Ensemble flood forecasting by neuro-fuzzy inferencing

Yu, Lan

2016

Yu, L. (2016). Ensemble flood forecasting by neuro-fuzzy inferencing. Doctoral thesis,
Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/68924>

<https://doi.org/10.32657/10356/68924>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**ENSEMBLE FLOOD FORECASTING BY NEURO-
FUZZY INFERENCE**

YU LAN

School of Civil and Environmental Engineering

2015

**ENSEMBLE FLOOD FORECASTING BY NEURO-FUZZY
INFERENCE**

YU LAN

School of Civil and Environmental Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2015

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to those people who support me during my four-year research and life.

I want to thank my first advisor, Associate Prof Lloyd Chua, who offered me the research opportunity into the interesting world of modeling. Professional advice, continuous support and insightful comments that he gave to me helped me get through those challenges in my research study.

I would also like to thank my supervisor, Associate Prof Tan Soon Keat, for helping me in my last two years of PhD research. His patience, immense knowledge and precious support helped me to finalize my PhD research.

The measured data and URBS results from Chapter 4- 5 were generously provided by the Mekong River Commission. The funding for the work from Chapter 6-7 was provided by the DHI-NTU Centre, NEWRI and National Chung Hsing University. I also wish to thank the TTFRI for providing data for this research.

I would like to thank Amin Talei for his professional suggestions in modeling and helping me get out of the confusion in my research.

I want to thank my family for giving me the constant and tremendous energy in my work and life. I want to thank Wang Qi, who went from being my girlfriend to my wife in my PhD life, for your love.

Last but not least, I would like to thank all my lovely friends in Singapore. Their company makes a piece of precious memory in my PhD life.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	I
SUMMARY	VI
LIST OF TABLES.....	VIII
LIST OF FIGURES	IX
LIST OF ABBREVIATIONS	XII
LIST OF PUBLICATIONS	XIV
CHAPTER 1 INTRODUCTION	1
1.1 BACKGROUND.....	1
1.2 MOTIVATION	2
1.3 OBJECTIVES	3
1.4 SCOPE	4
CHAPTER 2 LITERATURE REVIEW	6
2.1 INTRODUCTION	6
2.2 FLOOD FORECASTING	6
2.2.1 Physically Based Models	6
2.2.2 Statistical Models.....	12
2.2.3 Data-Driven Models.....	18
2.3 ENSEMBLE METHODS	24
2.3.1 Ensemble Methods in Weather Forecasting.....	24
2.3.2 Ensemble Methods for Optimization in Water Resources	26
2.3.3 Ensemble Methods in Flood Forecasting.....	26
CHAPTER 3 METHODOLOGY AND DATA USED	32
3.1 INTRODUCTION	32

3.2	SIMPLE AVERAGE METHOD	32
3.3	ADAPTIVE-NETWORK-BASED FUZZY INFERENCE SYSTEM.....	33
3.4	DYNAMIC EVOLVING NEURAL-FUZZY INFERENCE SYSTEM	35
3.5	STUDY AREAS	37
3.5.1	Lower Mekong Basin.....	38
3.5.2	Lanyang Creek Basin, Taiwan	39
3.6	ERROR ANALYSIS	42
 CHAPTER 4 WATER LEVEL FORECASTING FOR THE LOWER MEKONG USING A NEURO-FUZZY INFERENCE SYSTEM ENSEMBLE APPROACH 46		
4.1	INTRODUCTION	46
4.2	METHODOLOGY	47
4.2.1	ANFIS Ensemble Model (ANFIS-EN)	47
4.2.2	DENFIS Ensemble Model (DENFIS-EN)	49
4.2.3	Data Used.....	49
4.3	RESULTS AND DISCUSSIONS	52
4.3.1	ANFIS Ensemble Model.....	52
4.3.2	DENFIS Ensemble Model	56
4.3.3	Analysis of Results	59
4.4	CONCLUSIONS	63
 CHAPTER 5 ONLINE ENSEMBLE MODELING FOR REAL TIME WATER LEVEL FORECASTS FOR THE LOWER MEKONG RIVER 65		
5.1	INTRODUCTION	65
5.2	METHODOLOGY	66
5.2.1	Neural-fuzzy Model	66
5.2.2	Real Time Updating Approach.....	67
5.2.3	Study Site	69

5.3	RESULTS AND DISCUSSIONS	70
5.3.1	Offline Ensemble Model (EN-OFF)	70
5.3.2	Ensemble Model with Real Time updating using Online Learning (EN-RTON1).....	72
5.3.3	Ensemble Model with Real Time updating using Online Learning and Sub-Models (EN-RTON2)	77
5.4	CONCLUSIONS	81
CHAPTER 6 ENSEMBLE WATER LEVEL FORECASTING FOR LANYANG CREEK, TAIWAN		83
6.1	INTRODUCTION	83
6.2	METHODOLOGY	84
6.3	EVALUATION OF INPUT COMPONENT MODELS	84
6.3.1	Short- and Long-term Forecasts.....	84
6.3.2	Forecast Results at Different Water Level Regimes	89
6.4	RESULTS OF ENSEMBLE FORECASTS	93
6.5	CONCLUSIONS	95
CHAPTER 7 ENSEMBLE APPROACH USING MODIFIED OFFLINE MODELS FOR WATER LEVEL FORECASTING IN LANYANG CREEK, TAIWAN		97
7.1	INTRODUCTION	97
7.2	METHODOLOGY	97
7.2.1	Modified DENFIS with Linear Constraints	97
7.2.2	Modified DENFIS with Linear Constraints and Slopes	99
7.2.3	Data and Study Area	101
7.3	RESULTS AND ANALYSIS.....	102
7.3.1	Results of the Modified Offline Model.....	102
7.3.2	Results of the Modified Offline with Slope Model.....	109

7.4	CONCLUSION.....	116
CHAPTER 8 CONCLUSIONS AND RECOMMENDATIONS.....		118
8.1	CONCLUSIONS	118
8.2	PRACTICAL APPLICATIONS	122
8.3	RECOMMENDATIONS	123
REFERENCES.....		125

SUMMARY

Physically-based models, regressive models and data-driven models have been applied to flood forecasting to produce accurate predictions in many studies. However, there remain drawbacks in individual models, whether physically based or otherwise. For example, not all the phases of the hydrograph can be predicted well by any model, even though the global optimum may be reached. Therefore, in order to exploit the strengths of different models, the ensemble model approach can be used to improve forecast accuracy. This thesis began with a review on the research carried out on flood forecasting conducted at two levels: (i) flood forecasting by individual models (including physically-based models, statistical models and data-driven models) and (ii) methods used to develop ensemble flood forecasts by combining component model results. For the former, much work has been done by applying conceptual, distributed or lumped models, time series models and black box models to flood forecasting. For the latter, only limited studies have been attempted with some simple statistical methods and data driven models. Although limited in scope, these studies indicate that ensemble flood forecasting show improved accuracy over the individual models. In addition, when multiple predictions are available, it is common to calculate an average of the different models' results. However, there is often no basis to use an averaging procedure, and therefore, a better approach is needed.

Current ensemble methodologies adopted in flood forecast studies include simple average method (SAM), the weighted average method (WAM), Bayesian Model Averaging (BMA) and fuzzy logic. These ensemble models are applied to the component models with arbitrary weights or fixed weight allocation strategy and are not considering the performance of the component models at different stages of the hydrograph. This thesis presents the use of neuro-fuzzy inference system (NFIS) as an ensemble methodology exploiting the parameter learning from neural networks and interpretation from fuzzy logic. In particular, the Dynamic Evolving Neural-

Fuzzy Inference System (DENFIS) was used in this study with the clustering algorithm for its weight allocation strategies and online learning for model adaptation during testing stage. Here, two general cases of ensembling are investigated: (i) Different rainfall-runoff models with the same rainfall inputs and (ii) Different rainfall inputs but with the same rainfall runoff model. For the former, data from the Lower Mekong River was analyzed. For the Lower Mekong, flood forecasts are from two independent models, Adaptive-Network-Based Fuzzy Inference System (ANFIS) and the Unified River Basin Simulator (URBS) hydrological model. A real time updating ensemble model based on the online learning ability of DENFIS was proposed to provide an ensemble forecast. With the proposed ensemble model using real time updating, the ensemble model can adapt to higher water levels during testing stage than those in the training stage. By continuously updating the model, the model is able to better adapt to changes in the forecast by reducing the spikes from the component URBS model and the time shift error from the ANFIS component model. For the Taiwan case study, data from a catchment in Taiwan was analyzed. The Taiwan data includes runoff predictions based on 15 rainfall inputs, obtained from 15 different perturbations of an atmospheric model. A data processing procedure is suggested as a preliminary step to form a truncated input space for the ensemble model. The modified offline models which impose weight constraints and consider the effects of the slopes were proposed to highlight the interpretation of the ensemble process. Not only the peak and the shape of the hydrograph was better predicted compared with the benchmark SAM model, the fuzzy rules of the weight allocation were interpreted to show the mechanism of the ensemble approach based on NFIS model. The results from the two catchments show the possible ensemble solutions to optimizing the water level estimations for different cases of flood forecasting in practice.

LIST OF TABLES

Table 3.1 Inductive and deductive approaches.....	44
Table 4.1 Model comparison.....	60
Table 5.1 Model performance evaluation for (a) Thakhek (b) Pakse and (c) Kratie of test data from 20 Jun 2011 to 20 Oct 2011	80
Table 6.1 The distance of the fifteen component models to the “Perfect Model” in PCA reduced space.	86
Table 6.2 Component models classification	87
Table 6.3 Ranking of the fifteen component models for the training, validation and test phase.	89
Table 6.4 Selected component models for ensemble model	90
Table 6.5 Training and validation RMSE (m) of the ensemble offline model with different input selections	91
Table 6.6 Evaluation of 2013 results for different inputs selected of the ensemble offline model	91
Table 6.7 Evaluation of 2014 results for different inputs selected of the ensemble offline model	92
Table 7.1 Training and validation RMSE (m) of the modified offline model with different input selections	102
Table 7.2 Evaluation of 2013 results for different inputs selected of the modified offline model	103
Table 7.3 Evaluation of 2014 results for different inputs selected of the modified offline model	103
Table 7.4 Training and validation RMSE of the modified offline model with slope for different input selections	109
Table 7.5 Evaluation of 2013 results for different inputs selected of the modified offline with slope model.	110
Table 7.6 Evaluation of 2014 results for different inputs selected of the modified offline with slope model	111

LIST OF FIGURES

Figure 3.1 (a) Type-3 Fuzzy Reasoning (TS fuzzy if-then rules are used). (b) Equivalent ANFIS or Type-3 ANFIS (Jang, 1993).	34
Figure 3.2 Clustering process using ECM in 2-D space (a)x1: the creation of the first cluster 1(b) x2: update cluster 1; x3: creation of a new cluster 2; x4: belongs to cluster 1 (c)x5: update cluster 1; x6: belongs to cluster 1; x7: update cluster 2; x8: creation of a new cluster 3 (d)x9: update cluster 1 (Kasabov and Song, 2002)	37
Figure 3.3 (a) Location map for the Mekong Basin; (b) Sub-basin with gauging station Kratie: 28,815 (km ²). Source: (MRC, 2005; 2007)	39
Figure 3.4 Lanyang Creek Basin (Shih <i>et al.</i> , 2014)	41
Figure 4.1 The structure of the ANFIS-EN with three membership functions	48
Figure 4.2 Comparisons of observed water level, ANFS predictions and URBS predictions for Kratie station (Nguyen and Chua, 2012) of (a) 7 th Jun 2009 to 31 st Oct 2009, (b) 7 th Jun 2010 to 31 st Oct 2010 (c) 20 th Jun 2011 to 20 th Oct 2011	51
Figure 4.3 ANFIS-EN-V forecasts for (a) training data (7 th Jun 2009 to 31 st Oct 2009), (b) validation data (7 th Jun 2010 to 31 st Oct 2010) (c) test data (20 th Jun 2011 to 20 th Oct 2011)	54
Figure 4.4 ANFIS-EN without validation forecasts for (a) training data (7 th Jun 2009 to 31 st Oct 2009 and 7 th Jun 2010 to 31 st Oct 2010), (b) test data (20 th Jun 2011 to 20 th Oct 2011)	55
Figure 4.5 DENFIS-EN with validation forecasts for (a) training data (7 th Jun 2009 to 31 st Oct 2009), (b) validation data (7 th Jun 2010 to 31 st Oct 2010) (c) test data (20 th Jun 2011 to 20 th Oct 2011)	57
Figure 4.6 DENFIS-EN forecasts for (a) training data (7 th Jun 2009 to 31 st Oct 2009 and 7 th Jun 2010 to 31 st Oct 2010), (b) test data (20 th Jun 2011 to 20 th Oct 2011)	59

Figure 4.7 Comparison of the ANFIS-EN-V and DENFIS-EN results with the SAM results for the test data (20 th Jun 2011 to 20 th Oct 2011)	61
Figure 4.8 Time shift comparison of the ensemble models and the component models for the test data (20 th Jun 2011 to 20 th Oct 2011).....	62
Figure 4.9 Weights Change of the Component Models from DENFIS-EN for the test data (20 th Jun 2011 to 20 th Oct 2011).....	63
Figure 5.1 Structure of the ensemble model with real time updating using online learning (EN-RTON1).....	67
Figure 5.2 The structure of the ensemble model with real time updating using online learning and sub-models (EN-RTON2).	69
Figure 5.3 EN-OFF Results for (a) Thakhek, (b) Pakse and (c) Kratie for test data from 20 Jun 2011 to 20 Oct 2011	71
Figure 5.4 EN-RTON1 and EN-RTOFF compared with EN-OFF and SAM Results for (a) Thakhek, (b) Pakse and (c) Kratie for test data from 20 Jun 2011 to 20 Oct 2011	74
Figure 5.5 Error Analysis on creating new clusters of EN-RTON1 ensemble results for (a) Thakhek and (b) Pakse for test data from 20 Jun 2011 to 20 Oct 2011 .	75
Figure 5.6 Change in weights of EN-RTON1 for test data from 20 Jun 2011 to 20 Oct 2011, Pakse.	77
Figure 5.7 EN-RTON2 results compared with EN-OFF, EN-RTON1 and SAM model results for (a) Thakhek, (b) Pakse and (c) Kratie for test data from 20 Jun 2011 to 20 Oct 2011	79
Figure 6.1 Evaluation of component model forecast performance at different ranges of water levels for the training dataset	90
Figure 6.2 Comparison of the forecasts from the WASH123D and ensemble models: (a) 2013 test event, (b) 2014 test event.	94
Figure 7.1 The Structure of the modified offline mode of DENFIS with linear constraints and slopes.	100

Figure 7.2 Comparison of the forecasts from the WASH123D and Modified Offline model: (a) 2013 test event, (b) 2014 test event.	105
Figure 7.3 Weight allocation of the combined component models.....	106
Figure 7.4 Total weights of the normalized component model forecasts in the modified offline model for (a) 2013 test event (b) 2014 test event.	108
Figure 7.5 Comparison of the forecasts from the WASH123D and modified offline with slope model : (a) 2013 test event, (b) 2014 test event	112
Figure 7.6 Clusters with the highest weight allocation to each component model of the modified offline with slope model	114
Figure 7.7 Total weights of the normalized component model forecasts in the modified offline with slope model for (a) 2013 test event; (b) 2014 event.	115

LIST OF ABBREVIATIONS

Abbreviation	Description
1L1M1H	Best performing component models at Low, Medium and High water levels
1L2M3H	Best component model for low, best two component models for medium and best three component models for high water levels
2L2M2H	Best two component models for Low, Medium and High water levels
ANFIS	Adaptive-Network-Based Fuzzy Inference System
ANFIS-EN	ANFIS Ensemble Model
ANFIS-EN-V	ANFIS Ensemble Model with Validation
ANN	Artificial Neural Network
DENFIS	Dynamic Evolving Neural-Fuzzy Inference System
DENFIS-EN	DENFIS Ensemble Model
DENFIS-EN-V	DENFIS Ensemble Model with Validation
V	
D_{thr}	Threshold of the Distance
ECM	Evolving Clustering Method
EN-OFF	Offline Ensemble Model
EN-RTOFF	Ensemble Model with Real Time updating using Offline Learning
EN-RTON1	Ensemble Model with Real Time updating using Online Learning
EN-RTON2	Ensemble Model with Real Time updating using Online Learning with Sub-Models
LS	Least Squares
MaxDist	Maximum Distance

MRC	Mekong River Commission
NFIS	Neuro-Fuzzy Inference System
NSE	Nash-Sutcliffe Efficiency
PBIAS	Percent Bias
PE	Percentage Error at peak flow
PEP	Percent Error in Peak
PT	Peak Time difference
RLS	Recursive Least Squares
RMSE	Root Mean Square Error
SAM	Simple Average Method
URBS	Unified River Basin Simulator
WRF	Weather Research Forecasting

LIST OF PUBLICATIONS

Journals

Lan Yu, Soon Keat Tan, Lloyd H C Chua, Dong-sin Shih, Development of a Model Ensemble Approach for Flood Forecasts, (Under revision in Natural Hazards)

Lan Yu, Soon Keat Tan, Lloyd H C Chua, Water Level Forecasting for the Lower Mekong using an Ensemble Model Approach, (Submitted to Water Resources Management)

Lan Yu, Soon Keat Tan, Lloyd H C Chua, Online Ensemble Modeling for Real Time Water Level Forecasts, (Submitted to Water Resources Management)

Conferences

Lan Yu, Lloyd H C Chua, 2013. Ensemble Model Water Level Forecasts for the Lower Mekong, International Conference on Flood Resilience: Experiences in Asia and Europe, Exeter, United Kingdom. (Oral presentation)

Lan Yu, Lloyd H C Chua, Dong-sin Shih, 2014. An Ensemble Approach Typhoon Runoff Simulation with Perturbed Rainfall Forecasts in Taiwan, 11th International Conference on Hydroinformatics, New York City, USA.

CHAPTER 1 INTRODUCTION

1.1 Background

Floods pose one of the greatest hazards to mankind. Recent trends in the earth's weather have seen an increase in the occurrence of floods. According to IPCC (2012), "A changing climate leads to changes in the frequency, intensity, spatial extent, duration, and timing of extreme weather and climate events, and can result in unprecedented extreme weather and climate events". In addition to climate change, with rapid expansion of urban centers, our exposure to the dangers posed by floods has also increased (Du *et al.*, 2012; Poelmans *et al.*, 2011). For example, when Typhoon Morakot struck Taiwan in 2009, rainfall intensities exceeded 1,000 mm/hr, resulting in close to 700 deaths, and more than US\$ 3 billion in damages (Shen and Chang, 2012). The impacts of increased flood peaks and flow volume due to urbanization and population growth are expected to exacerbate the risk of floods in major cities in the future. Any significant increase in flood risk would have serious adverse socio-economic impacts. Thus, cost effective measures for mitigation and management of the risks associated with rising exposure to flooding are expected to become increasingly critical for policy makers and the industry. To facilitate the decision-makers in dealing with flood-risk assessment, there is an urgent need for the development of improved numerical modeling tools to better predict floods.

According to their causes, flooding can be classified into riverine flooding, flash floods, downstream flooding caused by urban drainage, floods due to dam failure, floods due to ground failures or coastal flooding. The primary effect of floods is physical damage to structures including buildings, drainage systems and roads and loss of human lives. In addition, water and food supplies will be adversely affected because of the contamination of clean water and loss to agriculture production and the outbreak of waterborne diseases is also common during flood events. Higher flood

peaks and shorter response time are brought about by urbanization. Vegetation removal and meander removal increase the flow velocity resulting in the risk of floods and increased erosion. Increase of impervious surface also leads to the increase of surface runoff discharge. To prevent losses of life and property, establishing accurate and fast flood forecasting and warning system is essential for the public to take actions with enough lead time (Yang *et al.*, 2015). Various methodologies, roughly categorized as physically-based models, statistical models and data-driven models are used to model floods (Chua, 2012). However, most models cannot provide the flexibility required to model all aspects of the hydrological process (Xiong *et al.*, 2001). For example, physically based models are advantageous in scenario based studies, however may not be well suited for real time applications, due to the computational time and extensive data requirements imposed by the models. Data driven models are faster to set up and arguably have a relatively lower demand on data needs; however, this class of models do not provide any physical interpretation and are often treated as black box models (Todini, 2007; Toth *et al.*, 2000).

In real time applications, it is important that any model be able to provide accurate forecasts at all stages during a flood. Clearly, such a model does not exist. Therefore, an ensemble approach has recently been proposed which considers outputs from a committee or collection of models that are combined in such a way so as to produce a forecast result which is more accurate than the individual models considered. In this approach, the premise is that each of the models constituting the ensemble provides distinct information resulting in the combination having a higher accuracy and reliability than that of individual models (Shamseldin and O'Connor, 2003).

1.2 Motivation

Physically-based models, statistical models and data driven models are used in flood prediction (Blöschl *et al.*, 2008; Chen *et al.*, 2006; Tingsanchali and Gautam, 2000).

Data driven models, in particular, neural networks, fuzzy inference system and hybrid neural-fuzzy systems have been adept in providing superior results in some studies. However, all these models working individually cannot be used to provide satisfactory predictions all the time and indeed all phases of flooding. A better solution is needed. In this thesis, the ensemble approach is investigated. The ensemble approach includes combining component model results to provide improved forecasts than a random combination of models. In addition, a more reasonable basis should be provided compared with the simple average method which uses equitable weight allocation through all phases of the prediction. Additional selection and combination of information from the component models is performed in model ensembling rather than combining model results arbitrarily (Seni and Elder, 2010). Limited research studies using the ensemble approach have been carried out (Duan *et al.*, 2007; Shamseldin and O'Connor, 1999; Xiong *et al.*, 2001). However, the ensemble methods used have mainly been confined to the use of simple and weighted averaging, Bayesian theory and data-driven approach. Yet, in spite of the simplicity of these ensemble approaches, the improvements obtained by various researchers, clearly point to the potential for the ensemble modeling approach.

1.3 Objectives

The objectives for this study include:

1. Investigate the use of (Neuro-Fuzzy Inference System) NFIS for ensemble modeling for flood forecasting applications and analyze the advantages and disadvantages of the ensemble approach over conventional flood forecasting methodologies by demonstrating the models on real world cases
2. Validate the flexibility during rules creation of the use of NFIS model with clustering algorithm for ensemble flood forecasting.

3. Validate the use of ensemble model approach for a real time flood forecast problem using the online learning algorithm to improve the model adaptation ability during testing stage by capturing the uncertainty that physics-based model cannot do so in the real-time forecasting.
4. Show the improvement from the input pre-processing of the component members before applying the ensemble model.
5. Interpret the structures and parameters in the NFIS ensemble model to provide mechanism description of the ensemble process.
6. Generalize the patterns of the ensemble flood forecasting by inductive approach.

1.4 Scope

The scope for this study includes:

1. Three individual groups (physically-based models, statistical models and data-driven models) will be reviewed and discussed in relation to their structure, parameter estimation as well as their advantages and disadvantages. This will help clarify the distinct qualities of individual models.
2. Apply the ensemble NFIS model with clustering algorithm for rules creation and the ensemble NFIS model with pre-defined structure to the water level forecasts for Kratie located in Lower Mekong where two rainfall-runoff models have been used in application. Evaluate the performance of the ensemble models against the component models and show the improvement of using the dynamically estimating the fuzzy rules
3. Real time updated ensemble approaches will be proposed and applied to three stations located in Lower Mekong and the improvement of the proposed ensemble model will be demonstrated compared with the ensemble model without real time updating

4. Apply ensemble approach to the water level forecasts of the same hydrological model with different rainfall forecasting inputs of Taiwan catchment. The pre-analysis and processing of the input data will be highlighted and shown to improve the ensemble water level forecasts
5. Modified ensemble models will be used for Taiwan catchment to improve the forecasting accuracy while trying to interpret the ensemble process. The parameters variation with real time forecasting will be shown in this part.
6. The validation of the proposed models: the first step was to divide the data set into training data set and test data set; the training data usually contains $\frac{2}{3}$ of the whole data set and is used to train and optimize the model parameters in which the measured water levels were available to the model as feedback; the trained models will be applied to the test data with the optimized parameters in which the measured water levels were used to calculate the errors of the model forecasts. Thus the model performance can be validated from the unknown data set.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

This chapter reviews the two main aspects of research on hydrological forecasting in the literature. The first aspect is more commonly found and includes forecast models which includes physically based models, statistical models and data driven models. The review of forecast models is added for completeness; however will not be overly emphasized since the focus of this thesis is on the second aspect, which is ensemble modeling. Although ensemble modeling has received relatively less attention in flood forecasting applications, ensemble models have been applied to different fields such as climate change, weather forecast and water resources management. The results of studies where the ensemble method has been adopted in these related fields is included in this review.

2.2 Flood Forecasting

2.2.1 Physically Based Models

Physically-based models have traditionally been used to predict river stage and discharge. This class of models is based on mathematical equations describing the physical characteristics of mass and momentum conservation in the flow in a river or overland plane. Physically based models can be categorized as conceptual, distributed or lumped models. Conceptual rainfall-runoff models approximate the physical mechanisms governing the hydrologic cycle with the identification of water budgets (Duan *et al.*, 1992). Distributed models use spatially distributed forcing and distributed watershed parameters to simulate watershed processes (Tang *et al.*, 2007). The most important difference of distributed models from other physically based models is the prescription of the highly heterogeneous media properties within a grid element (Blöschl *et al.*, 2008), while lumped models treat the catchment as a whole unit (Shultz, 2007) and are based on the uniform distribution of hydrological and

physical parameters such as rainfall, soil type, vegetation type and land-use practices over sub-basin scales.

Franchini and Pacciani (1991) compared several conceptual rainfall-runoff models with the four-month period data for the Sieve watershed. These models were carefully selected from various classes: Stanford Watershed Model IV (STANFORD IV), SACRAMENTO model, Tank model, APIC model, SSARR model, XINANJIANG model and ARNO model. The soil-level water balance and the transfer to the closure section of the watershed were identified as two distinct components for each model. The results showed that an inaccurate description of the drainage process in the soil-level water balance component would lead to overestimation of the concentration time. Besides, inadequate quantitative calculation of the flows in the soil exists in these conceptual models. In other research work, the results of the conceptual Tevere Flood Forecasting (TFF) model were compared with those of the Tevere neural network (TNN model). Both models gave similar performance for 12-hour ahead prediction but the TFF model showed high dependence on rainfall information which resulted in an underestimation of the rising limb of the hydrograph with an increase in the lead time (Napolitano *et al.*, 2010). Yao *et al.* (2014) improved the forecasting ability of a conceptual model Xinanjiang model (XAJ) by using the geomorphologic instantaneous unit hydrograph to replace the lag-and-route method in the conceptual model. The hybrid coupled model (XAJ-GIUH) and the conceptual XAJ model were evaluated at hourly scale for a mountainous catchment of six interior catchments with observed stream flow in the south of Anhui, China. Two experiments were conducted to calibrate the model parameters on each of the catchments or to transpose the model parameters from the downstream catchments. For the first experiment both the XAJ and XAJ-GIUH models produced comparable forecasting results. For the second experiment the XAJ-GIUH model improved the peak discharge and peak time estimation, which showed the major uncertainty of XAJ model comes from the runoff

routing. Wolfs *et al.* (2015) proposed a modeling approach from the storage cell concept to select the model structure and optimization methods based on the studied river system. In the configuration of the conceptual models, a semi-automatic tool Conceptual Model Developer (CMD) was used to improve the speed. A conceptual model based on the proposed modeling approach was tested on the Marke River in Belgium with the flood events of Nov 2010 and Mar 2008 for calibration and the event of Jan 2011 was used to validate the model performance. With the proposed approach the calculation time was proved to be reduced by more than 2000 times. Orth *et al.* (2015) used three conceptual models the simple water balance model (SWBM), Hydrologiska Byråns Vattenbalansavdelning model (HBV) and PREcipitation Runoff EVApotranspiration Hydrological response unit model (PREVAH) to model runoff and soil moisture for the catchments across Switzerland. After calibrating the models with the same runoff, the PREVAH and HBV which have higher complexity produced better runoff simulation than the SWBM while HBV produced worse soil moisture simulation than the other models. For extreme events, it was found SWBM worked better for droughts while the other two conceptual models were suitable for floods. It was concluded that weak correlation between the model complexity and the improved model performance and the forecasted hydrological variable and the conditions will lead to different forecasting performance.

One of the physically distributed hydrological models, the WASH123D numerical model was calibrated and validated for the Lanyang Creek basin in northeastern Taiwan with two years of data (Shih and Yeh, 2010). The hydraulic parameters including Manning's roughness coefficients, porosity and hydraulic conductivity were determined based on the mean absolute error and the root-mean-squared error. The good performance of this model was demonstrated but the lack of data, especially precipitation, and computation cost proved to be a limitation of distributed flood

modeling when applied to the 622 km² Kamp catchment in Austria for 4-24 h lead times (Blöschl *et al.*, 2008). The model structure based on multi-source model identification was built at the model element scale of 1 km². To decrease the impact from the biases in rainfall data, the Ensemble Kalman Filter (applicable to non-linear models) and the autocorrelation of the forecast error were implemented in a real time mode. Tang *et al.* (2007) evaluated the parameter sensitivities of a distributed model (HL-RDHM) and tested the model for all grid cells at 24-hour and 1-hour model time steps respectively for two watersheds in the Juniata River Basin in central Pennsylvania. Sobol's Method (Sobol', 1993) was used to evaluate the variance of the model outputs with changes of parameter vectors and Latin Hypercube Sampling (LHS) (McKay *et al.*, 1979) was applied to sample the feasible parameter space. The results revealed that HL-RDHM was mostly influenced by storage variation, spatial trends in forcing and cell proximity to the gauged watershed outlet. Trudel *et al.* (2014) proposed a data assimilation approach of outlet and interior stream flow observations for the distributed hydrological model CATHY (CATchment HYdrology) for the Des Anglais catchment in Canada. Latin hypercube sampling was used to replace Monte Carlo method to produce ensembles for reducing the number of the ensembles. With the proposed assimilation of stream flow observations, the lag between the forecasts and observations during rainfall events were corrected. Braud *et al.* (2010) used two distributed hydrological models for the flash flood event of the September 8-9 of Year 2002 in southern France. The first model was a spatially distributed rainfall-runoff model, MARINE model, to simulate extreme events. The other model was the CVN model built in the LIQUID modeling platform and the two models had different spatial discretization, infiltration, river flow transfer and representation of the water distribution. The analysis of the two models indicated the most important factor of the precipitation for the flash flood dynamics and the flood peak was also affected by the roughness of the river bed. Cole and Moore (2009) used the improved the radar precipitation by combining with rain gauge network data to produce gridded rainfall

estimates for a distributed hydrological model Grid-to-Grid Model (G2G). A lumped hydrological model the Probability Distributed Model (PDM) was used as the benchmark. The River Kent and the River Darwen in northwest England were selected for the hydrological modeling and the distributed model with the proposed gridded radar precipitation estimators produced better forecasts for the ungauged locations.

Yu and Yang (1997) implemented a rainfall forecast model into a lumped rainfall-runoff forecasting model derived from a transfer function and forecast flow with lead time of 1 to 4 hours. The rainfall forecast model used a probability-based approach with transition probability matrices and the comparison between the observed and forecasted hydrographs showed accurate output with 1 to 4 hours lead time. Another lumped hydrological model was applied to produce flood frequencies at the outlet of an un-gauged basin, which was followed by using the threshold runoff in computing the Flash Flood Guidance or FFG (Norbiato *et al.*, 2009). The FFG method and a semi-distributed continuous rainfall-runoff model were compared in a validation exercise over four basins in the central-eastern Italian Alps with size ranging from 75.2 km² to 213.7 km². The results showed that the threshold runoff produced improvements in gauged and un-gauged basins with inherent bias correction. Anctil *et al.* (2003) modified three lumped conceptual rainfall-runoff models (GR4J, IHAC, and TOPMO) with artificial neural networks as the output updating method. GR4J originates from GR3J (Edijatno *et al.*, 1999) with modified parameters. IHAC is a modified form of the IHACRES model (Jakeman *et al.*, 1990) with six parameters, and TOPMO is a modified version of TOPMODEL (Beven and Kirkby, 1979). The results of 1 day and 3 days stream flow forecasts indicated superiority over a parameter updating scheme and the simple output updating scheme that just takes the last observed forecast error. In addition, the 3 days forecast performance of the Artificial Neural Network (ANN)-output-updated models was better than that of the

ANN model forecast. In flood forecasting for the Wichianburi sub-basin of the Pasak River basin (6,250 km²) and the Tha Wang Pha sub-basin of the Nan River Basin (2,200 km²), Thailand, two lumped hydrological models tank and the Nedbør-Afstrømnings-Model or NAM model were compared with a neural network model (Tingsanchali and Gautam, 2000). The predictions of the Tank and NAM models were inaccurate because of erroneous rainfall input data but the neural network model performed better for both sub-basins with rainfall, evaporation and runoff data to make predictions. Stochastic models were also applied to improve the forecast accuracy of tank and NAM, which was proved to present high dependency on the persistence of the error time series. Martina *et al.* (2011) proposed a physical interpretation of the hysteresis which affected the relationship between the Saturated Area Ratio (SAR) and the Relative Water Content (RWC) through the cycles of wetting and drying phase. By incorporating the mechanism into the lumped TOPKAPI model and implementing the model on the Reno catchment of Casalecchio, both the outlet observed discharge and the time variability of the SAR and RWC were reproduced by the modified lumped model. Lerat *et al.* (2012) coupled the continuous lumped GR4J model which was used to calculate lateral inflows and the linearised diffusive wave hydraulic model for propagation. The coupled models were applied to the 10-year hourly data of Illinois River with point inflows only or with point and uniformly distributed inflows as well as 1-6 tributaries. The downstream end of the river reach and two interior points were assessed and it was shown that the coupled models including uniformly distributed inflows produced more robust and stable results. Including two or three tributaries produced good enough results for the main channel which indicated the less sensitivity of the model to the number of the tributaries. Seiller *et al.* (2015) applied twenty lumped conceptual models to twenty watersheds from the Model Parameter Estimation Experiment (MOPEX) international project in the United States of America. In the context of climate change, the performance of the individual models and the multi-model approaches of

weighted averaged results and simple averaged results was evaluated with the NSE efficiency criteria. It was found that the variety of the individual model performance based on the test period and catchments. And the weighted averaged results of the multi-models improved the temporal transposability for most watersheds than the simple averaged results. Nasr and Bruen (2008) included two lumped models of Simple Linear Model (SLM) and Soil Moisture Accounting and Routing (SMAR) model into a framework of the neuro-fuzzy model (NFM). To describe the temporal and spatial variation, two scenarios NFM_T and NFM_S were considered. For the temporal scenario, 11 catchments from the world were studied and the NFM_T_SMAR model produced better results than the NFM_T_SLM. For the spatial scenario, a subtractive clustering algorithm which used elevation, slope index, generalised land use types and soil types was adopted and the best results were reached by the NFM_S_SMAR model.

2.2.2 Statistical Models

2.2.2.1 Recursive methods

Young and Wallis (1985) pointed out that the dominant dynamic behavior of the hydrological system could be simulated or approximated with linear models. Due to the presence of a strong autocorrelation in the forecast errors of the conceptual models, some work has been done through applying stochastic models to simulate forecast errors. Vieira *et al.* (1993) applied a statistical model based on a multiple regression technique for the forecasting of floods in the Venice Lagoon. Forecasting up to 9 hours lead time showed good accuracy. Besides using the rainfall and discharge as the input, errors can also be estimated with statistical models. The errors between the forecasted and the observed discharge for the Upper Danube catchment were simulated by ARMAX with the combination of state-space models and wavelet transformations (Bogner and Kalas, 2008). Toth *et al.* (1999) analyzed the performance of using several stochastic models to adjust the forecasts of a

deterministic conceptual model ADM on the Sieve River basin near Fornacina in Italy with hourly rainfall and five years' discharge observations. The stochastic models tested were the linear and stationary stochastic Auto Regressive Moving Average (ARMA) models with the number of parameters ranging from 1 to 3, the fractionally differenced ARIMA model with one parameter and the ARIMA model with an additional parameter taking into account the autoregressive term. The results showed that the stochastic updating methods were effective within a given threshold of the forecast lead time and the use of complex models in the stochastic updating process would not improve efficiency. Shao *et al.* (2009) developed functional-coefficient time series models with periodic variation component by changing the coefficients with the designed periodic function, which possessed an invariance property under data differencing. Barron River in Queensland representing a summer rainfall area, Margaret River in the Western Australia for the winter rainfall area and Murray River in Victoria of no season rainfall area as well as Ying Luo Gorge (Y LX) in Hei River of North-Western China were studied. Compared with the traditional functional-coefficient model, the developed model produced better results in all three climate types. Mohammadi *et al.*(2006) adopted Goal Programming (GP) to estimate the parameters of ARMA model. The proposed model was evaluated with the 68-year measured stream flow data at the Shaloo Bridge station on the Karun River in the south-west of Iran. Compared with the maximum likelihood estimation, the GP method produced better results with increasing AM and MA parameters. The drawback of GP was also shown that the computation cost will be high for a lot of model parameters. Chao *et al.*(2008) proposed a robust recursive least squares (RRLS) with a forgetting factor to estimate the parameters of auto-regressive updating models by inserting a non-linear transformation of the residuals. Compared with the conventional recursive least squares (RLS). Synthetic data and real data from the Weishui reservoir basin in Hubei Province of China were used to test the proposed method. It was found that the effect of the outliers during estimating the parameters

was reduced with the proposed RRLS method. Amiri (2015) compared Threshold Autoregressive (TAR), Smooth Transition Autoregressive (STAR), Exponential Autoregressive (EXPAR), Bilinear Model (BL) and Markov Switching Autoregressive (MSAR) for river discharge forecasting. The optimization of the model parameters were completed with Least Squares (LS) and Maximum likelihood (ML) approaches and the water discharge at River Colorado in U.S.A. was used to evaluate the models. For the in-sample or calibration, three criteria of loglikelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used and the best results were obtained by the self-exciting TAR (SETAR) model. For the out-of-sample forecasting or validation, ten criteria including root-mean-square (RMSE) and 9 other methods were used to evaluate the model forecasting performance and the SETAR model performed the best among all the models.

Another important method used in recursive estimation is the use of filtering methods such as the Kalman filter for estimating state variables of a linear, stochastic-dynamic system (Young and Wallis, 1985). Wilke and Barth (1991) applied Wiener filtering and Kalman filtering at the Upper Rhine for the real time flood forecasting in February 1990. The Wiener filter was used to characterize the response of the catchment and the discharge. The response functions describing the response system were obtained on discharge changes of historical floods. The Kalman filter was used to update the parameters of the ARMA-models. The results showed that the model performance deteriorated with increasing lead time and the updating parameters could not characterize the forecast system. However, oscillation of the forecasted flow around the observations in real time led to the deficiency of the parameter updating with the filter. The extended Kalman filter which can deal with nonlinearities was used to investigate the application on conceptual watershed models (Puente and Bras, 1987). Stochastic noise components were added to both the dynamic and observation equations in the conceptual models so that the states and

observations were transformed to random variables. The catchment used was the Bird Creek basin in Oklahoma (2,344 km²) with discharge data available every six hours. The performance of this nonlinear filter was tested with one-step-ahead (6 hours) and multistep forecasts. The results showed that the extended Kalman filter was able to adjust the noise components instead of using smoothing algorithms which places a higher computational burden. In addition, delay in predictions would occur due to the filter detection of non-predicted rainfall at times of observations. Shamir *et al.* (2010) applied a Monte Carlo-based Kalman filter (Ensemble Extended Kalman Filter, ECKF) to update the states from discharge and reservoir levels together with an event-based storage function method enhanced by involving continuous soil water accounting. A 4,500-km² regulated basin of a tributary of the Nakdong River, Korea with hourly data from January 2006 to August 2008 was tested. The incorporation of ECKF led to improvements of the high flow forecasts, probability of detection and false alarm rate. Gragne *et al.* (2015) used a gain updating scheme (GU-COMP) for a complementary forecasting framework (COMP) which included a semi-distributed conceptual model HBV and error-forecasting models. Hourly flows of the Krinsvatn catchment in Norway were forecasted with the proposed model for a lead time of 24 hours. An adaptive Kalman Filter was used to update the gain coefficients every 24 hours and the root mean square error (RMSE) and the percentage volume error (PVE) metrics were used to evaluate the models. It was found that the proposed GU-COMP reduced the forecasting errors compared with the non-updating model COMP.

Young (2002) applied the data-based mechanistic (DBM) approach that incorporates a stochastic modeling based on the statistical identification and transfer function models as well as an adaptive Kalman filter forecasting algorithm. Of the two types of transfer functions, the nonlinear rainfall-flow component was given more attention than the linear flow-routing model. The formulation of the Kalman filter equations introduced the noise variance ratio (NVR) parameters to specify the stochastic inputs.

This approach was applied to the Hodder Place gauging station on the River Hodder in northwest England using hourly flow for 1993. The Captain Toolbox (<http://www.es.lancs.ac.uk/cres/captain/>) by the Center for Research on Environmental Systems (CRES) at Lancaster provided the estimation procedures and the results showed the ability of the approach to explain the rainfall-flow dynamics, especially in interpretation of the hydrological aspects. Specific steps about the DBM model process can be found in (Beven *et al.*, 2012).

2.2.2.2 Probability-based methods

Silvestro *et al.* (2011) used a probabilistic flood forecasting chain comprising the input from an expert forecaster's precipitation estimation, a rainfall downscaling module and a semi-distributed hydrological model called Discharge River Forecast (DRiFt) model (Giannoni *et al.*, 2000). The forecast chain was tested with the single-site and multi-catchment approaches on a series of events between December 2008 and December 2009 in the Liguria region, Italy. The forecast results were in the form of possible scenarios and probabilities of event occurrence, which was considered effective for flood-alert purposes. Uncertainty analysis is applied in hydrological system usually with Bayesian statistics as the framework. Biondi and De Luca (2012) applied the Bayesian Forecasting System (BFS) to the Turbolo creek catchment (29 km²) in Calabria, Southern Italy. The precipitation forecast was obtained from a stochastic model, Prediction of Rainfall Amount Inside Storm Events (PRAISE) and a distributed rainfall-runoff model was developed for small and medium sized basins. The precipitation uncertainty and hydrologic uncertainty were quantified separately, and integrated into a probabilistic forecast according to Bayesian theory. Even though the predictive distribution of the total uncertainty (2 hours in advance) comprised all the observations within 90% uncertainty bounds, the forecast performance deteriorated when the rainfall forecast was absent for the lead time required. A synthesis of the Bayesian forecasting system is summarized in (Krzysztofowicz, 2002) showing that the asymmetric and bimodal shapes of the predictive probability density

function resulted from many factors such as the nonlinearity of the hydrologic model. Zhang *et al.* (2011) proposed a Markov Chain Monte Carlo (MCMC) framework to enhance the uncertainty analysis of Bayesian Neural Networks (BNNs). Different treatments of input, parameter and model structure uncertainties for BNNs were tested on the Little River Experimental Watershed in the Tifton Upland Physiographic region. The uncertainty estimation was improved under the new framework but the contribution of individual uncertainty sources was hard to identify because of the interactions among these sources. Van Steenbergen *et al.* (2012) produced probabilistic water level forecasts by analyzing the differences between the predictions and observations at the gauge stations with a non-parametric method. The studied hydrological model was from a lumped conceptual NAM model and the hydrodynamic modes were from the Mike 11 river modeling system. The forecast residuals were split into different sets of the simulated time horizons and water levels and three river basins of Demer, Dender and Yser in the Flanders region of Belgium were studied. Compared with a Bayesian method, the proposed non-parametric model produced more realistic confidence intervals. Silvestro and Rebora (2014) set up a series of synthetic experiments for the uncertainty analysis of the effect of forecasted precipitation, initial conditions of soil moisture and ensemble size on probabilistic stream flow forecasts used at the Hydro-Meteoro-logical Functional Center of Liguria Region. The input to the flood forecasting system was from a wide range of synthesized expert quantitative precipitation forecasts (EQPFs) and it was found the errors with the EQPFs were increased of dry conditions of soil moisture and large rainfall events. Chen and Yu (2007) proposed a probabilistic flood forecasting method which combined deterministic forecasts with error probability distribution. The deterministic flood forecasts were from support vector regression and the parameters were calibrated with two-step grid searching and cross validation. The probability distribution of forecasting error was produced by a fuzzy inference system with basic defuzzification distribution (BADD) transformation method and importance

sampling technique. The developed forecasting system was applied to the forecasting with lead times of 1-6 hours in the Lan-Yang River in Taiwan and it was proven that the forecasting uncertainty was steady for different lead times. Reggiani and Weerts (2008) proposed the hydrological uncertainty processor (HUP) based on Bayesian revision of optimizing prior density on river stages considering multiple upstream observations. The proposed HUP was validated at the river Rhine for operational flood forecasting system and applied after running a hydrological and a hydrodynamical model. Measured water levels at the forecasted Lobith at the beginning of the forecasting and the measured water levels at upstream locations were considered as the conditioning variables. It showed that with the upstream measured water levels the prior distribution and the revised posterior were improved. Zhao *et al.*(2015) proposed a Bayesian joint probability (BJP) model to estimate the uncertainty of the forecasts from a deterministic stream flow forecasting system. A parametric variance-stabilizing log-sinh transformation was used to normalize the hydrological data and a bi-variate Gaussian distribution was adopted to calculate the dependence relationship, both of which the parameters were estimated by Bayesian inference with a Monte Carlo Markov chain. Real-time stream flow forecasts from 2004 to 2009 for the Three Gorges Reservoir in China from a deterministic forecasting system depending on upstream stream flow gauge stations and rainfall gauge stations were used for the BJP model. It was found that the forecasting uncertainty increased when the value of forecasts or the lead time increased.

2.2.3 Data-Driven Models

2.2.3.1 Artificial Neural Network

Shamseldin (1997) compared a neural network model with the seasonally based linear perturbation model (LPM) which is a simple linear model (SLM), and the nearest neighbor linear perturbation model (NNLPM). The input information was rainfall, historical seasonal and nearest neighbor information. The number of the neurons in

the input layer and output layers was determined by the numbers of the elements in the input and output arrays, respectively. The number of neurons in the hidden layer as well as the number layers was selected. The value of the weights and the structure of the network preserve the relationship between the input and output data. Thus the model can be used to forecast values for new sets of input data provided the relationship between the input and outputs remain unchanged. The results show that the discharge forecast with the neural networks is more accurate than that of the SLM, the LPM and the>NNLPM methods. Coulibaly *et al.* (2001) improved the performance of the neural network model by using a peak and low flow criterion (PLC) in the selection of the model input parameters. Data from the Chute-du-Diable watershed in northeastern Canada was analyzed and the model was used to forecast peak and low flows with a lead time of 1 week. The accuracy of forecasting extreme hydrologic events was improved with the new criteria. Chen and Chang (2009) presented the evolutionary artificial neural network (EANN) by evolving the network structure with genetic algorithm (GA) and optimizing the connection weights with the gradient algorithm. The inflow measurements from the watershed of the Shihmen reservoir in northern Taiwan were used for testing the forecasting performance. The forecast errors of EANN were decreased compared with those of the AR and ARMAX models. Furthermore, the automatic search algorithm of EANN does not need a predefined architecture that is required in the conventional neural network construction. Wei (2015) compared the lazy learning models of the locally weighted regression (LWR) and the k-nearest neighbor (KNN) model with the eager learning models of the ANN, support vector regression (SVR) and linear regression (REG) for river stage forecasting during typhoons. Fifty historical typhoon events from 1996 to 2007 from the Tanshui River Basin in Northern Taiwan with the forecasting horizon of 1 hour to 4 hours. After calculating the correlation coefficient, mean absolute error, root mean square, significance, computation efficiency and Akaike information criterion, ANN and SVR produced better results than REG in the eager learning

models while LWR was better than kNN in the lazy learning models. But neither of the eager and lazy learning models produced good results. Siou *et al.*(2011) applied a neural network model without rainfall forecasts to the Lez karst Basin in southern France. The model complexity selection was completed with the proposed cross-validation in which preselected the width of around the produced value from the cross-correlation and adjusted the width with cross-validation. It was shown that the designed model produced good discharge forecasts up to 1-day lead time. Elsafi (2014) applied the ANN to forecasting the River Nile at Dongola Station in Sudan, which was located downstream of the junction of the main tributaries to the Nile. Different stations were used as the input the ANN model and the data from 1970 to 1985 were used as the training data with the period from 1986 to 1987 used as the verification data. The root mean square error was used to evaluate the forecasting performance and the input selection of Eddeim, Tamaniat and Atbara produced the best forecasting results. Valipour *et al.*(2013) compared the forecasting performance of ARMA, ARIMA with the static and dynamic artificial neural networks. Monthly discharges from 1960 to 2007 at Dez basin in south-western of Iran were used as the studied area, in which the last five years were used as the validation data. Four and six parameters of the polynomial of the ARMA and ARIMA were tested and the radial and sigmoid activity functions were tested for the neural network. After calculating the root mean square error and mean bias error, the dynamic artificial neural network with sigmoid activity produced the best inflow forecasting while ARIMA produced better results than ARMA for the inflow forecasting of the last twelve months. Kalteh (2013) combined the discrete wavelet transform with ANN and support vector regression (SVR) respectively and compared the developed models with the regular ANN and SVR models. Monthly river flow data from 1966 to 2006 of Kharjegin and Ponal stations in Northern Iran were used to evaluate the model performance. The results showed that the coupling with wavelet transform improved the forecasting accuracy than the regular ANN and SVR models. And SVR model produced better

forecasting results than ANN model in both the coupled form and the regular form. Piotrowski and Napiorkowski (2013) investigated into the techniques to prevent over-fitting problem of ANN, which are noise injection, optimized approximation algorithm and early stopping. The upper part of Annapolis River in Canada up to Wilmot settlement with variant seasonal runoff was used as the studied area. It was found that the noise injection can achieve better results than early stopping at the expense of practical applicability while optimized approximation algorithm did not improve the results. The curse of dimensionality was also proved to impose great effect on the evolutionary computing approaches. Asadi *et al.* (2013) proposed a pre-processed evolutionary Levenberg-Marquardt neural networks (PELMNN) model with the combination of the genetic algorithms and feed forward neural networks. Data transformation, input variables selection and data clustering were used in the data pre-processing and the Aghchai watershed in northwest Iran at Azerbaijan province was used as the studied area with the data from 1995 to 2007. Compared with original ANN and adaptive neuro-fuzzy inference system (ANFIS), the proposed model produced more accurate runoff forecasting.

2.2.3.2 Neuro-Fuzzy Models

Nayak *et al.* (2004) used ANFIS to model the river flow of Baitarani River in Orissa, India. Discharge, selected through cross-validation analysis was used as input to the model. Compared with ANN and traditional ARMA models, ANFIS produced more a precise forecast. Talei *et al.* (2010) investigated the effect of different selection of inputs for the ANFIS model in rainfall-runoff modeling. The inputs used were a sequential rainfall time series, a pruned sequential rainfall time series by using narrower time window, non-sequential inputs of rainfall and the antecedent discharge. The models were tested using data from an outdoor experimental station for different forecast lead times. The results showed that models with only rainfall antecedents performed better at larger lead times, models with antecedent discharge were more accurate at shorter lead time forecasts and models with sequential rainfall antecedents

produced the worst results. The time shift error problem was discussed and the models with only rainfall information were affected less than those that included antecedent discharge information. In the river flow forecast in Dim Stream in the southern of Turkey, the ARMA models were used to produce synthetic flow series for the training data of the ANFIS model (Keskin *et al.*, 2006). Additional monthly mean flow data produced by the AR model was included in the training data set of the ANFIS and compared with an ANFIS model with limited number of observed flows in the training data set. The results showed that this extension of the training data produce significant improvements for the forecasts of both the low and high flows. Nayak *et al.* (2005) tested the performance of the ANFIS model for the river flow forecast of Kolar basin in India. The neural network and fuzzy models were used for comparison. The results show that the ANFIS model outperformed both the neural network and the fuzzy model for lead times of 1 to 6 hours. It was also shown that the 1-hour lead time forecast errors of the ANFIS model were clustered around the rising limb while no clear clusters were observed for ANN and fuzzy model, which may be caused by underestimating the rate of rising limb variation. Similar works can be found in the analysis of rainfall and flow data for Choshui River in central Taiwan (Chen *et al.*, 2006). A back-propagation neural network was used for comparison. It was suggested from the analysis that the persistence effect and upstream flow information affected the forecast the most and the accuracy can be improved by including the watershed's average rainfall in addition to water level data as input. Seo *et al.* (2015) proposed a wavelet-based artificial neural network (WANN) and a wavelet-based adaptive neuro-fuzzy inference system (WANFIS) by using the decomposed approximation and detail components from the time series as the inputs the ANN and ANFIS. The Andong dam watershed in the south-eastern region of South Korea was used as the studied area with the data from 2002 to 2010 as the training data and the data from 2011 to 2013 as the testing data set. The coefficient of efficiency, index of agreement, coefficient of determination, root mean squared

error, mean absolute error, mean squared relative error and the mean higher order error were used as the evaluation indexes. The results showed that with the wavelet decomposition the forecasting accuracy of both the ANN and ANFIS model were improved and the model performance is dependent on input sets and mother wavelets. He *et al.*(2014) compared the ANN, ANFIS and support vector machine (SVM) for the river flow forecasting in a semiarid mountainous Pailugou catchment in Qilian Mountains that is located in northwestern China. Different combinations of the antecedent river flows were tested as the input to the three models and the period from 2001 to 2003 were used as the training data with the period from 2009 to 2011 used as the validation data. After calculating the coefficient of correlation (R), root mean squared error (RMSE), Nash-Sutcliffe efficiency coefficient (NSE) and mean absolute relative error (MARE), reliable and accurate forecasting ability of the three models were proved and the SVM produced a bit better forecasting than the other models for the validation data. Badrzadeh *et al.*(2013) coupled the wavelet multi-resolution analysis with ANN and ANFIS for river flow forecasting of the Dingo road station in the Harvey River catchment in Western Australia with the period from 1972 to 1999 used as training data and the period from 2000 to 2011 used as validation data. The river flow and rainfall time series were decomposed into 3,4 and 5 levels of resolution with Daubechies and Symlet wavelets, of which different time lag combinations of the wavelet coefficients were used as the inputs to ANN and ANFIS model. The hybrid models improved the peak values and forecasting performance of the longer lead time than original ANN and ANFIS models. Talei *et al.* (2013) compared a local learning neuro-fuzzy system (NFS) with physically-based models Kinematic Wave Model (KWM), Storm Water Mangement Model (SWMM) and Hydrologiska Byråns Vattenbalansavdelning (HBV) model for rainfall-runoff modeling. A small experimental catchment, a small urbanized catchment in Singapore and a large rural watershed in Rönne basin were used to evaluate the model performance. It was proved that the proposed local learning model produced better

forecasting results with shorter training time. With the real-time mode, the local learning model can produce adapted forecasts without retraining the model. Nourani and Komasi (2013) proposed an Integrated Geomorphological Adaptive Neuro-Fuzzy Inference System (IGANFIS) coupled with the fuzzy C-means (FCMs) method. The Eel River catchment in California of USA was used as the studied area and the spatial and temporal variables of the sub-basins were used as the inputs to the proposed model. After calculating the index such as the Determination of Coefficient (DC) and root mean squared error (RMSE), the proposed model was proved to provide successful spatiotemporal hydrological modeling. Ghose et al.(2013) used the Non-Linear Multiple Regression (NLMR) and ANFIS to forecast the runoff during monsoon period of the basin upstream of Basantpur station of Mahanadi river in India. It was found the developed models can provide daily forecasts with the historical information the ANFIS model with three inputs were better than that with two inputs. By coupling the Genetic Algorithm (GA) with the NLMR model, the optimum process parameters for the maximum runoff were identified. Lohani *et al.*(2012) used autoregressive model (AR), ANN and ANFIS to forecast monthly reservoir inflow and the time series river flow of the Sutlej River at Bhakra Dam in India was used to evaluate the model performance. It was found that including the cyclic terms of monthly periodicity in addition to previous inflows in the input vectors to ANFIS model improved the forecasting accuracy and ANFIS model always produced better predictions than the AR and ANN models.

2.3 Ensemble Methods

The ensemble approach has been used in weather predictions and for a limited number of applications in flood forecasting.

2.3.1 Ensemble Methods in Weather Forecasting

Ensemble methods have been used in fields related to climate modeling. Hargreaves and Annan (2006) used an ensemble Kalman filter in predicting the response of the

North Atlantic thermohaline circulation (THC) to anthropogenic forcing. The ensemble members with different parameters setting were tuned and reliable medium term (100 years) forecasts were achieved. Höglind *et al.* (2012) applied a multi-model ensemble of 15 global climate models to develop local-scale climate scenarios for assessment of climate change impact on grass species in Northern Europe. With this ensemble model, the climate prediction uncertainties of the initial conditions, model parameters and structural variations of the global climate models were quantified. In risk assessment of biodiversity, Fordham *et al.* (2011) ranked commonly used atmosphere-ocean general circulation models (AOGCMs) according the performance of simulating the observed climate patterns. Without considering models' performance, the models were allocated with equal weights to produce averaged ensemble forecasts, which globally outperformed the forecasts of individual models. Yang and Wang (2012) compared Bayesian model averaging (BMA) ensemble with equal weight ensemble in reducing biases in regional climate downscaling. Both the models can reduce the circulation biases but only BMA ensemble was able to minimize the biases of precipitation. In weather forecasting, because of the nonlinearities from the extreme events, model bias exists even in a perfect model. The National Center for Environmental Prediction (NCEP) ensemble forecast system (Toth and Kalnay, 1997) has been used to produce ensemble forecasts with different initial perturbations. Goerss (2000) used a simple ensemble average method in forecasting the cyclone tracks for the 1995-1996 Atlantic hurricane seasons and the western North Pacific in 1997. The ensemble predictions reduced the standard deviation of the forecasting error and the 95th percentile of forecasting error than the best of the individual models. Yuan *et al.* (2009) used a time-lagged multi-model ensemble forecast system initialized with two mesoscale models in investigation of quantitative precipitation forecasts (QPFs) and probabilistic QPFs. The combination was proved to reduce forecast biases and more models constructing the ensembles were suggested.

2.3.2 Ensemble Methods for Optimization in Water Resources

Shabani (2009) applied the reinforcement learning algorithm to a multi-reservoir optimization problem, aimed at finding a trade-off between the present reward and the future value of the stored water in the reservoirs, which had to be within some operating rules and flood control constraints. The main sources of uncertainty in this large hydropower system were market prices and inflows. This study showed that stochastic large-scale problems can be solved within acceptable time and computation cost with the reinforcement learning algorithm. Similar application of reinforcement learning in multi-reservoir systems optimization problem was used by Lee and Labadie (2007). The two-reservoir Geum River system in South Korea was tested with the Q-Learning method in reinforcement learning. The results reflected the superiority of the Q-Learning over the operating rules by implicit stochastic dynamic programming and sampling stochastic dynamic programming (SSDP). Bhattacharya *et al.* (2003) used artificial neural networks and reinforcement learning with the Aquarius Decision Support System (DSS) to build a controller for multi-criteria optimization problems. The error of the ANN component was minimized by reinforcement learning and the water systems in the Netherlands were used for the test. The results showed the improvement of the computation accuracy when the hybrid algorithm combining ANN and RL was used in dynamic real-time control operations.

2.3.3 Ensemble Methods in Flood Forecasting

Recently, limited studies using ensemble models for the forecasting of river discharge have been attempted. Chang *et al.* (2010) used a Clustering-Based Hybrid Inundation Model (CHIM) to combine linear regression with neural networks. K-means clustering method was used in the data preprocessing stage to identify the cluster centers as the control points. In the model building stage, three options are available: the back-propagation neural network (BPNN) forecasting model is built for control

points, the linear regression model is built for those grids linearly correlated with control points, and a multi-grid BPNN is used for other grids in the study area. The ensemble results in a flood inundation map. This hybrid method was tested in the Dacun Township, Changhua County, Taiwan and the results showed that predictions of the inundation were accurate up to lead time of 1 hour. Goswami and O'Connor (2007) compared three ensemble techniques for a French and an Irish catchment. Five individual models: autoregressive (AR) model, NN models with observed data for direct and recursive multi-step forecasting, NN models with departures of observed data for direct and recursive multi-step forecasting, parametric simple linear model using rainfall as leading indicator (P-SLM-LI) and parametric linear perturbation model using rainfall departure as leading indicator (P-LPM-LI) were used in the ensemble. Three model ensemble techniques: simple average method (SAM), the weighted average method (WAM) and the neural network method were studied. In the neural network method, the forecasts of individual models were treated as inputs to the network and the weights were optimized by the Simplex search method. All of the combination methods produced better forecasts than those of the individual models. A weighted average method (WAM) was applied for the flow forecasting of the Blue Nile River. The models used were the Linear River Flow Routing Model (LRFRM) and the Neural Network River Flow Routing Model (NNRFRM). LRFRM is a non-parametric linear storage model and NNRFRM has the same structure as a multi-layer feed-forward neural network. An autoregressive model error updating procedure was used to update the ensemble forecasts by forecasting the errors in the simulation-mode forecasts (non-updating with recently observed discharges), the final updated real-time forecasts (sum of the simulation-mode discharge) and the error predictions. The weights were estimated with the ordinary least squares method, which minimized the sum of squares of the differences between the forecasts and the observed value. The results showed the best individual model was assigned the highest weight and the performance of the ensemble model was almost the same as

the best individual model (Shamseldin and O'Connor, 2003). Chen *et al.* (2015) studied a weighted average method as the multi-model combination (MC) together with the autoregressive method used for flood error correction (FEC). Based on the sequence of applying MC or FEC, three combination methods were created including FEC-MC, MC-FEC and global real-time combination method (GRCM) which calibrated and optimized the two methods together. The rainfall-runoff predictions were from four models Xinanjiang model, Tank model, Artificial Neural Network and antecedent precipitation index based model. After validating the combination results on Three Gorge Reservoir and Jinsha River in China, the GRCM method produced the best combination results based on the evaluation of the root-mean-square error, NSE and qualified rate.

The total prediction uncertainty of the ensemble approach can be calculated with probabilistic methods based on the likelihood measures of each models. The weights of individual models reflect the probability that a model can predict correctly with the observation data (Duan *et al.*, 2007). See and Openshaw (2000) compared four ensemble approaches to integrate the predictions of individual models with historical time series data from the River Ouse in Northern England. The combined models were a hybrid neural network, a fuzzy logic model, an ARMA model and naïve predictions. The ensemble approaches used were the simple average method, a Bayesian approach which selected the best performing model at the last time step to make the prediction at the next time step (Der Voort and Dougherty, 1996), a fuzzy logic model based only on current and past flow information and fuzzification of the crisp Bayesian method which was able to recommend more than one good model or any combination of models at each time step. The results showed that the fuzzified Bayesian model produced better overall results than other hybrid approaches and individual approaches based on the global evaluation measures. A Bayesian Model Averaging (BMA) scheme was applied by Duan *et al.* (2007) on three test basins in

the United States. Three conceptual hydrological models: the Sacramento Soil Moisture Accounting (SAC-SMA) model, the Simple Water Balance (SWB) model and the HYMOD model were combined with three objective functions, which resulted in a nine-member ensemble. The three objective functions: daily root mean square error (DRMS), heteroscedastic maximum likelihood estimator (HMLE) and Daily absolute error (DABS) were selected to place emphasis on the high flows, low flows and all parts of the hydrograph, respectively. Before the application of BMA, the conditional probability distribution was assumed as Gaussian by using the Box-Cox transformation and log-likelihood function was applied for computational convenience. The Expectation-Maximization (EM) algorithm was used to obtain the solutions of the probabilistic predictions through iterative procedures. A reasonable empirical probability density function (PDF) for each time step was obtained with 100 BMA ensemble predictions. The results showed that the BMA scheme produced better predictions compared with the best individual predictions on daily root mean square error and daily absolute mean error. In addition, the BMA predictions with multiple sets of weights (BMA_m) improved the ability of individual models to characterize different parts of the hydrograph.

Araghinejad *et al.* (2011) used a probabilistic method based on the nonparametric K-nearest neighbor regression to combine the forecasts of artificial neural networks with various training algorithms, which produced specific networks with better estimation ability. A performance function that is flexible to measure the goodness of fit at different hydrological values was introduced to create ensemble members. The data used were from the Red River in Canada and Zayandeh-rud River in Iran. Compared with other models such as K-nearest neighbor regression, multilayer perceptron network (MLP) and MLP trained by the Peak and Low flow Criteria (PLC) performance function, the ensemble method produced better point estimation forecasts in terms of performance statistics in training and validation data sets. Xiong

et al. (2001) used the first-order Takagi-Sugeno fuzzy system to combine the forecast results obtained from five different conceptual models on eleven catchments. Compared with other combination methods of the SAM, the weighted average method (WAM) and the neural network method (NNM), the results of the first-order Takagi-Sugeno method showed similar efficiency in improving the forecasting accuracy. However, correlation analysis showed that the combination results were greatly influenced by the best individual model in combination and the number of the if-then rules should be carefully designed to prevent over-parameterization. Kasiviswanathan *et al.* (2013) created a group of rainfall- runoff forecasts for an Indian basin by randomly perturbed the current artificial neural network parameters with respect to multiple optimization objectives. With 97.17% of the observed validation data lying in the spread of the ensemble forecasts, the mean value of the forecasts improved the accuracy of the peak forecasting. Rathinasamy *et al.* (2013) developed a new multi-wavelet based ensemble method which used Bayesian Model Averaging (BMA) to combine the forecasts from the wavelets with different properties for the river flow of two stations in the USA at daily, weekly and monthly scales. Improved BMA ensemble results were achieved compared with the single best wavelet model and the mean averaged ensemble. Erdal and Karakurt (2013) employed bagging and stochastic gradient boosting to generate 100 randomly drawn replica subsets of the learning data for a classification and regression trees (CARTs) model and the outputs of the multiple CART models were linearly combined for the final forecasting. The ensemble forecasting was tested with a 35-year measured data of the observation station on the Çoruh River in Turkey against a support vector regression (SVR) model used as benchmark. The monthly stream flow predictions were remarkably improved by using the ensemble learning.

The shortcomings in applying a single specific model to flood forecasting has been highlighted in this review. While it is generally accepted that no one single model can

claim to provide accurate predictions for all the stages of the hydrological process, the review revealed promising results obtained by the ensemble approach. Although there are benefits from the use of a combined models, but number of studies adopting the ensemble modelling approach in flood forecasting have been limited. Further studies using especially in the use of different ensemble approaches, such as modern computational intelligence tools, are therefore needed.

CHAPTER 3 METHODOLOGY AND DATA USED

3.1 Introduction

The Neuro-Fuzzy Inference System (NFIS) modeling approach is adopted as the ensemble model for this study. The NFIS model benefits from the learning ability of the neural network and the reasoning ability of fuzzy logic inferencing. Specifically, the NFIS models used in this study are the Adaptive-Network-Based Fuzzy Inference System (ANFIS) and Dynamic Evolving Neural-Fuzzy Inference System (DENFIS). Both of these models have been used in hydrological applications and are reported in the literature. While both of these models have the potential to be used either in batch or incremental learning mode, the ANFIS model, which has a longer history, is predominantly a batch model whereas DENFIS adopts an incremental learning algorithm. The incremental learning algorithm represents an advancement over the batch approach since incremental learning is amenable to real time applications, which is an important requirement in flood forecasting. Thus, the advantage of DENFIS over ANFIS is DENFIS allows for the flexibility of adjusting the parameters based on model inputs in real time for online applications. The novelty of exploiting incremental learning in the NFIS model thus lies in the ability of the ensemble model to be trained or adapted in real time (online), thus improving results.

3.2 Simple Average Method

SAM can be the simplest ensemble approach that only averages the predictions from individual models (Shamseldin *et al.*, 1997):

$$\hat{Q}_{c_i} = \frac{1}{N} \sum_{j=1}^N \hat{Q}_{j,i} \quad (3.1)$$

where j is the index of the individual model, N is the number of the combined models, $\hat{Q}_{c_i}$ is the combined estimate of the discharge at the i^{th} time period. The errors of a

model can be decomposed to the bias component representing the difference between the simulated values and the true values and the variance component describing the sensitivity of the model to input data (Bishop, 2006). SAM can reduce the variance of the ensemble models leading to improved predictions, especially for those models with low bias but high variance. In this study, the SAM is used as benchmark model, as this method is often adopted as a default ensemble approach (Goswami and O'Connor, 2007; See and Openshaw, 2000; Xiong *et al.*, 2001; Kasiviswanathan *et al.*, 2013).

3.3 Adaptive-Network-Based Fuzzy Inference System

The Adaptive-network-based fuzzy inference system or ANFIS (Jang, 1993) is suitable for modeling nonlinear systems and implements a fuzzy inference system within the structure of a neural network. A schematic of the ANFIS model is shown in Figure 3.1. In Figure 3.1, the first layer receives the external crisp signals and performs the fuzzification and the fuzzy member functions are contained in this layer. The second layer is the rule layer and receives the membership results from the previous layer. Then the firing strength for each neuron or rule is obtained by the operator product:

$$W_i = \mu_{A_i} * \mu_{B_i} \quad i = 1, 2 \quad (3.2)$$

where W_i is the firing strength of each rule and μ_{A_i} and μ_{B_i} are the degrees of membership. The third layer is the normalization layer and the fourth layer calculates the first-order Takagi-Sugeno rules for each rule. The fifth layer is the summation of the weighted output of the whole system. This process that produces a crisp output is

called defuzzification. Thus in Figure 3.1 (a), two fuzzy rules are formed to process the initial inputs x and y and the final results are obtained by an average weighting of the output of each rule and the ANFIS structure is shown in Figure 3.1(b). The backpropagation or gradient descent method is used to modify the membership function (premise parameters) and the least mean square error algorithm is adopted to identify the rule linear combination parameters (consequent parameters).

(a)

(b)

Figure 3.1 (a) Type-3 Fuzzy Reasoning (TS fuzzy if-then rules are used). (b) Equivalent ANFIS or Type-3 ANFIS (Jang, 1993).

3.4 Dynamic Evolving Neural-Fuzzy Inference System

The structure of the Dynamic Evolving Neural-Fuzzy Inference System or DENFIS (Kasabov and Song, 2002) is similar to ANFIS; however, DENFIS uses an online learning algorithm. The Evolving Clustering Method (ECM) is used to reach a dynamic estimation of the number of the clusters and the cluster centers in the input space. The parameters of the model include the maximum distance (MaxDist) between an example point and the center point and the threshold of the distance (D_{thr}) will affect the number of the clusters. The ECM is illustrated schematically in Figure 3.2. The initial condition is an empty set of clusters with an input data stream. When the first input point comes into the input space, a new cluster is formed with the radius set to zero. For subsequent data patterns, three scenarios are possible: (i) if the distance between the new point and the existing cluster centers D_{ij} is not more than at least one of the radius R_{uj} , the point will be clustered into the existing cluster with the minimum distance, (ii) after the calculation of the index $S_{ij} = D_{ij} + R_{uj}$, the minimum of the S_{ij} is found to be S_{ia} . If the S_{ia} is larger than twice the threshold D_{thr} , a new cluster is created, and (iii) if the S_{ia} is not greater than twice D_{thr} , the cluster C_a will be updated by moving its center and increasing its radius. The new radius is set to be $S_{ia}/2$ and the new center is located on the line connecting the new point and the center of the cluster C_a with the distance of the new radius from the new center to the new point. Besides the online mode, DENFIS can also be used in offline mode with offline ECMc which will optimize the cluster centers produced by ECM. The results of offline ECMc are obtained by applying global optimization to the resulted clusters of ECM. New clusters centers are found in order to minimize an objective function that is subject to given constraints. The cluster centers are initialized by ECM clustering and a binary membership matrix can be calculated according to:

$$\begin{aligned} \text{If } \|x_i - C_{c_j}\| \leq \|x_i - C_{c_k}\|, \text{ for each } j \neq k; \\ \mu_{ij} = 1, \text{ else } \mu_{ij} = 0 \end{aligned} \quad (3.3)$$

where x_i is the i^{th} data point and C_{c_j} are cluster centers, μ_{ij} is the element of the membership matrix. An objective function subject to given constraints is minimized for a vector x_i in cluster j with the cluster center C_{c_j} :

$$J = \sum_{j=1}^n J_j = \sum_{j=1}^n (\sum_{x_i \in C_j} \|x_i - C_{c_j}\|) \quad (3.4)$$

$$\text{Subject to } \|x_i - C_{c_j}\| \leq Dthr$$

where $i=1,2,\dots,p$; $j=1,2,\dots,n$

The constrained minimization method is employed for the objective function to obtain new cluster centers. The cluster centers, membership matrix and new objective function parameters are updated iteratively until improvement is within a threshold or the iteration reaches a certain limit. In this way, the summation of the distance from every point to its corresponding cluster is minimized after ECM is processed. Then the nearest several clusters to the input vectors will be used to form the basis of the fuzzy rules group and the rules in these clusters are used to process the input vectors. The first order Takagi-Sugeno (Takagi and Sugeno, 1985) inference procedure which adopts a linear function was used as the consequent function. The firing strength of cluster i or the degree to which the rules are applicable for the inputs is calculated by:

$$\omega_i = \prod_{j=1}^q \mu_{R_{ij}}(X_j) \quad (3.5)$$

where $j = 1,2,\dots,q$. $\mu_{R_{ij}}(X_j)$ is the membership function for input data $\mathbf{X}$ with q dimensions. The parameters in a fuzzy rule include the weights of the input dimensions and a constant term (Eq. 3.6). The final output is a weighted average of the clusters in the fuzzy rules group based on each firing strength. The parameters

updating algorithm is a weighted recursive least-square (RLS) with forgetting factor:

$$y = w_0 + w_1X_1 + w_2X_2 + \cdots + w_qX_q \quad (3.6)$$

where y is the output of a fuzzy rule, w_i is the weight for each input dimension $X_i(i=1,2,\dots,q)$ and w_0 is the constant term.

Figure 3.2 Clustering process using ECM in 2-D space (a)x1: the creation of the first cluster 1(b) x2: update cluster 1; x3: creation of a new cluster 2; x4: belongs to cluster 1 (c)x5: update cluster 1; x6: belongs to cluster 1; x7: update cluster 2; x8: creation of a new cluster 3 (d)x9: update cluster 1 (Kasabov and Song, 2002)

3.5 Study Areas

The review of the ensemble modelling in Chapter 2 indicates two general cases of

ensembling in real-world applications: (i) Different rainfall-runoff models with the same rainfall inputs and (ii) Different rainfall inputs but with the same rainfall runoff model. The Lower Mekong Basin is an example of the former scenario where several models have been developed in the past to provide up to 5 days lead time water level forecasts, but no attempt has ever been made to resolve the issue of the disparity between model results. The latter case concerns the uncertainties associated rainfall forecasts largely resulting from the use of atmospheric models to derive rainfall forecasts. For the Taiwan catchment, uncertainties in the weather model results in 15 possible scenarios for rainfall forecasts on a real-time basis. The Taiwan catchment was thus used to explore the ensembling procedure in this particular application.

3.5.1 Lower Mekong Basin

The Mekong River has a watershed of size 795,000 km² and is the 10th largest river in the world. The Mekong River flows through China, Myanmar, Laos, Thailand, Cambodia and Vietnam and can be divided into the upper part (about 2,200 km) and the lower part (2,700 km). The lower Mekong Basin with 600,000 km² area varies in climate and geography conditions. The flow of the Lower Mekong River comes mainly from the tributaries to the east of the Mekong between Luang Prabang and Kratie (www.mrcmekong.org). Fig 3.3 (b) shows the sub-basin containing the Kratie gauging station used in Chapter 4. Nguyen and Chua (2012) applied ANFIS to forecast the daily water levels at Pakse from 1 to 5 days ahead during the wet seasons from 1993 to 2003 and for 2009. The daily water levels at Pakse and an upstream station (Savannakhek) were used as inputs. Water levels at the current and two previous time steps were identified, along with average rainfall in sub-basins between Thakhek and Pakse, as inputs of ANFIS for Pakse and Savannakhek after correlation analysis.

Figure 3.3 (a) Location map for the Mekong Basin; (b) Sub-basin with gauging station Kratie: 28,815 (km²). Source: (MRC, 2005; 2007)

The ANFIS model used by Nguyen and Chua (2012) to forecast the water level at Pakse and the results from ANFIS were compared with forecasts by the Unified River Basin Simulator (URBS) hydrological model (Carroll, 2007), which is a lumped parameter model, adopted to provide operational forecasts. Forty nine sub-models of Mekong River Basin were created in URBS by combining rainfall-runoff modelling which converts the gross rainfall into excess rainfall and a runoff-routing modelling which calculates the flow based on the inputs of the excess rainfall (Tospornsampan *et al.*, 2009). The daily historical hydrological data were provided by the Mekong River Commission (MRC) for model calibration (MRC, 2009). The 5-day ahead water level predictions of ANFIS and URBS for Thakhek, Pakse and Kratie stations during the wet seasons of 2009 to 2011 were used for Chapter 4 and Chapter 5.

3.5.2 Lanyang Creek Basin, Taiwan

Taiwan is vulnerable to typhoons and the floods that occur as a result. With the steep

topography and large amount of the rainfall brought by typhoons, life and property has been under the threat of the resulting floods and other natural disasters, such as landslides and debris flow. Building an accurate flood forecasting system is necessary for policy makers and for effective flood management. The short hydraulic response time that results from the steepness of the landscape makes it impractical to predict floods with ground observations of rainfall (Shih *et al.*, 2014). Hence, the use of precipitation forecasts can be a practical approach for flood forecasts. However, precipitation forecasts are sensitive to small errors in the initial conditions because of the chaotic property of the atmospheric system, which results in uncertainties in the precipitation forecasts. The use of multiple precipitation forecasts with perturbed initial conditions is thus adopted to reduce the uncertainties (Hsiao *et al.*, 2013). Shih *et al.* (2014) have developed a forecast model to provide operational flood forecasts at the Lanyang Bridge, Yilan County, Taiwan. In the operational model, flood forecasts are predicted corresponding to the precipitation forecasts with widely varying water level estimates due to large differences in the precipitation forecasts.

The location of Lanyang Creek basin is shown in Figure 3.4. The basin is hilly and steep and covers approximately 652 km² mountainous terrain and 978 km² in total with an average river bed slope of 1/55, resulting in a short hydraulic response time.

Figure 3.4 Lanyang Creek Basin (Shih *et al.*, 2014)

Rainfall forecasts are provided by the Weather Research Forecasting (WRF) Model (Skamarock *et al.*, 2005). A total of fifteen forecasts under different perturbed initial and boundary conditions of the atmospheric states and cumulus scheme (Hsiao *et al.*, 2013) are obtained from the WRF model, and the results of the WRF model are used to generate basin flows using the WASH123D (Yeh *et al.*, 1998) distributed model which is a multi-process model suitable for watershed hydrology. The river and overland diffusive wave equations were used to calculate the watershed flows with the semi-Lagrangian and Galerkin finite-element methods. The finite element method is used in the WASH123D model at different temporal and spatial scales to simulate river networks, overland regime and subsurface media. Calibration of the

WASH123D model for Lanyang Creek Basin is reported in (Shih *et al.*, 2014). Numerical runs were started 72 hours ahead of time. However eight hours is required to complete the WASH123D model runs for all the WRF rainfall forecast, therefore, 64 hours ahead water level forecasts can be obtained by the system. Operationally, the water level forecasts are updated at 6-hourly intervals.

In all, 15 water level forecasts, derived from the 15 different perturbed initial conditions in the rainfall forecasts were used. The data for the period from 11th May 2012 to 3rd September 2013 was used for training the model with the data for the period from 21st September 2013 to 24th September 2013 and the data from 22nd July 2014 to 26th July 2014 was used as the test data. In order to avoid over-fitting, the first $\frac{2}{3}$ of the training dataset was used to determine the model parameters and the second $\frac{1}{3}$ used for validation. Taiwan catchment was used for Chapter 6 and Chapter 7.

3.6 Error Analysis

Model performance was quantified as root mean square error (RMSE),

$$RMSE = \left(\frac{1}{n} \sum_{t=1}^n (q_o(t) - q_s(t))^2 \right)^{1/2} \quad (3.7)$$

percent error in peak (PEP) (Green and Stephenson, 1986),

$$PEP = (q_{ps} - q_{po}) / q_{po} * 100\% \quad (3.8)$$

percent bias (PBIAS) (Gupta *et al.*, 1999),

$$PBIAS = \sum_{t=1}^n (q_o(t) - q_s(t)) / \sum_{t=1}^n q_o(t) * 100\% \quad (3.9)$$

Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970),

$$NSE = 1 - \sum_{t=1}^n (q_s(t) - q_o(t))^2 / \sum_{t=1}^n (q_o(t) - q_{mean})^2 \quad (3.10)$$

percentage error at peak flow (PE),

$$PE = (q_s(t_p) - q_{po}) / q_{po} * 100\% \quad (3.11)$$

peak time difference (PT)

$$PT = t_s - t_p \quad (3.12)$$

mean relative difference (MRD) (Lasserre *et al.*, 1999)

$$MRD = 100 * 1/n * \sum_{t=1}^n (q_s(t) - q_o(t)) / q_o(t) \quad (3.13)$$

mean absolute relative difference (MARD) (Stevens *et al.*, 1983)

$$MRD = 100 * 1/n * \sum_{t=1}^n |(q_s(t) - q_o(t)) / q_o(t)| \quad (3.14)$$

Mahalanobis distance (Dmah) (Lhermitte *et al.*, 2011; Mahalanobis, 1936)

$$Dmah = \sqrt{(q_s - q_o)' * \Sigma^{-1} * (q_s - q_o)} \quad (3.15)$$

and overall inconsistency (OI)

$$OI = MRD * RMSE / MARD * Dmah^2 / n \quad (3.16)$$

where $q_s(t)$ is the simulated water level at time t , $q_o(t)$ is the observed water level at time t , q_{mean} is mean of the observed water levels, q_{ps} is the simulated peak water level, q_{po} is the observed peak water level, t_s is the time of the simulated peak water level and t_p is the time of the observed peak water level, Σ is the error covariance matrix, were used to evaluate the model performance. PEP indicates the error of the maximum water levels in percentage and PE describes the percentage error of the water levels at the peak time. PEP and PE are positive when water levels are over-estimated while the negative values indicate under-estimation. Positive values of PBIAS indicated underestimation of the average tendency. PT describes the time difference of the peak between the predictions and the observed water levels, where negative values depict early warning. The OI between two perfectly consistent time series will be around 0.

The deductive approach was first adopted to test the performance of NFIS models as the ensemble approach for two selected catchments to validate the potential of the NFIS in ensemble flood forecasting. Then more analysis will be done with the inductive approach to look for the patterns of the ensemble flood forecasting based on NFIS model. A generalization of the ensemble approach will be achieved from analyzing the model performance for the two practical flood forecasting scenarios.

Table 3.1 Inductive and deductive approaches

Inductive approaches	Deductive approaches
Generalization, statistical syllogism, simple induction, argument from analogy, causal inference, prediction	Syllogism, contrapositive, detachment

Detachment was used as the deductive approach which tested the performance of the NFIS model combining the ANFIS and URBS model forecasts under the premise that the NFIS model can provide improved ensemble forecasts. Generalization and argument from analogy were used as the inductive approaches to generalize the patterns of the ensemble forecasting for ANFIS and URBS models from Lower Mekong and perturbed flood forecasts from Taiwan to similar scenarios.

CHAPTER 4 WATER LEVEL FORECASTING FOR THE LOWER MEKONG USING A NEURO-FUZZY INFERENCE SYSTEM ENSEMBLE APPROACH

4.1 Introduction

Even though there have been promising results with the ensemble approach as reviewed in Chapter 2, the number of studies into this approach is limited and more research with more flexible and reliable ensemble models is needed. In this Chapter, results of experiments using the ensemble approach for a flood forecasting application for the Lower Mekong are explored and presented. The measured water level data from the Lower Mekong and the 5-day ahead predictions from a physically-based model, the URBS hydrological model and a data-driven model, the ANFIS model (Nguyen and Chua, 2012) were used in the analysis. The ANFIS and URBS model performance was evaluated for the 2009 wet season using the coefficient of efficiency, mean percentage absolute error, mean absolute error and RMSE. These comparisons showed that ANFIS outperformed URBS model results for one- to three-lead-day forecasts. Comparison of the five-lead day results showed that improvements of the ANFIS model over URBS was not as significant. Outputs from the URBS and ANFIS models were analyzed with two NFIS ensemble models in order to ascertain their capability in improving flood forecasts using the ensemble approach. The first NFIS ensemble model is an ANFIS model, denoted as ANFIS-EN, which adopts a batch learning approach. The second ensemble method is the DENFIS proposed by (Kasabov and Song, 2002), denoted as DENFIS-EN, which employs an incremental learning approach. For ANFIS-EN, the number of fuzzy sets and rules need to be specified manually before training the model while DENFIS-EN uses a clustering algorithm to specify the number of clusters or rules from the learning process of the training data. The adoption of these two NFIS models as the ensemble approach first showed the difference between the user specified NFIS structure from

ANFIS-EN and a more flexible and dynamic clustering process for the FIS structure from DENFIS-EN. Then the benefits from the incremental learning of DENFIS-EN will be addressed in next chapter. The over-fitting issue where trained models may fail during testing due to improper data division is also considered, since this is rarely discussed for ensemble modeling.

4.2 Methodology

4.2.1 ANFIS Ensemble Model (ANFIS-EN)

The ANFIS model was adopted as the first ensemble model and modified for the ensemble approach, which was named after ANFIS-EN model. In initial trial and error experiments, it was found that having too many rules will lead to an increase in forecast error. Thus, even though the model can simulate the training data well, too many parameters in the rule base can cause over fitting to the training data. In this study, the number of rules in ANFIS-EN was kept the same as the same number of the membership functions based on the assumption that the predictions from the two individual models will not be too different (Xiong *et al.*, 2001). Three membership functions (high, medium, low) and two membership functions (high, low) were investigated in the analysis. Corresponding to the reduced membership functions, the rules were pruned by setting the ANFIS-EN model with three rules (high-high, medium-medium, low-low) and two rules (high-high, low-low), respectively. The structure of ANFIS-EN with three rules is shown in Figure 4.1. For the ANFIS-EN with three membership functions and three rules, the model contains nine linear parameters with 18 nonlinear parameters. Each rule in Figure 1 indicates the strategy to combine the inputs for the three conditions when both the models give high, medium and low predictions. With this modification of the ANFIS structure, the number of the fuzzy rules is decreased from nine to three and less parameters in the reduced membership functions and fuzzy rules will need to be optimized.

Figure 4.1 The structure of the ANFIS-EN with three membership functions

The forecasts from URBS model ($X1$) and ANFIS model ($X2$) are used as the inputs for the ANFIS-EN ensemble model as shown in Figure 4.1. The first layer receives the external crisp signals and performs the fuzzification and the fuzzy member functions are contained in this layer. The second layer is the rule layer and receives the membership results from the previous layer. Then the firing strength for each neuron or rule is obtained by the operator product of the membership results. The third layer is the normalization layer and the fourth layer calculates the first-order Takagi-Sugeno rules for each rule. The fifth layer is the summation of the weighted output of the whole system. The outputs of ANFIS-EN were compared with the measured water level data and the errors will be used to provide feedback for parameters optimization. In this study, grid partitioning is used in ANFIS-EN to generate FIS because of the applicability to fuzzy control with several state variables (Jang, 1997). The hybrid method of the back-propagation gradient descent method and least squares estimation is used as the optimization method. The membership function used is set as generalized bell curve membership function and a first-order

TSK model is used to derive the NFIS model output. The structure and the number of membership functions or fuzzy rules need to be defined before training the ANFIS-EN without considering the characteristics of the training data.

4.2.2 DENFIS Ensemble Model (DENFIS-EN)

DENFIS (Kasabov and Song, 2002) uses a clustering algorithm the ECM to dynamically estimate the number of the clusters and centers in the input space, which was used as the second ensemble model. In ANFIS-EN the number of fuzzy sets and fuzzy rules is required to be specified before training model. In DENFIS the number of fuzzy sets and fuzzy rules is determined during training by ECM. The offline mode of DENFIS was used in this chapter which optimized the clustering results. The details of DENFIS can be referred to Chapter 3.4. Thus, the ECM clustering algorithm in DENFIS considers the property of the input data when creating the clusters. The offline mode of DENFIS optimizes the clustering results of input data and was employed as the ensemble approach to improve the water level forecasts and denoted as DENFIS-EN model.

4.2.3 Data Used

The data used for this Chapter was obtained from the Lower Mekong Basin (Chapter 3.5.1). The 5-day ahead water level predictions of ANFIS and URBS for Kratie station during the wet seasons of 2009 to 2011 were used as inputs to the ANFIS-EN and DENFIS-EN ensemble models. The data used in the analysis are shown in Figure 4.2 as a time series, where it is observed that the predictions of ANFIS and URBS models behave rather differently. From Figure 4.2, the predictions from both the model URBS and ANFIS models can follow the observed trends in the water level for the three events quite well. However, the URBS model predictions are more spiky and over-predicts the peak water level, especially for 2009, while the ANFS model appears to be prone to time-shift errors, evident from the falling and rising phases of the 2010 and 2011 data, respectively.

(a)

(b)

(c)

Figure 4.2 Comparisons of Measured water level, ANFS predictions and URBS predictions for Kratie station (Nguyen and Chua, 2012) of (a) 7th Jun 2009 to 31st Oct 2009, (b) 7th Jun 2010 to 31st Oct 2010 (c) 20th Jun 2011 to 20th Oct 2011

Based on the usage of a validation data set, the ensemble approaches were adopted with or without validation. For the ensemble approaches with validation, the ANFIS and URBS predictions as well as the measured water levels from 2009 to 2011 were divided into three subsets: the training data are from 7th Jun 2009 to 31st Oct 2009, the validation data from 7th Jun 2010 to 31st 2010 and the testing data from 20th Jun 2011 to 20th Oct 2011. The usage of a validation data set is to prevent over-fitting by selecting the parameters $Dthr$ for DENFIS-EN or stopping training the model for ANFIS-EN at the lowest error of the validation data instead of the lowest training error. The ensemble approaches with validation were denoted as ANFIS-EN-V and DENFIS-EN-V. It is common situation that not all catchments are extensively monitored, so ensemble approaches without validation data were also studied, due to lack of data. For the ensemble approaches without validation, the data was divided

into the training data from 7th Jun 2009 to 31st Oct 2010 and the same testing data set as before, which were denoted as ANFIS-EN and DENFIS-EN. In the following sections, the four ensemble approaches of ANFIS-EN, ANFIS-EN-V, DENFIS-EN, and DENFIS-EN-V will be used to combine the ANFS and URBS models' predictions to obtain a synthesized result which will be compared to the results obtained from the component URBS and ANFIS models. In addition, SAM was used as benchmark ensemble procedure to be compared.

4.3 Results and Discussions

4.3.1 ANFIS Ensemble Model

For the ANFIS-EN-V with three membership functions, the validation RMSE increased rapidly after an initial decrease in the training phase, which indicated over-fitting. In this case, the validation method is not suitable. The reason may be that the number of parameters is too large to properly train the ANFIS-EN-V model and the ANFIS-EN-V model is unable to generalize as it is over-parameterized. For ANFIS-EN-V with two membership functions, the number of rules is reduced to two. During the training phase, the error on the validation dataset was found to continuously decrease before increasing after certain period of training. Training was stopped when the validation error reached a minimum value ($RMSE = 0.87$ m), just before it started to increase. Further training will produce lower RMSE in the training dataset but at the expense of higher model variance trying to fit more training data, leading to the limited applicability of the model in the testing phase due to over-training. Figure 4.3(a) and 4.3(b) show the results of the ANFIS-EN-V model, with two membership functions, for the training and validation data sets, respectively, at the end of the training phase. An inspection of Figure 4.3 (c) shows that ANFIS-EN-V produced RMSE of 0.94 m for the testing data but overestimated the water levels around 15th Aug. The time shift error around the period of 5th July 2011 (Figure 4.3 (c)) observed in the ANFIS model is reduced in the ensemble forecast.

(a)

(b)

(c)

Figure 4.3 ANFIS-EN-V forecasts for (a) training data (7th Jun 2009 to 31st Oct 2009), (b) validation data (7th Jun 2010 to 31st Oct 2010) (c) test data (20th Jun 2011 to 20th Oct 2011)

Results from the ANFIS-EN model without validation are presented next. The ANFIS-EN model with two membership functions was trained until the training error reached 1.0 m. ANFIS-EN models with 3 membership functions were also tested and produced similar results as ANFIS-EN with 2 membership functions and only the time series of water levels predicted by ANFIS-EN with 3 membership functions are shown in Figure 4.4. The decrease in the number of membership functions reduces both the number of the non-linear parameters in the membership functions and the number of the linear parameters in the rules. With more parameters, ANFIS-EN with three membership functions can fit the training data better, with RMSE = 0.97 m, compared with the training RMSE = 1.01 m achieved with of two membership functions. In the testing period, the two models produced similar RMSE = 0.95 m for three membership functions and RMSE = 0.97 m for two membership functions.

Figure 4.4 (b) shows that the shift in the predicted hydrograph is reduced compared with ANFIS and URBS model results.

(a)

(b)

Figure 4.4 ANFIS-EN without validation forecasts for (a) training data (7th Jun 2009 to 31st Oct 2009 and 7th Jun 2010 to 31st Oct 2010), (b) test data (20th Jun 2011 to 20th Oct 2011)

4.3.2 DENFIS Ensemble Model

The results from DENFIS-EN-V are shown in Figure4.5.

(a)

(b)

(c)

Figure 4.5 DENFIS-EN with validation forecasts for (a) training data (7th Jun 2009 to 31st Oct 2009), (b) validation data (7th Jun 2010 to 31st Oct 2010) (c) test data (20th Jun 2011 to 20th Oct 2011)

For DENFIS, the threshold distance controls the number of the clusters created and different values $Dthr$ were tested for the DENFIS-EN-V model in the training period. The number of rules group was set as three. The threshold of the distance $Dthr$ was set to 0.22 at the validation RMSE of 0.93 m before the validation error increased.

Similar to ANFIS-EN-V model, the DENFIS-EN-V produced good ensemble water level forecasts for validation and test data with RMSE of 0.93 m and 0.95 m at expense of higher training error of 1.26 m. With smaller value of $Dthr$, more clusters can be created with smaller error of the training data but at the expense of higher validation error. The DENFIS-EN-V forecasted better results for the peak around 15 Aug compared with the underestimation of the peak around 24 Sep in 2011. The time shift around 5 July was also reduced in Figure 4.5 (c) and the spikes in the URBS

forecasts were also reduced. Next, the DENFIS-EN without validation results are shown in Figure 4.6. Through trial and error, D_{thr} was set at 0.19. In the training period, the offline mode of the DENFIS-EN model adapts the model predictions to the observed data by optimizing the cluster centers and the two periods of training data are simulated well. The results from this analysis show that RMSE of DENFIS-EN = 0.86 m and the ensemble prediction curve followed the observed water levels well during the testing phase. Not only the time shift from ANFIS model around 5 July was corrected, the spikes of the URBS model around 13-15 Aug were also reduced in the DENFIS-EN forecasts as shown in Figure 4.6(b).

(a)

(b)

Figure 4.6 DENFIS-EN forecasts for (a) training data (7th Jun 2009 to 31st Oct 2009 and 7th Jun 2010 to 31st Oct 2010), (b) test data (20th Jun 2011 to 20th Oct 2011)

4.3.3 Analysis of Results

Results from the ensemble models are compared against the ANFS, URBS and SAM model results for the testing dataset in Table 4.1. It is observed that all ensemble methods (SAM, ANFIS-EN, ANFIS-EN-V and DENFIS-EN, DENFIS-EN-V) are able to reduce RMSE to values lower than the ANFS and URBS models. The DENFIS-EN gives the best performance, reducing RMSE to 0.86 m. The highest NSE was also obtained by DENFIS-EN with a value of 0.91 although the ANFIS-EN-V produced the lowest PBIAS. It is evident that the usage of a validation dataset did not produce significant improvement to the ANFIS-EN and produced even worse results in DENFIS-EN models. This may result from the limited size of the dataset which only included 294 days available for training. Partitioning of an already small dataset was not able to exploit the advantage from the validation procedure. A more extensive

dataset is required to investigate this phenomenon fully. ANFIS-EN-V and URBS model seemed to achieve better OI values nearer to 0 but the better OI values resulted from better MRD evaluation of the two models. However, both the ANFIS-EN-V and the URBS model produced spiky forecasts, which the underestimation and overestimation would offset each other. DENFIS-EN produced the lowest MARD of 3.51 among all the models.

Table 4.1 Model comparison

Criteria	ANFIS-EN	ANFIS-EN-V	DENFIS-EN	DENFIS-EN-V	SAM	ANFIS	URBS
NSE	0.88	0.89	0.91	0.88	0.89	0.84	0.87
PBIAS(%)	2	0	1	2	2	2	1
RMSE(m)	0.95	0.94	0.86	0.95	0.93	1.11	1.01
MRD	-1.59	-0.28	-1.36	-1.46	-1.71	-2.30	-1.12
MARD	4.12	4.12	3.51	3.98	3.97	4.76	4.41
OI(m ²)	-2.40	-0.39	-2.21	-2.40	-2.76	-3.91	-1.59

The predictions by the ANFIS-EN-V and DENFIS-EN are compared with SAM for the testing data in Figure 4.7. For the period around 5 July, it was discussed before that in DENFIS-EN and ANFIS-EN-V model reduced the time shift error from the ANFIS model, while SAM underestimated the water levels in Figure 4.7 for this period. On 13 Aug both the SAM and ANFIS-EN-V over-predicted the peak because of the much higher water level forecasts from URBS model while DENFIS-EN model reduced the spike from URBS as discussed before and produced better peak forecasting. For the peak on 24 Sep all the models failed to predict well and underestimated the water levels. The better ensemble model was obtained by the DENFIS-EN compared with ANFIS-EN-V, showing the advantage of using the clustering algorithm for specifying the number of the fuzzy sets and rules according the property of the input data over manually pre-defining these parameters in ANFIS-

EN-V model.

Figure 4.7 Comparison of the ANFIS-EN-V and DENFIS-EN results with the SAM results for the test data (20th Jun 2011 to 20th Oct 2011)

The effect of the ensemble models in reducing time shift errors that was observed especially for the component ANFIS and URBS models is analysed here. In this analysis, the time shift error of each model prediction was evaluated by shifting the predicted hydrograph and recalculating the NSE values after each shift. The time shift error corresponds to the number of shifts (days) which gave the highest NSE value. Figure 4.8 shows that the ANFIS model has highest NSE value with a time shift of three days, the highest among all the models considered. For the SAM model, the time shift is also three days. It can be observed from Figure 4.8 that the ANFIS-EN-V and DENFIS-EN models are able to reduce the time shift errors. In addition, DENFIS-EN model achieved the highest NSE values.

Figure 4.8 Time shift comparison of the ensemble models and the component models for the test data (20th Jun 2011 to 20th Oct 2011)

Unlike the SAM model that always allocates the same weight on the component models of the ANFIS and URBS, the proposed NFIS ensemble approach allocates the weight according the performance of the component models and the weights vary during the testing stage based on the forecasted water levels from the component models. In DENFIS model, the final output will be a linear combination of the component models ANFIS and URBS forecasts. Even though the weights for each model are not constrained within $[0,1]$, higher values of the weights indicate more trust on that component. The change of the weights of ANFIS and URBS model from DENFIS-EN for the 2011 testing stage was plotted in Figure 4.9. When at lower water levels before 20 July, URBS model was allocated much higher weight than that of the ANFIS model. The DENFIS-EN model was able to make use of the better forecasts from URBS model and corrected the underestimation of ANFIS model around 5 July. When water level increased, the URBS model produced more spiky forecasts with decreasing weight and the DENFIS-EN model reduced the oscillation of the URBS

model.

Figure 4.9 Weights Change of the Component Models from DENFIS-EN for the test data (20th Jun 2011 to 20th Oct 2011)

4.4 Conclusions

The following can be concluded from this study:

1. Together with the SAM, all ensemble model results showed improvements over the results obtained from the component models ANFIS and URBS models.
2. For the ANFIS-EN ensemble model, the over-fitting issue was addressed by considering pruning rules and validation data. By pruning the fuzzy rules, the number of parameters in the ANFIS-EN model was decreased and better results were obtained with only two membership functions and considering a validation data set. Less parameters in the ensemble model provide stronger

generalization ability with better performance in the testing phase.

3. For the DENFIS-EN ensemble model, the usage of validation produced slightly worse ensemble results than that without validation. This may come from a lack of enough data for the training stage.
4. The underestimation of the ANFS model and overestimation of the URBS model for the peak water levels are improved by the ensemble models. The time shift errors from the ANFS model are almost eliminated in the ensemble predictions and the strong oscillation present in URBS predictions is reduced. The best ensemble results were obtained by the DENFIS-EN model which reduced the RMSE from 1.11 m and 1.01 m to 0.86 m. For the DENFIS-EN ensemble model, an optimal value of the threshold distance D_{thr} was found by trial, which led to the formation of a lesser number of clusters, and stronger generalization ability.
5. The weight change based on the forecasted water levels from the component models of the NFIS ensemble approaches showed more flexible and reasonable weight allocation strategies than SAM. More research studies were possible of using real-time updating algorithm to update the structures and parameters of the ensemble model.

CHAPTER 5 ONLINE ENSEMBLE MODELING FOR REAL TIME WATER LEVEL FORECASTS FOR THE LOWER MEKONG RIVER

5.1 Introduction

The SAM is often adopted without any theoretical basis as no better alternative exists. The SAM assumes equal weighting between component models and does not differentiate between component models performance. More sophisticated ensemble techniques including the weighted averaged method, Bayesian models or neural network allow for a differentiation between the individual models. However, once trained or calibrated, the parameters of these ensemble models remain fixed and even though the weight allocation between individual models may be different, the model parameters remain unchanged during the entire testing stage. Thus, when future events with water level exceed the maximum value used in the training period occur, the ensemble model may not provide an accurate forecast since it will be used outside its range of applicability. In such a situation, the ensemble needs to be capable of adapting its model parameters in real time during the test period.

This chapter presents results obtained using a real time ensemble approach based on an online neural-fuzzy inference system which exploits the learning ability of neural network and reasoning process of fuzzy logic. The NFIS features incremental learning capabilities in order that the ensemble model can be updated in real time allowing for model parameters to be continuously adapted during the testing stage. This capability currently does not exist in ensemble models surveyed in the literature. The proposed real time ensemble approach was applied for a case study of three stations along the Lower Mekong River for 5 days ahead water level forecasts with the inputs of 5-day water level forecasts from two models, the Adaptive-Network-

based Fuzzy Inference System (ANFIS) and the URBS. A comparison of the real time forecast accuracy between these 2 component models did not reveal a clear superiority of one model over the other. The proposed real time ensemble approach was applied for a case study of three stations along the Lower Mekong River to provide 5 days ahead water level forecasts, consistent with the operational forecasts requirements for the Lower Mekong (Pagano, 2013). Specifically, the online capability of the NFIS ensemble model was tested to study the ability of the ensemble model to adapt to changed conditions during the testing stage, by subjecting the ensemble model to water levels that are higher during the testing stage compared to that used in model training and therefore validate the use of online ensemble model for real time forecasts. With real time updating using online learning therefore, the ensemble model will be adapted to larger events during the testing phase without having to retrain the ensemble model with the entire dataset. The objectives of this study were: 1.) Analyze the applicability and limitations of the NFIS as an ensemble approach, 2.) Evaluate the effectiveness of online learning in ensemble model adaptation in a real-time application , and 3). Validate the usefulness of the online NFIS in providing real time water level forecasts for the Lower Mekong.

5.2 Methodology

5.2.1 Neural-fuzzy Model

Kasabov and Song (2002) developed the dynamic evolving neural-fuzzy inference system or DENFIS model for time series prediction, which was adopted as the ensemble model in this paper. After the training stage, the NFIS parameters including the cluster centers, radius, membership functions and the consequent parameters in the fuzzy rules will be determined. The first ensemble approach applied in this study is the offline mode of DENFIS, denoted as EN-OFF. In the offline mode, the ensemble model is first trained using the ECMc local learning algorithm to optimize the input data clusters. Once training is completed, the model parameters remain fixed

during the testing stage.

5.2.2 Real Time Updating Approach

To overcome the deficiency of the ensemble model may fail during the test stage if much higher water levels occur that have never been learnt during the training stage, a real time updating approach is proposed. The incremental algorithm employed in the ECM allows for the online implementation of the NFIS model, where model parameters are updated in real time. In a practical application, the offline model is first applied to train the ensemble model, and populate the cluster space and determine the NFIS parameters. During the testing phase, the ensemble model is switched to online mode which enables the ensemble model to be continuously adapted by modifying the clusters and changing NFIS parameters, each time an observed water level becomes available. The structure of the ensemble model with real time updating using online learning or EN-RTON1 is shown in Figure 5.1.

Figure 5.1 Structure of the ensemble model with real time updating using online learning (EN-RTON1)

In Figure 5.1, the training part of EN-RTON1 uses the offline mode to create and optimize the clusters and rules of the DENFIS model and initialize the model parameters with the training data set. With the availability of each data point during the testing phase, the online mode of DENFIS is used to update the clusters and rules and subsequently produce the ensemble forecasts.

In addition to EN-RTON1, another real time updating ensemble approach which incorporates 5 sub-models was also proposed in this study, which is defined as EN-RTON2. The structure of EN-RTON2 is shown in Figure 5.2. For the EN-RTON2, the initialization is the same as that in EN-RTON1, which uses the same clusters and fuzzy rules trained from the offline mode of DENFIS. Then these clusters and rules will be used to produce the ensemble results for the first 5 days of the test data. When the measured water levels for each day of the first 5 days are available, the ensemble model will be updated to produce the 5 day forecasts independently. The idea behind EN-RTON2 is that the ensemble results of Day 6 is based on the data of Day 1 only and the updating should be only carried out with 5 days interval in the testing data and as a result, five sub-models in total will be created. The updating procedure for each of the sub-models in EN-RTON2 is the same as that of EN-RTON1. The only difference between these two approaches is that in EN-RTON1 there is only ensemble model to be updated each time the measured data is available, while in EN-RTON2 the five sub-models are updated at 5 days' interval. Thus the EN-RTON2 updates the clusters and rules of each sub-models only based on the feedback from the results on the days when the ensemble forecasts are made from those clusters and rules. With more reasonable updating process, each updating in EN-RTON2 is not depending on previous days. A third real time model which uses the offline mode of DENFIS to retrain the whole data set each time when measured water levels are available was used as a comparison against the online updating approach. This is indicated as EN-RTOFF.

Figure 5.2 The structure of the ensemble model with real time updating using online learning and sub-models (EN-RTON2).

5.2.3 Study Site

The 5 days forecast results from two component models, the URBS model (Carroll, 2007) and the Adaptive-Neural-based Fuzzy Inference System or ANFIS model (Jang, 1993), have been used to predict the 5-day ahead water levels for the three stations (Thakhek, Pakse and Kratie) during the wet seasons from 2009 to 2011. Data from 7 Jun 2009 to 31 Oct 2010 were used for training the ensemble models (EN-OFF, EN-RTON1, EN-RTON2 and EN-RTOFF) and the data from 20 Jun 2011 to 20 Oct 2011 were used as test data. During training, the measured water levels were used to optimize the model parameters in the ensemble model. An overall assessment of the results obtained from these two component models is that the predictions of the URBS model are spiky while the ANFIS model tended to under-estimate the peak. This study explores the ensemble approach based on DENFIS model to optimize the ANFIS and URBS model forecasts for the Lower Mekong River. SAM was adopted

as the benchmark model to be compared with the proposed ensemble models in this chapter.

5.3 Results and Discussions

5.3.1 Offline Ensemble Model (EN-OFF)

A comparison between results obtained from EN-OFF with ANFIS, URBS and SAM is shown in Figure 5.3. The results from URBS model are spiky and produces large spikes around the peak values such as around 5 Aug and 25 Sep 2011 for Thakhek, 25 Sep 2011 for Pakse and 10 Aug 2011 for Kratie. The ANFIS model underestimates the peak around 10 Aug 2011 for Thakhek. In general, it can also be said that ANFIS and URBS model results are affected by time shift problems while the SAM results generally does not result in a great improvement as it is an averaging these two results.

(b)

(c)

Figure 5.3 EN-OFF Results for (a) Thakhek, (b) Pakse and (c) Kratie for test data from 20 Jun 2011 to 20 Oct 2011

After the training phase, five clusters were created each for Thakhek and Pakse and three clusters were created for Kratie. An inspection of Figure 5.3 shows that the EN-OFF model does not seem to produce improved results for Thakhek and Pakse, especially for the peaks as shown in Figure 5.3(a) and Figure 5.3(b). For Kratie, the EN-OFF model performs better than the SAM especially around 5 July 2011 and 13 Aug 2011 where the overestimation of the peaks by URBS model is reduced. The poor performance by the EN-OFF model for the Thakhek and Pakse test data is due to the difference between the training data set and the testing data set. For Thakhek and Pakse, the maximum values of the water levels forecasted by ANFIS and URBS, and the measured water levels for the training period are smaller than those for the test period. This means that in real time forecasting, the EN-OFF model rules created based on historical data may not be applicable during the testing phase. To overcome this problem, real time updating was used. It is hypothesized that including the latest measured water levels continuously allows the ensemble model to adapt, enabling the online model to gradually adjust the clustering results and fuzzy rules.

5.3.2 Ensemble Model with Real Time updating using Online Learning (EN-RTON1)

The results of the real time updating ensemble model EN-RTON1 and EN-RTOFF are plotted in Fig 5.4 for the three stations and compared with the SAM and EN-OFF results.

(a)

(b)

Figure 5.4 EN-RTON1 and EN-RTOFF compared with EN-OFF and SAM Results for (a) Thakhek, (b) Pakse and (c) Kratie for test data from 20 Jun 2011 to 20 Oct 2011

The EN-RTON1 model improved the peak forecasting from 7 Aug 2011 to 15 Aug 2011 for Thakhek where EN-RTOFF and SAM under predicted the peak. For Pakse, the EN-RTON1 corrected the under estimation of the water levels from 19 Aug to 21 Aug. But significant time shift was found of the EN-RTON1 model for the peak around 11 Aug. The same time shift problem was also found for the peak on 24 Sep of Kratie.

Five clusters were created during the training of EN-RTON1 model for Thakhek. During the testing phase, two more clusters were created with the online updating procedure. For Pakse, in the testing phase the online updating procedure created one more cluster. For Kratie, even though the number of the clusters did not change, slight changes of the centers and radius were made with the cluster updating portion of the

online updating procedure. The predicted water levels and the difference between the ensemble forecasts and measured water levels for Thakhek and Pakse in the testing phase are plotted in Figure 5.5 to show the effect of the created clusters in test stage. The results for Kratie are not shown since clusters were not created during the testing phase.

Figure 5.5 Error Analysis on creating new clusters of EN-RTON1 ensemble results for (a) Thakhek and (b) Pakse for test data from 20 Jun 2011 to 20 Oct 2011

In Figure 5.5, the arrows indicate the date when new clusters were created for the EN-RTON1 model in the testing phase. In Figure 5.5(a), up to 4 Jul 2011, EN-RTON1 results are similar to EN-OFF results because no new clusters were created from EN-RTON1 when water levels were still within the range of the training data. But the EN-RTON1 cannot respond fast enough to improve the forecasts for the peak around 4 July. A new cluster on 6 Aug 2011, the EN-RTON1 model started to deviate from the EN-OFF model results and move closer to measured water levels, while the EN-OFF and SAM results showed an under-estimation of about 2 meters. For Pakse in Figure 5.5(b), after creating a new cluster on 6 Aug 2011 as a result of high water levels, the EN-RTON1 model is unable to respond fast enough and could not correctly predict the peak around 15 Aug 2011. However, from 18 Aug 2011 to 22 Aug 2011, the EN-RTON1 model corrected the EN-OFF forecasts from an under-estimation of about 2 meters to less than 0.5m over-estimation. This suggests that the model requires a finite time for spin-up before the changes to model parameters can take effect. From 28 Aug to 5 Sep, the EN-RTON1 model over-predicted the falling limb of the hydrograph.

Figure 5.6 indicates how EN-RTON1 calculates the final weights considering the firing strength of the fuzzy rules. From 21 Aug 2011 to 26 Aug 2011 in Figure 5.6, the URBS model showed very spiky forecasts and the weight of URBS model decreased close to zero, which imposed strong constraint on the URBS component in the EN-RTON1 ensemble model. This weight punishment of URBS model can also be seen from 19 Sep 2011 to 27 Sep 2011 when the ANFIS model was allocated higher weights.

Figure 5.6 Change in weights of EN-RTON1 for test data from 20 Jun 2011 to 20 Oct 2011, Pakse.

5.3.3 Ensemble Model with Real Time updating using Online Learning and Sub-Models (EN-RTON2)

Results obtained from the EN-RTON2 model are shown in Figure 5.7. For Thakhek, the peak from 7 Aug to 15 Aug is forecasted well by EN-RTON2 ensemble model and the peak estimation on 21 Sep is improved compared with the EN-RTON1 model. For Pakse, the time shift problem observed in EN-RTON1 is reduced in EN-RTON2 and the peak estimation around 11 Aug is improved, although. Even though the falling limb is over-estimated, the predictions of the rising limb from 8 Aug to 11 Aug is improved by correcting the underestimation of around 2 meters. For Kratie station, the peak around 24 Sep is now well predicted compared with EN-RTON1, EN-OFF and SAM results and the time shift found in EN-RTON1 appears to be improved. Hence, by using five sub-models, the EN-RTON2 ensemble model is able to improve on many of the deficiencies observed in earlier.

(a)

(b)

Figure 5.7 EN-RTON2 results compared with EN-OFF, EN-RTON1 and SAM model results for (a) Thakhek, (b) Pakse and (c) Kratie for test data from 20 Jun 2011 to 20 Oct 2011

Since the five sub-models in EN-RTON2 are updated independently, the clusters created during the testing phase may show distinct results. For Thakhek,, only sub-model 5 created two clusters during the testing phase. The performance of the EN-OFF, SAM, EN-RTOFF, EN-RTON1 and EN-RTON2 used in this paper and the forecasts from ANFIS and URBS model are compared in Table 5.1.

Table 5.1 Model performance evaluation for (a) Thakhek (b) Pakse and (c) Kratie of test data from 20 Jun 2011 to 20 Oct 2011

(a)

Thakhek	EN-OFF	EN- RTOFF	EN- RTON1	EN- RTON2	SAM	ANFIS	URBS
NSE	0.62	0.68	0.68	0.75	0.69	0.73	0.47
PBIAS(%)	8	5	1	1	4	5	4
RMSE(m)	1.15	1.07	1.07	0.94	1.04	0.98	1.36
PEP(%)	-6	1	4	1	7	-2	18
PE(%)	-11	-4	-7	-2	-7	-5	-8

(b)

Pakse	EN-OFF	EN- RTOFF	EN- RTON1	EN- RTON2	SAM	ANFIS	URBS
NSE	0.81	0.84	0.80	0.85	0.84	0.75	0.80
PBIAS(%)	5	3	0	-1	4	7	1
RMSE(m)	1.01	0.91	1.02	0.88	0.93	1.15	1.04
PEP(%)	0.00	-3	3	8	-2	-9	11
PE(%)	-26	-25	-22	-12	-23	-22	-24

(c)

Kratie	EN-OFF	EN- RTOFF	EN- RTON1	EN- RTON2	SAM	ANFIS	URBS
NSE	0.91	0.89	0.88	0.89	0.89	0.84	0.87
PBIAS(%)	1	1	1	1	2	2	1
RMSE(m)	0.86	0.91	0.96	0.93	0.93	1.11	1.01
PEP(%)	-3	1	-1	1	-1	-3	2
PE(%)	-7	-4	-8	-1	-7	-7	-7

From Table 5.1(a), EN-RTON2 model produced the lowest RMSE 0.94m and highest NSE 0.75. Both the EN-RTON1 and EN-RTON2 produced the lowest PBIAS with 1% underestimation compared with the 4%-8% underestimation of all the other models. When the peak occurs, the EN-RTON2 produced the lowest PE with 2% underestimation while SAM and EN-OFF model underestimated by 7% and 11%. For Pakse station, EN-RTON2 produced the highest NSE of 0.85 and the lowest RMSE of 0.88m. The EN-OFF model seemed to produce very good peak estimation with PEP of zero, but the measured peak occurred on 11 Aug while the EN-OFF produced maximum forecast on 25 Sep. The EN-RTON2 produced 12% underestimation while all the other models underestimated the peak from 22% to 26%. For Kratie station, all the models showed similar evaluation results for overall performance but only EN-RTON2 produced the lowest PE with 1% underestimation.

5.4 Conclusions

In this study an ensemble model based on neural-fuzzy inference system and three real time updating approaches were used to combine the water level forecasts of ANFIS and URBS model for Lower Mekong. A simple averaged model was used as benchmark.

1. The ensemble model which utilized the offline mode of DENFIS model (EN-OFF) can produce improved ensemble results when the test data is within the range of the training data. When higher water levels that have never been trained in the training period, the EN-OFF model failed to give good peak predictions.

2. The EN-RTOFF model which retrained the whole data set every time when the latest measured water levels were available cannot produce significant improvement to the EN-OFF model. The offline learning which optimized the all the clusters without considering the time order of the input data led to the small improvements when the data of the measured water levels used for updating were much less than

the training data set.

3. The real time updated ensemble approach used the clustering results and fuzzy rules trained with the offline mode of DENFIS (EN-OFF) as the basic knowledge. In the testing phase, the online mode of DENFIS was switched on to update existing cluster, create new clusters and update fuzzy rules when the latest measured water levels were available. By exploiting the latest accessible information, the real time updated ensemble model continued adapting to the new situations in the testing phase without retraining on the whole data set. Two real time updating approaches with online learning in the testing phase were proposed with different updating interval. EN-RTON1 updated the ensemble model results continuously with accessible daily measured water levels while EN-RTON2 updated the ensemble results by 5-day interval with five sub-models.

4. Statistical analysis of the models for all the three stations indicated the superiority of the EN-RTON2 model over EN-RTOFF, EN-RTON1 models, SAM and the EN-OFF model. Not only the spikes in the URBS model were eliminated, but also the time shift problems in the ANFIS model results were decreased.

CHAPTER 6 ENSEMBLE WATER LEVEL FORECASTING FOR LANYANG CREEK, TAIWAN

6.1 Introduction

The water level at Lanyang Bridge in Yilan County, Taiwan is currently forecasted using the WRF/WASH123D operational system (Hsiao *et al.*, 2013; Shih *et al.*, 2014). The watershed model takes as input 15 precipitation forecasts resulting in 15 forecasts of water levels in this study. However, this is impractical for engineers as there is a wide variation in the 15 forecasts, and engineers do not have sufficient basis to select the best result for implementation. Currently, the simple average method result is adopted, however this is done without justification as there is no better method. In order to overcome this problem, this study applied the NFIS ensemble modeling approach, where the component models to the ensemble model are the water level forecasts provided by WASH123D for each precipitation forecast. The results obtained from the NFIS ensemble model were compared with the SAM, which is currently implemented to provide operational forecasts (Hsiao *et al.*, 2013). Although the forecast of water level on Lanyang Bridge was adopted in this study, it is expected that the findings are applicable for similar applications elsewhere.

Few studies in ensemble modeling applications have dealt with the topic of component model selection and this analysis will allow for the identification of possible bias in component model performance at different forecast horizons or ranges of water levels and therefore arrive at a truncated input component model space for the ensemble model. The analysis includes investigations into the feasibility to train separate ensemble models for short- and long-term forecasts based on whether the forecasting horizon is beyond 24 hours. In addition, a preprocessing of the input component models will be conducted in order to remove the negative effects from those component models with very poor forecasting performance.

6.2 Methodology

The offline mode of Dynamic Evolving Neural-Fuzzy Inference System or DENFIS (Kasabov and Song, 2002) which optimized the cluster centers was used as the ensemble approach. The detailed description of DENFIS can be referred to Chapter 3.4. Fifteen water level forecasts of the component models for Lanyang Creek Basin in Taiwan derived from the 15 different perturbed initial conditions in the rainfall forecasts were used. The data for the period from 11th May 2012 to 3rd September 2013 was used for training the model with the data for the period from 21st September 2013 to 24th September 2013 and the data from 22nd July 2014 to 26th July 2014 was used as the test data. In order to avoid over-fitting, the first $\frac{2}{3}$ of the training dataset was used to determine the threshold of the distance D_{thr} and the second $\frac{1}{3}$ used for validation. The benchmark model is SAM which is currently adopted as the ensemble model in Taiwan.

6.3 Evaluation of Input Component models

Although it is possible to run the ensemble model using the complete set of fifteen WASH123D outputs as component models of the ensemble, it has been shown that errors associated with stream flow forecasts for different lead times can vary with different methods to generate the rainfall forecasts (Bennett *et al.*, 2014). Thus, before constructing the ensemble model, an initial study was conducted to investigate the performance characteristics of the fifteen component models as a function of the forecast horizon (short-term or < 24 hrs forecasts and long-term or > 24 hrs forecasts) and water level regime (L: below 3.5 m; M: between 3.5 m and 5.8m; H: above 5.8 m).

6.3.1 Short- and Long-term Forecasts

The six error statistics (NSE, PBIAS, RMSE, PEP, PE and PT) were calculated for the short-term and long-term operational runs in the training, validation and test data

respectively. A “perfect” model is added into the criteria matrix with all the six error statistics set to their ideal values assuming perfect model performance, ie $NSE = 1$, $PBIAS = 0$, $RMSE = 0$, $PEP = 0$, $PE = 0$ and $PT = 0$. To reduce redundancy, principal component analysis or PCA (Jolliffe, 2005) was used to remove the most correlation components between the different dimensions. This analysis revealed that 75% retention of variance could be achieved when the dimension was reduced to 3. The distance in the PCA between the dimension-reduced criteria vectors of the fifteen component models and the vector of the “perfect” model was calculated and the results are shown in Table 6.1 for the fifteen component models corresponding to the different WASH123D forecasts where a smaller distance indicates greater similarity with the “perfect” model. Further classification of the results obtained in Table 6.1 are shown in Table 6.2.

Table 6.1 The distance of the fifteen component models to the “Perfect Model” in PCA reduced space.

Component models	Training		Validation		Test	
	Short	Long	Short	Long	Short	Long
M1	4.04	6.15	4.17	6.48	4.49	3.51
M2	4.32	6.83	4.23	7.11	4.41	4.01
M3	4.23	6.49	4.04	6.84	10.27	3.92
M4	4.52	5.57	3.93	5.95	2.85	2.84
M5	4.57	6.49	4.25	6.51	3.15	3.07
M6	4.81	6.11	4.21	6.18	5.24	3.03
M7	4.26	6.19	3.95	6.01	4.91	2.30
M8	5.00	6.83	4.63	5.86	5.45	2.67
M9	3.88	5.32	3.71	5.42	8.29	2.81
M10	3.97	6.06	3.93	5.83	3.17	2.13
M11	4.81	6.61	4.68	5.95	5.05	2.52
M12	4.69	6.64	4.54	6.21	3.14	3.33
M13	5.54	6.43	4.81	5.98	4.29	3.26
M14	4.83	6.37	4.24	6.90	4.66	2.87
M15	5.28	6.26	4.17	6.74	4.35	2.68

Table 6.2 Component models classification

Component models	Training	Validation	Test
M1	SHORT	SHORT	LONG
M2	SHORT	SHORT	LONG
M3	SHORT	SHORT	LONG
M4	SHORT	SHORT	LONG
M5	SHORT	SHORT	LONG
M6	SHORT	SHORT	LONG
M7	SHORT	SHORT	LONG
M8	SHORT	SHORT	LONG
M9	SHORT	SHORT	LONG
M10	SHORT	SHORT	LONG
M11	SHORT	SHORT	LONG
M12	SHORT	SHORT	SHORT
M13	SHORT	SHORT	LONG
M14	SHORT	SHORT	LONG
M15	SHORT	SHORT	LONG

In Table 6.2, “SHORT” refers to the case where the distance for the short-term forecast is smaller than that of the long-term forecast, which means that the component model is better at short-term forecasting. Conversely, “LONG” indicates that the component model is better suited for long-term forecasting. This result indicates that with the exception of Component model 12, the performance of the component models are not consistent; all the other component models showed better long-term forecast performance for the test data, but performed better at short-term forecasts for the training and validation dataset. Next, the performance of the fifteen component models was ranked based on the distance calculated in Table 6.1, and the results are shown in Table 6.3. For the long-term forecasting, the best three

component models in the training were Component model 9, Component model 4 and Component model 10. In the test data, Component model 10 performed well for the long-term forecasts, however Component model 9 and Component model 4 are ranked in the middle. However, for the short-term forecasting, the best component model, Component model 9, showed the second worst performance in the short-term forecasting in the test data.

The results in the above-mentioned analysis showed that the component models are not able to consistently provide either short-term or long-term forecasts well. This leads to the conclusion that, in this case, it is infeasible to train separate ensemble models for short- and long-term forecasts as has been proposed by Han *et al.* (2007) and Lin *et al.* (2013). In order to further identify the characteristics of each component model, the component model performance was evaluated at different water levels.

Table 6.3 Ranking of the fifteen component models for the training, validation and test phase.

Rank	Training		Validation		Test	
	Short	Long	Short	Long	Short	Long
1	M9	M9	M9	M9	M4	M10
2	M10	M4	M4	M10	M12	M7
3	M1	M10	M10	M8	M5	M11
4	M3	M6	M7	M11	M10	M8
5	M7	M1	M3	M4	M13	M15
6	M2	M7	M1	M13	M15	M9
7	M4	M15	M15	M7	M2	M4
8	M5	M14	M6	M6	M1	M14
9	M12	M13	M2	M12	M14	M6
10	M11	M5	M14	M1	M7	M5
11	M6	M3	M5	M5	M11	M13
12	M14	M11	M12	M15	M6	M12
13	M8	M12	M8	M3	M8	M1
14	M15	M8	M11	M14	M9	M3
15	M13	M2	M13	M2	M3	M2

6.3.2 Forecast Results at Different Water Level Regimes

After calculating the RMSE of all the 15 component models at different water level regime (L: below 3.5 m; M: between 3.5 m and 5.8m; H: above 5.8 m).the results indicate that there are no clear patterns in component model performance at the different water level regimes (see RMSE results in Figure 6.1). After ranking the fifteen component models, three strategies were used for component model selection:

1L1M1H - the best performing component models at each range of water level was chosen; 2L2M2H - the best two component models for each range of water level was chosen; 1L2M3H - the best component model for low, best 2 component models for medium and best 3 component models for the high water level range were chosen. The selected component models are listed in Table 6.4.

Figure 6.1 Evaluation of component model forecast performance at different ranges of water levels for the training dataset

Table 6.4 Selected component models for ensemble model

Input Selections	Selected Component Models				
1L1M1H	1	13			
2L2M2H	1	7	13	14	
1L2M3H	1	7	10	13	14

The offline mode of DENFIS was adopted as the ensemble model and the number of fuzzy rules group was set as three. Different values of the threshold distance $Dthr$ were investigated by inspecting the RMSE in the validation data when it is at a minimum. After trial and error, the value of $Dthr$ which resulted in the lowest of all

the observed minimum validation errors was adopted. These values are listed in Table 6.5.

Table 6.5 Training and validation RMSE (m) of the ensemble offline model with different input selections

Input Selection	D_{thr}	Training	Validation
15 Component models	0.13	0.39	0.42
1L1M1H	0.09	0.52	0.53
2L2M2H	0.09	0.48	0.50
1L2M3H	0.11	0.46	0.48

The error statistics associated with each of the selected component models for the testing dataset are compared with the case where all 15 component models were used as inputs to the ensemble in Table 6.6 and Table 6.7. Results of the SAM are included as benchmark.

Table 6.6 Evaluation of 2013 results for different inputs selected of the ensemble offline model

Criteria	15 Component models	1L1M1H	2L2M2H	1L2M3H	SAM
Mean (m)	3.46	3.46	3.47	3.49	4.02
STD (m)	0.35	0.38	0.39	0.39	0.25
E	0.40	0.61	0.59	0.56	-0.08
PBIAS (%)	5	5	5	5	-10
RMSE (m)	0.33	0.27	0.27	0.28	0.44
PEP (%)	-2	-1	1	3	0
PE (%)	-7	-1	1	3	-1
PT (h)	-2	1	0	0	1

The mean and standard deviation of the measured water levels for the event in 2013 are 3.66 m and 0.42 m. The results of SAM model shows a higher mean value with a smaller standard deviation, which means that the hydrograph will be flat and overestimate the measured water levels. All the input selections of the offline model produced higher NSE values while SAM produced negative values. For PBIAS, the SAM model over-predicted the measured water levels by 10% and all the other ensemble model under-predicted by 5%. The improvement of the ensemble models for the overall evaluation can also be seen from the RMSE values which reduced from the SAM model from 0.44 m to around 0.3 m. For the peak evaluation of the percent error in peak (PEP), the percentage error at peak flow (PE) and peak time difference (PT), SAM model produced good results.

Table 6.7 Evaluation of 2014 results for different inputs selected of the ensemble offline model

Criteria	15 Component models	1L1M1H	2L2M2H	1L2M3H	SAM
Mean (m)	3.20	3.15	3.17	3.20	3.44
STD (m)	0.78	0.73	0.73	0.71	0.84
NSE	0.16	0.15	0.21	0.36	0.38
PBIAS (%)	8	9	9	8	1
RMSE (m)	0.49	0.49	0.48	0.43	0.42
PEP (%)	11	13	6	2	11
PE (%)	-3	-10	-10	-10	-3
PT (h)	-3	-3	-3	-3	-4

For the event in 2014, the mean and standard deviation of the measure water levels are 3.48 m and 0.54 m. From Table 6.7, the SAM results produced a much higher STD value which indicated the hydrograph of SAM model will show a high peak.

Only the offline ensemble model with 1L2M3H input selection and SAM model produced good results for the Nash efficiency. For the peak estimation, all the models over-estimated the maximum water levels where only the offline model with 1L2M3H inputs reduced the over estimation. The PT results showed a time shift issue which the ensemble offline models reduced the early warning problem by 1 hour. From the results of two events in 2013 and 2014 as shown in Table 6.6 and Table 6.7, the best ensemble offline model is the 1L2M3H input selection.

6.4 Results of Ensemble Forecasts

The ensemble offline model with input 1L2M3H produced the highest NSE and the lowest RMSE and PEP among all the input selections for the ensemble offline models and the results obtained from the best ensemble offline model, 1L2M3H were plotted in Figure 6.2 against the benchmark SAM model.

(a)

(b)

Figure 6.2 Comparison of the forecasts from the WASH123D and ensemble models: (a) 2013 test event, (b) 2014 test event.

The results of the ensemble model with real time updating using online learning approach EN-RTON1 proposed for Lower Mekong data in Chapter 5 which used the latest measured water levels to update the ensemble model are also tested for Taiwan catchment and shown in Figure 6.2. The 5th and 95th percentiles of the results from the 15 component models were shown by the grey shaded area in the figure. The large spread of the component model forecasts showed a strong divergence in the water level forecasts with different rainfall inputs, which makes it challenging for policy makers to select a correct component model. Most component models over-predicted the peak water levels and produced much higher forecasts of the falling limb of the hydrograph especially for the 2013 test event in Figure 6.2(a). In Figure 6.2 (a), the SAM results appear to match the peak very well, in spite of the wide variation in the WASH123D predictions around the peak. This is somewhat fortuitous. It is also

observed that the low water levels are not well predicted and is attributed to the inadequate treatment of infiltration in the WASH123D model. The poor performance of SAM for the low water levels showed the limitation of the weighted average ensemble method that poor performance cannot be avoided if most of the component models over-predicted or under-predicted in the same direction simultaneously. The offline model with 1L2M3H input selection showed a better shape of the hydrograph even though at the peak the ensemble results were slightly worse than the SAM model. The better peak results of the SAM model does not indicate the SAM approach is good for peak ensemble estimation and in Figure 6.2 (b) the SAM model produced much worse peak results for the event in 2014. The offline 1L2M3H ensemble model reduced the peak over-estimation and shifted the peak forward by 1 hour. For both the events in 2013 and 2014, the real time updated ensemble model did not show significant improvement over the offline 1L2M3h ensemble model. This is because in the training data set, the highest event is around 8 m while for the highest water level of the two events in the test data set is only 5.3 m. The real time updated approach will not make a big difference if no new information or higher events that have never been trained before occur in the testing phase.

6.5 Conclusions

In this chapter, the offline mode of DENFIS was applied to the water level forecasts for Lanyang Creek Basin. The inputs to the ensemble model are from a hydrological model WASH123D with different rainfall inputs. The rainfall forecasting with different perturbed initial conditions form the difference among the water level forecasts. Few studies in ensemble modeling applications have dealt with the topic of component model selection and this chapter considers the identification of possible bias in component model performance at different forecast horizons or ranges of water levels and therefore arrive at a truncated input component model space for the ensemble model. The analysis includes investigations into the feasibility to train separate ensemble models for short- and long-term forecasts based on whether the

forecasting horizon is beyond 24 hours and training the ensemble model with different input selections of the component models based on the performance at different water levels. Before importing the component model forecasts to the ensemble model, all the component models were analyzed and pre-processed:

1. The 15 component models were not able to consistently provide either short-term or long-term forecasts well. In this case, it is infeasible to train separate ensemble models for short- and long-term forecasts.
2. Different input selections based on the performance of the component models at different water levels were tried for the ensemble model and the input selection 1L2M3H which addressed more on the higher water levels produced the best ensemble results compared to other input selections.
3. The ensemble offline model with 1L2M3H results produced better results compared the benchmark model SAM. For the event in 2013, the ensemble model reduced the over-estimation of the SAM model and produced much better overall forecasting. For the event in 2014, the time shift and over-estimation of the SAM model for the peak were reduced in the ensemble model results.

CHAPTER 7 ENSEMBLE APPROACH USING MODIFIED OFFLINE MODELS FOR WATER LEVEL FORECASTING IN LANYANG CREEK, TAIWAN

7.1 Introduction

Even though positive results from the reviewed studies in Chapter 2 clearly point to the potential for the ensemble modeling approach, the approaches applied in the ensemble procedure have been limited and more sophisticated ensemble methodologies capable of evaluating the performance of the component models at each time step are lacking. The objective of this study was to adopt a methodology based on the neural fuzzy approach which benefits from the reasoning ability of the fuzzy inference system and the learning ability of neural networks and to try to interpret the ensemble process with the proposed modifications. In this chapter two modified offline model of DENFIS were proposed for the ensemble approach to interpret the combination process. The input data are the fifteen estimates of the predicted water levels at Lanyang Bridge, Yilan County, Taiwan from Chapter 6. The first ensemble approach is a modified offline ensemble model which imposed linear constraints and removed the requirement of the constant term in fuzzy rules. The other ensemble model will consider the effects of the slopes of the hydrographs in the ensemble process. Then the proposed modified ensemble models will be compared with the benchmark SAM. The visualization of the fuzzy rules and the process which shows how the weights change with the input data will be shown in the end.

7.2 Methodology

7.2.1 Modified DENFIS with Linear Constraints

According to (Kasabov and Song, 2002), a first-order Takagi-Sugeno fuzzy inference system is used so the fuzzy rules are linear functions of the input dimensions. The

parameters in a fuzzy rule include the weights of the input dimensions and a constant term (Eq. 7.1); however, the weights are not constrained between [0, 1]. This is not ideal in an ensemble model application since it is desirable to study the relative importance of various inputs through the weight allocation:

$$y = w_0 + w_1X_1 + w_2X_2 + \cdots + w_qX_q \quad (7.1)$$

where y is the output of a fuzzy rule, w_i is the weight for each input dimension $X_i(i=1,2,\dots,q)$ and w_0 is the constant term.

In the online model, the recursive least-square (RLS) algorithm is used to update the ensemble model parameters. Zhu and Li (2007) proposed the solution to the linear constraints for the RLS algorithm. For the linear-equality constraint, the initial values of RLS can be changed to satisfy the constraint. But for the linear-inequality constraint, at every step too many parameters need to be calculated for the online mode of DENFIS. Therefore, for this study weight constraints were only applied in the offline mode:

$$\sum_{i=1}^q w_i = 1 \quad (7.2)$$

where $w_i > 0$ for $i = 1, 2, \dots, q$ is the weight for each dimension of the input vectors in the fuzzy rules, q is the dimension of the input vectors. In addition to the weights constraints, the constant term in the fuzzy rules is also removed. Thus the output y of each fuzzy rule will be a weighted average of the WASH123D water level forecasts. The weights in the final normalized output Y will also satisfy the linear constraints after multiplication by the normalized firing strength for each fuzzy rule (Eq. 7.3).

$$Y = \frac{\sum_{i=1}^n \omega_i f_i(X_1, X_2, \dots, X_q)}{\sum_{i=1}^n \omega_i} \quad (7.3)$$

where $\omega_i = \prod_{j=1}^q \mu_{R_{ij}}(X_j)$; $i = 1, 2, \dots, n$, $j = 1, 2, \dots, q$. $\mu_{R_{ij}}(X_j)$ is the membership for the input $\mathbf{X}$ in the dimension of j of the fuzzy rule i with n fuzzy rules. The results of the model in chapter 7.2.1 will be denoted as Modified Offline.

7.2.2 Modified DENFIS with Linear Constraints and Slopes

Since the rising and falling limbs in a hydrograph are governed by different hydrological processes, hydrograph slope was taken into consideration. The slopes, S , of each component model forecasts at different time steps were calculated by (Eq. 7.4-7.6):

$$S_{i,t} = (X_{i,t+\Delta} - X_{i,t-\Delta}) / (2 * \Delta) \quad (7.4)$$

$$S_{i,1} = (X_{i,1+\Delta} - X_{i,1}) / \Delta \quad (7.5)$$

$$S_{i,T} = (X_{i,T} - X_{i,T-\Delta}) / \Delta \quad (7.6)$$

where $i = 1, 2, \dots, q$, $t = 2, 3, \dots, T-1$, Δ is the time interval. Lastly, the component model forecasts and slopes were treated as individual input dimension to the ensemble model for clustering. The structure of the modified offline model of DENFIS with linear constraints and slopes was shown in Figure 7.1.

Figure 7.1 The Structure of the modified offline mode of DENFIS with linear constraints and slopes.

In Figure 7.1, the input matrix to the modified offline model considering slopes consists of the water level forecasts of N component models and the N slopes of the hydrograph, which double the dimension of the input space for clustering. For the offline mode of DENIFS, ECMc (Section 3.4) is used which optimized the resulted clusters from ECM. Thus during the training period, several clusters considering the water levels and the slopes will be formed to characterize the input space. After the clustering process, the fuzzy rules will be created and optimized for each cluster. Only the N water level forecasts are considered for constructing the fuzzy rules and in creating fuzzy rules the weights are constrained and constant term is removed as described in Chapter 7.2.1. Thus the final water level output can be interpreted as the

combination based on the importance of each component model forecasts. The dimension of the weight vector of the fuzzy rules is reduced to N and the least square with linear constraints is used to optimize the parameters. After those changes, the model will select those rules which consider the effects of the water levels and slopes and then combine the component model forecasts with the learnt strategies of weight allocation. The results of the model in Chapter 7.2.2 will be denoted as Modified Offline with Slope. The following sections will show the results of the modified offline model, which will be followed by a part about interpreting the modified model. Then the results of the modified offline with slope will be shown and the visualization of the weights allocation comes after.

7.2.3 Data and Study Area

The ensemble approaches with modified offline model were used to forecast water levels for Lanyang Bridge, Yilan County, Taiwan (Chapter 3.5.2). The water level at Lanyang Bridge is currently forecasted using the WRF/WASH123D operational run system (Hsiao *et al.*, 2013; Shih *et al.*, 2014). The watershed model takes as input fifteen precipitation forecasts resulting in fifteen forecasts of water levels in this study. The results obtained from the neural fuzzy ensemble model were compared with the simple average method, which is currently implemented to provide preliminary operational forecasts (Hsiao *et al.*, 2013). The data for the period from 11th May 2012 to 3rd September 2013 was used for training the model with the data for the period from 21st September 2013 to 24th September 2013 and the data from 22nd July 2014 to 26th July 2014 was used as the test data. In order to avoid over-fitting, the first $\frac{2}{3}$ of the training dataset was used to determine the model parameters and the second $\frac{1}{3}$ used for validation.

7.3 Results and Analysis

7.3.1 Results of the Modified Offline Model

The same data pre-processing as elaborated in Chapter 6 was carried out and the models were trained with different input selections which are the 15 component models, 1L1M1H, 2L2M2H and 1L2M3H. The benchmark model is the SAM. The training results of different input selections are shown in Table 7.1.

Table 7.1 Training and validation RMSE (m) of the modified offline model with different input selections

Input Selection	Dthr	Training	Validation
15 Component models	0.11	0.38	0.40
1L1M1H	0.12	0.54	0.56
2L2M2H	0.14	0.53	0.54
1L2M3H	0.18	0.50	0.52

The threshold of distance $Dthr$ is set according to the model performance on the validation data set, as in Chapter 6. Smaller values for $Dthr$ will lead to less training error but the model will not generate the characteristics of the whole data set. The evaluation of the test events in 2013 and 2014 were shown in Table 7.2 and Table 7.3.

For the test event in 2013, the input selection 1L1M1H produced the highest Nash efficiency 0.74. The RMSE of 1L1M1H is the lowest 0.22m and reduced the RMSE of SAM to half. Overall the modified offline model with 1L1M1H input selection over estimated the measured water levels by 2% and other input selections reduced the 10% over-estimation of the SAM model. For the peak evaluation, the modified offline model produced similar evaluation results with 3%-6% under-estimation of the peaks and the water levels at the peak time.

Table 7.2 Evaluation of 2013 results for different inputs selected of the modified offline model

Criteria	SAM	15 Component models	1L1M1H	2L2M2H	1L2M3H
RMSE (m)	0.44	0.25	0.22	0.26	0.24
PEP (%)	0	-2	-3	-3	-6
PE (%)	-1	-5	-3	-3	-6
PT (h)	1	-1	1	0	0
PBIAS (%)	-10	1	-2	-4	-1
NSE	-0.08	0.65	0.74	0.63	0.67

Table 7.3 Evaluation of 2014 results for different inputs selected of the modified offline model

Criteria	SAM	15 Component models	1L1M1H	2L2M2H	1L2M3H
RMSE (m)	0.42	0.62	0.38	0.38	0.39
PEP (%)	11	38	3	2	3
PE (%)	-3	-10	-10	-11	-10
PT (h)	-4	-4	-3	-3	-3
PBIAS (%)	1	6	8	7	5
NSE	0.38	-0.34	0.50	0.50	0.48

The highest Nash efficiency was obtained by the input selection 1L1M1H and 2L2M2H which improved the SAM results from 0.38 to 0.5 as shown in Table 7.3. For the event in 2014, the model which used all the 15 component models failed to produce good ensemble results which produced the highest RMSE and negative Nash efficiency. The lowest RMSE was also produced by 1L1M1H and 2L2M2H which slightly improved the SAM results from 0.42 m to 0.38 m. For the estimation of the

maximum water levels, the 2L2M2H reached the lowest error with 2% over-estimation compared with the 11% over-estimation of the SAM model. The lowest PE value was obtained by SAM but it is because the falling limb of the SAM model just goes through the peak accidentally. The time shift of the SAM model was reduced from 4 hours to 3 hours in 1L1M1H, 2L2M2H and 1L2M3H input selections. Combining the evaluation results of the two events in 2013 and 2014, the modified offline model with 1L1M1H produced the best ensemble results. The results of 1L1M1H input selection were plotted in Figure 7.2.

(a)

(b)

Figure 7.2 Comparison of the forecasts from the WASH123D and Modified Offline model: (a) 2013 test event, (b) 2014 test event.

From Figure 7.2(a), it can be observed that the modified offline model can give a reasonably good prediction of the hydrograph and depicts the rising and falling limb better than the SAM. With the constraints on the weights and removal of the constant term, the modified offline model is not so sensitive to the input component models compared with the offline model in Chapter 6. Compared with the previous results in Figure 6.2(a), the initial ensemble results and the falling limb were greatly improved. With the constraints on the weights for each component model in the fuzzy rules and removal of the constant term, the modified offline model is less sensitive to the input variation and the spikes in component model forecasts will not be exaggerated in the modified offline ensemble outputs. The analysis performed in DENFIS is evaluated on normalized data. The data are first pre-processed and normalized using the historical highs and lows of the measured. The use of measured data for normalization is an acceptable procedure since measured data are generally available in most flood forecast applications. Although the raw (normalized) output of the DENFIS model

are subject to the weights (Eq. 7.3) satisfying the linear constraints; however, when scaled back to real-world values, it is possible for the ensemble model results to fall outside the range of the component models, because of use of the normalization parameters from the training data. For the test data in 2014 in Figure 7.2 (b), the modified offline model reduced the peak over-estimation and approached nearer to the measured peak. The visualization of the ensemble process of the modified offline model is shown in the next section.

With the linear constraints on the weights for the combined component models in each fuzzy rules, those parameters showed a clear pattern for the ensemble strategy. For the 1L1M1H input selection, the combined component models are Component model 1 and Component model 13. The optimized weight allocation for each rule was shown in Figure 7.3.

Figure 7.3 Weight allocation of the combined component models

With the Modified Offline model, the weights are constrained to between 0 and 1, indicating the relative importance of the component models. In the modified offline model, ten clusters were created in the normalized input space in Figure 7.3. Each pie chart in the figure is a cluster or fuzzy rule and the center of the clusters shows the position of each fuzzy rule in the normalized space. When the 2-dimensional input vectors are near to the cluster centers, the corresponding rules will be more applicable to the input data. The pattern is clear that when the normalized Component model1 forecasts higher water levels than those of the normalized Component model13, the weight of Component model1 will decrease and Component model13 almost dominates the weight allocation. This pattern was also shown in Figure 7.4 where the changes of the weights for each component model with the variation of the water levels were plotted.

(a)

(b)

Figure 7.4 Total weights of the normalized component model forecasts in the modified offline model for (a) 2013 test event (b) 2014 test event.

From Figure 7.4 (a), the weights of component model1 are almost zero at the peak in the second run when component model1 predicted higher water levels than those of Component model 13. After the peak, component model 13 predicted higher water levels and the weights of component model 1 increased to almost the same level as those of component model 13. component model 1 which produced better forecasting results was allocated increased weight. For the event of 2014 in Figure 7.4 (b), from 00 am to 06 am on July 23th before the peak, the component model 1 forecasted much higher water levels compared with component model 13. Correspondingly, the weight of component model 1 decreased almost to zero and the modified offline model allocated higher weights to component model 13 with better forecasting performance. In the modified offline model, the ensemble strategy was made based on not only the magnitude of the forecasts of the input component models, but also the relationship among the combined inputs.

7.3.2 Results of the Modified Offline with Slope Model

After training the modified offline with slope model, the results of the training and validation data as well as the optimized threshold of distance $Dthr$ were shown in Table 7.4.

Table 7.4 Training and validation RMSE of the modified offline model with slope for different input selections

Input Selections	Dthr	Training	Validation
15 Component models	0.19	0.58	0.43
1L1M1H	0.11	0.67	0.48
2L2M2H	0.1	0.64	0.48
1L2M3H	0.1	0.57	0.46

The error statistics associated with each of the selected component models for the testing dataset in 2013 and 2014 were compared with the case where all 15 component models were used as inputs to the ensemble in Table 7.5 and Table 7.6. Results of the SAM are included as benchmark. For 2013, all the ensemble models produced lower RMSE than the SAM results with around 50% decrease (see Table 7.5). The Nash-Sutcliffe efficiency was significantly improved to around 0.7 by the ensemble models and PBIAS was also greatly reduced by decreasing the over estimation from 10% to 3% at most. Both the 2L2M2H ensemble mode and SAM produced comparably good PE with a 1 hour lag of the peak water level. For the criteria of the overall evaluation, the ensemble models outperformed the SAM results. Even though for the event in 2013 the SAM model accidentally achieved very good peak estimation, the 2L2M2H input selection produced the comparable good peak forecasting results. The evaluation for the event in 2014 is shown in Table 7.6. The ensemble models produced similar overall statistics as the SAM results. The ensemble model with input selection 2L2M2H improved the peak value estimation

by decreasing the percent error in peak from 11% to 5% and time shift was decreased from -4 hours to -3 hours. The best ensemble results overall are from the input selection 2L2M2H, which may result from that the rising limbs and peaks of the two test events are within the range of middle water levels and the information from high water levels did not contribute much to the ensemble results for the test data. 2013 and 2014 were relatively drier years and it is possible that as a result, high water levels (H) did not contribute positively to the results and therefore 1L2M3H had slightly poorer performance.

Table 7.5 Evaluation of 2013 results for different inputs selected of the modified offline with slope model.

Criteria	SAM	15 Component models	1L1M1H	2L2M2H	1L2M3H
RMSE (m)	0.44	0.25	0.20	0.20	0.24
PEP (%)	0	-9	1	0	-5
PE (%)	-1	-11	1	-1	-7
PT (h)	1	-3	1	1	-1
PBIAS (%)	-10	-1	-2	-3	0
NSE	-0.08	0.66	0.79	0.78	0.69

Table 7.6 Evaluation of 2014 results for different inputs selected of the modified offline with slope model

Criteria	SAM	15 Component models	1L1M1H	2L2M2H	1L2M3H
RMSE (m)	0.42	0.43	0.43	0.41	0.47
PEP (%)	11	2	8	5	7
PE (%)	-3	-14	-8	-8	-8
PT (h)	-4	-4	-3	-3	-3
PBIAS (%)	1	7	7	6	4
NSE	0.38	0.35	0.36	0.41	0.23

Fifty-four clusters were created in the eight dimensional space consisting of the four water levels (M1, M7, M13, M14) and the four slopes (S1, S7, S13, S14) based on the 2L2M2H input selection. The results of the 2L2M2H ensemble model are evaluated in Figure 7.5. The figure includes the 5th and 95th percentiles of the results from the 15 component models (represented by the shaded area), and the SAM (averaging the results of all 15 component models) results. In Figure 7.5, the large range between the 5th percentile and 95th percentile represents the uncertainty associated with the water level forecasts provided by the 15 component models. The large disagreement of the component models' forecast is also discussed in (Shih *et al.*, 2014). Although the SAM seems to provide excellent predictions close to the peak in Figure 7.5(a), this is fortuitous and not expected all the time. The SAM is unable to provide reasonable estimates for the entire hydrograph largely because the fifteen component models are not able to predict the low water levels well in Figure 7.5(a).

Figure 7.5 Comparison of the forecasts from the WASH123D and modified offline with slope model : (a) 2013 test event, (b) 2014 test event

Some ensemble results fall out of the range of the combined component models, which is caused by the normalization factors from the training data. Compared with the results from the offline model in Figure 6.2(a) and the modified offline model in Figure 7.2(a), the modified offline with slope model produced the best ensemble

results for both the rising and falling limb as well as the peak. In Figure 7.5(b) the component models overestimated the rising limb and the peak of the test event in 2014, while the ensemble results produced some correction to better follow the hydrograph.

With the modifications to the offline ensemble model, the total weights in the output are constrained to between 0 and 1 which are calculated from the weights in the fuzzy rules and the firing strength, indicating the relative importance of the component models. The inputs were normalized so the value of the input for each dimension will be between 0 and 1. In each cluster, the weights are allocated to the four component models with sum to 1. For the dimensions of slopes, the boundaries of S1, S7, S13 and S14 are 0.30, 0.37, 0.37 and 0.40 respectively, which means the lower values than the boundaries are negative slopes. The clusters in the modified offline with slope model which allocate the highest weight to each component model were plotted in Figure 7.6. To visualize the location of each cluster center, radar plots were used with the eight lines from the center to the vertexes representing each dimension. Next to the radar plots of cluster centers, the fuzzy rule or weight allocation is plotted in pies correspondingly. So if the shape of an input vector in the radar plot is similar to one of the cluster center plot, the rule of the weight allocation in this cluster will be active with higher firing strength. The nearest several clusters are active with an input vector enters the space and the ensemble outputs are calculated based on the weight allocation and the firing strength.

Figure 7.6 Clusters with the highest weight allocation to each component model of the modified offline with slope model

Even though more clusters were used when the ensemble model processed the input data, some patterns in Figure 7.6 can also be found in real combination. From Figure 7.6, it is found that M7 was allocated the highest weight if M7 produced the lowest forecasts if most component models predicted it was on the rising limb. These patterns were found in Figure 7.7 which shows the total weights for the normalized component model forecasts for the testing event in 2013 and 2014 in the modified offline with slope model. In Figure 7.7(a), the increase of M7 appeared on the morning of Sep 22 when M13 and M14 forecasted a rising limb and M7 produced the lowest estimation. When all the component models predicted rising limb in Figure 7.7 (a) and 7.7 (b), M14 was always allocated higher weight, which corresponded to the pattern in Figure 7.6(d) where all the slopes exceed the slope boundaries meaning positive slopes.

(a)

(b)

Figure 7.7 Total weights of the normalized component model forecasts in the modified offline with slope model for (a) 2013 test event; (b) 2014 event.

From Figure 7.7(a), M1 always overestimated the measured water levels and the total weights stayed at a low level. The total weight of M14 reduced at the peak and the total weight of M13 increased, which is consistent with the observation that M13 gives a better prediction of the water level near the peak. The total weight of M7 stayed at a low level but increased at the falling limb with the water levels between around 3.6 m and 4.0 m. By automatically varying the total weight for each component model at different time steps, the ensemble model improved the forecasting results.

7.4 Conclusion

Two modified offline models of DENFIS were adopted as the ensemble approach to combine different estimates of precipitation corresponding to different perturbed initial and boundary conditions of the atmospheric states and cumulus scheme which were obtained from WRF model for the Lanyang Bridge, Yilan County, Taiwan.

1. A modified offline ensemble model which imposed linear constraints and removed the requirement of the constant term in fuzzy rules was proposed in the paper. Comparison of the results from the proposed modified ensemble model with the benchmark SAM showed that forecast results of the input 1L1M1H were superior to the SAM. Implementation of constraints in weight allocation highlighted the relative importance of the individual component models. The changes of weight allocation come from the knowledge learned about the performance at different forecasted water levels of the component models.

2. The effects of the slopes were considered in the modified offline with slope model by changing the clustering and rules creating process. The input selection based on 2L2M2H produced the best results. For the event in 2013, improvements in RMSE, PBIAS and the Nash-Sutcliffe efficiency were obtained by the 2L2M2H ensemble model over SAM results and all other error statistics were similar. For the event in

2014, the 2L2M2H ensemble model results showed improvements over SAM results for all the error statistics considered. The patterns of several clusters were visualized and demonstrated in the real time combination process. By visualizing the total weight changing at different time steps, it was observed that M13 was allocated higher total weight at the peak time with better performance than that of M14 and the total weights of M1 and M7 stayed at a low level because of the poor performance of the two component models.

3. The ensemble model results based on inputs consisting of all the fifteen component models were improved when compared to the SAM results. However, the identification of component model selection at different ranges of water levels resulted in a truncated input component model space, improving the ensemble model results.

4. Compared with the SAM, which allocates an average weight throughout all time steps, the modified neuro-fuzzy model provides a sound basis for weight apportionment in forecast predictions.

CHAPTER 8 CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

It is a common feature in flood modelling that existing models for flood forecasting are not able to model all phases of the hydrograph well, even though the global optimum may be reached. Therefore, in order to exploit the strengths of different models, the ensemble model approach can be used to increase forecast accuracy. A review of the literature revealed that there have been limited studies on ensemble methods in flood forecasting and statistical (SAM, WAM, BMA) and data-driven (FIS, ANN) methods have been used. These studies show that although global errors were decreased, the investigations carried out is still in its early stages and clearly much research is required.

Chapter 4-7 explored the ensemble approaches for two distinct cases for ensemble flood forecasting. The first catchment is a large river basin located in Lower Mekong which uses two different hydrological models with the same rainfall inputs. The other catchment is located in Taiwan where there is only one hydrological model but with different rainfall inputs. The conclusions from this study on ensemble approaches to optimize the water level forecasts for the two catchments can be summarized as:

1. User defined FIS structure versus clustering based FIS structure: The offline mode of DENFIS (DENFIS-EN) and ANFIS model with pruned rules (ANFIS-EN) were compared for the gauge station Kratie in Lower Mekong with a focus on the initial exploration into the model configuration for ensemble purpose. The underestimation of the ANFIS model and overestimation of the URBS model for the peak water levels are improved by the ensemble models with reduced RMSE compared to the component models. The time shift errors from the ANFIS model were almost eliminated in the ensemble predictions and the

strong oscillation present in URBS predictions is reduced. The best ensemble results were obtained by the DENFIS-EN model which reduced the RMSE from 1.11 m (ANFIS) and 1.01 m (URBS) to 0.86 m. Compared with ANFIS-EN of which the number of the membership functions were defined before model training, better results were obtained by DENFIS-EN model. Instead of arbitrarily defining the number of fuzzy sets for the ensemble model, more intelligent algorithm considering the data characteristics produced a more precise division of the forecasted values and the patterns of each combined models were easily distinguished.

2. Adaptation of the ensemble model can be enhanced by incremental learning: The ensemble model which utilized trained data driven models such as the offline mode of DENFIS model can produce improved ensemble results when the test data is within the range of the training data. When higher water levels that have never been used in the training period were used, the ensemble model failed to give good peak predictions. Retraining the ensemble model with more available data can be a solution. However, The EN-RTOFF model which retrained the whole data set each time using offline learning when the latest measured water levels were available could not produce significant improvements to the ensemble model. The offline learning which optimized all the clusters without considering the time order of the input data led to the small improvements when the data of the measured water levels used for updating were much less than the training data set. Incremental learning was attempted for the ensemble model with real time updating using online learning models, which would produce different clustering results based on the data sequence. In the testing phase, the online mode of DENFIS was switched on to update existing cluster, create new clusters and update fuzzy rules when the latest measured water levels were available. Statistical analysis of the models for all

the three stations indicated the superiority of the EN-RTON2 model over EN-RTOFF, EN-RTON1models, SAM and the ensemble model without real time updating. Not only were the spikes in the URBS model eliminated, but also the time shift problems in the ANFIS model results were decreased. By selecting correct updating interval for the ensemble model to be updated, including the incremental learning into the ensemble model makes the forecasting capable of handling events that are larger than historical floods with negligible computation time.

3. How to apply ensemble models to divergent model forecasts? For the Taiwan catchment in this thesis, the model forecasts to be combined are quite divergent because of the rainfall forecasts with different perturbation. Directly applying the ensemble model from previous chapters led to excessive sensitivity to the input data. By imposing linear constraints and removing the requirement of the constant term in fuzzy rules, the proposed modified offline model became less sensitive to the combined models and produced improved forecasts compared with the benchmark model SAM that is currently being adopted in real application. Implementation of constraints in weight allocation highlighted the relative importance of the component models. The changes of weight allocation come from the knowledge learned about the performance at different forecasted water levels of the component models. Compared with the SAM, which allocates an average weight throughout all time steps, the modified neuro-fuzzy models provide a sound basis for weight apportionment in forecast predictions.
4. Data pre-analysis and pre-processing for the ensemble model:
The over-fitting issue was addressed by considering pruning rules and creating

less clusters for the Lower Mekong catchment. Less parameters in the ensemble model provide stronger generalization ability with better performance in the testing phase. The input analysis and pre-processing were considered for the Lanyang Creek Basin. Before importing the component model forecasts to the ensemble model, all the component models were analyzed and pre-processed. It was found that the 15 component models are not able to consistently provide either short-term or long-term forecasts well. In this case, it is infeasible to train separate ensemble models for short- and long-term forecasts. Different input selections based on the performance of the component models at different water levels were tried for the ensemble model and the input selection 1L2M3H which addressed more on the higher water levels produced the best ensemble results compared to other input selections.

The innovation of this thesis can be summarized as:

1. Creatively propose real time updating algorithms based on the incremental learning, which showed strong adaptation and resilience.
2. Creatively propose modified NFIS ensemble models by incorporating linear constraints and the “black box” of the ensemble process was visualized.
3. Data pre-analysis and pre-processing were first proposed for the ensemble approach.

The significance of the studies can be summarized as:

1. The application of the ensemble approach based on more intelligent models for flood forecasting was validated.
2. The generalization of the ensemble flood forecasting was achieved from the experiments on two most often typical scenarios in practice. The proposed methodology can be extended to other catchments as well.

8.2 Practical Applications

The application of the proposed ensemble models in this thesis can be summarized as follows:

1. The first step should be data pre-analysis by evaluating the performance of the combinations of the component models for different forecast lead time. If there is any pattern among the combined models with respect to the lead time, separate ensemble models should be trained based on lead time. Then the combined models can be evaluated for different values of the forecasted hydrological variables and those models with poor or random performance should be removed from the input matrix.
2. According to different scenarios of the models to be combined, the suggested ensemble models would vary:
 - (i) If the scenario is that different hydrological models (different categories preferred) are adopted to forecast with the same weather or discharge inputs, the model performance may be quite different and correction with more flexibility may be needed.. The offline mode of the DENFIS is proposed of which the constant term is retained and the weights are free from constraints in the fuzzy rules to provide the ensemble model with higher sensitivity. Real time updating algorithm using online learning can be used so that new information can be included to make the model more flexible to the recent events, which produced the best ensemble results. The updating interval should be carefully selected based on whether the new information is corresponding to the forecasts from the ensemble models to be updated.
 - (ii) If there is only one hydrological model with different rainfall forecasts as the input and there was a large spread of the water level forecasts, the forecasts from the combined models have covered a large range of possible values of the forecasted variables In this case the ensemble

approaches need constraints instead of flexibility to avoid the effects brought by the randomness in the component inputs. With the modification of removing the constant term and adding constraints to the weights, the offline mode of DENFIS model can produce better ensemble results. As a result of these modification, the ensemble process can be interpreted more easily.

3. The error analysis of the hydrological models should cover both the overall statistics and peak evaluation. During the stage of input pre-analysis, too many criteria for the model evaluation may be too complex to find out the initial patterns. The use of algorithms to reduce the dimensions, such as PCA, the evaluation can be reached by calculating the distance from the model evaluation vectors to the vector of the “ideal” model in the input space with less dimensions.

8.3 Recommendations

1. This study focused on two types of ensemble forecasting (different rainfall-runoff models with the same rainfall inputs and different rainfall inputs but with the same rainfall runoff model) by analyzing the data from Lower Mekong and Taiwan catchment. The study can be extended to more basins in other locations with different catchment size and lead time to enrich the research on ensemble models for flood forecasting.
2. The NFIS used in this thesis as the ensemble approach can be used in real time updating version with stronger adaptability or modified offline version with model interpretation. The improved results of the ensemble models based on NFIS indicate the possibility of using other data driven models for ensemble approach.

3. For the case of different rainfall inputs but with the same rainfall runoff model, the pre-analysis and pre-processing used in this thesis for the component models showed improved ensemble results. Further studies can focus on more methods of the pre-analysis and pre-processing before applying the ensemble model.

REFERENCES

- Amiri, E., 2015. Forecasting daily river flows using nonlinear time series models. *Journal of Hydrology*, 527(0): 1054-1072.
- Anctil, F., Perrin, C., Andréassian, V., 2003. Ann output updating of lumped conceptual rainfall/runoff forecasting models. *Journal of the American Water Resources Association*, 39(5): 1269-1279.
- Araghinejad, S., Azmi, M., Kholghi, M., 2011. Application of artificial neural network ensembles in probabilistic hydrological forecasting. 407(1-4): 94-104.
- Asadi, S., Shahrabi, J., Abbaszadeh, P., Tabanmehr, S., 2013. A new hybrid artificial neural networks for rainfall–runoff process modeling. *Neurocomputing*, 121(0): 470-480.
- Badrzadeh, H., Sarukkalige, R., Jayawardena, A.W., 2013. Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting. *Journal of Hydrology*, 507(0): 75-85.
- Bennett, J.C. et al., 2014. A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9 days. *Journal of Hydrology*, 519,Part D: 2832-2846.
- Beven, K., Kirkby, M., 1979. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal*, 24(1): 43-69.
- Beven, K.J., Leedal, D.T., Smith, P.J., Young, P.C., 2012. Identification and Representation of State Dependent Non-linearities in Flood Forecasting Using the DBM Methodology, *System Identification, Environmental Modelling, and Control System Design*, Springer London. In: Wang, L., Garnier, H. (Eds.). Springer London, pp. 341-366.
- Bhattacharya, B., Lobbrecht, A.H., Solomatine, D.P., 2003. Neural networks and

- reinforcement learning in control of water systems. *Journal of Water Resources Planning and Management*, 129(6): 458-465.
- Biondi, D., De Luca, D.L., 2012. A Bayesian approach for real-time flood forecasting. *Physics and Chemistry of the Earth, Parts A/B/C*, 42–44(0): 91-97.
- Bishop, C.M., 2006. *Pattern recognition and machine learning*, 4. springer New York.
- Blöschl, G., Reszler, C., Komma, J., 2008. A spatially distributed flash flood forecasting model. *Environmental Modelling & Software*, 23(4): 464-478.
- Black, T.L., 1994. The New NMC Mesoscale Eta Model: Description and Forecast Examples. *Weather and Forecasting*, 9(2): 265-278.
- Bogner, K., Kalas, M., 2008. Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. *Atmospheric Science Letters*, 9(2): 95-102.
- Braud, I. et al., 2010. The use of distributed hydrological models for the Gard 2002 flash flood event: Analysis of associated hydrological processes. *Journal of Hydrology*, 394(1–2): 162-181.
- Bravo, J.M. et al., 2009. Incorporating forecasts of rainfall in two hydrologic models used for medium-range streamflow forecasting. *Journal of Hydrologic Engineering*, 14(5): 435-445.
- Calvo, B., Savi, F., 2009. Real-time flood forecasting of the Tiber river in Rome. *Natural Hazards*, 50(3): 461-477.
- Carroll, D., 2007. *The manual for URBS: A Rainfall Runoff Routing Model for Flood Forecasting and Design*.
- Chang, L.C., Shen, H.Y., Wang, Y.F., Huang, J.Y., Lin, Y.T., 2010. Clustering-based hybrid inundation model for forecasting flood inundation depths. *Journal of Hydrology*, 385(1-4): 257-268.
- Chao, Z., Hua-sheng, H., Wei-min, B., Luo-ping, Z., 2008. Robust recursive estimation of auto-regressive updating model parameters for real-time flood forecasting. *Journal of Hydrology*, 349(3–4): 376-382.

- Chen, L. et al., 2015. Real-time error correction method combined with combination flood forecasting technique for improving the accuracy of flood forecasting. *Journal of Hydrology*, 521(0): 157-169.
- Chen, S.-H., Lin, Y.-H., Chang, L.-C., Chang, F.-J., 2006. The strategy of building a flood forecast model by neuro-fuzzy network. *Hydrological Processes*, 20(7): 1525-1540.
- Chen, S.-T., Yu, P.-S., 2007. Real-time probabilistic forecasting of flood stages. *Journal of Hydrology*, 340(1–2): 63-77.
- Chen, Y.-h., Chang, F.-J., 2009. Evolutionary artificial neural networks for hydrological systems forecasting. *Journal of Hydrology*, 367(1–2): 125-137.
- Chua, L.H., 2012. Considerations for data-driven and physically-based hydrological models in flow forecasting. *System Identification*, 16(1): 1025-1030.
- Cole, S.J., Moore, R.J., 2009. Distributed hydrological modelling using weather radar in gauged and ungauged basins. *Advances in Water Resources*, 32(7): 1107-1120.
- Coulibaly, P., Bobee, B., Anctil, F., 2001. Improving extreme hydrologic events forecasting using a new criterion for artificial neural network selection. *Hydrological Processes*, 15: 1533-1536.
- Der Voort, M., Dougherty, M., 1996. Combining Kohonen maps with arima time series models to forecast traffic flow. *Transportation Research: Part C*, 4(5): 307.
- Du, J. et al., 2012. Assessing the effects of urbanization on annual runoff and flood events using an integrated hydrological modeling system for Qinhuai River basin, China. *Journal of Hydrology*, 464–465(0): 127-139.
- Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, 30(5): 1371-1386.
- Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization

- for conceptual rainfall- runoff models. *Water Resources Research*, 28(4): 1015-1031.
- Edijatno, NASCIMENTO, N.D.E.O., YANG, X., MAKHLOUF, Z., MICHEL, C., 1999. GR3J: a daily watershed model with three free parameters. *Hydrological Sciences Journal*, 44(2): 263-277.
- Elsafi, S.H., 2014. Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. *Alexandria Engineering Journal*, 53(3): 655-662.
- Erdal, H.I., Karakurt, O., 2013. Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms. *Journal of Hydrology*, 477(0): 119-128.
- Fares, A. et al., 2014. Rainfall-runoff modeling in a flashy tropical watershed using the distributed HL-RDHM model. *Journal of Hydrology*, 519, Part D(0): 3436-3447.
- Fordham, D.A., Wigley, T.M.L., Brook, B.W., 2011. Multi-model climate projections for biodiversity risk assessments. *Ecological Applications*, 21(8): 3317-3331.
- Franchini, M., Pacciani, M., 1991. Comparative analysis of several conceptual rainfall-runoff models. *Journal of Hydrology*, 122(1-4): 161-219.
- Garambois, P.A., Roux, H., Larnier, K., Labat, D., Dartus, D., 2015. Parameter regionalization for a process-oriented distributed model dedicated to flash floods. *Journal of Hydrology*, 525(0): 383-399.
- Ghose, D.K., Panda, S.S., Swain, P.C., 2013. Prediction and optimization of runoff via ANFIS and GA. *Alexandria Engineering Journal*, 52(2): 209-220.
- Giannoni, F., Roth, G., Rudari, R., 2000. A semi-distributed rainfall-runoff model based on a geomorphologic approach. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 25(7-8): 665-671.
- Goerss, J.S., 2000. Tropical cyclone track forecasts using an ensemble of dynamical models. *Monthly Weather Review*, 128(4): 1187-1193.

- Goswami, M., O'Connor, K.M., 2007. Real-time flow forecasting in the absence of quantitative precipitation forecasts: A multi-model approach. *Journal of Hydrology*, 334(1-2): 125-140.
- Gragne, A.S., Alfredsen, K., Sharma, A., Mehrotra, R., 2015. Recursively updating the error forecasting scheme of a complementary modelling framework for improved reservoir inflow forecasts. *Journal of Hydrology*, 527(0): 967-977.
- Green, I.R.A., Stephenson, D., 1986. Criteria for comparison of single event models. *Hydrological Sciences Journal*, 31(3): 395-411.
- Gupta, H., Sorooshian, S., Yapo, P., 1999. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. *Journal of Hydrologic Engineering*, 4(2): 135-143.
- Höglind, M., Thorsen, S.M., Semenov, M.A., 2012. Assessing uncertainties in impact of climate change on grass production in Northern Europe using ensembles of global climate models. *Agricultural and Forest Meteorology*, 170(0): 103-113.
- Han, D., Kwong, T., Li, S., 2007. Uncertainties in real-time flood forecasting with neural networks. *Hydrological Processes*, 21(2): 223-228.
- Hargreaves, J.C., Annan, J.D., 2006. Using ensemble prediction methods to examine regional climate variation under global warming scenarios. *Ocean Modelling*, 11(1-2): 174-192.
- He, Z., Wen, X., Liu, H., Du, J., 2014. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*, 509(0): 379-386.
- Hsiao, L.-F. et al., 2013. Ensemble forecasting of typhoon rainfall and floods over a mountainous watershed in Taiwan. *Journal of Hydrology*, 506(0): 55-68.
- IPCC, 2012. Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of Working Groups I and II of the

- intergovernmental panel on climate change.[Field, C.B., V. Barros, T.F. Stocker,D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen,M. Tignor, and P.M. Midgley (eds.)] Cambridge University Press, Cambridge, UK, and New York, NY, USA.
- Jakeman, A., Littlewood, I., Whitehead, P., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of Hydrology*, 117(1): 275-300.
- Jang, J.S.R., 1993. ANFIS: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3): 665-685.
- Jang, J.S.R., 1997. *Fuzzy inference systems*. Upper Saddle River, NJ: Prentice-Hall.
- Jasper, K., Gurtz, J., Lang, H., 2002. Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. *Journal of Hydrology*, 267(1–2): 40-52.
- Jolliffe, I., 2005. *Principal Component Analysis*, *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd.
- Kalteh, A.M., 2013. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Computers & Geosciences*, 54(0): 1-8.
- Kasabov, N.K., Song, Q., 2002. DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *Fuzzy Systems, IEEE Transactions on*, 10(2): 144-154.
- Kasiviswanathan, K.S., Cibin, R., Sudheer, K.P., Chaubey, I., 2013. Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations. *Journal of Hydrology*, 499(0): 275-288.
- Keskin, M.E., Taylan, D., Terzi, Ö., 2006. Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series. *Hydrological Sciences Journal*, 51(4): 588-598.
- Krzysztofowicz, R., 2002. Bayesian system for probabilistic river stage forecasting.

- Journal of Hydrology, 268(1–4): 16-40.
- Lasserre, F., Razack, M. and Banton, O., 1999. A GIS-linked model for the assessment of nitrate contamination in groundwater. *Journal of hydrology*, 224(3): pp.81-90.
- Lee, J.H., Labadie, J.W., 2007. Stochastic optimization of multireservoir systems via reinforcement learning. *Water Resources Research*, 43(11): W11408.
- Lerat, J., Perrin, C., Andréassian, V., Loumagne, C., Ribstein, P., 2012. Towards robust methods to couple lumped rainfall–runoff models and hydraulic models: A sensitivity analysis on the Illinois River. *Journal of Hydrology*, 418–419(0): 123-135.
- Lhennitte, S., Verbesselt, J., Verstraeten, W. W., & Coppin, P. 2011. A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. *Remote Sensing of Environment*, 115(12): 3129-3152. doi:10.1016/j.rse.2011.06.020
- Lin, G.-F., Jhong, B.-C., Chang, C.-C., 2013. Development of an effective data-driven model for hourly typhoon rainfall forecasting. *Journal of Hydrology*, 495(0): 52-63.
- Lohani, A.K., Kumar, R., Singh, R.D., 2012. Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *Journal of Hydrology*, 442–443(0): 23-35.
- Mahalanobis, P. C. 1936. On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1): 49-55.
- Martina, M.L.V., Todini, E., Liu, Z., 2011. Preserving the dominant physical processes in a lumped hydrological model. *Journal of Hydrology*, 399(1–2): 121-131.
- McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2): 239-245.
- Mohammadi, K., Eslami, H.R., Kahawita, R., 2006. Parameter estimation of an

- ARMA model for river flow forecasting using goal programming. *Journal of Hydrology*, 331(1–2): 293-299.
- Moreno, H.A., Vivoni, E.R., Gochis, D.J., 2012. Utility of Quantitative Precipitation Estimates for high resolution hydrologic forecasts in mountain watersheds of the Colorado Front Range. *Journal of Hydrology*, 438-439: 66-83.
- MRC, 2005. Overview of the Hydrology of the Mekong Basin. In: Commission, M.R. (Ed.), Vientiane, pp. 73.
- MRC, 2007. Annual Mekong Flood Report 2006, Mekong River Commission, Vientiane.
- MRC, 2009. System Performance Evaluation Report. In: The MRC Technical Task Group for verification of the new MRC Mekong Flood Forecasting System (FEWS-URBS-ISIS), M.R.C.R.F.M.a.M.C. (Ed.), Phnom Penh, Cambodia.
- Napolitano, G., See, L., Calvo, B., Savi, F., Heppenstall, A., 2010. A conceptual and neural network model for real-time flood forecasting of the Tiber River in Rome. *Physics and Chemistry of the Earth, Parts A/B/C*, 35(3-5): 187-194.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3): 282-290.
- Nasr, A., Bruen, M., 2008. Development of neuro-fuzzy models to account for temporal and spatial variations in a lumped rainfall–runoff model. *Journal of Hydrology*, 349(3–4): 277-290.
- Nayak, P.C., Sudheer, K.P., Rangan, D.M., Ramasastri, K.S., 2004. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291(1–2): 52-66.
- Nayak, P.C., Sudheer, K.P., Rangan, D.M., Ramasastri, K.S., 2005. Short-term flood forecasting with a neurofuzzy model. *Water Resour. Res.*, 41(4): W04004.
- Nguyen, P.K.T., Chua, L.H.C., 2012. The data-driven approach as an operational real-time flood forecasting model. *Hydrological Processes*, 26(19): 2878-2893.
- Norbiato, D., Borga, M., Dinale, R., 2009. Flash flood warning in ungauged basins

- by use of the flash flood guidance and model-based runoff thresholds. *Meteorological Applications*, 16(1): 65-75.
- Nourani, V., Komasi, M., 2013. A geomorphology-based ANFIS model for multi-station modeling of rainfall–runoff process. *Journal of Hydrology*, 490(0): 41-55.
- Orth, R., Staudinger, M., Seneviratne, S.I., Seibert, J., Zappa, M., 2015. Does model performance improve with complexity? A case study with three hydrological models. *Journal of Hydrology*, 523(0): 147-159.
- Pagano, T.C., 2013. Evaluation of Mekong River Commission operational flood forecasts, 2000–2012. *Hydrol. Earth Syst. Sci. Discuss*, 10(14):433-14.
- Piotrowski, A.P., Napiorkowski, J.J., 2013. A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling. *Journal of Hydrology*, 476(0): 97-111.
- Poelmans, L., Rompaey, A.V., Ntegeka, V., Willems, P., 2011. The relative impact of climate change and urban expansion on peak flows: a case study in central Belgium. *Hydrological Processes*, 25(18): 2846-2858.
- Puente, C.E., Bras, R.L., 1987. Application of nonlinear filtering in the real time forecasting of river flows. *Water Resources Research*, 23(4): 675-682.
- Rathinasamy, M., Adamowski, J., Khosa, R., 2013. Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. *Journal of Hydrology*, 507(0): 186-200.
- Reggiani, P., Weerts, A.H., 2008. A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine. *Journal of Hydrology*, 356(1–2): 56-69.
- Seann, R., John, S., Ziya, Z., 2007. A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *Journal of Hydrology*, 337: 402-420.
- See, L., Openshaw, S., 2000. A hybrid multi-model approach to river level forecasting.

- Hydrological Sciences Journal, 45(4): 523-536.
- Seiller, G., Hajji, I., Anctil, F., 2015. Improving the temporal transposability of lumped hydrological models on twenty diversified U.S. watersheds. *Journal of Hydrology: Regional Studies*, 3(0): 379-399.
- Seni, G., Elder, J.F., 2010. Ensemble methods in data mining: improving accuracy through combining predictions. *Synthesis lectures on data mining and knowledge discovery*. Morgan & Claypool Publishers, San Rafael, Calif.
- Seo, Y., Kim, S., Kisi, O., Singh, V.P., 2015. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology*, 520(0): 224-243.
- Shabani, N., 2009. Incorporation flood control rule curves of the columbia river hydroelectric system in a multireservoir reinforcement learning optimization model. Master Thesis, University of British Columbia.
- Shamir, E., Lee, B.J., Bae, D.H., Georgakakos, K.P., 2010. Flood forecasting in regulated basins using the ensemble extended Kalman filter with the storage function method. *Journal of Hydrologic Engineering*, 15(12): 1030-1044.
- Shamseldin, A.Y., 1997. Application of a neural network technique to rainfall-runoff modelling. *Journal of Hydrology*, 199: 272-294.
- Shamseldin, A.Y., O'Connor, K.M., 1999. A real-time combination method for the outputs of different rainfall-runoff models. *Hydrological Sciences Journal*, 44(6): 895-912.
- Shamseldin, A.Y., O'Connor, K.M., 2003. A "consensus" real-time river flow forecasting model for the Blue Nile River, *Water Resources Systems-Hydrological Risk, Management and Development*, Sapporo.
- Shamseldin, A.Y., O'Connor, K.M., Liang, G.C., 1997. Methods for combining the outputs of different rainfall-runoff models. *Journal of Hydrology*, 197(1-4): 203-229.
- Shao, Q., Wong, H., Li, M., Ip, W.-C., 2009. Streamflow forecasting using functional-

- coefficient time series model with periodic variation. *Journal of Hydrology*, 368(1–4): 88-95.
- Shen, H., Chang, L., 2012. On-line multistep-ahead inundation depth forecasts by recurrent NARX networks. *Hydrol. Earth Syst. Sci. Discuss*, 9: 11999-12028.
- Shih, D.-S., Chen, C.-H., Yeh, G.-T., 2014. Improving our understanding of flood forecasting using earlier hydro-meteorological intelligence. *Journal of Hydrology*, 512(0): 470-481.
- Shih, D.S., Yeh, G.T., 2010. Identified model parameterization, calibration, and validation of the physically distributed hydrological model WASH123D in Taiwan. *Journal of Hydrologic Engineering*, 16(2): 126-136.
- Shultz, M.J., 2007. Comparison of Distributed Versus Lumped Hydrologic Simulation Models Using Stationary and Moving Storm Events Applied to Small Synthetic Rectangular Basins and an Actual Watershed Basin. PhD Thesis, The University of Texas at Arlington.
- Silvestro, F., Rebora, N., 2014. Impact of precipitation forecast uncertainties and initial soil moisture conditions on a probabilistic flood forecasting chain. *Journal of Hydrology*, 519, Part A(0): 1052-1067.
- Silvestro, F., Rebora, N., Ferraris, L., 2011. Quantitative flood forecasting on small- and medium-sized basins: A probabilistic approach for operational purposes. *Journal of Hydrometeorology*, 12(6): 1432-1446.
- Siou, L.K.A., Johannet, A., Borrell, V., Pistre, S., 2011. Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). *Journal of Hydrology*, 403(3–4): 367-380.
- Skamarock, W.C. et al., 2005. A Description of the Advanced Research WRF Version 2, NCAR Tech. Note NCAR/TN-468+STR. <http://box.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf>. 88: 7--25.
- Sobol', I.M., 1993. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.*, 1(4): 407-417.

- Şorman, A.A., Şensoy, A., Tekeli, A.E., Şorman, A.Ü., Akyürek, Z., 2009. Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of Turkey. *Hydrological Processes*, 23(7): 1031-1040.
- Sorooshian, S., Gupta, V.K., 1983. Automatic calibration of conceptual rainfall-runoff models: the question of parameter observability and uniqueness. *Water Resources Research*, 19(1): 260-268.
- Stevens, B.H., Treyz, G.I., Ehrlich, D.J. and Bower, J.R., 1983. A new technique for the construction of non-survey regional input-output models. *International Regional Science Review*, 8(3):271-286.
- Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics, IEEE Transactions on, SMC-15*(1): 116-132.
- Talei, A., Chua, L.H.C., Quek, C., Jansson, P.-E., 2013. Runoff forecasting using a Takagi–Sugeno neuro-fuzzy model with online learning. *Journal of Hydrology*, 488(0): 17-32.
- Talei, A., Chua, L.H.C., Wong, T.S.W., 2010. Evaluation of rainfall and discharge inputs used by Adaptive Network-based Fuzzy Inference Systems (ANFIS) in rainfall-runoff modeling. *Journal of Hydrology*, 391: 248-262.
- Tang, Y., Reed, P., Van Werkhoven, K., Wagener, T., 2007. Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. *Water Resources Research*, 43(6): W06415.
- Tingsanchali, T., Gautam, M.R., 2000. Application of tank, NAM, ARMA and neural network models to flood forecasting. *Hydrological Processes*, 14(14): 2473-2487.
- Todini, E., 2007. Hydrological catchment modelling: past, present and future. *Hydrology and Earth System Sciences*, 11(1): 468-482.
- Tospornsampan, M.J., Malone, T., Katry, P., Pengel, B., An, H.P., 2009. Short and Medium-term Flood Forecasting at the Regional Flood Management and

- Mitigation Center. AMFF-7 Intergrated Flood Risk Management in the Mekong River Basin, Bangkok, Thailand, pp. 155-164.
- Toth, E., Brath, A., Montanari, A., 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of Hydrology*, 239(1–4): 132-147.
- Toth, E., Montanari, A., Brath, A., 1999. Real-time flood forecasting via combined use of conceptual and stochastic models. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 24(7): 793-798.
- Toth, Z., Kalnay, E., 1997. Ensemble Forecasting at NCEP and the Breeding Method. *Monthly Weather Review*, 125(12): 3297-3319.
- Trudel, M., Leconte, R., Paniconi, C., 2014. Analysis of the hydrological response of a distributed physically-based model using post-assimilation (EnKF) diagnostics of streamflow and in situ soil moisture observations. *Journal of Hydrology*, 514(0): 192-201.
- Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of Hydrology*, 476(0): 433-441.
- Van Steenberghe, N., Ronsyn, J., Willems, P., 2012. A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication. *Environmental Modelling & Software*, 33(0): 92-105.
- Vieira, J., Føns, J., Cecconi, G., 1993. Statistical and hydrodynamic models for the operational forecasting of floods in the Venice Lagoon. *Coastal Engineering*, 21(4): 301-331.
- Wei, C.-C., 2015. Comparing lazy and eager learning models for water level forecasting in river-reservoir basins of inundation regions. *Environmental Modelling & Software*, 63(0): 137-155.
- Wilke, K., Barth, F., 1991. Operational river-flood forecasting by Wiener and Kalman

- filtering. Hydrology for the water management of large river basins. Proc. symposium, Vienna, 1991: 391-400.
- Wolfs, V., Meert, P., Willems, P., 2015. Modular conceptual modelling approach and software for river hydraulic simulations. *Environmental Modelling & Software*, 71(0): 60-77.
- Xiong, L.H., Shamseldin, A.Y., O'Connor, K.M., 2001. A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system. *Journal of Hydrology*, 245(1-4).
- Yang, H., Wang, B., 2012. Reducing biases in regional climate downscaling by applying Bayesian model averaging on large-scale forcing. *Climate Dynamics*, 39(9-10): 2523-2532.
- Yang, T.-H. et al., 2015. Flash flood warnings using the ensemble precipitation forecasting technique: A case study on forecasting floods in Taiwan caused by typhoons. *Journal of Hydrology*, 520(0): 367-378.
- Yao, C., Zhang, K., Yu, Z., Li, Z., Li, Q., 2014. Improving the flood prediction capability of the Xinanjiang model in ungauged nested catchments by coupling it with the geomorphologic instantaneous unit hydrograph. *Journal of Hydrology*, 517(0): 1035-1048.
- Yeh, G.T., Cheng, H.P., Cheng, J.R., Lin, J.H., 1998. A numerical model to simulate flow and contaminant and sediment transport in watershed systems (WASH123D). Technical Rep. CHL-98-15, Waterways Experiment Station, U.S. Army Corps of Engineers, Vicksburg, MS 39180-6199.
- Young, P., Wallis, S., 1985. Recursive estimation: A unified approach to the identification estimation, and forecasting of hydrological systems. *Applied Mathematics and Computation*, 17(4): 299-334.
- Young, P.C., 2002. Advances in real-time flood forecasting. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 360(1796): 1433-1450.

- Yu, P.S., Yang, T.C., 1997. A probability-based renewal rainfall model for flow forecasting. *Natural Hazards*, 15(1): 51-70.
- Yuan, H. et al., 2009. Evaluation of short-range quantitative precipitation forecasts from a time-lagged multimodel ensemble. *Weather and Forecasting*, 24(1): 18-38.
- Zhang, X., Liang, F., Yu, B., Zong, Z., 2011. Explicitly integrating parameter, input, and structure uncertainties into Bayesian Neural Networks for probabilistic hydrologic forecasting. *Journal of Hydrology*, 409(3–4): 696-709.
- Zhao, T. et al., 2015. Quantifying predictive uncertainty of streamflow forecasts based on a Bayesian joint probability model. *Journal of Hydrology*, 528(0): 329-340.
- Zhu, Y., Li, X.R., 2007. Recursive Least Squares with Linear Constraints. *Communications in Information & Systems* 7(3): 287-312.