WATER LEVEL FORECASTING MODEL USING IMPROVED ARTIFICIAL NEURAL NETWORK ARCHITECTURE

MUHAMMAD @ S A KHUSHREN BIN SULAIMAN

THESIS SUBMITTED IN FULFILLMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF ENGINEERING AND BUILT ENVIRONMENT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2012

MODEL PERAMALAN PARAS AIR MENGGUNAKAN ARKITEKTUR RANGKAIAN NEURAL BUATAN YANG DI TINGKAT BAIK

ABSTRAK

Membangunkan aplikasi ramalan aras air boleh membantu pengurusan sumber air yang cekap dan optimum serta dapat meminimumkan kerosakan akibat banjir. Pada masa ini, rangkaian neural tiruan atau *artificial neural network* (ANN) telah berjaya diuji dalam banyak kajian ramalan, termasuk aliran sungai. Walau bagaimanapun, ketepatan dan kebolehpercayaan ramalan aliran sungai berasaskan ANN memerlukan penyelidikan yang berterusan. Dalam konteks tersebut, kajian ini adalah bertujuan untuk membangunkan model ramalan yang berketepatan tinggi dan yang boleh dipercayai berasaskan model ANN untuk meramal aras air khususnya untuk kes aras air yang tinggi. Tujuan ini dicapai dengan memperkenalkan empat pendekatan baru dalam permodelan berasaskan ANN. Pertama, arkitektur dalaman ANN dipertingkatkan dengan penggunaan pekali kecuraman yang optimum (OSC) dalam fungsi sigmoid. Kedua, arkitektur luaran ANN di tambah baik dengan menggunakan pendekatan pengezonan sepadan (ZMA) di mana data yang dipilih untuk latihan ANN adalah berdasarkan aras air yang di sasarkan dalam ramalan. Ketiga, penilaian ketepatan keputusan ramalan model di tingkatkan dengan pendekatan kejuruteraan di mana ralat dengan kadar yang dibenarkan digunakan untuk mengesahkan ketepatan keputusan ramalan tersebut. Keempat, ramalan turun/naik aras air yang betul di perkenalkan untuk menilai keupayaan ramalan dan kebolehpercayaan model ANN. Kawasan kajian adalah stesen Rantau Panjang di sungai Johor di mana data bagi setiap jam aras air yang direkodkan sejak tahun 1964 sehingga 2008 digunapakai untuk ramalan aras air bagi kes-kes ramalan aras air sehari dan beberapa jam sebelumnya. Hasil kajian ini menunjukkan bahawa penggunaan teknik OSC dan ZMA bukan sahaja meningkatkan ketepatan model ANN berbanding pendekatan ANN yang biasa dipakai tetapi juga mencapai keputusan ramalan dengan ketepatan yang tinggi. Selain itu, penilaian berdasarkan ralat yang dibenarkan serta penilaian ramalan turun/naik aras air berjaya memperkayakan penilaian ketepatan dan kebolehpercayaan model ramalan. Kesan kajian ini ialah teknik-teknik yang diperkenalkan boleh menjadi asas pendekatan yang baru dalam permodelan ramalan berasaskan ANN khususnya dalam memantau kejadian banjir.

ABSTRACT

Reliable water level forecasting can help achieve efficient and optimum use of water resources and minimize flooding damage. Currently, artificial neural network (ANN) has been successfully tested in many forecasting studies, including river flow. However, the accuracy and reliability of the river flow forecasting in such application requires continuous research. In this context, this study aims at developing a high accuracy and reliable forecasting model using the ANN to predict high water level events. The aim is achieved by introducing four new approaches in the ANN modeling. Firstly, the internal architecture of ANN is enhanced by utilization of optimal steepness coefficient (OSC) in sigmoid function. Secondly, the external architecture of ANN is improved by applying zoning matching approach (ZMA) where data training selected is based on the target water level to be forecasted. Thirdly, model evaluation of forecasting results is improved by engineering approach where allowable offset errors are used to demonstrate the accuracy of forecasting results. Lastly, the correct prediction of up/down of water level is a new evaluation model to evaluate forecasting capability and reliability of the ANN model. A case study has been applied at the Rantau Panjang station at Johor River where hourly water level data dated from 1963 to 2008 have been examined to forecast daily and several hourly intervals lead-times. The result showed that the use of the OSC and ZMA techniques had not only improved the accuracy of ANN models compared to standard approach (SA) but also achieved high accuracy forecasting results. The allowable offset errors and up/down prediction measures have enriched measures in evaluating forecasting model's accuracy and reliability. The impact of the study is that the techniques can be the basis of future development approach in ANN based data forecasting model specially in monitoring flood events.

CONTENTS

		Page
DECLARATION		ii
ACKNOWLEDGMENTS		iii
ABSTRAK		iv
ABSTRACT		v
CONTENTS		vi
LIST OF TABLES		ix
LIST OF ILLUSTRATIONS		x
LIST OF ABBREVIATIONS		xii
CHAPTER I	**INTRODUCTION**	
1.1	Introduction	1
1.2	Motivation	2
1.3	Research Questions	4
1.4	Research Aims & Objectives	5
1.5	Scope of the Research	6
1.6	Structure of Thesis	6
CHAPTER II	**LITERATURE REVIEW**	
2.1	Introduction	8
2.2	Importance of ANN	9
2.3	Accuracy and Reliability Studies in ANN	10
2.4	Conclusion	16
CHAPTER III	**METHODOLOGY**	
3.1	Introduction	18
3.2	Artificial Neural Network (ANN) as Nonlinear Solver	20
3.3	Optimal Steepness Coefficient (OSC)	24
3.4	Zoning Matching Approach (ZMA)	26
3.5	Performance Measures	28
	3.6.1 Common Statistical Performance Measures	28

	3.6.2	Range of Allowable Errors Measure	29
	3.6.3	Up/Down Prediction Measure	30
3.6	The ANN Computer Program		31
3.7	Set-up No.1		43
	3.7.1	Introduction	43
	3.7.2	Implementation	44
3.8	Set-up No.2		46
	3.9.1	Introduction	46
	3.9.2	Implementation	47
	a)	Data and Datasets	47
	b)	Stages in the Analysis	48
3.9	Set-up No.3		51
	3.10.1	Introduction	51
	3.10.2	Implementation	52
	a)	First stage	53
	b)	Second stage	57
	c)	Third stage	59
	d)	Fourth stage	61
3.10	Additional Set-up		61
	3.11.1	Introduction	61
	3.11.2	Implementation	62
CHAPTER IV	**CASE STUDY AREA**		
4.1	Background		63
4.2	The Relevance of Rantau Panjang Station as a Case Study Area		65
CHAPTER V	**RESULTS AND DISCUSSION**		
5.1	Set-up No.1		68
5.2	Set-up No.2		80
5.3	Set-up No.3		96
5.4	Additional Set-Up		112
CHAPTER VI	**CONCLUSION**		
6.1	Introduction		115
	6.1.1	Conclusion Set-up No.1	115
	6.1.2	Conclusion Set-up No.2	116

	6.1.3 Conclusion Set-up No.3	116
	6.1.4 Conclusion Additional Set-up	117
6.2	Conclusion of the study.	118
6.3	Contribution of the Research	118
6.4	Recommendation for Future Works	121

REFERENCES 123

APPENDICES

A	Source Code - ANN Data Normalization	131
B	Source Code - Sigmoid Functions in ANN	132
C	Source Code - ANN Data Forecasting	133
D	Source Code - ANN Performance Measure	135
E	Snapshot of ANN program user interface in the study	136
F	Snapshot of ANN sample datasets in the study	137
G	List of Publications	139

LIST OF TABLES

Table No.		Page
3.1	Steepness coefficients in sigmoid function in the study.	25
3.2	ANN architectures in the set-up no.1.	45
3.3	Datasets in the set-up no.1.	46
3.4	Network models in the set-up no.2.	50
4.1	Flooding records in Kota Tinggi and Rantau Panjang ares.	65
5.1	Data training performances (CE) of the six training datasets.	73
5.2	Forecasting performances (CE) of dataset DS_7.	73
5.3	Forecasting performances for the five SA network models.	82
5.4	Forecasting performances for the four ZMA models and NET_5 for water levels greater than 8000 mm.	85
5.5	Forecasting performances for ZMA_9 and NET_5 for 1-hr to 6-hr intervals.	87
5.6	Forecasting performance at 6-hr ahead based on SA with 6-hr interval data inputs.	97
5.7	Forecasting performance at 6-hr ahead for water level higher than 8000 mm based on ZMA and 6-hr interval data inputs.	99
5.8	Forecasting performance at 12-hr ahead for water level higher than 8000 mm based on ZMA and 12-hr interval data inputs.	102
5.9	Forecasting performance at 12-hr ahead for water level higher than 8000 mm based on 6-hr interval data inputs.	104
5.10	Partial results of 6-hr ahead up/down predictions based on ZMA_8 and NET_4.	106
5.11	Partial results of 12-hr ahead up/down predictions based on ZMA_8 and NET_4.	107
5.12	Forecasting performances at 3-hr ahead based on ZMA_9 and NET_5 for SSC and MSC_8.	113
5.13	Forecasting performances at 6-hr ahead based on ZMA_8 and NET_4 for SSC and MSC_8.	113
5.14	Forecasting performances at 12-hr ahead at 6-hr interval with multi-lead approach based on ZMA_8 and NET_4 for SSC and MSC_8.	113
5.15	Partial results of 3-hr ahead up/down predictions based on ZMA_9 and NET_5.	114

LIST OF ILLUSTRATIONS

Figure No.		Page
3.1	MLP-BP neural network architecture in the study.	21
3.2	An active ANN neuron with sigmoid function and threshold input.	22
3.3	Sigmoid function with 10 steepness coefficients.	25
3.4	Zoning matching approach model.	27
3.5	The data and steps in data training process.	33
3.6	The data and steps in feed-forward module (FFM).	35
3.7	The data and steps in back-propagation module (BPM).	38
3.8	The data and steps in model performance module (MPM).	39
3.9	The data and steps in data forecasting process.	41
3.10	The data and steps in performance evaluation process.	42
3.11	Time series data model inputs.	45
3.12	The stages in developing of forecasting models based on ZMA.	49
3.13	Stage 1 - steps in development of ANN model based on SA.	54
3.14	Stage 2 – steps in development of multi-lead ANN model based on ZMA	58
3.15	Multi-lead water level forecasting model.	59
3.16	Stage 3 - steps in development of multi-lead ANN model based on ZMA.	60
4.1	Johor River basin.	64
4.2	Water level elevation status at Rantau Panjang station.	64
4.3	Historical daily water level data at Rantau Panjang station	67
5.1	Data training performances for the six datasets with five ANN architectures and ten steepness coefficients.	69
5.2	Enlargement of data training performance for Net5.	70
5.3	Comparison between data training with SSC and optimal MSC.	71
5.4	Forecasting performance of ANN model with DS4, Net1 and SSC	75
5.5	Forecasting performance of ANN model with DS4, Net5 and MSC_9	76
5.6	Forecasting performance of ANN model with DS6, Net5 and SSC	77

5.7	Forecasting performance of ANN model with DS6, Net5 and MSC_9	78
5.8	Weight of bias connected to output neuron (a) Using SSC (b) Using MSC_9	80
5.9	Comparisons between observations and model outputs based on ZMA_7 to ZMA_10 for 3-hr intervals using 5 inputs network model.	86
5.10	Forecasting results at 3-hr intervals for water levels above 8 m using ZMA_9 and NET_5.	88
5.11	Forecasting results at 6-hr intervals for water levels above 8 m using ZMA_9 and NET_5.	89
5.12	Comparisons between observations and model outputs for 3-hr intervals during 2006–2007 flood events.	90
5.13	Comparisons between observations and model outputs for 6-hr intervals during 2006–2007 flood events.	91
5.14	Distribution of errors for forecasting results (a) SA at 3-hr intervals (b) ZMA_9 at 3-hr intervals (c) ZMA_9 at 6-hr intervals.	93
5.15	Comparisons between observations and model outputs for 3-hr intervals during flood events for training dataset (a) Dec-1969 (b) Nov-1979 (c) Dec-1983 (d) Dec-1989.	95
5.16	Six hourly interval validation results at 6-hr ahead based on NET_4 with ZMA_7, ZMA_8, and ZMA_9, respectively.	100
5.17	Validation results: (a) 6-hr intervals at 6-hr ahead based on NET_3 with DS_STD; (b) 6-hr intervals at 6-hr ahead based on NET_4 with ZMA_8; (c) 6-hr intervals at 12-hr ahead based on NET_4 with ZMA_8.	109
5.18	Training performance for 6-hr ahead at 6-hr intervals based on NET_4 with ZMA_8.	110
5.19	Error distributions for validation dataset: (a) 6-hr intervals at 6-hr ahead based on NET_3 with DS_STD; (b) 6-hr intervals at 6-hr ahead based on NET_4 with ZMA_8; (c) 6-hr intervals at 12-hr ahead based on NET_4 with ZMA_8	111

LIST OF ABBREVIATION

AI	Artificial Intelligence
ANN	Artificial Neural Network
ARMA	Auto-Regressive Moving Average
ATF	Activation Transfer Function
BPM	Back-Propagation Module
CE	Coefficient Efficiency
CI	Confidence Interval
DDM	Data De-normalization Module
DNM	Data Normalization Module
FFM	Feed-Forward Module
MLP-BP	Multi-Layer Perceptron Back-Propagation
MPM	Model Performance Module
MSC	Milder Steepness Coefficient
OSC	Optimal Steepness Coefficient
R2	Correlation Coefficient
RMSE	Root Mean Squared Error
SA	Standard Approach
SC	Steepness Coefficient
SI	Scatter Index
SSC	Standard Steepness Coefficient
VPM	Validation performance module

CHAPTER I

INTRODUCTION

1.1 INTRODUCTION

Water is an essential resource to sustain the lives in this planet. This is described in the Al-Quran about 1400 years ago in chapter 23 verse 30 - "We made from water every living thing. Will they not then believe?" (Ali 2001). Water defines the landscape of a region and it is important to the earth eco-system. Water covers 70% of the earth surface and it is continously moving below, above and on the surface of the earth. This movement of water is called the water cycle. The ocean locates where ninety percent of water on earth is and eighty percent of global evaporation occurs. Water stores in the atmosphere in a form of clouds and in the process of precipitation, water falls on the land and on the sea. Most water that fall on the land will either become surface water such as lakes and streams and others infiltrate into the ground that become ground water. Evaporation also occurs on the surface water. The water that moves to the land is either cycle back to the ocean or evaporate to the atmospere. The balance of the water on earth has been constant although the water changes to different states and moves to different locations over time.

The scarcity of water in the current time underscores the needs for better water management. The water scarcity even occurs in the areas where there is plenty of rainfall or freshwater. This could be due to poor management of the water resources. The situation is getting worse as the needs for water rise along with population growth, urbanization, and increases in household and industrial uses. Some countries already have made an attempt to find water in other planets as solution to the problem of water scarcity. One of the main sources of water is the river and effective management of the river resources has become very important to resolve or to minimize the effect of water scarcity problem.

Another concern regarding water resources is flooding. Historically, flooding events have been a major concern over many generations due to the loss of lives and damage to properties caused by flooding. Flooding events normally occur when there is a heavy rainfall over a short period of time or when there is an extensive rainfall over a long period of time in flood plain areas. Besides rainfall, poor maintenance of drainage networks, deforestation and inappropriate development in the area are the human factors contributing to the causes of flooding. These human factors are preventive which means that it should be handled accordingly to minimize the effects of flooding. The common area of the flooding events occur in the coastal regions and low level areas where a high tide can slow or stop the water flow from river stream to the sea. Nowadays, the city is also a common area of flooding which could be due to deforestation, inappropriate development of the area, and poor maintenance of drainage networks. Global warming is another big factor to the causes of flooding events where excessive rainfall is recorded in many countries. Many current flooding events have recorded high in the past hundred years and the excessive flooding events occur more frequently.

1.2 MOTIVATION

To manage the river resources effectively, river flow forecasting is essential because it can help facilitate the management of water resources, thereby optimizing the use of water. The ability to forecast flooding events is hugely important, since it can help predict the occurrence of future flooding, to enable better preparation of flood events which is to avoid the loss of lives and to minimize the damages to property.

One of the approaches for developing modeling river flow and flood forecasting is based on a hydraulic design using hydrological parameters that affect river flow. This means that many physical factors of a river, such as the cross sectional area, soil types along the banks, sedimentation, elevation, geographical area, and rainfall need to be determined. This approach is not only complex, but also requires data coordination in real-time, which can be very costly and complex to engineers and water authorities. Other approaches, such as regression, have failed because the river flow patterns are neither constant, nor can they be predicted by fixed regression lines.

Currently, artificial neural network (ANN) based models have been successfully applied in many river flow forecasting studies. Extensive reviews on ANN applications in hydrological simulation and forecasting have been reported in ASCE (2000a,b). An ANN-based model does not presuppose a detailed understanding of the river's physical characteristics, nor does it require an extensive data of pre-processing (Dawson et al. 2002). Instead, the ANN uses a sample or historical dataset of river flow to establish the forecast model. The distinct function of an ANN is that it can find hidden patterns in the sample dataset, the knowledge from which is used in the data forecasting. The first neural model was introduced by McCulloch & Pitts (1943) to simulate neurons in the human brain. The use of error propagation and computer to simulate the neural processes (Rumelhart & McClelland 1986) has popularized the use of ANN in many forecasting and data classification studies in recent years. ANNs have been used in many studies, including engineering, science, mathematics, social science, and business studies, for data classification, data forecasting, data mining, robotics, and data processing. In water resources, ANNs have been applied in forecasting river flow, water levels, flood events, hydrograph, sedimentation, and water quality.

There is little doubt that ANNs have the potential to be a useful tool for the prediction and forecasting of water resources variables (Maier & Dandy 2000). The primary focus should be on achieving good results, rather than statistical optimality (Maier & Dandy 2000). Future research efforts should also be directed towards the extraction of the knowledge that is contained in the connection weights of trained ANN models (Maier & Dandy 2000). The scope for improving the way the ANN models are evaluated by applying various measures in a consistent and informed manner should be given attention (Dawson et al. 2007). Not the least is an attention to good practice in model development as it is vitally important in all modeling efforts (Jakeman et al. 2006; Robson et al. 2008; Welsh 2008).

1.3 RESEARCH QUESTIONS

Various approaches have been developed in the ANN based modeling studies to improve the basic modeling capability of ANN so that high forecasting accuracy can be achieved with suitable lead-time. The study of improving forecasting accuracy of river flow is a continuous research that contributes positively to the science society and the public. Based on the modeling approaches reviewed, there seems always computation and/or modeling approaches that can always be explored to improve forecasting accuracy. The question is can the ANN model still be improved by working on new approach in the way data are modeled in ANN or the way the computing algorithm or parameters are used?

Many data forecasting studies for extreme events such as monitoring flood events still lag of accuracy. This could be due to the approach that one fits all approach where all data are used to forecast different sections or zones of data to be forecasted. Recently, more studies have used hybrid ANN model to utilize the self-organized maps to improve the forecasting at the extreme events. The hybrid approach is complex and time consuming since adjustment is needed to categorize the data for training and forecasting. Is there a simpler approach and practice to produce high accuracy or even better at the extreme events?

How does one assess the reliability of the ANN model? The forecasting accuracy should have boundary or confidence limit but current studies do not define this limit. Most studies have used common statistical measure and relative errors to describe the reliability of the model developed. It is important to the water authorities to determine the probability of the fail conditions of forecasting model and the range of errors in making decisive decisions. In engineering design, there is always safety factor and probability of failure so as to show the fail condition. The issue is can these factors be associated with the ANN model?

Convincing the water authorities on the capability of the ANN based model in making predictions has not been easy. Many engineers or water authorities are not familiar with ANN forecasting based on historical data and non-physical conditons of

1.3 RESEARCH QUESTIONS

Various approaches have been developed in the ANN based modeling studies to improve the basic modeling capability of ANN so that high forecasting accuracy can be achieved with suitable lead-time. The study of improving forecasting accuracy of river flow is a continuous research that contributes positively to the science society and the public. Based on the modeling approaches reviewed, there seems always computation and/or modeling approaches that can always be explored to improve forecasting accuracy. The question is can the ANN model still be improved by working on new approach in the way data are modeled in ANN or the way the computing algorithm or parameters are used?

Many data forecasting studies for extreme events such as monitoring flood events still lag of accuracy. This could be due to the approach that one fits all approach where all data are used to forecast different sections or zones of data to be forecasted. Recently, more studies have used hybrid ANN model to utilize the self-organized maps to improve the forecasting at the extreme events. The hybrid approach is complex and time consuming since adjustment is needed to categorize the data for training and forecasting. Is there a simpler approach and practice to produce high accuracy or even better at the extreme events?

How does one assess the reliability of the ANN model? The forecasting accuracy should have boundary or confidence limit but current studies do not define this limit. Most studies have used common statistical measure and relative errors to describe the reliability of the model developed. It is important to the water authorities to determine the probability of the fail conditions of forecasting model and the range of errors in making decisive decisions. In engineering design, there is always safety factor and probability of failure so as to show the fail condition. The issue is can these factors be associated with the ANN model?

Convincing the water authorities on the capability of the ANN based model in making predictions has not been easy. Many engineers or water authorities are not familiar with ANN forecasting based on historical data and non-physical conditons of river. Many forecasting results are based on fitness of fit by common statistical measure which demonstrate the accuracy value of the performance index, not the predictable capability of the ANN model. Is there a way to demonstrate that the ANN based model has the ability to make the data prediction? This is a challenge that needs to be addressed so that the developed ANN models are not only used for scientific studies but also use in practice by the water authorities and water engineers. These are the issues and questions examined in this study.

1.4 RESEARCH AIMS & OBJECTIVES

The general aim of this study is to develop high accuracy and reliable forecasting model that is based on Artificial Neural Network (ANN) to forecast flooding events in Johor River specifically Rantau Panjang station & also to elevate the confidence level in the developed model.

The general aims are achieved by the following specific aims:

1. Improve ANN internal architecture by studying the computing part of the ANN model;

2. Improve ANN external architecture by exploring different ways data are modelled in the ANN;

3. Improve ANN performance evaluation by evaluating the current approach and exploring other approaches that can verify the accuracy and reliability of the ANN model.

The success in this study hopefully can provide the water authorities the tools that are not only accurate but also reliable in the forecasting of water level specifically to monitor flood events.

1.5 SCOPE OF THE RESEARCH

The scope of this study is to improve ANN based modeling in the forecast of high water level events specifically for the case of monitoring flood events at Rantau Panjang station. The modeling tool used is a multi-layer perceptron neural network, which is the most commonly use of the ANN based model type. The lead-times tested in the foreacasting are hourly and daily.

1.6 STRUCTURE OF THESIS

The thesis is composed of six chapters. Chapter 1 introduces the motivation, research questions, aims and objectives, scope of of the study, and the structure of the thesis. Chapter 2 covers the literature review on the research topic. The literature covers the current and previous studies in river flow forecasting in the areas of extreme events, activation transfer functions that are used and studied in the ANN model, the importance of data type in the ANN model, the importance of lead-times, confidence limit and level. Chapter 3 is the methodology used in the study. This includes the modeling approach, the modeling tools, the improvement to the modeling approach and the set-ups conducted in the study. Four set-ups were conducted and studied where they evolved from one set-up to another and eventually leading to the fourth set-up. Chapter 4 describes the case study area and its relevance to the future study of flood monitoring at Kota Tinggi, Johor. Chapter 5 is the ANN performance evaluation on data training and forecasting results with discussion on the four test set-ups conducted. Chapter 6 is the conclusion on each of the four test set-ups and a summary of the conclusions. The in-text citations in the thesis are referred to in the references section and appendices include the computer codes in the ANN program and list of publications .

In general the topics in the chapters are;

1. Introduction;
2. Research Background;
3. Methodology;

4. Case Study Areas;
5. Results & Discussion;
6. Conclusion.

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION

River flow forecasting is essential in water resources management since it can help optimize the usage of water. Flood forecasting is necessary in flood management to avoid the loss of lives and to minimize damages to properties. In both areas, forecasting with high accuracy and reliability are very important.

One of the approaches for developing flood forecasting model is based on a hydraulic design using hydrological parameters that affect the river flow. The parameters that can effect the modeling are sedimentation, dam operation, soil type, cross sectional area of the river, and rainfall. This means that many of the physical factors of the river need to be considered where this can lead to many undetermined relationship between the dynamic processes of the river flow. This approach is not only complex, but also requires data coordination in real-life situations, which can be very costly. Currently, ANN based model is a popular modeling tool for the river flow forecasting where it does not relied in physical properties of the study area but instead relied on historical data events in the area.

In the data forecasting study, the target is to find the best forecasting model that can provide high accuracy and reliable forecasting results. Many approaches have been tested in the ANN based model to improve the target forecasting of the model developed and these approaches are discussed in the following sections.

2.2 IMPORTANCE OF ANN

Artificial Neural Network (ANN) is a computing model that can solve non-linear problems from the time-series data that the ARMA based model or traditional statistical models cannot provide. Other problems such as regression, classification, prediction, system identification, feature extraction and data clustering can also be solved through ANN computing. In water resources studies, ANN has become a well established research area over the past 15 years for the prediction and forecasting of water resources variables (Maier et al. 2010). ANNs are applied in forecasting daily river flows (Atiya et al. 1999; Coulibaly et al. 2000; Ahmed & Sarma 2007; Wu et al. 2009), water level (Alvisi et al. 2006; Bustami et al. 2007; Chang & Chen 2003; Leahy et al. 2008; Romanowicz et al. 2008), flood events (Tareghian & Kashefipour, 2007; Kerh & Lee, 2006), rainfall-runoff patterns (Chiang et al. 2004; Agarwal & Singh 2004; Rahnama & Barani 2005), optimization (Cancelliere et al. 2002; Chandramouli & Deka 2005) and sedimentation (Krishnaswamy et al. 2000; Raveendra & Mathur 2008). Based on the study by Maier et al. (2010), 230 papers were published in high ranking journals that are within the water resources studies up to 2008. In the previous study conducted by Maier et al. (2000), about 43 papers were published. This shows that the popularity of ANN based models have grown due to the simplicity and flexibility to manage the model and its ability to produce data forecasting on the non-linear and dynamic data pattern.

Many early studies tried to develop forecasting model by determining the pattern in historical time series data. This means that the prediction of future events is based on past events. Most time-series techniques traditionally used modeling water resources series fall within the framework of the auto-regressive moving average (ARMA) class of linear stochastic processes (Toth et al. 2000). However, the inability of an ARMA-type model to account for a non-linear relationship between the various aspects of the dynamic process, non-linear statistical models have been used for the analysis of real-world temporal data. Nevertheless, the formulation of reasonable the non-linear models is an extremely difficult task (Chakraborty et al. 1990). Time series forecasting can also be based on statitistical models. Such models include simple linear regression, projection pursuit regression, polynomial regression, nonparametric regression, logistic regression, linear discriminant functions, classification trees, finite mixture models, kernel regression, and smoothing splines (Cheng & Titterington, 1994; Sarle, 1994). However, they often emphasized certain rules or limitations which restrict the condition of data forecasting. This is due to the data pattern that is non-linear and stochastic in nature. ANN modelling approaches have been embraced enthusiastically by practitioners in water resources, as they are perceived to overcome some of the difficulties associated with traditional statistical approaches. (Maier & Dandy 2000). The differences in the modeling approach between artificial intelligence (AI) and traditional statistical models, coupled with a lack of strict rules governing the development of the former, are probably the major reasons for the increased popularity of neural networks and ANNs have placed such sophisticated models within the reach of practitioners (Maier & Dandy 2000). It is necessary to consider the alternative models that have skills to disseminate the non-linearity and non-stationarity nature of the inflow data series (El-Shafie et al. 2008a).

2.3 ACCURACY AND RELIABILITY STUDIES IN ANN

The objective in data forecasting study is to find the best forecasting model that can give accurate forecasting results, if possible. In the ANN based model, the approach is by improving the generalization of the network model in data training. Modification to essential parameters in the ANN model during data training could achieve the results. These essential parameters are the number and type of data inputs, the number of hidden layer, the number of hidden neurons in each ANN layer, type of activation transfer function used in the neurons, and the optimization method of finding the weight in the ANN model.

In river flow forecasting, data inputs can be based on historical river flow data, rainfall, precipitation, and sedimentation. El-Shafie (2007, 2008b) had conducted a river flow forecasting study in the Nile River, using river flow data from a single station. Turan & Yurdusev (2009) used a multiple upstream river flow station for forecasting the river flow. Chiang et al. (2004) and Rahnama & Barani (2005) had included rainfall and run-off data for river flow forecasting, and Zealand et al. (1999) included precipitation, rainfall, and flow data in their forecasting study. Fernando et al. (2005) had suggested several methods to identify the proper inputs to a neural network. Other studies (Alvisi et al. 2006; Toth & Brath 2007) had investigated the effects of the number and type of inputs on the ANN performance.

Once data inputs have been determined, the selection of the number of hidden layers and neurons plays a vital role in achieving the best forecasting performance. Forecasting studies, in determining the number of layers and neurons, have been based on trial and error approach (Coulibaly et al. 2000; Joorabchi et al. 2007; Solaimani & Darvari 2008; Turan & Yurdusev 2009). For the number of hidden layers, many studies have shown that one hidden layer is enough. Hornik et al. (1989) had shown that the multilayer feed forward network with one hidden layer is capable of approximating to any desired degree of accuracy provided the sufficient hidden units are available based on universal approximation theorem. Zhang et al. (1998) had produced a synthesis of published research papers on ANN, showing that a single hidden layer is the most popular and widely used in layer selection. Two hidden layers are also able to produce best forecasting performance in certain problems (Barron 1994). However, for the selection of the number of ANN neurons, the method has been based on trial and error approach (Chauhan & Shrivastava 2008).

Activation transfer function (ATF), the main computing element of ANN, plays an important role in achieving the best forecasting performance. The most common type of ATF is sigmoid function (Zhang et al. 1998). However, several studies had used different types of ATF within the ANN to improve forecasting performance. Shamseldin et al. (2002) used logistic, bipolar, hyperbolic tangent, arc-tan and scaled arc-tan to explore the potential improvement of ANN forecasting. Joorabchi et al. (2007) applied log-sigmoid and hyperbolic tangent sigmoid transfer functions to produce their output. Han et al. (1996) introduced optimization of variant sigmoid function using genetic algorithm to optimize ANN convergence speed and generalization capability.

Many researchers had studied the ANN's exterior architecture using trial and error approach. Others had studied the ANN's interior architecture, but had been limited to testing different types of ATF in the ANN architecture. Sigmoid function, which is the most commonly computing function for ATF, has been widely used because of its ability to affect the performance of ANN. The objective of this study is to evaluate the effectiveness of steepness coefficient in sigmoid function in improving ANN data training and forecasting in a river flow study. The second objective is to investigate the effectiveness of the optimal steepness coefficient approach, compared to traditional approach based on trial and error on the exterior architecture in river flow forecasting. The improved performance of ANN water level forecasting could assist water authorities in managing water resources.

Another approach in improving the generalization of network model is the extrapolation method especially in the case of missing data or small dataset for learning process. Cigizoglu (2003) used scaling of data training in the range of 0.2 to 0.8 to predict flows beyond the training range. The forecasting results in the study were within 20% error bound, which is still high even though it is a better channel than the conventional statistical and stochastic models that are used in the study. Coulibaly et al. (2001) also used the extrapolation strategy with the use of time delay dynamic neural network to improve multi-layer perceptron based data forecasting. Imrie et al. (2000) improved generalization during training by applying cubic polynomial function in the output layer which may be necessary for ANNs to capture extreme values but not necessarily in other cases since the function is correlated to data training pattern of the extreme events. Hu et al. (2005) successfully applied what is called goal programming (GP) neural network in improving generalization. Giustolisi and Laucelli (2005) suggested using avoiding over-fitting technique that may be a helpful approach to increase the generalization performance of ANNs.

Improved performance in extreme events in water level forecasting utilizing ANN model is quite important since the extreme event is the critical point in providing effective flood monitoring and preventive measures. A common problem is the under-estimation of river flow (Alvisi et al. 2006; Shrestha et al. 2005; Thirumalaiah & Deo 1998). The reasons given for this under-estimation include the lack of water level events in the sample dataset (Toth et al. 2000). This is true for several highest points of water level during big flooding events, although it is not true for water level events that are within the flood water level. Several studies had analyzed extreme events in different approaches such as spiking of the calibration data, data partitioning, and extrapolation.

Shrestha et al. (2005) attempted to overcome the problem of under-estimation by multiplying the high flow events a factor of 1.5. This approach could be suitable in a specific case study, but may not work in others. Cheung et al. (1990) suggested the use of spiking of the calibration data to improve the learning speed and generalization ability of back-propagation ANN based model especially in extreme events. However, the study by Maier et al. (1998) found that the use of the spiking method did not improve the generalization. Instead, Maier et al. (1998) suggested the application of appropriate learning rate and momentum with continuous evaluation of forecasting performance at each small interval of data training to reach the best network generalization. The approach described in this study, however, is time consuming where during each change of training interval, the learning rate and momentum are modified and the network performance is evaluated.

Self-organizing map (SOM) is a data partitioning/splitting, which is a novel approach to improve the generalization of the network model (May et al. 2010). Various techniques have been utilized to improve the SOM based model since it has the potential to perfect the generalization of the network model. Among the studies, May et al. (2010) surveyed the SOM-based stratified sampling approach in data splitting for the development of the ANN model. SOM is an un-supervised type of ANN-based model that can hide the pattern from sample dataset without specifying sample data output. SOM organized sample dataset into groups of dataset that have similar data pattern as output. In terms of river forecasting, SOM organizes the flow or water level into groups of similar flow or water level pattern in the testing and validation datasets, which can be used as hybrid model with the ANN-based model. Jain & Srinivasulu (2006) suggested two approaches in data partitioning using rainfall run-off data for river flow forecasting. First is the data partitioning by segmenting the rising and falling of the effective rainfall run-off data using statistical measures by trial and error. The other is the partitioning of the data based on the high, medium, and low flows using SOM method. The study concluded that forecasting is improved by the partitioning of the rainfall run-off data but the processes of the development of the