepatan ramalan yang cukup tinggi. Walau bagaimanapun, untuk ramalan waktu sebenar, ketepatannya perlu ditingkatkan.

Untuk meningkatkan ramalan bagi model-model rangkaian saraf tiruan yang dibinakan, satu model hibrid baru telah dibinakan. Pengoptimuman koloni semut dan rangkaian saraf

Elman telah diintegrasikan untuk membentuk satu model hibrid ACO-ENN baru. ACO melaksanakan tugas membuat keputusan dan ENN melaksanakan tugas pembelajaran. ACO memilih kombinasi parameter dalaman yang terbaik dan memberikannya kepada ENN untuk titik permulaan lebih baik, yang membawa kepada hasil yang lebih baik. Empat model yang sama dikembangkan berdasarkan data sejarah yang sama menggunakan teknik pemodelan hibrid. Hasil ramalan yang diperoleh dari model-model hibrid mempunyai ketepatan yang lebih tinggi daripada yang diperoleh dari model-model rangkaian saraf tiruan tersendiri.

Kata kunci: Nitrate-Nitrogen Bulanan, Amonia-Nitrogen Bulanan, optimasi Koloni Semut, Rangkaian Neural Elman, Rangkaian Neural Pelbagai Lapisan

ACKNOWLEDGMENT

I thank the almighty GOD with all my heart for his blessing and kind support throughout my life.

I would like express my heartfelt thanks my supervisors Dr. Ahmed Hussein Kamel Ahmed Elshafie, Dr. Lai Sai Hin, Dr. Nuruol Syuhadaa Binti Mohd and Dr. Md Rowshon Kamal for their eternal support and immeasurable patience for me during my PhD. With the help of their support it has only been possible for me to achieve this degree.

I extend my gratitude to Department of Irrigation and Drainage (DID), Malaysia for providing me the required data for my research studies.

I would like to thank my family members for their supporting me in achieving my PhD degree overseas. Their support and blessings have helped me a lot in my research work.

I wish to thank my friends Nwakaire Chidozie Maduabuchukwu and Mohammed K. H. Radwan for helping at various stages in my research works and class works.

Once again, I thank you all for making me capable enough to complete this thesis and hence, my PhD degree.

TABLE OF CONTENTS

<i>ABSTRACT</i>	<i>iii</i>
<i>ABSTRAK</i>	<i>v</i>
<i>ACKNOWLEDGMENT</i>	<i>vii</i>
<i>List of Figures</i>	<i>xi</i>
<i>List of Tables</i>	<i>xiv</i>
<i>List of Symbols and Abbreviations</i>	<i>xv</i>
CHAPTER 1. GENERAL INTRODUCTION	1
1.1 Background	1
1.1.1 Nitrogen Sources in Streams.....	2
1.1.2 Effects of Nitrogen.....	4
1.2 Problem Statement	6
1.3 Aims and Objectives	7
1.4 Study area	7
1.5 Thesis Outline	10
CHAPTER 2. LITERATURE REVIEW	11
2.1 Nitrogen monitoring	11
2.2 ANN	15
2.3 Application of ANN	19
CHAPTER 3. ANN MODEL FOR RIVER-NITROGEN PREDICTION	38
3.1 Introduction	38

3.2 Literature Review	39
3.3 Methodology	40
3.3.1 Data collection and interpolation	40
3.3.2 Data division	44
3.3.3 ANN training and parameter selection	45
3.3.4 Performance Criteria	49
3.4 Results	51
3.5 Discussion	65
3.6 Conclusion	71
<i>CHAPTER 4. ENHANCEMENT OF NITROGEN PREDICTION ACCURACY</i>	73
4.1 Introduction	73
4.2 Background	74
4.2.1 ACO	74
4.2.2 ENN	76
4.2.3 Hybrid model	77
4.3 Literature Review	78
4.4 Methodology	80
4.4.1 Data Acquisition	80
4.4.2 Modification of ACO	80
4.4.3 Integrating ACO and ENN	80
4.4.4 Application of ACO-ENN Hybrid model on nitrogen data	88
4.4.5 Performance criteria	90
4.5 Results	91
4.6 Discussion	99

4.6.1 Software Interface.....	103
4.7 Conclusion.....	110
<i>CHAPTER 5. CONCLUSION</i>.....	111
5.1 Introduction.....	111
5.2 Conclusion.....	111
5.3 Recommendations for future work.....	113
<i>References:</i>.....	115
<i>List of Publications and Papers Presented</i>.....	123
<i>Co-Author Consent</i>.....	124

Universiti Malaysia

List of Figures

Figure 1.1 Langat River basin (Zomorodian et al., 2017). Reprinted from (Zomorodian et al., 2017) under a CC BY license, with permission from PLOS ONE, original copyright 2017.....	9
Figure 2.1 Comparison of Nitrate-Nitrogen in rivers for two decades data (Rekacewicz, 2006).....	12
Figure 2.2 (a) Classification of different point sources showing their contribution of Ammonia-Nitrogen in the Langat River; (b) comparison of the contributions of Ammonia-Nitrogen from point and non-point sources (Farid et al., 2016; Juahir et al., 2011).....	14
Figure 2.3 Basic structure of neural network.....	16
Figure 2.4 Classification of neural network(Jain et al., 1996).....	18
Figure 2.5 Comparison of 13 different models results with the DONcal	33
Figure 3.1 Plot of monthly average interpolated data of Nitrate and Ammonia.....	42
Figure 3.2 Plot of monthly average data of rainfall, water level and discharge	43
Figure 3.3 Flow chart for the model selection for Nitrate-Nitrogen at Lui station.....	52
Figure 3.4 Flow chart for the model selection for Ammonia-Nitrogen at Lui station....	53
Figure 3.5 Flow chart for the model selection for Nitrate-Nitrogen at Kajang station...	54
Figure 3.6 Flow chart for the model selection for Ammonia-Nitrogen at Kajang station.....	55
Figure 3.7 Taylor diagram for Nitrate-Nitrogen (Station Lui).....	57
Figure 3.8 Taylor diagram for Ammonia-Nitrogen (Station Lui).....	57
Figure 3.9 Taylor diagram for Nitrate-Nitrogen (Station Kajang).....	58

Figure 3.10 Taylor diagram for Ammonia-Nitrogen (Station Kajang).....	58
Figure 3.11 Comparison of mean square error for different input vectors for Nitrate-N at station Lui.....	61
Figure 3.12 Comparison of mean square error for different training data division percentage for Nitrate-N at station Lui	62
Figure 3.13 Plot of variation of mean square error against number of nodes in hidden layers for Nitrate-N at station Lui	62
Figure 3. 14 Plot of variation of mean square error against number hidden layers for Nitrate-N at station Lui	63
Figure 3.15 Plot of variation of testing mean square error against spread values for general regression model for Nitrate-N at station Lui	64
Figure 3.16 Plot of variation of testing mean square error against spread values for RBFNN model for Nitrate-N at station Lui	64
Figure 3.17 Plot of variation of mean square error against epochs for Nitrate-N at station Lui	65
Figure 3.18 Relative error percentage plot for optimum selected model.....	67
Figure 3.19 Observed versus predicted plot for optimum selected models.....	68
Figure 3.20 Average uncertainty of different models	69
Figure 4.1 Basic concept of ACO.....	75
Figure 4.2 Structure of ENN	77
Figure 4.3 Plot for ant movement	82
Figure 4.4 Movement of queen ant (Blue path)	84
Figure 4.5 Flow chart of model structure.....	88
Figure 4.6 Convergence of MSE in ACO for Nitrate-Nitrogen model at station Lui.....	93
Figure 4.7 Convergence of MSE in ACO for Ammonia-Nitrogen model at station Lui	93

Figure 4.8 Convergence of MSE in ACO for Nitrate-Nitrogen model at station Kajang..	94
Figure 4.9 Convergence of MSE in ACO for Ammonia-Nitrogen model at station Kajang.....	94
Figure 4.10 Plot of target and output for each hybrid model.....	96
Figure 4.11 Plots of relative percentage error obtained from ANN.....	98
Figure 4.12 Plots of relative percentage error obtained from hybrid model.....	99
Figure 4.13 ACO-ENN hybrid model software interface	103
Figure 4.14 Model Description	104
Figure 4.15 Creating Elman Neural Network	104
Figure 4.16 Window after creating Elman neural network.....	105
Figure 4.17 Window for choosing training parameters	106
Figure 4.18 Window for ACO iterations.....	106
Figure 4.19 ACO Training	107
Figure 4.20 Elman neural network training	108
Figure 4.21 Results of the Trained network.....	109

List of Tables

Table 2.1 Advantages and disadvantages of different ANN models	18
Table 2.2 A summary of studies that utilize ANN model for nitrogen prediction, including their specific area, location, and methods used	21
Table 2.3 Details of methodology of the reviewed research work	24
Table 2.4 A summary of studies that utilize ANN model for nitrogen prediction, including their input variables, prediction variables, and accuracy.....	28
Table 3.1 Statistical analysis of the data for both stations.....	44
Table 3.2 Different Input and internal parameters for different ANN models	46
Table 3.3 Optimum ANN models for Lui and Kajang stations	59
Table 3.4 Comparison of current study results with literature.....	70
Table 4.1 Configurations of hybrid models for Lui and Kajang stations.....	95
Table 4.2 Comparison of the result of ANN and Hybrid model.....	97
Table 4.3 Performance improvement of hybrid model	100
Table 4.4 Comparison of the result of ENN and Hybrid model	101

List of Symbols and Abbreviations

Abbreviations	Description
ACO	Ant Colony Optimization
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BPNN	Back Propagation Neural Network
DID	Department of Irrigation and Drainage
DON	Dissolved Organic Nitrogen
DON _{cal}	Calculated DON
Dt	Squared Differences
ENN	Elman Neural Network
EPR	Evolutionary Polynomial Regression
FFNN	Feed Forward Neural Network
GBM	Generalized Busted Model
GLM	Generalized Linear Model with Stepwise Feature Selection
GPL	Gaussian process with linear kernel
GPR	Gaussian process with radial basis function kernel
GRNN	General Regression Neural Network
I_{rem-tn}	Average removal rate by treatment plants
I_{sew}	Percentage of sewage effluent treated by treatment plants
KNN	k-nearest neighbours
LFC	Low Flow Criteria
MAE	Mean Absolute Error
MLIT	Ministry of Land, Infrastructure, Transport and Tourism
MSE	Mean Square Error

$NANI_n$	Net Anthropogenic Nitrogen Input from non-point source
$NANI_p$	Net Anthropogenic Nitrogen Input from point source
NBM	Naïve Bayes Model
N_{chem}	Chemical nitrogen from the fertilizers applied
N_{dep}	Oxidized nitrogen deposition from atmosphere
NES	National Eutrophication Survey
N_{fix}	Nitrogen contribution from natural nitrogen fixation in crops
N_{ind}	Nitrogen from Industries
N_{r-im}	Net food and feed import in rural areas
NSE	Nash-Sutcliffe Efficiency
N_{urban}	Nitrogen from urban inhabitants
PFC	Peak Flow Criteria
Q	Discharge
RBFNN	Radial Basis Function Neural Network
RE	Relative Error Percentage
RF	Rainfall
RMSE	Root mean square error
SVML	support vector machine with linear kernel
SVMR	support vector machine with radial basis function kernel
WL	Water Level

CHAPTER 1. GENERAL INTRODUCTION

1.1 Background

Human activities have provoked serious effects on nutrient cycle, streams ecological functioning, and water quality (Jacobs et al., 2017; Kilonzo et al., 2014; Maloney & Weller, 2011). Presently, agriculture production consummately depends on the amount of fertilizers and pesticides used. Fertilizers mainly contain nitrogen compared with other chemicals. Crops require nitrogen for their growth and for the production of fruits or grains. Some agricultural specialists have also recommended using the fertilizers that carry a higher percentage of nitrogen (Hessong, 2019). However, only 40%–70% of nitrogen compounds applied as fertilizers are absorbed by the crops. The remaining nitrogen compounds either percolate downward with water to join groundwater or flow along with the runoff water to join the streams (Salehi et al., 2000; Sharma et al., 2003). In both cases, the nitrogen concentration in water escalates, which can affect human health (Fewtrell, 2004; Gallo et al., 2015; Ward et al., 2005). If pesticides and fertilizers are added to the fields at a high rate, there is more chance for nitrate to percolate to the aquifer, increasing the nitrate level in groundwater (Hamed et al., 2013; Rahmati et al., 2015; Reddy et al., 2009). In warmer countries, the loss of total nitrogen is more, as mineralization rate is probably higher due to the higher temperature; thus, the percolation of total nitrogen is increased (Räike et al., 2003).

The major proportion of the surplus nitrogen is transported by the runoff water to the streams, and consequently, nitrogen compounds such as ammonia-nitrogen, nitrite, and nitrate are escalated in the streams. A surfeit of nitrogen in streams seems to be deleterious for both human beings and aquatic lives. In water bodies, it may lead to the magnification of aquatic plants and algae, which can result in the depletion of dissolved oxygen and

hinder the contact of water with air and light. Thus, affecting the aquatic life. Nitrogen causes serious health issues for human, such as various types of cancer, thyroid related problems and several types of birth defect.

1.1.1 Nitrogen Sources in Streams

Nitrogen is a vital element for plants, as it helps them in their growth and productivity. Nitrogen present as N_2 in the atmosphere cannot be utilized directly by plants until it is converted to its reactive compounds, such as NH_3 , NH_4^+ , NO_2^- , or NO_3^- (Suo et al., 2012).

This process is naturally done by bacteria present in the soil and in the root nodules of legume crops. Additionally, nitrogen compounds are provided to the soil in the form of fertilizers. Nitrate is the main constituent of fertilizers, but ammonia, ammonium, urea, and amines are also present in minor proportions. Nowadays, fertilizers contain more percentage of nitrogen compounds in order to boost the agricultural productivity.

In addition, the landscapes of the farmlands have been modified extensively. Farmlands are now designed to drain off the excess rainwater or irrigation water (USGS, 2018). This drained water is rich in nitrogen compounds, which had been applied to the field for crops nourishment. The drained water then joins either running rivers or still water bodies such as lakes, leading to a surfeit of nitrogen entering the water system.

Sources of nitrogen to streams are not confined only to agricultural fields. Industries and municipal and residential areas also contribute nitrogen compounds to streams. Comprehensively, the sources of nitrogen are classified into two:

a. Point Sources

A point source of nitrogen pollution is any single identifiable source of nitrogen pollution into rivers. Point sources include industries and municipal sewage treatment plants (Farid et al., 2016; Yi et al., 2017; W. S. Zhang et al., 2015). In urban areas, the contribution of nitrogen from point sources is dominant. Industries and municipal sewage treatment plants deliver more than 50% of the total nitrogen

in rivers (Suo et al., 2012). W. S. Zhang et al. (2015) proposed a formula to calculate the total anthropogenic nitrogen contribution in rivers from point source.

$$NANI_P = (N_{urban} + N_{ind})(1 - I_{sew}I_{rem-tn}) \quad (1.1)$$

where:

$NANI_P$ = Net Anthropogenic Nitrogen Input from point source

N_{urban} = Nitrogen from urban inhabitants

N_{ind} = Nitrogen from Industries

I_{sew} = Percentage of sewage effluent treated by treatment plants

I_{rem-tn} = Average removal rate by treatment plants

b. Non-point sources

Non-point sources are the sources of nitrogen pollution whose specific locations of input to rivers are not defined. They mainly consist of agricultural fields and atmospheric and biological nitrogen fixation (Farid et al., 2016; Yi et al., 2017; W. S. Zhang et al., 2015). In rural areas, the contribution by non-point sources is dominant. In different regions of rural areas, different parts of non-point sources contribute major amounts of nitrogen in streams; for example, in farming regions, agricultural fields provide significant nitrogen to the streams, and in the regions of rivers surrounded by dense forests, atmospheric nitrogen deposition dominates (Suo et al., 2012). According to W. S. Zhang et al. (2015), contributions of nitrogen from non-point source can be determined by following formula:

$$NANI_n = N_{chem} + N_{fix} + N_{dep} + N_{r-im} \quad (1.2)$$

where:

$NANI_n$ = Net Anthropogenic Nitrogen Input from non-point source

N_{chem} = Chemical nitrogen from the fertilizers applied

N_{fix} = Nitrogen contribution from natural nitrogen fixation in crops

N_{dep} = Oxidized nitrogen deposition from atmosphere

N_{r-im} = Net food and feed import in rural areas

Net food and feed import in rural areas is calculated based on the total consumption, by humans and animals, and production, by crops, of nitrogen. N_{r-im} comes out to be negative if production of nitrogen is more than its consumption, representing that there will be a net export of nitrogen to the streams.

1.1.2 Effects of Nitrogen

Nitrogen, if present in river water causes different disorders, which are deleterious for both human and aquatic animals. Nitrogen present in streams are mainly found in three compound states: ammonia, nitrate, and nitrite. Some amounts of ammonia present in the river water get converted to nitrate depending on the dissolved oxygen concentration in the water (Nuruzzaman et al., 2017). As stated earlier, nitrate is not much deleterious, but if present in surplus amount, it starts converting into nitrite, which is very harmful even in minute concentration. The Environmental Protection Agency has set standards that state that for water that is to be distributed for public use, the maximum acceptable nitrate concentration is 10 mg/l (Salehi et al., 2000; Suen & Eheart, 2003) and that for nitrite is 1 mg/l.

There are two major effects of ammonia on whole ecosystem: eutrophication of marine and terrestrial ecosystems (Rabalais & Turner, 2006; Rabalais et al., 2002) and increase in the acidity of water bodies (Hessen et al., 1997). Excessive nutrients such as nitrogen and phosphorus when present in water bodies lead to the growth of algae on the top surface of water; this process is termed as eutrophication. Excess grown algae cover the whole water surface, blocking the contact of water from sunlight and air. Additionally,

the algae growth decreases the oxygen level in water body, which affects the aquatic lives. Stream eutrophication was recognized as a major problem years ago, and the United States along with other countries commenced nutrient control measures in rivers (W. Dodds & Smith, 2016; W. K. K. Dodds & Welch, 2000).

Streams may get acidified due to the presence of surfeit ammonia. The most common form of ammonia, ammonium sulphate, leads to formation of considerable amount of acid, as hydrogen ions are released during nitrification. Also, nitrite ions present in the streams lead to the formation of nitric acid under different situations along with sulfate ions, consequently acidifying the stream water (Murdoch & Stoddard, 1992). Acidic stream water is not even suitable for reuse to satisfy human water requirements. As stated by Gündüz (2015), one day, reuse of treated water would be a reality for the rural population, and this would result in serious problems such as human health issues. Compared with urban areas, agricultural areas are more susceptible to health risks by the presence of nitrate-nitrogen in groundwater (S. He & Wu, 2019; Su et al., 2013).

Nitrite has been found to be more toxic than nitrate and if present in drinking water can cause human health problems such as liver damage and in worst cases can lead to various types of cancer (Hossain et al., 2010) and two types of birth defects (Chen et al., 2017; Gulis et al., 2002). Nitrite present in surplus quantity in drinking water will eventually lower the ability of bloodstreams to carry oxygen, leading to the lack of oxygen in the body. As per Salehi et al. (2000), the presence of such excess nitrogen in drinking water reduces the amount of oxygen transported in the blood. Infants and young livestock are lamentably affected, as this causes “blue baby syndrome” (Hossain et al., 2010), a disease causing bluish skin of the babies because of poorly oxygenated blood. The reaction of nitrites with amines either enzymatically or chemically leads to the formation of potent carcinogenic nitrosamines (Hossain et al., 2010; Sawyer CN et al., 2003).

Consumption of nitrates leads to various tumors in the human body (Aslan & Turkman, 2003; Hossain et al., 2010). In the digestive system, nitrate leads to the formation of N-nitroso compounds (Della Rocca et al., 2005; Hossain et al., 2010), which are considered to be carcinogenic. Iodine uptakes can be restricted by nitrates, causing thyroid-related problems (Hossain et al., 2010).

1.2 Problem Statement

Most treatment plants are not designed for the full removal of nitrogen compounds from river water. In China, sewage treatment systems remove total nitrogen by 40%–70% (Qiu et al., 2010; W. S. Zhang et al., 2015). In Malaysia, sewage treatment plants are not designed for ammonia removal (Indah Water, 2019). Recently, several water treatment plants have been forced to shut down (Pavitra Kumar et al., 2020) when, after testing the samples, it was found that ammonia-nitrogen pollution has crossed the acceptable limit of 1.5 ppm in different rivers in Malaysia. The abrupt closure of the water treatment plant affects the water supply to the consumers. Thus, adding additional pressure on the government for arranging alternate source of water supply.

The lack of monitoring systems leads to an abrupt increase in pollution, which can result in the closure of the water treatment plants. Monitoring systems should contain a proper predictive system: which works based on the historical data; and a treatment system: that deals with the nitrogen pollutant, should be developed in treatment plants. Predictive systems could provide the daily data of pollutants and thus save the daily effort of quantifying such data in the laboratory. Moreover, predictive systems would create an alert for nitrogen surge in rivers before it actually happens. Hence, the government would have ample time to optimize various nitrogen inputs in the rivers. Different river basins require separate predictive model, trained on historical data of basin's parameters. Because a model well-trained on historical data of one particular basin, not necessarily will perform with same accuracy on different basins. Hence, the government require

separate predictive model for each basin. Also, to consider the upcoming seasonal changes the predictive models need to be re-trained with the real-time data quarterly or yearly basis. Observing the increased pollution of nitrogen in rivers, this topic becomes important to be evaluated.

1.3 Aims and Objectives

The aim of this research is to develop a real time prediction model for nitrate-nitrogen and ammonia-nitrogen for Langat River Basin based on the historical data of the basin. The discretized objectives of this research are as follows:

1. To develop four artificial neural network models for prediction of nitrate-nitrogen and ammonia-nitrogen corresponding to stations Lui and Kajang. Three different artificial neural networks (GRNN, Multilayer, RBFNN) are explored with different set of internal combinations for developing optimized model.
2. To develop a new hybrid model by integrating an optimization algorithm with an artificial neural network. ACO is used as optimization algorithm for optimizing weights and biases of ENN (the artificial neural network used in hybrid model).
3. To enhance the prediction accuracy of the developed artificial neural network model using the new hybrid model. The results of four artificial neural network models are to be compared with the results of four hybrid models for possible enhancement of the prediction accuracy.

1.4 Study area

This study is based on the Langat River basin in Selangor, Malaysia. This basin has been selected as the Langat River has been facing the problem of high nitrogen content between 2012 to 2015, which led to the frequent suspension of different water treatment plants during this time period. As stated by Selangor Water Management Authority, Malaysia, the level of ammonia-nitrogen in the Langat River has exceeded the high risk level of 7.0

mg/l several times between 2012 and 2015 (Farid et al., 2016), resulting in the suspension of treatment plant operations. A study by AYERS et al. (1999), stated that the atmospheric deposition of oxides of sulphur and nitrogen in Petaling Jaya, a city near the Langat River basin, lies within the range 277-480 meq-m⁻²yr⁻¹, with nitrogen species contribution of 56%. This basin has a catchment area of about 2400 km². The Langat River supplies about 65% of the total water usage in the Selangor state. The Langat Dam (area 41.0 km²) and the Semenyih Dam (area 56.6 km²) are the two major reservoirs supplying water to the state (Soh et al., 2018). As per the 2013 analysis, the Langat River basin has a forest area of about 48,285.0 ha, an agricultural area of about 142,387.916 ha and the developed area of about 69,056.1 ha (Elfithri et al., 2017). About 72% of the soil in Malaysia is acidic and highly weathered (Ultisols and Oxisols) (Land and Plant Nutrition Management Service & Land and Water Development Division, 2004), which requires fertilizers for agriculture. The main fertilizers used in Malaysia are urea, ammonium sulphate, calcium ammonium nitrate, phosphate rock, super phosphates, ammonium phosphate, potassium chloride, potassium sulphate and NPK, NP and PK compound fertilizers (Land and Plant Nutrition Management Service & Land and Water Development Division, 2004). Along with the agricultural runoff, livestock wastes too have its role in increasing the nitrogen content in rivers. Livestock production in Malaysia consists of pork, poultry meat and eggs; and it has to import milk, beef and mutton.

The Langat River basin has a hot and humid tropical climate with a 27°C average annual temperature, which is uniform throughout the year and 2470 mm average annual rainfall distributed throughout the year (Ebrahimian et al., 2016).

Within the course of the Langat River flow, data from two water quality stations (Lui and Kajang) were acquired from the Department of Irrigation and Drainage (DID), Kuala Lumpur, Malaysia. The water quality station, Lui, is situated at the river Lui, in the upstream region of the Langat River basin, as shown in figure 1.1. This region is mainly

mountainous and less populated and hence, has less agricultural and industrial activities. The water quality station, Kajang, is situated at the Langat River in Kajang town, as shown in figure 1.1. This town is densely populated and is located near the capital city of Malaysia, Kuala Lumpur. Within the path of the flow from Lui to Kajang, the Langat River receives inflow from various agricultural fields of rubber, paddy and coconuts, and from various industries as well. These inflows increase the nitrogen content in the Langat River, which is clearly reflected in the water quality data of station Kajang. As discussed in the chapter 3, in statistical analysis, the average nitrate-nitrogen and ammonia-nitrogen show significant increase at the station Kajang, in comparison to that of station Lui. This clearly indicate the addition of nitrate-nitrogen and ammonia-nitrogen, in the Langat river, within the course of its flow from station Lui to station Kajang and perhaps more addition downstream of station Kajang. These addition between both the stations may be from both the sources, i.e., the point sources and non-point sources.

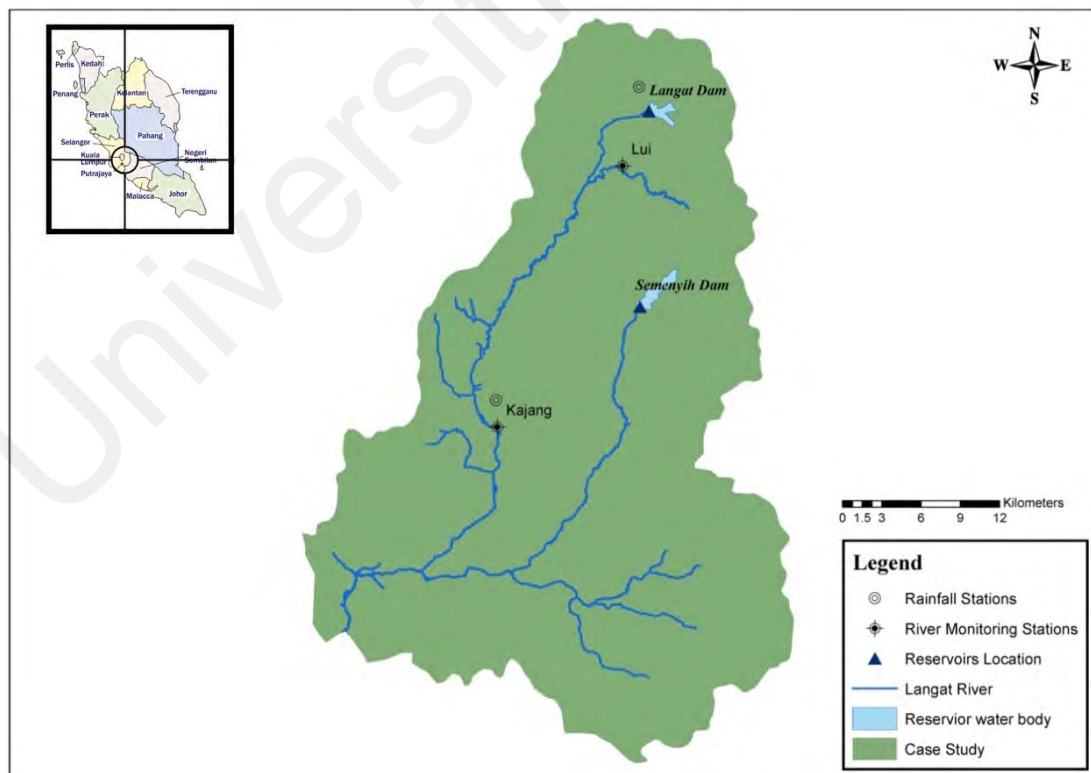


Figure 1.1 Langat River basin (Zomorodian et al., 2017). Reprinted from (Zomorodian et al., 2017) under a CC BY license, with permission from PLOS ONE, original copyright 2017

1.5 Thesis Outline

This thesis is compiled in article style format consisting of five chapters. The current chapter (chapter 1) consists of the general introduction, problem statement, objective, study area and the thesis outline.

Chapter 2 consists of literature review where use of artificial neural network and other predictive models has been reviewed. A list of authors has used ANN for prediction of various nitrogen compounds at different geographical locations around the world. Some of the reviewed research work have the exposure of hybrid modelling. This chapter also contains the nitrogen monitoring details and also the nitrogen concentration details of the globe.

Being an article format thesis, chapter 3 contains description of an article, “ANN modelling for river nitrogen prediction. It describes methods, results, discussion and conclusion of the artificial neural network model development for prediction of nitrate-nitrogen and ammonia-nitrogen for the stations Lui and Kajang in Langat river basin in Malaysia. This chapter contains the selection process of the optimum model among the different ANN models.

Chapter 4 also contains an article, which reviews ACO, ENN and hybrid model and has the description of development of a new hybrid model created by integration of ACO and ENN. The enhancement of the results of previously developed ANN models using the new hybrid model are also presented in this chapter.

Chapter 5 contains the overall conclusion of the thesis and also contains the recommendations of future works.

CHAPTER 2. LITERATURE REVIEW

The current research work is based on nitrogen compounds prediction in water bodies using ANN and other predictive models. In this chapter, in the section of ‘application of ANN’, analysis of the sources of data collection, methods used, internal parameters of the predictive model have been processed, followed by the final results of the previous research work in literature.

As used by Fiyadh et al. (2019), the relevant search engines such as Google Scholar and Science Direct have been used for obtaining the literature research works. Also, Martín-Martín et al. (2020) concluded, in their study, that Google Scholar is most comprehensive source. Through these search engines, the existing research works have been obtained from all around the world. While searching the relevant literature research works, following keywords have been used: nitrogen compounds prediction, use of ANN in nitrogen prediction and nitrogen prediction in water bodies.

2.1 Nitrogen monitoring

More than 60% of the world rivers are affected by pollution (Nuruzzaman et al., 2017), either from point sources or from non-point sources. Wastes generated by industrial, municipal, and agricultural activities are discharged into the rivers; which are extensively polluting them (Kannel et al., 2007; Nuruzzaman et al., 2017). Over time, human activities have escalated nitrogen species concentration in water bodies. Nitrate concentrations in many European rivers have surged by 5- to 10-fold since the 20th century (Suo et al., 2012). In Malaysia, because of the excessive chemical pollution in rivers, more than one among the nine water treatment plants in Langat River basin has been closed several times between 2012 and 2015 (Farid et al., 2016). According to Selangor Water Management Authority, Malaysia, between 2012 and 2015, the ammonia concentration level in the Langat River exceeded the high risk level of 7.0 mg/l, which

led to the repeated closure of many water treatment plants during this period (Farid et al., 2016). Moreover, in the Johor River basin, nearly five treatment plants were repeatedly closed between 2017 and 2019 due to the high concentration of ammonia in the Johor River (New Straits Times, 2017; The Star, 2019).

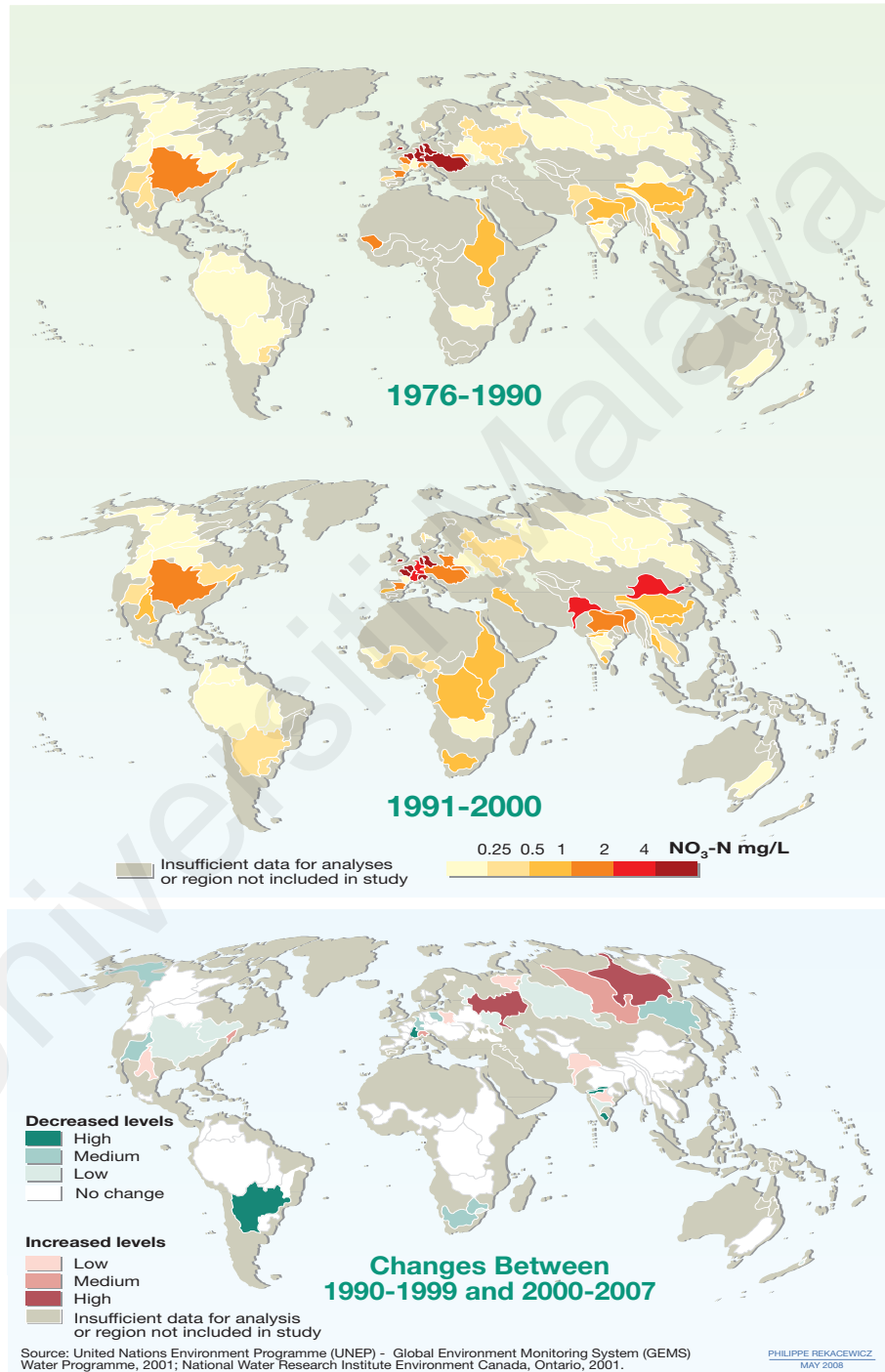


Figure 2.1 Comparison of Nitrate-Nitrogen in rivers for two decades data (Rekacewicz, 2006)

There is no specific standard set for ammonia discharge in water bodies, but different agencies have provided separate guidelines for ammonia concentration in water bodies. Canadian water quality guidelines for the protection of aquatic lives, (Canadian Council of Ministers of the Environment, 2010) states that the guideline value for unionized ammonia discharge in freshwater is a concentration of 0.019 mg/l. The guidelines for drinking water quality (2003) published by WHO states that natural levels of ammonia in groundwater are usually below 0.2 mg/l, and this level may go up to 12 mg/l for surface waters.

For analyzing nitrate variations, Rekacewicz (2006) designed a map, as shown in Figure 2.1, by considering all the river data at continental level, which represent the concentration of nitrate-nitrogen in streams at various locations around the world. Rekacewicz (2006) compared the data of two decades and observed that rivers in North America and Europe were fairly stable, but those of south-central Asia and Southeast Asia showed high nitrate concentrations.

Furthermore, Basheer et al. (2017) studied the water quality of the Langat River in Malaysia. They utilized 10 samples from different locations to quantify different water quality parameters. Their results showed that the pH range for the Langat River was between 5.91 and 6.79. The average value of ammonia for the Langat River was measured to be 0.24 mg/l. The total ammonia-nitrogen amounts added to the Langat River from point and non-point sources were calculated to be 9.51 ton/day and 12.67 ton/day, respectively (Farid et al., 2016; Juahir et al., 2011), as displayed in Figure 2.2.

Moreover, W. S. Zhang et al. (2015) tried to calculate nitrogen input to the Huai River in China from anthropogenic point and non-point sources and also the impact of nitrogen discharge on the riverine ammonia-nitrogen flux. They used the data from Yan et al. (2010), which stated that the average nitrogen concentration in the sewage discharged from industries in the Changjiang River basin was 25 mg/l. From the previous studies,

they could conclude that ammonia-nitrogen in the river was about 10% (or less) of the total nitrogen (S. Li et al., 2009; Singh et al., 2005; W. S. Zhang et al., 2015), and it could be as high as 70% in heavily polluted Asian rivers in the urban areas (W. Li et al., 2014; Pernet-Coudrier et al., 2012; W. S. Zhang et al., 2015). They used the data of Liang Zhang et al. (2011), which suggested that nitrate had become a major constituent of riverine nitrogen flux; the data was obtained from measurement in 2008, at several stations in the Huai River basin; the values of riverine nitrate concentration was found to vary between 0 and 15.7 mg/l nitrate-nitrogen, with a mean of 2.1 mg/l nitrate-nitrogen. When (W. S. Zhang et al., 2015) measured the ammonia-nitrogen in the same river basin, they found that the average ammonia-nitrogen concentration varied between 0.2 and 3.3 mg/l N, with an average of 1 mg/l N, which was half of the average nitrate-nitrogen concentration measured in 2008. The calculation of nitrogen input to the Huai River showed that on average, $27200 \pm 1100 \text{ kg km}^{-2}\text{y}^{-1}$ of nitrogen was added to the river from 2003 to 2010 as the net anthropogenic nitrogen input.

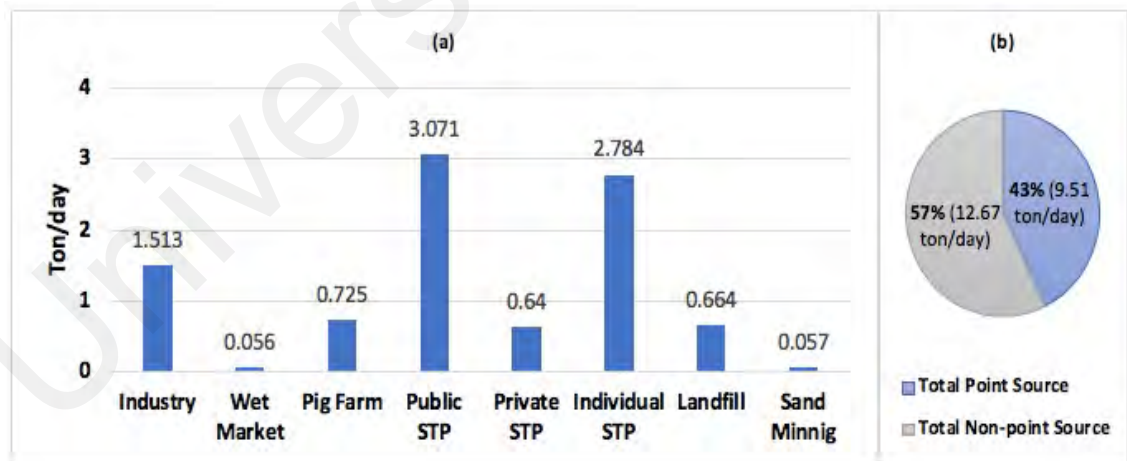


Figure 2.2 (a) Classification of different point sources showing their contribution of Ammonia-Nitrogen in the Langat River; (b) comparison of the contributions of Ammonia-Nitrogen from point and non-point sources (Farid et al., 2016; Juahir et al., 2011)

2.2 ANN

ANN is a black-box computational model (Akrami et al., 2013) that contains interconnected network-like structures passing values to other nodes of the connections. It contains an input layer, hidden layers as required, and an output layer. It is well known for its capability of predicting the non-linear variables (Farzad & El-Shafie, 2017). ANN has the capability of learning various patterns in the target data and it also has the ability to generalize the relation between the input and the target vectors. According to the generalized relations of the target data, ANN provides its future prediction. ANN forms the same structure as neurons in the human brain (Anctil et al., 2009; Lek et al., 1999; Sharma et al., 2003). It functions like a biological neuron, receiving the input as stimulus, evaluating the stimulus, and then providing the output as the response to the stimulus. Figure 2.3 represents a simple example of the neural network having four input vectors, two hidden layers and two output vectors. ANN receives inputs from user through input layer. The nodes pass the values of input data to the nodes in hidden layer 1 via interconnecting links. As the values are passed from input nodes to the following nodes, it is multiplied with the weights and then passed to the corresponding layer through a user defined transfer function (A. H. El-Shafie et al., 2011). Hidden layer 2 receives the values, after the transfer function of hidden layer 1, and processes them by multiplying them with its weights. The multiplying values are then forwarded to adjacent layer using the user defined transfer function. Likewise, it is passed up to the output layer, where the error is calculated using target vector. Based on this error, weights get updated to obtain minimum error. Once the minimum error, according to user defined criteria, is achieved the ANN stores the combination of weights to use it for future prediction of target vector based on the future set of inputs.

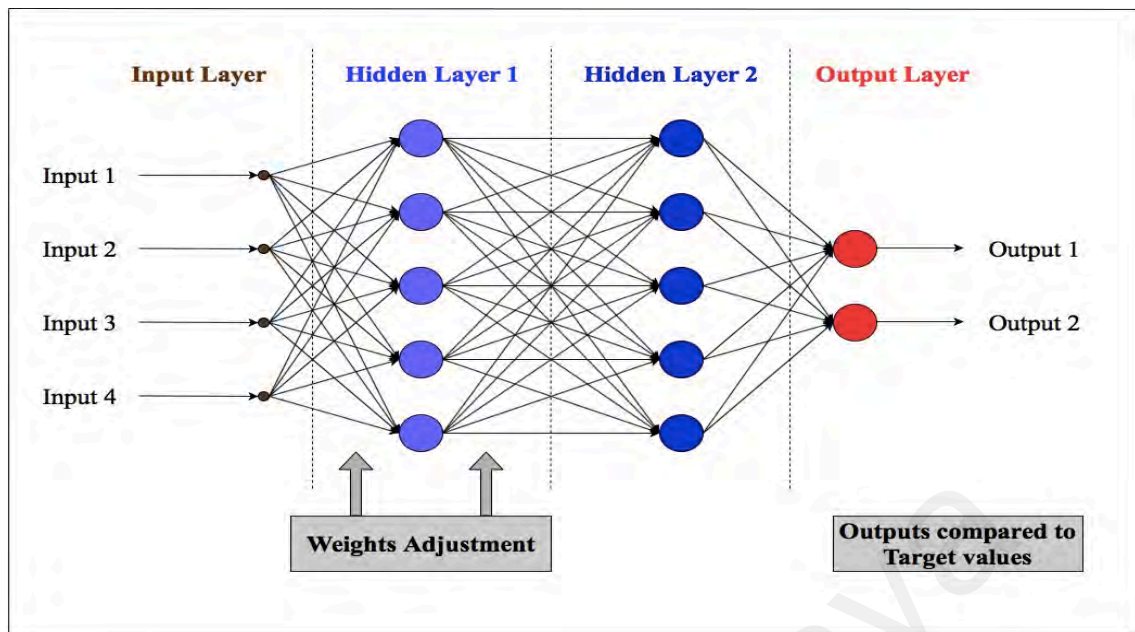


Figure 2.3 Basic structure of neural network

The major advantage of application of ANN model, over traditional model, such as statistical model, is that it learns itself the complexity of nature, without being explicitly transformed into mathematical form (El Shafie et al., 2012; Raju et al., 2011). Statistical models, have a limitation of assuming additional information to derive a sharp conclusion (Xie, 2011). The major disadvantage of ANN is that it is susceptible to overfitting. Overfitting is the state in training, beyond which, training error decreases but the model starts losing its ability of generalizing the relation between input and output for the new data set i.e. the testing set data. This results in increasing the testing error and decreasing the overall performance of the model. There are several ways to prevent the model from overfitting, among which well-known method is early-stopping; in which training process is stopped early. But if the training is stopped too early then model fails to learn important information. Hence, training should be stopped accordingly to learn all important information without overfitting.

Many types of ANNs feature different concepts of data processing. Each type is designed differently to obtain a more precise output with less data processing time. This is achieved

by changing the networks architecture. According to Jain et al. (1996), based on the network connection pattern, i.e., their architecture, ANN is classified into two categories:

a. Feed-Forward Neural Networks (FFNNs)

FFNN has the simplest network connection pattern in which data flow in the forward direction only, starting from the input layer to hidden layers, and then to the output layer. No loops are formed in the paths of the data flow. As shown in Figure 2.4, FFNN is classified into three subcomponents: single-layer perceptron, multilayer perceptron, and radial basis function neural network (RBFNN). Single-layer perceptron, which consists of one layer, i.e., the output layer, is the simplest form of neural network. It is mainly used for classifying the linearly separable cases that use binary targets. The connection patterns of multilayer perceptron and RBFNN are the same: an input layer, as many hidden layers as required, and an output layer. The only difference between these two is the use of the data processing function. Multilayer perceptron utilizes either threshold function or sigmoidal function (A. El-Shafie et al., 2012) in each of its computational units, whereas RBFNN utilizes radial basis function as the activation function in each unit of its hidden layers. The table 2.1 presents the advantages and disadvantages of different models of FFNN. These models are generally used for time series prediction, system control, and data classification.

b. Recurrent or Feedback Neural Networks

Recurrent or feedback neural networks experience the backward flow of data in some computational cells. The data flow is not unidirectional; loops within the cells transfer back the feedback of the errors encountered in computations, with reference to the target values. The feedback of errors helps in updating the weights of the corresponding inputs. As shown in Figure 2.4, feedback neural network is classified into four subcomponents: adaptive resonance theory model, Hopfield

networks, Kohonen's networks, and competitive networks. Table 2.1 presents their advantages and disadvantages. These networks form very complex architectures, composed of a number of loops. These networks are utilized for complex computations, such as speech recognition, image processing, robotics, and process controls. This study is limited to the review of the FFNN.

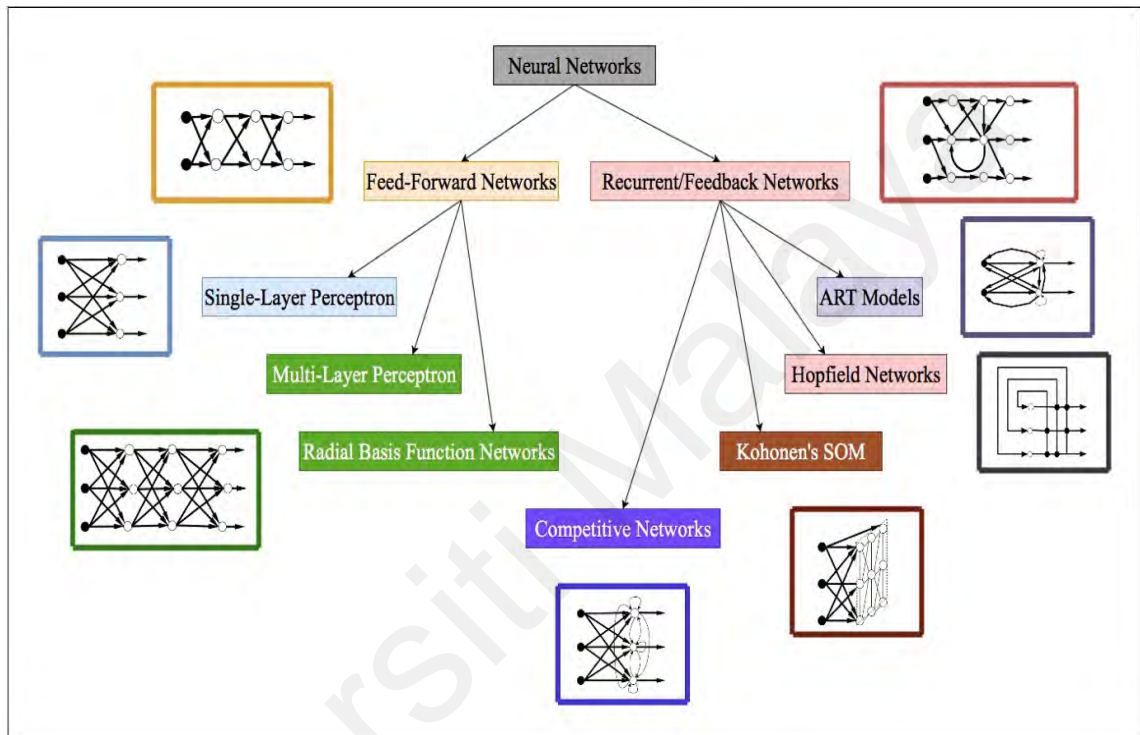


Figure 2.4 Classification of neural network(Jain et al., 1996)

Table 2.1 Advantages and disadvantages of different ANN models

Main Type	Model Name	Advantages	Disadvantages
FFNN	Single Layer Perceptron	<ul style="list-style-type: none"> • Less Computation- time • Easy to setup 	<ul style="list-style-type: none"> • Can only be used in linearly separable data
	Multi-Layer Perceptron	<ul style="list-style-type: none"> • Can be used for complex problems 	<ul style="list-style-type: none"> • Need more time for training • Can get stuck in local minima

Table 2.1, continued

Main Type	Model Name	Advantages	Disadvantages
	RBFNN	<ul style="list-style-type: none"> • Less susceptible to be stuck in local minima • Can tolerate with input noise 	<ul style="list-style-type: none"> • Classification is slow, as the network have to calculate the radial basis function for each input vector during classification
Recurrent/ Feedback Network	ART (Adaptive Resonance Theory) model	<ul style="list-style-type: none"> • Can be integrated with other models to enhance the performance 	<ul style="list-style-type: none"> • Some ART models are inconsistent. They depend upon the order of the training data
	Hopfield Network	<ul style="list-style-type: none"> • No training needed 	<ul style="list-style-type: none"> • Handles a smaller number of memories. • More number of patterns results in spurious output
	Kohonen's SOM	<ul style="list-style-type: none"> • Provides deterministic and reproducible results • Simplicity of computation 	<ul style="list-style-type: none"> • Performance depends on initialization
	Competitive Network	<ul style="list-style-type: none"> • Groups the similar pattern based on the data correlation 	<ul style="list-style-type: none"> • Susceptible to stability issue

2.3 Application of ANN

ANN models have been utilized for developing better-precision water quality predictive models (Fogelman et al., 2005; B. He et al., 2011; Holmberg et al., 2006; Lek et al., 1999; Sarangi & Bhattacharya, 2005). The computational intelligence, among which ANN is one, has become a fast-evolving area (Ehtram et al., 2017). The applications of ANN are not limited to water quality prediction. According to B. He et al. (2011), ANNs have been successfully used for reservoir operations (Aguilera et al., 2001; Chang Li-Chiu & Chang Fi-John, 2001; Suen & Eheart, 2003; Tayfur et al., 2005; Zaheer & Bai, 2003), water resources management (Bin He & Takase, 2006; Mazvimavi et al., 2005), and hydrological processes (Cigizoglu & Alp, 2004; Riad et al., 2004). Application in water

resources management includes river flow forecasting (Shamseldin et al., 2002; Teschl & Randeu, 2006), rainfall-runoff modeling (Riad et al., 2004; Wu & Chau, 2011), and water quality predictions (Gazzaz et al., 2012; Khalil et al., 2011; Palani et al., 2008; Singh et al., 2009). The present study is confined to water quality predictive systems only.

ANNs have been extensively used worldwide in the past as a predictive model for nitrogen prediction in streams. Table 2.2 lists studies on the use of ANN by various authors. Various authors had utilized different methodology, as shown in table 2.3. For nitrogen prediction, ANN was utilized for the first time probably by Lek et al. (1999). They used ANN to predict inorganic and total nitrogen concentration in streams using eight input parameters from the catchments along with the historical data of inorganic and total nitrogen. The input database was obtained from U.S. National Eutrophication Survey (NES); which had many variables in record but according to the scope of the research (prediction of stream nitrogen concentration) following eight variables were included: average annual flow; animal unit density; mean annual streamflow; the percentages of forest cover, wetland, urban areas, and agriculture areas; and the percentages of the remaining area in the catchment. Sensitivity analysis showed five different types of variation in total nitrogen concentration and three different types of variation in inorganic nitrogen concentration. The sensitivity types (or contribution) for total nitrogen concentration are: (i) Increasing sigmoid contribution: wetland and animal unit density. Low values of these independent variables lead to low (minimum) value of total nitrogen; which then enhances to reach its maximum value with the independent variable. (ii) Weakly growing contribution: agricultural areas. For low values of agricultural areas, the total nitrogen is less and likewise increasing gradually. (iii) Decreasing contribution: average annual flow and percentage of remaining area. (iv) Gaussian: Urban areas. (v) Weak contribution: percentage of forest cover. For inorganic nitrogen: (i) Growing contribution: urban and agricultural areas. For low values of urban

and agricultural areas inorganic nitrogen concentration is less and then rapidly increases with these independent variables. (ii) Gaussian: percentage of wetland areas. (iii) Decreasing contribution: Percentage of forest cover, animal unit density and remaining area. Forest cover rapidly and constantly decreases the inorganic nitrogen concentration. Other two independent variables also reduce the inorganic nitrogen concentration but at low levels only. Input variables were auto-scaled by centered and reduced variables. Autoscaling reduces the chance of domination of any one particular input variable over the prediction. This input database was divided into training and independent testing set (2/3rd and 1/3rd of the total database, respectively). Using data from 927 sites from different parts of the United States, Lek et al. (1999) developed a multilayer feed-forward ANN model having 10 neurons and 1 hidden layer, with a correlation coefficient of 0.82 for total nitrogen concentration and 0.8 for inorganic nitrogen concentration. Examining the results obtained, they concluded that the urban areas produced most of the inorganic nitrogen and animal husbandry contributed the most to the total nitrogen concentration in streams. It was assumed that fertilizers were used in less quantities as its contribution was less in stream nitrogen. Forest cover lowered the inorganic nitrogen concentration in streams and has less effect on total nitrogen concentration. Percentage of wetland areas helped in reducing the inorganic nitrogen in streams, but they increased the total nitrogen.

Table 2.2 A summary of studies that utilize ANN model for nitrogen prediction, including their specific area, location, and methods used

No.	Authors	Specific Area	Location	Method
1	Anctil et al. (2009)	Streams	Melarchez, France	Stacked multilayer perceptron
2	B. He et al. (2011)	Streams	Japan	Feed-forward model
3	Holmberg et al. (2006)	Streams	Finland	Backpropagation algorithm
4	Lek et al. (1999)	Streams	The United States	Multilayer feed-forward

Table 2.2, continued

No.	Authors	Specific Area	Location	Method
5	Suen and Eheart (2003)	Streams	Illinois, The United States	Backpropagation and radial basis
6	Sharma et al. (2003)	Drainage water	Canada	Fast backpropagation and self-organizing radial basis
7	Wang et al. (2016)	Groundwater	Australia	13 machine learning models
8	L. Zhang et al. (2016)	Lake	China	ARIMA, radial basis, and hybrid
9	M. Markus et al. (2010)	Streams	Illinois	Backpropagation, Evolutionary Polynomial Regression (EPR), and Naïve Bayes Model (NBM)
10	Amiri and Nakane (2009)	Stream	Japan	Backpropagation and Multiple Linear Regression (MLR)
11	Zeleňáková et al. (2012)	Streams	Slovakia	Dimensional analysis

The condition of the United States seemed to be critical in terms of nitrogen in streams, as four years after the study by Lek et al. (1999), a research work published by Suen and Eheart (2003) stated that nitrate has become an important problem. They conducted a study in the Upper Sangamon River, Illinois, and pointed out the use of chemical fertilizers in agriculture to be responsible for the high nitrate concentration in streams. In their study, they developed two models, RBFNN and backpropagation neural network (BPNN), and compared the models on the basis of accuracy. The parameters used for modeling were daily highest temperature, seven-day cumulative daily rainfall, daily streamflow, and Julian date. To include the common practice of fertilizer application, Julian date was used as an input parameter to the model. They used dataset of eight years, i.e., 1993–2000. To divide the dataset into the training set and testing set, two methods were adopted. In the first method, data from 1993 to 1996 were used as the training dataset and the remaining were used for testing. For the second method, the data of odd years

(i.e., 1993, 1995, 1997, and 1999) were used for training, and those of even years were used for testing. Comparing the results obtained from the models, they concluded that the odd-even years method proved to be more accurate. The overall accuracy of the first method was obtained to be 0.784 and 0.752 for BPNN and RBFNN, respectively, and that of the second method was 0.832 for both the networks. Neural network models predicted with greater precision when tested for Boolean output considering the second method. The network signaled 1 when the nitrate concentration exceeded 10 mg/l and 0 when the nitrate concentration was below 10 mg/l. Considering Boolean output, they concluded that RBFNN had a higher accuracy (0.893) than BPNN (0.866).

In 2003, a research work published in Canada by Sharma et al. (2003) stated that subsurface waters in Canada were being polluted by the nitrate from the fertilizers used in agricultural fields. Their experimental site was a field, of area 14 ha, located at the Greenbelt Research Farm of Agriculture and Agri-Food Canada, near Ottawa. The authors proposed a neural network model to assist in optimizing the use of fertilizers. The input database was collected from the experimental field for the period of 1991-1994, except the temperature and precipitation data. Data of these two variables were collected station of Agriculture and Agri-Food Canada, located 12 km from the site. Two neural network models, fast BPNN and self-organizing RBFNN, were examined aiming to select the superior network. Inputs to the model used were treatment (tillage or no tillage, i.e., whether the land was prepared or not), Julian day, rainfall per day, cumulative rainfall, total nitrogen applied, snowfall per day, and maximum and minimum temperature. Sensitivity analysis was performed to determine the optimum internal parameters of both the networks. The input data were divided into two sets: training and testing set. Training set consisted of eight input variables and two output and the testing set consisted of only the unexposed inputs from the replicate plots. For fast BPNN the parameters varied for sensitivity analysis were learning rate and number of hidden neurons. This analysis

comprised of two stages, first stage was to keep the number of hidden neurons constant at 20 and vary the learning rate from 0.02 to 0.08. Analysis of the fluctuation of error on every variation led to the selection of optimum learning rate as 0.02. In second stage, learning rate was kept constant to 0.02 and number of hidden neurons were varied from 5 to 25. Analyzing the similar way, optimum number of hidden neurons were selected as 20. Similarly, sensitivity analysis was performed for RBFNN, in two stages, by varying the tolerance and spread values from 5 to 20 and 1 to 20, respectively. The selected optimum value for tolerance and spread values were 20 and 15, respectively. Using these parameter values, both the models were further trained. Comparing the results of both networks, the authors concluded that the self-organizing RBFNN, with a correlation coefficient of 0.8079 for conventional tillage and 0.6911 for no tillage, outperformed the fast BPNN, with a correlation coefficient of 0.8017 for conventional tillage and 0.6635 for no tillage, for nitrate-nitrogen concentration prediction in drainage water.

Table 2.3 Details of methodology of the reviewed research work

No.	Authors	Duration of Data	Data Pre-processing	Internal Parameters
1	Anctil et al. (2009)	1975-1993 (Daily)	Standardization (linearly)	2 Inputs, 12 hidden neurons
2	B. He et al. (2011)	1995 (Monthly)	Sensitivity Analysis	8 Inputs, 7 hidden neurons
3	Holmberg et al. (2006)	1990-2000 (Daily)	--	13 Inputs, 1 hidden layer, 7 nodes
4	Lek et al. (1999)	One year	Sensitivity Analysis, Autoscaling	8 Input, 10 hidden neurons
5	Suen and Eheart (2003)	1993-2000, (Daily)	--	--
6	Sharma et al. (2003)	1991-1994, (Daily)	Sensitivity Analysis	Fast Backpropagation: 8 Inputs, 20 hidden neurons, learning rate: 0.02 RBFNN: Tolerance 20, spread 15

Table 2.3, continued

No.	Authors	Duration of Data	Data Pre-processing	Internal Parameters
7	Wang et al. (2016)	2006-2014, (401 samples)	--	--
8	L. Zhang et al. (2016)	2006-2011, (Monthly)	--	ARIMA: <ul style="list-style-type: none"> • Nitrogen: p=1, d=1, q=1 • Phosphorus: p=2, d=1, q=1 RBFNN: <ul style="list-style-type: none"> • 2 hidden layers • Training width $\sigma = 0.6$
9	M. Markus et al. (2010)	1994-1999, (Weekly)	--	ANN: 4 Input, 2 hidden nodes, epochs: 100,000; performance gradient: 1E-10; goal: zero EPR equations: $N_{t+1} = 0.827N_t$ $N_{t+1} = 0.659N_t + 0.560N_t\sqrt{Q_t}$ NBM equations: $N_{t+1} = f[N_t, Q_t, P_t, T_t]$ $N_{t+1} = f[N_t, Q_t, Q_{t-1}, P_t, P_{t-1}, T_t, T_{t-1}]$
10	Amiri and Nakane (2009)	2001, (Monthly)	Statistical Analysis	6 Input nodes, 2 hidden nodes, 1 output nodes, 11,600 epochs
11	Zeleňáková et al. (2012)	2003-2010, (Monthly)	Sensitivity Analysis	Dimensional analysis equations: $\pi_1 = 0.0039\pi_2^{13.805}$ $\pi_1 = 0.1868\pi_2^{9.7892}$

Holmberg et al. (2006), predicted the future data of total organic carbon, total nitrogen, and total phosphorus in streams, considering the climate change effect and utilizing the data of three streams (Kelopuro, Hietapuro and Valkea-Kotinen) located in two catchments of same name (Hietajärvi) in Finland. They developed a BPNN model employing the database of 13 input variables: month of data sampling, mean temperatures of 3 and 10 preceding days, runoff of sampling day, maximum and minimum runoffs of

three preceding days, days of peak flow, days of low flow, catchment area, fractions of lake area and peatland area with respect to catchment area, catchment latitude, and elevation. This database was collected from the catchment, except the daily temperature and precipitation; which was collected from the nearby Finnish Meteorological Institute weather station, Lammi, from 1990 to 2000. Samples of these variables were divided into two sets: training set and testing set. The samples were allocated into these sets by randomly choosing, provided it was ensured that the highest and lowest 10-percentile data were included in both the sets. While training, they were to test all the possible set of models with the available inputs, hence, they varied the number of inputs from 2 to 16, fixing the number of hidden layer to one and the neurons in the hidden layer was set as the integer part of $(1 + \text{number of inputs})/2$. Training 10 sessions for each combination, resultant models were analyzed on the basis of their efficiency. The model resulted the best efficiency with 13 input variables and one hidden layer with 7 nodes, having the values of flux efficiencies of total organic carbon, total nitrogen, and total phosphorus as 0.94, 0.92, and 0.90, respectively. Using this model, they forecasted the total nitrogen data until 2050. They stated that if there is a low change in climate, then the total nitrogen flux will be near the value in 2005, but for a scenario of high change in climate, the nitrogen flux will increase by 26%, with respect of the value in 2005.

Similar conditions have been stimulated in Melarchez, a catchment near Paris, France, where Anctil et al. (2009) investigated an agricultural catchment area, to develop a neural network model for predicting the nitrate-nitrogen flux. Considering the soil moisture at different depths as the input parameter, the authors analyzed its effect on the nitrate-nitrogen flux. They developed a stacked multilayer perceptron model focusing mainly on the selection of best performing model among the list of models developed based on different combinations of input variables and neurons in hidden layers. 50 models were trained for each combination of inputs and neurons in hidden layers. Neurons in hidden

layers were varied from 2 to 20. Every issue was tested discretely to make the final decision on the basis of the model accuracy. They had 12 different options for the input parameter: same-day stream flow, previous-day stream flow, increment in the flow from the previous day, same-day precipitation, previous-day precipitation, same-day historical mean flux, increment in the historical mean flux from the previous day, same-day 10 cm-, 20 cm-, 40 cm-, 80 cm-, and 120 cm-depth soil moisture indices. These input variables were collected from the gauge station for the period of 1975 to 1993. Since the important step, in pre-processing of data, is standardization (Akrami et al., 2014), all the input variables were ensured to be on the same scale by standardizing them linearly such that their standard deviation as one and mean as zero. After optimizing, the final model had two input parameters (same-day stream flow and same-day 80 cm-depth soil moisture index), 12 neurons in hidden layers, and Levenberg-Marquardt with Bayesian regulation as the calibration procedure, which performed well with an efficiency index of 0.888. The utilization of soil moisture content at different depths revealed that the soil moisture also had an effect on nitrate-nitrogen flux generated from the agricultural field.

Since a large number of input variables are available to decide for the neural network, these inputs should be chosen using sensitivity analysis (May & Sivakumar, 2009). Numerous authors have provided models with different sets of input parameters, which according to them, were suitable for their models (Table 2.4). B. He et al. (2011) investigated 59 river basins all over Japan and developed an FFNN to predict the monthly total nitrogen concentrations in streams. They had to choose the most important independent input variables from a set of 16 input variables: the area of each basin, amount of fertilizer applied in each basin, average temperature, precipitation, sunshine duration and river discharge of each basin, ratio of paddy area, farmland area, forest area, bare land area, urban area, road area, river area, lake area, seashore area, and other land areas in the total basin area. This input database was collected from different sources. The

land use variables were collected from Ministry of Land, Infrastructure, Transport and Tourism (MLIT land use database), digital database of Japanese. Total nitrogen concentration was collected from MLIT water information system. Sunshine duration, precipitation and temperature data were obtained from Automated Meteorological Data Acquisition System. The input data were divided into three subsets: Training, overfitting test and validation subsets. Among the data of 59 river basins, 40 river basin data were used for training and overfitting test (80% and 20%, respectively). The remaining 19 river basin data were never exposed to the network for training and were used for validation only. FFNN was trained with backpropagation algorithm with different combinations of input variables and internal parameters: input variables were varied from 7 to 9, number of hidden layer was fixed to one with number of neurons in it fixed to 7 and 8. Analyzing the results of all the trained network on the basis of coefficient of regression, authors found that the model with eight input variables (river discharge, average temperature and precipitation of the each basin, amount of fertilizer applied in each basin, the proportions of forest land area, urban land area, road area, and other areas in the total basin area) and one hidden layer with seven nodes provided the best accuracy with R^2 for training as 0.96, R^2 for validation as 0.84, and R^2 for overfitting as 0.90.

Table 2.4 A summary of studies that utilize ANN model for nitrogen prediction, including their input variables, prediction variables, and accuracy

No.	Authors	Input Variables	Prediction Variables	Accuracy
1	Anctil et al. (2009)	<ul style="list-style-type: none"> • Same-day stream flow • Same-day 80 cm-depth soil moisture index 	Nitrate-nitrogen flux	Efficiency index = 0.888

Table 2.4, continued

No.	Authors	Input Variables	Prediction Variables	Accuracy
2	B. He et al. (2011)	<ul style="list-style-type: none"> • River discharge • Average temperature and precipitation of each basin • Amount of fertilizer applied in each basin • Proportions of forest land area, urban land area, road area, and other areas in the total basin area 	Monthly total Nitrogen concentrations	$R^2_{training}=0.96$ $R^2_{Validation}=0.84$ $R^2_{Overfitting}=0.9$
3	Holmberg et al. (2006)	<ul style="list-style-type: none"> • Month of data sampling • Mean temperatures of 3 and 10 preceding days • Runoff of sampling day • Maximum and minimum runoffs of 3 preceding days • Days of peak flow, days of low flow • Catchment area • Fractions of lake area and peatland area with respect to catchment area • Catchment latitude and elevation 	<ul style="list-style-type: none"> • Total organic carbon • Total nitrogen • Total phosphorus 	Flux efficiency: <ul style="list-style-type: none"> • Total organic carbon = 0.94 • Total nitrogen = 0.92 • Total phosphorus = 0.90
4	Lek et al. (1999)	<ul style="list-style-type: none"> • Average annual flow • Animal unit density • Mean annual streamflow • Percentage of forest cover, wetland, urban, agriculture and the percentage of remaining area in the catchment 	Inorganic and Total Nitrogen concentration	Correlation coefficient: Total Nitrogen = 0.82 Inorganic Nitrogen = 0.8

Table 2.4, continued

No.	Authors	Input Variables	Prediction Variables	Accuracy
5	Suen and Eheart (2003)	<ul style="list-style-type: none"> • Daily highest temperature • Seven-day cumulative daily rainfall • Daily streamflow • Julian date 	Nitrate concentration	Overall accuracy: <ul style="list-style-type: none"> • Method one: <ul style="list-style-type: none"> • BPNN = 0.784 • RBFNN = 0.752 • Method two: <ul style="list-style-type: none"> • BPNN = 0.832 • RBFNN = 0.832 • Boolean output (Method two) <ul style="list-style-type: none"> • BPNN = 0.866 • RBFNN = 0.893
6	Sharma et al. (2003)	<ul style="list-style-type: none"> • Treatment • Julian day • Rainfall per day • Cumulative rainfall • Total Nitrogen applied • Snowfall per day • Maximum and Minimum Temperature 	Nitrate concentration	Correlation coefficient <ul style="list-style-type: none"> • RBFNN <ul style="list-style-type: none"> • Tillage = 0.8079 • No tillage = 0.6911 • BPNN <ul style="list-style-type: none"> • Tillage = 0.8017 • No Tillage = 0.6635
7	Wang et al. (2016)	<ul style="list-style-type: none"> • Scenario 1 <ul style="list-style-type: none"> • Nutrients (Dissolved Organic Nitrogen (DON), total nitrogen, NH_4^+, NO_x^-) • Landscape (vegetation, land use, and soil) • Hydrological conditions (surface water subarea, groundwater subarea, and catchment area) • Sampling condition (temperature, sample depth, and sampling date, pH) • Scenario 2 <ul style="list-style-type: none"> • Total nitrogen • All other non-nutrient data 	DON	R ² of best models: <ul style="list-style-type: none"> • Scenario 1 <ul style="list-style-type: none"> • Cubist = 0.897 • Bagged multivariate adaptive regression spline (Bagged mars) = 0.882 • Random forest (RF) = 0.856 • Scenario 2 <ul style="list-style-type: none"> • Cubist = 0.849 • Bagged mars = 0.887 • RF = 0.858

Table 2.4, continued

No.	Authors	Input Variables	Prediction Variables	Accuracy
8	L. Zhang et al. (2016)	Monthly data for total nitrogen	<ul style="list-style-type: none"> • Monthly total nitrogen • Monthly total phosphorus 	<p>Mean absolute percentage error:</p> <ul style="list-style-type: none"> • Nitrogen <ul style="list-style-type: none"> • ARIMA = 18.194% • RBFNN = 34.633% • Hybrid = 7.017% • Phosphorus <ul style="list-style-type: none"> • ARIMA = 27.299% • RBFNN = 126.957% • Hybrid = 14.528%
9	M. Markus et al. (2010)	<ul style="list-style-type: none"> • Observed weekly river discharge • Precipitation • Air temperature • Nitrate-nitrogen concentration 	<ul style="list-style-type: none"> • Weekly nitrate-nitrogen 	<p>Root mean square error (RMSE) for ANN:</p> <ul style="list-style-type: none"> • Training = 0.787 mg/l • Testing = 0.935 mg/l <p>RMSE for EPR:</p> <ul style="list-style-type: none"> • Training = 0.991 mg/l • Testing = 1.010 mg/l <p>Critical success index for NBM:</p> <ul style="list-style-type: none"> • NBM1: <ul style="list-style-type: none"> • Training = 0.214 • Testing = 0.200 • NBM2: <ul style="list-style-type: none"> • Training = 0.286 • Testing = 0.188
10	Amiri and Nakane (2009)	<ul style="list-style-type: none"> • Percentage land use <ul style="list-style-type: none"> • Urban • Forest • Agriculture • Grassland • Water body • Population density 	Total nitrogen	<p>R² Value:</p> <ul style="list-style-type: none"> • BPNN = 0.94 • MLR = 0.85
11	Zeleňáková et al. (2012)	<ul style="list-style-type: none"> • Stream discharge • Catchment area • Stream velocity • Temperature of air and water • Concentration of pollutant 	Nitrogen and phosphorus concentration	<p>Average Uncertainty:</p> <ul style="list-style-type: none"> • Nitrogen = 31.33% • Phosphorus = 32.30%

In addition to ANN, other machine learning methods can also be used to predict nonlinear environmental variables. Wang et al. (2016) compared 13 machine learning models, including ANN, on the basis of precision in the prediction of DON in groundwater in urban areas in southwestern Australia. These 13 machine learning models are classified into five different groups: (1) tree-based and rule-based model (generalized boosted model (GBM), RF, conditional inference random forest (cforest), and cubist); (2) kernel-based machine learning model (Gaussian process with radial basis function kernel (GPR), Gaussian process with linear kernel (GPL), support vector machine with radial basis function kernel (SVMR), and support vector machine with linear kernel (SVML)); (3) generalized stepwise linear regression models (bagged mars, multivariate adaptive regression spline (mars), and generalized linear model with stepwise feature selection (GLM)); (4) instance-based model (k-nearest neighbors (KNNs)); and (5) ANNs. Using 401 groundwater samples (60% for training and 40% for testing), the models were examined based on two scenarios: (1) to train the models with all the data available such as nutrients (DON, total nitrogen, NH_4^+ , and NO_3^-), landscape (vegetation, land use, and soil), hydrological conditions (surface water subarea, groundwater subarea, and catchment area), and sampling conditions (temperature, sample depth, sampling date, and pH); (2) to train the models with only total nitrogen and all other non-nutrient data. Database of nutrients were obtained from Western Australian Department of Water for the period of 2006-2014. ArcGIS spatial mapping feature provided the data of soil type, land use and vegetation type. These models were analyzed on the basis of their RMSE and R^2 values and compared with the manually calculated DON (DONcal) (Figure 2.5). Analysis of all the results revealed that scenario 1 produced lower errors in models than scenario 2, stating that nutrients can improve the performance of models. Among the 13 tested models, three models showed higher R^2 value. For scenarios 1 and 2, the cubist

model had R^2 values of 0.897 and 0.849; bagged mars, 0.882 and 0.887; random forest, 0.856 and 0.858; and ANN, about 0.72 and 0.65, respectively.

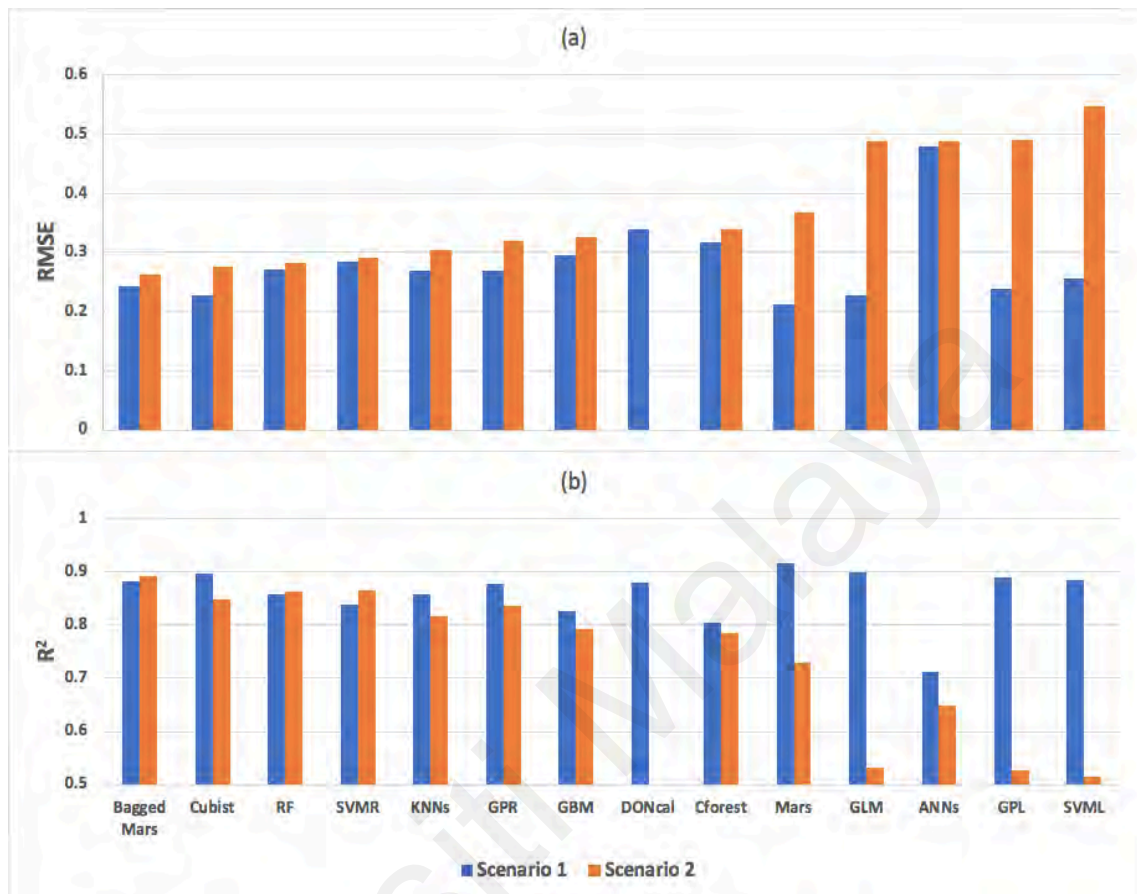


Figure 2.5 Comparison of 13 different models results with the DONcal

L. Zhang et al. (2016) compared ARIMA model, RBFNN model, and hybrid ARIMA-RBFNN model based on the analysis and prediction of water quality in Chagan lake, China. Database of water quality was collected from “The Second Songhua River Diversion Project Record” from Chinese Academy of Science. The water quality parameters utilized for analysis were monthly total nitrogen and total phosphorus for the period of 2006–2011. The parameters of ARIMA model for total nitrogen were $p=1$, $d=1$ and $q=1$ and for total phosphorus were $p=2$, $d=1$ and $q=1$. Water quality data from 2006 to 2010 were used for training and the trained model was used for prediction of water quality data of 2011. The width of training, σ , was 0.6 for RBFNN model with 2 nodes in hidden layers. ARIMA-predicted values were linearly super-positioned with RBFNN-

derived ARIMA residual prediction values to generate the hybrid ARIMA-RBFNN model. These models were analyzed on the basis of their RMSE and mean absolute percentage error. Results showed that RBFNN model had bad prediction results for total phosphorus. Though, this model had learned the pattern of total nitrogen, but the predicted values were not satisfactory. Though ARIMA model did not had high prediction accuracy but it had successfully learned various trends for both total nitrogen and total phosphorus. Analyzing the results obtained, the mean absolute percentage error for the monthly total nitrogen was 18.194%, 34.633%, and 7.017% for ARIMA, RBFNN, and hybrid ARIMA-RBFNN, respectively, and the mean absolute percentage error for the monthly total phosphorus was 27.299%, 126.957%, and 14.528% for ARIMA, RBFNN, and hybrid ARIMA-RBFNN, respectively. Following the results, it was stated that hybrid models had more capacity in predicting nonlinear variables.

M. Markus et al. (2010) developed three models—BPNN, EPR and NBM—for predicting weekly nitrate-nitrogen in a small agricultural watershed in Illinois. For the ANN part, the authors utilized observed weekly river discharge, precipitation, air temperature, and nitrate-nitrogen concentration as input variables. The study used the historical data of nitrate-nitrogen concentration was collected from the Upper Sangamon River near Decatur for the period of 1994-1999. Employing half of the data for training and the other half for testing, they predicted the weekly data of nitrate-nitrogen in streams. The input selection was performed on the basis of trial and error with two sets of variables and their time lags. First set consisted of four variables: N_t, Q_t, T_t, P_t ; and the second set consisted of four variables and three time lags $N_t, Q_t, T_t, P_t, Q_{t-1}, T_{t-1}, P_{t-1}$. First set predicted better results and hence was used for ANN modeling. ERP model has the capability of selecting the input subset, hence it is fed with the larger input set, the second set. In case of NBM both the set were used for modeling. For modeling in ANN part, the internal parameters selected were: epochs: 100,000; performance gradient: 1E-10; goal: zero;

number of hidden nodes: 1, 2, 3, 4 and 5; input variables: 4 (air temperature, discharge, nitrate-N concentration and precipitation) and output variable: 1 (next week nitrate-N concentration). The results indicated that the ANN with 2 nodes showed more accurate results in terms of RMSE as 0.787 mg/l and 0.935 mg/l for training and testing, respectively. For EPR, two models (EPR1 and EPR2) were generated which had their equations as: $N_{t+1} = 0.827N_t$ and $N_{t+1} = 0.659N_t + 0.560N_t\sqrt{Q_t}$, respectively. The RMSE obtained for EPR1 was 1.092 mg/l for training and 1.170 mg/l for testing. The RMSE obtained for the EPR2 was more accurate: 0.991 mg/l and 1.010 mg/l for training and testing, respectively. The NBM model utilized two categories: high and low values for variables. Each variable, except for nitrate-N concentration, had its categories divided by the average values as threshold. For nitrate-N concentration the separation point was the emergency cutoff level (8.5 mg/l). NBM1 and NBM2 were the two models tested with the equations as: $N_{t+1} = f[N_t, Q_t, P_t, T_t]$ and $N_{t+1} = f[N_t, Q_t, Q_{t-1}, P_t, P_{t-1}, T_t, T_{t-1}]$, respectively. The results of these models indicated that, for low concentration, NBM1 had accurately predicted 79 of 80 concentrations, but for high concentrations the prediction rate was 2 of 9. For NBM2, the predicted high flows (10) were somewhat similar to the observed ones (9). However, the false alarm rate for NBM2 was higher (7) than NBM1 (1). The critical success index for NBM1 was obtained as 0.214 and 0.200 for training and testing, respectively and that for NBM2 was 0.286 and 0.188 for training and testing, respectively. Authors concluded that none of these models can be considered superior based on these analysis criteria, hence, suggesting multi-tool approach. In their previous study, Momcilo Markus et al. (2003), they compared the ANN model and linear regression model to calculate the uncertainty in forecasting the weekly nitrate-nitrogen in the Sangamon River, Illinois. They stated that the ANN model was more accurate than the linear regression model. The ANN model surpassed the linear regression model by 3.30% and 4.42% of RMSE in testing and training phases, respectively.

Amiri and Nakane (2009) compared BPNN and MLR on the basis of the total nitrogen prediction in streams. The study was conducted in the Chugoku district of Japan, which contains 21 river basins. Total nitrogen database, for year 2001, was collected from prefecture offices from Okayama, Shimane, Hiroshima, Tottori and Yamaguchi. Six input variables were used for the prediction, which included five variables for land cover percentage (urban area, forest area, agriculture area, grassland, and water body) and the last variable for population density. The total nitrogen was predicted by utilizing 60% of the data for training, 25% for controlling, and the remaining 15% for testing. BPNN consisted of six input nodes for the corresponding six input variable, one hidden layer and one node in output layer for total nitrogen prediction. The optimum number of nodes in hidden layer were selected by varying the nodes from 0 to 13 and training the network 5 times for each variation and evaluating them on the basis of correlation coefficient. The selected optimum BPNN had the following internal parameters: input nodes:6, hidden layer:1, hidden layer node:2, output node:1, epochs:11,600. MLR model had the same inputs as for the BPNN. For MLR modeling, a normality test was conducted for total nitrogen and land cover data using Sharpio-Wilk test having p-value less than 0.05. Models were analyzed on the basis of regression statistics and coefficient of the model (if the resultant was normally distributed). Final regression model was developed by using backward approach. The goodness of fit of the models was evaluated by regression of observed versus predicted and scatter plot. Comparison of the results for both the models showed that the backpropagation model ($R^2 = 0.94$) predicted the results more precisely than the multiple regression model ($R^2 = 0.85$)

Zeľňáková et al. (2012) predicted nitrogen and phosphorus concentrations in river Laborec in Slovakia, employing dimensional analysis method. They used Buckingham theorem to develop a prediction model utilizing important variables such as stream discharge, area of catchment, stream velocity, temperatures of air and water, and pollutant

concentration. The equation established for nitrogen concentration was: $\pi_1 = 0.0039\pi_2^{13.805}$ and for phosphorus was: $\pi_1 = 0.1868\pi_2^{9.7892}$. These models were tested for the data of eight years (2003-2010); which was collected from Slovak Hydrometeorological Institute and Slovakian Water Management Company in Košice. Sensitivity analysis of the model stated that air and water temperature have major influence on the prediction of concentration of nitrogen and phosphorus. Velocity and flow of water have less influence and the catchment area has no influence on the prediction. By exploring the results of the model, it was found that the model equations calculated the prediction values with an average uncertainty of 31.33% for nitrogen and 32.30% for phosphorus.

Information of these reviewed literature were used for the development of ANN models for prediction of nitrate-nitrogen and ammonia nitrogen for the stations Lui and Kajang, as described in chapter 3.

CHAPTER 3. ANN MODEL FOR RIVER-NITROGEN PREDICTION

3.1 Introduction

Chapter 3 presents an article, which contains the development process of ANN models for prediction of nitrate-nitrogen and ammonia-nitrogen in Langat River basin in Malaysia. This chapter elaborates the development process of four prediction models for two different pollutant at two different geographical locations. It contains introduction, literature review, methodology, results, conclusion and discussion sections.

This article presents a training and a selection approach to develop an optimum artificial neural network model for predicting monthly average nitrate-nitrogen and monthly average ammonia-nitrogen. Several studies have predicted these compounds, but most of the proposed procedures do not involve testing various model architectures in order to achieve the optimum predicting model. Additionally, none of the models have been trained for hydrological conditions such as the case of Malaysia (P. Kumar et al., 2020). This article presents models trained on the hydrological data from 1981 to 2017 for the Langat River in Selangor, Malaysia. The model architectures used for training are General Regression Neural Network (GRNN), Multilayer Neural Network and Radial Basis Function Neural Network (RBFNN). These models were trained for various combinations of internal parameters, input variables and model architectures. Internal parameters consisted of hidden layers, nodes in hidden layers, spread values, epochs and data division. Input variables used for modelling are rainfall, stream water level and stream discharge. Post-training, the optimum performing model was selected based on the regression and error values, plot of predicted versus observed values and Nash-Sutcliffe Efficiency. Taylor diagram was also considered during optimum model selection process. Four different optimum modes were selected for nitrate-nitrogen and ammonia-nitrogen for both the stations (i.e. station Lui and station Kajang).

3.2 Literature Review

Artificial Neural Network (ANN) models, a computational intelligence model, have been extensively used for prediction over the last few decades (Ehteram et al., 2017). Various types of ANN models have been developed by researchers for different fields of prediction and forecasting. Every model was developed with some special feature to make it compatible for different levels of complexity of modelling such as prediction, image processing, face recognition, voice recognition, etc. ANN models are capable to mathematically relate the input to the desired output, forming a completely data-driven model. An ANN model trains itself with the historical data of the desired output and using the training parameters, it predicts the upcoming data. It has various internal parameters (such as hidden layers, nodes in hidden layers, maximum epochs, spread values, etc.) that need to be adjusted to get the results with high accuracy. ANN has the unique feature of learning the crests and troughs of the historical data used for a model training. B. He et al. (2011) reported that, ANN models are used for reservoir operations (Aguilera et al., 2001; Chang Li-Chiu & Chang Fi-John, 2001; Suen & Eheart, 2003; Tayfur et al., 2005; Zaheer & Bai, 2003), water resources management (Bin He & Takase, 2006; Mazvimavi et al., 2005) and hydrological processes (Cigizoglu & Alp, 2004; Riad et al., 2004).

Several studies, including (Anctil et al., 2009; Czernuszenko, 1987; M. Markus et al., 2010; Sharma et al., 2003; Suen & Eheart, 2003), used ANN for predicting nitrogenous compounds in rivers across the world. Most of these studies have not considered the application of different architectures of ANN, such as multilayer, RBFNN and GRNN. In addition, none of the models have been trained for the Malaysian hydrological conditions. An ANN model trained for a particular set of input data for some locations cannot be used efficiently at different locations as the pattern of the historical input data may not be same as the previous ones. In other words, such ANN models are site specific

and may not be implemented before further training on other sites. Hence, there is a need for the development of an efficient model for the Malaysian rivers.

3.3 Methodology

3.3.1 Data collection and interpolation

Water quality (mainly comprising of nitrate-nitrogen and ammonia-nitrogen concentrations), water level (WL) and discharge (Q) data of Lui and Kajang water quality stations and rainfall (RF) data of the nearest rainfall gauge stations of Lui and Kajang were collected. WL and Q data collected was the data of Langat river at both the stations. These data were obtained from Department of Irrigation and Drainage (DID), Malaysia, for the period of 1981 – 2017. The target variables (i.e. nitrate-nitrogen and ammonia-nitrogen) obtained were measured on monthly basis. To align with the target variables, rest of the data were converted from daily data to monthly data, by considering the 30-day average values as an average value for a particular month. The input variables selected for the current study are RF, WL and Q, as the concentrations of nitrate-nitrogen and ammonia-nitrogen in rivers depend on rainfall, water flow (ANI et al., 2010) and depth (Czernuszenko, 1987). Nitrate-nitrogen concentration reduces when river receives short and intense rainfall water and it may increase if the rainfall is prolonged one, as water leaches through the soil in the latter case, collecting nitrate-nitrogen from the soil. Water flow controls the transformation processes of nitrate-nitrogen and ammonia-nitrogen i.e. nitrification and denitrification (ANI et al., 2010). Czernuszenko (1987) reported that the concentration of pollutants depends on depth of the river. Those rivers that flow at greater depth has lower concentration of pollutant in comparison to those flowing at lower depth.

Being an important step in data standardization (Akrami et al., 2014), data received was pre-processed as the data received for the study had some gaps with respect to time. There were also few irrelevant data points such as, exceptionally high values. Such values were adjusted by replacing them with the values that were somewhat relevant to the surrounding values. During data standardization, the gaps in the data, with respect to time, was completed using interpolation technique. For interpolating the missing data, spline curve, normalized spline curve and ANN model were used. Spline curve and normalized spline curve did not provide satisfactory results. The interpolated values obtained from these curves consisted of some negative values for nitrate-nitrogen and ammonia-nitrogen concentrations. These negative values were not acceptable for processing further, as practically, the concentration cannot be a negative value. Hence, feed-forward ANN model was used for interpolation of missing data points, which proved to be more accurate in interpolating the values. The ANN model opted for interpolation was a simple feed-forward model with backpropagation training technique. The model consisted of two hidden layers with 10 nodes in each hidden layer. The model was trained with 100 epochs, TRAINLM training function, LEARNINGDM learning function and TANSIG transfer function to obtain overall regression values of 0.75. The model was trained on the complete dataset except the missing data points, which were later used to calculate the missing values. Hence, the dataset was completed after fixing the missing data points using the interpolation technique.

Figure 3.1 represents the plots of chronological data points of rainfall, water level and discharge for station Lui and station Kajang. The interpolated monthly average data of nitrate-nitrogen and ammonia-nitrogen concentrations for station Lui and station Kajang are presented in Figure 3.2, with the data points arranged chronologically.

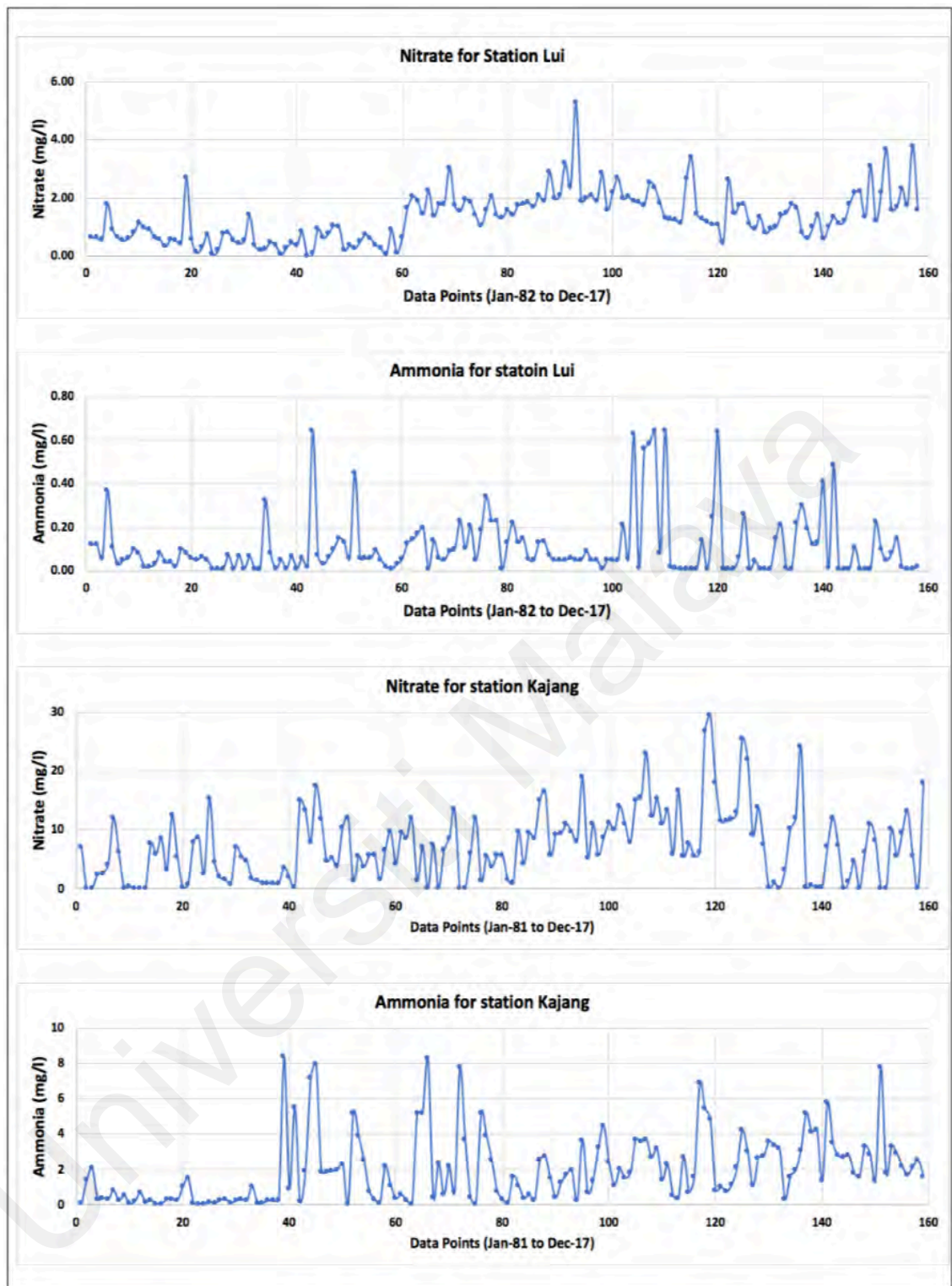


Figure 3.1 Plot of monthly average interpolated data of Nitrate and Ammonia



Figure 3.2 Plot of monthly average data of rainfall, water level and discharge

Statistical Analysis of the data, as presented in Table 3.1, reported that the average rainfall received at station Lui and station Kajang were approximately same (6.85 mm and 6.89 mm, respectively); with the maximum rainfall received at both the stations as 16.70 mm and 18.75 mm, respectively. Water level and discharge differed at the station Lui and the station Kajang due to different geographical locations. Station Lui, situated in mountainous region, has the water level of 76.17 m. Station Kajang, situated in almost plane region, has the water level of 22.7 m. Nitrate-nitrogen at the station Lui has an average value of 1.34 mg/l, which increases to an average value of 7.32 mg/l at the station Kajang. In addition to this, ammonia-N at the station Lui has an average value of 0.11 mg/l, which reaches to 1.96 mg/l, at station Kajang. As explained in above chapter, this

increment in nitrate-nitrogen and ammonia-nitrogen concentrations between the two stations represents the addition of nitrogen pollutant the path of the Langat river.

Table 3.1 Statistical analysis of the data for both stations

Station	Parameter	Average	Standard Deviation	Skewness	Maximum	Minimum
Lui Station	RF (mm)	6.85	3.57	0.43	16.7	0.1
	WL (m)	76.17	1.05	-0.69	77.41	74.41
	Q (m ³ /s)	2.19	1.02	1.29	7.66	0.68
	Nitrate-N (mg/l)	1.34	0.87	1.01	5.3	0.01
	Ammonia-N (mg/l)	0.11	0.14	2.41	0.64	0.01
Kajang Station	RF (mm)	6.89	3.87	0.61	18.75	0
	WL (m)	22.7	0.34	0.34	23.8	22
	Q (m ³ /s)	12.53	7.47	1.21	40.6	2.5
	Nitrate-N (mg/l)	7.32	6.09	0.998	29.5	0.02
	Ammonia-N (mg/l)	1.96	1.93	1.39	8.4	0.05

3.3.2 Data division

For ANN multilayer modelling, input data has to be divided into three sets: training, validation and testing set (Ahmed et al., 2012). The training set is used for adoption of the weights of neural network (de Gennaro et al., 2013; Najah et al., 2011), whereas the validation set is used for preventing the model from overfitting. ANN does not adjust its weights on the validation set. The testing set is used only for testing the final solution in order to confirm the actual predictive power of the network.

By default, ANN modeling system divides the input data as: 70% for training set, 15% for validation set and remaining 15% for testing set; by selecting randomly from the input set. Setting the division function as random, the network will randomly select different training, validation and testing set every time the network is trained. Hence, any conclusion cannot be drawn on the basis of accuracy by changing any internal parameter

because training, validation and testing set keeps on changing every time the network is trained. Hence, for this study, the division function was selected as division index; in which the separate index numbers were provided for the three sets. These index numbers were selected from the input list such that all the three sets were statistically identical. These indices were selected randomly such that the mean values of all the three sets were close to each other. As suggested by Lagos-Avid and Bonilla (2017) and Lu et al. (2019), while selecting, it was ensured that the maximum and minimum output values were lying in the training set, so that network is trained for all patterns of the data available. After selecting the best set, it was stored and then used for all the network training for particular pollutant and station. Selection of indices was done separately and before training the neural network. Four set of data division were created which had the following percentage division:

1. Training = 75%, Validation = 12.5% and Testing = 12.5%
2. Training = 80%, Validation = 10% and Testing = 10%
3. Training = 85%, Validation = 7.5% and Testing = 7.5%
4. Training = 90%, Validation = 5% and Testing = 5%

3.3.3 ANN training and parameter selection

Three different ANN models were trained for this research. The results of all the three models were then compared to obtain the optimum model. GRNN, multilayer and RBFNN models were trained at different set of internal parameters. Separate training was carried out for nitrate-nitrogen and ammonia-nitrogen for station Lui and station Kajang. After training and testing the models on all combinations of the internal parameters, the optimum model was selected based on the regression values, mean square error, mean absolute error and Nash-Sutcliffe Efficiency. Table 3.2 represents different values of internal parameters that were tested for ANN to get the most accurate model. Monthly

average rainfall, water level and discharge were three inputs used in the model and also three different combinations of two inputs were used for training. Manually selected spread values were used for GRNN and RBFNN models. In multilayer, different models were developed having hidden layers 1, 2 and 3; having nodes in each hidden layer ranging from 2 to 10. Multilayer models were trained with epochs ranging from 100 to 1000. All the models were trained on Matlab platform; in which certain set of codes made it possible to train thousands of ANN models with each possible combination of different input variables and internal parameters.

Table 3.2 Different Input and internal parameters for different ANN models

	Input:	Spread Values:	Data Division:
GRNN	<ol style="list-style-type: none"> 1. Three Input (RF, WL, Q) 2. Two Input (RF, WL) 3. Two Input (RF, Q) 4. Two Input (WL, Q) 	0.001; 0.002; 0.004; 0.005; 0.006; 0.007; 0.008; 0.009; 0.01; 0.02; 0.04; 0.06; 0.08; 0.09; 0.1; 0.2; 0.4; 0.6; 0.8; 1; 2; 3; 4; 5	<ol style="list-style-type: none"> 1. Training =75% 2. Training =80% 3. Training =85% 4. Training =90%
Multilayer	<ol style="list-style-type: none"> 1. Three Input (RF, WL, Q) 2. Two Input (RF, WL) 3. Two Input (RF, Q) 4. Two Input (WL, Q) 	Hidden Layers: 1, 2, 3 Nodes: 2, 3....10 Epochs: 100, 200, 300 ...1000	<ol style="list-style-type: none"> 1. Training =75% 2. Training =80% 3. Training =85% 4. Training =90%
RBFNN	<ol style="list-style-type: none"> 1. Three Input (RF, WL, Q) 2. Two Input (RF, WL) 3. Two Input (RF, Q) 4. Two Input (WL, Q) 	Spread Values: 0.01; 0.02; 0.04; 0.06; 0.08; 0.09; 0.1; 0.2; 0.4; 0.6; 0.8; 1; 2; 3; 4	<ol style="list-style-type: none"> 1. Training =75% 2. Training =80% 3. Training =85% 4. Training =90%

In comparison to the problems associated with the selection of the size of the input and output layers the issues associated with the size and number of the hidden layer are significantly more difficult to resolve. There are no strict guidelines available to select the correct number of hidden layers required or the needed number of hidden neurons as well. The exact requirements for each layer remain very application-specific despite the development of rule-o-thumb guidelines derived from the experience. This situation is in direct contrast to the process of defining the number of neurons in the input and output

layer, where the stimulus and the desired response provide considerable guidance as to the number of input and output neurons required to perform a specified task.

The size of the hidden layer including the hidden neurons, more specifically the number of neurons (hidden) require a specified task that is intimately linked to the role of hidden neurons. In fact, the size of the hidden neurons affects not only how well the network is able to detect important features of the risk curves, but also its ability to generalize and make decisions based on curves which are not encountered during training. An indication of the importance of the architecture of the hidden layers is that hidden layers intermediately form the first response of the input data patterns. In case that there is an extra number of hidden neurons available within the layer, the final architecture might not be able to achieve generalization. On the other hand, a few numbers of neurons might lead to the inability to custom satisfactory and tolerate middle representations to be able to encode the final architecture to perceive and sense the important characteristics and attributes of the input pattern.

In the extreme, the loss of generalization due to too many hidden neurons can result in the grand-mothering effect. The grand-mothering effect refers to the condition where, if the number of hidden neurons is equal to the number of stimulus patterns employed during training, the network is capable in theory of perfectly memorizing these input patterns. However, in this situation, the network does not learn to detect patterns in the stimulus, but rather uses each neuron in the hidden layer to memorize the desired response of one of the training stimuli. Without the ability to detect important features of a stimulus, the network is unable to generalize.

Currently, the most common approach available to identify the appropriate number of hidden neurons in the hidden layer is the trial-and-error approach. Using the trial-and-error approach is mainly to try a training process with a different number of neurons in the hidden layer and evaluate the model's outputs compared with the desired actual

outputs since the feature of the input data and the aptitude to generalize these results. The optimal architecture of the network is the network that could achieve good results and sense the important characteristics of the input pattern with a minimal number of hidden neurons.

While the experimental approach to find the optimal number of hidden neurons can be implemented successfully, it is very time consuming and requires the investigation of a large number of neural networks. An alternative procedure for finding the optimal number of neurons could be adjusted. This procedure, referred to as the dynamic-node-creation method, progressively adds neuron to the hidden layer whenever the network can no longer be improved using the current number of hidden neurons. A practical metric to determine how close the network's output is to the desired response is the sum of the squared differences (D_t). This progressive addition to neurons is accomplished by adding a new neuron when any improvement to the training metric D_t , is insignificant. Letting D_t denotes the value of the training metrics at iteration t , the following equation shows the process for adding new neuron:

$$\frac{|D_t - D_{t-\varepsilon}|}{D_{t_0}} < \Delta T, (t \geq t_0 + \varepsilon) \quad (3.1)$$

Where t_0 is iteration index at the prior neurons number, ε represents the number of iterations through the error curve searching slope D_t could be computed, and ΔT denotes the slope of the trigger. The optimal final condition as presented in Eq. (3.1) guarantees that at best training iterations ε have been carried out before any further new additional neuron is appended. The stopping criteria for this procedure are achieved when D_t is adequately small or the performance goal of convergence is attained.

The convergence of the neural network (when the number of neurons in the hidden layer is at its optimum) is best assessed using the maximum squared difference (errors) at any time t . Mathematically, the largest squared error is:

$$D_{max} = \max_{\mu,i} [(S_{L_i}^{\mu} - \zeta_i^{\mu})^2], \quad (3.2)$$

When the largest squared error experiences a drastic drop, the optimal number of neurons has been identified. The objective of the training session is to obtain an output response $S_{L_i}^{\mu}$, $i = 1, \dots, NL$, that is ideally the same as the desired response ζ_i^{μ} , $i = 1, \dots, NL$, where NL is the number of neurons required to define the response.

3.3.4 Performance Criteria

For a neural network, to produce accurate result, the selection of hidden layers and its neurons and number of inputs are essential. Selection of most accurate model, from the pool of models developed having different set of hidden layers and number of inputs, were to be done based on certain performance criteria. Analysis of the results of developed models was based on the regression values (equation 3.3) of training, validation and testing. However, complete accuracy of the model cannot be decided based on the regression values alone (Sousa et al., 2007). The regression value of the model provides the statistical measure of the data fitting to the best fit line but cannot indicate the deviation of the predicted data from the observed data. In order to include the deviation factor in the analysis of the results, some error variables need to be included in the analysis. Hence, mean absolute error (MAE) (equation 3.4), mean square error (MSE) (equation 3.5), plot of the observed and the predicted values, plot of relative error

percentage values (equation 3.6), Nash-Sutcliffe Efficiency (NSE) (equation 3.7) and plot of models on Taylor diagram were also considered in the process of optimum model selection. Taylor diagrams were drawn on the basis of the testing standard deviation, testing mean square error and testing correlation. In Taylor diagram, the model that is close to the actual point is the optimum model. The actual point is the observed value of the pollutants (nitrate-nitrogen or ammonia-nitrogen), which has a definite standard deviation, a correlation value of 1 and a mean square error of zero. The closest model to the actual point has the standard deviation near to the observed values and correlation, with the observed values, close to 1 and least mean square error; making the model best fit for predicting the actual values. NSE measures the performance of prediction models on the scale of $-\infty$ to 1. The model having its NSE value approaching to 1 is considered as the best fit model for prediction. Equations for the performance criteria are given hereafter:

- Regression Values:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (3.3)$$

- Mean Absolute Error:

$$MAE = \frac{1}{n} \sum_{i=1}^n |x - y| \quad (3.4)$$

- Mean Square Error:

$$MSE = \frac{1}{n} \sum_{i=1}^n (x - y)^2 \quad (3.5)$$

- Relative Error Percentage:

$$RE = \frac{|x - y|}{x} * 100 \quad (3.6)$$

- NSE

$$NSE = 1 - \frac{\sum(y - x)^2}{\sum(x - \bar{x})^2} \quad (3.7)$$

Where, in this study, n = number of data points, x = observed data points, and y = predicted data points

3.4 Results

Training of GRNN, multilayer and RBFNN models with different set of internal parameters and input variables resulted in tens of thousands of networks, each with different combinations of internal parameters and having results with varying accuracy. These models were analyzed based on the performance criteria, as discussed in above section, to bring out the optimum models. These set of performance criteria were used in a particular order. Initially, the regression values were used to filter out thousands of low regression valued models, followed by examining high regression valued models on other analysis parameters to sort out the optimum ones. Low regression valued models were filtered out because the models having low regression values clearly indicate that such models have failed to generalize the relation between the input and target vectors, delivering poor accuracy results. And, hence, those models were not required to be further analyzed on other performance criteria. The main aim of the analysis of the model results was to bring out four optimum neural network models for nitrate-nitrogen and ammonia-nitrogen each for both the stations (i.e. station Lui and station Kajang). Figure 3.3 represents the flow chart for the selection procedure of the optimum model for nitrate-nitrogen at station Lui. Three different types of ANN models (i.e. GRNN, Multilayer and RBFNN) used in this research were trained separately with different set of internal parameters and input variables. For each type of ANN models, thousands of models were filtered out to get four models for each set of input variables. Among those four models one optimum model was selected for further comparison with other types of ANN models.

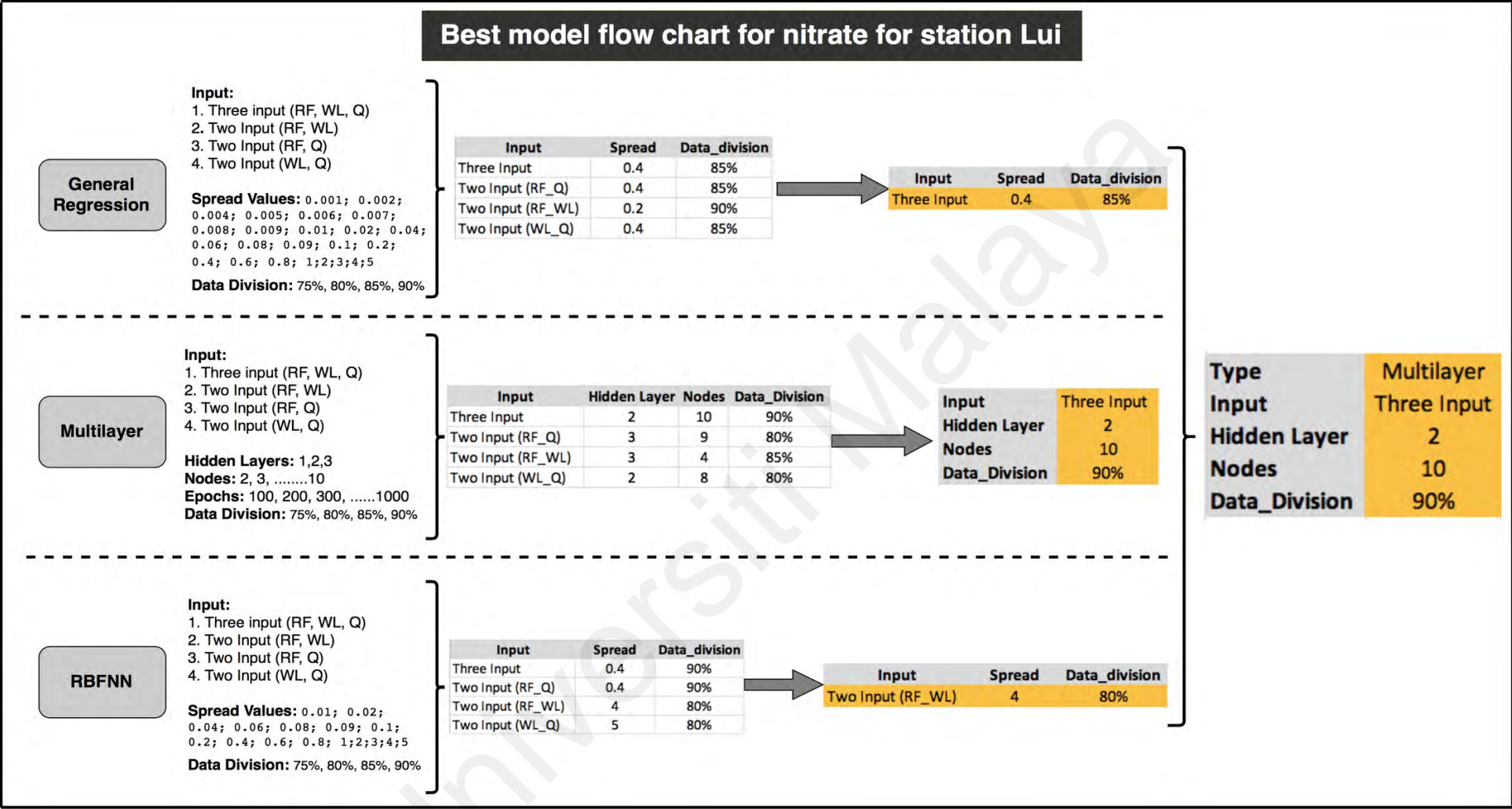


Figure 3.3 Flow chart for the model selection for Nitrate-Nitrogen at Lui station

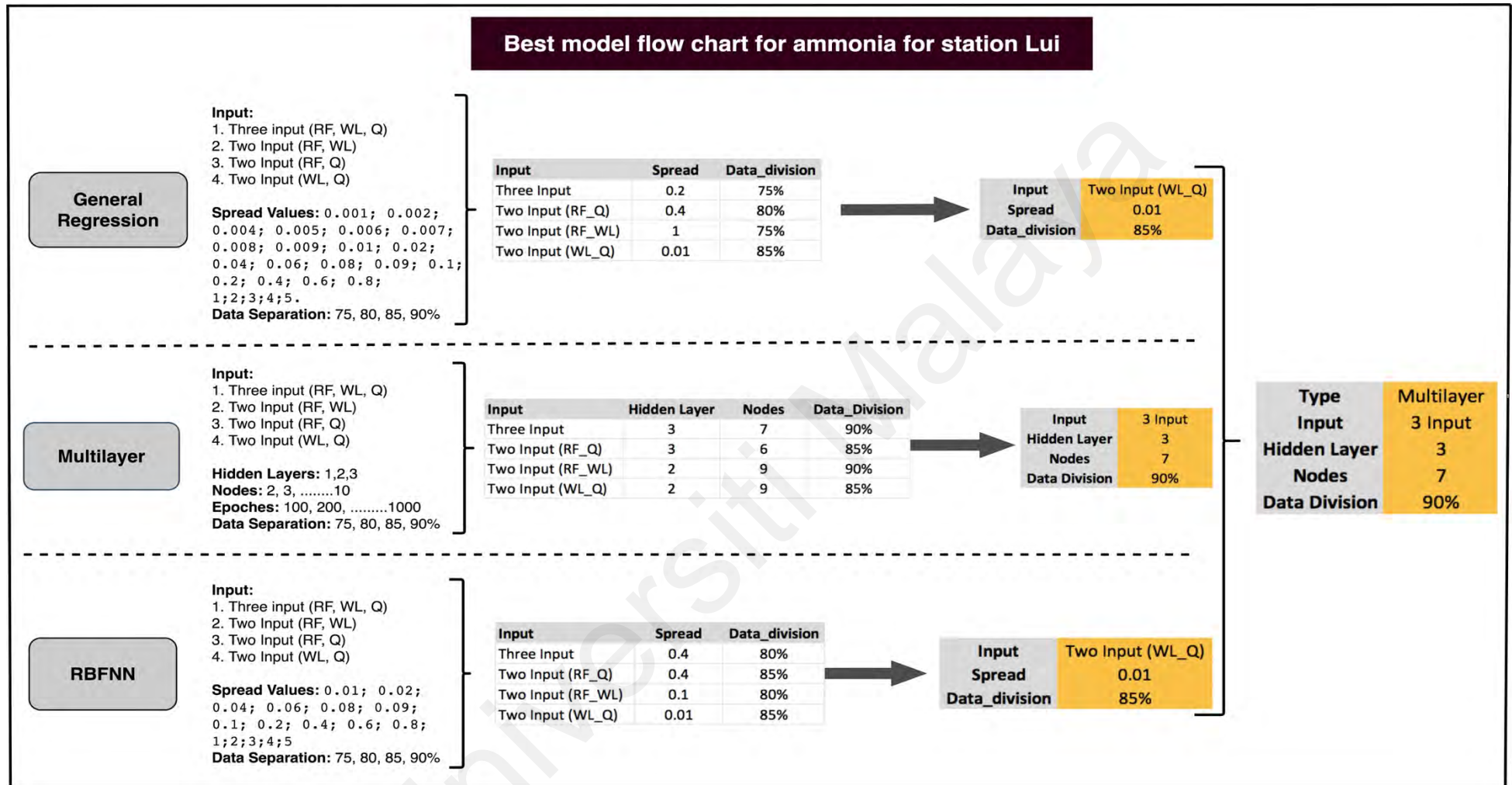


Figure 3.4 Flow chart for the model selection for Ammonia-Nitrogen at Lui station

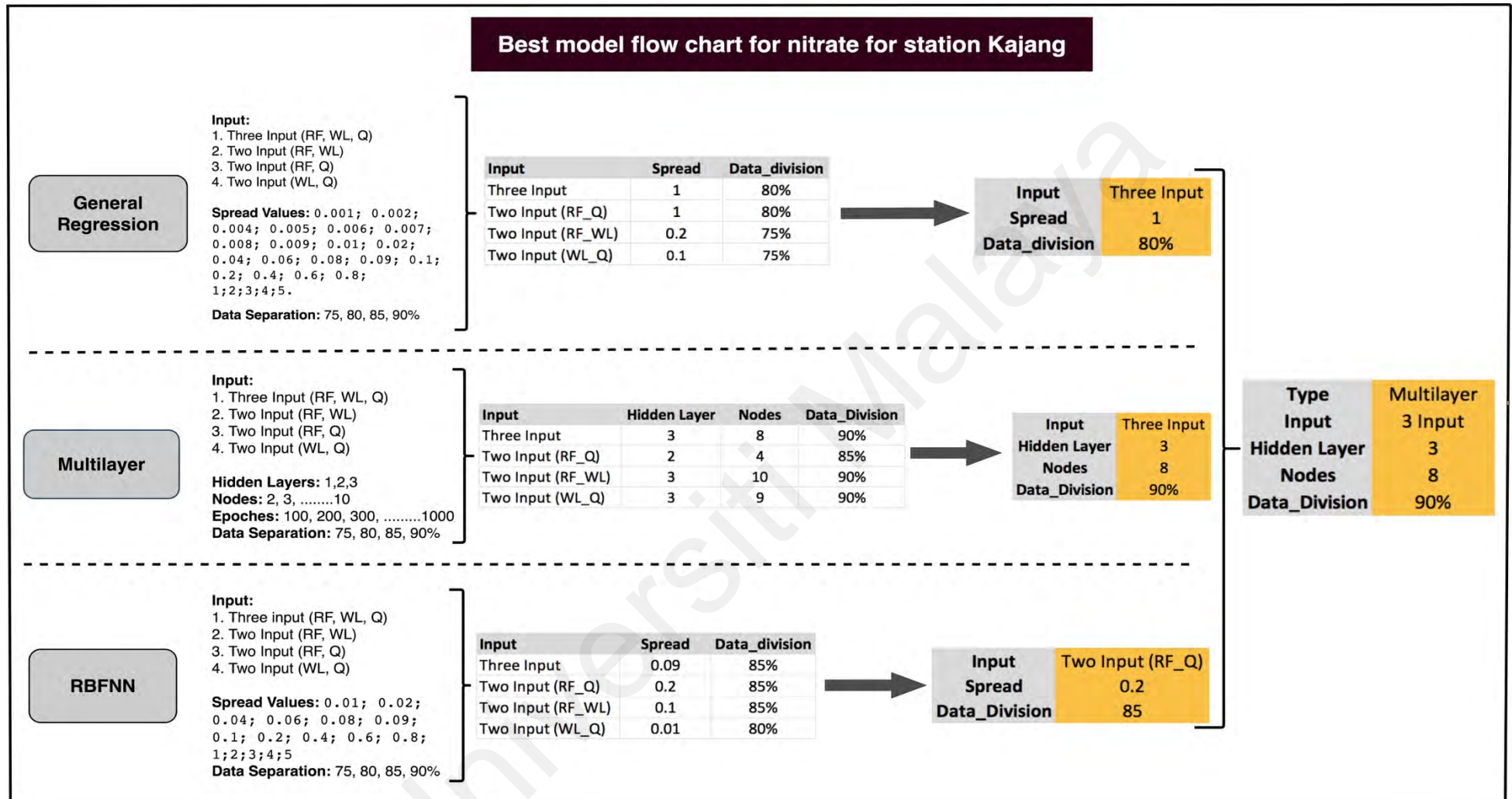


Figure 3.5 Flow chart for the model selection for Nitrate-Nitrogen at Kajang station

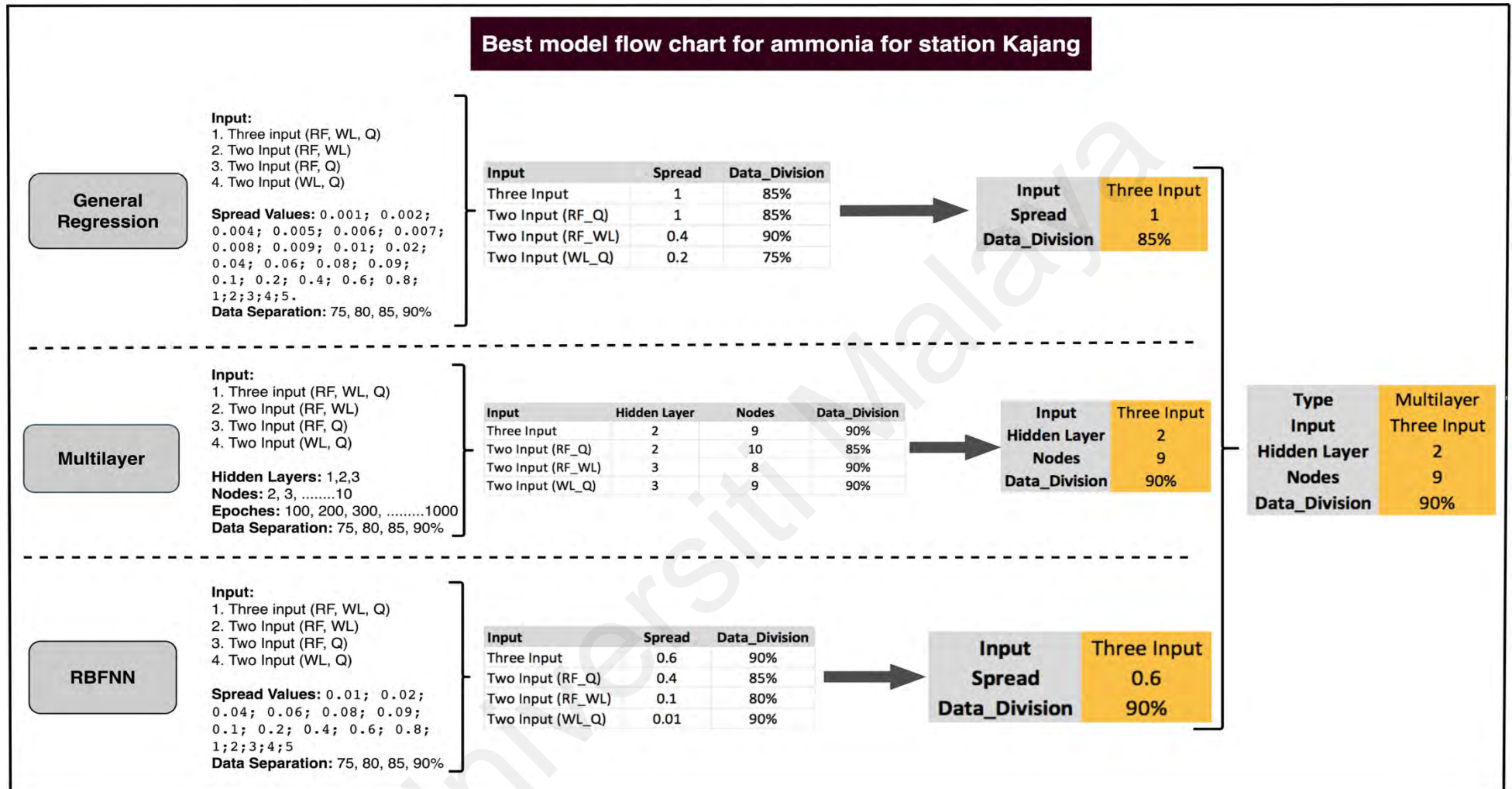


Figure 3.6 Flow chart for the model selection for Ammonia-Nitrogen at Kajang station

Three models obtained, corresponding to three different types of ANN models used, are then compared with each other to get the most accurate model. Hence, the optimum model was selected for prediction of nitrate-nitrogen at station Lui (figure 3.3). Same procedure was followed for the selection of optimum ANN model for ammonia-nitrogen at station Lui, and nitrate-nitrogen and ammonia-nitrogen at station Kajang. Figure 3.4-3.6 represent the optimum model selection procedure for ammonia-nitrogen model at station Lui, nitrate-nitrogen model at station Kajang and ammonia-nitrogen model at station Kajang, respectively. Figure 3.7 represents the Taylor diagram of models for nitrate-nitrogen prediction at station Lui; which clearly displays that the multilayer model with three input and general regression model with input as RF and WL, are close to the actual point. But the relative error percentage plot, and the plot of observed vs predicted values for multilayer model were acceptable over general regression model. Hence, the multilayer model with three inputs is considered to be the optimum model, in comparison with other models, for the prediction of nitrate-nitrogen at station Lui. Figure 3.8 represents the Taylor diagram of models for ammonia nitrogen prediction at station Lui. It clearly displays that the multilayer model with three inputs lies close to the actual point. Also, the relative error percentage plot, and the plot of observed vs predicted values were in acceptable range. Hence, the multilayer model with three input was selected as the optimum model against other models for prediction of ammonia-nitrogen at station Lui. Figure 3.9 represents the Taylor diagram of models for nitrate-nitrogen prediction at station Kajang. It displays that the RBFNN model with inputs as RF and Q lies close to the actual point than the other two models. Comparing the other performance criteria (i.e. relative error percentage plots and plot of observed vs predicted values) of all the three models, it was found that the multilayer model with three inputs outperformed the other two models. Hence, the multilayer model with three inputs was selected as optimum model for prediction of nitrate-nitrogen at station Kajang. Figure 3.10 represents the

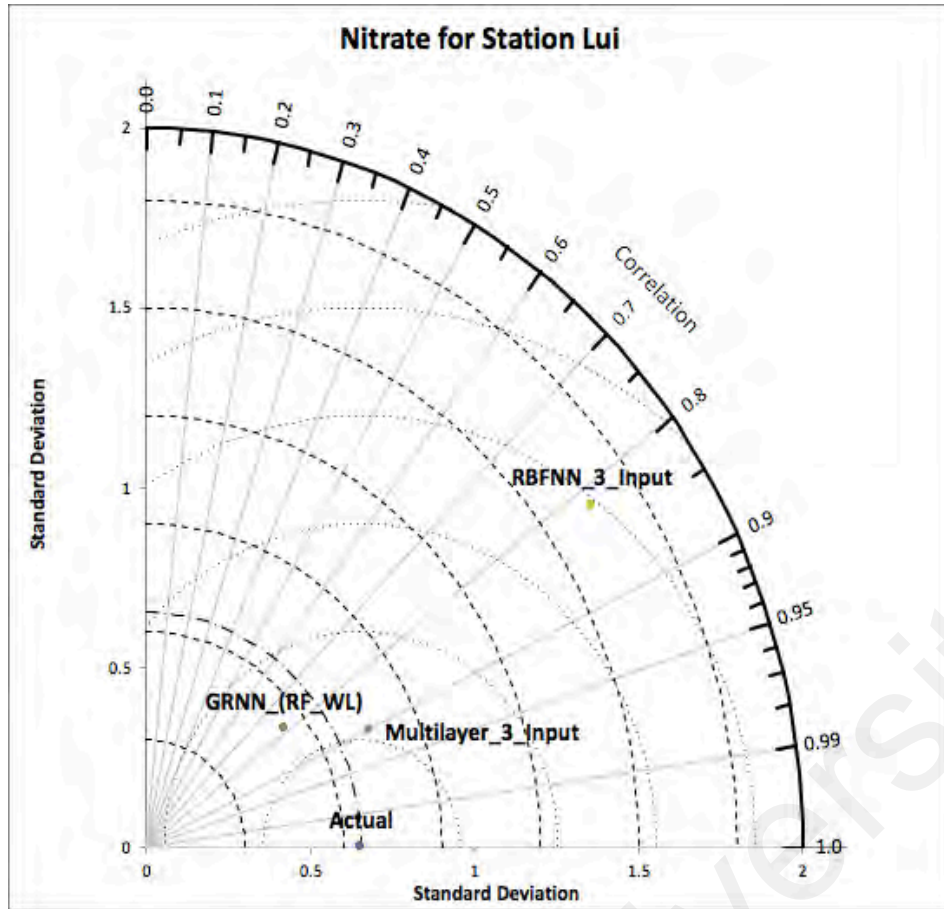


Figure 3.7 Taylor diagram for Nitrate-Nitrogen (Station Lui)

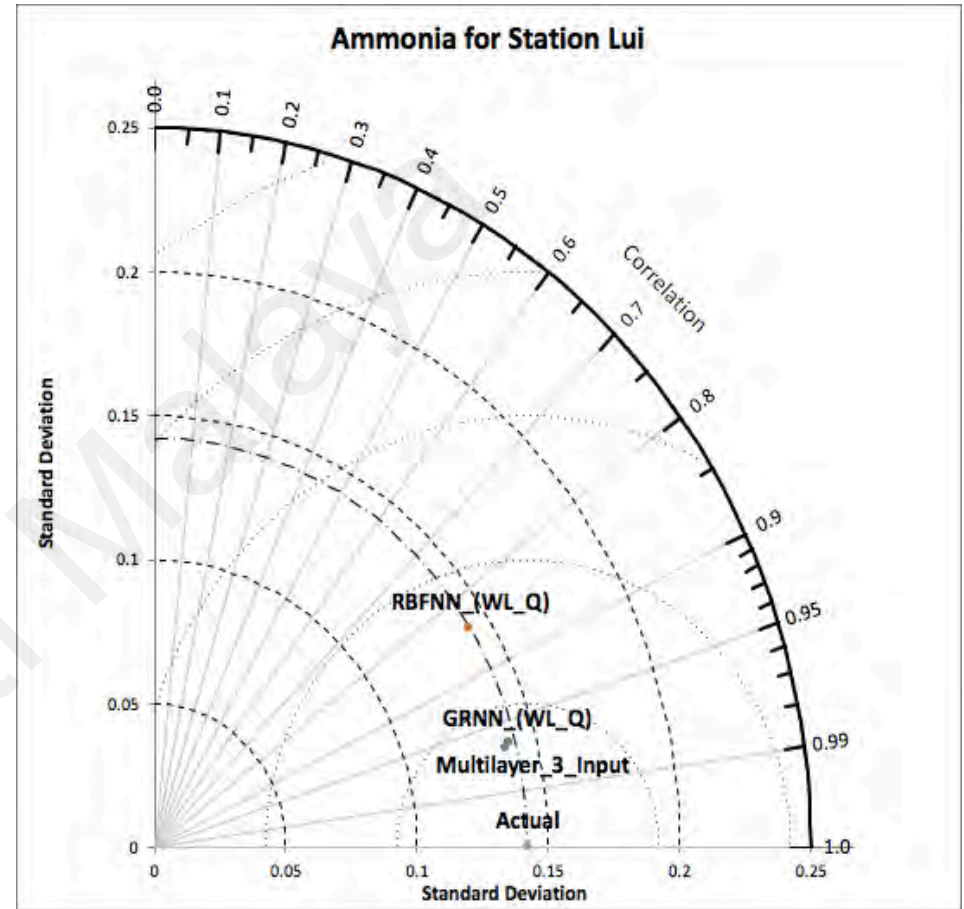


Figure 3.8 Taylor diagram for Ammonia-Nitrogen (Station Lui)

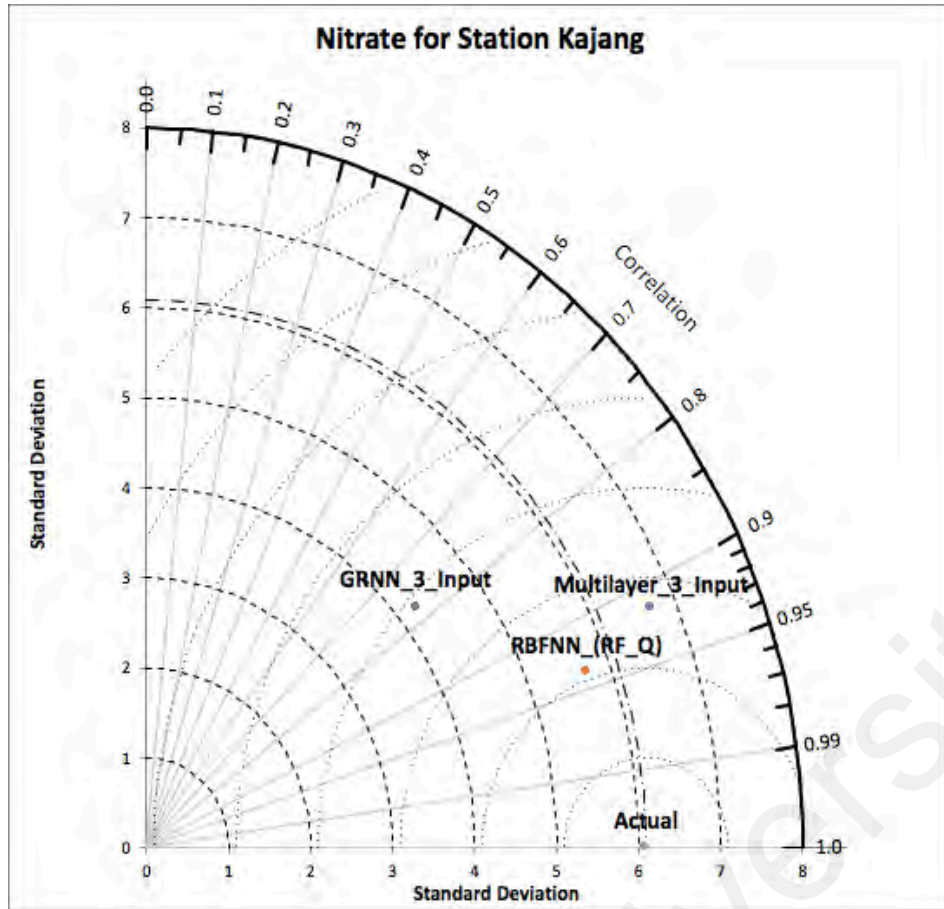


Figure 3.9 Taylor diagram for Nitrate-Nitrogen (Station Kajang)

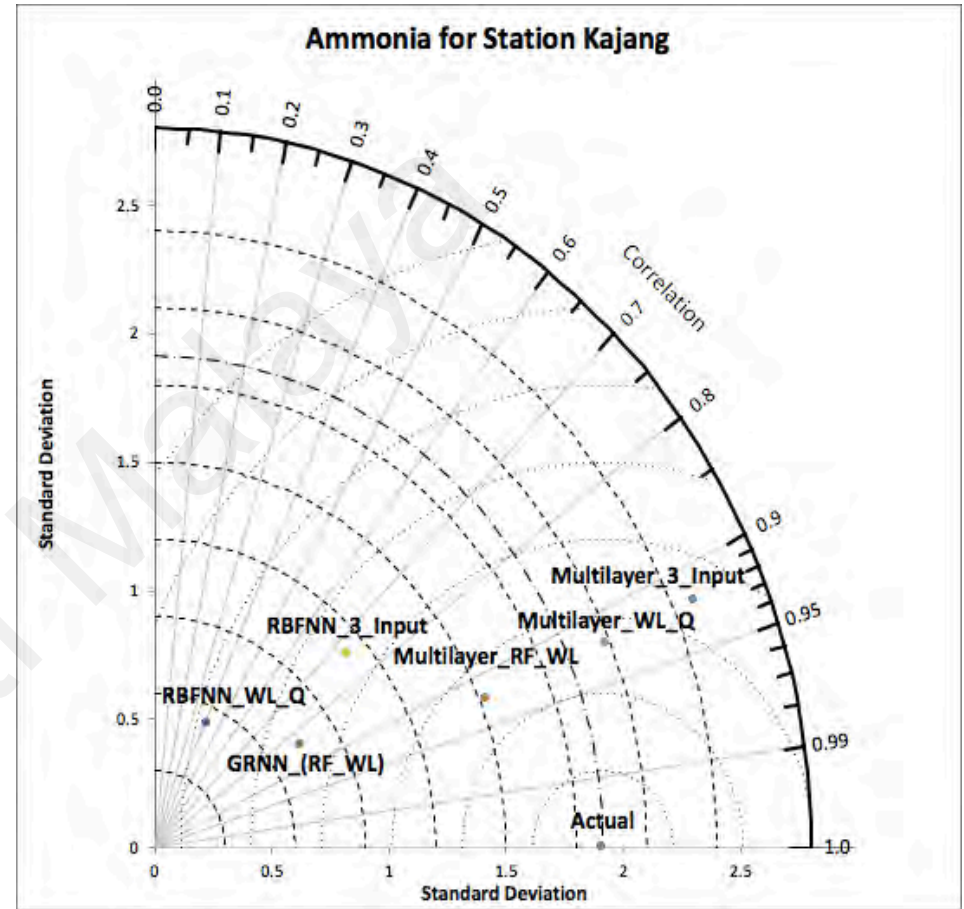


Figure 3.10 Taylor diagram for Ammonia-Nitrogen (Station Kajang)

Taylor diagram of models for ammonia-nitrogen prediction at station Kajang. It represents that multilayer models with three inputs, with input as RF and WL and with input as WL and Q are close to the actual point. On analyzing the relative error percentage plots and plot of observed vs predicted values, it was found that the multilayer model with three inputs has the promising results over other models. Hence, this model was considered as the optimum model, in comparison with others, for the prediction of ammonia-nitrogen at station Kajang.

It is evident that there cannot be one universal model which predicts the desired hydrological parameters for different geographical locations. Model trained on the data of one particular location can predict the desired variable of other locations, but the accuracy may vary, because all locations differ hydrologically, and their historical data have different patterns which the model trained at different location may have not been exposed to. Hence to avoid this variation in accuracy, four different models have been selected, two for each location corresponding to nitrate-nitrogen and ammonia-nitrogen. Table 3.3 represents the configuration and regression values of final selected models for Lui and Kajang stations for nitrate-nitrogen and ammonia-nitrogen. All the selected models are multilayer ANN with overall regression value more than 0.90 and input data division as 90% for training, 5% for validation and 5% for testing. Nash-Sutcliffe Efficiency for all the four optimum models are close to 1; which indicates that models have efficiently predicted the actual values.

Table 3.3 Optimum ANN models for Lui and Kajang stations

Station		Lui		Kajang	
Nitrogen Compound		Nitrate-N	Ammonia-N	Nitrate-N	Ammonia-N
Parameters	Type	Multilayer	Multilayer	Multilayer	Multilayer
	Inputs	Three Input (RF, WL, Q)	Three Input (RF, WL, Q)	Three Input (RF, WL, Q)	Three Input (RF, WL, Q)
	Hidden Layer	2	3	3	2
	No. of Nodes	10	7	8	9

Table 3.3, continued

Station		Lui		Kajang	
Nitrogen Compound		Nitrate-N	Ammonia-N	Nitrate-N	Ammonia-N
	Epochs	300	200	1000	1000
	Training data	90%	90%	90%	90%
Accuracy	Overall Regression	0.98	0.968	0.92	0.98
	Training Regression	0.99	0.988	0.99	0.99
	Validation Regression	0.95	0.704	0.92	0.95
	Testing Regression	0.90	0.65	0.61	0.92
	Mean Absolute Error (mg/l)	0.0978	0.017	0.771	0.173
	Mean Square Error (mg/l)	0.027	0.0013	7.09	0.121
	Nash-Sutcliffe Efficiency	0.9666	0.9457	0.9588	0.9715

Models were tested for different combination of input vectors and internal parameters, as given in table 3.2. Model performance, measured with mean square error, varied with variations in different internal parameter and input vectors. Analyzing the model performance by varying number of inputs, it was observed that model has least mean square error when all the three input vectors were used. Hence, three inputs (RF, WL, Q) were selected for optimum models. One of the comparisons between the four set of input vectors on the basis of mean square error of the model for nitrate-nitrogen at station Lui, is shown in Figure 3.11. Variation of performance of the model on the basis of percentage data division seems to follow a pattern of training a model with more percentage of data will lead to better results. Hence, the model with 90% training data has least mean square error and was used for optimum models. The comparison between the percentage data divisions on the basis of mean square error of the model for nitrate-N at the station Lui, is shown in Figure 3.12. Variation of performance of the models on the basis of number of nodes in hidden layers is presented in Figure 3.13 and the variation of the performance

of the models on the basis of number of hidden layers is shown in Figure 3.14. The concept of increasing the number of hidden layers and number of nodes in the model, as explained earlier, was to increase the complexity of the network which helps the model to learn different patterns in the target data. Beyond a certain number of hidden layer and nodes in it, network becomes over complexed leading to the decrease in the performance of the model. Within the selected range of number of nodes, for this research work, it was observed that the mean square error was decreasing with increase in the nodes. And for the hidden layers, the minimum mean square error was obtained at two hidden layers, beyond which network seems to have become over complexed as the mean square error increased for three hidden layers.

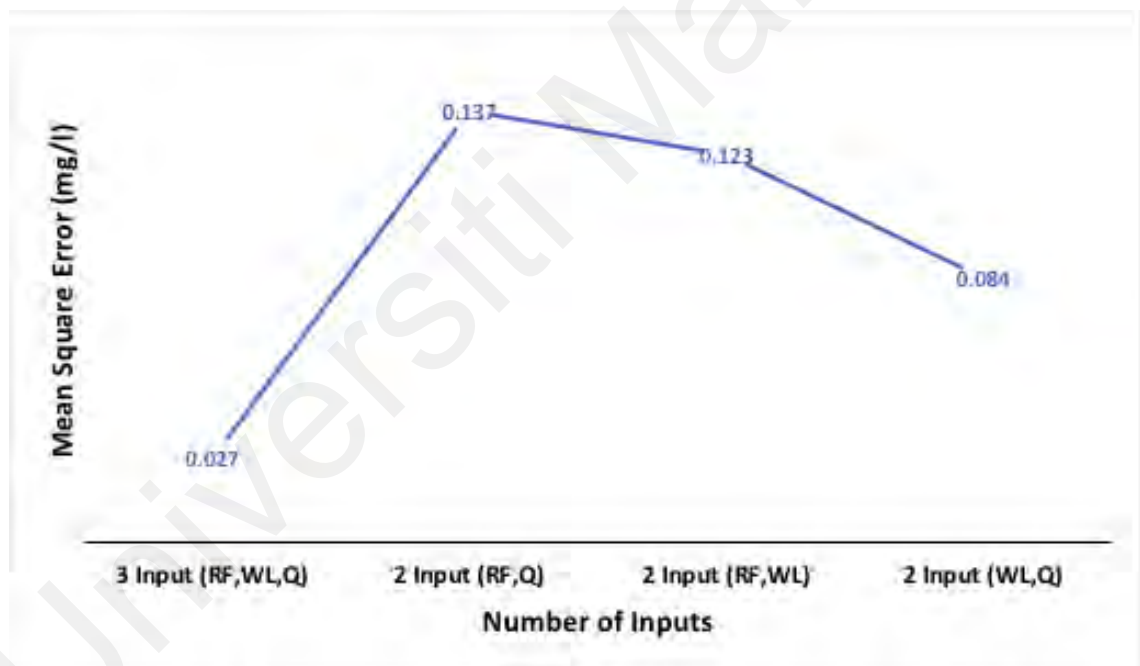


Figure 3.11 Comparison of mean square error for different input vectors for Nitrate-N at station Lui

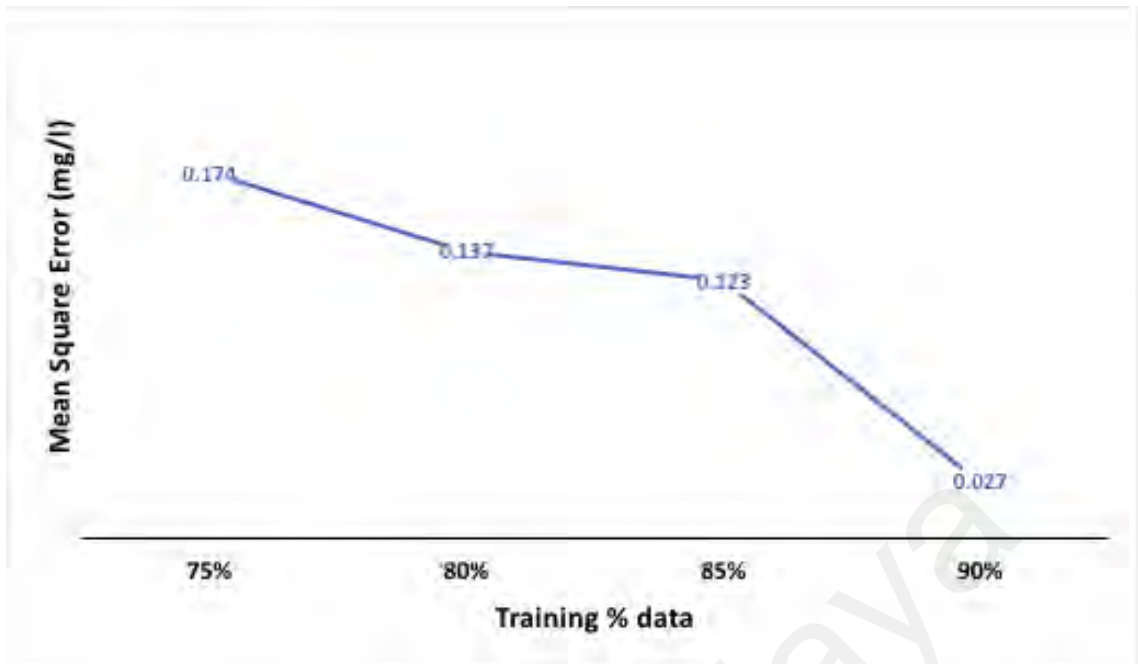


Figure 3.12 Comparison of mean square error for different training data division percentage for Nitrate-N at station Lui

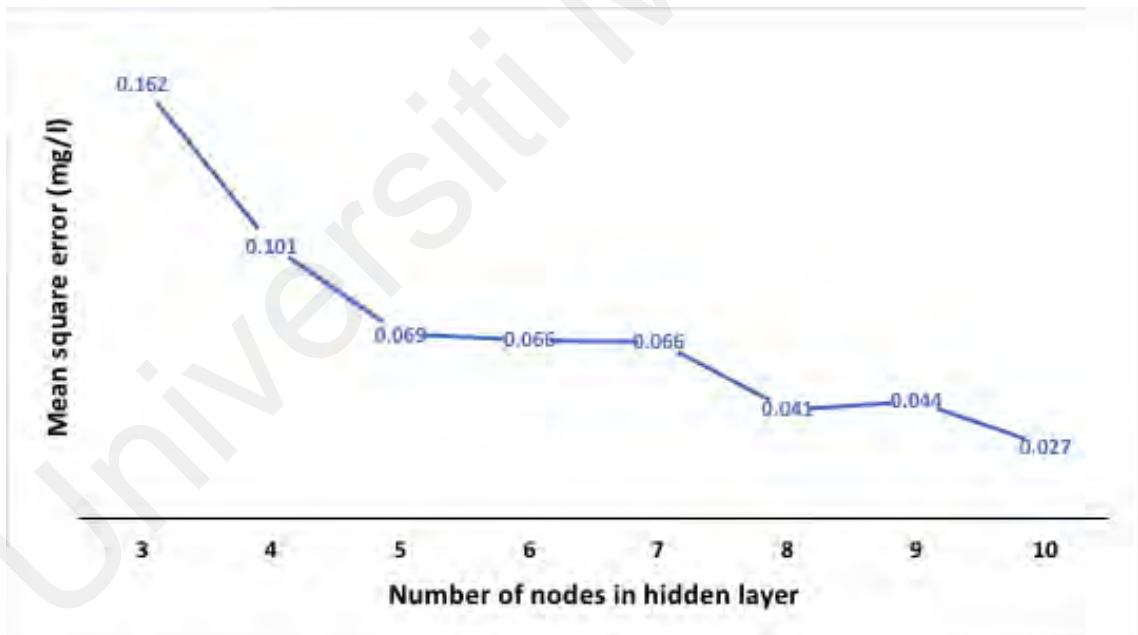


Figure 3.13 Plot of variation of mean square error against number of nodes in hidden layers for Nitrate-N at station Lui

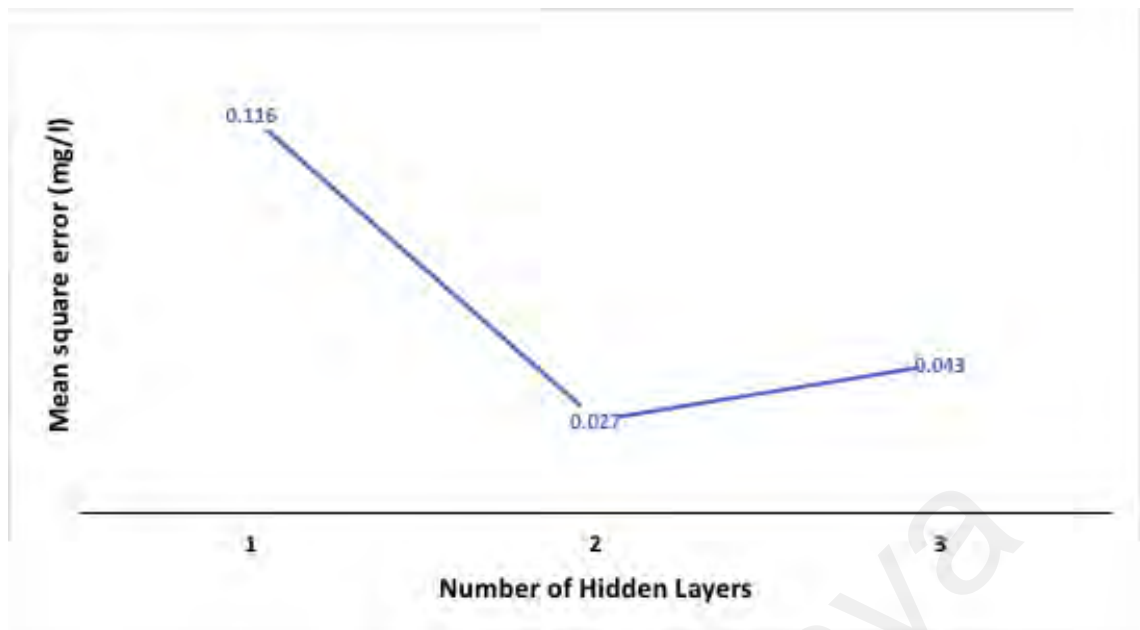


Figure 3. 14 Plot of variation of mean square error against number hidden layers for Nitrate-N at station Lui

Variation of performance of models on the basis of spread values for general regression and RBFNN models are shown in Figure 3.15 and Figure 3.16, respectively. As shown in the Figure 3.15 and Figure 3.16, the testing mean square error for these models are decreasing with increase in the spread values and after a certain point it increases with further increase in spread values, leading to the identification of a spread value having better accuracy and suitable for optimum model. Figure 3.17 shows the plot of the variation of mean square error against the number of epochs. The concept of changing training epochs is to allow the model to train sufficient number of iterations and also to stop before the model begins overtraining. For the model predicting nitrate-N at station Lui, the optimum epochs obtained from Figure 3.17 is 300, as the model delivers least mean square error indicating that model is trained with sufficient number of iterations without being over-trained. The number of epochs beyond which model starts overtraining depends on the complexity of the network.

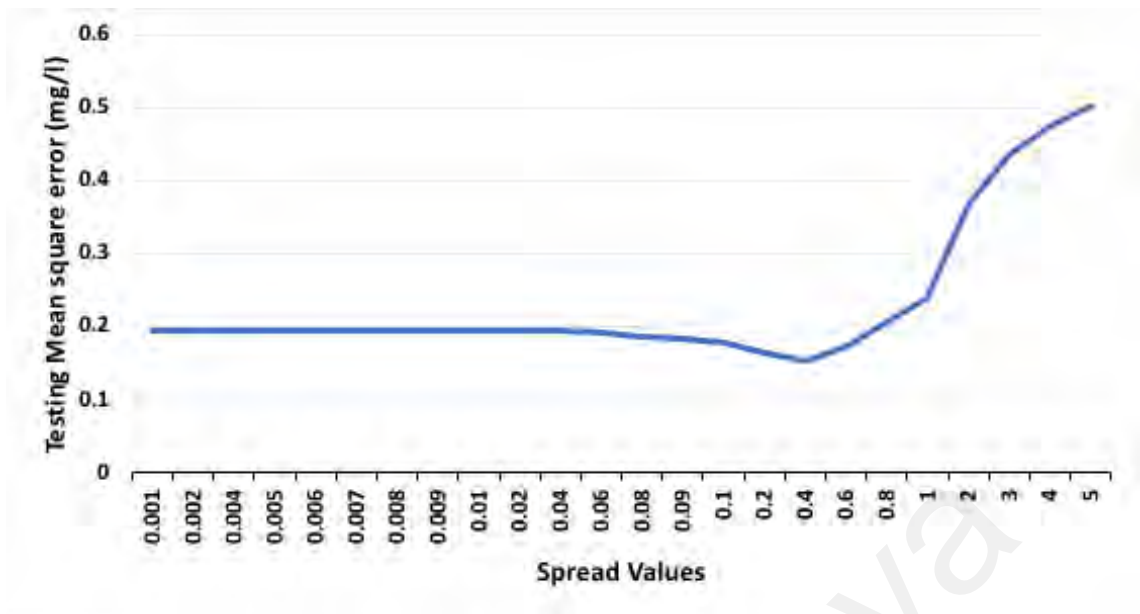


Figure 3.15 Plot of variation of testing mean square error against spread values for general regression model for Nitrate-N at station Lui

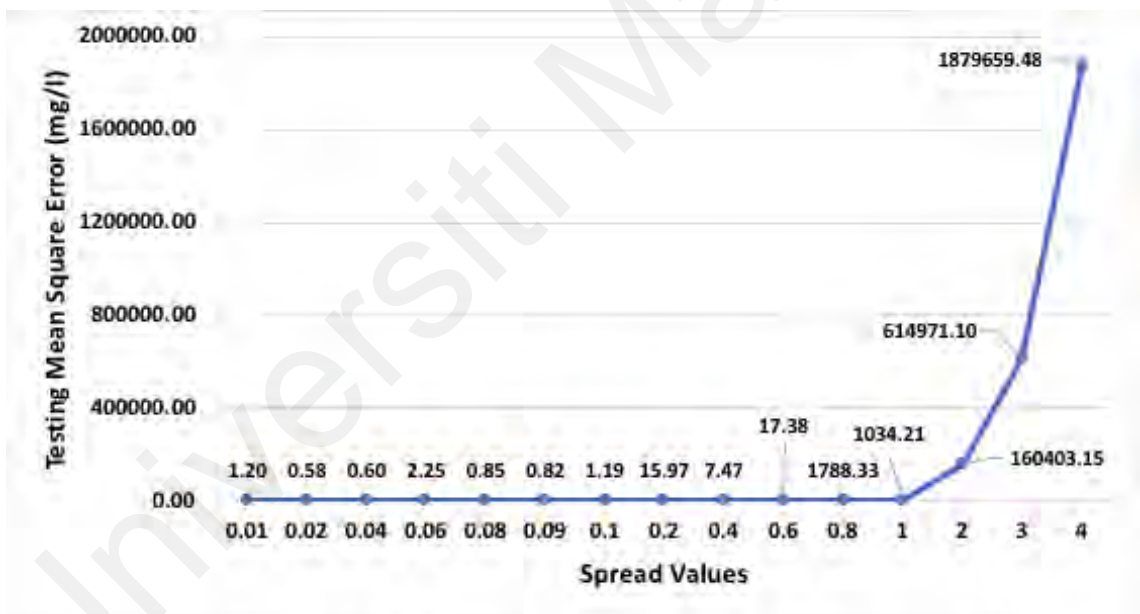


Figure 3.16 Plot of variation of testing mean square error against spread values for RBFNN model for Nitrate-N at station Lui

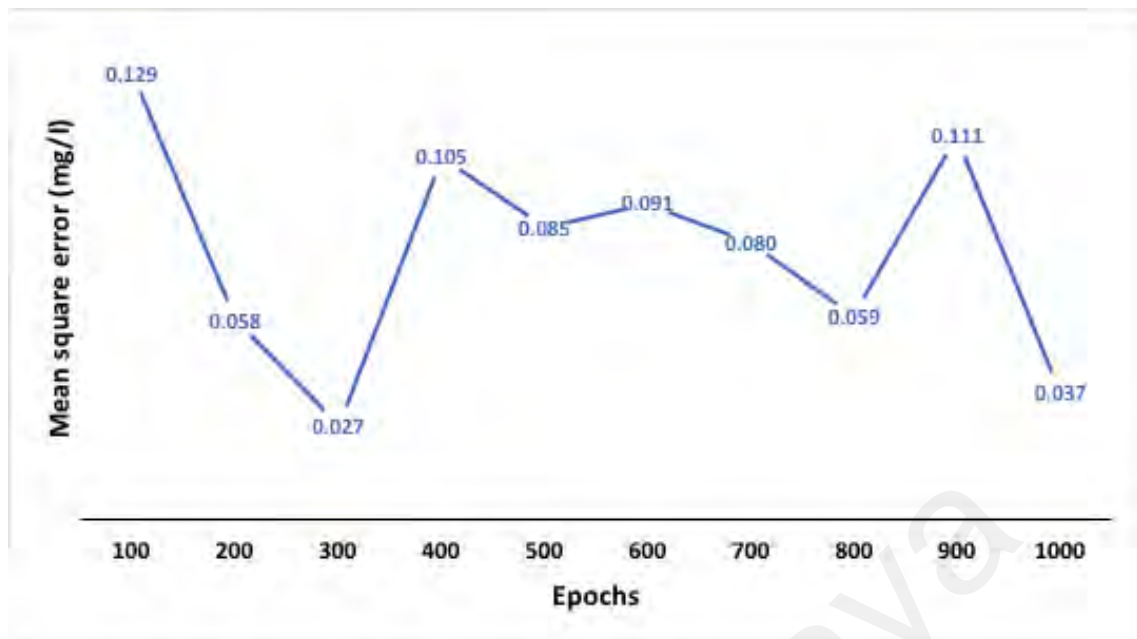


Figure 3.17 Plot of variation of mean square error against epochs for Nitrate-N at station Lui

3.5 Discussion

While filtering out thousands of models, it was observed that some models of GRNN and RBFNN performed well with training regression of more than 0.98 but did not perform satisfactorily in testing when new input data was fed into the model, which the model was not exposed to in the training process. This led to low regression values for testing and high mean square error values. In the selection process, the main focus was on the testing results of the model, which represents the exact ability of the model to predict the actual values. The possible explanation of the low testing regression and high mean square error of those GRNN and RBFNN models is overfitting, which generally lead to high training regression values and low testing regression values.

As shown in Table 3.3, the testing regression values for ammonia-nitrogen for the station Lui and for nitrate-nitrogen for the station Kajang were 0.65 and 0.61, respectively, which are considerably low in comparison with testing regression values for other models. The reason for the low testing regression values lies in the correlation of the input variables

mainly with the output variables. The data obtained for this research work showed good correlation for nitrate-nitrogen for the station Lui and satisfactory correlation for ammonia-nitrogen for the station Kajang but low values for the nitrate-nitrogen for the station Kajang and for ammonia-nitrogen for the station Lui. The correlation for the station Lui for RF, WL and Q with ammonia-nitrogen were 0.57, 0.61 and 0.61 respectively and for the station Kajang for RF, WL and Q with nitrate-nitrogen were 0.69, 0.75 and 0.67 respectively. Corresponding to the low values of the correlation and other unaccountable natural parameters, upon which concentration of these compound depends, the model failed to establish the relation between the input variables and the output variables, leading to low testing regression.

Figure 3.18 represents the percentage relative error of the four optimum models selected for stations Lui and Kajang for nitrate-nitrogen and ammonia-nitrogen. Data points in these figures are arranged chronologically. Relative error figures represent that the model generated more error for the data recorded in earlier days i.e. in 1980s. Some of these errors reached near 100%, but the maximum number of errors were close to zero-percentage line. High error values could be brought close to the zero-percentage line using deep learning methods.



Figure 3.18 Relative error percentage plot for optimum selected model

Figure 3.19 represents the plot of the observed vs predicted values for the optimum selected models. The trend line formed approximately 45° for all the selected models and also, nearly all the points lied near to the trend line. This indicated that the predicted values were very close to the observed values. Hence, making these models optimum for predicting the monthly average nitrate-nitrogen and monthly average ammonia-nitrogen for the Langat River.

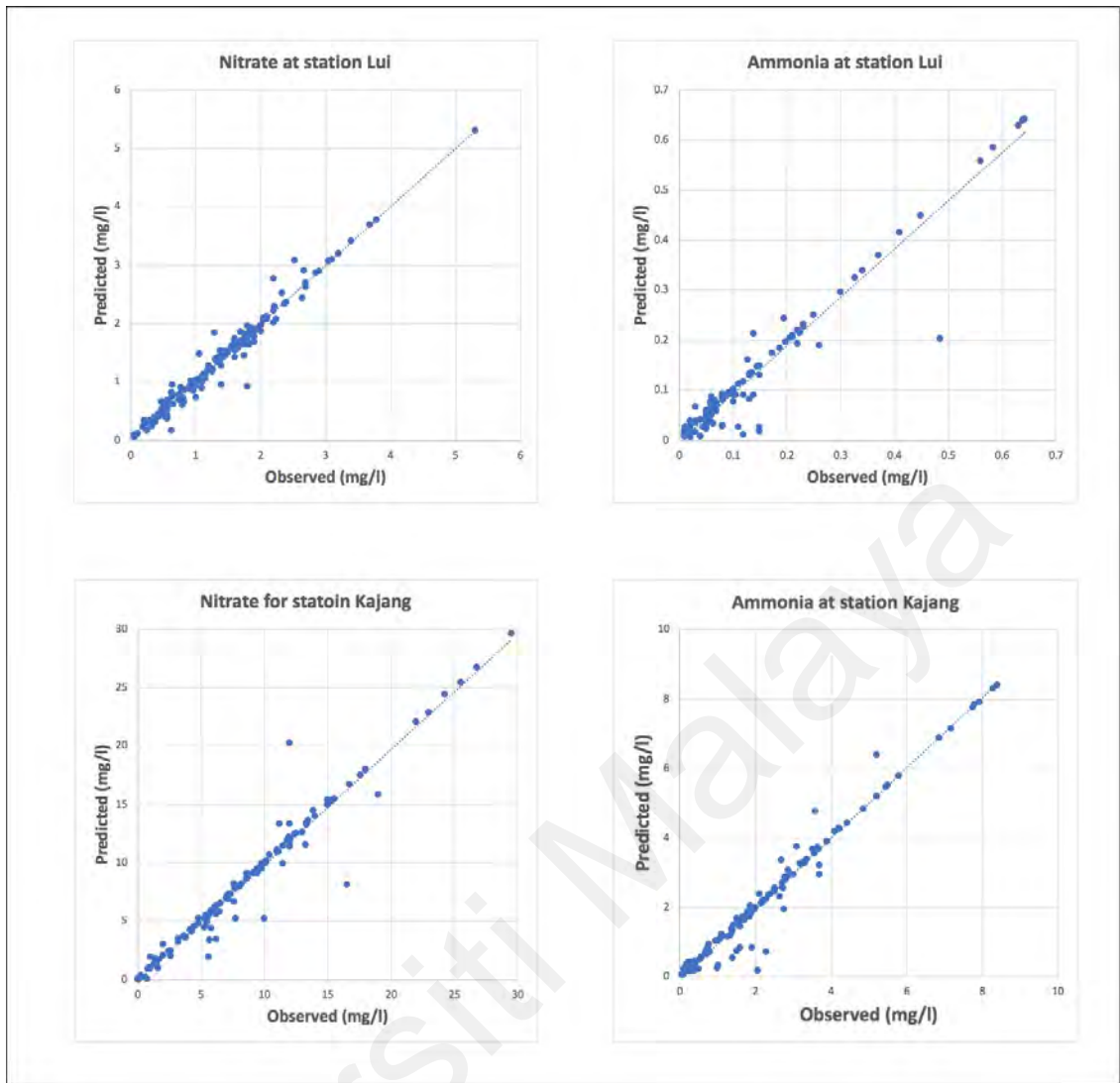


Figure 3.19 Observed versus predicted plot for optimum selected models

According to Chitsazan et al. (2015), the sources of uncertainty in model prediction lies in the uncertainty in model inputs, model structure, weights and biases. However, most important source is the uncertainty in the model inputs. In the current study, model inputs had few time gaps. Some of those minor time gaps were covered with interpolated values, thus introducing some amount of uncertainty in model inputs. Average uncertainty in the prediction can be calculated using the following equation(Zeleňáková et al., 2012):

$$\sigma = \frac{1}{n} \sum_{i=1}^n \left(\frac{|x-y|}{x} \right) * 100 \quad (7)$$

where: σ = average uncertainty percentage, n = number of data points, x = observed data points, and y = predicted data points

Uncertainty increases at every level of calculation or prediction performed using the data already having some amount of uncertainty. Interpolation of the data, used in this study, for obtaining the missing values had introduced some amount of uncertainty in the input data, which may have multiplied in the output values after prediction. To reduce the amount uncertainty in the output values it is advised to try to minimize it from the initial stage of processing the raw data obtained for the study.

Average uncertainty of all the four selected optimum models, calculated by equation (7), are shown in Figure 3.20. Model predicting nitrate-nitrogen for both the stations, Lui and Kajang, show less uncertainty of 9.5%. Ammonia-nitrogen model at station Lui shows highest uncertainty of 23.9%. These models seem appropriate for nitrate-nitrogen and ammonia-nitrogen prediction at stations Lui and Kajang.

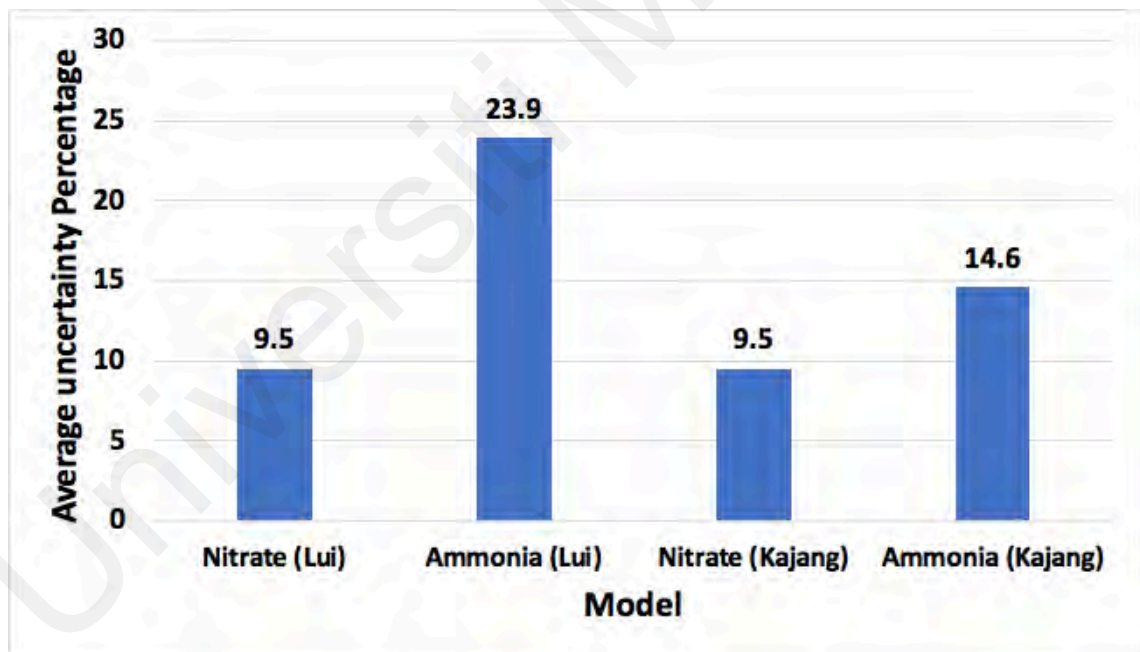


Figure 3.20 Average uncertainty of different models

Table 3.4 Comparison of current study results with literature

	Current Study				Literature			
Station/ Author	Lui	Lui	Kajang	Kajang	Ancil et al. (2009)	Suen and Eheart (2003)	Sharma et al. (2003)	Markus et al. (2010)
Prediction Variable	Monthly Average Nitrate-N	Monthly Average Ammonia-N	Monthly Average Nitrate-N	Monthly Average Ammonia-N	Nitrate-nitrogen flux	Nitrate concentration	Nitrate concentration	Weekly nitrate-nitrogen
Accuracy	<p>Regression:</p> <ul style="list-style-type: none"> • Overall: 0.98 • Training: 0.99 • Validation: 0.95 • Testing: 0.90 <p>MAE: 0.0978</p> <p>MSE: 0.027</p> <p>Nash-Sutcliffe Efficiency: 0.9666</p>	<p>Regression:</p> <ul style="list-style-type: none"> • Overall: 0.968 • Training: 0.988 • Validation: 0.704 • Testing: 0.65 <p>MAE: 0.017</p> <p>MSE: 0.0013</p> <p>Nash-Sutcliffe Efficiency: 0.9457</p>	<p>Regression:</p> <ul style="list-style-type: none"> • Overall: 0.92 • Training: 0.99 • Validation: 0.92 • Testing: 0.61 <p>MAE: 0.771</p> <p>MSE: 7.09</p> <p>Nash-Sutcliffe Efficiency: 0.9588</p>	<p>Regression:</p> <ul style="list-style-type: none"> • Overall: 0.98 • Training: 0.99 • Validation: 0.95 • Testing: 0.92 <p>MAE: 0.173</p> <p>MSE: 0.121</p> <p>Nash-Sutcliffe Efficiency: 0.9715</p>	<p>Efficiency index = 0.888</p>	<p>Overall accuracy:</p> <ul style="list-style-type: none"> • Method one: BPNN = 0.784 • RBFNN = 0.752 • Method two: BPNN = 0.832 • RBFNN = 0.832 • Boolean output (Method two) BPNN = 0.866 • RBFNN = 0.893 	<p>Correlation coefficient</p> <ul style="list-style-type: none"> • RBFNN Tillage = 0.8079 • No tillage = 0.6911 • BPNN Tillage = 0.8017 • No Tillage = 0.6635 	<p>RMSE for ANN:</p> <ul style="list-style-type: none"> • Training = 0.787 mg/l • Testing = 0.935 mg/l <p>RMSE for evolutionary polynomial regression (EPR):</p> <ul style="list-style-type: none"> • Training = 0.991 mg/l • Testing = 1.010 mg/l <p>Critical success index for NBM:</p> <ul style="list-style-type: none"> • Training = 0.286 • Testing = 0.188

Selected models provide improved results when compared with the existing models available in literature. Analyzing the accuracy of the nitrate-nitrogen-predicting models (Table 3.4), existing in literature, it can be observed that current research work models provide results with better regression values. Anctil et al. (2009) used stacked multilayer perceptron to model nitrate-nitrogen flux in streams and had the efficiency index of 0.888. Suen and Eheart (2003) implemented back-propagation and radial basis function neural network for predicting nitrate-N concentration in streams. Sharma et al. (2003) predicted nitrate-N concentration in drainage water. M. Markus et al. (2010) predicted weekly nitrate nitrogen, in streams, using evolutionary polynomial regression, Naïve Bayes model and back-propagation neural network.

3.6 Summary

Selection of the appropriate internal parameters for the ANN models along with the relevant input variables are essential to ensure accuracy. This article have discussed the selection procedure of those internal parameters and input variables for the ANN model for predicting the monthly average nitrate-nitrogen and monthly average ammonia-nitrogen levels in the Langat River in Selangor, Malaysia. Also, the discussion about the variation of performance response of the model against the variation of different internal parameters and input variables is also included. Among the three model architectures (i.e. GRNN, multilayer and RBFNN), the multilayer model performed very well for nitrate-nitrogen and ammonia-nitrogen prediction. Among the various sets of internal parameters and inputs, selected models have three input variables (RF, WL, and Q) and the data division for training as 90%, validation as 5% and testing as the remaining 5%. The minimum overall regression of the four selected optimum models is 0.92. Nash-Sutcliffe Efficiency for the selected optimum models are very close to 1. Based on the results and their comparison between different sets of training data divisions, it can be stated that

higher percentage of data for training will eventually lead to better accuracy of the model. Maximum relative error percentage points are close to zero-percentage line, with few data point approaching more than 100%; which can be brought back to the zero-percentage line by using hybrid modeling, as described in chapter 4.

Universiti Malaya

CHAPTER 4. ENHANCEMENT OF NITROGEN PREDICTION ACCURACY

4.1 Introduction

Chapter 4 presents the article of enhancement of nitrogen prediction accuracy. This chapter contains the detailed steps of development of new hybrid model using ant colony optimization and Elman neural network models. A software interface is also discussed in this article, which provides an interface for the users to use the newly developed ACO-ENN hybrid model. The sections included in this chapter are introduction, literature review, methodology, results, discussion and conclusion.

Artificial neural networks have already been deployed for prediction of nitrogen compounds, in various countries. But standalone ANN have several limitations, resulting to the lower accuracy of the prediction results. At some stages, standalone ANN fails to generalize the relation of input and target vectors in complex natural variables. However, the limitation of ANN is resolved using hybrid models. Because the hybrid models contain more than one models integrated with each other which helps in learning the complexities of the natural variables. This article proposes a new ACO-ENN hybrid model developed by integrating Ant Colony Optimization (ACO) with Elman Neural Network (ENN). In the new ACO-ENN hybrid model, ACO is used as a decision-making model and ENN is used as a learning model. ACO decides the best set of weights and biases for ENN model. ENN model uses the same weights and biases, selected by ACO, to learn the patterns of target vectors. The developed new ACO-ENN hybrid model was used to improve the existing results of the nitrate-nitrogen and ammonia-nitrogen prediction models. The prediction accuracy of existing model was improved by the new hybrid model. There was a significant improvement in the mean square error, mean absolute error, the maximum relative percentage error, Nash-Sutcliffe Efficiency, peak

flow criteria and low flow criteria. The new hybrid model had an outstanding performance over the classical one.

4.2 Background

4.2.1 ACO

ACO is one of the nature-inspired algorithms for discrete optimization (Mavrovouniotis & Yang, 2013), proposed by Dorigo et al. (1996). ACO algorithm is based on the foraging behaviour of the real ants in nature (Juniora et al., 2013). Such behaviour helps them to find the shortest path to the food source from their initial place (nest) (Kucukkoc & Zhang, 2013). Ants deposit chemical, called pheromone, on their paths while moving (Shah et al., 2012; Tehrani & Khodayar, 2010), forming a pheromone trail. The following ant smells the deposited pheromone and chooses paths in probability (Kalinli et al., 2010). The path at which intensity of pheromone is more has the higher probability of being selected by the following ants. Ants choose paths randomly when a set of paths are given to them that has equal or no pheromone deposited on paths. Ants choosing the shortest path will return back soon, thus, increasing the intensity of the pheromone on the shortest path. Hence, the possibility of selecting the shortest path among the set of paths increases. Figure 4.1 represents an example of ants selecting the shortest path. At time $T=0$, all the ants are at their nest. Each ant leaving the nest will have three paths (A-B-C-F-H, A-B-D-F-G-H, A-B-D-E-G-H) to the food sources. At time $T=0$, there is no pheromone deposited on the paths, hence, the selection possibility of each path is same, and the ants will select the paths randomly. At time $T=1$, ants have started moving in search of food source, selecting every possible path randomly. At this point, those ants that have chosen the shortest path (A-B-D-F-G-H) will return faster than those ants following other two paths. Hence, by time, intensity of pheromone will increase on the shortest path leading

to more ants choosing the shortest path. At time $T=2$, most of the ants are seen moving on the shortest path. Hence, the shortest path is chosen by the ants.

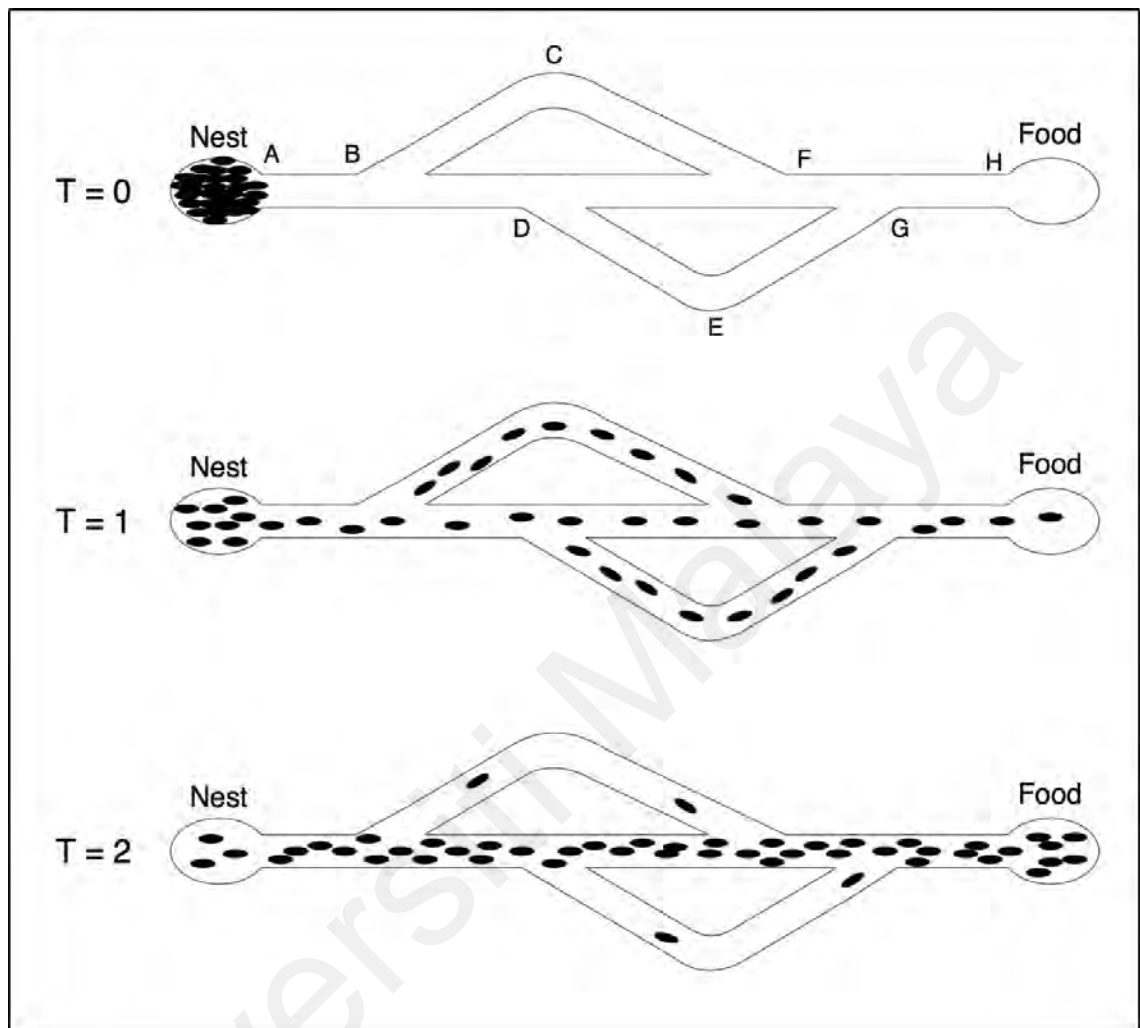


Figure 4.1 Basic concept of ACO

ACO has been continuously used for many discrete optimizations. Most common use of ACO is in Travel Salesman Problem (Hingrajiya et al., 2012), where the minimum cost is to be calculated for the sales person to visit many cities. Many authors have modified ACO to perform certain task, as in the case of ElSaid et al. (2020), they have developed ant based neural topology search based on ACO for optimizing recurrent neural network topology. ACO has been used to determine the connections of recurrent neural network and long short-term memory recurrent neural network for predicting the flight parameters

(ElSaid et al., 2019). Juang and Yeh (2017) used advanced continuous ACO to optimize fully connected recurrent neural network to generate multi-objective gait of biped robot. Sheikhan and Mohammadi (2012) used hybrid model of genetic algorithm and ant colony optimization for selecting efficient features. Zheng-da et al. (2005) used hybrid of ACO and recurrent neural network to overcome the shortcoming of backpropagation algorithm. In this research work, ACO has been used for optimizing the weights of the ENN learning model.

4.2.2 ENN

ENN was developed by Elman in 1990 for resolving the voice processing problems (C. Li et al., 2019). ENN is one of the typical dynamic recurrent neural networks. It is similar to the other ANN models except for the context layer introduced into the structure that receives input from the output of the hidden layers. The stored-values of the context layer is fed back to the same hidden layers (Mahdaviyani et al., 2008), in next iteration along with the model inputs, as shown in figure 4.2. Context layer adds short memory to the network as it stores activations of hidden layers of last iteration, thus, adding one-step delay of hidden layers (Sheela & Deepa, 2013). Context layers intervene between the hidden layer and output layer, unlike the Jordan network, in which context layers receives values from output layers (Song, 2010). In ENN, each hidden layer has their separate context layers, storing the user-defined delays. Inputs of the ENN model are divided into two, the user defined input and the input provided by context layers. ENN has mainly feedforward connections, except for the context layers, and the model is trained using back-proportion algorithm (Tampelini et al., 2011).

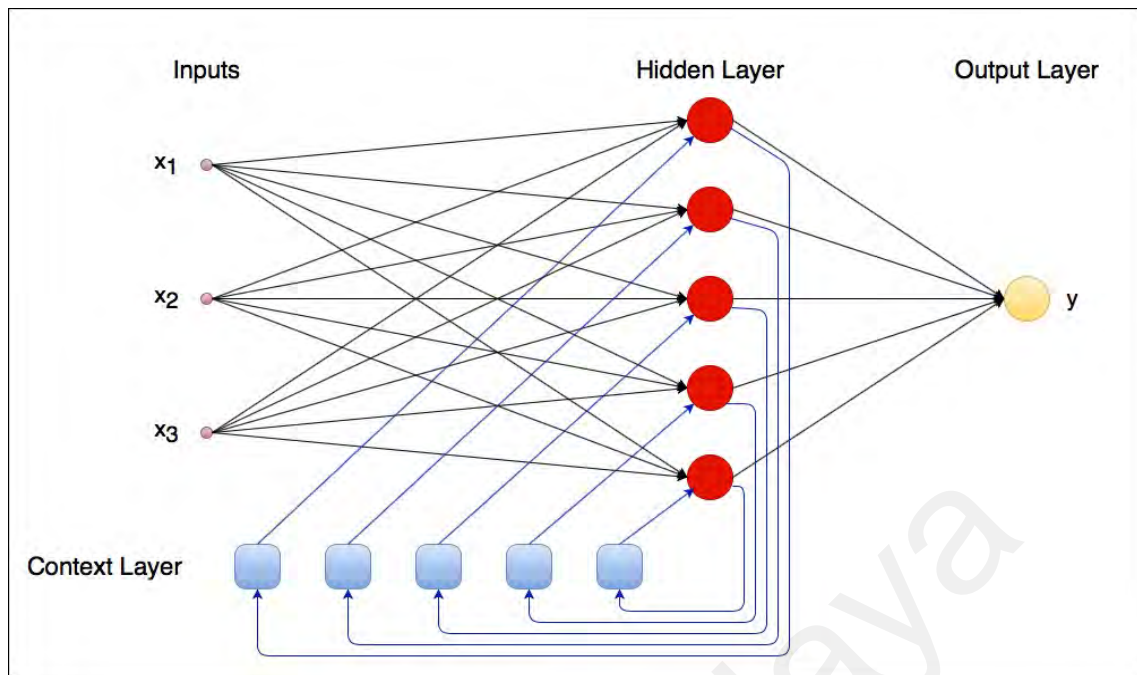


Figure 4.2 Structure of ENN

As shown in the figure 4.2, the ENN model has three user-defined inputs (x_1, x_2, x_3), one hidden layer with five nodes and one output layer with one output (y). The context layer, attached to the hidden layer, receives the activations of the hidden layers and stores it for the next iteration, which is fed along with the input as the memory of the previous time step. This behaviour of the model makes it more suitable, in comparison to other ANN models, for time series prediction. In this research work, ENN model is used along with the ACO to develop a new hybrid model.

4.2.3 Hybrid model

Hybridization is the integration of two or more ANN and other models. Standalone ANN models are proved to be very efficient in some special cases and inefficient in some other cases (Voyant et al., 2012). For instance, some of the ANN models are used for linear time series prediction, which cannot be used for non-linear time series prediction. The concept of the hybridization is to overcome such limitations of the standalone ANN

models. Integrating both the models for linear and non-linear time series prediction will create a hybrid model making it efficient for both the cases.

The main advantage of hybridization is that it helps learning the complex relation of input and target vector, thus, improving the prediction accuracy. The disadvantage of hybridization includes the increased complexity of the network, higher time consumption for training the model and requirement of the modern computational resources for training.

In this research, ACO and ENN have been integrated to develop a new hybrid model. ACO has been used as a decision-making model, which decides the better starting values of the internal parameter of the learning model, i.e. ENN. The developed hybrid model significantly improves the accuracy of the prediction of the nitrate-nitrogen and ammonia-nitrogen in the Langat River in Malaysia.

4.3 Literature Review

Ant Colony Optimization has been used for optimization in various fields since years. Number of authors have used ACO for optimizing different variables by slightly modifying the architecture of ACO. Basis ant's behaviour is to find the shorted path between the nest and the food source. Researches use this behaviour by replacing the shorted path by different features they want to optimize. As discussed above, the most common use of ACO is for the travelling salesman problem. Hingrajiya et al. (2012) used ACO for determining the shortest trip with minimal total cost for visiting all the cities given in the list. Ebrahimpour et al. (2016) used ACO for optimizing the instructions and parameters for multilayer ANN model in a hybrid model for software cost estimation. They used the movement of the ants to select the instructions and parameters for multilayer ANN model. Likewise, Kucukkoc and Zhang (2013) used ant colony optimization for solving assembly line balancing problems. They used ant's behaviour to

optimize the various parameters such as capacity, precedence and other constraints. ACO was used by Tehrani and Khodayar (2010) to optimize the connection weights and threshold values for feed-forward network for prediction of variation of stock price index. They employed the ACO algorithm to improve the learning algorithm and also to reduce the limitations of gradient descent algorithm. J et al. (2018) used ACO in two phases for financial crisis prediction model. In First phase, ACO was used for feature selection and in second phase, it was used for data classification. Their study concluded that the ant colony optimization based financial crisis prediction was most efficient and competitive than other ANN models. Kalinli et al. (2010) compared performance of ANN with ACO for prediction of ultimate bearing capacity of shallow foundation. They concluded that the application of ACO with Meyerhof formula yielded better prediction results in comparison with the prediction results of ANN.

Elman neural network has been successfully used by several researchers for predicting various natural variables. Tampelini et al. (2011) used ENN for prediction and treatment of hydrologic time series. They developed rainfall-runoff model and with the help of their research work they certified the ability of ENN model for producing the reliable predictions of river water based on rainfall data only. Similarly, runoff prediction model was developed by C. Li et al. (2019) using ENN based on normalized mutual information and kernel principal component analysis. In their study, they reduced the parameters of ENN by using less features and concluded that the overfitting problem of ENN model is alleviated by using less features. Coulibaly, Anctil, et al. (2001) investigated three models for multivariate reservoir inflow forecasting, which are: input delayed neural network and a recurrent neural network (mainly Elman neural network) with and without input time delays. Upon investigation they concluded that, broadly, the recurrent neural network was proved to be more efficient in multivariate reservoir inflow prediction.

4.4 Methodology

4.4.1 Data Acquisition

For this article, the hybrid models have been trained on the same dataset as used in the previous article (chapter 3). The dataset consisted of rainfall, water level and discharge as input and nitrate-nitrogen and ammonia-nitrogen as target variables. Monthly average data was obtained from two different stations i.e. Station Lui and Station Kajang in the Langat River basin in Malaysia. Statistical analysis of the obtained dataset has been discussed in chapter 3.

4.4.2 Modification of ACO

General behaviour of ACO is to find the shortest path between the nest of the ants to the food source for the ants. For this research work, ACO is modified to find the optimized hidden layer weights and biases for the ENN on the basis of MSE of the output of ENN. As explained earlier, ants are capable of choosing their paths, among various paths, at a certain node based on some criteria. The criteria for natural ants are the shortest path between the nest and the food source. In various problems, artificial ants are given different criteria. For instance, the criteria for travel salesman problem is the path with minimum cost. This feature of ACO have been used in this research work and the artificial ants have been appointed the criteria of minimum MSE.

4.4.3 Integrating ACO and ENN

The new ACO-ENN Hybrid model is developed on the MATLAB platform. The process of the integrating ACO with ENN involves the different processes explained below.

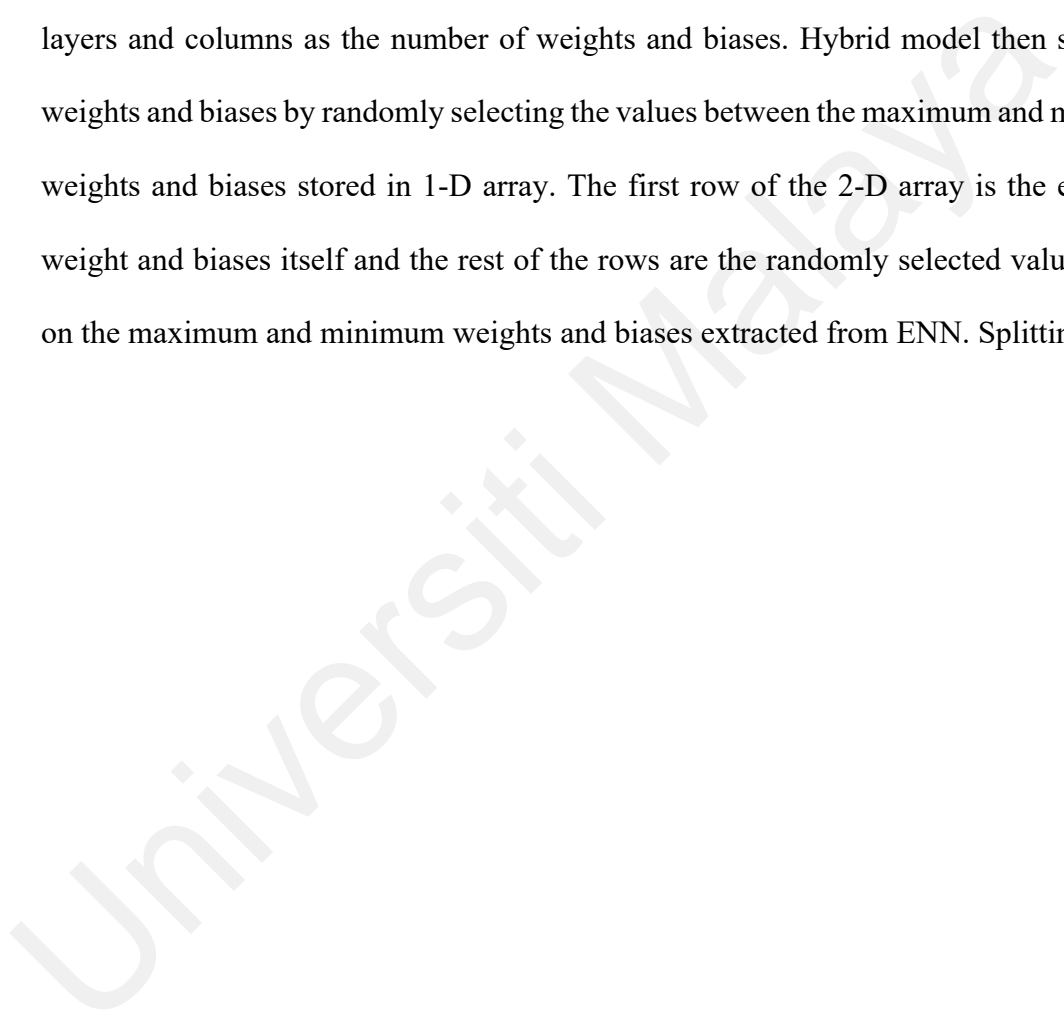
(a) Creation of Elman Network and weights extraction

The process of the hybridization starts with creation of Elman network with hidden layers and their corresponding nodes defined by user. The created Elman network is then configured with the inputs and target variables provided by the user. On configuring, the

network assigns the weights and biases to the nodes of hidden layers. Hybrid model then extracts the weights and biases of the network and stores them into a single 1-D array by appending them one after the other.

(b) Splitting the weights

The extracted weights and biases, stored in a single 1-D array, are then split into user-defined number of layers, to form a 2-D array with rows as the number of weight-split layers and columns as the number of weights and biases. Hybrid model then splits the weights and biases by randomly selecting the values between the maximum and minimum weights and biases stored in 1-D array. The first row of the 2-D array is the extracted weight and biases itself and the rest of the rows are the randomly selected values based on the maximum and minimum weights and biases extracted from ENN. Splitting of the



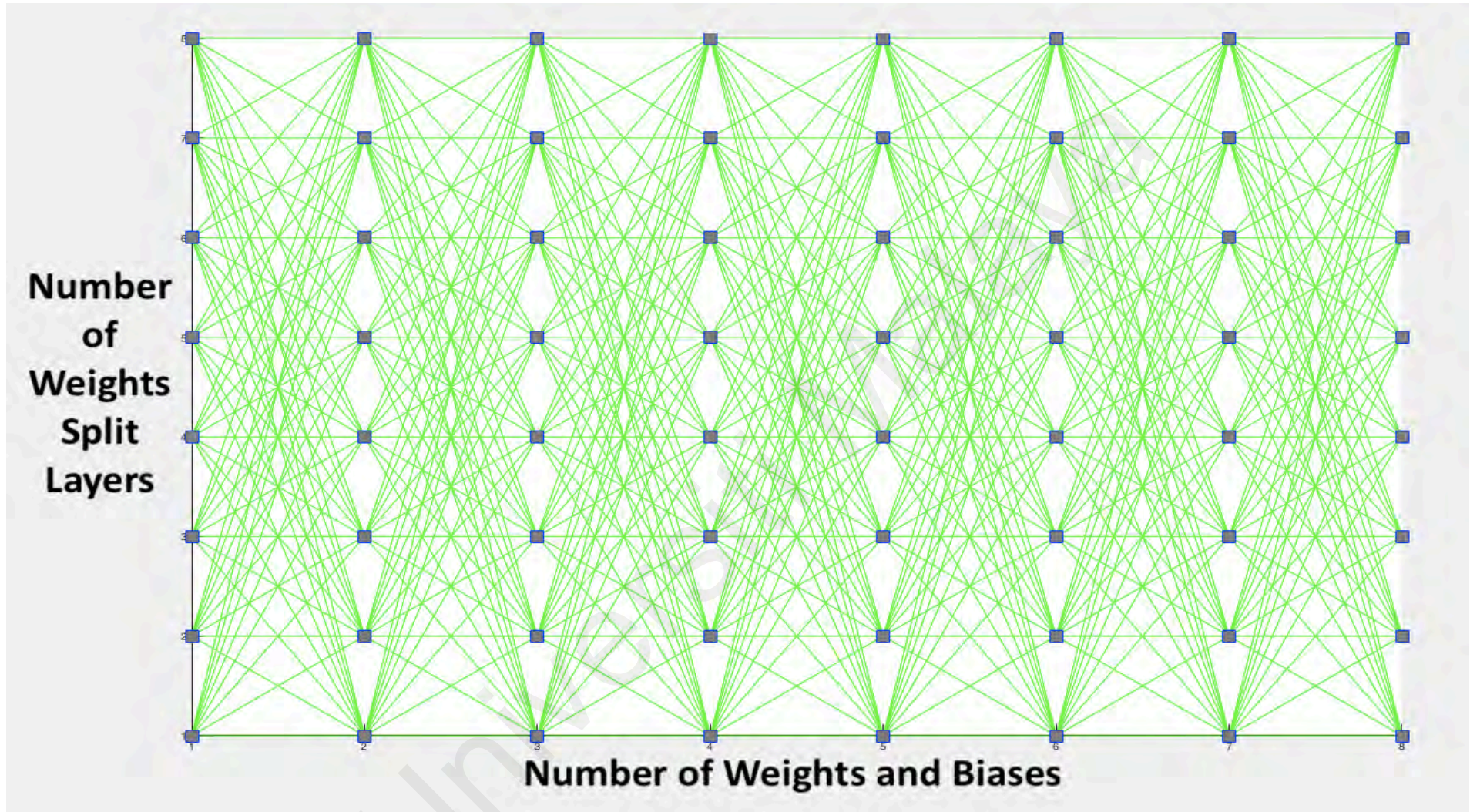


Figure 4.3 Plot for ant movement

weights and biases form a 2-D array which can be converted into the network of paths by making the split weights as nodes and joining them with all the node of ahead layers, as shown in figure 4.3.

(c) Ant Movement

Hybrid model then attaches a nest to the left and a food source to the right of the network of the paths obtained by splitting the weights. Attaching the nest and the food source completes the plot for the ants to start the movement for the search of food source. User-defined number of ants then start the motion from the nest through each node to the food source, depositing the pheromone alongside. Pheromone for each path is stored in a 3-D array with rows as the number of weights split layers, column as the number of weights and biases and the third dimension as number of the weights split layers (representing the connection of the node with each node in the adjacent layers). The initial values of pheromone are assigned as one. The path of each ant is stored in a separate array. Every ant reaching the food source has a definite path, which represents a definite set of weights and biases. These sets of weights and biases for every ant is then sent to ENN, where these weights and biases are restructured, in the same manner as these were extracted, and assigned to the ENN. This network uses the input data to predict the output and calculates the MSE using the target values. Hence, MSE for each ant is obtained. Once all the ants reach the food source through a definite path, each ant get their MSE. Hybrid model then selects the ant with minimum MSE and assigns it as queen ant (i.e. best ant). MSE value of the queen ant is stored for comparison in the upcoming iterations. The path of the queen ant is also stored for displaying it on the graph, as shown in figure 4.4, and for reference after the iterations of ACO ends.

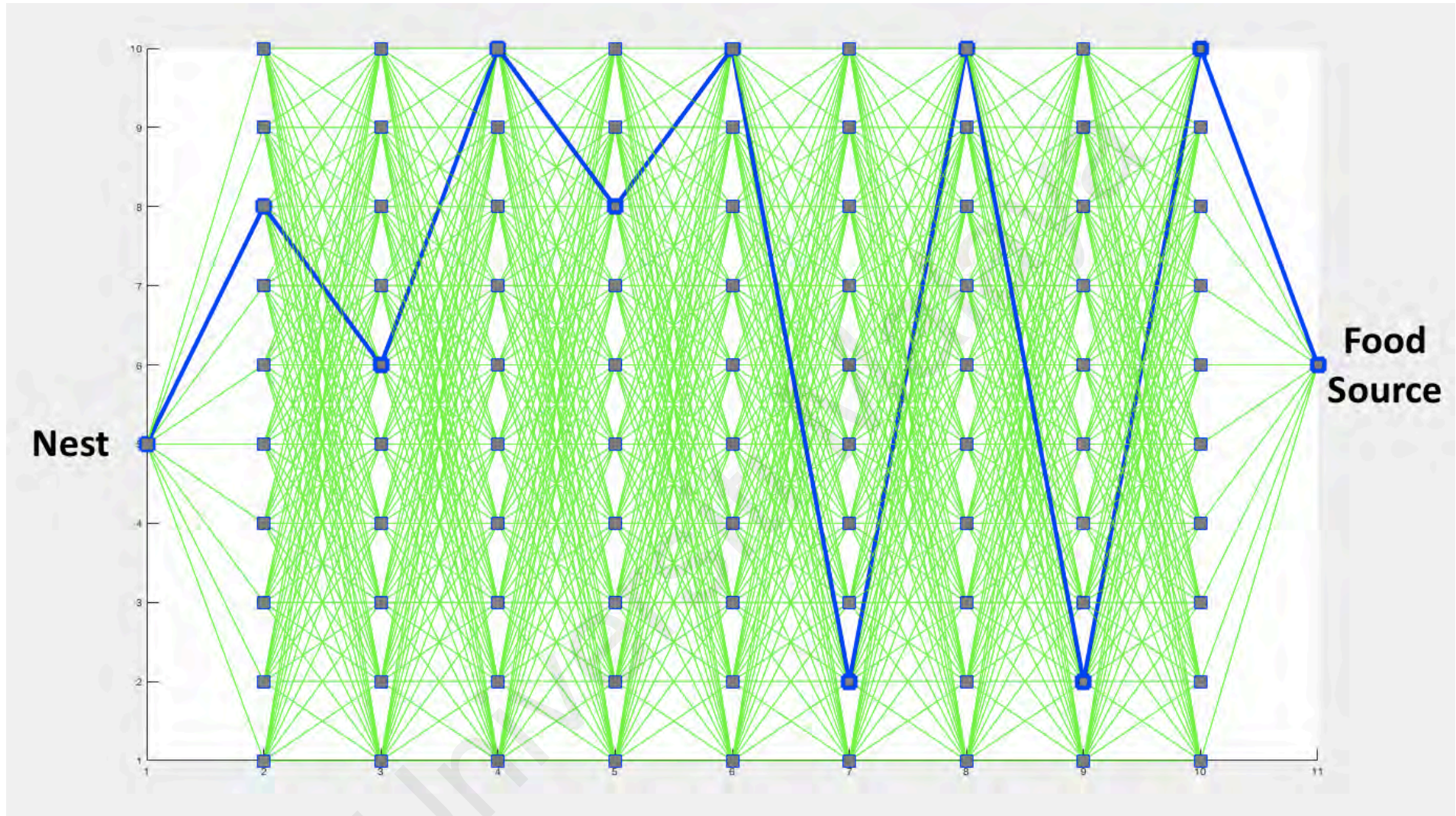


Figure 4.4 Movement of queen ant (Blue path)

(d) Updating Pheromone

Hybrid model then allows only the queen ant to deposit the pheromone while returning to the nest. Rest of the ants are forced to return to the nest without depositing pheromone. The 3-D pheromone array is then updated according to the MSE value of the queen ant (equation. (4.1), as used by Mavrovouniotis and Yang (2013)).

$$\text{Updated Pheromone Value} = \text{Previous Pheromone Value} + \frac{1}{\text{Mean Square Error of Best Ant}} \quad (4.1)$$

The lesser the value of the MSE the greater the value of pheromone addition.

(e) Limiting Pheromone

In some cases where MSE is less than one, the inverse of such value will result in higher value. Hence, a higher pheromone value will be updated (equation 4.1), leading to one particular path having high intensity of pheromone in comparison to other paths. As a result, almost all ant will choose that path because the probability of choosing that particular path will be very high. As a result, all other paths which may result in even lesser value of MSE will be left unexplored, leading to lower accuracy results. Hence, to avoid this limitation of hybrid model, the pheromone values are limited to a certain value according to equation 4.2 (as used by Mavrovouniotis and Yang (2013)).

$$\text{Max Pheromone} = \frac{1}{\text{Evaporation rate} * \text{Mean Square Error of Queen ant}} \quad (4.2)$$

The deposited pheromone keeps on evaporating, after every iteration, with the user-defined evaporation rate. Hybrid model allows only queen ant to deposit the pheromone, leaving the pheromone of other paths (followed by ants or not) undeposited. But the pheromone of all the paths are subjected to evaporation. After few iterations, the pheromone of the path which received no pheromone deposition but only evaporation, will observe major reduction in the intensity of the pheromone. As a result, the intensity of the pheromone of queen ant path will be higher in comparison of the other paths,

leading to the higher probability of the selection of the paths by ants. As a result, all other paths which may lead to even lesser value of MSE will be left unexplored, leading to lower accuracy results. Hence, to overcome this situation, the lower limit of the pheromone is applied according to the equation 4.3 (as used by Mavrovouniotis and Yang (2013)).

$$\text{Min Pheromone} = \frac{\text{Max Pheromone}}{2 * \text{no. of weights and biases}} \quad (4.3)$$

Minimum pheromone is less than half of the maximum pheromone to provide limited but sufficient gap between minimum and maximum pheromone. Minimum pheromone depends on the number of weights and biases which in turn depends on the size of the Elman network.

(f) ENN Training

At the end of the ACO iterations, hybrid model collects the final queen ant path from the ACO part of the model. Based on the queen ant path, the corresponding weights and biases are obtained from the 2-D weight split array. Hybrid model then re-structures these weights and biases and assigns it in the Elman network. Hybrid model then allows the Elman network to start the training using the weights and biases received from ACO and inputs and target values received from the user. Beginning the training of the Elman neural network from the lower MSE, obtained from ACO, will provide the Elman network a better start, which will lead the network to enhanced results. The trained network is then provided as the output of the hybrid model.

Figure 4.5 represents the flow chat of overall process of hybrid structure development. Flow chart briefly describes each step, starting from the initialization of Elman Neural Network and all the way through ACO, to finally training the ENN with the optimized weights and biases. After training the ENN, a hybrid model is developed

and can be used for further analysis. Algorithm used to develop the hybrid structure is presented below. Steps presented in the algorithm is similar to the steps in the flow chart in figure 4.5.

- *Initialise the Elman Neural Network*
- *Configure the Elman Neural Network*
- *Extract the weights and biases of the configured ENN*
- *Reshape the weights and biases to 1 – D array*
- *Split the weights and biases to form a 2 – D array*
- *Attach the nest at the beginning and a food source at the end of the 2 – D array*
- *Start the ACO iterations*
 - *Start the movement of the ants from the nest*
 - *Ants reaching the food source through any path has a set of weights and bises*
 - *Send this set to ENN*
 - *Reshape and assign the weight and biases to ENN*
 - *Obtain the MSE value for each ant*
 - *Every ant return to their nest with their MSE*
 - *Update pheromone values (Only Queen ant is allowed to deposit the pheromone)*
 - *Limit the pheromone within maximum and minimum values*
- *Weights and biases of the queen ant is sent to ENN*
- *Weights ands biases are reshaped and assigned to ENN*
- *Train the ENN with new weights and biases*

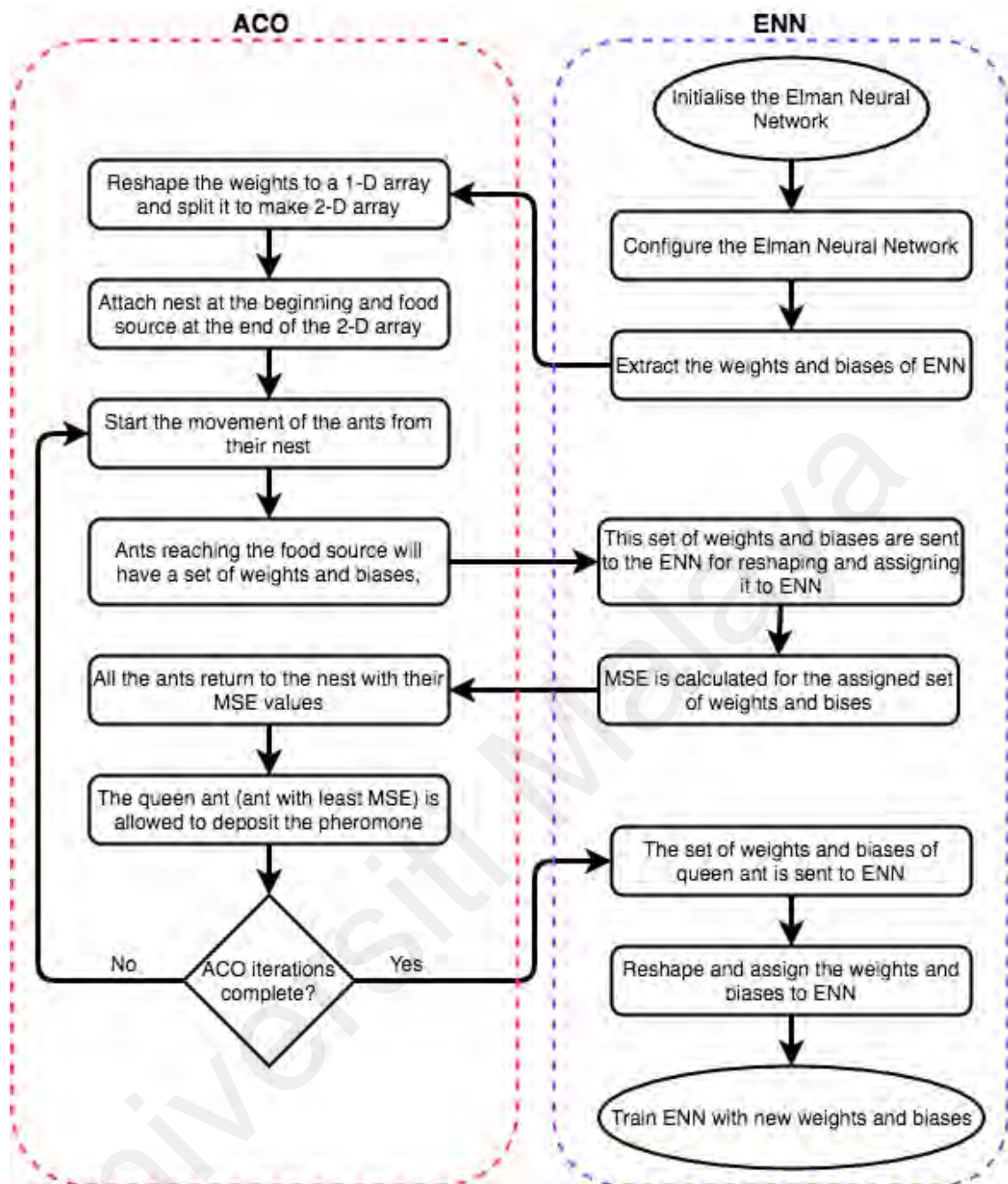


Figure 4.5 Flow chart of model structure

4.4.4 Application of ACO-ENN Hybrid model on nitrogen data

The developed ACO-ENN Hybrid model was tested on the nitrate-nitrogen and ammonia-nitrogen prediction data obtained for this study. Four separate model were trained for nitrate-nitrogen and ammonia-nitrogen for corresponding stations, i.e. Station Lui and Station Kajang. Input and target data were provided to the hybrid model with 90% data assigned for training the model and remaining 10% data for the testing the trained model.

The data was divided for training and testing in such a way that training and testing data are statistically same (i.e. their average values are approximately same) with the maximum and the minimum data points present in the training portion. The following configurations of the model was used for training the hybrid model:

ACO Iteration: 50-200

Number of ants: 20-50

Evaporation rate: 0.5

Number of weight-split layers: 20-50

Elman hidden layers: varied from 2 to 3

Elman nodes in each hidden layer: varied from 7 to 10

Elman train function: “trainlm”

Elman Performance Function: “mse”

Elman Epochs: 1000-2000

ACO-ENN hybrid model begins the training with the deployment of ACO for optimization of weights and biases extracted ENN. ACO performs the optimization process for the defined number of iterations. Post completion of optimization process ACO provides ENN the best set of weights and biases to train upon. ACO-ENN hybrid model then trains the ENN model, for defined number of epochs, to deliver the predicted model results. The default training function “trainlm” is used for training the hybrid model. “trainlm” is the Levenberg-Marquardt backpropagation training algorithm which updates weight and bias values according to Levenberg-Marquardt procedure. The results obtained using this configuration of hybrid model yielded improved results in comparison to that obtained from multilayer ANN model, as discussed in chapter 3. The results

obtained from the hybrid model for nitrate-nitrogen and ammonia-nitrogen prediction at corresponding stations (i.e. Station Lui and Station Kajang) are presented in the following section.

4.4.5 Performance criteria

The performance criteria for hybrid model is the same as that used in previous article, except for few additions. The results of hybrid model are analysed on the basis of six different performance criteria, which are listed as: MSE (equation 4.4), MAE (equation 4.5), NSE (equation 4.6), the maximum relative percentage error (equation 4.7), Peak Flow Criteria (PFC) (equation 4.8) and Low Flow Criteria (LFC) (equation 4.9). MSE and MAE measure the error between the predicted and the target values. NSE quantifies the accuracy of the model on the scale of $-\infty$ to 1, where 1 represents the perfect model predicting the exact target values. PFC and LFC represent the ability of the model to predict the peak and low values, respectively. These two criteria measure the error of peak values and low flow values. Zero value of PFC and LFC represents the perfect model (Coulibaly, Anctil, et al., 2001), predicting exactly the peak and low values. The equations used for calculation of different parameters are as follows:

MSE

$$MSE = \frac{1}{n} \sum_{i=1}^n (x - y)^2 \quad (4.4)$$

MAE

$$MAE = \frac{1}{n} \sum_{i=1}^n |x - y| \quad (4.5)$$

NSE

$$NSE = 1 - \frac{\sum(y - x)^2}{\sum(x - \bar{x})^2} \quad (4.6)$$

Maximum relative percentage error

$$\text{Maximum error} = \max \left(\left| \frac{x_i - y_i}{x_i} \right| * 100 \right), i = 1 \text{ to } n \quad (4.7)$$

PFC

$$PFC = \frac{(\sum_1^{T_P} (x-y)^2 * x^2)^{0.25}}{(\sum_1^{T_P} x^2)^{0.5}} \quad (4.8)$$

LFC

$$LFC = \frac{(\sum_1^{T_L} (x-y)^2 * x^2)^{0.25}}{(\sum_1^{T_L} x^2)^{0.5}} \quad (4.9)$$

where: for this study, n = number of data points, x = observed data points, y = predicted data points, T_P = the number of peak flow greater than one-third of observed mean peak flow, and T_L = the number of low flows lower than one-third of observed mean low flow (Coulibaly, e, et al., 2001).

4.5 Results

Training the hybrid model at the above-mentioned configuration and trying all the possible configurations in the given range of ACO iterations, number of ants, number of weights split layers, hidden layers and nodes in hidden layers, hybrid model improved the accuracy of the results significantly. One of the results of the hybrid model is the plot of

the MSE convergence, which the queen ant has achieved at the end of every ACO iterations. Figure 4.6 shows the plot of the MSE convergence in the ACO iterations. The plot shows how the ants, by their pheromone communication, have managed to find the best path having minimum MSE value. Figure 4.6 is the plot of the MSE convergence obtained while training the model for predicting nitrate-nitrogen at Station Lui. This figure represents how the ants in the ACO had found the path, among the network of paths, which have the least MSE. Towards the end of the iteration all the ants have converged to the path having least MSE. As can be seen in the figure 4.6, ants have tried to converge several times at different paths (the portion of the plot having constant MSE), but after few iterations queen ant found the path which have even lesser MSE. In similar steps queen ant managed to find the ultimate least MSE path, leading other ants to converge on that path. This MSE and its set of weights and biases was then assigned to the weights and biases of the ENN learning model. ENN trained on this MSE provided enhanced results. Figure 4.7-4.9 are the plots of the MSE convergence for the models predicting ammonia-nitrogen at station Lui, nitrate-nitrogen at station Kajang and ammonia-nitrogen at station Kajang, respectively. These plots show how the ACO iterations in all the models have converged to the least MSE.

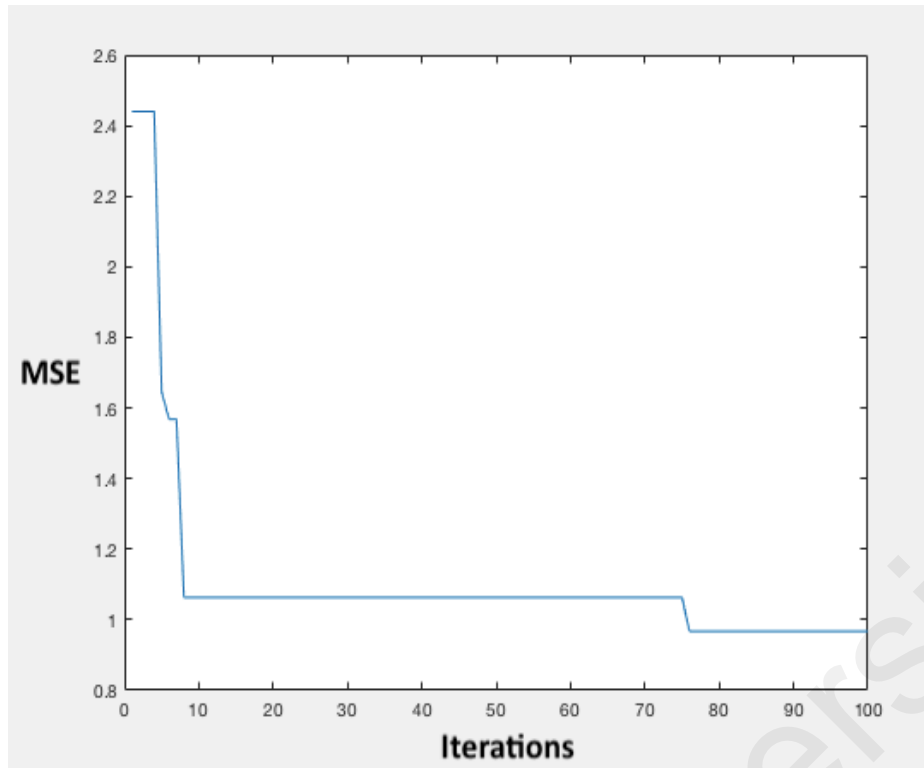


Figure 4.6 Convergence of MSE in ACO for Nitrate-Nitrogen model at station Lui

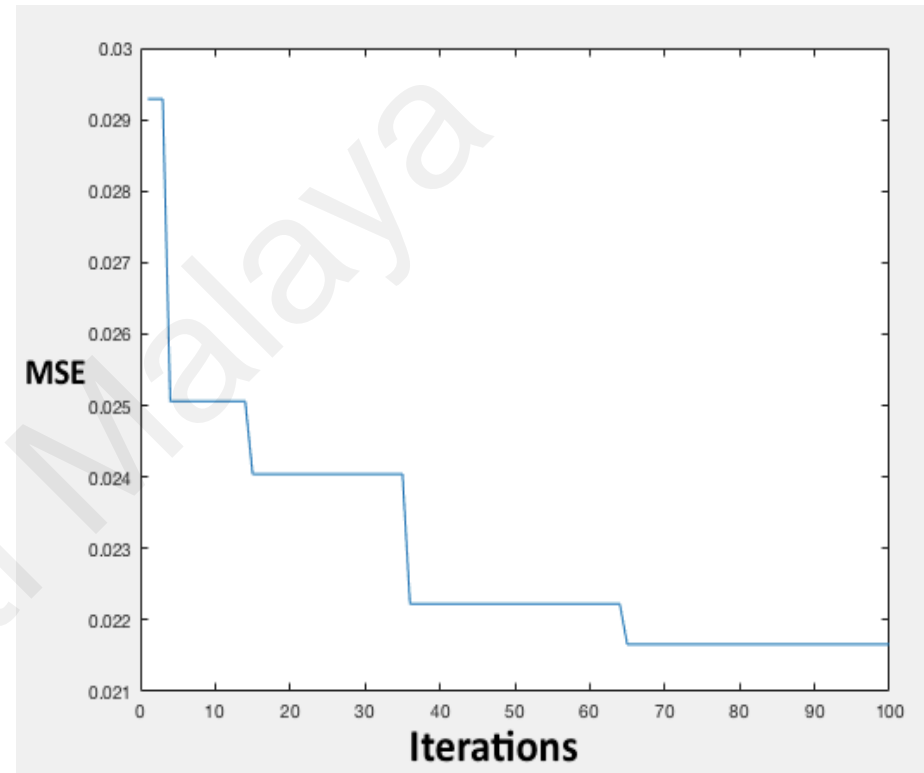


Figure 4.7 Convergence of MSE in ACO for Ammonia-Nitrogen model at station Lui

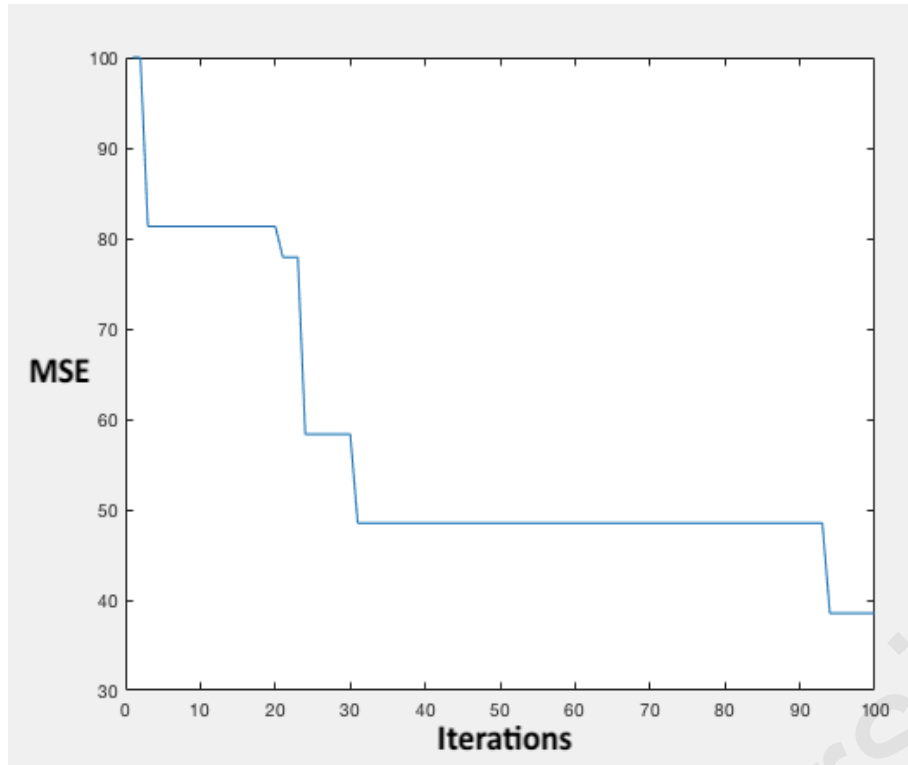


Figure 4.8 Convergence of MSE in ACO for Nitrate-Nitrogen model at station Kajang

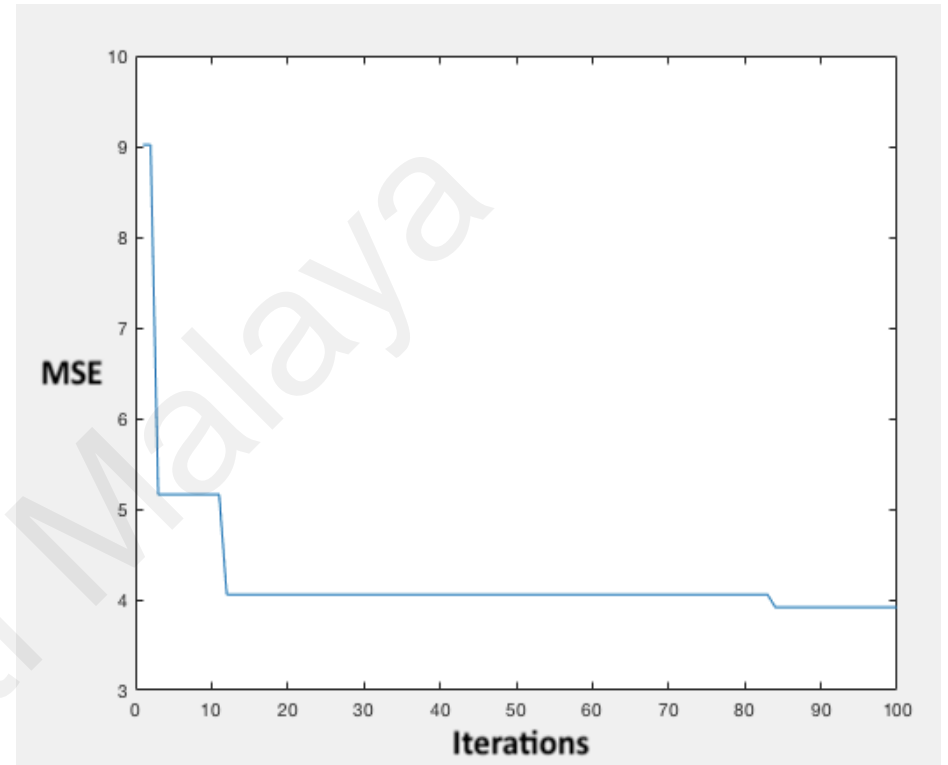


Figure 4.9 Convergence of MSE in ACO for Ammonia-Nitrogen model at station Kajang

Table 4.1 Configurations of hybrid models for Lui and Kajang stations

Station		Lui		Kajang	
Nitrogen Compound		Nitrate-N	Ammonia-N	Nitrate-N	Ammonia-N
Parameters	ACO Iterations	100	100	100	100
	Number of Ants	50	50	50	50
	Number of weight-split layers	50	50	50	50
	Elman network hidden layer	2	2	3	2
	Nodes	10	7	10	9
Accuracy	Mean Square Error (mg/l)	0.012	0.0008	1.21	0.112
	Mean Absolute Error (mg/l)	0.069	0.0169	0.396	0.16
	Nash-Sutcliffe Efficiency	0.984	0.957	0.965	0.97
	Maximum Error	52.21%	91.79%	89.63%	75.38%
	PFC	0.070	0.118	0.103	0.117
	LFC	0.330	0.293	0.335	0.509

Table 4.1 shows the configurations and model accuracy of all the four hybrid models trained for prediction of nitrate nitrogen and ammonia nitrogen at station Lui and station Kajang. Configuration of the hybrid model represents the parameters of both ACO and Elman neural network. For ACO, the parameters are the ACO iterations, number of ants and number of weight-split layers. For Elman neural network, the parameters are hidden layers and nodes in hidden layers. Accuracy of the hybrid models are measured on the basis of the above-mentioned performance criteria. Mean square error and mean absolute error represent the overall error (difference between the target and predicted values) produced by the model while training. NSE criteria measures the model's capability of predicting the values close to the target values. As stated earlier, for model's capability measurement, NSE uses the scale of $-\infty$ to 1, where 1 represents the perfect generalized model. NSE values of the hybrid models are more than 0.95, which are close to 1. The

maximum error criteria represent the lack of proper generalization of the model. The maximum error of the hybrid models is less than 100%. PFC represents the ability of the model to learn the pattern, in the target data provided to the model, when there is peak flow. In other words, it represents the ability of the model to predict accurately the peak flows values. Similarly, LFC represents the ability of the model to predict accurately the low flow values.

Figure 4.10 represents the plot of target and output for all the hybrid models developed. The output plot seems to cover the target plots for all the hybrid models, which means that hybrid models are predicting almost the same as target values.

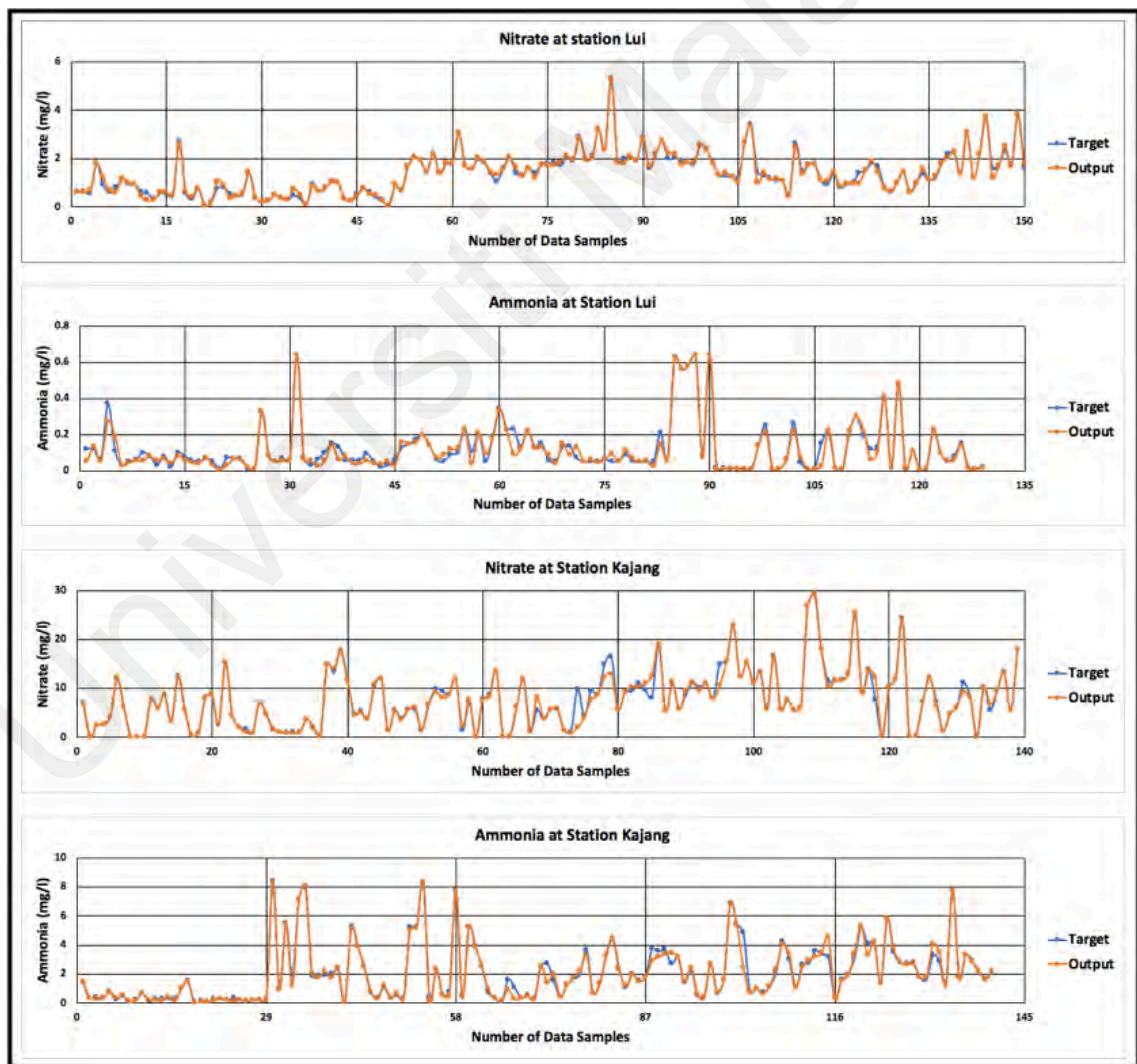


Figure 4.10 Plot of target and output for each hybrid model

Table 4.2 shows the comparison of the results obtained from the hybrid models with the existing results obtained from the existing standalone ANN models (from the article in chapter 3). Both the model results are compared on the basis of all the six performance parameters, mentioned earlier i.e. MSE, MAE, NSE, maximum error, PFC and LFC. Results from the new hybrid model showed a significant improvement in almost all performance criteria computed. With the help of new hybrid model, MSE and MAE values have been reduced up to a certain extent. The new hybrid model has brought the NSE values closer to 1, making the hybrid model more perfect in comparison to the existing standalone ANN model. In some cases of the current study, the maximum error for existing standalone ANN model was more than 100% which have been reduced below 100% with the help of hybrid model.

Hybrid model helped in reducing PFC and LFC closer to zero; which means after training the models with new hybrid algorithm, models are now capable of predicting more accurately the values of nitrate-nitrogen and ammonia-nitrogen when they are at their peak flow and low flow conditions. Thereby, increasing the prediction capability of the models at both the extreme values of the pattern of target data.

Table 4.2 Comparison of the result of ANN and Hybrid model

Stations	Lui				Kajang			
	Nitrate		Ammonia		Nitrate		Ammonia	
Type	ANN	Hybrid	ANN	Hybrid	ANN	Hybrid	ANN	Hybrid
MSE (mg/l)	0.027	0.012	0.0013	0.0008	7.09	1.21	0.121	0.112
MAE (mg/l)	0.098	0.069	0.017	0.0169	0.771	0.396	0.173	0.16
NSE	0.967	0.984	0.946	0.957	0.959	0.965	0.97	0.97
Maximum Error	73.41%	52.21%	129.5%	91.79%	97.81%	89.63%	130.93%	75.38%
PFC	0.091	0.070	0.173	0.118	0.118	0.103	0.098	0.117
LFC	0.336	0.330	0.370	0.293	0.670	0.335	0.723	0.509

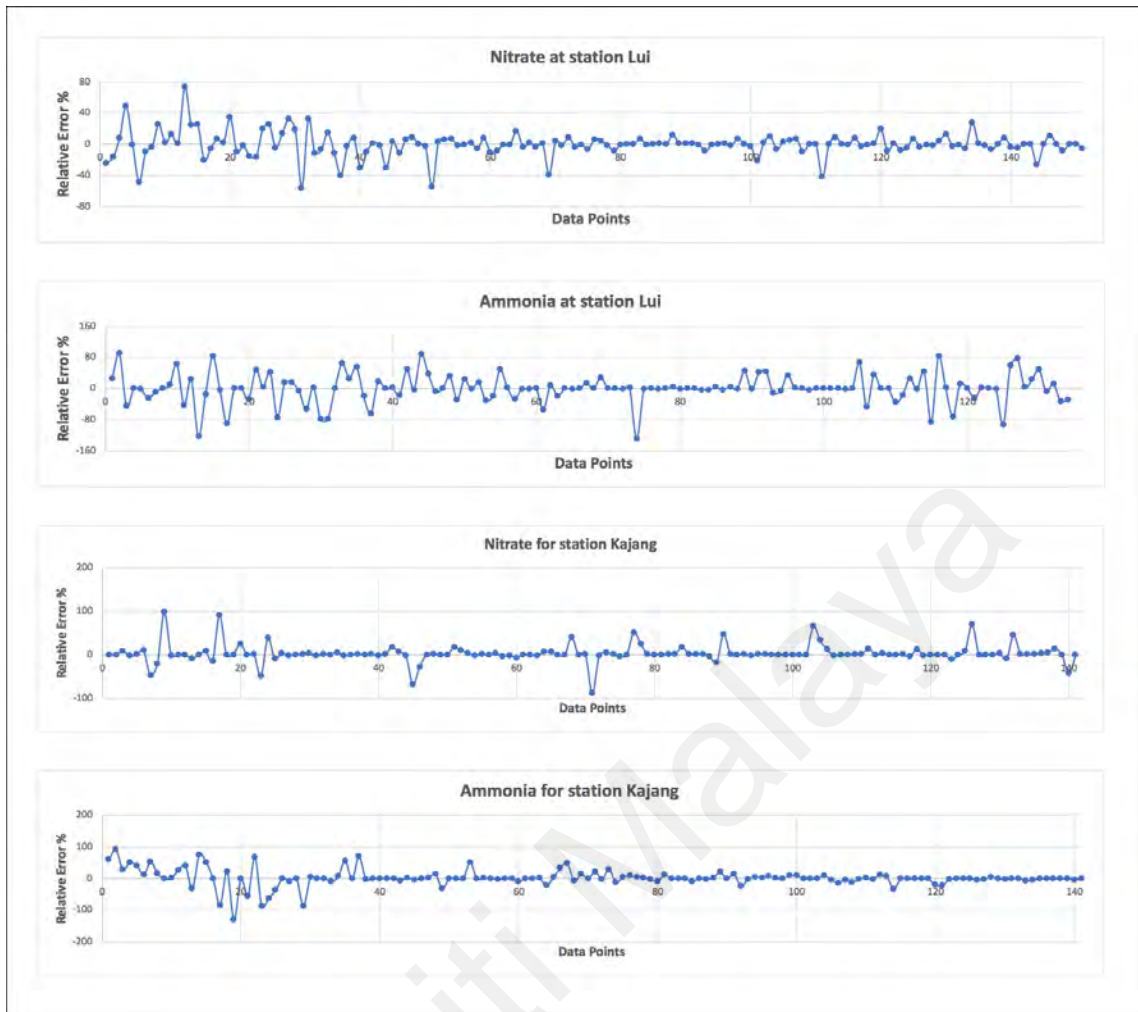


Figure 4.11 Plots of relative percentage error obtained from ANN

Comparing the plots of relative percentage error of hybrid models with that of multilayer ANN models state that the maximum error points have been reduced. For comparison the plots of relative percentage error, obtained through multilayer ANN in previous article, have been reused in this section. Figure 4.11 represents the plot of relative percentage error for the all the four models (nitrate-nitrogen and ammonia-nitrogen for corresponding stations i.e. Station Lui and Station Kajang) obtained from the standalone ANN models. In comparison to these plots, figure 4.12 represents the plot of improved relative percentage error of all the four hybrid models.

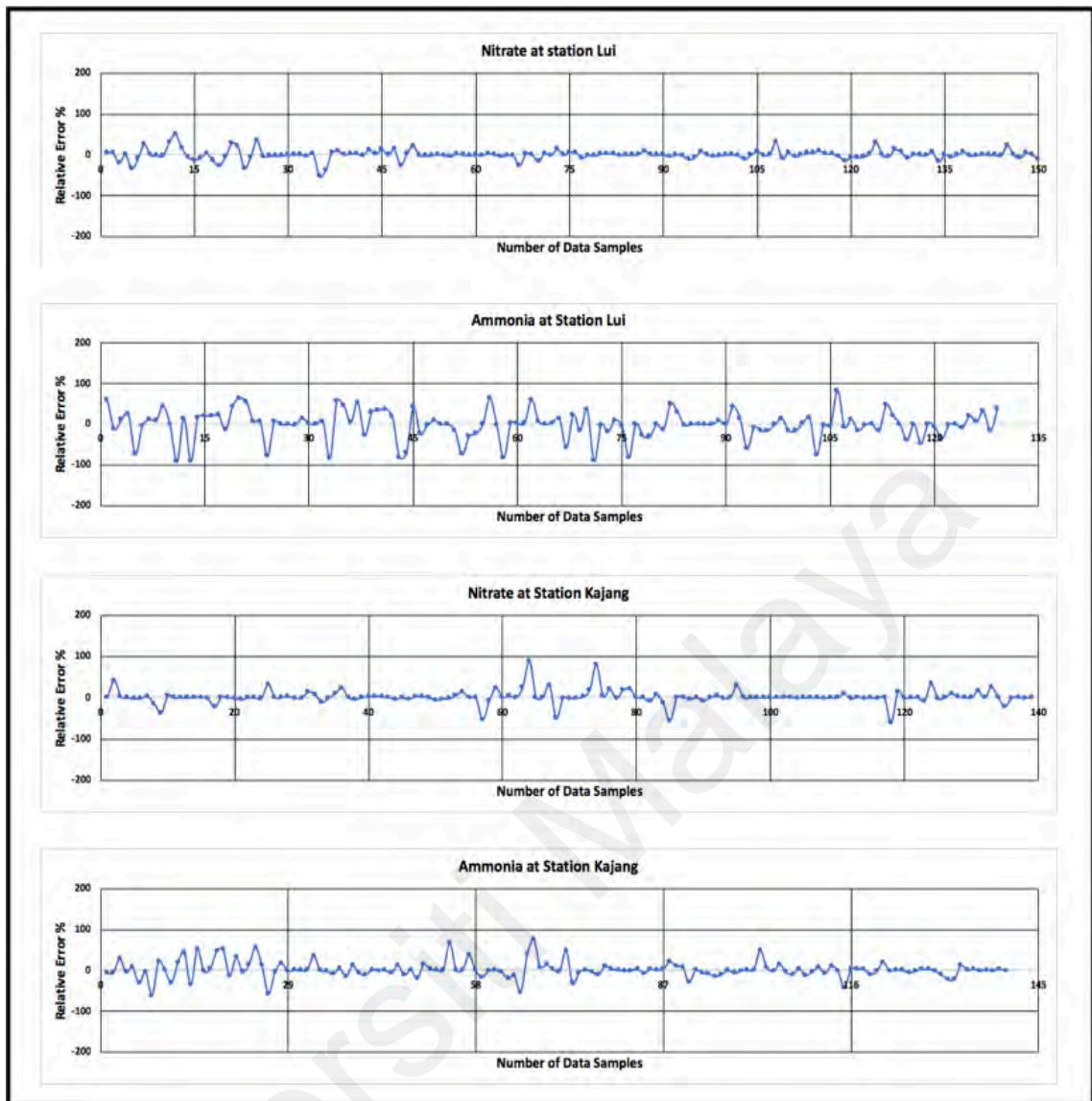


Figure 4.12 Plots of relative percentage error obtained from hybrid model

Observing plots of relative percentage error for both ANN and hybrid model, it can be stated that new hybrid training algorithm enhances the generalization capability of the model. Thus, improving the accuracy of the prediction results as the errors of the predicted values have been reduce up to a certain extent.

4.6 Discussion

The hybrid model developed by integrating ACO with Elman network shows improvement in the existing results when trained on the same data. Table 4.3 represents

the percentage improvement of different performance criteria when compared with the existing standalone ANN model results. As shown in the table 4.3, the new hybrid training algorithm have improved the model capability in almost all respect. The error parameters, such as MSE, MAE and maximum relative percentage error (maximum error), have been improved significantly, leading to the reduction of prediction error up to a certain extent. The PFC and LFC have been enhanced closer to zero, leading to the increased accuracy of the model for predicting the peak flow values and low flow values. Hence, the new ACO-ENN hybrid model finds itself to be more suitable, in comparison with other ANN models, for developing a model with enhanced capabilities.

The results of the ACO-ENN hybrid model were also compared with the results of standalone Elman neural network model. This comparison proved the effectiveness of the ACO-ENN hybrid model over the ANN models. The comparison is presented in the table 4.4. Elman network was trained on the same dataset with the same Elman network configurations, as used by hybrid model. The standalone ENN model provided inferior results compared to the ACO-ENN hybrid model. As presented in table 4.4, the maximum percentage error values were way higher than those of hybrid models. The error parameters (i.e. MSE, MAE) were also higher than the corresponding hybrid models. The

Table 4.3 Performance improvement of hybrid model

Station	Lui		Kajang	
	Nitrate	Ammonia	Nitrate	Ammonia
ANN MSE (mg/l)	0.027	0.0013	7.09	0.121
Hybrid MSE (mg/l)	0.012	0.0008	1.21	0.112
% MSE Improvement	55.56 %	38.46 %	82.93 %	7.44 %
ANN MAE (mg/l)	0.098	0.017	0.771	0.173
Hybrid MAE (mg/l)	0.069	0.0169	0.396	0.16
% MAE Improvement	29.59 %	0.59 %	48.64 %	7.51 %
ANN NSE	0.967	0.946	0.959	0.97

Table 4.3, continued

Station	Lui		Kajang	
Nitrogen Compound	Nitrate	Ammonia	Nitrate	Ammonia
Hybrid NSE	0.984	0.957	0.965	0.97
% NSE Improvement	1.76 %	1.16 %	0.63 %	0.00 %
ANN Max Error %	73.41	129.5	97.81	130.93
Hybrid Max Error %	52.21	91.79	89.63	75.38
% Max Error Improvement	21.2 %	37.71 %	8.18 %	55.55 %
ANN PFC	0.091	0.173	0.118	0.098
Hybrid PFC	0.070	0.118	0.103	0.117
% PFC Improvement	23.41 %	32.00 %	13.02 %	-19.44 %
ANN LFC	0.336	0.370	0.670	0.723
Hybrid LFC	0.330	0.293	0.335	0.509
% LFC Improvement	1.56 %	20.97 %	49.93 %	29.60 %

Table 4.4 Comparison of the result of ENN and Hybrid model

Stations	Lui				Kajang			
	Nitrate		Ammonia		Nitrate		Ammonia	
Type	Hybrid	ENN	Hybrid	ENN	Hybrid	ENN	Hybrid	ENN
Regression	0.967	0.967	0.979	0.841	0.983	0.831	0.986	0.918
MSE (mg/l)	0.012	0.049	0.0008	0.0066	1.21	12.8	0.112	0.696
MAE (mg/l)	0.069	0.094	0.0169	0.033	0.396	1.75	0.16	0.286
NSE	0.984	0.933	0.957	0.676	0.965	0.640	0.97	0.815
Maximum Error	52.21%	289.41%	91.79%	777.2%	89.63%	710.4%	75.38%	208.9%

regression values and the NSE values were found to be closer to that of hybrid models yet inferior to hybrid models. Thus, confirming the ability of ACO-ENN hybrid model for producing enhanced prediction results.

The accuracy of ACO-ENN hybrid model depends on number of ACO iterations, number of ants, number of weight-split layers, Elman network hidden layers, nodes in hidden

layers and Elman network epochs. Increasing the number of ACO iterations will increase the probability of exploring all the paths. However, after sufficient iterations all the ants starts converging to a particular path. Increasing the number of ants will increase the probability of each paths to be explored at least once. Lesser number of ants may lead to poor accuracy. Increasing the number of weight-split layers will increase the combination of weights to be explored which may lead to higher accuracy. Number of hidden layers in Elman network, nodes in hidden layers and epochs depend on the data and the pattern in it to learn. Network with lesser number hidden layers, nodes and epochs may not be capable of learning all the patterns of the target data. However, higher number of hidden layers, nodes and epochs may lead the model to overfit, hence, reducing the model accuracy.

The rules and properties of the new ACO-ENN hybrid model are as listed below:

1. ACO-ENN hybrid model searches for the optimal values of the internal parameters of the Elman neural network
2. It accelerates the convergence process during the training stage to achieve the performance goal, the faster convergence rate the better for attaining the better model output and hence it proves suitable for real-time application for the proposed model as well

The developed hybrid model has the advantage of increasing the learning capability of the ANN model and hence increasing the prediction accuracy of the model. However, the disadvantage of the developed hybrid model is that it needs higher computation time. However, the computation time can be reduced with modern computation resources.

4.6.1 Software Interface

Based on the above discussed methodology, a software interface was designed based on MATLAB platform. This software provides an interface for the world-wide users to use the newly developed ACO-ENN hybrid model. The software can be used by installing it in the MATLAB application.

As discussed in above sections, the application of ACO-ENN hybrid model is not limited to nitrogen prediction or to hydrological parameters predictions. This hybrid model can be used to train any sort of model in any field, provided the training process should follow the supervised learning method. Figure 4.13 represents the home screen window of the interface. It consists of a short description of hybrid model along with the graphical abstract. Model description button pops-out the window containing the detailed explanation of the model (Figure 4.14). Proceed button leads to the window designed for creating the Elman neural network (Figure 4.15).

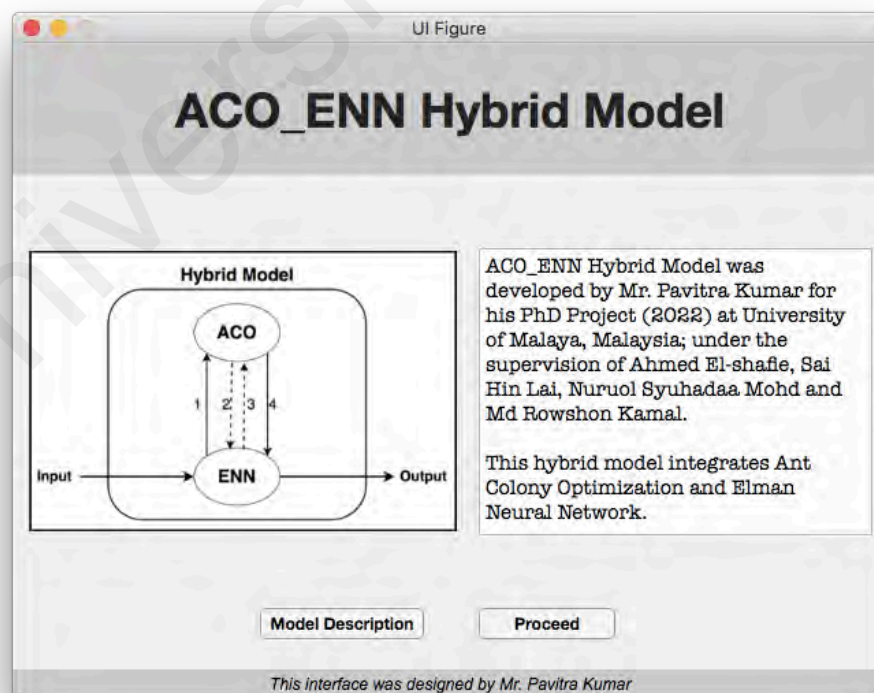


Figure 4.13 ACO-ENN hybrid model software interface

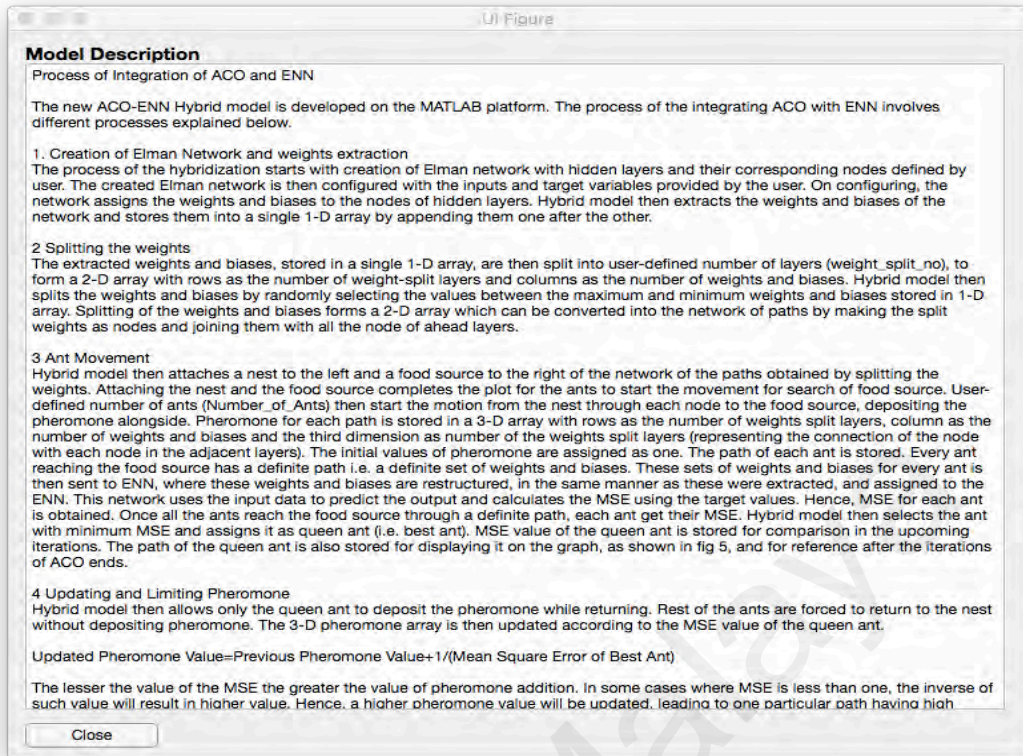


Figure 4.14 Model Description

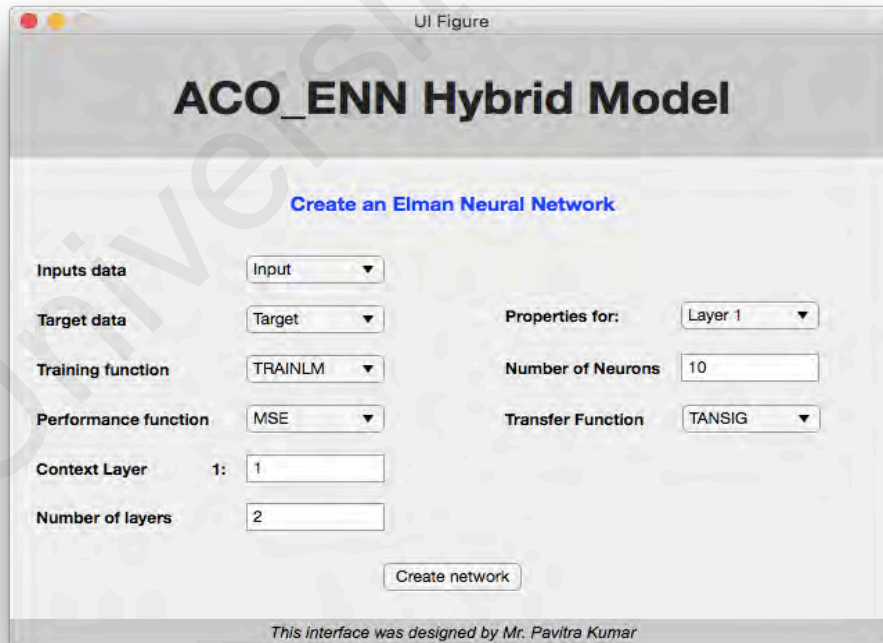


Figure 4.15 Creating Elman Neural Network

User needs to provide the inputs and target data in the “Create an Elman Neural Network” window (Figure 4.15). User can alter rest of the configurations of the Elman neural network, however, otherwise the software will pick the default values. The default values of the training function and performance function are assigned as “TRAINLM” and “MSE”, respectively. By default, the network has 1:1 context layer and two layers (one hidden layer and one output layer) with 10 nodes in hidden layer. The transfer function has a default value of “TANSIG”. The “Create network” button creates the network according to the configuration provided by the user or picked up by default. Creating the network, updates the window, as shown in figure 4.16. At this window user can view the network or proceed to train the network. User can also change the network configuration and create a new network accordingly.

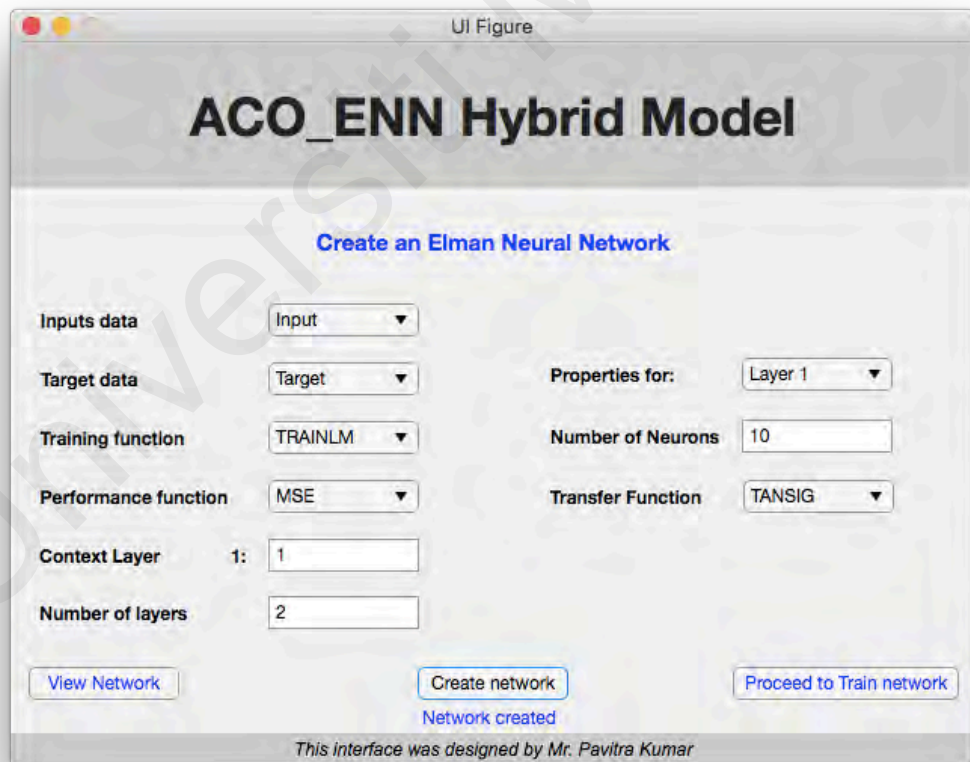


Figure 4.16 Window after creating Elman neural network

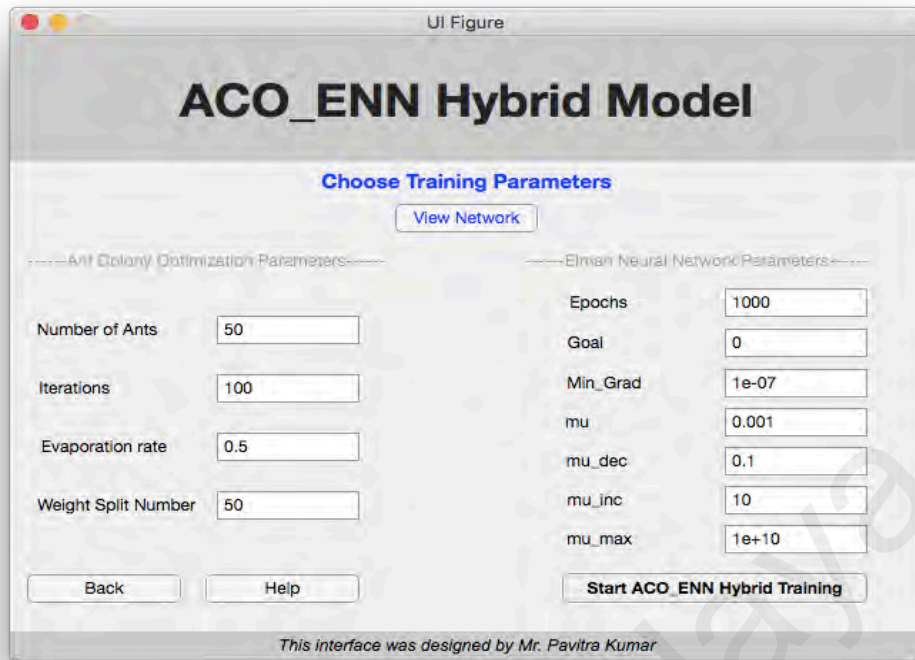


Figure 4.17 Window for choosing training parameters

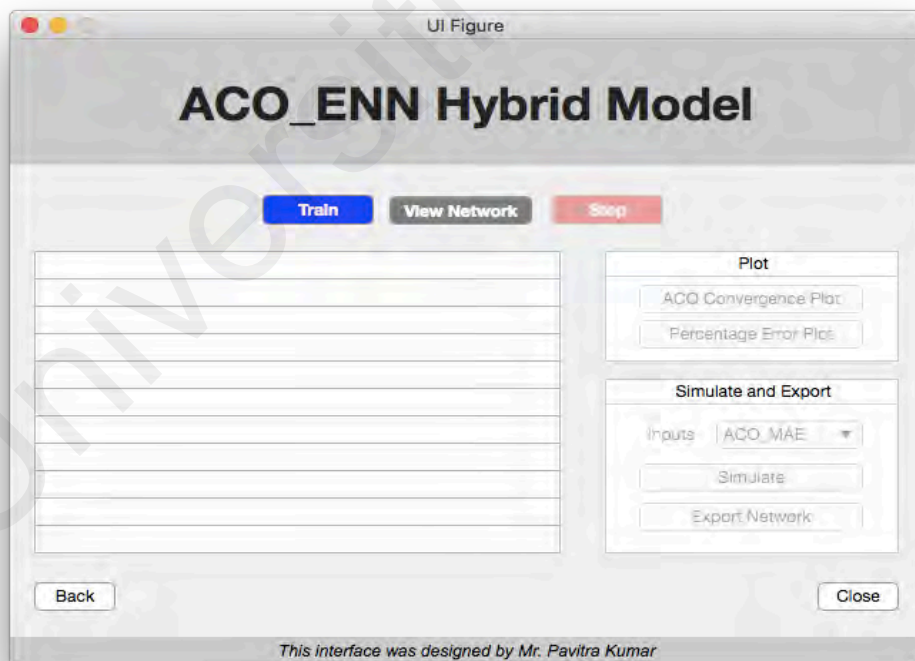


Figure 4.18 Window for ACO iterations

Proceeding to train the network bring the “Choose the training parameters” window, (figure 4.17). This window allows the user to provide the training parameters of ant colony optimization and Elman neural network. However, there are default values assigned to each parameter. For ant colony optimization parameters, user needs to provide the number of ants, number of iterations, evaporation rates and weight split number. For Elman neural network, user needs to provide the epochs, goal, minimum gradient, mu, mu decreasing gradient, mu increasing gradient, and maximum value of mu. Providing the values for each parameter, user can proceed to the training window by clicking “Start ACO-ENN Hybrid Training” button. On the training window (figure 4.18), user can start the training by clicking the “Train” button. Once the train button is clicked, the software interface starts the ACO iterations by displaying the MSE values for each iteration (figure 4.19). However, user can stop the ACO iterations at any time by clicking the Stop button.

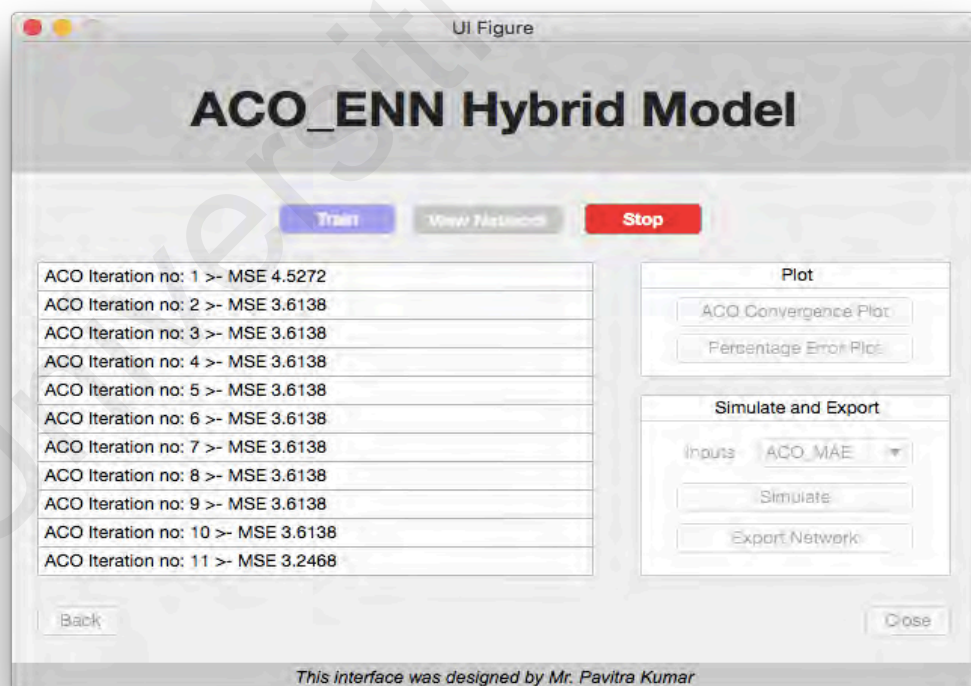


Figure 4.19 ACO Training

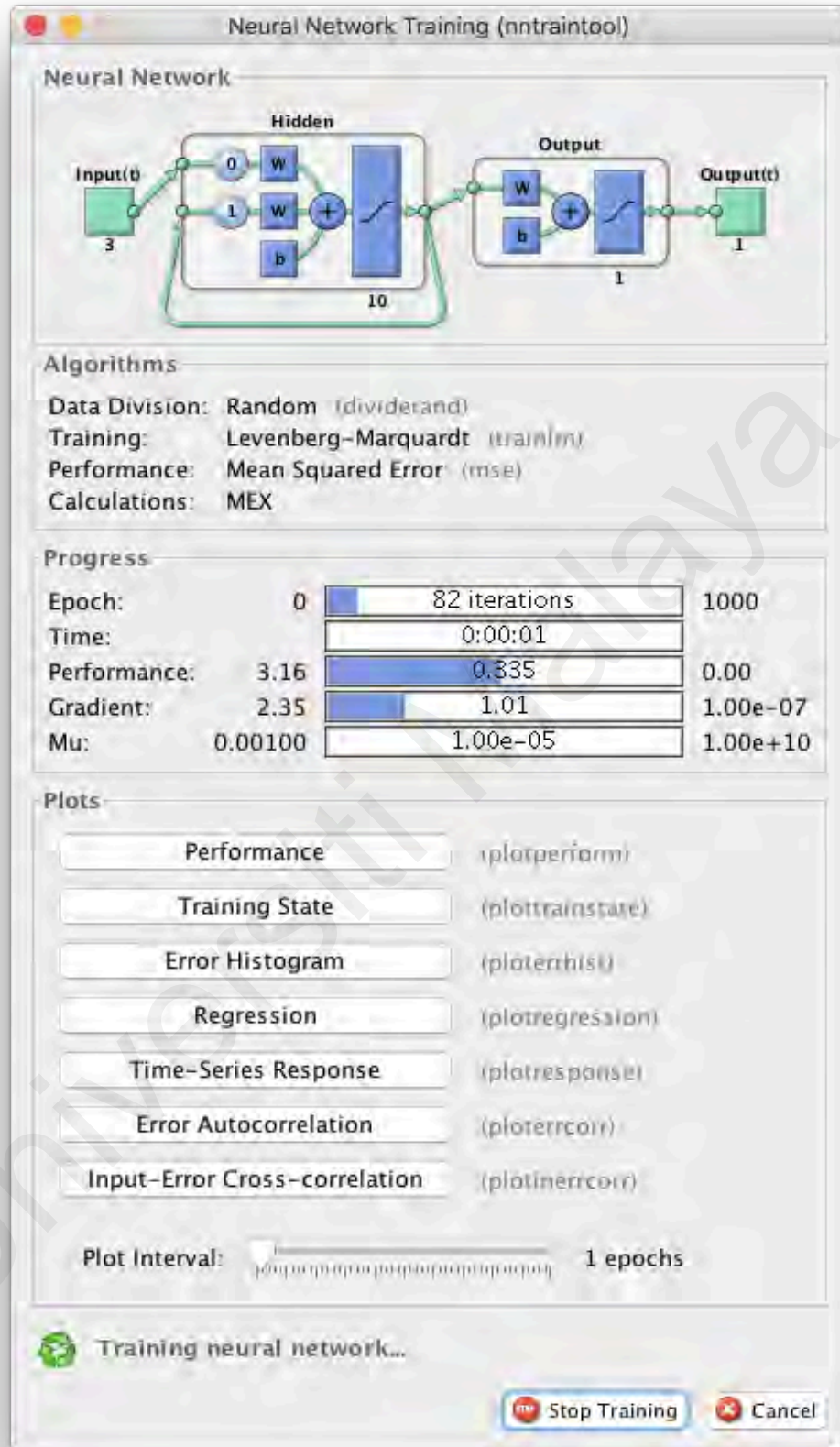


Figure 4.20 Elman neural network training

Upon completion of ACO iterations, the software interface then starts the Elman neural network training, as shown in figure 4.20. Training of Elman neural network can be stopped at any time by clicking “Stop Training” button. Upon completion of all the training, the “Training” window (figure 4.21) appears again, but with plot and simulate and export options enabled. User can train the network again by clicking on “Train” button. User can also view the ACO convergence plot or percentage error plot. The software interface provides its users the option of exporting the network and also the option of simulating and exporting the results on user defined inputs. Upon simulating or exporting the network, the simulated results and the network are stored in the current workspace of the MATLAB application, where user can use it for further calculations.

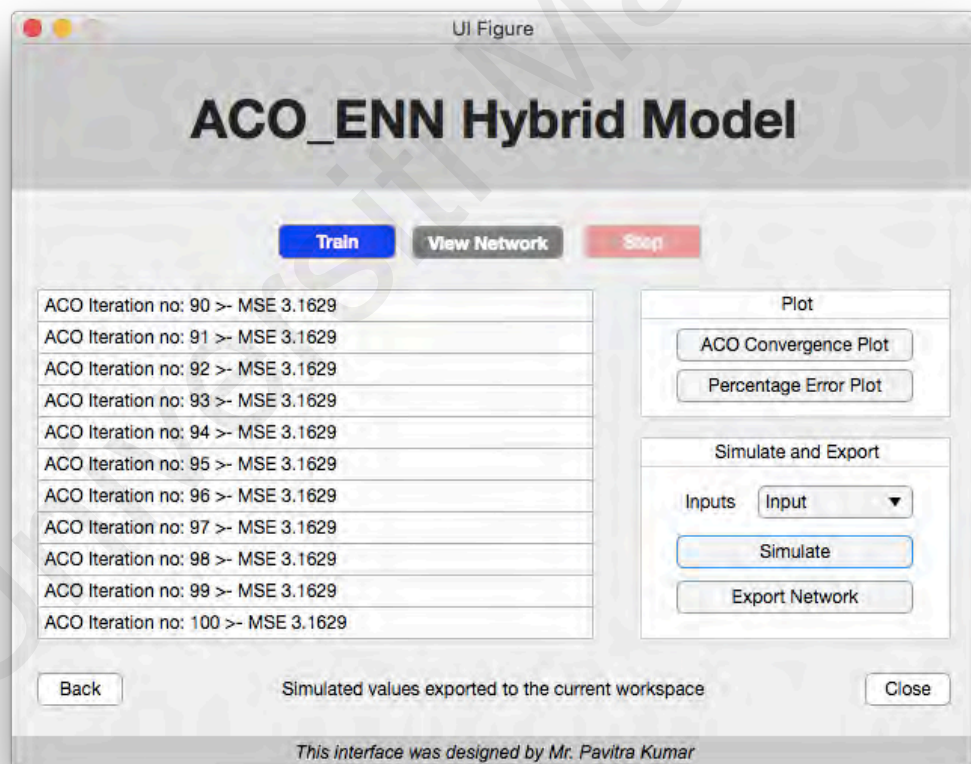


Figure 4.21 Results of the Trained network

4.7 Summary

In this article, a new hybrid model is successfully developed by integrating ACO with the ENN. Here, ACO is used as a decision-making model which decides the initial weights and biases of the ENN. In this hybrid model, ENN is used as a learning model which learns various patterns of the target variable and provides a network for data prediction. Selecting the initial weights and biases for the ENN provides it a better start which leads the ENN to achieve high accuracy results. The ACO-ENN hybrid model is tested on the nitrate-nitrogen and ammonia-nitrogen prediction data for the two different stations i.e. Station Lui and Station Kajang in Langat river basin in Malaysia. The hybrid model yielded improved results, in comparison to the previously existing results. There was significant improvement in the MSE, MAE, the maximum relative percentage error, NSE, PFC and LFC. Training the model on new ACO-ENN hybrid model algorithm resulted in enhanced capabilities of prediction model, especially in predicting the conditions when the nitrogen compound was in low flow and peak flow. The new hybrid model has the advantage of enhancing the model's prediction capabilities. However, the new hybrid model has some disadvantage, such as: it requires more training time and it requires modern computational resources.

The final conclusion and recommendation of future work related to the current study is described in chapter 5.

CHAPTER 5. CONCLUSION

5.1 Introduction

This chapter contains the overall conclusion of the current research work. Based on this study recommendation of future works have been proposed in this chapter.

5.2 Conclusion

Artificial Neural Network models have been developed for the prediction of nitrate-nitrogen and ammonia-nitrogen at two different geographical locations in Langat River basin in Malaysia. Additionally, a new hybrid modelling technique have been presented in this research work which can be used for developing prediction models in various other fields. However, in this research work the new hybrid modelling technique have been used to enhance the prediction accuracy of the nitrogen pollutants in rivers. Models developed using the hybrid modelling technique had the outstanding performance.

Following are the main points of the conclusion obtained from the results of this study:

1. Complete and accurate information of the study area leads to better prediction accuracy of the model. More the information of the site will lead to the better generalization of the target variable. As shown in chapter 3, figure 3.11, the mean square error of the predicted results is least for the model having three inputs, in comparison with the models having two inputs. Hence, it can be concluded that the model having three inputs has more generalizing capability leading to better results and least mean square error. Also, the quantity of the data (i.e. number of data samples or in other words the number of years of historical data) has an important role in generalization of the model. Models trained on huge amount of

data will always deliver results with higher accuracy. Missing data will always be the source of uncertainty in the predicted results. Missing data is the gap in continuation of the data, which leads to the loss of the patterns in the data affecting the generalization capability of the model. Sometimes, interpolation of the missing data may complete the gap but still there remains some uncertainty in the interpolated values, which is reflected in the predicted results.

2. Proper configuration of internal parameters of the artificial neural network model has a key role in the model performance. Choosing the internal parameters of the model is a trial-error process and usually time-consuming process. To achieve the better combination, each combinations must be explored and tested for the performance accuracy. As shown in chapter 3, figure 3.12-3.17 present the mean square error comparison against the variation of different internal parameters. The percentage training data provides the model important information while training. Models trained on larger set of training data will always provide higher accuracy, provided there must be enough data left for testing and validation. Number of nodes and hidden layers in the model increases the complexity of the model network, thus, strengthening it to learn various patterns in the training data. If the complexity of the network is more or less than the optimum required then the accuracy of the predicted results declines. Less complex network fails to learn all the patterns present in the training data and more complex network usually overfits, leading to the reduction in the prediction accuracy. Number of training epochs allows the network to learn all the important information in iterations. To allow the network to learn sufficient information it should be allowed run sufficient epochs. Less number of epochs will reduce the learning capability of the model and higher number of epochs will over-train the network, leading to reduction in prediction accuracy.

3. Natural hydrological variables can be modelled using artificial neural network, by choosing optimum model among the models with different set of internal parameters as mentioned above. This study proposes a new training approach to obtain the optimum model with better prediction accuracy. Using this approach, four optimum models have been trained for prediction of nitrate-nitrogen and ammonia-nitrogen for the stations LUI and KAJANG, situated in Langat River basin in Malaysia. These models provide sufficient prediction accuracy of the nitrogen pollutant.
4. Model performance of the standalone artificial neural network model can be enhanced by using hybrid models. This study proposes a new hybrid modelling technique developed by integrating ACO with ENN. ACO has been used as a decision making model which decides the best set of weights and biases for ENN. ENN has been used as a learning model which learns different patterns of the training data. The models obtained from the ACO-ENN modelling technique have enhanced the prediction results in comparison with the results obtained from the standalone artificial neural network developed above.

5.3 Recommendations for future work

On the basis of the results obtained from this study, the following recommendations are being proposed:

1. In this research work, only three types of ANN, i.e. GRNN, RBFNN and multilayer, have been tested. However, there are several types of ANN in the list with their own speciality in specific fields. It is recommended to explore other types of ANN models for developing the prediction models. Other ANN models

may have the tendency to generalize the data in much better way providing better prediction results.

2. As explained above, more information for provided as inputs will lead to better accuracy of the prediction results. For this study, three variables, i.e. rainfall, water level and discharge, have been used as inputs. Models have provided sufficient prediction accuracy, which is further enhanced using hybrid modelling technique. It is recommended to include other variable upon which nitrogen concentration in the rivers depend such as amount of fertilizers monthly used in the farms, geographical variables, amount of sewage monthly dumped into the rivers and many more variables. It is worth exploring other variables as inputs to achieve better prediction results.
3. Hybrid model is developed by integrating different models together for increasing the prediction capability of the standalone models. In this study, ant colony optimization have been integrated with Elman neural network to develop a new hybrid model. However, there still remains scope of integrating other models together to form a new hybrid model. Swarm algorithm, firefly algorithm and other algorithm can be integrated with different ANN models to explore the possibility of developing new hybrid model for prediction accuracy enhancement.

References:

- Aguilera, P. A., Frenich, A. G., Torres, J. A., Castro, H., Vidal, J. L. M., & Canton, M. (2001). Application of the kohonen neural network in coastal water management: Methodological development for the assessment and prediction of water quality. *water Resources*, 35(17), 4053-4062.
- Ahmed, A. N., El-Shafie, A., Karim, O. A., & El-Shafie, A. (2012). An augmented wavelet de-noising technique with neuro-fuzzy inference system for water quality prediction. *International Journal of Innovative Computing, Information and Control*, 8(10), 7055-7082.
- Akrami, S. A., El-Shafie, A., & Jaafar, O. (2013). Improving Rainfall Forecasting Efficiency Using Modified Adaptive Neuro-Fuzzy Inference System (MANFIS). *Water Resour Manage*. doi:10.1007/s11269-013-0361-9
- Akrami, S. A., El-Shafie, A., Naseri, M., & Santos, C. A. G. (2014). Rainfall data analyzing using moving average (MA) model and wavelet multi-resolution intelligent model for noise evaluation to improve the forecasting accuracy. *Neural Comput & Applic*, 25, 1853-1861. doi:10.1007/s00521-014-1675-0
- Amiri, B. J., & Nakane, K. (2009). Comparative prediction of stream water total nitrogen from land cover using artificial neural network and multiple linear regression approaches. *Polish Journal of Environmental Studies*, 18(2), 151-160. Retrieved from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-65249159650&partnerID=40&md5=6184c403bfc348fabac3d6a6034d7a2e>
- Ancil, F., Filion, M., & Tournebize, J. (2009). A neural network experiment on the simulation of daily nitrate-nitrogen and suspended sediment fluxes from a small agricultural catchment. *Ecological Modelling*, 220(6), 879-887. doi:10.1016/j.ecolmodel.2008.12.021
- ANI, E.-C., HUTCHINS, M. G., KRASLAWSKI, A., & AGACHI, P. S. (2010). Assessment of pollutant transport and river water quality using mathematical models. *Revue Roumaine de Chimie*, 55(4), 285-291.
- Aslan, S., & Turkman, A. (2003). Biological denitrification of drinking water using various natural organic solid substrates. *Water Science Technology*, 48(11-12), 489-495. doi:10.2166/wst.2004.0898
- AYERS, G. P., PENG, L. C., FOOK, L. S., KONG, C. W., GILLET, R. W., & MANINS, P. C. (1999). Atmospheric concentrations and deposition of oxidised sulfur and nitrogen species at Petaling Jaya, Malaysia, 1993-1998. *Tellus B: Chemical and Physical Meteorology*, 52(1), 60-73. doi:10.3402/tellusb.v52i1.16082
- Basheer, A. O., M. Hanafiah, M., & J. Abdulhasan, M. (2017). A Study on Water Quality from Langat River, Selangor. *Acta Scientifica Malaysia*, 1(2), 01-04. doi:10.26480/asm.02.2017.01.04
- Canadian Council of Ministers of the Environment. (2010, 26-06-2019). Canadian Water Quality Guidelines for the Protection of Aquatic Life: Ammonia. Retrieved from <http://ceqg-rcqe.ccme.ca/download/en/141>
- Chang Li-Chiu, & Chang Fi-John. (2001). Intelligent control for modelling of real-time reservoir operation. *Hydrological Processes*, 15(9), 1621-1634. doi:10.1002/hyp.226
- Chen, J., Wu, H., Qian, H., & Gao, Y. (2017). Assessing Nitrate and Fluoride Contaminants in Drinking Water and Their Health Risk of Rural Residents Living in a Semiarid Region of Northwest China. *Exposure and Health*, 9(3), 183-195. doi:10.1007/s12403-016-0231-9
- Chitsazan, N., Nadiri, A. A., & Tsai, F. T. C. (2015). Prediction and structural uncertainty analyses of artificial neural networks using hierarchical Bayesian model averaging. *Journal of Hydrology*, 528, 52-62. doi:10.1016/j.jhydrol.2015.06.007

- Cigizoglu, H. K., & Alp, M. (2004, June 7-11). *Rainfall-Runoff Modelling Using Three Neural Network Methods*. Paper presented at the ICAISC: International Conference on Artificial Intelligence and Soft Computing, Zakopane, Poland.
- Coulibaly, P., Anctil, F. o., & e, B. B. (2001). MULTIVARIATE RESERVOIR INFLOW FORECASTING USING TEMPORAL NEURAL NETWORKS. *JOURNAL OF HYDROLOGIC ENGINEERING*, 6, 367-376.
- Coulibaly, P., e, B. B., & Anctil, F. o. (2001). Improving extreme hydrologic events forecasting using a new criterion for artificial neural network selection. *Hydrological Processes*, 15, 1533-1536. doi:10.1002/hyp.445
- Czernuszenko, W. (1987). Dispersion of pollutants in rivers. *Hydrological Sciences Journal*, 32(1), 59-67. doi:10.1080/02626668709491162
- de Gennaro, G., Trizio, L., Di Gilio, A., Pey, J., Perez, N., Cusack, M., . . . Querol, X. (2013). Neural network model for the prediction of PM10 daily concentrations in two sites in the Western Mediterranean. *Sci Total Environ*, 463-464, 875-883. doi:10.1016/j.scitotenv.2013.06.093
- Della Rocca, C., Belgiorno, V., & Meric, S. (2005). Cotton-supported heterotrophic denitrification of nitrate-rich drinking water with a sand filtration post-treatment. *Water S Afr*, 31(2), 229-236. doi:10.4314/wsa.v31i2.5177
- Dodds, W., & Smith, V. (2016). Nitrogen, phosphorus, and eutrophication in streams. *Inland Waters*, 6(2), 155-164. doi:10.5268/iw-6.2.909
- Dodds, W. K. K., & Welch, E. B. (2000). Establishing nutrient criteria in streams. *Journal of the North American Benthological Society*, 19(1), 186-196.
- Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant System: Optimization by a Colony of Cooperating Agents. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART B: CYBERNETICS*, 26(1). doi:<https://doi.org/10.1109/3477.484436>
- Ebrahimian, M., Nuruddin, A. A., Soom, M. A. M., Sood, A. M., Neng, L. J., & Galavi, H. (2016). Trend analysis of major hydroclimatic variables in the Langat River basin, Malaysia. *Singapore Journal of Tropical Geography*. doi:10.1111/sjtg.12234
- Ebrahimpour, N., Gharehchopogh, F. S., & Khalifehlou, Z. A. (2016). New Approach with Hybrid of Artificial Neural Network and Ant Colony Optimization in Software Cost Estimation. *Journal of Advances in Computer Research*, 7(3), 1-12.
- Ehteram, M., Karami, H., Mousavi, S.-F., El-Shafie, A., & Amini, Z. (2017). Optimizing dam and reservoirs operation based model utilizing shark algorithm approach. *Knowledge-Based Systems*, 122, 26-38. doi:<https://doi.org/10.1016/j.knosys.2017.01.026>
- Ehtram, M., Karami, H., Mousavi, S.-F., El-Shafie, A., & Amini, Z. (2017). Optimizing Dam and Reservoirs Operation Based Model Utilizing Shark Algorithm Approach. *Knowledge-Based Systems*. doi:10.1016/j.knosys.2017.01.026
- El Shafie, A. H., El-Shafie, A., Almkhtar, A., Taha, M. R., El Mazoghi, H. G., & Shehata, A. (2012). Radial basis function neural networks for reliably forecasting rainfall. *Journal of Water and Climate Change*, 3(2), 125-138. doi:10.2166/wcc.2012.017
- El-Shafie, A., Noureldin, A., Taha, M., Hussain, A., & Mukhlisin, M. (2012). Dynamic versus static neural network model for rainfall forecasting at Klang River Basin, Malaysia. *Hydrol. Earth Syst. Sci.*, 1151-1169. doi:10.5194/hess-16-1151-2012
- El-Shafie, A. H., El-Shafie, A., El Mazoghi, H. G., Shehata, A., & Taha, M. R. (2011). Artificial neural network technique for rainfall forecasting applied to Alexandria, Egypt. *International Journal of Physical Sciences*, 6(6), 1306-1316.
- Elfithri, R., Mokhtar, M. B., Abdullah, M. P., & Toriman, M. E. (2017). Restoring and Managing Langat River Basin, Malaysia: Challenges for a Sustainable Future. *International Journal of Environment and Sustainability [IJES]*, 6(4), 1-10. doi:10.24102/ijes.v6i4.897
- ElSaid, A., Benson, S., Patwardhan, S., Stadem, D., & Desell, T. (2019). *Evolving recurrent neural networks for time series data prediction of coal plant parameters*. Paper presented at the International Conference on the Applications of Evolutionary Computation.

- ElSaid, A., Ororbia, A. G., & Desell, T. J. (2020). *Ant-based Neural Topology Search (ANTS) for Optimizing Recurrent Networks*. Paper presented at the International Conference on the Applications of Evolutionary Computation.
- Farid, A. M., Lubna, A., Choo, T. G., Rahim, M. C., & Mazlin, M. (2016). A Review on the Chemical Pollution of Langat River, Malaysia. *Asian Journal of Water, Environment and Pollution*, 13(1), 9-15. doi:10.3233/ajw-160002
- Farzad, F., & El-Shafie, A. H. (2017). Performance Enhancement of Rainfall Pattern – Water Level Prediction Model Utilizing Self-Organizing-Map Clustering Method. *Water Resources Management*, 31(3), 945-959. doi:10.1007/s11269-016-1556-7
- Fewtrell, L. (2004). Drinking-water nitrate, methemoglobinemia, and global burden of disease: a discussion. *Environmental Health Perspect*, 112(14), 1371-1374. doi:10.1289/ehp.7216
- Fiyadh, S. S., AlSaadi, M. A., Jaafar, W. Z., AlOmar, M. K., Fayaed, S. S., Mohd, N. S., . . . El-Shafie, A. (2019). Review on heavy metal adsorption processes by carbon nanotubes. *Journal of Cleaner Production*, 783-793. doi:10.1016/j.jclepro.2019.05.154
- Fogelman, S., Blumenstein, M., & Zhao, H. (2005). Estimation of chemical oxygen demand by ultraviolet spectroscopic profiling and artificial neural networks. *Neural Computing and Applications*, 15(3-4), 197-203. doi:10.1007/s00521-005-0015-9
- Gallo, E. L., Meixner, T., Aoubid, H., Lohse, K. A., & Brooks, P. D. (2015). Combined impact of catchment size, land cover, and precipitation on streamflow and total dissolved nitrogen: A global comparative analysis. *Global Biogeochemical Cycles*, 29(7), 1109-1121. doi:10.1002/2015gb005154
- Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H., & Ramli, M. F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Mar Pollut Bull*, 64(11), 2409-2420. doi:10.1016/j.marpolbul.2012.08.005
- Gulis, G., Czompolyova, M., & R Cerhan, J. (2002). An Ecologic Study of Nitrate in Municipal Drinking Water and Cancer Incidence in Trnava District, Slovakia. *Environmental Research*, 88(3), 182-187. doi:10.1006/enrs.2002.4331
- Gündüz, O. (2015). Water Quality Perspectives in a Changing World. *Water Quality, Exposure and Health*, 7(1), 1-3. doi:10.1007/s12403-015-0161-y
- Hamed, Y., Awad, S., & Ben Sâad, A. (2013). Nitrate contamination in groundwater in the Sidi Aïch–Gafsa oases region, Southern Tunisia. *Environmental Earth Sciences*, 70(5), 2335-2348. doi:10.1007/s12665-013-2445-5
- He, B., Oki, T., Sun, F., Komori, D., Kanae, S., Wang, Y., . . . Yamazaki, D. (2011). Estimating monthly total nitrogen concentration in streams by using artificial neural network. *J Environ Manage*, 92(1), 172-177. doi:10.1016/j.jenvman.2010.09.014
- He, B., & Takase, K. (2006). Application of the Artificial Neural Network Method to Estimate the Missing Hydrologic Data. *J. Japan Soc. Hydrol. & Water Resour.*, 19(4), 249-257.
- He, S., & Wu, J. (2019). Hydrogeochemical Characteristics, Groundwater Quality, and Health Risks from Hexavalent Chromium and Nitrate in Groundwater of Huanhe Formation in Wuqi County, Northwest China. *Exposure and Health*, 11(2), 125-137. doi:10.1007/s12403-018-0289-7
- Hessen, D. O., Hindar, A., & Holtan, G. (1997). The Significance of Nitrogen Runoff for Eutrophication of Freshwater and Marine Recipients. *Royal Swedish Academy of Science*, 26(5), 312-320.
- Hessong, A. (2019). The Composition of Fertilizers. *Home Guides | SF Gate*. Retrieved from <http://homeguides.sfgate.com/composition-fertilizers-48898.html>
- Hingrajiya, K. H., Gupta, R. K., & Chandel, G. S. (2012). An Ant Colony Optimization Algorithm for Solving Travelling Salesman Problem. *International Journal of Scientific and Research Publications*, 2(8).
- Holmberg, M., Forsius, M., Starr, M., & Huttunen, M. (2006). An application of artificial neural networks to carbon, nitrogen and phosphorus concentrations in three boreal streams

- and impacts of climate change. *Ecological Modelling*, 195(1-2), 51-60. doi:10.1016/j.ecolmodel.2005.11.009
- Hossain, F., Chang, N.-B., Wanielista, M., Xuan, Z., & Daranpob, A. (2010). Nitrification and Denitrification in a Passive On-site Wastewater Treatment System with a Recirculation Filtration Tank. *Water Quality, Exposure and Health*, 2(1), 31-46. doi:10.1007/s12403-010-0022-7
- Indah Water, M. (2019). Ammonia. Retrieved from <https://www.iwk.com.my/do-you-know/ammonia>
- J, U., Metawa, N., Shankar, K., & Lakshmanprabu, S. K. (2018). Financial crisis prediction model using ant colony optimization. *International Journal of Information Management*. doi:10.1016/j.ijinfomgt.2018.12.001
- Jacobs, S. R., Breuer, L., Butterbach-Bahl, K., Pelster, D. E., & Rufino, M. C. (2017). Land use affects total dissolved nitrogen and nitrate concentrations in tropical montane streams in Kenya. *Sci Total Environ*, 603-604, 519-532. doi:10.1016/j.scitotenv.2017.06.100
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial Neural Networks: A Tutorial. *Computer*, 29, 31-44. doi:arXiv:2004.14329
- Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. I., Armi, A. S., Toriman, M. E., & Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environ Monit Assess*, 173(1-4), 625-641. doi:10.1007/s10661-010-1411-x
- Juang, C.-F., & Yeh, Y.-T. (2017). Multiobjective Evolution of Biped Robot Gaits Using Advanced Continuous Ant-Colony Optimized Recurrent Neural Networks. *IEEE TRANSACTIONS ON CYBERNETICS*. doi:10.1109/TCYB.2017.2718037
- Juniora, L. S., Nedjahb, N., & Mourelle, L. d. M. (2013). Routing for applications in NoC using ACO-based algorithms. *Applied Soft Computing*, 13, 2224-2231. doi:10.1016/j.asoc.2013.01.009
- Kalinli, A., Acar, M. C., & Gündüz, Z. (2010). New approaches to determine the ultimate bearing capacity of shallow foundations based on artificial neural networks and ant colony optimization. *Engineering Geology*, 117, 29-38. doi:10.1016/j.enggeo.2010.10.002
- Kannel, P. R., Lee, S., Lee, Y. S., Kanel, S. R., & Pelletier, G. J. (2007). Application of automated QUAL2Kw for water quality modeling and management in the Bagmati River, Nepal. *Ecological Modelling*, 202(3-4), 503-517. doi:10.1016/j.ecolmodel.2006.12.033
- Khalil, B., Ouarda, T. B. M. J., & St-Hilaire, A. (2011). Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *Journal of Hydrology*, 405(3-4), 277-287. doi:10.1016/j.jhydrol.2011.05.024
- Kilonzo, F., Masese, F. O., Van Griensven, A., Bauwens, W., Obando, J., & Lens, P. N. L. (2014). Spatial-temporal variability in water quality and macro-invertebrate assemblages in the Upper Mara River basin, Kenya. *Physics and Chemistry of the Earth, Parts A/B/C*, 67-69, 93-104. doi:10.1016/j.pce.2013.10.006
- Kucukkoc, I., & Zhang, D. Z. (2013). *On Applications of Ant Colony Optimisation Techniques in Solving Assembly Line Balancing Problems*. Paper presented at the OR55 Annual Conference, University of Exeter.
- Kumar, P., Lai, S. H., Mohd, N. S., Kamal, M. R., Afan, H. A., Ahmed, A. N., . . . El-Shafie, A. (2020). Optimised neural network model for river-nitrogen prediction utilizing a new training approach. *PLoS One*, 15(9), e0239509. doi:10.1371/journal.pone.0239509
- Kumar, P., Lai, S. H., Wong, J. K., Mohd, N. S., Kamal, M. R., Afan, H. A., . . . El-Shafie, A. (2020). Review of Nitrogen Compounds Prediction in Water Bodies Using Artificial Neural Networks and Other Models. *Sustainability*, 12(11), 4359. doi:10.3390/su12114359
- Lagos-Avid, M. P., & Bonilla, C. A. (2017). Predicting the particle size distribution of eroded sediment using artificial neural networks. *Sci Total Environ*, 581-582, 833-839. doi:10.1016/j.scitotenv.2017.01.020

- Land and Plant Nutrition Management Service, & Land and Water Development Division. (2004). Fertilizer use by crop in Malaysia. Retrieved from <http://www.fao.org/3/y5797e/y5797e00.htm#Contents>
- Lek, S., Guisresse, M., & Giraudel, J.-L. (1999). Predicting stream nitrogen concentration from watershed features using neural networks. *Water Resources Research*, 33(16), 3469-3478.
- Li, C., Zhu, L., He, Z., Gao, H., Yang, Y., Yao, D., & Qu, X. (2019). Runoff Prediction Method Based on Adaptive Elman Neural Network. *Water (MDPI)*, 11, 1113. doi:10.3390/w11061113
- Li, S., Cheng, X., Xu, Z., Han, H., & Zhang, Q. (2009). Spatial and temporal patterns of the water quality in the Danjiangkou Reservoir, China. *Hydrological Sciences Journal*, 54(1), 124-134. doi:10.1623/hysj.54.1.124
- Li, W., Li, X., Su, J., & Zhao, H. (2014). Sources and mass fluxes of the main contaminants in a heavily polluted and modified river of the North China Plain. *Environ Sci Pollut Res Int*, 21(8), 5678-5688. doi:10.1007/s11356-013-2461-8
- Lu, H., Li, H., Liu, T., Fan, Y., Yuan, Y., Xie, M., & Qian, X. (2019). Simulating heavy metal concentrations in an aquatic environment using artificial intelligence models and physicochemical indexes. *Sci Total Environ*, 694, 133591. doi:10.1016/j.scitotenv.2019.133591
- Mahdaviani, K., Mazyar, H., Majidi, S., & Saraee, M. H. (2008). *A Method to Resolve the Overfitting Problem in Recurrent Neural Networks for Prediction of Complex Systems' Behavior*. Paper presented at the International Joint Conference on Neural Networks (IJCNN 2008).
- Maloney, K. O., & Weller, D. E. (2011). Anthropogenic disturbance and streams: land use and land-use change affect stream ecosystems via multiple pathways. *Freshwater Biology*, 56(3), 611-626. doi:10.1111/j.1365-2427.2010.02522.x
- Markus, M., Hejazi, M. I., Bajcsy, P., Giustolisi, O., & Savic, D. A. (2010). Prediction of weekly nitrate-N fluctuations in a small agricultural watershed in Illinois. *Journal of Hydroinformatics*, 12(3), 251-261. doi:10.2166/hydro.2010.064
- Markus, M., Tsai, C. W.-S., & Demissie, M. (2003). Uncertainty of Weekly Nitrate-Nitrogen Forecasts Using Artificial Neural Networks. *Journal of Environmental Engineering*, 129(3), 267-274. doi:10.1061/(ASCE)0733-9372(2003)129:3(267)
- Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & López-Cózar, E. D. (2020). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *arXiv*.
- Mavrovouniotis, M., & Yang, S. (2013). Evolving Neural Networks using Ant Colony Optimization with Pheromone Trail Limits. *IEEE*, 13.
- May, D. B., & Sivakumar, M. (2009). Prediction of urban stormwater quality using artificial neural networks. *Environmental Modelling & Software*, 24(2), 296-302. doi:10.1016/j.envsoft.2008.07.004
- Mazvimavi, D., Meijerink, A. M. J., Savenije, H. H. G., & Stein, A. (2005). Prediction of flow characteristics using multiple regression and neural networks: A case study in Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 30(11-16), 639-647. doi:10.1016/j.pce.2005.08.003
- Murdoch, P. S., & Stoddard, J. L. (1992). The Role of Nitrate in the Acidification of Streams in the Catskill Mountains of New York. *Water Resources Research*, 28(10), 2707-2720.
- Najah, A., El-Shafie, A., Karim, O. A., & Jaafar, O. (2011). Integrated versus isolated scenario for prediction dissolved oxygen at progression of water quality monitoring stations. *Hydrological Earth Syst. Sci.*, 15, 2693-2708. doi:10.5194/hess-15-2693-2011
- New Straits Times. (2017). Another Johor water treatment plant shuts down over ammonia pollution. Retrieved from <https://www.nst.com.my/news/nation/2017/11/304914/update-another-johor-water-treatment-plant-shuts-down-over-ammonia>

- Nuruzzaman, M., Mamun, A. A., & Salleh, M. N. B. (2017). Determining ammonia nitrogen decay rate of Malaysian river water in a laboratory flume. *International Journal of Environmental Science and Technology*, 15(6), 1249-1256. doi:10.1007/s13762-017-1482-0
- Palani, S., Liong, S. Y., & Tkalich, P. (2008). An ANN application for water quality forecasting. *Mar Pollut Bull*, 56(9), 1586-1597. doi:10.1016/j.marpolbul.2008.05.021
- Pernet-Coudrier, B., Qi, W., Liu, H., Muller, B., & Berg, M. (2012). Sources and pathways of nutrients in the semi-arid region of Beijing-Tianjin, China. *Environ Sci Technol*, 46(10), 5294-5301. doi:10.1021/es3004415
- Qiu, Y., Shi, H.-c., & He, M. (2010). Nitrogen and Phosphorous Removal in Municipal Wastewater Treatment Plants in China: A Review. *International Journal of Chemical Engineering*, 2010, 1-10. doi:10.1155/2010/914159
- Rabalais, N. N., & Turner, R. E. (2006). Oxygen depletion in the gulf of mexico adjacent to the mississippi river. *Past and Present Water Column Anoxia; Springer*, 225-245.
- Rabalais, N. N., Turner, R. E., & Scavia, D. (2002). Beyond Science into Policy: Gulf of Mexico Hypoxia and the Mississippi River. *Bioscience*, 52, 129-142.
- Rahmati, O., Samani, A. N., Mahmoodi, N., & Mahdavi, M. (2015). Assessment of the Contribution of N-Fertilizers to Nitrate Pollution of Groundwater in Western Iran (Case Study: Ghorveh–Dehgelan Aquifer). *Water Quality, Exposure and Health*, 7(2), 143-151. doi:10.1007/s12403-014-0135-5
- Räike, A., Pietiläinen, O. P., Rekolainen, S., Kauppila, P., Pitkänen, H., Niemi, J., . . . Vuorenmaa, J. (2003). Trends of phosphorus, nitrogen and chlorophyll a concentrations in Finnish rivers and lakes in 1975–2000. *Science of The Total Environment*, 310(1-3), 47-59. doi:10.1016/s0048-9697(02)00622-8
- Raju, M. M., Srivastava, R. K., Bisht, D. C. S., Sharma, H. C., & Kumar, A. (2011). Development of Artificial Neural-Network-Based Models for the Simulation of Spring Discharge. *Advances in Artificial Intelligence*, 2011, 1-11. doi:10.1155/2011/686258
- Reddy, A. G. S., Niranjana Kumar, K., Subba Rao, D., & Sambashiva Rao, S. (2009). Assessment of nitrate contamination due to groundwater pollution in north eastern part of Anantapur District, A.P. India. *Environmental Monitoring and Assessment*, 148(1), 463-476. doi:10.1007/s10661-008-0176-y
- Rekacewicz, P. (2006). Nitrate levels: concentrations at river mouths. Retrieved from <http://www.grida.no/resources/5650>
- Riad, S., Mania, J., Bouchaou, L., & Najjar, Y. (2004). Rainfall-runoff model using an artificial neural network approach. *Mathematical and Computer Modelling*, 40(7-8), 839-846. doi:10.1016/j.mcm.2004.10.012
- Salehi, F., Prasher, S. O., Amin, S., Madani, A., Jebelli, S. J., Ramaswamy, H. S., . . . Drury, C. F. (2000). Prediction of annual nitrate-n losses in drain outflows with artificial neural networks. *American Society of Agricultural Engineers*, 43(5), 1137-1143.
- Sarang, A., & Bhattacharya, A. K. (2005). Comparison of Artificial Neural Network and regression models for sediment loss prediction from Banha watershed in India. *Agricultural Water Management*, 78(3), 195-208. doi:10.1016/j.agwat.2005.02.001
- Sawyer CN, McCarty PL, & Parkin GF. (2003). Chemistry for environmental engineering and science. In (5th ed., pp. 667). New York, NY, USA: McGraw Hill.
- Shah, H., Ghazali, R., & Nawi, N. M. (2012). *Hybrid Ant Bee Colony Algorithm for Volcano Temperature Prediction*. Paper presented at the International Multi Topic Conference.
- Shamseldin, A. Y., Nasr, A. E., & O'Connor, K. M. (2002). Comparison of different forms of the Multi-layer Feed-Forward Neural Network method used for river flow forecasting. *Hydrology and Earth System Sciences Discussions*, 6(4), 671-684.
- Sharma, V., Negi, S. C., Rudra, R. P., & Yang, S. (2003). Neural networks for predicting nitrate-nitrogen in drainage water. *Agricultural Water Management*, 63(3), 169-183. doi:10.1016/s0378-3774(03)00159-8

- Sheela, K. G., & Deepa, S. N. (2013). Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Hindawi Publishing Corporation, Mathematical Problems in Engineering*, 2013. doi:10.1155/2013/425740
- Sheikhan, M., & Mohammadi, N. (2012). Time series prediction using PSO-optimized neural network and hybrid feature selection algorithm for IEEE load data. *Neural Comput & Applic.* doi:10.1007/s00521-012-0980-8
- Singh, K. P., Basant, A., Malik, A., & Jain, G. (2009). Artificial neural network modeling of the river water quality—A case study. *Ecological Modelling*, 220(6), 888-895. doi:10.1016/j.ecolmodel.2009.01.004
- Singh, K. P., Malik, A., & Sinha, S. (2005). Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. *Analytica Chimica Acta*, 538(1-2), 355-374. doi:10.1016/j.aca.2005.02.006
- Soh, Y. W., Koo, C. H., Huang, Y. F., & Fung, K. F. (2018). Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia. *Computers and Electronics in Agriculture*, 144, 164-173. doi:10.1016/j.compag.2017.12.002
- Song, Q. (2010). On the Weight Convergence of Elman Networks. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 21(3). doi:10.1109/TNN.2009.2039226
- Sousa, S., Martins, F., Alvimferraz, M., & Pereira, M. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, 22(1), 97-103. doi:10.1016/j.envsoft.2005.12.002
- Su, X., Wang, H., & Zhang, Y. (2013). Health Risk Assessment of Nitrate Contamination in Groundwater: A Case Study of an Agricultural Area in Northeast China. *Water Resources Management*, 27(8), 3025-3034. doi:10.1007/s11269-013-0330-3
- Suen, J.-P., & Eheart, J. W. (2003). Evaluation of Neural Networks for Modeling Nitrate Concentrations in Rivers. *Journal of Water Resources Planning and Management, ASCE*, 129(6), 505-510. doi:10.1061//ASCE/0733-9496/2003/129:6/505
- Suo, W. Q., Dong-Bao, S., Wei-Ping, H., Yu-Zhong, L., Xu-Rong, M., & Yan-Qing, Z. (2012). Human activities and nitrogen in waters. *Acta Ecologica Sinica*, 32(4), 174-179. doi:10.1016/j.chnaes.2012.04.010
- Tampelini, L. G., Boscarioli, C., Peres, S. M., & Sampaio, S. C. s. (2011). AN APPLICATION OF ELMAN NETWORKS IN TREATMENT AND PREDICTION OF HYDROLOGIC TIME SERIES. *Journal of the Brazilian Neural Network Society*, 9(3), 148-156. doi:10.21528/LNLM-vol9-no3-art1
- Tayfur, G., Swiatek, D., Wita, A., & Singh, V. P. (2005). Case Study: Finite Element Method and Artificial Neural Network Models for Flow through Jeziorsko Earthfill Dam in Poland. *Journal of Hydraulic Engineering*, 131(6), 431-440. doi:10.1061//ASCE/0733-9429/2005/131:6/431
- Tehrani, R., & Khodayar, F. (2010). Optimization of the Artificial Neural Networks Using Ant Colony Algorithm to Predict the Variation of Stock Price Index. *Journal of Applied Sciences*, 10(3), 221-225.
- Teschl, R., & Randeu, W. L. (2006). A neural network model for short term river flow prediction. *Natural Hazards and Earth System Sciences*, 6, 629-635.
- The Star. (2019). Five water treatment plants shut down due to ammonia pollution fully operational. Retrieved from <https://www.thestar.com.my/news/nation/2019/04/06/five-water-treatment-plants-shut-down-due-to-ammonia-pollution-fully-operational/>
- USGS. (2018). Nitrogen and Water. Retrieved from https://www.usgs.gov/special-topic/water-science-school/science/nitrogen-and-water?qt-science_center_objects=0#qt-science_center_objects

- Voyant, C., Muselli, M., Paoli, C., & Nivet, M.-L. (2012). Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. *Energy*, 341-355. doi:10.1016/j.energy.2012.01.006
- Wang, B., Oldham, C., & Hipse, M. R. (2016). Comparison of Machine Learning Techniques and Variables for Groundwater Dissolved Organic Nitrogen Prediction in an Urban Area. *Procedia Engineering*, 154, 1176-1184. doi:10.1016/j.proeng.2016.07.527
- Ward, M. H., deKok, T. M., Levallois, P., Brender, J., Gulis, G., Nolan, B. T., . . . International Society for Environmental, E. (2005). Workgroup report: Drinking-water nitrate and health--recent findings and research needs. *Environmental Health Perspect*, 113(11), 1607-1614. doi:10.1289/ehp.8043
- Wu, C. L., & Chau, K. W. (2011). Rainfall-runoff modeling using artificial neural network coupled with singular spectrum analysis. *Journal of Hydrology*, 399(3-4), 394-409. doi:10.1016/j.jhydrol.2011.01.017
- Xie, Y. (2011). Values and Limitations of Statistical Models. *Res Soc Stratif Mobil*, 29(3), 343-349. doi:10.1016/j.rssm.2011.04.001
- Yan, W., Mayorga, E., Li, X., Seitzinger, S. P., & Bouwman, A. F. (2010). Increasing anthropogenic nitrogen inputs and riverine DIN exports from the Changjiang River basin under changing human pressures. *Global Biogeochemical Cycles*, 24(4), n/a-n/a. doi:10.1029/2009gb003575
- Yi, Q., Chen, Q., Hu, L., & Shi, W. (2017). Tracking nitrogen sources, transformation and transport at a basin scale with complex plain river networks. *Environmental Science & Technology*, 51, 5396-5403.
- Zaheer, I., & Bai, C.-G. (2003). Application of artificial neural network for water quality management. *Lowland Technology International*, 5(2), 10-15.
- Zeľňáková, M., Āarnogurská, M., Šlezinger, M., & Stýš, D. (2012). Model based on dimensional analysis for prediction of nitrogen and phosphorus concentration in the River Laborec. *Hydrology and Earth System Sciences Discussions*, 9(4), 5611-5634. doi:10.5194/hessd-9-5611-2012
- Zhang, L., Song, X., Xia, J., Yuan, R., Zhang, Y., Liu, X., & Han, D. (2011). Major element chemistry of the Huai River basin, China. *Applied Geochemistry*, 26(3), 293-300. doi:10.1016/j.apgeochem.2010.12.002
- Zhang, L., Zhang, G. X., & Li, R. R. (2016). Water Quality Analysis and Prediction Using Hybrid Time Series and Neural Network Models. *J. Agr. Sci. Tech.*, 18, 975-983.
- Zhang, W. S., Swaney, D. P., Li, X. Y., Hong, B., Howarth, R. W., & Ding, S. H. (2015). Anthropogenic point-source and non-point-source nitrogen inputs into Huai River basin and their impacts on riverine ammonia-nitrogen flux. *Biogeosciences*, 12(14), 4275-4289. doi:10.5194/bg-12-4275-2015
- Zheng-da, Z., Ya-ming, S., & Zhi-sheng, Z. (2005). Short-term load forecasting based on recurrent neural network using ant colony optimization algorithm. *Power System Technology*, 3, 59-63.
- Zomorodian, M., Lai, S. H., Homayounfar, M., Ibrahim, S., & Pender, G. (2017). Development and application of coupled system dynamics and game theory: A dynamic water conflict resolution method. *PLoS One*, 12(12), e0188489. doi:10.1371/journal.pone.0188489