An Exploration of Neural Network Modelling Options for the Upper River Ping, Thailand

Tawee Chaipimonplin

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Geography

August, 2010

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to give my sincere thanks to Dr. Linda See and Professor Pauline Kneale, who accepted me as a PhD student, also thanks for your supporting, advising and helping me to develop my research. You are incredible and perfect combination supervisors for my research as Linda is expert on Neural Network and Pauline is expert on Hydrology. I am really appreciated your times and patience for spending lots of times for reading, checking and editing my writing. I am in debt both of you. You are like a candle in the dark who guide me walk through my research, with out your help I would not complete it.

I also wish to thank Professor Phill Rees for his advice when I was the first year PhD student and the RSG member teams; Dr. Katherine Arrell, Dr. Andy Evans and Dr. Avijit Gupta for your comments. Thanks to Mrs. Jacqui Manton for her conversation class and my fellow postgraduate colleagues, in particular to Giulia Napolitano who is my buddy of learning Matlab, Faith Chan, Eran Sadek, Neil Dickson, Pakorn Meksangsouy, Rukchanok & Chansak Karcharnubarn, Prapaipan Phingchim, Bundit Boonkao, Sikarin Yookung, Yan Hong and Young Mi Song who provide the friendship and the useful discussion.

This thesis was undertaken with the financial support from Faculty of Social Sciences and Chiang Mai University, Thailand. I would like to express my appreciation to my colleagues in Geography Department, Chiang Mai University especially, Mr. Chira Prangkio who has supported me since I worked as a lecturer at Chiang Mai University. I also thank Mr. Thada Sukhapunnaphan, the director of Hydrology and Water Management Centre for Upper Northern Region, Royal Irrigation Department for the hydrology data, Mr. Chanti Detyohin for the Radar Extracted program and Mr. Sinchai Peungtambon for taking me to meet the director of the Bureau of Royal Rainmaking and Agricultural Aviation who gave me access to the radar imagery.

Finally I wish to thank my parents; Mr. Pheera and Mrs. Nitdawan, also all my family members, my brother and his wife; Winyu and Niracha for their support during the research and encouragement when I was very depressed.

Abstract

This thesis reports results from a systematic experimental approach to evaluating aspects of the neural network modelling process to forecast river stage for a large, 23,600 km^2 catchment in northern Thailand. The research is prompted by the absence of evidenced recommendations as to which of the array of input processes, validations and modelling procedures might be selected by a neural network forecaster.

The flood issue for forecasters at Chiang Mai derives from the monsoon rainfall, which leads to serious out-of-bank flooding two to four times a year. Data for stage and rainfall is limited as the instrumentation is sparse and the historical flood record is limited in length. Neural network forecasting models are potentially very powerful forecasters where the data are limited. The challenge of this catchment is to provide adequate forecasts from data for relatively few storm events using three stage gauges and one rain gauge. Previous studies have reported forecasts with lead times of up to 18 hours. Thus, one research driver is to extend this lead time to give more warning.

Eight input determination methods were systematically evaluated through thousands of model runs. The most successful method was found to be correlation and stepwise regression although the pattern was not consistent across all model runs.

Cloud radar imagery was available for a few storm events. Rainfall data from a network was not available so it was decided to explore the value of the raw cloud reflectivity data as a catchment-wide surrogate for rainfall, to enhance the data record and potentially improve the forecast. The limited number of events makes drawing conclusions difficult, but for one event the forecast lead time was extended to 24-30 hours. The modelling also indicates that for this catchment where the monsoon may come from the south west or the north east, the direction of storm travel is important, indicating that developing two neural network models may be more appropriate.

Internal model training and parameterisation tests suggest that future models should use Bayesian Regularization, and average across 50 runs. The number of hidden nodes should be less than the number input variables although for more complex problems, this was not necessarily the case. Ranges of normalisation made little difference. However, the minimum and maximum values used for normalisation appear to more important.

The strength of the conclusions to be drawn from this research was recognised from the start as being limited by the data, but the results suggest that neural networks are both helpful modelling processes and can provide valuable forecasts in catchments with extreme rainfall and limited hydrological data. The systematic investigation of the alternative input determination methods, algorithms and internal parameters has enabled guidance to be given on appropriate model structures.

Table of contents

List of figures

List of tables

Glossary

Mathematical symbols

A_c	Area of cloud
D	Drop diameter, Degree of seasonal differencing
R_v	Volumetric rain rate
Vc	Pulse volume
$Wv(t)$	updating of a neuron with weight vector
Z	Radar reflectivity factor
a, b	The value depend on type of rainfall
a_o, a_1	Empirical coefficients
dA_c/d_t	Rate of change in could area

Acronyms

A	All inputs
AME	Absolute Mean Error
ANN	Artificial Neural Network
ASCE	American Society of Civil Engineers
BPNN	Backpropagation Neural Network
BR	Bayesian Regularization
C	Method of selecting input variable based on correlation
CAPPI	Constant Altitude Plan Position Indicator
CCF	Cross Correlation Function
CCNNM	Cascade Correlation Neural Network Model
CE	Coefficient of Efficiency
CENDRU	Civil Engineering Department Chiang Mai University Natural Disasters Unit
CS	Method of selecting input variable combination between correlation and stepwise regression
D	Method of selecting input variable based on data mining
dBZ	A measure of the radar reflectivity factor (Z) in logarithmic units
G	Method of selecting input variable base on genetic algorithm
GA	Genetic Algorithm
GAGNN	Genetic Algorithm General Regression Neural Network
GRNN	General Regression Neural Network
M	Method of selecting input variable based on M5 model tree
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
MSE	Mean Squared Error
NN	Neural Network
P	Method of selecting input variable based on PMI
P1	Water level station number and it is the primary target to predict
P21, P4a	Upper water level station numbers located at tributary
P67, P75	Upper water level station numbers located at the main river
PDF	Probability Density Function
PDIFF	Peak Difference
PMI	Partial Mutual Information
PPI	Plan Position Indictor
Pr	Method of selecting input variable based on pruning algorithm
R	Correlation Coefficient
R1	Rain gauge station
R^2	Coefficient of Determination
RBF	Radial Basis Function
RBFN	Radial Basis Function Network

RMSE	Root Mean Squared Error
S1	Storm in 2003
S2-S6	Storms in 2005
S7, S8	Storms in 2006
S	Stepwise Regression
SHE	Type of physically based model (System Hydrologique European)
SOLO	Self Organizing Linear Map
SOM	Self Organizing Maps
WinNN	Window Neural Network
Z	Radar reflectivity
Z-R	The empirical relationships between radar reflectivity (Z) and rainfall rate (R)

Definitions

Architecture
: Description of the number of the layers in a neural network, each layer apply with transfer function, each layer contain numbers of node and connect node between layers.

Backpropagation learning rule
: Weights and biases are adjusted by error-derivative backpropagated through the network.

Bayesian framework
: Assumes that the weights and biases of the network are random variables with specified distributions.

Bias
: Node's parameter that is summed with the node's weighted inputs and passed through the node's transfer function to generate the node's output.

C band
: Radar wavelength of 5 cm.

Doppler radar
: A type of radar that measures not only the intensity of a returned signal but Doppler shift and hence the radial velocity of the target.

Early stopping
: Technique to avoid the overfit, it based on dividing the data into three subsets. The first subset is the training set, used for computing the gradient and updating the network weights and biases. The second subset is the validation set. When the validation error increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned. The third subset is the test set. It is used to verify the network design.

Epoch
: Presentation of the set of training.

Feedforward network
: Layered network in which each layer only receives inputs from previous layers.

Generalization
: Attribute of a network whose output for a new input vector tends to be close to outputs for similar input vectors in its training set.

Global minimum
: Lowest value of a function over the entire range of its input parameters.

Gradient descent
: Process of changing the weights and biases, where the changes are proportional to the derivatives of network error with respect to those weights and biases.

Initialization
: Process of setting the network weights and biases to their original values.

Input layer
: Layer of node receiving inputs directly from outside the network.

Input vector
: Input of a neural network. Each element of the input vector is the input of next layer node. Input of this study is water stage, radar data and rainfall.

Input weights
: Weights connecting network inputs to layers.

Kohonen learning rule layer
: Learning rule that trains a selected node's weight vectors to take on the values of the current input vector.

Learning or Training
: Process by which weights and biases are adjusted to achieve some desired network behaviour.

Learning rate
: Training parameter that controls the size of weight and bias changes during learning.

Levenberg-Marquardt
: Algorithm that trains a neural network 10 to 100 times faster than the usual gradient descent backpropagation method. Ti always computes the approximate Hessian matrix, which has dimensions *n*-by-*n*.

Linear transfer function
: Transfer function that produces its input as its output.

Local minimum
: Minimum of a function over a limited range of input values. A local minimum might not be the global minimum.

Log-sigmoid transfer function
: Squashing function of the form shown below that maps the input to the interval (0, 1).

Lumped models
: Hydrological model which relates the characteristics of the river hydrograph to the physiographic factors.

Mean square error function
: Performance function that calculates the average squared error between the network outputs **a** and the target outputs **t**.

Momentum
: Technique often used to make it less likely for a backpropagation network to get caught in a shallow minimum.

Node
: Basic processing element of a neural network. Includes weights and bias, a summing junction, and an output transfer function.

Neural network
: Information processing systems, inspired by biological neural systems but not limited to modelling such systems.

Output layer
: Layer whose output is passed to the world outside the network.

Output vector
: Output of a neural network. Each element of the output vector is the output of a node. Output of this study is water stage at P1.

Overfitting
: Case in which the error on the training set is driven to a very small value, but when new data is presented to the network, the error is large.

Perceptron
: Single-layer network with a hard-limit transfer function. This network is often trained with the perceptron learning rule.

Perceptron learning rule
: Learning rule for training single-layer hard-limit networks. It is guaranteed to result in a perfectly functioning network in finite time, given that the network is capable of doing so.

Performance
: Behavior of a network.

Performance function
: Commonly the mean squared error of the network outputs. However, the toolbox also considers other performance functions.

Radar
: Radio detecting and ranging.

Radar reflectivity factor
: A measure of the fraction of the energy reflected back to a radar by the targets in a volume of air.

Radial basis networks
: Neural network that can be designed directly by fitting special response elements where they will do the most good.

Regularization
: Modification of the performance function, which is normally chosen to be the sum of squares of the network errors on the training set, by adding some fraction of the squares of the network weights.

S band
: Radar wavelength of 10 cm.

Sigmoid
: Monotonic S-shaped function that maps numbers in the interval (-∞, ∞) to a finite interval such as (-1, +1) or (0, 1).

Sum-squared error
: Sum of squared differences between the network targets and actual outputs for a given input vector or set of vectors.

Tan-sigmoid transfer function
: Squashing function of the form shown below that maps the input to the interval (-1, 1).

Transfer function
: Function that maps a node's (or layer's) net output **n** to its actual output.

Update
: Make a change in weights and biases.

X band
: Radar wavelength of 3 cm.

Chapter 1 Introduction

1.1 Background

Regular flooding occurs in Thailand during the monsoon season, which results in both loss of life and infrastructural damage. During the last 10 years alone, there have been severe floods reported in most years (e.g. BBC, 2001; 2002; 2007; NineNews, 2006; UPI, 2008). There is also a great deal of uncertainty surrounding the effects of climate change. Since the middle of the last century, the occurrence of heavy precipitation events has already increased. A warmer climate may lead to increased summer precipitation in the Asian monsoon region and therefore result in increased flooding (IPCC, 2007).

Timely and accurate flood warnings are crucial in order to minimise the loss of life and to be able to put into place operational measures to minimise the flood damage. A flood warning system usually has some type of physical or conceptual hydrological model to produce flood forecasts, which are based on physical and/or empirical relationships. Neural Networks (NNs) and other data-driven methods (e.g. fuzzy logic, support vector machines, M5 model trees) offer an alternative to traditional hydrological models as they do not require any knowledge of these physical relationships. Instead these methods can learn them from the data. Neural network also have other advantages such as the ability to generalise to unseen datasets and to perform distributed parallel processing (Kasabov, 1996).

There are now hundreds of papers in the academic literature that demonstrate that neural networks or other data-driven methods can be used successfully for rainfall-runoff modelling and other hydrological applications as evidenced by two recent reviews (Abrahart *et al*., 2010; Maier *et al*., 2010). Unlike physically-based or conceptual hydrological models which require physical parameters, these models require some historical input data, usually stage, discharge or precipitation for their development (ASCE, 2000a; Wilby, 1997). Moreover, neural network models are relatively easy to develop, and to update when new data become available (Haykin, 1999; Wood and Connell, 1985). There are many examples of hybrid, data fusion or soft computing applications, where different technologies are used together to produce a better forecast than an individual model (See, 2008; Solomatine *et al*., 2008). However, the black box nature of neural network modelling and many other kinds of data-driven approaches have meant that hydrologists have been reluctant to use these technologies operationally. A rainfall-runoff neural network model integrated with

telemetered flow and rainfall data to produce real-time forecasts was developed by Kneale *et al*. (1999) but this system was never used for operational flood forecasting.

This thesis reports the results from a systematic experimental approach to evaluating the use of neural network modelling to forecast river stage for the Upper Ping catchment in northern Thailand. The flood issue for local forecasters derives from the monsoon rainfall which leads to serious out-of-bank flooding two to four times a year. As might be expected, the historic flood record is limited in both length and number of stage gauging stations. The Hydrology and Water Management Centre for the Upper Northern region (2005, 2007a, b) has responsibility for flood warning in the Upper Ping catchment. The technique currently in use is based on a correlation between the water stage at an upstream station (P67) and the downstream station (P1) at Chiang Mai. The maximum flood warning is currently 6-7 hours. There are also several hydrological applications in the Ping catchment using conceptual and physical hydrological models but most of these predict either monthly or daily discharge (Schreider *et al*., 2002; Vongtanaboon *et al*., 2008; Taesombat and Sriwongsitanon, 2006, 2010; Mapiam and Sriwongsitanon, 2009). Moreover, the disadvantage of these models is that they require a lot of data for their development, and they are not forecasting at an hourly level, which is needed for effective operational flood forecasting. Neural network forecasting models are potentially very powerful forecasters where the data are limited. One of the unique aspects of trying to develop a flood forecasting model for Chiang Mai is the limited amount of data available, in particular hourly rainfall, so the challenge is to see whether good performing models can be produced with such a limited data set. For this thesis, there are 7 years of data available from 2001 to 2007 for stage and rainfall data, and 3 years, i.e. 2003 and 2005 to 2006, for the radar data.

There have also been neural network models developed for this catchment (Patsinghasanee *et al*., 2004; Thaisawasdi *et al*., 2007; Ninprom and Chumchean, 2009; Chidthong *et al*., 2009). However, Thaisawasdi *et al*. (2007) only modelled daily flow at ungauged station. Patsinghasanee *et al*. (2004) applied neural networks to predict a storm in the period September to November 2000 for a 12 hour lead time at the station P1 but the results were that the peak was underestimated. However, a separate study by Patsinghasanee (2004) found that the performance of the neural networks was better when compared with MIKE11. Ninprom and Chumchean (2009) built neural network models using water stage at 3 hour intervals for a lead time of 72 hours. However, they did not show any hydrograph or model performance results in terms of peak prediction. The models by Chidthong *et al.* (2009) are the most relevant. They built a series of hybrid forecasting models to predict the flood at Chiang Mai in

2005 using hourly river level data but they only used daily rainfall as an input. It was possible to obtain hourly rain gauge information from one station for this study so the effect of having better information will be evaluated.

One of the problems with many of the neural network and data-driven rainfall-runoff models reported in the literature and also the ones listed above for the Ping catchment is that they have been developed for short lead times. To determine whether models with longer lead times can be developed, investigations with weather radar data have been undertaken. Weather radar data are normally used for calibrating rainfall using rain gauges (Chumchean, 2007; Collier, 1996; Hitch, 1991; Kalinga and Gan, 2007; Rachaneewan, 2006a, b) in order to predict rainfall (Cole and Moore, 2008; Georgrakakos *et al*., 2000; Hsu *et al*., 2000). Other researchers have specifically applied this to flood forecasting (Borga, 2002; Knowles, 1987; Moore, 1987, Schultz, 1987; Smith *et al*., 2007; Wardah *et al*., 2008). However, predicting floods using raw radar reflectivity as an input to a neural network has not been attempted before. There is only one hourly rain gauge right near Chiang Mai so using the radar data with this one rain gauge would not have been very useful. Instead, the method used here takes advantage of the spatial and temporal coverage of the radar images. This thesis will investigate the effect of adding raw radar data as an input to the neural network to determine whether the accuracy and the lead time of the forecast can be improved. With the excellent temporal and spatial coverage of radar data, this is potentially an important source of information in this data sparse catchment.

Another major issue addressed in this thesis is the lack of guidance in neural network model development, in particular, which input variables to use. Maier *et al*. (2010) continue to highlight the importance of this issue as an area where further research is needed. Hydrological knowledge can help inform this choice but the input variable inclusion is often determined through trial and error, simple statistical methods (Li *et al*., 2009; Shamseldin, 1997; Jia *et al*., 2009; Kim *et al*., 2009, Shevnina, 2009) or sensitivity analysis (Kalra and Ahmad, 2009; Sudheer, 2005). However, there are other methods available such as partial mutual information (Bowden *et al*., 2005b; Fernando *et al*., 2009, May *et al*., 2008a), neural network pruning (Chen and Yu, 2007; Corani and Guariso, 2005), saliency analysis (Abrahart *et al*., 2001) and the use of data-driven methods such as a self-organizing map (Bowden *et al*., 2005b; Hsu *et al*., 2002) or genetic algorithms (Abrahart *et al*., 1999). This thesis will review a range of different input determination methods available for the development of neural network hydrological models and choose a subset for subsequent experimentation.

This lack of guidance in the literature is also evident in other areas of neural network model development. For example, making choices such as the number of hidden nodes, the learning algorithms, the learning rate, the momentum, the activation function and what normalization to use are all choices that the modeller must make. These choices are either determined through trial-and-error or are based on heuristics provided by different researchers (e.g. Campolo *et al*., 1999; Dawson and Wilby, 1998, 1999; Golob *et al*., 1998; Maier and Dandy, 1996, 1998, 2000). This thesis will also address the lack of guidance in some of these areas through experimentation to determine if changes to these internal parameters will improve the neural network models developed for this catchment. These types of experiments will also add to the literature on neural network model development heuristics.

1.2 Aims and Objectives

The overall aim of this study is to determine the most effective neural network approach to forecasting stage in a large monsoon-fed river system through multiple experiments with different inputs and neural network parameterisations. There are few neural network applications to forecast stage in large monsoon-fed rivers. As Chapter 2 discusses, there are individual river studies but the rationale for the methods selected are rarely explored or explained. Rainfall data in northern Thailand comes from a very sparse rain gauge network. This thesis explores the practicalities and potential for integrating radar data to improve the forecasts.

The overall aim will be achieved through the following objectives:

1. To review the relevant literature on neural network modelling in hydrology. This review is presented in Chapter 2.
2. To review and critically evaluate existing input determination techniques, which is the subject of Chapter 4.
3. To experiment with a range of different input determination techniques for selecting input variables by training a series of neural network models and evaluating the results with various goodness of fit statistics and visual methods. These experiments are presented in Chapter 5.
4. To experiment with radar data as an input to the neural network models as a way of improving the model accuracy and extending the lead time of the forecasts. A series of experiments are undertaken in Chapter 6.

5. To investigate model improvements through experimentation with the training algorithm and internal neural network parameters. A series of experiments are described in Chapter 7.
6. To highlight the limitations of the study and to make recommendations for areas for further research. These follow the conclusions in Chapter 8.

The value of this thesis derives from a) its systematic experimental approach to evaluating all aspects of the modelling process and b) its inclusion of radar imagery in a novel way not used before in neural network modelling. The ambition is to make suggestions for future users of neural network models for forecasting stage in large monsoon-fed catchments. The outcomes cover parameterisation and training and should be relevant for similar large catchments with sparse stage and rainfall records.

1.3 Thesis Structure

Following this introductory chapter, a literature review on neural networks and hydrological modelling is presented. This chapter begins with a review of the relevant literature on hydrological modelling in general and then focuses specifically on neural networks. The main issues with neural network development are highlighted using a range of hydrological applications to illustrate these issues. Hydrological modelling in Thailand and the Upper Ping catchment in particular, both physical and data-driven, are then critically evaluated. Chapter 3 provides an overview of the study area and the data that are available for modelling. As mentioned already, this is a data sparse catchment so there are particular challenges with hydrological modelling on such a large monsoon-fed catchment. Chapter 4 then provides an in-depth review of existing input determination methods that have been used in neural network modelling. These methods are evaluated and a subset is selected for experimentation in Chapter 5. This chapter critically evaluates the accuracy of forecasts given different input variables as determined by the approaches coming out of the review in Chapter 4. The subject of Chapter 6 is evaluating the value of radar imagery for stage forecasting. This chapter focuses on the use of raw radar data as an input to the neural network models to determine the potential for increasing the forecasting accuracy and the extension of the forecast lead times. In Chapter 7, three issues with neural network development are addressed. The first is the use of a particular neural network training algorithm (Bayesian Regularization) and the need to produce many training runs to provide an acceptable averaged solution. This is an area that is rarely discussed in the literature and therefore receives treatment in this study. The second and third parts consider issues to do with the internal parameters of the network, specifically determination of

the number of hidden nodes and the range of normalization of the data prior to modelling. There is little guidance in the literature on determining these internal parameters so experimentation is undertaken to see whether any useful patterns emerge that may translate into guidance for other modellers. Chapter 8 concludes the thesis with a summary of the main research results, relating these back to the original aims and objectives of the thesis. The chapter also discusses some of the problems and limitations that were encountered, and then makes suggestions for areas of further research.

1.4 Conclusions

This chapter has introduced the background, aims and objectives and research rationale for this study. The next chapter presents a literature review of neural networks and their application to hydrology. Studies specific to the Upper Ping catchment are also critically evaluated.

Chapter 2 Neural Network Hydrological Modelling and the Ping Catchment

2.1 Introduction

This chapter establishes the academic background for the thesis through reviewing the literature on modelling hydrological flood events using neural network (NN) techniques. There is first a brief consideration of styles of modelling in hydrology to set the context for neural network modelling, which is then considered in detail with particular emphasis on neural network structures and what internal parameters to select. Neural network modelling is a relatively new technique so the third section considers some of the recent research evaluating neural networks for hydrological forecasting. These techniques, and their value for forecasting in the context of Thailand, are the focus of the research experiments presented in Chapters 5 to 7 where neural networks are applied to one large tropical river catchment, the Upper Ping River catchment in northern Thailand. Chapter 3 gives details of the catchment area, and the data available for the project. This chapter concludes with a survey of the research reported to date on neural networks and flood forecasting on the Upper Ping River.

2.2 Hydrological Modelling

The aim of hydrological modelling is to forecast flows or levels and water quality in real time for practical flood forecasting, and without a time reference for planning purposes (Anderson and Burt, 1985). A model requires one or more input variables which are processed to forecast one or multiple output variables. Applications for hydrological modelling include forecasting of rainfall, runoff, water supply and discharge, water quality, extreme flood events, snow melt, evaporation, soil water, and sediment movement and deposition.

Approaches to modelling in hydrology have been classified in a number of different ways including three system types by Wilby (1997), Wood and Connell (1985), O'Connor (2002), Anderson and Burt (1985), ASCE (2000a) and a four group system from Wheater *et al*. (1993). The four groups are metric models, conceptual models, physics-based models and hybrid metric-conceptual (HMC) models that are a mixture of statistical and conceptual models. Each system has its own advantages and drawbacks, and accuracy in forecasting is a continuing issue. The neural network model can be viewed as a black box input-output model. Data are fed into the processor, and the model learns the relationships in the data to produce the forecast. This is a process that does not necessarily appeal to modellers who prefer the core of the model to be a dynamic, physically-based, representation of the processes involved.

However, in very large catchments, where process representation would require considerable instrumentation, historical records and detailed data, a neural network offers strengths through its relatively light demands for data.

2.2.1 Conceptual or lumped models

The conceptual model (Wilby, 1997), the lumped conceptual model (Wood and Connell, 1985) and the geomorphology-based model (ASCE, 2000a), and similar models of this type are “quasi-physical in nature” (Wood and Connell, 1985). They are structured to represent the watershed and the stream network, rather than the physical equations of, for example, conservation of mass or movement. An example is the United States National Weather Service River Forecasting Model. Conceptual models are more detailed than input-output black box models, and not less demanding than physically-based models. Conceptual models aim to capture dominant catchment dynamics while remaining parsimonious and computationally efficient (Kavetski *et al*., 2006). They do not consider the physical processes in detail and retain the processing speed that empirical models exemplify. Conceptual or lumped models are generally seen as being the most successful in rainfall-runoff simulation where the parameters are limited to between 8 and 20 (Blackie and Eeles, 1985). The success of these models has made them well known, and popular forecasting tools for river flow and flooding applications. Typical models include HYRROM (Haigh, 1997) and TOPMODEL (Beven and Kirkby, 1979). One advantage of TOPMODEL is that it can deal with changing conditions on the catchment, such as land use changes (Anderson and Burt, 1985). Other successful models of this type include applications for large-scale modelling in southern Iran (Noruani and Mano, 2007), Tank modelling (Tingsanchali and Gautam, 2000), the United States National Weather Service River Forecasting Model (Wood and Connell, 1985), the degree-day snow model (Martinec, 1975) and Lee and Chang (2005) surface and subsurface flow model.

2.2.2 Physically-based or deterministic models

Physically-based models (Wilby, 1997) or deterministic models (Anderson and Burt, 1985) are based on complex physical parameters. The aim is to mirror numerically the characteristics and physical processes that take place in the catchment. Typically, these models consider the equations of energy and the movement of the water on the surface, and through the unsaturated and saturated zones to the watercourses, and build the flood hydrograph dynamically from the runoff (Wood and Connell, 1985). The catchment system is represented by a three-dimensional grid. The advantage of these types of model is in giving the operator a fuller understanding of the catchment processes. Where catchments are developing and changing, the operator can alter

some of the catchment characteristics and then evaluate the impact on runoff or water quality (Anderson and Burt, 1985). In contrast, the disadvantage is that these models require rich and complex field data from a range of sources and in a variety of timeframes. They are also computationally complex and time consuming to run. For these reasons they frequently have more value in planning than in real-time forecasting. Kuchment and Gelfan (2009) used a physically-based model for long term forecasting of spring flood flows on the Vyatka River. This example showed that the physically-based model was more accurate than black box regression approaches, but that it needs a good understanding of the flow process in the watershed to correctly parameterise it.

Probably the best known model of this type is the SHE model, where a wide range of applications have been published for flood forecasting, ground water supply, stream flow depletion, changing land use impacts on hydrology and many others (Abbott *et al*., 1986a). However, distributed models like SHE, when applied to large catchments, require many input parameters, and are therefore expensive to calibrate and can take considerable time to run (Abbott *et al*., 1986b).

2.2.3 Empirical or black box models

Empirical, black box and input-output models use historical data for flow, rainfall and other parameters to make forecasts without reference to physical parameters such as catchment characteristics (Wilby, 1997; ASCE, 2000a, Anderson and Burt, 1985; Wood and Connell, 1985). This family of models includes statistical approaches, and are generally straightforward to run once the data have been verified. Their processing speed makes them particularly useful for real time forecasting. However, the weakness of these models is their static nature. They cannot take account of dynamic changes in the catchment, such as land use changes or changing soil moisture patterns. However, Wood and Connell (1985) remarked that, while these models do not consider physical relationships, these models can all be updated when new data become available. An example application of this model is by Phien and Nda (2003) who succeeded in using this type of model to predict water level with 18 hour lead times, Rajurkar *et al*. (2004) who applied black box modelling to daily rainfall runoff and Vargas *et al*. (2003) who employed this approach for the development of a flood warning system.

Neural network modelling is considered to be in the empirical or black box model family. The Upper River Ping, which is the focus for this study, is a large catchment with very limited long-term data records. Calibrating a physical or conceptual model for a catchment of this size is inherently very difficult; many parameters would have to be

estimated rather than known quantities within such a model. There are limited data available for stage at stations on the River Ping that can be used to drive a neural network model. The city of Chang Mai in the lower part of the catchment is subject to flooding, and would benefit from effective real-time flood warning. The black box neural network model approach offers the opportunity to create models that can be run in real-time, and which can be updated relatively easily as new data become available each year.

2.3 Neural Networks (NN) Models

The human brain consists of billions of nerve cells or neurons interconnected in a large network (Murre and Sturdy, 1995). A neural network is very loosely based on this human brain architecture, using the concept of a neuron or node as a processing element and weights for connecting the neurons into a network as shown in Figure 2.1. The most common neural network arrangement is three layers: the input layer, where data are supplied to the model, a hidden layer, necessary for processing, and an output layer, where the forecast or prediction is produced. In addition there are bias nodes necessary for the calculations to work. The number of hidden layers and hidden layer nodes is determined by the operator. Some neural networks have more than one hidden layer although it has been shown that any continuous function can be approximated with this neural network structure using only one hidden layer (Hornik, 1993). Using an input-output data set, the neural network operator runs the data through a 'training process' to learn the relationships in the data, adjusting the weights in an iterative process until a stopping condition is satisfied, i.e. the error between the observed and predicted values is minimised or a certain number of iterations have passed. This training process may involve a large number of runs or iterations.

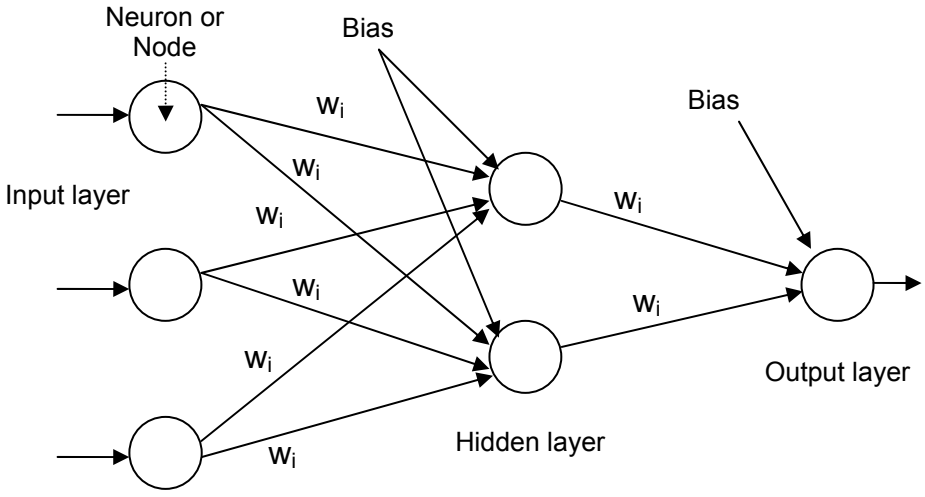


Figure 2.1: Classic neural network structure with one hidden layer.

There are two main classes of neural network architecture: feedforward and recurrent networks (Haykin, 1999). In a feedforward network, all the interconnections are in one direction, i.e. forward. The flow of information starts from the input layer of nodes, connects to the hidden layer and finally to the output layer, as illustrated in Figure 2.1, moving in a forward direction. More sophisticated recurrent networks process information in both directions. This allows model outputs to be looped back as new inputs to the model (see Figure 2.2).

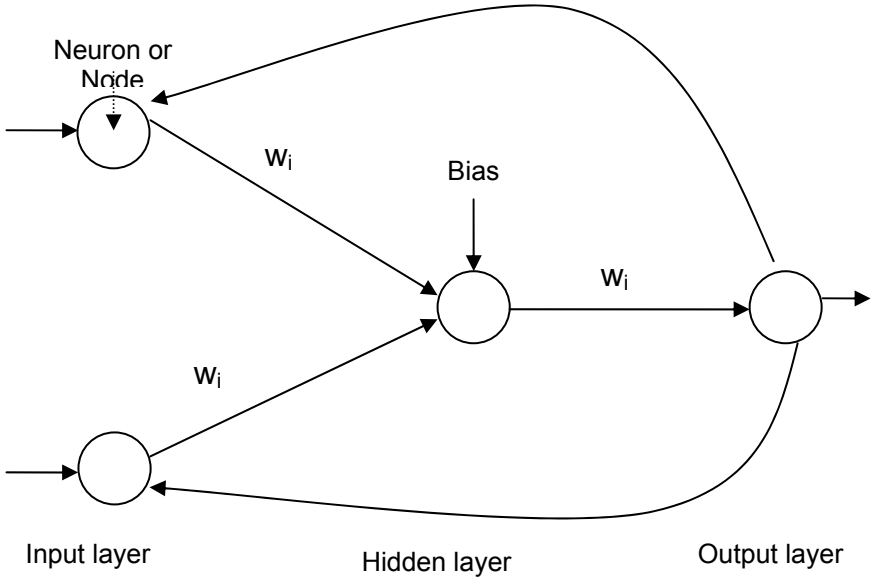


Figure 2.2: Recurrent neural network structure.

There are two types of network training procedure: supervised and unsupervised. Unsupervised training will be discussed further in section 2.4 in relation to a type of neural network called a Self Organizing Map. Supervised training is used to learn the relationships between a set of input and output variables, which in the context of this research, would be generally comprised of historical data. An example of an input data set would be river stage data at upstream points and precipitation data and river stage data at the station for which the forecasts are made. For each input, there is a corresponding output, which in this research would be the river stage at a lead time in the future.

There are a range of supervised training algorithms available, each of which has advantages and disadvantages. It is common for analysts to experiment with a number of different training algorithms to optimise their modelling results. The mostly commonly used supervised training algorithm is called backpropagation (Rumelhart *et al*., 1986). As shown in Figure 2.3, the algorithm starts with a ‘forward pass’ to produce an output prediction. The error between the output prediction and the known output is then

calculated, which is propagated backwards towards the input layer in a ‘backward pass’. During this backward pass, the weights connecting the neurons are updated. This process is repeated over many iterations until the relationships in the data are learned.

One problem which can occur during training is overfitting. To minimise this problem, early stopping is the most commonly used method (Haykin, 1999). Early stopping involves the use of a validation data set. In general, the full data set available for modelling is typically divided into three sub-sets: training, validation and testing. The network is trained on the training dataset and simultaneously examined for performance on the validation data set. The error will generally keep decreasing on the training dataset but will begin to increase on the validation data set as overfitting occurs. Training is stopped at the point where the error in the validation data set increases.

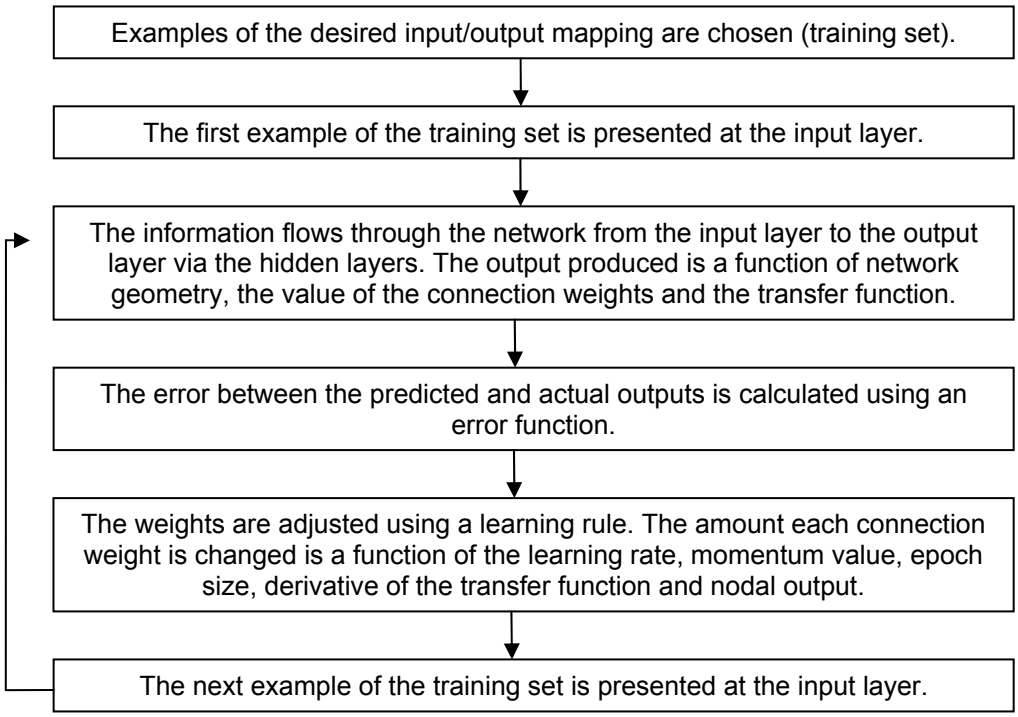


Figure 2.3: Summary of the backpropagation training process.
Source: Taken from Maier and Dandy (1998, p.196)

There are many types of backpropagation algorithm, which can be classified into two main groups. The first group is described as slow learning and uses batch gradient descent or simple gradient descent with momentum. The second, fast learning, group may employ conjugate gradient descent (CGD), quasi-Newtonian algorithms (which require more storage and computation for each iteration) and the Levenberg-Marquardt numerical optimization technique. The fast learning techniques can perform at 10 to 100 times the speed of the slow learning processes. The difference between CGD and simple gradient descent is that it does not proceed along the direction of the error

gradient but in a direction orthogonal to the one in the previous time step. The advantage for the modeller is a relatively fast training time. It has been used in hydrology for short and longer term stream flow forecasts (Kisi, 2007). Where the model has a very large number of hidden layer nodes, this process can be very efficient. However, where the input nodes and weights are small, there does not appear to be any advantage in using CGD over first order backpropagation methods. Levenberg-Marquardt is probably the most commonly employed algorithm (Demuth *et al*., 2009). Fun and Hagan (1996) compared the performance of three different algorithms including Levenberg-Marquardt and found that it produced the best performing result.

Bayesian regularization (BR) is another type of training procedure that can be applied to Levenberg-Marquardt (Mackay, 1992; Forsee and Hagan, 1997) and essentially reduces the impact of the network weights. It simultaneously minimises the overall error function and the sum of the squared weights. A big advantage of the BR algorithm is that it does not require a validation data set, unlike in early stopping and therefore makes BR a potentially suitable algorithm where data are limited (Demuth *et al*., 2009). This means that more data are available for the training process. The BR algorithm has been used in a few recent hydrological modelling applications (Anctil and Lauzon, 2004; Anctil *et al*., 2004a, b; Anctil *et al*., 2006; Zhang and Govindaraju, 2000).

Another algorithm is Cascade Correlation (CC), which was developed by Fahlman and Lebiere (1990). Unlike the networks described above, this network starts with the input and output nodes only. As training progresses, hidden layer nodes are added so that the network grows during training. This potentially takes away some of the subjectivity regarding how many hidden nodes to use. Each time a hidden node is added into the network, the model is trained to minimise the output error. The advantages of this training process are fast training times, robust results and no need to specify the number of hidden nodes. However, experience suggests that while these models are good at fitting in training exercises, they are less good when faced with unseen data in forecasting mode (Dawson and Wilby, 2001; Kisi, 2007).

In Chapter 5, backpropagation trained with Levenberg-Marquardt is used in the initial neural network experiments because it has been used successfully by previous hydrologists, suggesting that comparisons in quality of forecast may be possible. Experimentation with Bayesian Regularization is then evaluated because it does not require a validation dataset to avoid overfitting. This attribute is very attractive because there are limited data available for training and testing the neural network models for

the Upper Ping catchment.

2.4 Types of Neural Network

The most commonly used type of neural network is a Multi-layer Perceptron (MLP), where the structure was provided previously in Figure 2.1. Many researchers have applied MLPs to hydrological applications and generally found good or superior performance when the results were compared with other approaches. For example, Dawson *et al*. (2006a) employed an MLP to estimate flood frequencies at ungauaged catchments. Smith and Eli (1995) employed an MLP to predict peak discharge. Golob *et al*. (1998) used an MLP to predict natural water inflow 2, 4 and 6 hours ahead for the Soca River basin. Raman and Sunikumar (1995) applied an MLP and an autoregressive moving average model (ARMA) to predict monthly inflows to two study reservoirs. Lorrai and Sechi (1995) used an MLP to model rainfall-runoff using rainfall data and temperature for 30 years in the Araxisi catchment in Sardinia while Campolo *et al.* (1999) developed an MLP for the River Tagliamento in Italy. Dawson and Wilby (1998) used 15 min rainfall-runoff data to model runoff for a 6 hour lead time on the Rivers Amber and Mole in the UK. There are many other examples in the literature, many of which have been reviewed by Maier *et al*. (2010). Out of the 210 papers published between 1999 and 2007, Maier *et al*. (2010) found that 178 papers employed MLPs.

Another type of neural network that was used along with MLPs but less frequently is a Radial Basis Function (RBF) network. They are similar in architecture to MLPs. The difference is that the hidden layer layers use radial basis functions (of a Gaussian form) in the neurons while the output neuron activation function is linear. There is a more detailed explanation given on activation functions in section 2.5. Training usually involves determining the parameters of the radial basis functions first followed by determination of the weights. It has been noted that RBF models can be particularly useful where there are large training datasets involved (Achela *et al.*, 2009; Shamseldin, *et al.*, 2007). Dawson and Wilby (1999; 2000) added RBF networks and linear regression to their modelling of the Rivers Amber and Mole. However, they concluded that the MLP was more accurate for flow forecasting than the other techniques employed. Sahoo and Ray (2006) employed both MLPs and RBFs, with similar findings to Dawson and Wilby (1999; 2000). According to Maier *et al*. (2010), RBF networks were not employed that frequently, i.e. less than 20 out of 210 papers over the period 1999 to 2007.

Another type of neural network that has been used in hydrological modelling but less frequently is a recurrent neural network, as illustrated previously in Figure 2.2. The difference between recurrent neural networks and MLPs is that there are feedbacks allowed between the layers in the network and different learning algorithms are required. Chang *et al*. (2002) applied recurrent neural networks to rainfall-runoff modelling in Taiwan. They compared the results with ARMA models and found very good performance for the recurrent neural networks. In a further study, Chang *et al*. (2004) extended the model to forecast two steps ahead and once again found very good performance when compared to an ARMA model. Another example is a study by Carcano *et al*. (2008) where a conceptual model and MLP were compared to a recurrent neural network to forecast daily streamflow in two small catchments. The results showed that when good input data were unavailable, then the data-driven methods outperformed the conceptual model with the recurrent network performing very well. In the review by Maier *et al*. (2010), recurrent neural networks were used less than 20 times over the period 1999 and 2007.

Self Organizing Maps (SOMs) were first introduced by Kohonen (1984). As mentioned in section 2.3, SOMs are generally trained using unsupervised training where only an input dataset is necessary. This algorithm takes the input data and classifies it into groups with similar properties (ASCE, 2000a). This algorithm, can however, be used in forecasting or predictive mode. It efficiently deals with very large gridded datasets by reducing dimensional space through classification of data into groups. This method is very useful where the data are map-based and it is important to preserve the topology of the data (Haykin, 1999). It is an interesting approach but it is generally used for data pre-processing. For example, See and Openshaw (1999) used a SOM to classify river level data into events, e.g. low river levels, the rising limb of the hydrograph, the falling limb, etc. and they then trained individual MLPs to predict these events separately. Hsu *et al*. (2002) have developed the SOLO (Self-Organizing Linear Output) map which simultaneously classifies the input data and makes a prediction. The authors developed this model to predict daily streamflow for the Leaf River basin in Mississippi. The results were compared to ARMA, MLP, recurrent neural network and conceptual models and the SOLO model outperformed them all. Moradkhani *et al*. (2004) modified the SOLO architecture to add a radial basis function to the approach and showed that performance was further improved over the SOLO model, MLP and linear regression for daily streamflow forecasting. Chapter 6 of this thesis deals with inputs from radar data. It would be possible to use a SOM to classify the radar data by storm type, but as there was so little data available, this approach has not been used. The review by Maier *et al*. (2010) places SOMs into a category called stratified, unsupervised

methods for pre-processing of the data prior to use by another type of neural network. They found that SOMs were only used on about 10 occasions for this purpose during the period 1999 to 2007.

In summary it should be noted that the MLP is the most commonly used type of neural network for hydrological modelling (Maier *et al*., 2010). There are several studies that have also used ARMA models (e.g. see Raman and Sunikumar, 1995; Abrahart and See, 2000; Chang *et al*. 2002; 2004, Wang *et al*., 2009). However, Tang *et al*. (1991) found that ARMA models were better suited for short term forecasting, while neural networks are more effective for long term forecasting. Therefore MLPs will be used in this research study.

2.5 Internal Parameters

There are a series of decisions that must be made about the neural network including the internal parameters that must be set. These decisions include: which input variables to choose; how many epochs or training cycles to use; which error function to use; initialisation of the weights; choice of parameters associated with the training algorithm (e.g. the learning rate and the momentum); which transfer or activation function to choose; how many hidden nodes and hidden layers to choose; and the range of normalisation of the data. A disadvantage of neural network modelling is how to determine the most effective combination of these internal parameters (Maier and Dandy, 1996).

All input-output models depend on the quality of the data. Inappropriate selection of input variables will affect the quality of the output forecasts (ASCE, 2000a). If there are long term trends in any dataset, or the data record is not long enough to encapsulate the range of data to be expected, then model performances can be impaired. As the number of inputs in the neural network is increased, the size of the network increases, i.e. the number of nodes and weights expands and therefore the training time also increases. This can be very expensive computationally, possibly limiting opportunities for real time forecasting, which is relevant for the study catchment. This has led a number of hydrology researchers to focus on methods for determining the inputs to the neural network, with the aim of reducing training times without compromising the forecast quality (Dawson and Wilby, 2001; Bowden *et al*., 2005a, b). As this is a very important topic and one that even Maier *et al*. (2010) still recognise as needing further research, input determination techniques are a key area of experimentation in this thesis. Methods of input determination are reviewed separately in Chapter 4.

Moving onto the other decisions that need to be made regarding a neural network, there have been some studies that have focussed on internal parameter determination. In terms of the number of epochs or training cycles, Maier and Dandy (1998) found that there was no difference in using a larger or small epoch size for reaching the global minimum. They found that increasing the epoch size did not significantly affect model performance. However, a validation data set will be used for 'stop training' as is recommended in many neural network textbooks (e.g. Haykin, 1999). The research will also use Bayesian Regularization where no validation data set is required and the algorithm will converge on its own.

The error function most commonly used is the sum of the error squared. There have been some studies looking at changing the objective function of the neural network (e.g. Abrahart *et al*., 2007; De Vos and Rientjes, 2007). However, this research will continue to use the most commonly used error function.

For the initialisation of the network weights, Maier and Dandy (1998) used random weights between -0.1 and 0.1. They found no effect on the performance or generalization ability of the network. For this reason, a similar initialisation of weights has been adopted in this research.

In terms of learning rate and momentum, different values have been used in the literature. Maier and Dandy (1998) note that the momentum must be less than 1.0 to achieve convergence and that there were little differences in overall performance if different momentum values were used. They did note, however, that increasing the momentum speeds up the learning time. In an earlier paper, Maier and Dandy (1996) used learning rates of 0.02, 0.05 and 0.1 and found that 0.02 gave the best model performance. Dawson and Wilby (1998) used a learning rate of 0.1 and suggested that if the training rate is too small then training could be trapped in a local minimum. However, if too large, training can move back and forth between one non-optimal set of weights to another. Raman and Sunikumar (1995) and Lorrai and Sechi (1995) set the learning rate at 0.5 and the momentum at 0.9 while Danh *et al*. (1999) set the learning rate and momentum at 0.5. Advice from a neural network software producer suggests that the learning rate should be close to zero for backpropagation (Neural Ware Inc, 1991). Dawson *et al*. (2000) and Dawson and Wilby (1999) set the learning rate to vary between 0.1 to 0.01 and the momentum at 0.9. For this reason the learning rate was set to 0.01 and the momentum was set to 0.9. This follows the advice in the literature but these are also the recommended default settings provided by the Matlab software.

An activation function is a nonlinear transformation for converting the inputs to a given neuron into an output signal (Dawson and Wilby, 1999). There are many activation functions used in neural network forecasting, including logistic sigmoidal, linear, bipolar threshold, binary threshold, stochastic, Gaussian and hyperbolic tangent. The choice to a certain extent is based on trial and error. The following section reviews some of the options adopted in hydrological contexts. The most common activation function is sigmoidal although others include the logistic and hyperbolic tangent function (Maier and Dandy, 2000). There are a number of applications that have used the logistic sigmoidal activation function such as Lorrai and Sechi (1995), Campolo *et al.* (1999), Dawson and Wilby (1998; 1999), and Danh *et al*. (1999), while Crespo and Mora (1993) used the hyperbolic tangent function. Maier and Dandy (1998) used an MLP to forecast salinity 14 days ahead for the Murray River. They compared linear, logistic sigmoidal and hyperbolic tangent transfer functions. Training was fastest with the hyperbolic tangent function and slowest with the logistic sigmoidal. In terms of performance measures, the RMSE was best for the hyperbolic transfer function and worst for the linear. Therefore, they suggested that it was best to use a non-linear transfer function. Varoonchotikul (2003) investigated four different activation functions: linear, hyperbolic tangent, bipolar binary and logistic sigmoidal in four catchments. He found that both the linear and hyperbolic tangent functions outperformed the other functions. For this reason, the hyperpolic tangent function was used in the hidden layer and the linear transfer function was used in the output layer.

The next two sections discuss the number of hidden nodes and hidden layers to choose and which normalisation scheme to adopt. These are two areas that have been explored in further detail in Chapter 7.

2.5.1 Number of hidden nodes and hidden layers

The number of nodes in the hidden layer and the number of hidden layers is generally determined by trial-and-error procedures (ASCE, 2000a). If there are too many nodes in the hidden layer, then the model will overfit the training data and be unable to generalise to data it has not seen before. Moreover, the model will not be very parsimonious. However, where there are too few nodes, the neural network may be unable to model the underlying function sufficiently (Dawson and Wilby, 1999). There have been a series of heuristics suggested in the literature regarding the choice of the number of hidden nodes. For example, Smith and Eli (1995) adapted the advice of Widrow (1989), who suggested that the training size should be 10 times the number of nodes in the input layer. Other heuristics include: the optimum number of hidden layer nodes should be smaller than the number of inputs (Maren *et al*., 1990 cited in Maier

and Dandy, 1998); the number of hidden nodes should be 75% of the number of input nodes (Lenard *et al*., 1995; Jain and Nag, 1995; Walczak and Cerpa, 1999); the number of hidden nodes should be twice the number of input nodes plus 1 (Caudill, 1991; Fletcher and Goss, 1993; Patuwo *et al*., 1993; Walczak and Cerpa, 1999); and the number of nodes in the hidden layer should be half the number of nodes in the input layer (Minns and Hall, 1996; Walczak and Cerpa, 1999). An example of the latter heuristic was employed by Kerh and Lee (2006) who used 15 hidden nodes in a neural network with 30 inputs. They showed that increasing the number of nodes in the hidden layer from 10 to 15 improved the network performance. Danh *et al*. (1999) predicted daily river flow using neural networks in two different sized basins in Vietnam. Three models were used: case 1 (2-2-1), case 2 (3-2-1) and case 3 (7-4-1) where the numbers indicate the number of inputs, the number of neurons in the hidden layer and 1 output node. The authors found the best model performance was for case 3 for both basins. This shows that using a ratio of 2:1 for inputs to hidden nodes produced the best result, in line with the findings of Minns and Halls (1996).

There are, however, other examples of research in the literature on the impact of the number of hidden nodes, generally through experimentation. For example, Campolo *et al.* (1999) established a ratio of 8:1 between the input and hidden nodes through trial and error. Golob *et al*. (1998) found that using a large number of hidden nodes, i.e. 40, produced better results than a smaller number, i.e. 7, even though inaccuracies at high flow prediction were still evident. Dawson and Wilby (1999) investigated model performance using 5, 10, 20 and 30 hidden nodes for a neural network with 7 input nodes. They found that 20 hidden nodes was the best choice to predict river flow for the River Mole. Maier and Dandy (1996) also tried out a series of combinations and found the best performance was achieved using 45 nodes in the first hidden layer. Maier and Dandy (1998) investigated the ratio of the number of nodes between the first and second hidden layer using a range of combinations. Although the results did not point to a clear winner in terms of performance, 3:1 appeared to be the best choice. In addition, when using only one hidden layer, they found that using 30 nodes gave the best result. Lorrai and Sechi (1995) used various combinations of hidden nodes and numbers of hidden layers. However, their results were inconclusive. They did find, however, that increasing the length of the training period was more important in improving the model performance than adjusting these hidden node and layer parameters.

This review clearly indicates that there is no single heuristic that can be applied to determine the number of hidden nodes or number of hidden layers. Many of these may

only be applicable to the catchments in which they were tried. However, as it has been proven that a single hidden layer can approximate any continuous function (Hornik, 1993), only one hidden layer will be used in this research. This also simplifies the network structure and decreases the training time. The impact of the number of hidden layer nodes, however, will be explored for the Upper Ping River in Chapter 7.

2.5.2 Normalisation of the input data

Normalisation or standardisation transforms the input data to a common range, where [0, 1] and [-1, 1] are the most frequently employed ranges. For example, Allen and Le Marshall (1994), Dawson and Wilby (1998), Hsu *et al*. (1997) and Raman and Sunilkumar (1995) all used [0, 1] while Crespo and Mora (1993) and Sahoo and Ray (2006) used [-1, 1]. Other modellers such as Danh *et al*. (1999) and Sureerattanan and Phien (1997) have used [0.05, 0.95], Braddock *et al*. (1998) employed [-0.9, 0.9], Shamseldin (1997) used [0.1, 0.85], and [0.1, 0.9] was used by Campolo *et al*. (1999), Dawson *et al*. (2006b), Dawson and Wilby (1999), Hsu *et al*. (1995) and Smith and Eli (1995). Normalisation has also been shown to improve the model performance when the testing data exceeds the range of the training data (Shrestha *et al*., 2005) and where the training data was set in the range [0.8 X_{min}, 1.2 X_{max}]. An even tighter range of [0.3, 0.7] has been employed by Varoonchotikul (2003). In this research a broad range of [-1, 1], the most popular range of [0.1, 0.9] and the tightest range of [0.3, 0.7] will be employed.

2.6 Use of Neural Networks in Hydrology and Water Resource Mangement

The use of neural networks can be found throughout different areas of hydrology and water resource management. The ASCE (2000b) review into the use of neural networks in hydrology considered applications across a range of areas including rainfall-runoff modelling, modelling streamflows, water quality modelling, ground water applications, estimation of precipitation and miscellaneous areas such as reservoir operation and flood wave propagation. A more recent review by Abrahart *et al*. (2010) also considered applications across many water resource domains but this review was organised by methodological development rather than area. The review by Maier *et al*. (2010) looked at the steps that have been taken in developing neural networks in different water resource applications and then attempted to build a taxonomy so that trends in modelling can be understood and to highlight where further areas require more attention.

Table 2.1 shows the results of searching the Thomson Reuters ISI database using keywords that included the area and the expression 'neural network'. The table clearly shows that there is a considerable body of literature in this area, particularly within international peer-reviewed journals. The year in which the maximum number of papers was published is also quite recent showing the gathering momentum. However, additional, more refined searches would probably yield a greater number of papers than listed in Table 2.1, and adding other data-driven methods to the search would increase the number of papers considerably.

Table 2.1: The number of papers across different areas of water resource management that utilise neural networks.

Area	Conference Papers	Journal Papers	Review Papers	Total	Year of Max Number of Papers
Groundwater	54	147	3	204	2009
Irrigation	38	67	0	105	Tied for 2007, 2008, 2009
Rainfall-runoff modelling	54	197	2	253	2008
Rainfall estimation	42	115	0	157	2009
Water distribution	101	281	9	391	2009
Water quality	213	392	16	621	2006
Wastewater / Water treatment	64	185	4	253	2009

Source: Thompson Reuters ISI database (2010).

Several examples have already been provided in this literature review in previous sections that demonstrated the use of neural networks in rainfall-runoff modelling and flood forecasting (e.g. Campolo *et al*., 1999; Dawson *et al*., 2006a: Dawson *et al*., 2006b; Hsu *et al*., 2002; Kerh and Lee, 2006; Minns and Hall, 1996; Sahoo and Ray, 2006, etc). Almost all studies to date have shown that neural networks provide good model performance or superior performance over existing models.

Another aspect of the studies in this area has focussed on finding ways of improving neural networks for rainfall-runoff and flood forecasting, either through the data used in the model or issues related to the neural network itself. For example, Anctil *et al.* (2006) used mean daily areal rainfall for improving one day ahead rainfall-runoff forecasting. Their experiment first involved randomly selecting the rain gauge samples and then applying a genetic algorithm for selecting the rain gauges that improved the forecasting performance. The result showed that twelve rain gauges were chosen out of a total of 23. They also remarked that this method would be useful for decreasing the size of the total rain gauge network. Sahoo *et al.* (2006) used an MLP for stream flow prediction. To improve the forecast, they concluded that air temperature, water

temperature, stream flow, stream stage, ocean tide height, evapotranspiration and rainfall are also required inputs. Karunanithi *et al*. (1994) applied a different type of training algorithm, i.e. cascade correlation, to select a suitable network architecture for forecasting flow for the River Huron. Cannas *et al*. (2006) experimented with different data pre-processing algorithms including wavelet analysis and SOMs prior to training with a neural network for the Tirso basin in Sardinia. They found that the best results were obtained with pre-processing via the SOM. Other researchers have applied hybrid techniques such as See and Openshaw (1999), who used a SOM to classify historical input data into five event types (low flat, medium, rising, peaks and falling). They then trained all event types with individual MLPs. The results showed much better performance when compared with a fuzzy logic model, an ARMA model and naive prediction. There are many other examples of hybrid modelling applied to rainfall-runoff and flood forecasting, e.g. neuro-fuzzy and neuro-evolutionary approaches (Dawson *et al*., 2006b; Mukerji *et al*., 2009; Pramanik and Panda, 2009; Remesan *et al*., 2009; Firat and Turan, 2010). According to the review by Maier *et al*. (2010), more than 80 papers used some type of hybrid modelling in the papers published between 1999 and 2007. This is clearly an area of increasing interest. However, before hybrid modelling techniques are used, it is necessary to establish whether simple neural networks will produce viable forecasting models for the catchment under study in this thesis.

2.7 Hydrological Forecasting in the Ping Catchment

There are also several hydrological applications in the Ping catchment using conceptual and physical hydrological models but most of these predict either monthly or daily discharge. For example, Schreider *et al*. (2002) used a physical model to predict monthly discharge while Vongtanaboon *et al*. (2008) used a data-based mechanistic approach for daily discharge forecasting. Another example of daily forecasting is the work by Taesombat and Sriwongsitanon (2006), who used MIKE11-HD and NWS-FLDWAV. Mapiam and Sriwongsitanon (2009) applied both an URBS model and the Nedbor-Afstromings model for flood estimation but both models did not perform well in hydrograph prediction, most likely due to the inaccuracy of the daily rainfall data. Most recently, Taesombat and Sriwongsitanon (2010) investigated two models: IHACRES (rainfall-runoff model) and FLDWAV (a hydrodynamic model). However, the big disadvantage of these models is that they require a lot of data for their development, and they are generally not forecasting at an hourly level, which is needed for flood forecasting.

The technique for flood forecasting used at present by the Hydrology and Water

Management Centre for the Upper Northern Region (2007b) on the Ping catchment is a derived relationship between the upper stage stations and the main station. The average travel time is 6-7 hours as shown in Figure 2.4.

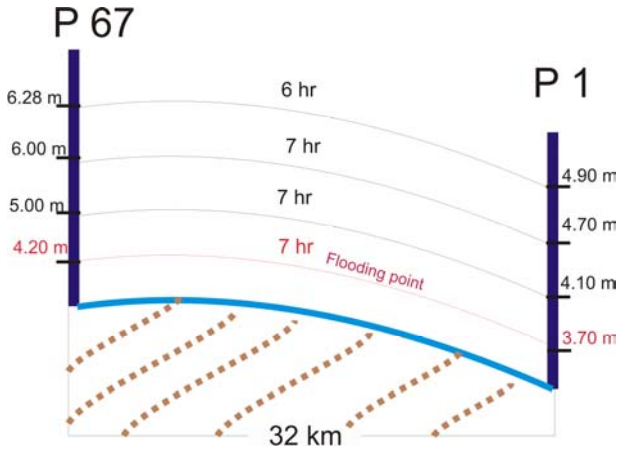


Figure 2.4: The relationship between water level and travel time between the P67 (upstream) and P1 (Chiang Mai) stations.
Adapted from: Hydrology and Water Management Centre for Upper Northern Region (2007b).

In addition to this approach, the CENDRU (Civil Engineering Department Chiang Mai University Natural Disasters Research Unit) uses a support vector machine and hydrodynamic model (called Infoworks RS) to predict stage 7 hrs ahead at Chiang Mai (P1 station) (Natural Disasters Research Unit, 2007b). The results are very good for this lead time.

To date the use of neural network modelling using data from Thai rivers has been limited, but the data are available to run such models (Tingsanchali and Gautam, 2000; Patsinghasanee *et al*., 2004; Sukka, 2005; Thaisawasdi *et al*., 2007; Ninprom and Chumchean, 2009; Chidthong *et al.*, 2009). In a catchment which is both large and lacking data for optimising TOPMODEL or SHE style modelling, data-driven approaches such as a neural network are theoretically promising. An early study by Tingsanchali and Gautam (2000) involved the development of a neural network model to predict floods 1 day ahead using the average daily rainfall of 10 stations, evaporation based on 4 meteorological stations and runoff data as the inputs. The model underestimated the peak which they claimed was due to the rainfall data not being representative of the actual values across the catchment. Sukka (2005) used a neural network trained with backpropagation to predict daily water inflows to a reservoir one and two days ahead using daily precipitation and discharge. He concluded that a 7-4-1 neural network should be applied for one day forecasting, and a 4-3-1 structure for two day ahead predictions. However, the results were poor. The maximum recorded discharge is 5.94 (10^6 m^3), whereas the best model predicted 2.47, i.e. an

underestimation of 3.47. The models by Thaisawasdi *et al*. (2007) were for flow forecasting one day ahead and Patsinghasanee *et al*. (2004) developed neural networks for a 12 hour lead time but the flood peaks were underestimated. However, a separate study by Patsinghasanee (2004) showed that the performance of the neural networks was better when compared with MIKE11. Ninprom and Chumchean (2009) built neural network models for a lead time of 72 hours. However, they did not show any graphical results or provide any information about the performance of the model in terms of peak prediction. The most relevant piece of research is the study by Chidtong *et al*. (2009), who built a series of hybrid forecasting models (i.e. neuro-fuzzy models optimised with a genetic algorithm) to predict the flood at Chiang Mai in 2005 using hourly river level data. The models were also applied to a large flood in Koriyama in Japan. However, for Chiang Mai, the authors used only daily rainfall as an input as hourly was not available. The results showed that the hybrid model outperformed other models to which it was compared (i.e. neuro-genetic and an ANFIS NN model), and that the hybrid system could produce a good forecast with a lead time of 12 hours.

Although there have clearly been some applications of both physical/conceptual and neural network forecasting models to the Upper Ping catchment, they either forecast daily data, predict for short lead times or do not use hourly precipitation or radar data. This research attempts to determine what combination of inputs and what parameters are important in the development of neural network models that are most likely to generate forecasts that would be acceptable for real-time flood warning with a sufficient lead time.

2.8 Conclusions

This chapter has reviewed the key ideas and concepts in neural network modelling for hydrological forecasting. One of the strengths of neural networks is that they can determine the relationships between the input and output variables without any explicit physical consideration of the process. This can reduce subjective errors introduced by operator-determined parameters in other forms of modelling. They can also work well despite a certain amount of noise and errors in the training data set. Based on this review, it is clear that neural networks have the potential to develop forecasting models of river stage for the Upper Ping River catchment where the data are limited and the monsoon rainfall generates high flows. Neural network forecasting is relatively new in hydrology. The research so far has shown that there have been many applications of neural network modelling that have proved effective in different catchments. Clearly there is no one approach suggested in the literature which is appropriate for flood

forecasting on a large monsoon influenced river system. The focus of this research is therefore essentially experimental. The data available for the Upper Ping River will be subjected systematically to a series of modelling experiments to determine if viable neural network models can be produced. The next chapter outlines the study area, i.e. the Upper Ping catchment, and the data that are available for modelling.

Chapter 3 Study Area and Data

3.1 Introduction

From the previous chapter a literature review on hydrological modelling, neural networks (NNs) and the use of NNs in the Upper Ping catchment was provided. This chapter presents background information on the study catchment and the data used in this research. The catchment chosen is the Upper Ping River basin, located in northern Thailand (Figure 3.1). The Ping River is the main river in this catchment, which flows from North to South and passes through Chiang Mai city. Stage data have been obtained for a number of stations, as well as rain gauge data but few stations record hourly data. Radar images are also available. The Upper Ping is a large catchment subject to monsoon rainfall, causing out-of-bank flooding events in most years. The details of the big flood in 2005 are described. Moreover, the method used to calculate missing data in the records in the stage and rainfall records is detailed.

3.2 The Upper Ping River Basin

Monsoon rainfall in Thailand comes from northeast weather systems (November to February), which brings moisture from the South China Sea, and from the southwest monsoon (May to September), which brings rain from the direction of the Indian Ocean (Boochabun *et al*., 2004). The southwest monsoon brings the larger events causing the major flooding in the Ping basin.

The Ping catchment is located in the Northern part of Thailand and covers 5 provinces: Chiang Mai, Lamphun, Kamphaengphet, Tak and Nakhonsawan. The annual average rainfall varies from 900 to 1,900 mm, with an annual average rainfall for the entire catchment of 1,125 mm (Rodratana and Piamsa-nga, 2008). More than half of the total area has an elevation range from 500 to 1500 m (Sharma *et al*., 2007). The Ping River is the main river in this catchment with a length of 740 km (Mapiam and Sriwongsitanon, 2009). The Ping catchment is divided into two parts: the Upper and the Lower Ping. The entire Ping catchment covers approximately 33,898 km^2 as shown in Figure 3.1. This catchment is mainly covered by forest at 46.5%, 31.2% by agriculture and 12.6% by paddy fields.

The geology of Thailand is complex, due to its location at the junction of a number of active plate margins. The detailed geology has still to be mapped, but in general the underlying rocks in the Ping region are heavily faulted; sedimentary rock, igneous rock, metamorphic rock, semi-consolidated deposits and folded Precambrian, Palaeozoic, Mesozoic and Cainozoic with overlying Quaternary sediments. The rocks of the

Precambrian age such as gneiss, schist and calc silicate rock lay at the west of the Chiang Mai basin (Wongprayoun and Sangsrijang, 2005).

Land use includes pasture and rice paddy agriculture, bamboo forest, dry dipterocarp forest, dry evergreen forest, hill evergreen forest, mixed deciduous forest, pine forest, grassland, savannah, secondary regrowth forest, settlement urban areas, water bodies, old clearings and plantations of eucalyptus, pine and teak (The World Bank, 2006).

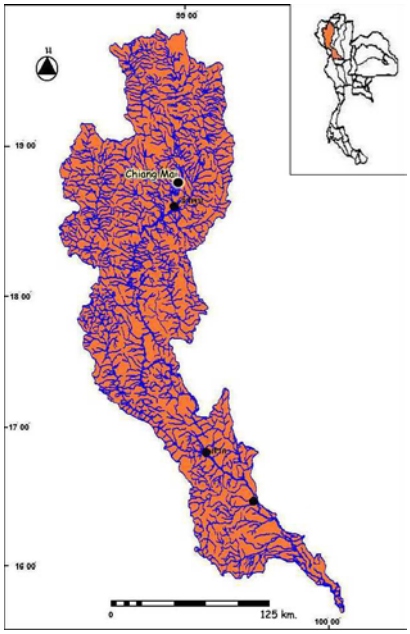


Figure 3.1: The Ping catchment.
Source: Department of Water Resources (2007)

The Upper Ping is a large complex river basin covering two provinces (17° 14′ 30″ – 19° 47′ 52″ N, 98° 4′ 30″ – 99° 22′ 30″ E); Chiang Mai and Lam Phun (Mapiam and Sriwongsitanon, 2009). It has an area of approximately 23,600 km^2 (Figure 3.2) with 15 sub-catchments. The distance from the source of the river to Chiang Mai city is 190 km (Hydrology and Water Management Centre for Upper Northern Region, 2007b). The areas of each sub-catchment are shown in Table 3.1. It can be seen that Nam Mae Tun is the largest sub-catchment in the lower part of the catchment, followed by the Third Part of Mae Nam Ping. The smallest sub-catchments are Nam Mae Hat and Nam Mae Lao.

Table 3.1: 15 sub-catchment areas of The Upper Ping basin.

Name of sub-catchment	Area (km^2)
Upper Part of Nam Mae Chaem	1,912
Upper Part of Mae Nam Ping	2,018
Third Part of Mae Nam Ping	3,071
Second Part of Mae Nam Ping	1,624
Nam Mae Tun	3,143
Nam Mae Rim	584
Nam Mae Ngat	1,260
Nam Mae Ngan	1,711
Nam Mae Li	1,956
Nam Mae Lao	535
Nam Mae Kuang	1,165
Nam Mae Klang	600
Nam Mae Hat	535
Mae Nam Mae Taeng	1,761
Lower Part of Nam Mae Chaem	1,926

Source: Hydrology and Water Management Centre for Upper Northern Region (2007a)

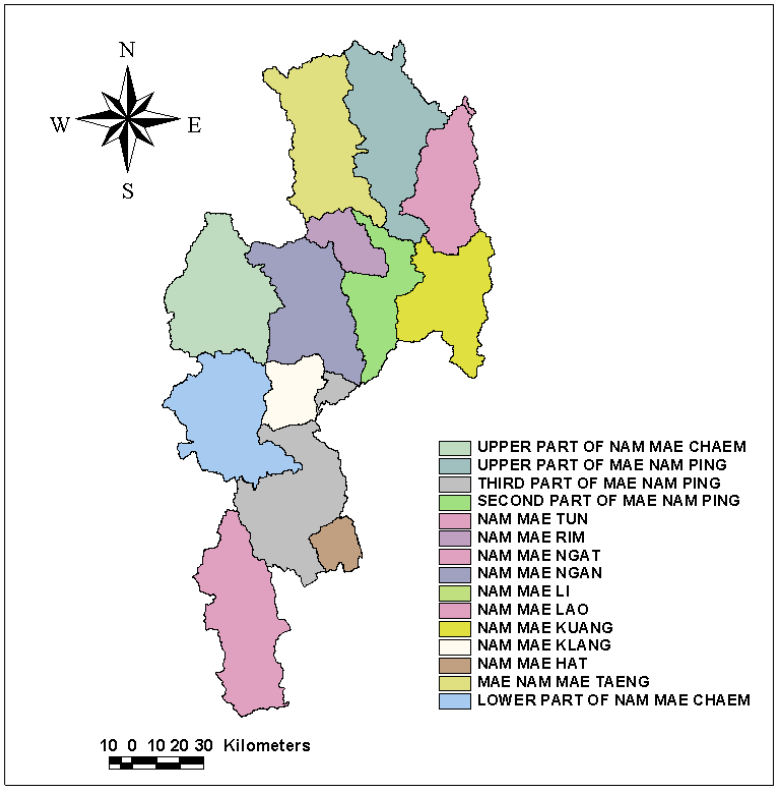


Figure 3.2: The 15 sub-catchments of the Upper Ping basin.
Source: Regional Centre of Geo-Informatics and Space Technology (2006)

This study focuses on forecasting river stage at a station called P1 at Chiang Mai in the Second Part of the Mae Nam Ping sub catchment (shown in Figure 3.3). Stage data are available for P1 and three upstream water level stations; P67 located in the Second Part of Mae Nam Ping, P4a in the Mae Nam Mae Taeng and P75 in the Upper Part of Mae Nam Ping. Station P1 is located in the centre of Chiang Mai city. Chiang Mai is in

a catchment with a relatively steep terrain (approximately 80%) with elevation higher than 500 m above sea level (Rodratana and Piamsa-nga, 2008). The elevation of the Ping River basin ranges from 380 m to 2,275 m above sea level.

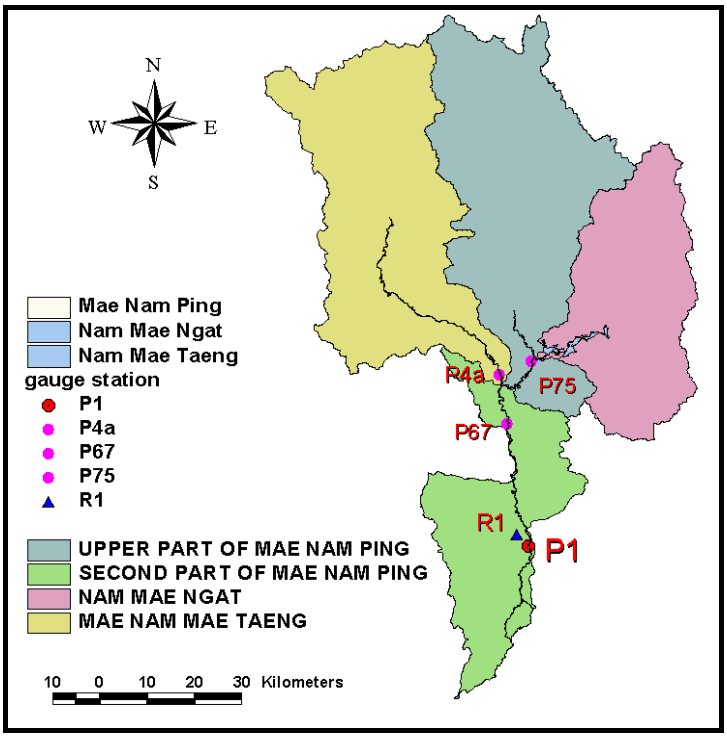


Figure 3.3: Study area sub catchments.
Source: Regional Centre of Geo-Informatics and Space Technology (2006)

In the Upper Ping catchment the average regional temperature is approximately 25.4°C, the maximum is 41.4°C in May and the minimum is 3.7°C in January (Natural Disaster Research Unit, 2007c). Figure 3.4 presents the 30 year average rainfall by month in Thailand and clearly shows that the monsoon season is from May to October. The wettest month is August, which has an average rainfall of approximately 224.4 mm. The driest month is January with 7.7 mm. Moreover, the average annual rainfall is 1,180 mm and the range of average annual rainfall is 900 to 1,900 mm (Hydrology and Water Management Centre for Upper Northern Region, 2007a).

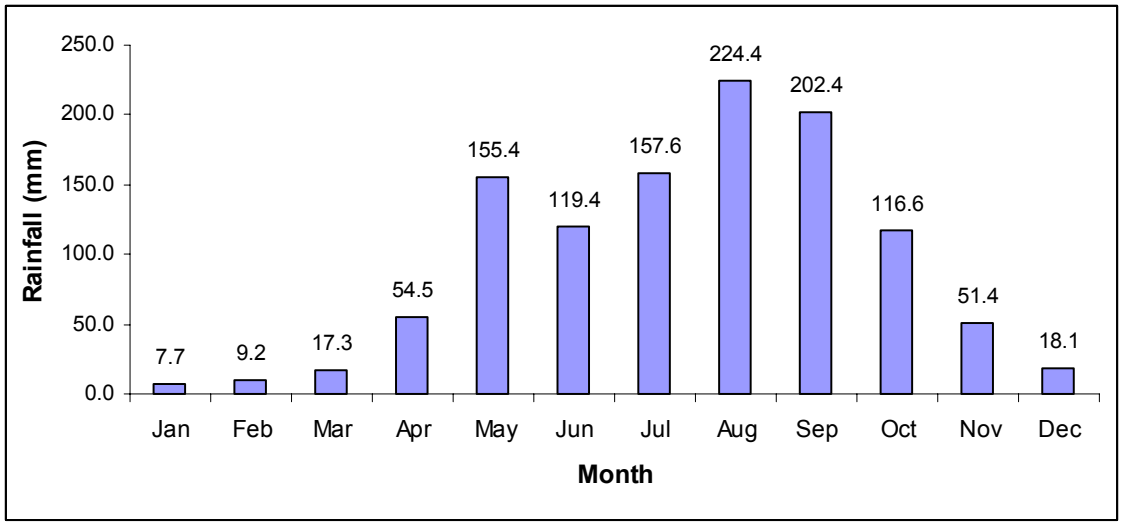


Figure 3.4: Average rainfall at Chiang Mai 1971- 2000 (30 years).
Source: Northern Meteorological Centre (2007)

Figure 3.5 presents the changes to the River Ping crosssection at P67 and P1 stations from 1994 to 2007 and 1972 to 2006, respectively. At P1 there is not much change when compared to P67.

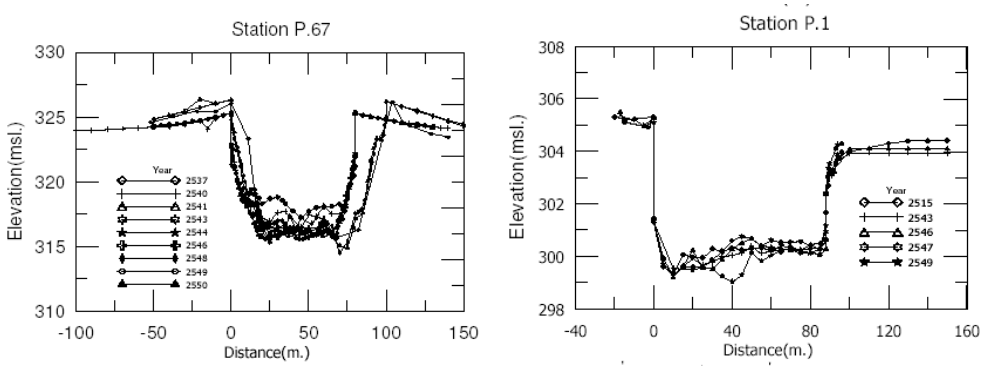


Figure 3.5: Changes to the crosssectiona; shape at P67 and P1.
Source: Eiamkarn, *et al*. (2009, p 1218)

The changes to the river channel have influenced the conveyance factor, which has increased by approximately 7% and 3% at P67 and P1, respectively, according to a recent study by Eiamkarn, *et al*. (2009). As a result, the water stages at P67 and P1 for flood discharge have decreased by approximately 12 cm. and 1 cm. per year, respectively (Figure 3.6). In addition, the water stage has dramatically declined after 2004 to 2007 because of the renovation of the Ping River channel. Flood defences were built after the big flood in 2005 and the government excavated the Ping River channel to increase the conveyance capacity.

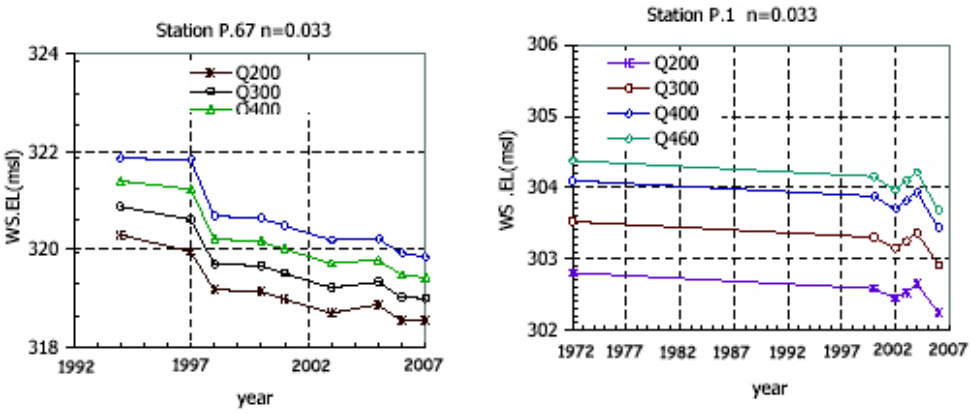
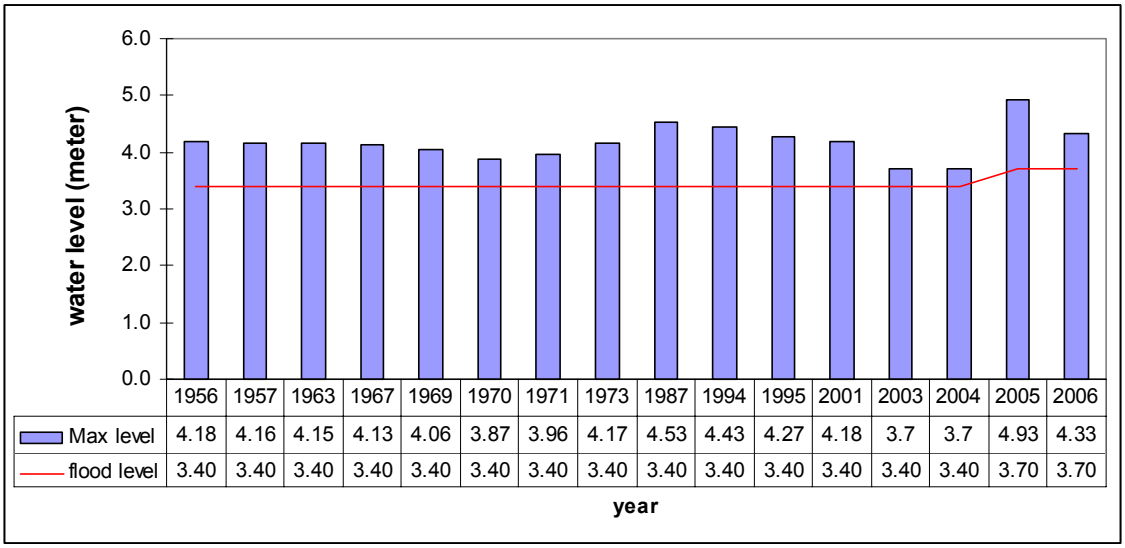


Figure 3.6: The correlation between water discharge and water stage at P67 and P1. Source: Eiamkarn, *et al*., 2009, p 1220 (n is Meanning'n)

3.3 Flooding at Chiang Mai

The flooding in Chiang Mai at P1 station is recorded by the Hydrology and Water Management Centre for the Upper Northern Region. In addition to the main Ping River channel, there are seven minor rivers in the Chiang Mai area, two of which feed into the Ping above station P1, where the average annual discharge is 59.38 m^3/s. The main causes of flooding in the city are water discharges of greater than 460 m^3/s and water stage levels in the main channel exceeding 3.70 m above local datum (304.2 msl) (Hydrology and Water Management Centre for Upper Northern Region, 2007a). Flooding in the city happens annually in response to monsoon rainfall. According to the flood records for the past 50 years in the Chiang Mai city area, the four highest monsoon flood events occurred in 1987, 1994, 1995 and 2005 with water levels of 4.53, 4.43, 4.27 and 4.93 m, respectively as shown in Figure 3.7.

The main causes of flooding in this catchment are considered to be meteorological, fed by monsoonal rainfall. However, Sophhonphattanakul *et al*. (2009) investigated the effect of changes in land use on stream flow in this catchment and they found that changing land use through urbanization, industrialization and deforestation have contributed to flooding in the Upper Ping catchment. Chatchawan (2005) states that Chiang Mai's land use has changed rapidly in response to the National Economic and Social Development Plan volume 5 (1982-1986), which developed Chiang Mai city into a 'Primate city' (Chatchawan, 2005). As a result of development, there has been deforestation in the catchment and the building of infrastructure in the city and along the Ping River has increased.



	1956	1957	1963	1967	1969	1970	1971	1973	1987	1994	1995	2001	2003	2004	2005	2006
Max level	4.18	4.16	4.15	4.13	4.06	3.87	3.96	4.17	4.53	4.43	4.27	4.18	3.7	3.7	4.93	4.33
flood level	3.40	3.40	3.40	3.40	3.40	3.40	3.40	3.40	3.40	3.40	3.40	3.40	3.40	3.40	3.70	3.70

Figure 3.7: Record of maximum flooding at P1 station from 1956-2007.
Source: Chatchawan (2005)

Engineering work on the Ping channel as part of flood control works has changed the datum flooding level. Prior to 2004 the flooding datum level at P1 was 3.40 m. After excavation in 2004 of the Ping River channel and the construction of permanent flood defences along the river, the out-of-bank flooding level changed to 3.70 m. Figure 3.7 shows that the maximum flood events more recently have been higher when compared with previous decades.

The current flood forecasting technique, which is operated by the Hydrology and Water Management Centre for the Upper Northern Region, requires only the water stage record at P67 and they operate on a maximum flood warning time of is 6-7 hrs (Figure 2.4). However, it was clear after the big storm in 2005 that 6 hours is not enough time for preparation and evacuation from Chiang Mai. To improve the flood warning system in this area, real time acquisition and improved quality of the rainfall and water stage data are required. Consequently, more rain gauges, water stage stations and the online systems for the real time collection of the data were installed after 2005 (Natural Disaster Research Unit, 2007a). New rain gauges and float gauge stations were installed above and behind the Mea Taeng Weir, rain gauge and float gauge stations were added at P20, P21 and P1, a rain gauge and bubble gauge were added at P67, and a float gauge at P75. However, the extra information for flood forecasting such as the changing of conveyance capacity in the River Ping is also necessary. For example, a water discharge of 460 m^3/s at P1 would cause flooding in 1972, 2004 and 2005 but this did not happen in 2002 and 2006 (Figure 3.6).

The most important information for flood warning is the accuracy of the timing of the flood, so that the government agency tasked with flood defence and the residents in the area would have more time to prepare, defend and evacuate the area in order to reduce loss of life and minimise the economic damage. Moreover, Chatchawan (2005) pointed out that apart from an effective flood forecasting system, the communication system between the government agency responsible for flood forecasting and defence and the people who live along the Ping River should be improved. During the 2005 flood, the water stage at P67 reached 4.20 meter at 8.00 pm, which meant that the Chiang Mai city area (P1) would be flooded in the next 6-7 hours. However, only the government agency and some people recognised this. It was clear that 6 hours for flood warning and preparation in a city the size of Chiang Mai is insufficient time. If there had been a greater lead time for flood preparation and defence, this may have reduced the massive economic loss that resulted in 2005.

The second important piece of information for a flood forecaster is the maximum height of the water stage at P1, so that they can predict the area that would be flooded based on the elevation of Chiang Mai city. They could then better prepare for the rescue and develop a mitigation plan. Figure 3.8 presents a detailed picture of the flooded area based on the seven sequences of water stage and water discharge, rising up by approximately 0.1 meter per sequence. Therefore, the error in forecasting the water stage at P1 that would be acceptable is less than 0.1 m. Moreover, Nemec (1986) recommended the desirable precision of the error between the actual and model predictions for river stage to be 0.01 m, however, this also depends on the sensitivity of the stage change and stage discharge.

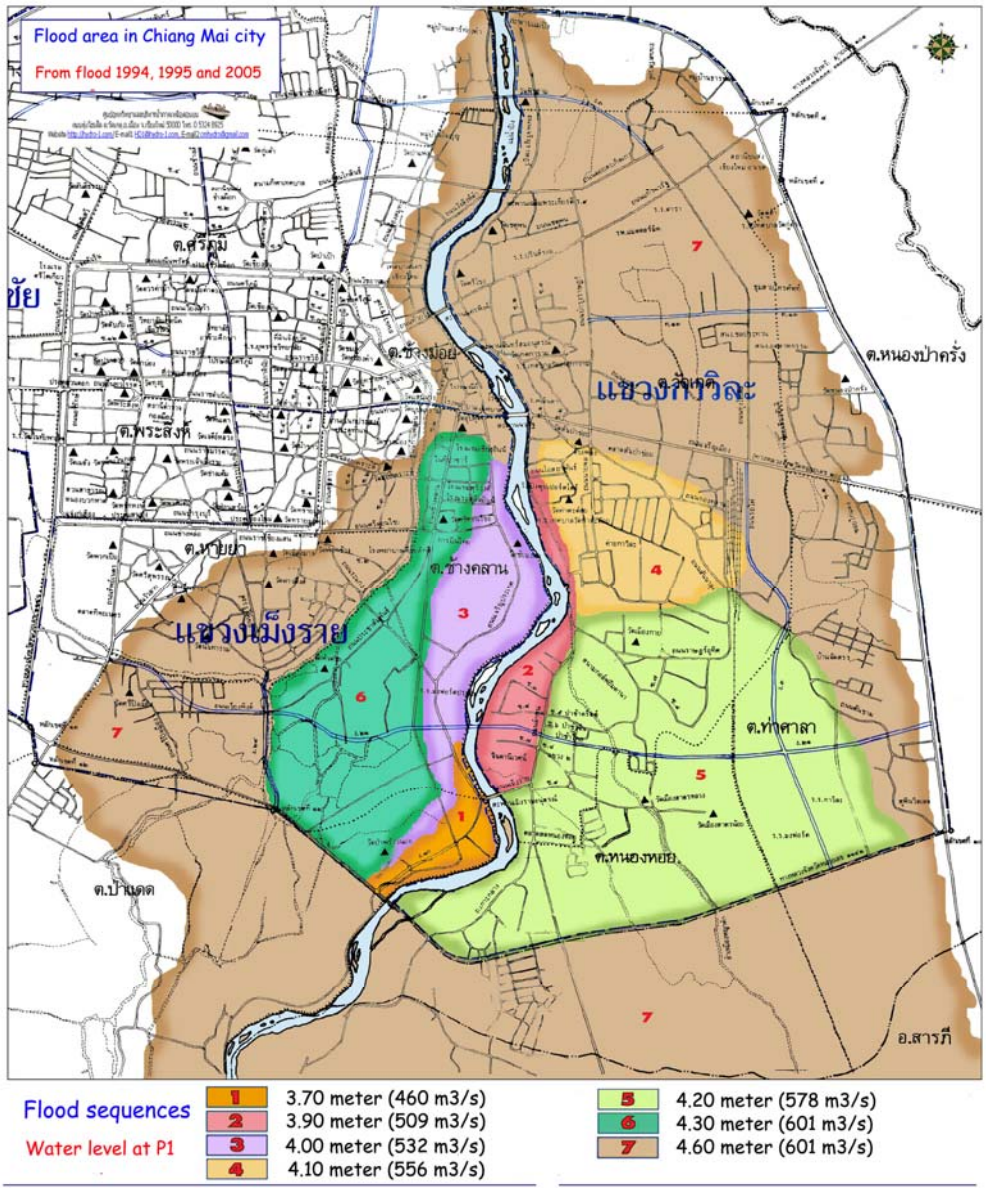


Figure 3.8: Map of flood area in Chiang Mai city.
Adapted from Hydrology and Water Management Centre for Upper Northern Region (2007a)

Data records were available for modelling from year 2001 to 2007. In that time period there are 19 storm events: 2001 (2), 2002 (4), 2003 (1), 2004 (3), 2005 (5), 2006 (4) and 2007 (0), which are used in this research. Taking 2005 as a typical year and one which is used in this research, there were five significant events triggered by heavy and prolonged rainfall (see photo taken during the 2005 flood, Figure 3.9):

- **12-17 August 2005:** The tropical depression moved from the South China Sea and passed through the northern part of the catchment. There was heavy rain on 12 August, with an average of 128 mm in 24 hr in the north Chiang Mai region. It started to flood on 13 August with the river rising at a rate of 12-14 cm/hr. The maximum water level at P1 was 4.90 m at 7.00 pm on 14th August, with a maximum water discharge of 747 m^3s^{-1}. This water level remained for 8

hours. As a result, the flood water covered a very wide area of the city for up to 51 hours. This was the biggest flood event of 2005.

- **12-13 September 2005:** Heavy rain occurred on 9 September. The maximum water level rose to 3.79 m, causing 6 hours of flooding. The river rose at a rate of 1-2 cm/hr and the maximum water discharge was 488 m^3s^{-1}.
- **20-22 September 2005:** Tropical cyclone Vicente. The average rainfall over the catchment was 80.72 mm. The 24hr and maximum water discharge was 700 m^3s^{-1}. The water rose to a maximum of 4.71 m on 21 September and this water level remained for 4 hours. This led to 57 hours of flooding across the city.
- **29 September-2 October 2005:** Typhoon Damrey moved in from the South China Sea causing heavy rainfall in the north of the Ping catchment on 27 September. This time the highest water level reached 4.93 m, the maximum water discharge was 754 m^3s^{-1} and the total flooding time was the longest experienced at 82 hours. This flood was exacerbated by the lack of storage in the catchment and high soil moisture levels from the late September weather. This meant that there was limited capacity to absorb further rainfall and greater runoff.
- **1-3 November 2005:** This event arose from Typhoon Kai Tak, another Northeast monsoon from the South China Sea. The maximum water level for this event was 3.79 m and the flooding time was 9 hours.

Figure 3.9: Photographs of flooding at Chiang Mai during the 2005 storm.
Source: Hydrology and Water Management Centre for Upper Northern Region (2007a)

Since the 29 September flood event in 2005 was the highest on record, this event was one of those selected for testing the neural network models in this study. However, the data records were not entirely complete and the data have been cleaned by SPSS software with linear interpolation (Section 3.). Therefore data from other events have also included in the testing process. Data from low flows in the period between 2001 and 2007 were included in the modelling process.

3.4 Data for the Upper Ping basin, Thailand

Figure 3.10 presents a diagrammatic representation of the topography of the Upper Ping catchment. The red lines show the location of the stage gauge stations. In Figure 3.11 the links between rainfall and water stage recorders in the basin are plotted. Rainfall stations are denoted by blue circles, and water level stations by triangles. Rainfall data are available at hourly and daily intervals. The rainfall station 'Met1' recorded at 3 hr intervals and the rainfall station 'R1' is recorded every 1 hr. Met1 was excluded from the study because the location of the two stations was so close that it did not offer useful additional information. Table 3.2 shows the data that are used in each chapter.

Table 3.2: Available data.

Data type	Available	Chapter 5	Chapter 6	Chapter 7
Water level				
- P20	1976-2004 (hr)/2005 (3hr)	-	-	-
- P56a	2001-05 (hr)	-	-	-
- P75	1998-2007 (hr)	2001-07 (hr)	2003/05-06 (hr)	2001-07 (hr)
- P4a	1970-2007 (3hr)	-	2005-06 (hr)	2005-06 (hr)
- P67	1996-2007 (hr)	2001-07 (hr)	2003/05-06 (hr)	2001-07 (hr)
- P21	1989-2007 (3hr)	-	2005-06 (hr)	2005-06 (hr)
- P1	1982-2007 (hr)	2001-07 (hr)	2003/05-06 (hr)	2001-07 (hr)
Rainfall				
- R1	2001-2007 (hr)	2001-07 (hr)	2003/05-06 (hr)	2005-06 (hr)
- Met1	Aug-Oct 2001-06 (3hr)	-	-	-
Radar image	2003/05-06 (15min – 1 hr)	-	2003/05-06 (hr)	2005-06 (hr)

The data used in the modelling experiments is taken from three water level gauging stations (P1, P75 and P67), one rainfall station (R1) and radar images at one hr intervals. In the first set of modelling experiments, water level and rainfall data are used to explore model types and variable selection processes (Chapter 5). The second phase of the research focused on the potential of radar images to enhance the input record and improve forecasts (Chapter 6) while the final phase considered experimentation with the parameters of a neural network (Chapter 7).

Figure 3.10: Landscape of Study area.
Source: Hydrology and Water Management Centre for Upper Northern Region (2007b).

Patsinghasanee (2004) investigated the correlation between the travel time between upstream stations and the station at P1 based on data between 1998 and 2001. However, the averages of the travel times, which are plotted from three big storm hydrographs in 2005 (S2, S4 and S5), are slightly different (see Table 3.3). This may be due to the channel physical condition changing or different storm patterns.

Table 3.3: The correlation of travel time to P1 station and distance.

Station	**Correlation (hr) (1998-2001)**	**Correlation (hr) (2005)**	**Channel distance (km)**
P20	25	33	103
P4A	24	20	52
P75	-	16	50
P67	-	6	34
P21	14	19	23

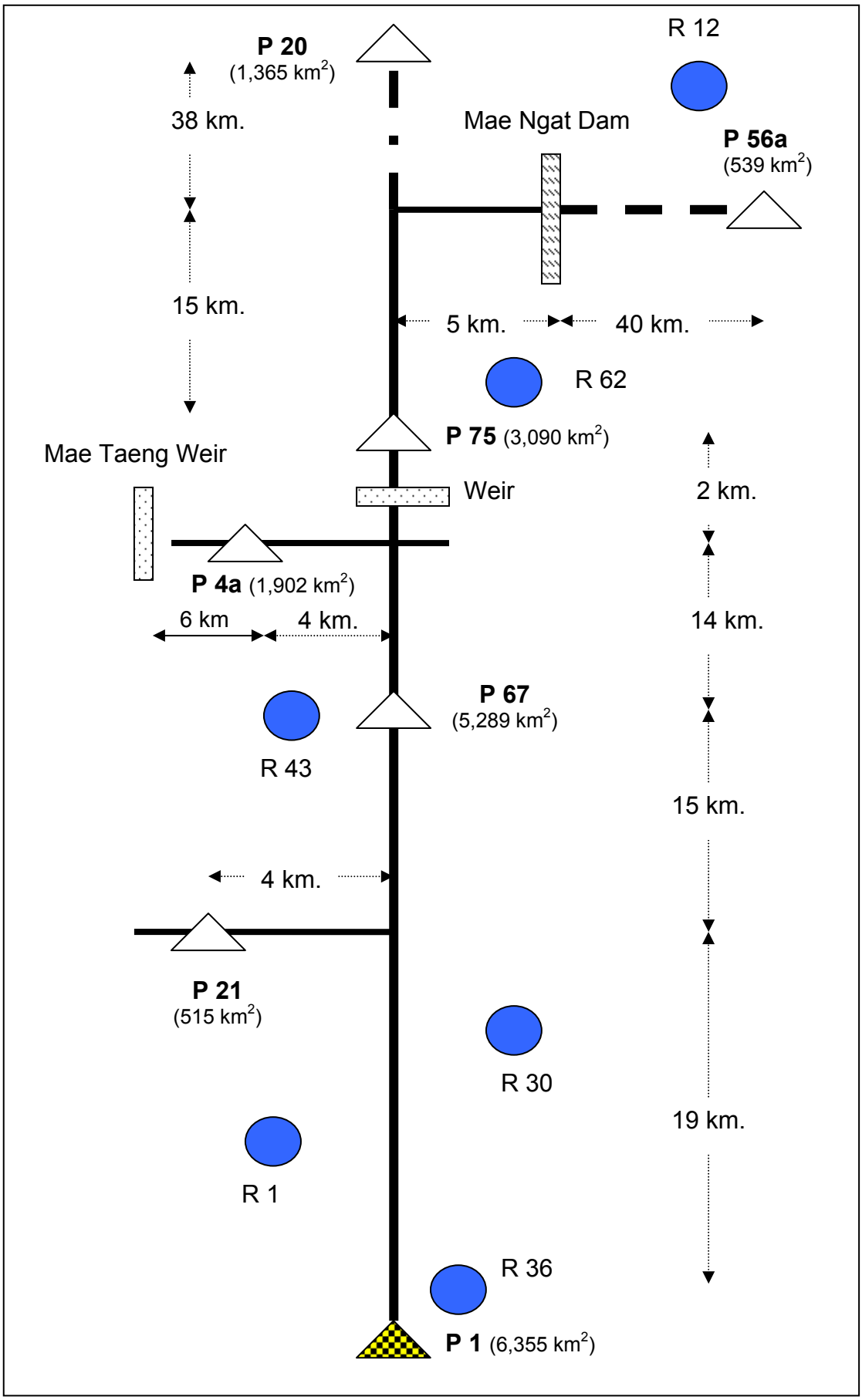


Figure 3.11: Schematic diagram of the rainfall and water level stations.
Source: Hydrology and Water Management Centre for Upper Northern Region (2007a).

3.5 Radar Data

The bureau of Royal Rainmaking and Agriculture Aviation operates 5 radar stations over Thailand (Figure 3.12). Radar images are the CAPPI (Constant Altitude Plan Position Indicator) type, which detect moisture in the cloud and radar reflectivity in units of dBZ. The spatial resolution of the radar image is one km, and the temporal resolution is between 15 minute and 1 hr with a ground coverage radius of 240 km.

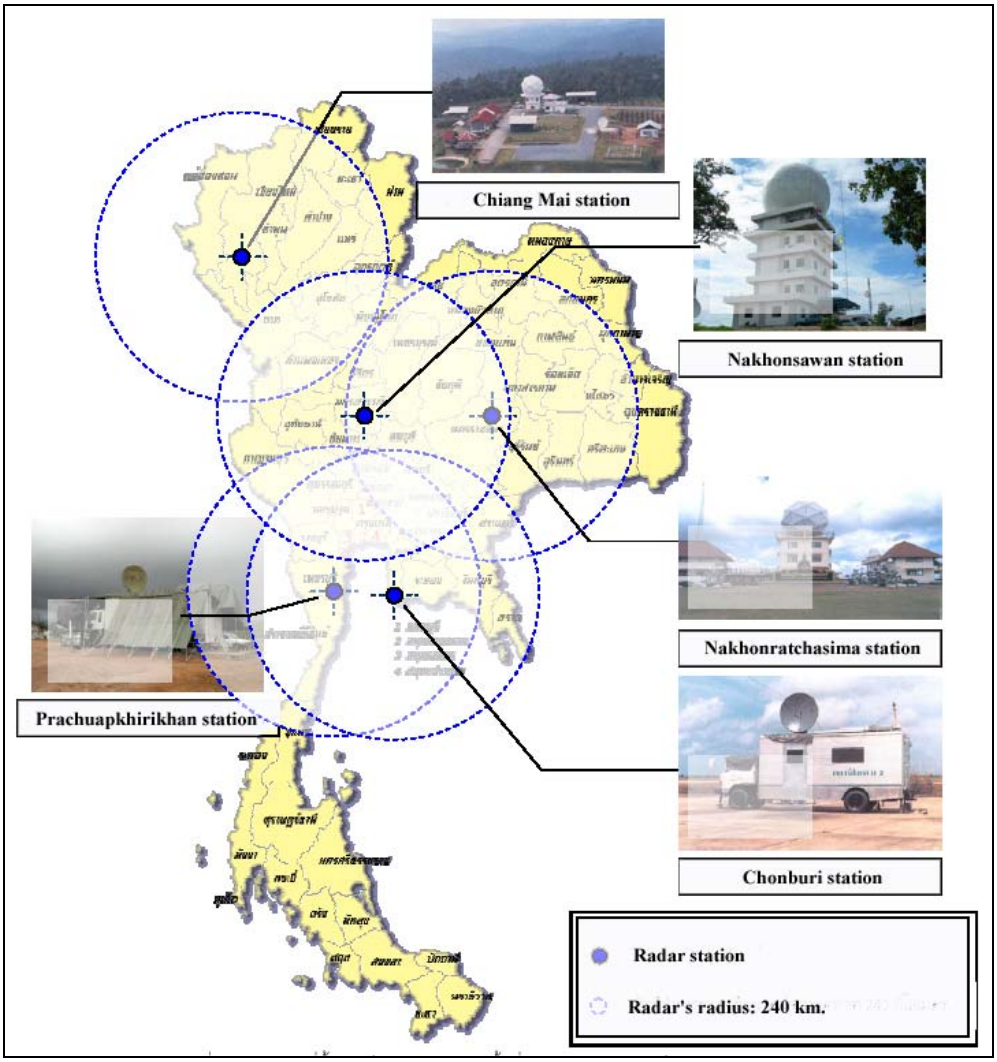


Figure 3.12: Location of five radar stations in Thailand.
Source: Jankad (2007).

The radar images from Chiang Mai station (Om Koi station) were obtained at a relatively late stage in the research, but are used in this study (Chapter 6 and 7). The Om Koi station operates S band (10 cm wavelength), 2.8 GHz. The minimum radius is 25 km at the minimum height 10 km above ground and the maximum is 240 km with an altitude greater than 20 km above ground level (Figure 3.13). As the Chiang Mai station is located on the top of mountain it is not affected by ground clutter (Okumura *et al*., 2003).

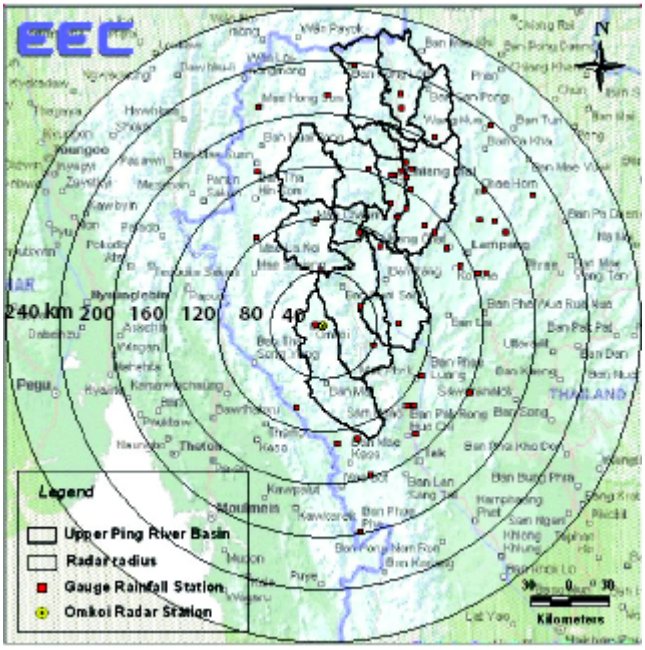


Figure 3.13: Location of the rain gauge station and the radius of Om Koi radar station. Source: Mapiam and Sriwongsitanon (2008).

3.6 Handling Missing Data

Missing data can occur as a result of instrument error, instrument malfunction and human error. As would be expected, there were missing values in the hourly rainfall and water level records, as well as missing radar images.

To deal with the missing water level data, the “Linear interpolation” function of SPSS was used to calculate the missing values. For the missing rainfall at R1, the data were compared with the nearest rain gauge station that had daily records and averaged into 1 hr intervals. The missing radar images were very difficult to interpolate. The best way to fill in the missing images was by using images from the same hour or the nearest hour. Fortunately during the storm events used in this study, there were very few missing images. The exception is for 2004 when the absence of radar data precluded modelling floods using the radar images.

3.7 Conclusions

The Upper Ping catchment presents an opportunity to explore the accuracy and potential of neural network forecasting opportunities in a large-scale monsoon catchment. There is limited hydrological data because of the lack of long-term records, but flooding is a serious problem in the region and more accurate forecasting will be beneficial for the region. Initially the data were limited to 2000-2005. Some of the earlier modelling concerns this shorter dataset, which was enhanced when the 2006 and 2007 data were made available (Chapter 5, 6 and 7).

As with all hydrological data sets there are gaps in the record, and the timescales do not always match. However, careful examination of the records made it is possible to use algorithms to interpolate data when the gaps were small. The presence of hourly data at a number of sites was also particularly helpful for this process. In the next chapter, input determination methods are reviewed, which will help determine which inputs to use in the neural network modelling experiments.

Chapter 4 Input Variable Determination Techniques

4.1 Introduction

The literature review in Chapter 2 highlighted that there is little guidance on many aspects of neural network (NN) development, in particular on how to choose the input variables. This was further highlighted in the recent review of neural network modelling by Maier *et al.* (2010). This chapter examines the range of approaches available for input determination. These methods are applicable to NNs and any other type of data driven modelling method. The chapter starts with a review of the commonly used method of trial and error, and then progresses through an examination of statistical, neural network specific and data-driven approaches. These methods are evaluated together with the hydrological contexts in which they have been implemented, and then a final set is chosen for subsequent experimentation in this research. One of the main criteria for the choice of a method will be the ability to automate the variable selection process in order to remove the subjectivity around the choice of input variables.

4.2 Trial and Error or an *Ad Hoc* Approach

The trial and error, or *ad hoc* method, involves systematically presenting the neural network with different combinations of input variables that are thought to have a relationship with the output variable. For example, rainfall is accepted as being a physical driver of runoff, while upstream river levels are indicative of downstream levels at a later time. These two variables are therefore acceptable inputs for neural network rainfall-runoff models. The trial and error method has been employed in a number of hydrological applications involving the development of NNs. For example, an early paper by Shamseldin (1997) on applying NNs to rainfall-runoff modelling involved testing 4 scenarios, which differed in terms of the input variables chosen. In the recent review by Maier *et al.* (2010), this method is referred to as a 'model based' approach since different sets of inputs are chosen and then neural network models are developed to determine which yields the best combination of inputs. Their review showed that 37 of the 210 papers published between 1999 and 2007 employed this type of approach. Papers published since 2007 have continued to use this method. For example, Li *et al*. (2009) used this approach when developing a neural network model to forecast river discharge on the Huaihe River in China, while Partal (2009) systematically tried a range of 1 to 5 different input variables in order to develop different neural network river flow forecasting models.

The main problem with this method is that many different combinations must be tried to

ensure that an acceptable set is found. Technically, optimized neural network models must also be developed for each case (Maier *et al.,* 2010), which adds further experimentation to the process. A very thorough application of trial and error can be regarded as applying brute force and is likely to be a computationally infeasible approach, especially when the number of input variables is large. Therefore, although the evidence from the literature suggests it is a valued and much used methodology, it is appropriate to research alternative, automated input determination methodologies to avoid having to use trial and error.

4.3 Sensitivity Analysis

Sensitivity analysis is a method used to measure the effect on the outputs after making changes to the input variables (Breierova and Choudhari, 2001). The variables that have little or no effect on the output can then be eliminated. There are different ways to undertake sensitivity analysis, e.g. by calculating the slopes between the inputs and the output (Montano and Palmer, 2003) or based on visual interpretation by looking at the weights of the connecting nodes (Ozesmi and Ozesmi, 1999). Maier and Dandy (1996) employed sensitivity analysis to develop a NN and they reduced the number of potential inputs by 50%. Moreover, when they trained neural network models using the full input dataset compared to the neural network used with the smaller number of inputs, the resulting neural network models with the smaller input dataset performed better across a range of performance measures. An additional benefit was the reduction in the training time of the neural network. The review by Maier *et al.* (2010) indicated that 7 out of the 210 papers reviewed had used sensitivity analysis to reduce the number of input variables. The work by Sudheer (2005) provides a good example, where perturbation analysis was performed on a runoff model for the River Narmada in India. Sudheer (2005) used this type of sensitivity analysis to show how different inputs have a direct influence on producing the shape of the hydrograph. This type of analysis can inform the choice of which inputs are needed in further neural network model development. A more recent paper published by Kalra and Ahmad (2009) used sensitivity analysis to select inputs to a support vector machine application of stream flow forecasting.

This method is not used in this thesis because it requires an element of trial and error while the approaches of interest in this research are those that can be automated to render the process as objective as possible.

4.4 Statistical Approaches to Input Determination

Statistical approaches offer some potential methods of automation that can be used to determine the inputs, e.g. calculation of a correlation coefficient. Other techniques include stepwise linear regression and partial mutual information, which are reviewed in the sections that follow.

4.4.1 Correlation Analysis

The correlation between two variables indicates to what degree there is a linear relationship between them. The correlation coefficient ranges between -1 to + 1 where -1 denotes a perfect negative relationship between the variables while + 1 indicates a perfect positive relationship. Values approaching 0 indicate no strong relationship. Before running a neural network application, it is suggested that variables that have correlations close to zero are removed, a process known as prewhitening, in order that the true relationships are obtained (Chatfield, 2001; Wei, 2006; Maier and Dandy, 1997).

The correlation coefficient is easy to calculate in packages such as Excel or SPSS and this method has been used frequently to aid in the choice of input variables. According to the review by Maier *et al.* (2010), correlation was used in 60 out of 210 papers reviewed. An example includes the work by Dawson *et al*. (2006b), who calculated correlation coefficients between the discharge at various lead times and historical data from upstream discharge and rainfall stations to determine which ones to use in their genetically optimised neural network rainfall-runoff models. Given the large number of potential inputs, this method provided a way of drastically reducing the number of input variables. More recent examples include the work by Liu *et al.* (2009), who applied correlation to select variables for the development of neural network models of water pollution. Kim *et al.* (2009) used correlation in the development of two kinds of rainfall-runoff model. Jia *et al.* (2009) used correlation in the development of an online rainfall-runoff forecasting and real-time water resources assessment system, and Shevnina (2009) selected stream gauge inputs based on high correlations for a forecasting model of the spring floods with a lead time of 20-24 days in advance.

Although this method is one of the most widely used (Huang *et al*., 2009; Maier *et al*., 2010), it has some disadvantages. For example, it does not result in any physical understanding of the variables in the model (Bowden *et al*., 2005a). Moreover, it assumes the relationship between the variables is linear, which is not often the case in rainfall-runoff modelling. However, because it is widely used, it provides a good

benchmark method against which to compare the other approaches. Therefore, it will be used in this research. If a threshold for what is considered to be a good correlation is chosen, then the method is essentially automated. Selection of an acceptable correlation for an automated process is always a trade off between including too many variables, which makes the computational time very long, and too few, which risks finding a less than optimal solution. In this research the catchment is very large and the rainfall is monsoonal, so linear correlations are also unlikely to be strong. However the variables that will be used as the modelling inputs, i.e. stage and rainfall, are physical drivers and are related. There are almost no modelling papers where the authors state the correlation coefficients that they used in their studies that can be used to guide this research. Various experiments were undertaken to look at the correlations. Originally a correlation of 0.9 was used as a cutoff as this still left many input variables remaining. However, in later experiments when 0.9 produced no input variables, a lower value of 0.7 was used.

4.4.2 Multiple Regression

Multiple regression attempts to quantify how one or more independent variables (x_1 to x_n) affect the dependent or response variable, y, in a linear way (Draper and Smith, 1966). A multiple regression equation takes the following form:

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon \tag{4.1}$$

Where a is a constant, b_1 to b_n is a vector of coefficients and ε is an error term. Given the input and output vector, a least squares methodology is used to estimate the constant and the vector of coefficients. There are three main methods for building multiple linear regression models: forward, backward and stepwise (Brace *et al*., 2006). The forward method adds one variable to the model at a time while the backward method adds all variables to the model at the same time and then removes those with a low correlation. Stepwise regression is developed from the forward selection procedure (Draper and Smith, 1966). This method finds the best input first, and then adds successive inputs as they improve the performance of the model. Since the main interest is in determining the input variables with the most significance, stepwise linear regression is the method that is most appropriate for determining variables for neural network model development. This method is not explicitly mentioned in the review by Maier *et al.* (2010). They do refer to stepwise methods but mean something entirely different, e.g. pruning. There is little evidence of this approach being used in the literature. The work by May and Sivakumar (2009) is the only recent example found. The authors used stepwise linear regression to select the input variables to a neural network stormwater quality model for urban catchments in the US. However, when

compared with linear regression models developed on the same dataset, they found worse performance with the neural network models. It is possible that the modelling problem was simply linear in this situation and that the application of non-linear neural network models was unable to add anything useful to this modelling exercise. Thus the method of input variable determination in this example was not overly useful to the development of better performing neural network models.

As this method is easy to implement in SPSS and also offers an automated approach, it will be used as one of the input determination methods in this research.

4.4.3 Partial Mutual Information (PMI)

Neural networks have the capacity to include a wide variety of variables, some of which may be highly correlated. Maier *et al.* (2010) highlight the potential problems associated with the inclusion of too many input variables, especially where there is redundant information due to correlation. The main problems include the potential for overfitting, since the number of inputs increases the ratio of training data to weights that need updating during neural network training, as well as the introduction of additional local minima, which makes it more difficult to find an optimal set of neural network weights. The Partial Mutual Information (PMI) method seeks to mitigate this issue by taking into account the interaction between variables. The PMI algorithm was first introduced by Sharma (2000) and Sharma *et al.* (2000) as an extension to the concept of mutual information, to find inputs for rainfall prediction at lead times of 3 months to two years. Bowden *et al*. (2005a, b) then used PMI and other input determination methods for neural network modelling of salinity in the River Murray in South Australia.

The PMI between two variables can be formulated as:

$$PMI = \frac{1}{n}\sum_{i=1}^{n} \ln\left[\frac{P_{X',Y'}(x_i', y_i')}{P_{X'}(x_i')P_{Y'}(y_i')}\right] \tag{4.2}$$

where x_i' and y_i' are the i[th] residuals in a sample data set of size *n*, $P_{X'}(x_i')$ and $P_{Y'}(y_i')$ are the marginal probability distributions and $P_{X',Y'}(x_i', y_i')$ is the joint probability distribution.

Application of the algorithm works as follows:

1. The set of inputs x_i (where i = 1 to n variables) and the output y is identified. Examples of x_i for this research would be previous values of river levels at the gauging station of interest and upstream while the output would be the river

level in the future for a given lead time.

2. The PMI between y and each x_i variable is calculated using equation 4.2.
3. The PMI values are ranked and the x_i variable with the highest PMI is identified.
4. The vector of x_i is then randomly reshuffled to create a set of randomised inputs.
5. The PMI values are calculated again and the 95th percentile PMI score is extracted.
6. If the maximum PMI determined in step 3 is higher than the 95th percentile PMI score, then the variable is selected by the method as an input variable to the neural network model.
7. This input is then removed from the set of inputs x_i. Steps 2 to 6 are repeated until the condition in step 6 fails and no further inputs are selected.

The resulting set contains those inputs that are most significant and independent of one another and which will be used in the neural network model development.

In the recent review by Maier *et al.* (2010), a non-linear approach like PMI was used in only 7 out of the 210 papers reviewed. However, the method is now appearing with more frequency. More recent examples include the work of May *et al*. (2008a, b), Fernando *et al*. (2009), Hejazi and Cai (2009) and Corzo *et al*. (2009). May *et al.* (2008a) applied the PMI algorithm to the development of neural network models to forecast water quality within two water distribution systems. Compared to trial and error, the PMI approach produced much better performing and parsimonious neural network models. They also argued that the procedure could provide insights into the relationships that exist between variables in the system. Both papers by May *et al.* (2008b) and Fernando *et al.* (2009) address the computational efficiency and accuracy of the PMI algorithm whilst simultaneously showing the superiority of the PMI algorithm over other linear or *ad hoc* input determination methods. Hejazi and Cai (2009) used a variation of PMI to select from 121 potential variables in order to predict daily reservoir release on 22 reservoirs in California. The authors concluded that the model performance was improved when using this method to choose the input variable dataset to the neural network. Finally, Corzo *et al.* (2009) used PMI to select the input variables to a neural network used to replace both process-based models and the routing component for the River Meuse catchment. They did not evaluate other input determination methods but used this one alone based on recommendations in the literature.

The advantage of PMI is that it takes variable dependency into account and is technically a non-linear approach, but it is a more complicated algorithm to implement and requires knowledge of programming or high level scripting. This has implications

for its use in operational forecasting and neural network model development. However, once programmed, it is a highly automated approach so this method is used in this research.

4.5 Neural Network Specific Input Determination Methods

This section reviews methods that have been specifically developed for NNs, i.e. pruning algorithms and saliency analysis.

4.5.1 Pruning Algorithm

A pruning algorithm is a neural network technique that removes unimportant or weak connections between nodes as well as the nodes themselves. The concept behind this algorithm is to start with a fully connected network and to then remove the least significant connections between the inputs and outputs. Pruning algorithms fall within model-based stepwise approaches in the taxonomy of input variable determination methodologies devised by Maier *et al*. (2010). These approaches (of which pruning is only one) have been used in 12 out of 210 papers. One of the earliest neural network hydrological examples is the work by Abrahart *et al*. (1999), who tried two different types of pruning algorithms, i.e. magnitude-based pruning (eliminating unwanted links) and skeletonization (eliminating unwanted nodes). Rainfall-runoff models were built using seasonal information, rainfall, potential evapotranspiration and flow to predict flow one step ahead. The authors found that both types of pruning algorithm could reduce the total number of connections between 10% - 43% whilst retaining good model performance.

A more recent example includes the work by Corani and Guariso (2005), who compared the performance of neural network flood forecasting models with and without pruning on two different catchments in Italy. Between 30 to 40% of the inputs were removed and the NNs trained on these smaller input datasets performed better on unseen data. More recently, Chen and Yu (2007) used pruning in a flood forecasting support vector machine application, which has some similarities to NNs. The authors showed that the pruned models resulted in a reduction of model complexity while still retaining good model performance.

Although not frequently used in hydrological applications, there is a freeware pruning algorithm available for Matlab (Ravn, 2003). This means that the algorithm is easy to implement and it is also an automated approach. Pruning will therefore be tested in this research.

4.5.2 Saliency Analysis

Saliency analysis is a method whereby a neural network is first trained and then one input variable is removed at a time by adding a weight of zero between the input and the hidden nodes. This operation is then repeated for each input variable and the output is analysed. Comparison of the performance when individual inputs are removed determines which inputs have had little effect on the overall result and which can therefore be removed.

Saliency analysis does not appear as a method of input determination or neural network model construction in the review by Maier *et al.* (2010). In fact the only known neural network rainfall-runoff modelling example was that undertaken by Abrahart *et al.* (2001). The authors found that this method provided a useful tool to gain knowledge about the relationship between the inputs and the output, in particular between previous river flow inputs and current river flow levels, seasonal variation and individual catchment response. However, Bowden *et al*. (2005a) highlight a disadvantage of this method, which is that the network is not retrained after each input variable is removed. Given that this technique is not automated and requires an element of trial and error, it was not considered further in this research.

4.6 Data-driven Input Determination Methods

The final set of input determination approaches to be considered here come from the field of artificial intelligence. Self-organizing maps (SOMs) are a type of unsupervised neural network that can be used for input variable determination. Other approaches reviewed in this section include: genetic algorithms, M5 model trees and a data mining algorithm embedded in the WEKA software (Witten and Frank, 2005).

4.6.1 Self-Organizing Map (SOM)

The self-organizing map (SOM), developed by Kohonen (1984), was reviewed as a type of neural network in Chapter 2. However, it is also possible to use a SOM as a method for input determination. Maier *et al.* (2010) consider the SOM to be part of a clustering-based approach to input determination. Of the 210 papers reviewed, only 6 papers used a clustering approach, of which the SOM is only one type. From a search of the literature, the only true example of a SOM being used to determine the input variables is the study by Bowden *et al.* (2005a). The SOM was used to cluster the inputs and then one input from each subsequent cluster was sampled reducing the overall number of variables. This reduced set was then fed to a hybrid genetic

algorithm / general regression neural network (GAGRNN), which was then used to determine the final set of input variables. The SOM-GAGRNN was compared with the PMI algorithm (Bowden *et al*. 2005b) and neural network models developed previously. The results showed that both methods produced better performing models than previous neural network models reported in the literature but that the PMI produced the most robust predictions on unseen data.

There are other studies that use a SOM to reduce a large number of inputs but they do not produce a smaller subset for subsequent training with a supervised neural network. For example, Hsu *et al.* (2002) developed the Self-Organizing Linear Map (SOLO) approach in which a SOM classifies the input data and linear regression is then used to produce a forecast. It is possible to visualise the output of the SOM and see which of the variables are the most important. However, this feature is not used by Hsu *et al*. (2002) to further develop supervised neural network models. The same type of approach was taken by Chang *et al.* (2007), who developed a SOM forecasting model for the Da-Chia River in Taiwan. A reduced input set for use by supervised neural network models was not the purpose of the exercise. Rather the SOM implicitly reduces the inputs while outputting a flood forecast.

Following the literature review of this technique, it was decided to exclude the SOM from this research because it is not an automated approach. There is an element of trial and error involved in choosing the size of the network, the network training parameters and the number of iterations to train the network, making it a relatively difficult approach compared to some of the others reviewed so far. Moreover, it is likely that a flood forecaster working in Chiang Mai would not find this a practical methodology.

4.6.2 Genetic Algorithms (GA)

A genetic algorithm (GA) is based on the concepts and language of biological evolution and natural selection, and was first developed by Holland (1975). The main difference between GAs and other optimisation methods is that a GA can search a population of possible solutions, whereas classical optimisation techniques work on a single solution (Goldberg, 2003). A GA can be broken down into a series of steps as follows (Pai, 2006):

1. Initialise a random population of chromosomes or solutions. These are generally strings that contain model parameters or settings.
2. Evaluate the fitness of each chromosome using an objective function.
3. Select the best performing chromosomes from the population for reproduction

and expose them to crossover and mutation operators to create new offspring. Crossover is the process whereby two parents swap genetic material at a randomly chosen crossover point and mutation is a random change at some point along the chromosome.

4. Use a selection method to create the next generation.
5. Repeat steps 2 to 4 until a stopping condition is reached where an optimal or acceptable solution is generated.

There are several applications of GAs in hydrology. One of the main uses has been in the calibration of conceptual rainfall-runoff models, with the earliest work undertaken by Wang (1991) and Franchini (1996). Other studies have employed the shuffled complex evolution algorithm to calibrate a range of different conceptual models (Khu and Madsen, 2005; Sardinas and Pedreira, 2003). GAs have also been used to find the starting weights of a neural network rainfall-runoff model (Whiteley *et al.*, 1990), for making updates to a trained neural network (Shamseldin and O'Connor, 2001), to actually train the neural network (Jain and Srinivasulu, 2004; Wu and Chau, 2006), and to determine a more representative training dataset for an neural network hydraulic model (Kamp and Savenije, 2006).

The review by Maier *et al.* (2010) classed GAs under global input determination methods. Of the 210 papers reviewed, global methods have only been used 5 times for input determination, where a GA is only one type of global method. There are actually very few examples to report. As reviewed earlier in section 4.6.1 on SOMs, Bowden *et al.* (2005b) used a GA in combination with a GRNN and a SOM to determine the input variables to a neural network water quality model. Anctil *et al.* (2006) used a GA to determine which rainfall inputs to use for a neural network stream forecasting model from amongst a set of 23 rain gauges. Finally, Heo and Oh (2008) showed that a GA can be used to prune a neural network and reduce the size by 8-25%. The authors showed that the GA implementation led to good model performance when compared with other construction methods or pruning techniques.

The WEKA software (Witten and Frank, 2005) contains a GA that can be used for input variable determination. Given the ease of use of the WEKA software and the automated nature of this approach, this method will be used in this thesis.

4.6.3 M5 model trees

M5 model trees were first developed by Quinlan (1992) as a data mining technique for classifying data. The first step in the M5 model tree is to partition the input space into local sub-spaces or smaller problems and then fit regression equations to each sub-

space. There is no mention of M5 model trees as an input determination method in the review by Maier *et al.* (2010). This is most likely due to the fact that the use of M5 model trees used in hydrological applications so far have been as standalone models rather than as input determination methods. For example, Solomatine and Xue (2004) built M5 model tree and neural network models for the Huai River in China. However, they used correlation analysis to select the inputs and then used the M5 model tree as a standalone model against which the neural network model was compared. A hybrid model was then built which used the M5 model tree to partition the solution space into smaller problems, for which individual neural network models were then trained. The hybrid model outperformed the M5 model tree and the NNs used on their own. Bhattacharya and Solomatine (2005) also applied M5 model trees and NNs as standalone discharge rating curve models but not as a method for input determination.

Given that M5 model trees can be used for input determination but have not as yet been used in this way, they will be used in this research. Just as with the genetic algorithm, the WEKA software has an implementation of M5 model trees, which makes this method both easy to implement and highly automated.

4.6.4 Data Mining Approach based on Hill Searching Methodology

In addition to GAs, there are many different search and optimisation algorithms available. The WEKA software (Witten and Frank, 2005) has some in-built attribute selection procedures of which the GA is one. Another one of these is exhaustive search but this is not practical for the requirements of this research as the search time increases exponentially with the number of inputs. However, there is a greedy hill-climbing search procedure in the WEKA software that provides a simple yet potentially powerful automated approach. Moreover, it has not been used for input determination in the context of neural network flood forecasting before. The search procedure starts by selecting a random subset to the solution. The algorithm then makes small changes through many iterations until no further improvements can be found. Given the automated nature of the approach and its simplicity to implement, this data mining search approach will be used in this research.

4.7 Conclusions

The review of input determination methods has shown that hydrological forecasters have adopted a range of approaches. However, the main criteria for use in this research are: (a) ease of implementation and (b) the automated nature of the approach. The Upper Ping catchment is large, the data are limited and real time forecasting is

required. The ease of implementation is an important consideration because there are already enough difficulties in neural network model development due to the lack of guidelines (ASCE 2000a, b; Dawson and Wilby, 2001). If the data input determination aspect of neural network modelling can be made easier, then there is a greater chance that hydrologists will use this technique, making NNs a potential operational choice by forecasters. The techniques chosen for experimentation include: correlation; stepwise regression; a genetic algorithm; M5 model trees, a data mining search algorithm; a pruning algorithm and PMI. In addition, it was decided to combine correlation and stepwise regression to see whether this hybrid approach would yield more parsimonious models.

Building on this review and evaluation of data input determination options, the next chapter reports an initial set of neural network modelling experiments to forecast stage for the Upper Ping River at Chiang Mai. The effects of using different input determination techniques are evaluated.

Chapter 5 Exploring Effective Neural Network Forecasting Procedures for the Upper Ping River

5.1 Introduction

This chapter presents four case studies using a variety of input determination techniques in order to discuss the relative efficiency of different neural network (NN) approaches for flood forecasting at P1 on the Upper Ping River (Figure 3.11). These cases were selected for discussion in this chapter from amongst multiple experimental runs because they are either typical of the results obtained in this study, or provide potentially useful results for practical forecasting.

The NNs in this research were developed using the Neural Network Toolbox in Matlab. This research focuses on comparing a number of different input determination techniques to better understand the relative effectiveness of these different approaches including: linear correlation, stepwise regression, a method combining correlation and stepwise regression, M5 model trees, a data mining algorithm from the WEKA software, a genetic algorithm, partial mutual information (PMI) and a neural network pruning algorithm, all of which were described in detail in chapter 4. The results from each model were then evaluated using the HydroTest website (Dawson *et al*., 2007) as outlined in the next section.

5.2 Evaluation of the Model Results

Methods for the evaluation of hydrographs and stage prediction have been developing since the 1970s. More recently, Dawson *et al.* (2007) classified evaluation metrics into three distinct categories. The first set provides statistical descriptors, e.g. the minimum, maximum, mean and standard deviation of the actual and predicted dataset. These measures can provide a general guide to the overall model performance, but they are less helpful to a flood forecaster who is interested in a small number of extreme and less frequent large events. The second category covers statistical goodness-of-fit measures that capture information about the residual errors between the actual and predicted values from a model. The units are normally expressed in the variable of interest, i.e. errors in river levels would be expressed in metres or centimetres. These measures are very helpful for comparing model results from a range of sites and time periods. Examples include the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE), and model performance is generally best when these errors are minimised. The third category covers dimensionless coefficients such as the coefficient of determination or the Nash-Sutcliffe coefficient of efficiency (CE) (Nash and Sutcliffe,

1970). Values for these types of measures often range between 0 and 1 or -1 and 1 where model performance is best at a value of 1.0.

The methods that are used regularly in the literature should theoretically be the most helpful, although they may also represent a researcher's need to provide an evaluation measure that can be compared to work at other sites and times. Karunanithi *et al*. (1994) drew attention to the most commonly employed error measurements: Mean Squared Error (MSE), Mean Squared Relative Error (MSRE), Coefficient of Efficiency (CE) and Coefficient of Determination (R^2). Examples of NN hydrological research that have used these different measures include:

- MSE (Karunanithi *et al*., 1994; Raman and Sunilkumar, 1995; Cigizoglu, 2005; Bowden *et al*., 2006; Dawson *et al*., 2006b; Kisi, 2006; Sahoo and Ray, 2006; Sahoo *et al*., 2006; Leahy *et al*., 2008; Ei-Shafie *et al*., 2009; Partal, 2009).
- MSRE (Dawson and Wilby, 1998, 1999; Dawson *et al*., 2000, 2006a).
- CE (Minns and Hall, 1996; Chang and Hwang, 1999; Dawson *et al*., 2006a; Kerh and Lee, 2006; Leahy *et al*., 2008; Lin and Chen, 2008; Yang and Chen, 2009).
- R^2 (Lorrai and Sechi, 1995; Campolo *et al*., 1999; Dawson and Wilby, 1999; Corani and Guariso, 2005; Giustolisi and Laucelli, 2005; Kumar *et al*., 2005, Cannas *et al*., 2006; Dawson *et al*., 2006a, b; Kerh and Lee, 2006; Kisi, 2006; Hung *et al*., 2009; Kim and Ahn, 2009; Mukerji *et al*., 2009).

In addition, other error measures have also been widely used, which include:

- RMSE (Hsu *et al*., 1995; Smith and Eli, 1995; Maier and Dandy, 1996, 1998; Dawson and Wilby, 1998, 1999; Campolo *et al*., 1999, Danh *et al*., 1999; Dawson *et al*., 2000, 2006b, Bowden *et al*., 2005b, 2006; Corani and Guariso, 2005; Kumar *et al*., 2005; Cannas *et al*., 2006; Kerh and Lee, 2006; Sahoo and Ray, 2006; Sahoo *et al*., 2006; Seidou and Ouarda, 2007; Chidthong *et al*., 2009; Fernando *et al*., 2009; Hung *et al*., 2009; Hejazi and Cai, 2009; Kim and Ahn, 2009; Mapiam and Sriwongsitanon, 2009; Mukerji *et al*., 2009; Remesan *et al*., 2009; Wang *et al*., 2009).
- MAE (Liong *et al*., 2000; Supharatid, 2003; Chang *et al*., 2004; Cannas *et al.*, 2006; Chen *et al*., 2006; Dawson *et al*., 2006b; Karunasinghe and Liong, 2006; Kisi, 2006).
- Correlation Coefficient (R) (Kerh and Lee, 2006; Sahoo and Ray, 2006; Sahoo *et al*., 2006; Leahy *et al*., 2008; Mapiam and Sriwongsitanon, 2009; Partal, 2009; Remesan *et al*., 2009; Wang *et al*., 2009).

- Error of Peak Discharge-EQp and Time to Peak-ETp (Chang and Hwang, 1999; Kerh and Lee, 2006; Lin and Chen, 2008).

The HydroTest website (Dawson *et al*., 2007) allows a user to calculate a large number of performance measures from an uploaded dataset containing the actual and predicted values from any model. Although designed for hydrological models, the system could actually be applied to any model results supplied by a user. These measures all have different properties and uses, some of which are more or less relevant to flood forecasting, and therefore a decision must be made in terms of which measures provide the most effective evaluation of a model in forecasting terms. As a flood forecaster is most interested in the extreme values, measures such as RMSE, MAE and CE will be theoretically, and practically, the most powerful. After reviewing Dawson *et al.* (2007) and looking at other statistics employed in the literature, it was decided to use the following performance measures in this thesis:

- Root Mean Squared Error (RMSE), where 0 indicates a perfect match. This process is used to evaluate the overall agreement between the actual and the predicted values. This measure gives more weight to larger errors, which are generally of more interest to flood forecasters.
- Mean Absolute Error (MAE), where 0 indicates a perfect match. Stephenson (1979, cited in Dawson *et al*., 2007) recommends using MAE to compare single event models. Unfortunately, this measure does not provide information on the extent of under or over estimation.
- Coefficient of Efficiency (CE), where 1 represents the best performing model. Originally presented in Nash and Sutcliffe (1970), this measure is widely used by hydrologists when reporting model performance.
- Peak Difference (PDIFF), where 0 indicates a perfect result, a positive value indicates an underestimate, and a negative value indicates an overestimate in the peak value. This measure is not suitable for multiple flood events but works well on single events of the type found in this study where there are a limited number of monsoon-fed river stage peaks to analyse and forecast.

The RMSE, MAE and PDIFF are all expressed in the units of stage while CE is a dimensionless coefficient. However, these statistical measures only have so much value, e.g. a good MAE may hide a consistent over or underestimation of the flood peak and the timing of the rising limb. Inspection of the hydrographs is therefore essential in conjunction with these statistical measures.

In this chapter the results are reported from trialling neural network models in a first pilot phase of the study. These experiments, which are organised into 4 case studies, explore a range of model inputs over different lead and travel times in order to determine which models provide useful forecasts.

5.3 Case Study 1: Early Experiments Using an MLP (Multilayer Perceptron) with Limited Data for the Upper Ping Catchment

This case study seeks to find the minimum number of input variables required to predict the water level at Chiang Mai with 6, 12, 18 and 24 hour lead times using station data at P1, P67 and P75 in the Upper Ping basin. The input data are hourly water levels for four years (2001-04). The forecasting model used in these experiments is a feedforward multilayer perceptron (MLP). The training algorithm is backpropagation implemented using a Levenberg-Marquardt (LM) approach, which is more efficient and faster than conventional backpropagation (Demuth *et al*., 2009). This type of training algorithm requires three datasets: one for training, one for validation in order to determine when to stop training, and one for independent testing to determine the real performance of the model. The 4 years of available data were then divided into 2 years for training, 1 year for validation and 1 year for testing. Figure 3.7 shows that flooding at P1 is the highest in 2001; therefore, 2001 was used in the testing dataset, 2002-2003 were employed for training and 2004 was used for validation. The neural network structure was set to 20 nodes in the hidden layer (after Demuth *et al*., 2009). The output layer contained 1 node to predict the water level at 6, 12, 18 or 24 hours ahead, where a separate neural network was trained for each lead time.

5.3.1 Input variables and input determination methodologies

The following 36 input variables were assembled: P1t, P1t-3, P1t-6, P1t-12, P1t-15, P1t-18, P1t-21, P1t-24, P67t, P67t-3, P67t-6, P67t-12, P67t-15, P67t-18, P67t-21, P67t-24, P75t, P75t-3, P75t-6, P75t-12, P75t-15, P75t-18, P75t-21, P75t-24 and the moving averages at P1, P67 and P75 for 6, 12 and 24 hours (i.e. MAP1_6, MAP1_12 and MAP1_24, etc.). Table 5.1 summarises the inputs chosen by the following six input determination techniques coded as:

- A: inclusion of all inputs
- C: inclusion of inputs if the correlation with the output is greater than 0.90
- S: inclusion of inputs determined by stepwise regression
- CS: inclusion of inputs when both the correlation is greater than 0.90 and the variables are selected by stepwise regression
- M: inclusion of inputs selected by running an M5 model tree algorithm
- D: inclusion of inputs selected by running a data mining algorithm
- G: inclusion of inputs selected by running a genetic algorithm.

Input	Input Determination Techniques and Lead Times																											
	1A	2A	3A	4A	1C	2C	3C	4C	1S	2S	3S	4S	1CS	2CS	3CS	4CS	1M	2M	3M	4M	1D	2D	3D	4D	1G	2G	3G	4G
P75t-24	X	X	X	X						X	X	X					X	X	X	X								
P75t-21	X	X	X	X													X	X	X	X								
P75t-18	X	X	X	X	X												X	X	X	X								
P75t-15	X	X	X	X	X												X	X	X	X								
P75t-12	X	X	X	X	X	X			X				X	X			X	X	X	X								
P75t-9	X	X	X	X	X	X				X		X					X	X		X								
P75t-6	X	X	X	X	X	X	X				X	X		X	X		X	X	X	X						X		X
P75t-3	X	X	X	X	X	X	X							X	X		X	X	X	X						X		X
P75t	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			X	X	X		X	X
MAP75_6	X	X	X	X	X	X	X			X		X	X				X	X	X	X		X					X	
MAP75_12	X	X	X	X	X	X	X				X			X	X		X	X	X		X						X	
MAP75_24	X	X	X	X	X	X			X	X		X	X	X			X	X	X									
P67t-24	X	X	X	X					X	X	X	X					X	X	X	X								
P67t-21	X	X	X	X					X	X	X	X						X	X	X								
P67t-18	X	X	X	X					X	X	X	X					X	X	X	X								
P67t-15	X	X	X	X	X												X	X		X								
P67t-12	X	X	X	X	X								X				X		X	X								
P67t-9	X	X	X	X	X	X			X		X	X	X				X	X		X								
P67t-6	X	X	X	X	X	X				X	X	X		X			X	X	X	X								
P67t-3	X	X	X	X	X	X	X										X	X	X	X								
P67t	X	X	X	X	X	X	X		X	X	X	X	X	X	X		X	X	X	X	X						X	X
MAP67_6	X	X	X	X	X	X	X			X	X	X		X	X		X									X	X	X
MAP67_12	X	X	X	X	X	X	X		X	X			X	X	X		X											
MAP67_24	X	X	X	X	X	X																			X			
P1t-24	X	X	X	X	X	X			X	X	X	X	X	X			X	X	X	X								X
P1t-21	X	X	X	X	X	X			X	X	X		X	X				X		X								
P1t-18	X	X	X	X	X	X	X			X	X	X		X			X	X	X	X								
P1t-15	X	X	X	X	X	X	X										X		X	X								X
P1t-12	X	X	X	X	X	X	X	X		X				X		X		X		X					X	X		
P1t-9	X	X	X	X	X	X	X	X									X	X	X						X	X	X	
P1t-6	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X		X								X	X
P1t-3	X	X	X	X	X	X	X	X									X							X			X	X
P1t	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				X	X	X	X	X	X	X	X
MAP1_6	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X								X	X	X	X	X
MAP1_12	X	X	X	X	X	X	X	X																	X	X	X	X
MAP1_24	X	X	X	X	X	X	X	X	X				X			X	X	X									X	
# of inputs	36				31	27	19	9	15	19	17	17	14	17	10	6	30	26	23	24	3	2	2	4	7	8	12	12

Table 5.1: Input variables selected by each method as indicated by an X. 1 is t+6, 2 is t+12, 3 is t+18 and 4 is t+24.

Models 1 to 4 in Table 5.1 relate to the 4 different lead times, i.e. 6, 12, 18 and 24 hours, respectively. SPSS was used to calculate the correlations and implement the stepwise regression. The WEKA software (Witten and Frank, 2005) was used to run the M5 model tree, the data mining procedure and the genetic algorithm.

Table 5.2 lists the total number of input variables selected by each input determination methodology. In these experiments, the data mining algorithm (1D to 4D) reduced the inputs the most heavily, with only 2 to 4 chosen for a given lead time. In contrast, the correlation (1C to 4C) and the M5 model tree technique (1M to 4M) retained the majority of the variables, making these potentially the least useful approaches as they did not produce very parsimonious models.

Table 5.2: The number of inputs selected by each input determination method for each lead time.

Model	1 (6hr)	2 (12hr)	3 (18hr)	4 (24hr)
A (All inputs included)	36	36	36	36
C (Correlation > 0.9)	31	27	19	9
S (Stepwise regression)	15	19	17	17
CS (Correlation > 0.9 and Stepwise)	14	17	10	6
M (M5 model tree)	30	26	23	24
D (Data mining)	3	2	2	4
G (Genetic algorithm)	7	8	12	12

5.3.2 Results

The experimentation process involved training neural network models using inputs chosen by each input determination method and for each lead time. This section of the case study shows summary tables of the model performance measured using RMSE, MAE, CE and PDIFF as outlined previously. In terms of the number of input variables chosen, a more parsimonious model will be helpful to the modeller as it requires less data and permits shorter neural network training times. However, more inputs may give a better hydrograph forecast. Therefore, model evaluation requires a thoughtful balance between the model development times and the forecasting accuracy. Selected hydrographs illustrating the model results are presented in the sections that follow. However, all of the hydrographs can be found in Appendix A.

6 hour lead time

Table 5.3 is an extraction from Table 5.1 which shows the variables chosen by the different input determination methods for just a lead time of 6 hours. It clearly shows that all the techniques selected P1 at time t, which shows the importance of this variable. Virtually every neural network paper that has ever been published in this area

includes the level or flow at the station where the predictions are being made at time t, confirming this input choice. The variables that were selected by the S and CS model were similar although the evaluation measures discussed next indicate that CS was better overall. Although the different methods chose different numbers of input variables, they all selected variables at the upstream stations and some moving average variables. This is not surprising as there are no rainfall drivers so the model must rely on upstream information.

Table 5.3: Input variables selected by each input determination technique (denoted by an X) for a 6 hour lead time.

Inputs	Input Determination Techniques							Inputs	Input Determination Techniques						
	A	C	S	CS	M	D	G		A	C	S	CS	M	D	G
P75t-24	X				X			P67t-24	X	X	X		X		
P75t-21	X				X			P67t-21	X		X				
P75t-18	X				X			P67t-18	X		X		X		
P75t-15	X	X			X			P67t-15	X	X			X		
P75t-12	X	X	X	X	X			P67t-12	X	X		X	X		
P75t-9	X	X			X			P67t-9	X	X	X	X	X		
P75t-6	X	X			X			P67t-6	X	X			X		
P75t-3	X	X			X			P67t-3	X	X			X		
P75t	X	X	X	X	X		X	P67t	X	X	X	X	X	X	
MAP75_6	X	X		X	X			MAP67_6	X	X			X		
MAP75_12	X	X			X	X		MAP67_12	X	X	X	X	X		
MAP75_24	X	X	X	X	X			MAP67_24	X	X					X
P1t-24	X	X	X	X	X										
P1t-21	X	X	X	X											
P1t-18	X	X			X										
P1t-15	X	X			X										
P1t-12	X	X					X								
P1t-9	X	X			X		X								
P1t-6	X	X	X	X	X										
P1t-3	X	X			X										
P1t	X	X	X	X	X	X	X								
MAP1_6	X	X	X	X			X								
MAP1_12	X	X					X								
MAP1_24	X	X	X	X	X										

From the hydrographs shown in Figure 5.1, it is clear that all the model forecasts have a delay and that they underestimate the peak stage. Both model S and CS, which selected fewer than 50% of the inputs, showed similar results to models A and C. In addition, model CS showed a better result than model C in terms of time to peak prediction (Table 5.4). Models A and M, in which almost all the inputs were selected, predicted peaks that lagged by only 1.5 hours compared to longer delays in peak prediction for the other methods (Table 5.4 and Figure 5.1). Model D predicted hydrographs with the greatest time delay (i.e. 6.5 hours) but this data mining method selected only 3 inputs: P67 at time t, a 12 hour moving average at P75 and P1 at time t (Table 5.3). However, it also produced an outstanding result at the peak as PDIFF was only 0.07 m, which is the smallest value when compared to the other models. At the

same time, this is the only model that had the highest overestimation on another storm event as shown in Figure A.1 in Appendix A.

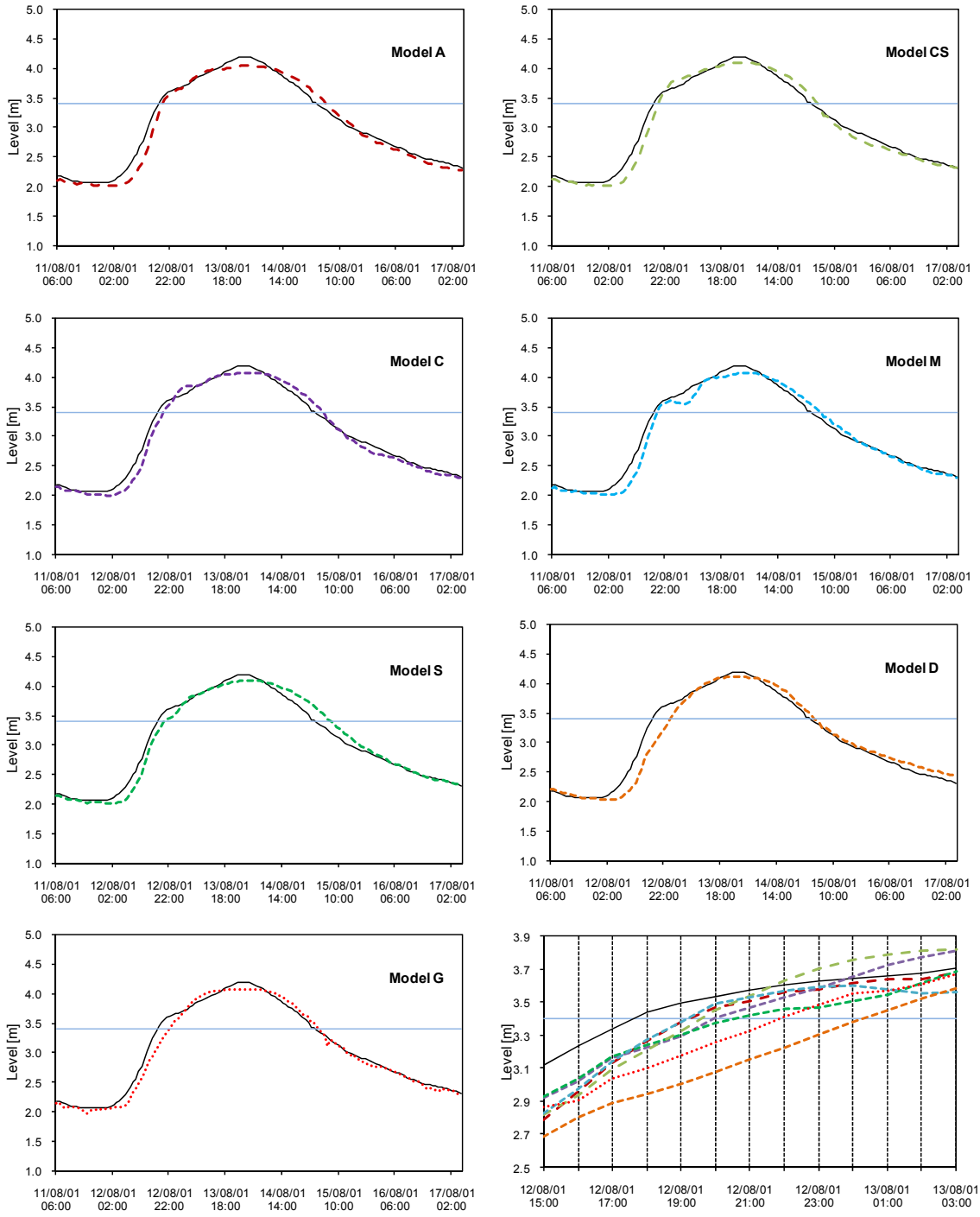


Figure 5.1: Observed and predicted water level at P1 station (Chiang Mai) from 11-17 Aug 2001 for a lead time of 6 hours.

Referring to the statistical results, Table 5.4 shows that model G gave the best result overall as measured by the CE, the RMSE and the MAE, with approximately 80% of the input variables eliminated. For the peak stage, model G predicted very close to the actual height (PDIFF = 0.11), but it had the second highest time delay of 4.5 hours for

the peak prediction. The model used only 7 inputs: one from P67, one from P75 and five from P1 (Table 5.3). Model M had the worst performance based on the CE, the RMSE and the MAE. However, the peak prediction and the time delay of 1.5 hours were good when compared with the other models.

Table 5.4: The number of inputs selected and the goodness of fit measures for a lead time of 6 hours.

Statistic	Input Determination Technique						
	A	**C**	**S**	**CS**	**M**	**D**	**G**
Number of inputs	36	31	15	14	30	3	7
PDIFF (m)	0.14	0.12	0.10	0.09	0.12	**0.07**	0.11
MAE (m)	0.0310	0.0212	0.0207	0.0230	0.0391	0.0219	**0.0163**
RMSE (m)	0.0452	0.0316	0.0343	0.0348	0.0577	0.0410	**0.0290**
CE	0.9845	0.9924	0.9910	0.9908	0.9747	0.9872	**0.9936**
Time delay (hr)	**1.5**	2.5	3	2	**1.5**	6.5	4.5

(Numbers in bold indicate the best performing model)

There are very small differences between the different input determination methodologies. The genetic algorithm (G) performed well according to most evaluation measures but the time delay is too large for such a lead time. Model CS appears to offer the best technique because the inputs are reduced by 50% whilst still retaining good model performance. This is similar to model S but the combination technique predicts both the time of the rising limb and the peak more effectively than the individual techniques alone. These initial runs also showed that input data from P67, P75 and P1 are required to produce the best model performance. However, different input determination methods produced different variable selections and there was no consistent pattern. It may be that this is a simple modelling problem that is virtually linear so the neural network is equally able to forecast 6 hours ahead based on a different combination of input variables. It is therefore concluded from these initial tests that the best performance for flood warning at a lead time of 6 hours was obtained by using the combined CS method to select the input variables.

12 hour lead time

The 12 hour ahead forecasts behaved in a similar manner to the 6 hour forecasts in terms of the number of variables chosen by the different input determination techniques. Once again, models S and CS reduced the number of inputs by approximately 50%, model D reduced the number of inputs by 90% and model G reduced the number of inputs by 80% as shown in Table 5.5.

Table 5.5: Input variables selected by each input determination technique (denoted by an X) for a 12 hour lead time.

Inputs	Input Determination Techniques							Inputs	Input Determination Techniques						
	A	C	S	CS	M	D	G		A	C	S	CS	M	D	G
P75t-24	X		X		X			P67t-24	X		X		X		
P75t-21	X				X			P67t-21	X		X		X		
P75t-18	X				X			P67t-18	X		X		X		
P75t-15	X				X			P67t-15	X				X		
P75t-12	X			X	X			P67t-12	X						
P75t-9	X	X	X		X			P67t-9	X	X			X		
P75t-6	X	X		X	X		X	P67t-6	X	X	X	X	X		
P75t-3	X	X		X	X		X	P67t-3	X	X			X		
P75t	X	X	X	X	X			P67t	X	X	X	X	X		
MAP75_6	X	X	X		X	X		MAP67_6	X	X	X	X			X
MAP75_12	X	X		X	X			MAP67_12	X	X	X	X			
MAP75_24	X	X	X	X	X			MAP67_24	X	X					
P1t-24	X	X	X	X	X										
P1t-21	X	X	X	X	X										
P1t-18	X	X	X	X	X										
P1t-15	X	X													
P1t-12	X	X	X	X	X		X								
P1t-9	X	X			X		X								
P1t-6	X	X	X	X											
P1t-3	X	X					X								
P1t	X	X	X	X		X	X								
MAP1_6	X	X	X	X											
MAP1_12	X	X					X								
MAP1_24	X	X			X										

Models A and CS provided the best time to peak prediction with a 3 hour delay as shown in Figure 5.2 and Table 5.6. Model CS provided the best overall RMSE, CE and delay in the peak as shown in Table 5.6 with a reasonable performance according to the other measures. Model G predicted very close to the actual peak stage (PDIFF=0.05) but there was a 4 hour delay in predicting the peak. Model G selected more input variables from P75 than P67, suggesting that the travel time between P67 to P1 is less than 12 hours. The 6 hour moving average at P67 was selected instead of P67 at time t. In addition, all of the models overestimated the lower peak event, and underestimated the highest peak as shown in Figure A.2 in Appendix A.

Table 5.6: The number of inputs selected and the goodness of fit measures for a lead time of 12 hours.

Statistic	Input Determination Technique						
	A	C	S	CS	M	D	G
Number of inputs	36	27	19	17	26	2	8
PDIFF (m)	0.09	0.07	**0.05**	0.12	0.12	0.19	**0.05**
MAE (m)	0.0272	0.0394	**0.0254**	0.0264	0.028	0.0685	0.0296
RMSE (m)	0.0492	0.0659	0.0521	**0.0489**	0.0515	0.0966	0.0547
CE	0.9817	0.9671	0.9794	**0.9819**	0.9798	0.9291	0.9773
Time delay (hr)	**3**	8	3.5	**3**	3.5	11	4

(Numbers in bold indicate the best performing model.)

Model M retained 26 variables (a 28% reduction) and give similar results to model C. Model CS was slimmed down to 17 variables (50%) with a predicted rising limb better than model C. It is clear that model D gave the worst performance (with an 11 hr time delay and the lowest peak stage prediction), probably because it is over-parsimonious in selecting just P1 at time t and the 12 hour moving average at P75 as shown in Table 5.5. This does not provide enough information for the neural network to learn and predict the unseen flood events. In contrast model A retained all the variables but did not substantially improve on model CS.

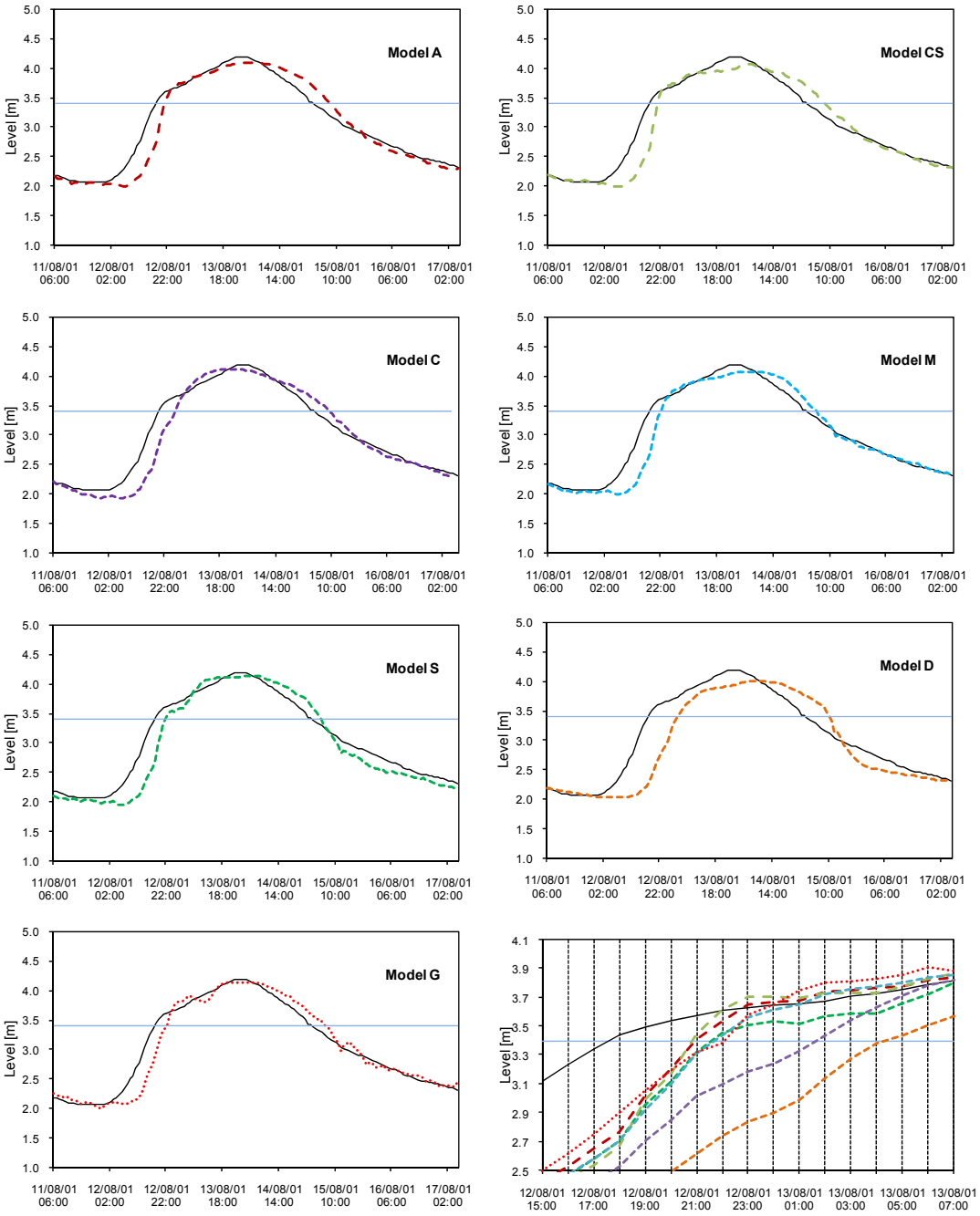


Figure 5.2: Observed and predicted water level at P1 station (Chiang Mai) for 11-17 Aug 2001 for a lead time of 12 hours.

While models G and S produced the most accurate peak forecast, model CS was still better in terms of most of the goodness-of-fit statistics and the time delay whilst eliminating approximately 50% of the variables. As with the previous 6 hour lead time, model CS provided the best input determination approach for neural network model development of flood prediction at a lead time of 12 hours.

18 hour lead time

At a lead time of 18 hours, the quality of the forecasts has clearly declined, as indicated by the hydrographs (Figure 5.3) and the statistics (Table 5.8). As before, model D was the most reductive approach, while model M was the least. This is the first experiment in which the input variable P75 at time t was selected by all 6 input determination techniques unlike in the models developed for 6 and 12 hour lead times (Table 5.7). Also model D selected only P1 and P75 at time t which might be a result of the travel time to P1 station from P75 being longer than from P67 (Patsinghasanee, 2004; Hydrology and Water Management Centre for Upper Northern Region, 2005).

Table 5.7: Input variables selected by each input determination technique (denoted by an X) for an 18 hour lead time.

Inputs	Input Determination Technique							Inputs	Input Determination Technique						
	A	**C**	**S**	**CS**	**M**	**D**	**G**		**A**	**C**	**S**	**CS**	**M**	**D**	**G**
P75t-24	X		X		X			P67t-24	X		X		X		
P75t-21	X				X			P67t-21	X		X		X		
P75t-18	X				X			P67t-18	X		X		X		
P75t-15	X				X			P67t-15	X						
P75t-12	X				X			P67t-12	X				X		
P75t-9	X							P67t-9	X		X				
P75t-6	X	X	X	X	X			P67t-6	X		X		X		
P75t-3	X	X		X	X			P67t-3	X	X			X		
P75t	X	X	X	X	X	X	X	P67t	X	X	X	X	X		X
MAP75_6	X	X			X		X	MAP67_6	X	X	X	X			X
MAP75_12	X	X	X	X	X		X	MAP67_12	X	X		X			
MAP75_24	X				X			MAP67_24	X						
P1t-24	X		X		X										
P1t-21	X		X												
P1t-18	X	X	X		X										
P1t-15	X	X			X										
P1t-12	X	X													
P1t-9	X	X			X		X								
P1t-6	X	X	X	X	X		X								
P1t-3	X	X					X								
P1t	X	X	X	X		X	X								
MAP1_6	X	X	X	X			X								
MAP1_12	X	X					X								
MAP1_24	X	X					X								

The summary provided in Table 5.8 suggests that model G gives the best overall performance overall according to all the goodness-of-fit statistics and had the smallest

delay in peak prediction of 7.5 hours. Model D, on the other hand, was the worst performer with only 2 inputs selected. Models CS and S were not as good as previous lead times although both models and model G correctly predicted the height of the maximum peak stage.

Table 5.8: The number of inputs selected and the goodness of fit measures for a lead time of 18 hours.

Statistic	Input Determination Technique						
	A	**C**	**S**	**CS**	**M**	**D**	**G**
Number of inputs	36	19	17	10	23	2	12
PDIFF (m)	0.09	0.03	**0**	**0**	0.2	0.34	**0**
MAE (m)	0.0571	0.055	0.07	0.0474	0.0527	0.0722	**0.0377**
RMSE (m)	0.091	0.1002	0.1853	0.0988	0.0916	0.1222	**0.081**
CE	0.9371	0.9238	0.7392	0.9259	0.9363	0.8866	**0.9502**
Time delay (hr)	8.5	17	14	14	9.5	14	**7.5**

(Numbers in bold indicate the best performing model.)

While these statistics are promising, visual inspection of the hydrograph shows that the model predictions are not useful in forecasting the sequence of events or mapping the time of rise (see Figure 5.3 and Appendix A, Figure A.3). Model G does get the peak stage correct but the shape of the hydrograph is not helpful. As expected, model D has the worst peak stage forecast, because of its minimal inputs, and possibly because this was the only model that did not select moving average variables. Model M also had poor peak prediction as only the moving averages at P75 were selected. Possibly the greater accuracy in forecasting the peak stage is allied to the inclusion of the moving average variables, a point that would benefit from further research but is outside the scope of this study.

Regarding the hydrographs in Figure 5.3, especially models CS, C, S and G, the forecasts exhibit an unusual pattern. They all rise late and then drop dramatically before rising again. Some also showed other strange behaviour by remaining at a high level for more than 24 hr rather than capturing the flood recession. This may be a function of incorrect inputs being chosen or a modelling problem that is too difficult at this lead time.

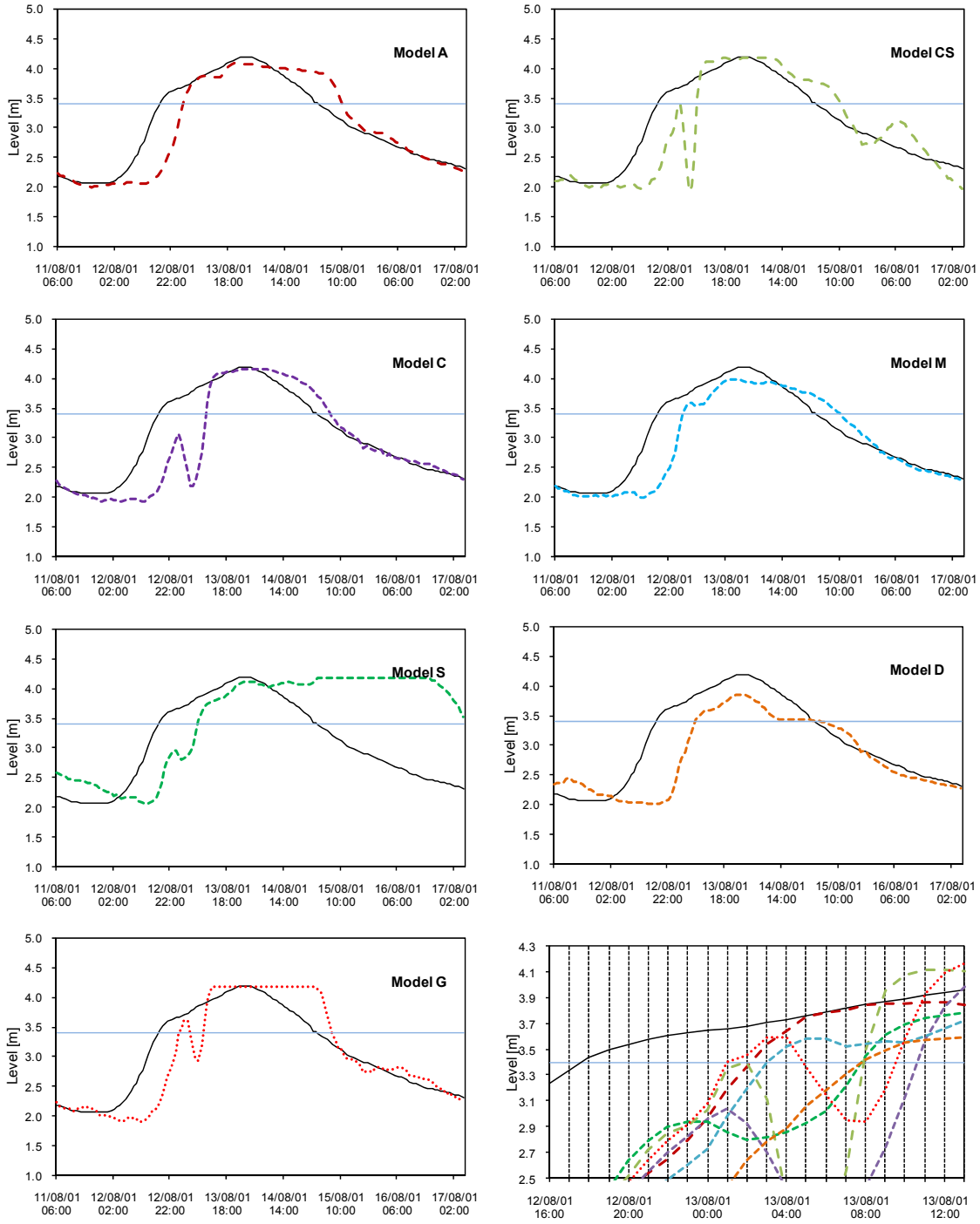


Figure 5.3: Observed and predicted water level at P1 (Chiang Mai) 11 -17 Aug 2001 at a lead time of 18 hours.

24 hour lead time

As with an 24 hour lead time, the input variable P75 at time t was selected by all the models, and the variable P67 at time t was selected less often when compared to the models at other lead times (Table 5.9). Model C, for example, did not select any variables from station P67. It is potentially unrealistic to forecast stage at P1 at a lead time of 24 hours from just two upstream stations (P75 and P67) because the travel time

of the flood wave from P67 to P1 is known to be less than 24 hours. The model was run to see at what point forecasting becomes impractical with this minimal, three site, input dataset.

Table 5.9: Input variables selected by each input determination technique (denoted by an X) for a 24 hour lead time.

Inputs	Input Determination Techniques							Inputs	Input Determination Techniques						
	A	**C**	**S**	**CS**	**M**	**D**	**G**		**A**	**C**	**S**	**CS**	**M**	**D**	**G**
P75t-24	X		X		X			P67t-24	X		X		X		
P75t-21	X				X			P67t-21	X		X		X		
P75t-18	X				X			P67t-18	X		X		X		
P75t-15	X				X			P67t-15	X				X		
P75t-12	X				X			P67t-12	X				X		
P75t-9	X		X		X			P67t-9	X		X		X		
P75t-6	X		X		X		X	P67t-6	X		X		X		
P75t-3	X				X		X	P67t-3	X				X		
P75t	X	X	X	X	X	X	X	P67t	X		X		X		X
MAP75_6	X		X		X			MAP67_6	X		X				X
MAP75_12	X							MAP67_12	X						
MAP75_24	X		X					MAP67_24	X						
P1t-24	X		X		X		X								
P1t-21	X				X										
P1t-18	X		X		X										
P1t-15	X				X		X								
P1t-12	X	X		X	X										
P1t-9	X	X													
P1t-6	X	X		X			X								
P1t-3	X	X				X	X								
P1t	X	X	X	X		X	X								
MAP1_6	X	X	X	X		X	X								
MAP1_12	X	X					X								
MAP1_24	X	X		X											

The statistics are included here for consistency in Table 5.10 but the hydrographs show the futility of working with them in detail. Model A is the least accurate according to the goodness-of-fit statistics while model G was the best. Model S predicted the correct peak stage followed by model D.

Table 5.10: The number of inputs selected and the goodness of fit measures for a lead time of 24 hours.

Statistic	Input Determination Technique						
	A	**C**	**S**	**CS**	**M**	**D**	**G**
Number of inputs	36	9	17	6	24	4	12
PDIFF (m)	0.24	0.03	**0**	0.23	0.06	0.01	0.13
MAE (m)	0.1003	**0.0583**	0.0657	0.0773	0.0616	0.0643	0.0668
RMSE (m)	0.1395	0.1196	0.1252	0.1334	0.1271	0.115	**0.1118**
CE	0.8523	0.8914	0.881	0.865	0.8773	0.8995	**0.9052**
Time delay (hr)	13	16.5	13	18	14	16	**12.5**

(Numbers in bold indicate the best performing model.)

All the models generated hydrographs with noise, especially overestimations of the small peaks (see Figure 5.4 and Appendix A, Figure A.4). The same unusual dip in the rising limb can be seen in Model M as occurred at a lead time of 18 hours.

All of these model results for a lead time of 24 hours are unsuitable for practical forecasting. It was however important to the research to define the point at which forecasting becomes impractical. The travel times are such that the input data from P75 and P67 are not able to provide enough information to give a reliable forecast at P1. For forecasts that can give a useful 24 hour flood warning, additional data will be required as discussed in more detail in Chapter 6.

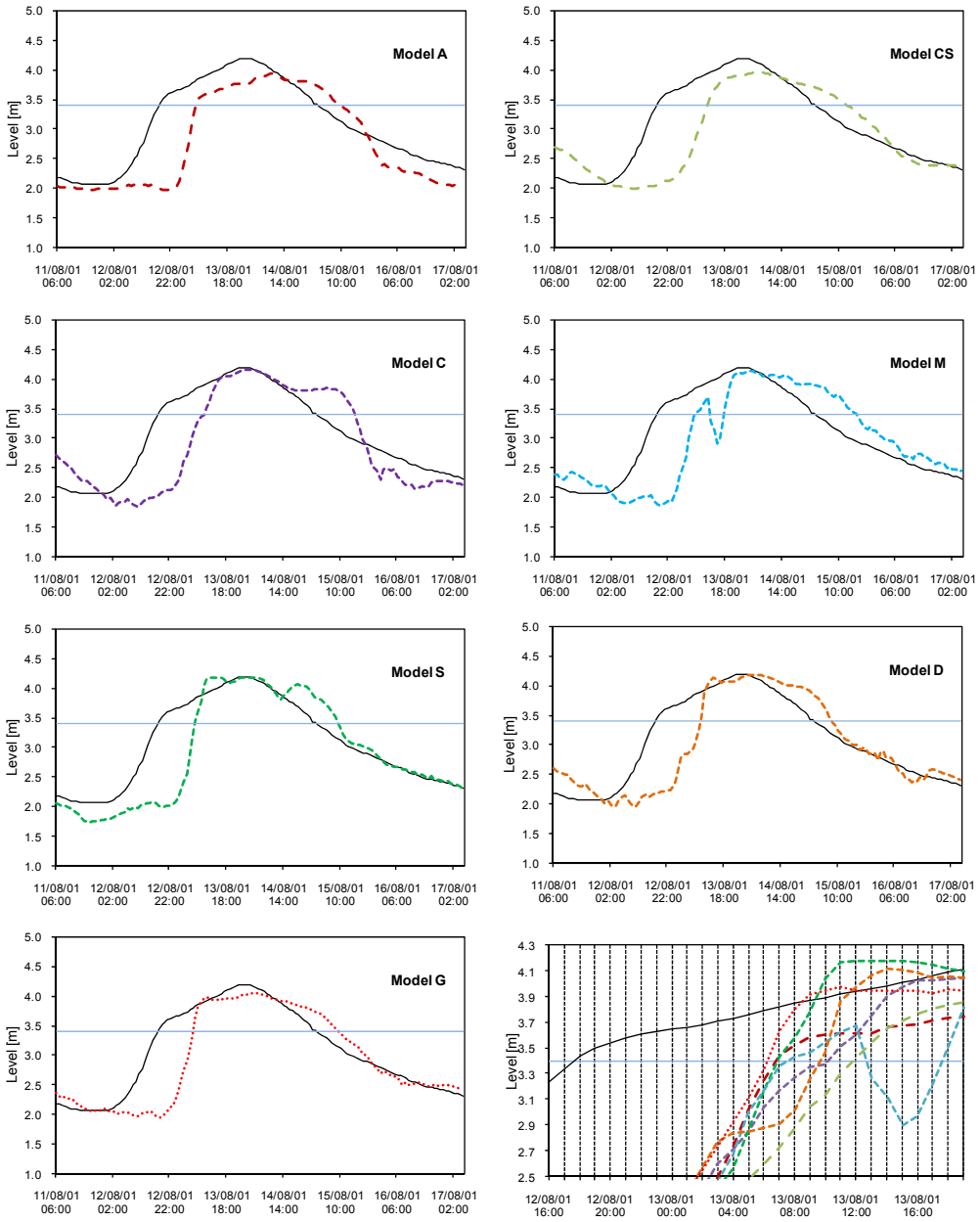


Figure 5.4: Observed and predicted water level at P1 (Chiang Mai) for 11-17 Aug 2001 for a lead time of 24 hours.

5.3.3 Summary of case study 1

Table 5.11 summarises the percentage of variables removed by each determination method for each lead time. From these initial pilot experiments it is concluded that the input determination methods of correlation and stepwise regression (CS), correlation greater than 0.90 (C) and the genetic algorithm (G) have value for this catchment and these particular conditions while the data mining technique (D) seems to produce the worst model for flood warning perhaps because it removes the majority of the variables. The CS and S methods remove more than half the variables whilst still maintaining good performance.

Table 5.11: The percentage of variables removed by each input determination technique by lead time.

Model	Lead time of 6 hours	Lead time of 12 hours	Lead time of 18 hours	Lead time of 24 hours
C	13.8	25.0	47.2	75.0
S	58.3	47.2	52.7	52.7
CS	61.1	52.7	72.2	83.3
M	16.6	27.7	36.1	33.3
D	91.6	94.4	94.4	88.8
G	80.5	77.7	66.6	66.6

It can be seen that the model forecasting 24 hrs ahead give a time delay of more than 10 hours. Therefore the overall model performance will only be evaluated for 6, 12 and 18 hr lead times. Comparative performance and ranking are shown in Table 5.12 based on the correct time to peak prediction. The top three scoring models are A, M, G and CS. For the peak stage prediction, model S seems to be the best choice followed by model G and C. In addition, based on CE, RMSE and MAE models G, CS, A, C and S are the top three scorers. In terms of the ability to reduce the number of inputs, the top three models are D, G and CS.

Model G seems to be the best method as it scored in the top three in all categories (i.e. time to peak prediction, PDIFF, CE, MAE, RMSE and the ability to reduce the number of inputs). The CS technique reduced the inputs by half and produced a slightly better forecast than either the stepwise regression (S) or correlation alone (C) but model S is the best model to predict peak stage. In contrast, the model M retained more input variables than other models but does not predict better results. The model that retains all variables does not produce the best results at the peak stage either, indicating that giving too many variables to the neural network is also not the best approach. Although model C has a very low score in terms of reducing the number of inputs and the correct

time to peak prediction in Table 5.12, it is a classic method that is commonly used. Therefore, in case study 2, only 4 techniques (i.e. C, CS, S and G) will be employed.

It should be noted that all 28 models over estimated the smaller peak and underestimated the higher stage in 2001 (Appendix A). This was most likely due to the fact that the neural network had not seen an event as large as that present in the testing data sets, i.e. these experiments were effectively forcing the network to extrapolate. Theoretically models perform better when they have events of a similar size in their training data sets. It is however a valuable hydrological test of a model to look at the accuracy of a forecast for an unobserved larger event because hydrologists seek to forecast extreme events. Later experiments will address different training data options and effects (Section 5.6).

Table 5.12: Ranking of model performance based on score (21 is the maximum score and 1 is the minimum score).

Statistic	Lead time	A	C	S	CS	M	D	G
Predict flood time	6 hours	21	19	18	20	21	14	15
	12 hours	18	12	17	18	17	9	16
	18 hours	11	7	8	8	10	8	13
	Total (**Rank**)	50 (**1**)	38 (**6**)	43 (**5**)	46 (**3**)	48 (**2**)	31 (**7**)	44 (**4**)
PDIFF	6 hours	12	14	16	17	14	18	15
	12 hours	17	18	19	13	14	11	19
	18 hours	17	20	21	21	10	9	21
	Total (**Rank**)	46 (**5**)	52 (**3**)	56 (**1**)	51 (**4**)	38 (**6**)	38 (**6**)	55 (**2**)
CE	6 hours	16	20	19	18	10	17	21
	12 hours	14	9	12	15	13	5	11
	18 hours	7	3	1	4	6	2	8
	Total (**Rank**)	37 (**2**)	32 (**3**)	32 (**3**)	37 (**2**)	29 (**4**)	24 (**5**)	40 (**1**)
Reduce input	6 hours		4	11	12	5	17	16
	12 hours	-	6	9	10	7	18	15
	18 hours		9	10	14	8	18	13
	Total (**Rank**)	(**7**)	19 (**6**)	30 (**4**)	36 (**3**)	20 (**5**)	53 (**1**)	44 (**2**)

From these experiments we can highlight the difficulty of forecasting 24 hours ahead, most likely because the travel time between station P67 and P1 is less than 24 hours. Models CS, C and D did not select any input variables from P67, and those chosen were not adequate to produce a valuable forecast. In practice this experiment, which was based on limited data, suggests that forecasting for P1 is possible with a maximum lead time of 18 hours.

5.4 Case Study 2: Forecasting with Bayesian Regularization (BR) for the Upper Ping Catchment

During the course of this research, the effectiveness of Bayesian algorithms in flood forecasting emerged (Demuth *et al*., 2009). The power of Bayesian techniques to forecast effectively where data sets are limited is why this technique is worth exploring for the Upper Ping. There are 36 input variables as per case study 1 and the Bayesian Regularization (BR) and Levenberg-Marquardt (LM) algorithm were run 50 times, taking the average of the 50 runs as the model forecast after Anctil and Lauzon (2004). New data for the year 2005 became available and were used for testing because it contains the biggest event in the record. Once again, the neural network models are being forced to extrapolate.

The input determination methods were applied to three different variations of the input dataset to determine if this affected the variables that were selected:

- **Dataset 1:** Storm events in 2001-2004 were used to select the inputs
- **Dataset 2:** Storm events in all the years (2001-2005) were used for selecting the inputs
- **Dataset 3:** The entire historical record for the period 2001-2004 was used for selecting the inputs, which is the same data set as that used in case study 1 (Table 5.1).

Once the inputs were selected by the input determination methods, data from only the main monsoon season were then used to develop and test the neural network models, in particular these periods:

- 2001, 2002, 2005 1/08 - 31/10
- 2003 1/09 - 31/10
- 2004 1/05 - 31/10

As the 24 hour ahead forecasts were shown to be of limited value for practical forecasting purposes, this section will only report on the models developed at lead times of 6, 12 and 18 hours ahead. As mentioned at the end of case study 1, only the most useful methods which emerged are reported here, i.e. correlation > 0.90 (C), correlation > 0.90 + stepwise regression (CS), stepwise regression (S) and the genetic algorithm (G).

Unlike the MLP trained with the Levenberg-Marquardt (LM) algorithm, the Bayesian Regularization (BR) algorithm does not require a validation dataset. Therefore the longer 2001-2004 dataset was used for training, and 2005 was used for testing.

However, training with LM requires a validation dataset so the training set was restricted to 2001 to 2003 and 2004 was used for validation. For these tests the number of hidden nodes was reduced from 20 to 10 because the initial experiments with 20 hidden nodes took a very long time to process, and the 20 node models showed no significant improvement in performance over the 10 node model results.

5.4.1 Selecting the input variables

Summary tables below show the input variables selected by the four different input determination methods for Dataset 1 (Table 5.13), Dataset 2 (Table 5.14) and Dataset 3 (Table 5.15).

Table 5.13: Input variables selected by each input determination technique (denoted by an X) for Dataset 1 (i.e. flood events in 2001-2004).

Input variables	Input Determination Techniques											
	6 hour lead time				12 hour lead time				18 hour lead time			
	C	CS	S	G	C	CS	S	G	C	CS	S	G
P67t-24											X	
P67t-21												
P67t-18												
P67t-15	X	X	X				X					
P67t-12	X											
P67t-9	X				X	X	X					
P67t-6	X				X	X						X
P67t-3	X				X				X			
P67t	X	X	X	X	X	X	X	X	X	X	X	X
MAP67_6	X				X	X	X		X	X	X	X
MAP67_12	X	X	X	X	X				X			X
MAP67_24	X				X							
P75t-24			X								X	
P75t-21							X					
P75t-18												
P75t-15												
P75t-12			X									
P75t-9											X	
P75t-6												
P75t-3			X					X				X
P75t			X	X			X	X			X	X
MAP75_6							X				X	X
MAP75_12	X	X										
MAP75_24												
P1t-24				X			X				X	
P1t-21	X										X	
P1t-18	X							X				X
P1t-15	X				X							
P1t-12	X			X	X							
P1t-9	X				X			X	X			X
P1t-6	X	X	X	X	X	X		X	X			X
P1t-3	X			X	X			X	X			X
P1t	X	X	X	X	X	X	X	X	X	X	X	X
MAP1_6	X	X	X	X	X	X	X	X	X	X	X	X
MAP1_12	X		X	X	X		X		X		X	X
MAP1_24	X				X				X	X		
Total (36)	**21**	**7**	**11**	**10**	**16**	**7**	**11**	**9**	**11**	**5**	**12**	**14**

As can be seen from Table 5.13, it is clear that for Dataset 1 (i.e. taking only flood events 2001-2004), model C and CS did not select variables at P75 except for the moving average variable selected by model C at a lead time of 6 hours.

Table 5.14: Input variables selected by each input determination technique (denoted by an X) for Dataset 2 (i.e. flood events in 2001-2005).

Input variables	Input Determination Techniques											
	6 hour lead time				12 hour lead time				18 hour lead time			
	C	CS	S	G	C	CS	S	G	C	CS	S	G
P67t-24												
P67t-21												
P67t-18												
P67t-15	X											
P67t-12	X	X	X									
P67t-9	X				X							
P67t-6	X				X	X					X	
P67t-3	X				X				X			
P67t	X	X	X	X	X	X	X	X	X	X	X	X
MAP67_6	X	X	X		X	X	X	X	X	X	X	
MAP67_12	X				X							X
MAP67_24	X											
P75t-24			X				X				X	
P75t-21												
P75t-18												
P75t-15			X				X				X	
P75t-12	X		X									
P75t-9	X			X								
P75t-6	X				X	X	X				X	
P75t-3	X			X	X			X				X
P75t	X	X	X	X	X	X	X	X	X	X	X	X
MAP75_6	X	X	X	X	X		X				X	
MAP75_12	X				X	X	X				X	
MAP75_24	X							X				
P1t-24												
P1t-21				X								
P1t-18	X										X	
P1t-15	X										X	
P1t-12	X				X			X				
P1t-9	X				X							X
P1t-6	X	X	X	X	X	X	X	X	X	X		X
P1t-3	X			X	X				X	X	X	X
P1t	X	X	X	X	X	X	X	X	X	X	X	X
MAP1_6	X	X	X	X	X	X	X	X	X			
MAP1_12	X	X	X	X	X			X	X	X		
MAP1_24	X			X	X						X	
Total (36)	**27**	**9**	**12**	**12**	**19**	**9**	**11**	**10**	**9**	**7**	**14**	**8**

In contrast, variables from P75 were selected by all the techniques for Datasets 2 and 3 (see Tables 5.14 and 5.15). The variable P1 at time t remains the important variable for all three datasets as it is selected by all the models. Moreover, all models selected P67 at time t for both Datasets 1 and 2 while this was only true of P75 at time t for Dataset 2. In the case of Dataset 1, only models S and G selected it. To disentangle this issue requires more information about the location of rainfall across the catchment.

It may be because there was rainfall in the P75 drainage basin in 2005, and limited or no rainfall over the P75 station over the storm period between 2001 and 2004. However, this is speculation at this stage.

Table 5.15: Input variables selected by each input determination technique (denoted by an X) for Dataset 3 (i.e. all historical level data in 2001-2004).

Input variables	Input Determination Techniques											
	6 hour lead time				12 hour lead time				18 hour lead time			
	C	CS	S	G	C	CS	S	G	C	CS	S	G
P67t-24			X				X				X	
P67t-21			X				X				X	
P67t-18			X				X				X	
P67t-15	X											
P67t-12	X	X										
P67t-9	X	X	X		X						X	
P67t-6	X				X	X	X				X	
P67t-3	X				X				X			
P67t	X	X	X		X	X	X		X	X	X	X
MAP67_6	X				X	X	X	X	X	X	X	X
MAP67_12	X	X	X		X	X	X		X	X		
MAP67_24	X			X	X							
P75t-24							X				X	
P75t-21												
P75t-18	X											
P75t-15	X											
P75t-12	X	X	X		X	X						
P75t-9	X				X		X					
P75t-6	X				X	X		X	X	X	X	
P75t-3	X				X	X		X	X	X		
P75t	X	X	X	X	X	X	X		X	X	X	X
MAP75_6	X	X			X		X		X			X
MAP75_12	X				X	X			X	X	X	X
MAP75_24	X	X	X		X	X	X					
P1t-24	X	X	X		X	X	X				X	
P1t-21	X	X	X		X	X	X				X	
P1t-18	X				X	X	X		X		X	
P1t-15	X				X				X			
P1t-12	X			X	X	X	X	X	X			
P1t-9	X			X	X			X	X			X
P1t-6	X	X	X		X	X	X		X	X	X	X
P1t-3	X				X			X	X			X
P1t	X	X	X	X	X	X	X	X	X	X	X	X
MAP1_6	X	X	X	X	X	X	X		X	X	X	X
MAP1_12	X			X	X			X	X			X
MAP1_24	X	X	X		X				X			X
Total (36)	**31**	**14**	**15**	**7**	**27**	**17**	**19**	**8**	**19**	**10**	**17**	**12**

5.4.2 Comparison of the LM and BR training algorithms

The first item for discussion in this case study is a comparison of the results when using the training algorithm from case study 1 (i.e. Levenberg-Marquardt – LM) compared with the Bayesian Regularization algorithm (BR) at the three different lead times.

6 hour lead time

Figure 5.5 contains the plots of actual and predicted stage for three storm events in 2005 (artificially linked together) and Table 5.16 provides the evaluation measures. These results clearly show that the overall performance of the BR algorithm was better than NNs trained with LM as the MAE, RMSE and CE are all better. Visual inspection of the hydrographs shows that models developed with the BR algorithm all have more accurate peak stage and rising time results (see also Appendix B for the full range).

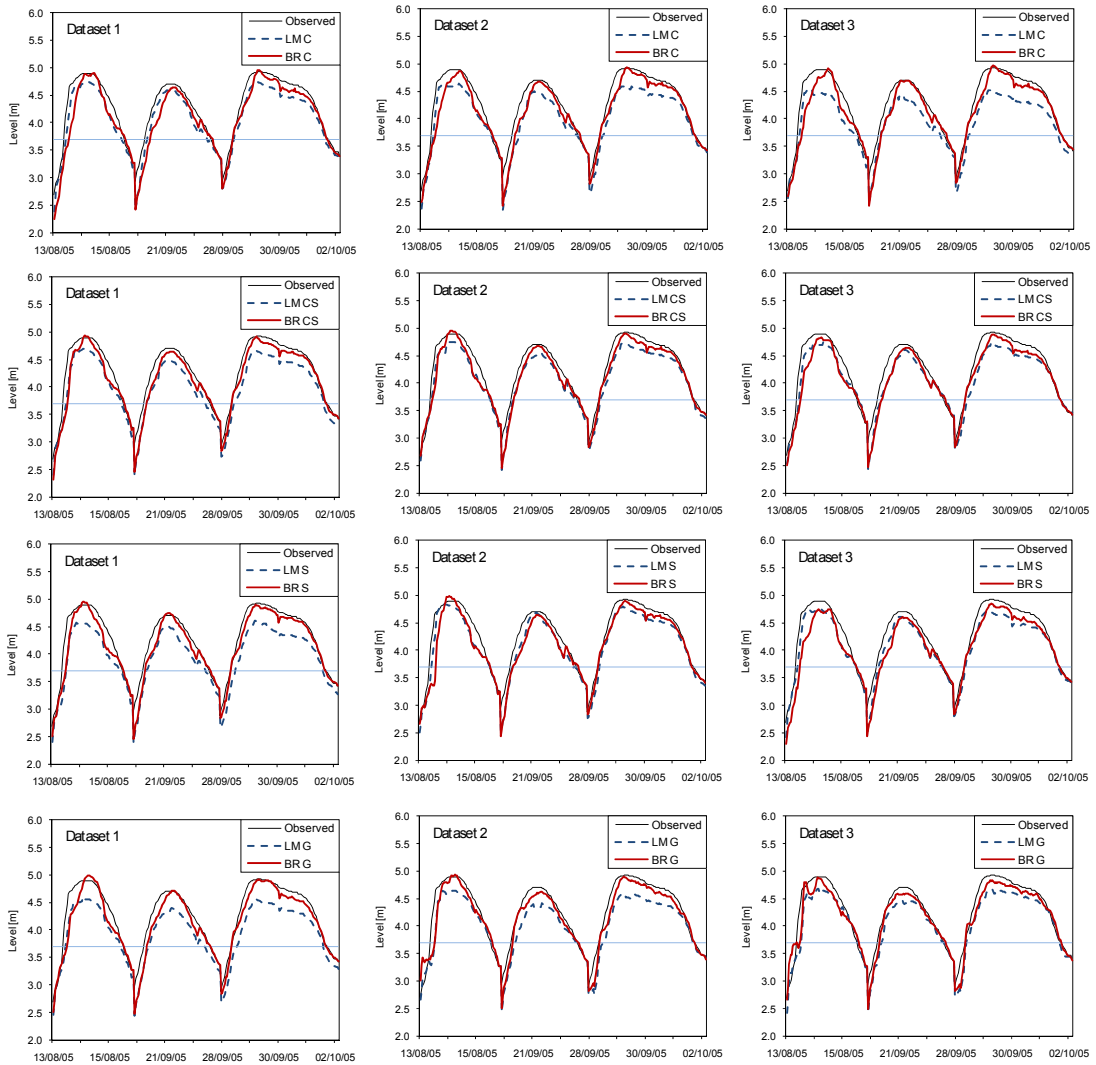


Figure 5.5: Comparison of the BR and LM algorithms all three peaks (1 Aug -10 Oct 2005), t+6 hr.

Table 5.16: Goodness of fit statistics for all models at a lead time of 6 hours

Stat	Alg	C			CS			S			G		
		Dataset			Dataset			Dataset			Dataset		
		1	2	3	1	2	3	1	2	3	1	2	3
MAE	BR	0.042	0.038	0.038	0.036	0.035	0.039	0.035	0.038	0.047	0.036	0.036	0.038
	LM	0.061	0.079	0.091	0.075	0.053	0.047	0.086	0.051	0.045	0.080	0.084	0.064
RMSE	BR	0.091	0.081	0.080	0.074	0.070	0.081	0.072	0.086	0.093	0.076	0.072	0.073
	LM	0.094	0.119	0.135	0.115	0.087	0.087	0.139	0.083	0.081	0.135	0.123	0.101
CE	BR	0.989	0.991	0.992	0.992	0.993	0.991	0.993	0.990	0.988	0.992	0.993	0.993
	LM	0.988	0.981	0.975	0.982	0.990	0.990	0.974	0.991	0.991	0.975	0.980	0.986

12 hour lead time

The evaluation measures are listed in Table 5.17 and the hydrographs are displayed in Figure 5.6. Visual inspection shows that the BR algorithm performs better than the LM algorithm as with the previous lead time, and the statistical measures once again demonstrate the superiority of the BR algorithm over LM.

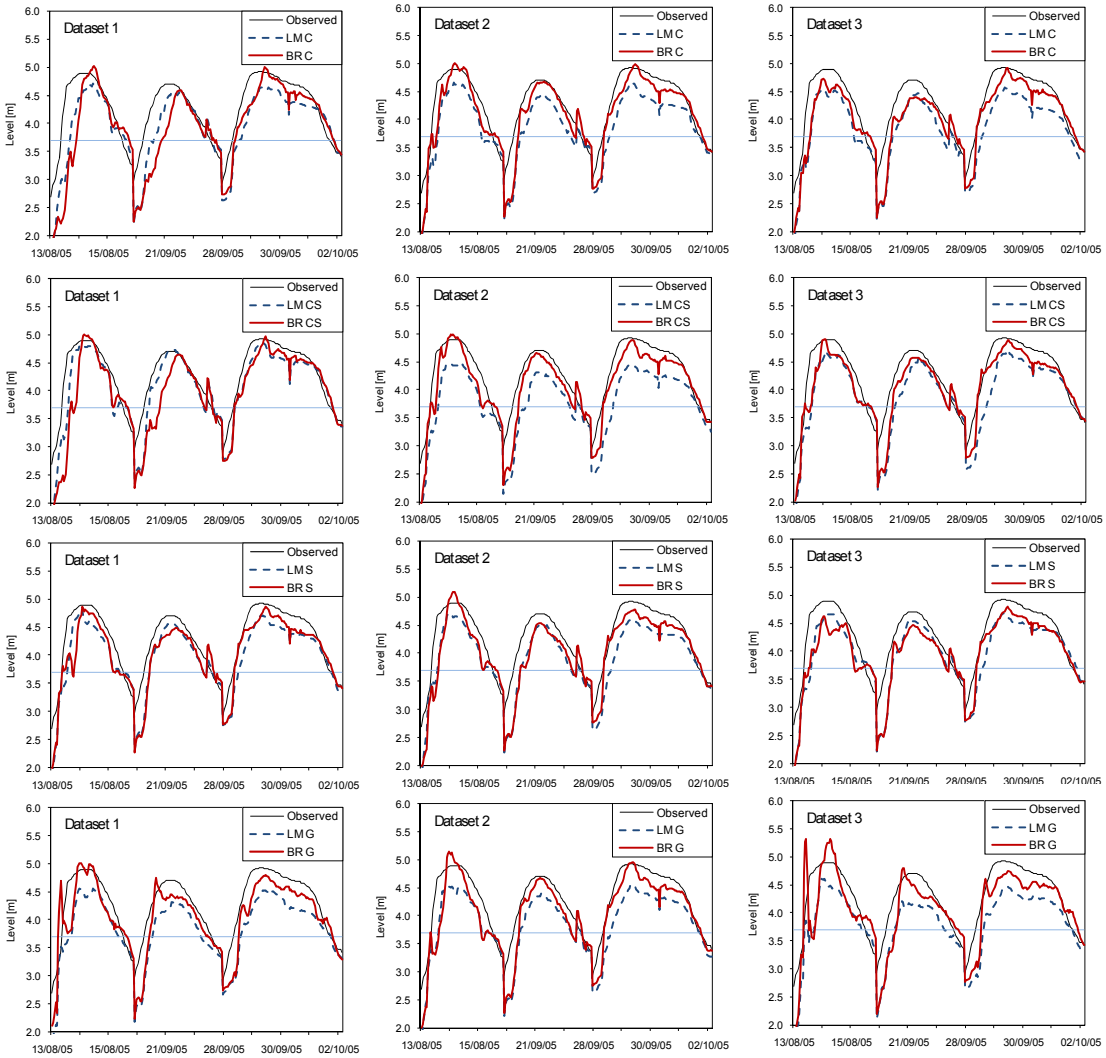


Figure 5.6: Comparison of the BR and LM algorithms all three peaks (1 Aug -10 Oct 2005), t+12 hr.

Table 5.17: Goodness of fit statistics for all models at a lead time of 12 hours.

Stat	Alg	C			CS			S			G		
		Dataset			Dataset			Dataset			Dataset		
		1	2	3	1	2	3	1	2	3	1	2	3
MAE	BR	0.089	0.063	0.073	0.081	0.065	0.069	0.071	0.070	0.079	0.079	0.066	0.086
	LM	0.112	0.097	0.100	0.067	0.173	0.119	0.176	0.103	0.100	0.119	0.109	0.119
RMSE	BR	0.214	0.125	0.141	0.186	0.126	0.131	0.140	0.142	0.151	0.151	0.137	0.166
	LM	0.191	0.173	0.179	0.136	0.246	0.190	0.142	0.170	0.166	0.195	0.181	0.199
CE	BR	0.937	0.979	0.973	0.952	0.978	0.976	0.973	0.972	0.969	0.969	0.974	0.962
	LM	0.950	0.959	0.956	0.975	0.917	0.951	0.972	0.961	0.962	0.948	0.955	0.946

18 hour lead time

At a lead time of 18 hours, the outcome of these model tests is less clear cut and the model is reaching its limits. In summary, the models trained with BR generally overpredict the peaks, while the models trained with LM underestimate the peaks for all three datasets (Figure 5.7 and Appendix B). Table 5.18 indicates that the models trained with LM for Dataset 1 gave the best forecasts for a lead time of 18 hours, followed by Datasets 2 and 3. However, the models trained with BR still gave good results with Datasets 2 and 3.

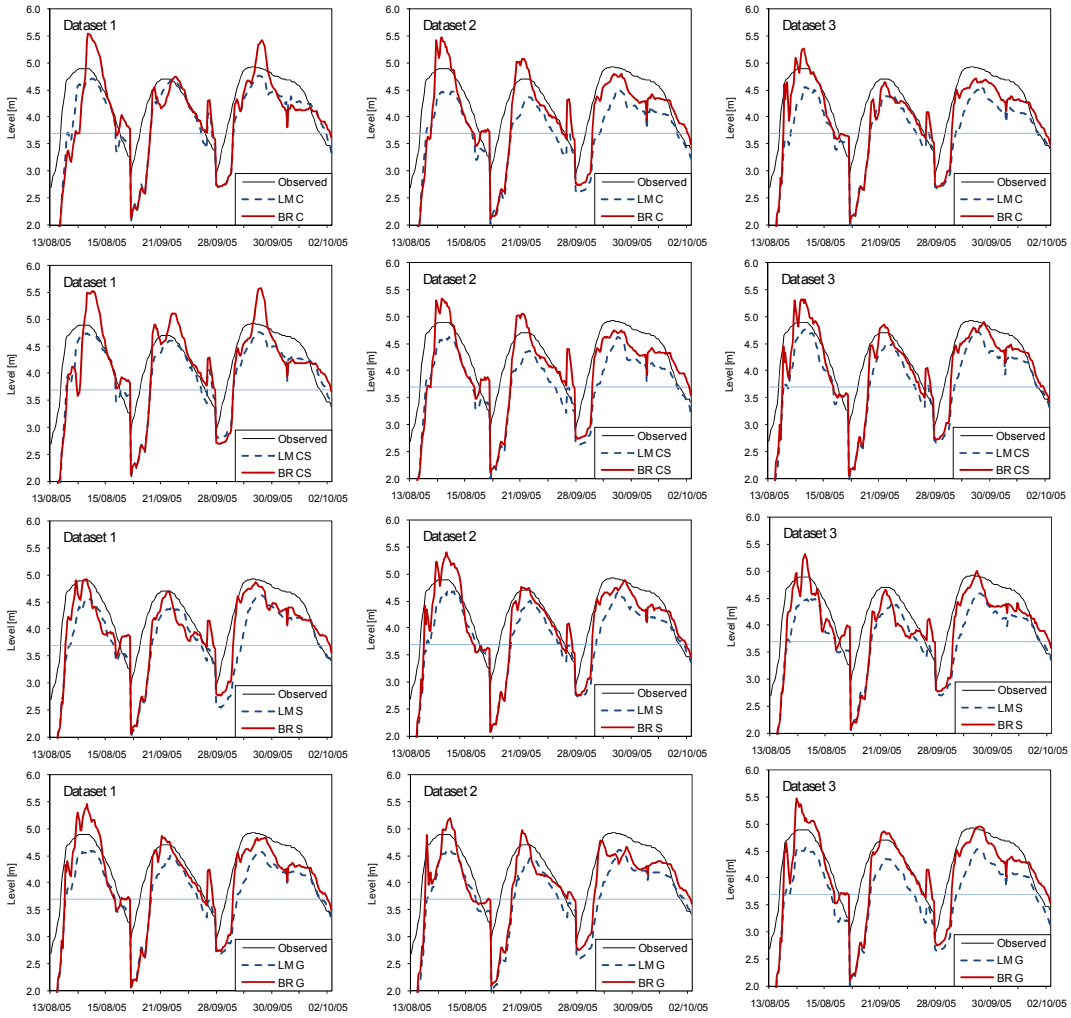


Figure 5.7: Comparison of the BR and LM algorithms all three peaks (1 Aug -10 Oct 2005), t+18 hr.

Table 5.18: Goodness of fit statistics for all models at a lead time of 18 hours.

Stat	Alg	C			CS			S			G		
		Dataset			Dataset			Dataset			Dataset		
		1	2	3	1	2	3	1	2	3	1	2	3
MAE	BR	0.127	0.120	0.105	0.123	0.122	0.272	0.125	0.109	0.125	0.109	0.120	0.108
	LM	0.113	0.137	0.136	0.116	0.137	0.264	0.140	0.115	0.125	0.119	0.142	0.139
RMSE	BR	0.247	0.235	0.203	0.237	0.231	0.489	0.226	0.203	0.227	0.205	0.217	0.202
	LM	0.228	0.264	0.250	0.222	0.257	0.486	0.256	0.225	0.239	0.227	0.258	0.252
CE	BR	0.917	0.925	0.943	0.923	0.927	0.674	0.930	0.944	0.930	0.942	0.936	0.944
	LM	0.930	0.905	0.914	0.932	0.910	0.677	0.910	0.931	0.922	0.930	0.910	0.913

The outcome of this work suggests that at each forecast lead time, the BR algorithm is more accurate at forecasting the stage at P1 compared to using the LM algorithm. The BR algorithm will therefore be used in future neural network model development.

5.4.3 Comparing model performance of the BR algorithm with different input datasets for different lead times

Having concluded that the BR algorithm produces a better performing model for the Upper Ping River compared to the LM algorithm, this section examines the effect of using only individual storm events to select the inputs (Datasets 1 and 2) versus input selection using data from a complete year (Dataset 3). Only results for the BR algorithm are reported.

6 hour lead time

Figure 5.8 provides hydrographs for the models developed using the BR algorithm for three flood events in the testing dataset. Appendix B provides the full set. Table 5.19 provides the evaluation measures for each dataset and input determination method. According to the hydrographs in Figure 5.8, all the models make reasonable forecasts of the actual peak stage in the three large flood events. As might be anticipated, there are small overestimations in the peak and time delays in some models.

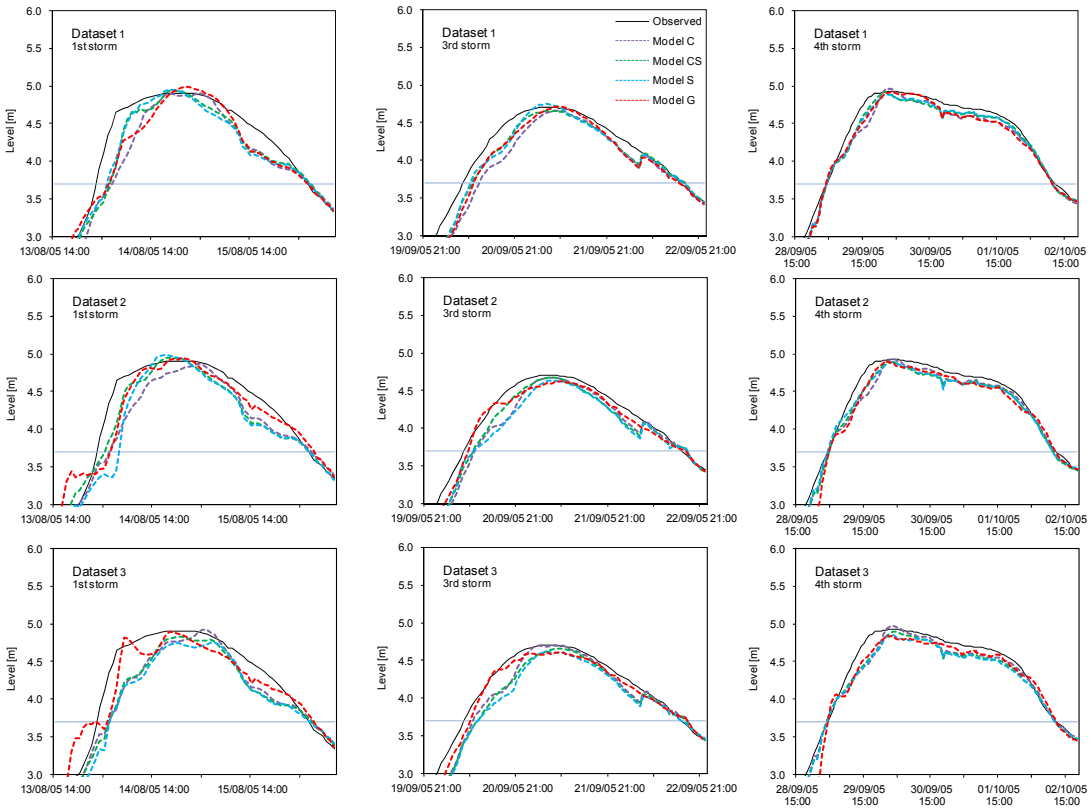


Figure 5.8: Comparison of the BR algorithms with different models of three datasets at t+6 hr, three peaks (13 Aug - 2 Oct 2005).

Table 5.19: Statistical comparison of the all three datasets with all methodologies for a lead time of 6 hours.

Statistic	Dataset 1				Dataset 2				Dataset 3			
	C	CS	S	G	C	CS	S	G	C	CS	S	G
MAE	0.042	0.036	**0.035**	0.036	0.038	**0.035**	0.038	0.036	**0.038**	0.039	0.047	**0.038**
RMSE	0.091	0.074	**0.072**	0.076	0.081	**0.070**	0.086	0.072	0.080	0.081	0.093	**0.073**
CE	0.989	0.992	**0.993**	0.992	0.991	**0.993**	0.990	**0.993**	0.992	0.991	0.988	**0.993**

(Bold indicates the best statistic value)

The goodness-of-fit statistics in Table 5.19 indicate that selecting inputs using Dataset 1 seems to produce better results than Dataset 3. Therefore, selecting input variables using information from storm events seems to produce better results than using data from the whole year. However, using Dataset 2 produces the lowest RMSE and MAE and better hydrograph predictions (as shown in Figure 5.8). Unfortunately, Dataset 2 includes data from the year 2005, which is also in the testing data set. Normally this would not be a permissible modelling operation but it was undertaken to examine the effect of including the largest storms on record in the determination of the inputs. It is clear that inclusion does improve the performance of the resulting neural network models.

The goodness of fit statistics and hydrographs between the three datasets show that model S produced the best result for Dataset 1, model CS produced the best model overall for Dataset 2 while model G gave the best result for Dataset 3, in particular for the first and third peak (Table 5.19 and Figure 5.8).

12 hour lead time

Figure 5.9 provides the hydrographs for three of the four storm events in 2005 and Table 5.21 lists the goodness-of-fit measures. The full set of hydrographs can be found in Appendix B. As can be seen from the hydrographs in Figure 5.9 (and Appendix B), the models where inputs were selected from Dataset 1 tend to give the most delayed forecasts, particularly for the third storm event. All models in all three datasets overestimated the first peak but underestimated the third and fourth peak. There is no correlation greater than 0.90 between the output and variables at P75 in Dataset 1. As a result, model C and CS did not select all the P75 variables (Table 5.13). This may be because no rain fell over P75 during the storm events between 2001 and 2004. Moreover, the hydrographs in Figure 5.9 show that model C and CS predicted the fourth storm similarly to other models but was very poor at predicting the first and third storm events, with a delay of approximately 12 and 18 hr respectively. Perhaps rain fell only over P75 during the first and third storm in 2005.

Model G seems to be able to predict the onset of the flood early on Datasets 1 and 3. However, it is also the most delayed in terms of peak prediction for the third and fourth storm but only for Dataset 3. This may be due to the fact that model G did not select P67 and P75 at time t while model G for Datasets 1 and 2 did. There may simply be a lack of information needed to predict the time of rise of the hydrograph. The same is true of model C and CS, which did not select P75 at time t for Dataset 1 and predicted a late time to rise. Therefore, the variables P67 and P75 at time t appear to be crucial input variables that influence the model performance.

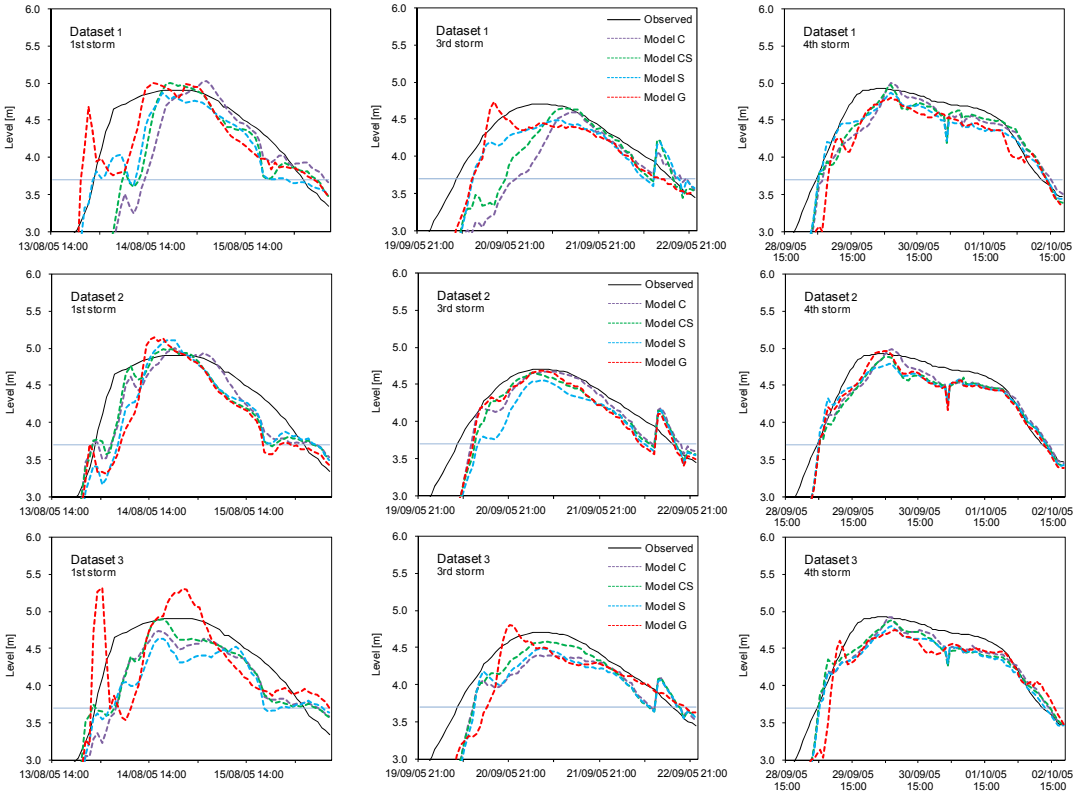


Figure 5.9: Comparison of the BR algorithms with different models of three datasets at t+12 hr, three peaks (13 Aug - 2 Oct 2005).

Table 5.20: Statistical comparison of the all three datasets with all methodologies for a lead time of 12 hours.

Statistic	Dataset 1				Dataset 2				Dataset 3			
	C	CS	S	G	C	CS	S	G	C	CS	S	G
MAE	0.089	0.081	**0.071**	0.079	**0.063**	0.065	0.070	0.066	0.073	**0.069**	0.079	0.086
RMSE	0.214	0.186	**0.140**	0.151	**0.125**	0.126	0.142	0.137	0.141	**0.131**	0.151	0.166
CE	0.937	0.952	**0.973**	0.969	**0.979**	0.978	0.972	0.974	0.973	**0.976**	0.969	0.962

(Bold indicates the best statistic value)

According to Table 5.20, model S gave the best results using Dataset 1 similar to the 6 hour lead time while model C gave the best performance for Dataset 2. Model CS gave the best result for Dataset 3. Once again models developed using Dataset 2 and model C were the best overall but this was followed by Dataset 3 and Dataset 1. This time the

use of the full year of data for selecting the inputs did appear to improve the results although they are generally quite similar between Datasets 1 and 3.

18 hour lead time

As might be anticipated from the experiments reported previously, the hydrograph predictions show more noise than at previous, shorter lead times as shown in Figure 5.10. There is greater overprediction of the first peaks than with the shorter forecasting times for all three storm events. The first storm on the far left of Figure 5.10 shows a late prediction of the rising limb followed by a drop before rising again. This pattern was seen in case study 1 at this lead time. However, this pattern does not appear as strongly in the next two storms. This may be due to the condition of the catchment, i.e. it may be much wetter once the later storms arrived. All models over predicted the first and third peak when developed using Dataset 1 except for model S, which selected input variables at P75 (t-9 and t-24) and P1 (t-24). In addition, only model C and CS selected a 24 hour moving average from P1 for Dataset 1.

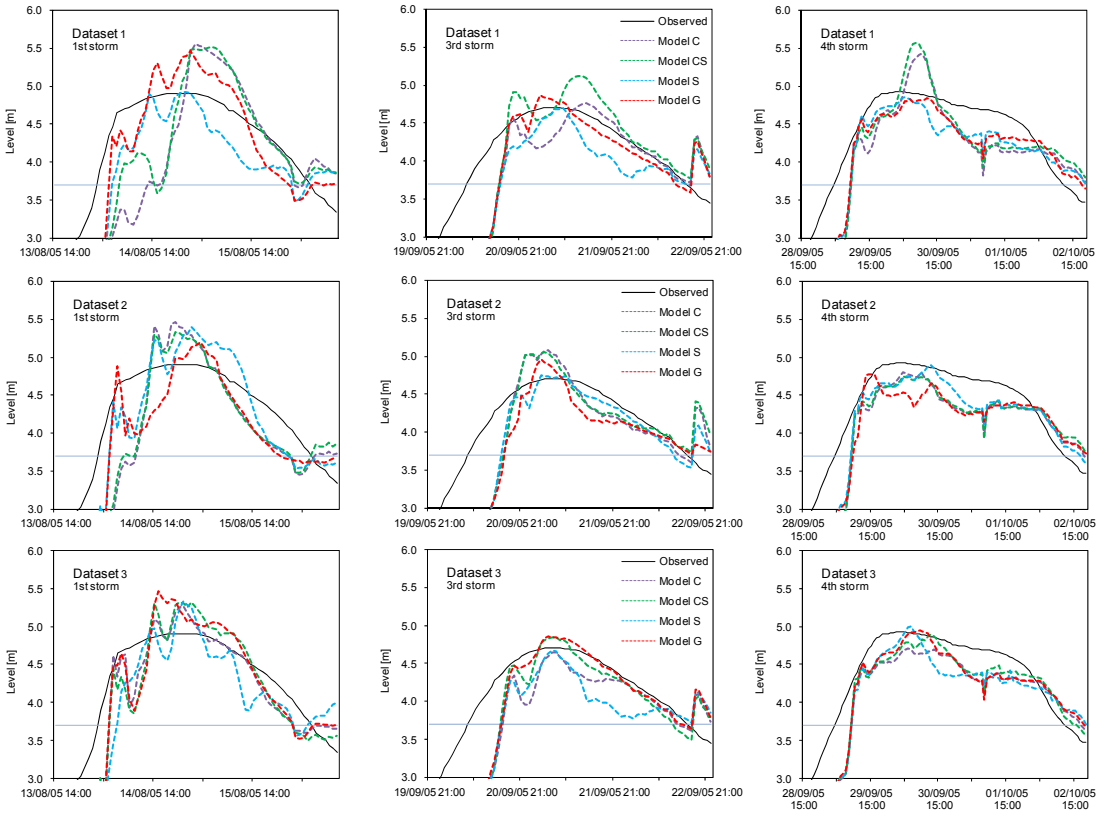


Figure 5.10: Comparison of the BR algorithms with different models of three datasets at t+18 hr for three different storm events (from 13 Aug - 2 Oct 2005).

Table 5.21 contains the goodness-of-fit statistics and suggests that the best model performance for a lead time of 18 hours for Datasets 1 and 3 is model G while for Dataset 2 it is model S. Models C and G provide good results when selecting input

variables from Dataset 3. Models CS and S gave the best result for Dataset 2. Stepwise regression always produces the best result when selecting input variables from Dataset 1 for a lead time of 6 and 12 hours but performs the best for Dataset 2 and a lead time of 18 hours. Statistically Dataset 3 seems to be the best dataset for selecting inputs to predict at a lead time of 18 hours as the goodness of fit measures are the highest.

Table 5.21: Statistical comparison of the all three datasets with all methodologies for a lead time of 18 hours.

Statistic	Dataset 1				Dataset 2				Dataset 3			
	C	CS	S	G	C	CS	S	G	C	CS	S	G
MAE	0.127	0.123	0.125	**0.109**	0.120	0.122	**0.109**	0.120	**0.105**	0.272	0.125	0.108
RMSE	0.247	0.237	0.226	**0.205**	0.235	0.231	**0.203**	0.217	0.203	0.489	0.227	**0.202**
CE	0.917	0.923	0.930	**0.942**	0.925	0.927	**0.944**	0.936	0.943	0.674	0.930	**0.944**

(Bold indicates the best statistic value)

5.4.4 Summary of case study 2

The Bayesian regularization (BR) algorithm clearly proved to have an overall advantage over the Levenberg-Marquardt (LM) algorithm on its own in terms of forecasting accuracy. When the modelling was based on events rather than annual data, the model performance was generally better or was not much different in terms of goodness-of-fit. Therefore, using the smaller storm event data set is recommended because it reduces the size of the dataset, reduces the number of low stage data points and decreases the time to train the neural network. Table 5.22 shows the percentage of variables removed by the different input determination techniques, which is similar to that found in case study 1.

Table 5.22: The percentage of variables removed by the input determination techniques.

Model	Dataset 1			Dataset 2			Dataset 3		
	t+6	t+12	t+18	t+6	t+12	t+18	t+6	t+12	t+18
C	41.6	55.5	69.4	25.0	47.2	75.0	13.8	25.0	47.2
CS	80.5	80.5	86.1	75.0	75.0	80.5	61.1	52.7	72.2
S	69.4	69.4	66.6	66.6	69.4	61.1	58.3	47.2	52.7
G	72.2	75.0	61.1	66.6	72.2	77.0	80.5	77.7	66.6

It can be seen in Table 5.23 that overall there is no difference between selecting inputs from the entire year or using data from only storm events. However, models developed using correlation as the input determination technique did better when presented with data from a full year. The stepwise regression technique produced the best model when selecting input variables from only the storm events. However, model performance was improved by including the 2005 data or including the testing data for selecting the input variables. Normally this would not be good practice but it does

illustrate that having the largest event on record in the dataset used for input determination would improve the model results.

Table 5.23: Summary of best BR model performance for Datasets 1, 2 and 3.

Models	Comparing all three datasets				Comparing only dataset 1 and 3			
	C	**CS**	**S**	**G**	**C**	**CS**	**S**	**G**
t+6								
CE	3	2	1	2, 3	3	1	1	3
RMSE (m)	3	2	1	2	3	1	1	3
t+12								
CE	2	2	1	2	3	3	1	1
RMSE (m)	2	2	1	2	3	3	1	1
t+18								
CE	3	2	2	3	3	1	3	3
RMSE (m)	3	2	2	3	3	1	3	3

(Number 1, 2 and 3 denote as dataset 1, 2 and 3 respectively.)

Table 5.24 represents a ranking of model performance based on only RMSE, MAE, CE and the ability to reduce the number of inputs because PDIFF is not available when more than one peak is present in the testing data. Stepwise regression is the best method for selecting inputs from Dataset 1, correlation and stepwise regression is the best for Dataset 2, while model G is the best input determination method. Moreover, model C seems to be the best method for selecting inputs from Dataset 3. It can also be concluded that model S is a suitable method for selecting inputs from storm event data.

Table 5.24: Ranking of model performance (1 is the best and 4 is the worst).

Statistic	Lead time	Dataset 1				Dataset 2				Dataset 3			
		C	**CS**	**S**	**G**	**C**	**CS**	**S**	**G**	**C**	**CS**	**S**	**G**
RMSE	6 hours	9	11	12	10	10	12	9	11	11	10	9	12
	12 hours	4	6	8	7	8	7	5	6	7	8	6	5
	18 hours	1	2	3	5	1	2	4	3	3	1	2	4
	Total (**Rank**)	14 (**4**)	19 (**3**)	23 (**1**)	22 (**2**)	19 (**3**)	21 (**1**)	18 (**4**)	20 (**2**)	21 (**1**)	19 (**2**)	17 (**3**)	21 (**1**)
MAE	6 hours	10	11	12	11	10	12	10	11	12	11	10	12
	12 hours	6	7	9	8	9	8	6	7	8	9	7	6
	18 hours	2	4	3	5	4	3	5	4	5	2	3	4
	Total (**Rank**)	18 (**3**)	22 (**2**)	24 (**1**)	24 (**1**)	23 (**1**)	23 (**1**)	21 (**3**)	22 (**2**)	25 (**1**)	22 (**2**)	20 (**3**)	22 (**2**)
CE	6 hours	10	11	12	11	11	12	10	12	11	10	9	12
	12 hours	5	7	9	8	9	8	6	7	6	8	5	4
	18 hours	2	3	4	6	2	3	5	4	2	7	1	3
	Total (**Rank**)	17 (**4**)	21 (**3**)	25 (**1**)	24 (**2**)	22 (**2**)	23 (**1**)	21 (**3**)	23 (**1**)	19 (**2**)	25 (**1**)	15 (**3**)	19 (**2**)

Reduce input	6 hours 12 hours 18 hours	3 4 8	11 11 12	8 7 6	9 10 5	4 5 10	10 10 12	7 8 6	7 9 11	3 4 5	8 6 10	7 5 6	12 11 9
	Total (**Rank**)	15 (**4**)	34 (**1**)	21 (**3**)	24 (**2**)	19 (**4**)	32 (**1**)	21 (**3**)	27 (**2**)	12 (**4**)	24 (**2**)	18 (**3**)	32 (**1**)

In the next section pruning algorithms are investigated to see whether this method performs better as an input determination method than the other simple ones tried out so far.

5.5 Case study 3: Inclusion of a Pruning Algorithm

In this set of experiments the opportunity is taken to evaluate a neural network pruning algorithm to see how this compares with the previous approaches. These experiments focus on testing the quality of the stage forecasts when there are different numbers of nodes in the hidden layer. Six experiments were undertaken as follows to select the input variables:

1. Pruning with 5 hidden nodes, with all remaining inputs (5A)
2. Pruning with 5 hidden nodes, where the remaining inputs have more than one connections to the hidden node (5B)
3. Pruning with 10 hidden nodes, with all remaining inputs (10A)
4. Pruning with 10 hidden nodes, where the remaining inputs have more than one connections to the hidden node (10B)
5. Pruning with 15 hidden nodes, with all remaining inputs (15A)
6. Pruning with 15 hidden nodes where the remaining inputs have more than one connections to the hidden node (15B)

The testing and training datasets used in this case study are the same as for case study 2, i.e. training on 2001 to 2004 and testing on 2005. Only storm events were used in both input determination and model development using the same input variables as in the previous two case studies.

Table 5.25 shows the input variables that remain after pruning with 5, 10 and 15 hidden nodes. Pruning with 5 hidden nodes for a lead time of 6 and 12 hours removed approximately 80% of the variables while for a lead time of 18 hours ahead, only 8% were removed. A much larger number of variables remained when pruning with 10 and 15 hidden nodes for all three lead times.

Table 5.25: Input variable selected by Pruning algorithms.

Input Variable	Lead time of 6 hours						Lead time of 12 hours						Lead time of 18 hours					
	5 A	5 B	10 A	10 B	15 A	15 B	5 A	5 B	10 A	10 B	15 A	15 B	5 A	5 B	10 A	10 B	15 A	15 B
P67t-24					X				X	X			X	X	X		X	X
P67t-21					X	X			X		X	X	X	X			X	X
P67t-18			X								X		X	X	X			
P67t-15	X		X	X	X	X					X	X	X		X	X	X	X
P67t-12					X	X	X	X	X	X	X		X	X	X	X	X	X
P67t-9			X	X	X	X	X	X	X	X	X		X	X			X	X
P67t-6					X	X	X	X	X	X	X	X	X	X	X	X		
P67t-3			X	X	X	X			X	X	X	X	X	X			X	X
P67t			X	X	X				X	X	X				X	X	X	
MAP67_6					X	X			X	X	X	X	X	X	X		X	X
MAP67_12	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	
MAP67_24					X	X			X	X	X		X	X	X	X	X	
P75t-24					X	X			X	X	X		X	X	X			
P75t-21									X	X	X	X	X	X	X	X	X	
P75t-18									X	X	X	X	X	X	X	X	X	
P75t-15			X	X	X				X	X	X		X	X	X	X	X	X
P75t-12					X	X			X	X			X	X	X	X	X	X
P75t-9			X	X	X				X	X	X	X	X	X	X	X		
P75t-6					X	X	X		X								X	X
P75t-3					X	X			X	X	X	X	X	X	X	X	X	
P75t					X	X			X	X			X		X	X	X	X
MAP75_6					X	X			X	X	X	X	X	X	X	X	X	
MAP75_12			X	X	X	X			X	X	X	X	X		X	X	X	X
MAP75_24	X		X		X	X	X	X	X	X	X		X	X	X		X	
P1t-24			X		X				X	X			X	X				
P1t-21			X		X	X	X	X	X	X	X	X	X		X	X	X	
P1t-18	X				X	X	X	X	X	X	X	X	X	X	X	X	X	X
P1t-15					X	X			X	X	X				X		X	X
P1t-12	X		X	X	X				X	X	X	X	X	X	X	X	X	
P1t-9			X		X	X			X	X	X	X	X	X	X	X	X	X
P1t-6			X	X					X	X	X	X	X		X	X	X	
P1t-3			X	X	X	X			X				X	X				
MAP1_6			X		X	X			X	X	X		X	X			X	X
MAP1_12			X		X	X	X	X	X	X	X	X	X		X	X	X	
MAP1_24			X	X	X	X			X	X	X	X	X	X	X	X		
P1t	X		X	X	X	X			X	X	X	X	X	X	X	X	X	
Total (36)	**6**	**1**	**20**	**13**	**32**	**20**	**9**	**8**	**34**	**31**	**30**	**20**	**33**	**26**	**29**	**23**	**29**	**16**

(X indicates the variables selected for the model after pruning with 5, 10 and 15 hidden nodes)

Neural network models were then developed for the three lead times using the input variables from the six pruning operations using the BR algorithm and 50 runs averaged as before.

5.5.1 Results

6 hour lead time

Figure 5.11 shows the results of the neural network models developed using the inputs selected by the six pruning methods. The complete set of hydrographs is provided in Appendix C. The hydrographs suggest that forecasts with a lead time of 6 hours are best produced with model 5A. The worst result is for model 5B so allowing multiple input connections to the hidden nodes is not a good feature. Also it retained only one input variable, which would be insufficient information to predict the correct output. More variables were also removed with 5 hidden nodes compared to 10 and 15. Thus,

increasing the number of hidden nodes results in a higher number of variables retained by this method. In contrast, model 10B showed better performance than model 10A, i.e. the peak was underestimated less and there was less delay in the rising limb. Table 5.26 provides goodness-of-fit measures that confirm this result. Models 15A and B predicted similar hydrographs but the statistical measures indicate that model 15A is better than 15B.

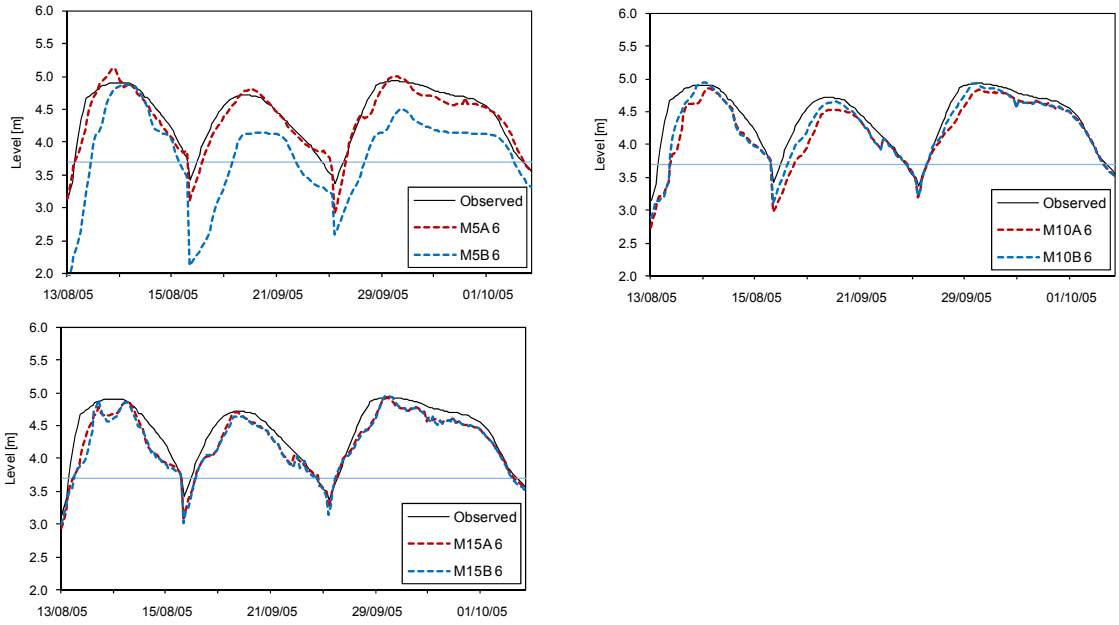


Figure 5.11: Results of the six pruning experiments for three storm events from 13 Aug-1 Oct 2005 for a 6 hour lead time.

Table 5.26: Goodness-of-fit statistics for the six pruning models at a lead time of 6 hours.

Statistic	5 hidden nodes		10 hidden nodes		15 hidden nodes	
	A	B	A	B	A	B
MAE (m)	0.0431	0.223	0.0465	0.038	0.0425	0.0453
RMSE (m)	0.083	0.3393	0.1122	0.0903	0.0833	0.0915
CE	0.9905	0.8415	0.9827	0.9888	0.9904	0.9885

Six inputs were selected by model 5A. These same six inputs were also selected for models 10A and 15A but the results (in terms of hydrographs and goodness-of-fit statistics) are not as good. It seems likely that both models 10A and 15A selected too many input variables, which could have reduced the overall performance. In addition, model 5A selected the moving average variables at P67 and P75, which was also the case with the best model produced in case study 2 for a 6 hour lead time, confirming the importance of these variables.

12 hour lead time

The hydrograph results for the six pruning experiments are shown in Figure 5.12. All six models show similar patterns, i.e. underestimation of the peak and a delay in the time of rise. Table 5.27 contains the goodness-of-fit statistics. Examining the statistics shows that model 15B seems to be the best model, followed by models 15A and 10A. Therefore, allowing inputs to have more than one link to a hidden node may improve the overall results. Model 5B shows an overestimation of the peak for the first storm event in Figure 5.12. Model 5B, in contrast to model 5A, did not select any variables from P75, which may explain this observed result.

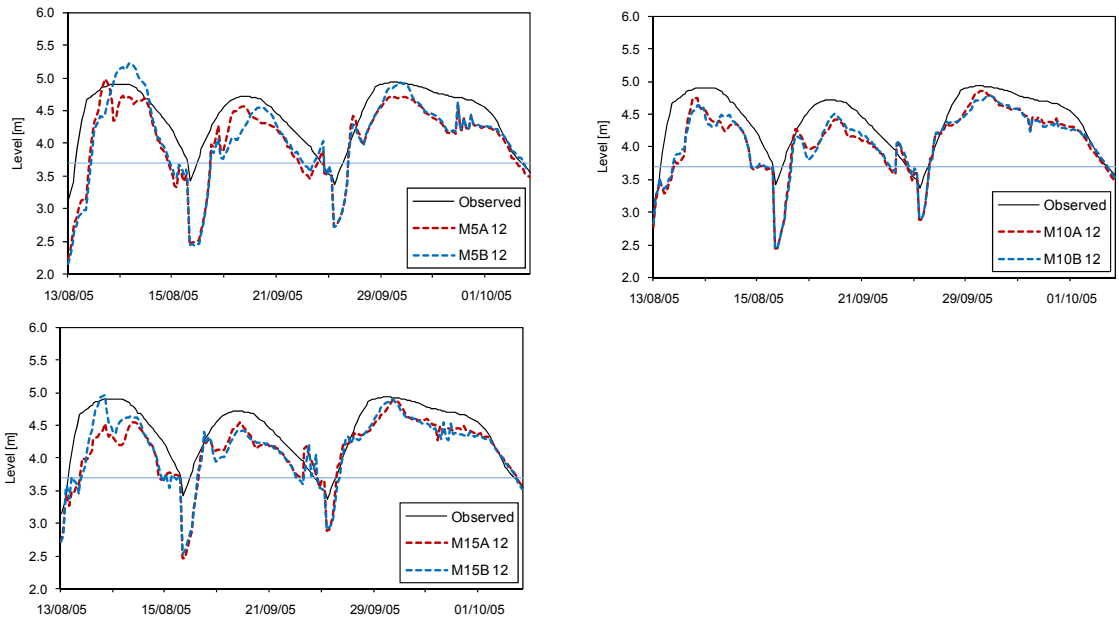


Figure 5.12: Results of the six pruning experiments for three storm events from 13 Aug-1 Oct 2005 for a 12 hour lead time.

Table 5.27: Goodness-of-fit statistics for the six pruning models at a lead time of 12 hours.

Statistic	5 hidden nodes		10 hidden nodes		15 hidden nodes	
	A	B	A	B	A	B
MAE (m)	0.098	0.0977	0.0839	0.0878	0.081	0.0764
RMSE (m)	0.1986	0.2034	0.1637	0.1648	0.1584	0.1506
CE	0.9458	0.9431	0.9632	0.9626	0.9655	0.9688

18 hour lead time

Figure 5.13 contains the model predictions for an 18 hour lead time while Table 5.28 contains the goodness-of-fit statistics. Figure 5.13 shows that all the models predict the rising limb late with a large overestimation of the first storm event. From the hydrographs it appears as if model 15B once again produces the best results although the statistics in Table 5.28 indicate that model 10B is slightly better.

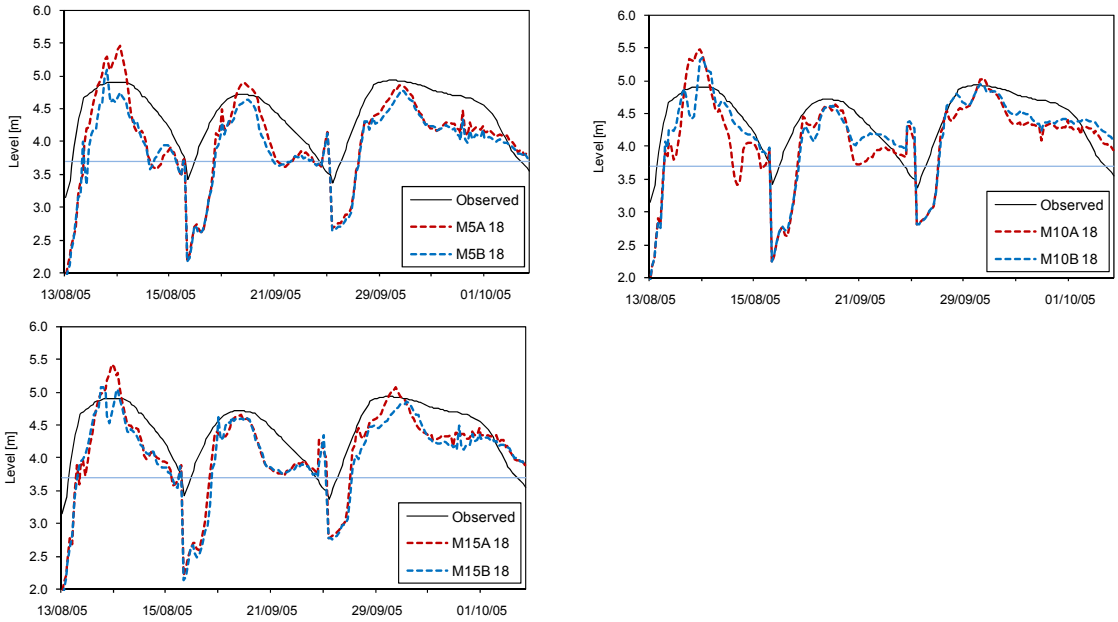


Figure 5.13: Results of the six pruning experiments for three storm events from 13 Aug-1 Oct 2005 for an 18 hour lead time.

Table 5.28: Goodness-of-fit statistics for the six pruning models at a lead time of 18 hours.

Statistic	5 hidden nodes		10 hidden nodes		15 hidden nodes	
	A	B	A	B	A	B
MAE (m)	0.1256	0.1377	0.1257	0.1163	0.122	0.1256
RMSE (m)	0.2376	0.2528	0.2311	0.2169	0.2268	0.237
CE	0.9227	0.9124	0.9268	0.9355	0.9295	0.923

5.5.2 Summary of case study 3

This case study has experimented with pruning algorithms to select the input variables. Table 5.29 shows the percentage of variables removed by the different experiments. Using 5 hidden nodes in the pruning increases the number of variables selected as the lead time increases, which may simply be a reflection of increasing complexity in the modelling problem. However, using 15 hidden nodes shows the opposite result while a 10 hidden node model shows a less clear pattern as the lead time increases.

Table 5.29: The percentage of variables removed by the three numbers of different hidden nodes (Model A).

Number of hidden nodes	t+6	t+12	t+18
5	83.3	75.0	8.0
10	44.4	5.5	19.4
15	11.1	16.6	19.4

Different results were also obtained depending on whether option A or B was used. These experiments suggest that future forecasting models could be more effectively and efficiently performed by pruning with 5 hidden nodes at a lead time of 6 hours and

15 hidden nodes at longer lead times. In the next case study, pruning algorithms will be used with a 15 node model.

5.6 Case Study 4: Adding Rainfall Data to the Models of the Ping Catchment

This case study continues to examine the efficiency of the forecasting techniques explored in case studies 1 to 3 by adding in data from one rain gauge located at Chiang Mai, which was obtained relatively late in the research process. This gauge is adjacent to P1 and cannot be considered as representative of such a large catchment, but its relative value was thought worthy of exploration. At this time the 2006 and 2007 stage data also became available.

In the following experiments, input determination methods C and S are used. Method C is retained as it is commonly used by many other studies while model S showed the best performance for Dataset 1. The pruning method was used, referred to as model Pr, as well as another more complicated method called partial mutual information (PMI), referred to as model P in this case study. The case study was trained and tested with the BR algorithm and run 50 times for each experiment to predict 6, 9, 12, 15 and 18 hours ahead. As previously, the 50 runs were averaged to produce a final forecast.

As with the previous case study, only data from the monsoon periods were used for both input variable determination and neural network model development. The length of each data set and the start dates were determined by the monsoon onset dates, which change between seasons:

- 2001, 2004, 2007 2/08 – 31/10
- 2002 2/08 – 9/11
- 2003 1/09 – 31/10
- 2005 2/08 – 4/11
- 2006 2/07 – 31/10

The training dataset included the years 2001 to 2005 while 2006 and 2007 were used for testing.

5.6.1 Input variables

The modelling undertaken in this section employed data from the three water level stations (P1, P67 and P75) and 1 rain gauge station, giving 54 input variables in total. As this case study uses more data than Case Study 3, pruning with 15 hidden nodes

was originally implemented. However, this took an incredibly large amount of time to complete, making this an infeasible solution, which is a serious issue in practical forecasting. As 10 hidden nodes still produced a satisfactory performance in Case Study 3, this number was used in this case study instead of 15.

Table 5.30: Input variables selected by each input determination technique (denoted by an X) for various lead times.

Input Variables	Input Determination Techniques																			
	Lead time of 6 hours				Lead time of 9 hours				Lead time of 12 hours				Lead time of 15 hours				Lead time of 18 hours			
	C	S	P	Pr	C	S	P	Pr	C	S	P	Pr	C	S	P	Pr	C	S	P	Pr
P75L-24				X				X				X		X						X
P75L-21				X				X		X		X		X		X				X
P75L-18								X												
P75L-15		X		X				X						X						
P75L-12	X	X						X						X						
P75L-9	X				X			X		X				X				X		
P75L-6	X			X	X			X	X											
P75L-3	X	X			X	X		X	X	X			X	X				X		X
P75L	X	X			X	X		X	X	X		X	X	X		X	X	X		X
MV6P75				X		X		X		X		X				X				X
MV12P75								X												
MV18P75		X						X												
MV24P75				X		X		X						X						
MV30P75																				X
MV36P75						X		X						X		X				X
P67L-24				X		X		X												X
P67L-21																		X		X
P67L-18								X								X		X		X
P67L-15	X	X				X		X												X
P67L-12	X				X	X		X		X										X
P67L-9	X				X			X	X	X				X				X		
P67L-6	X	X			X	X		X	X	X			X	X		X		X		
P67L-3	X	X		X	X	X		X	X	X			X	X			X	X		X
P67L	X	X		X	X	X		X	X	X			X	X		X	X	X		X
MV6P67	X			X	X			X	X	X				X				X		X
MV12P67	X	X		X	X	X		X												X
MV18P67	X			X				X				X						X		
MV24P67												X								X
P1L-24								X				X								
P1L-21								X												
P1L-18																X				X
P1L-15	X		X	X				X				X								X
P1L-12	X		X	X	X		X	X												X
P1L-9	X		X		X		X	X	X											X
P1L-6	X	X	X		X	X	X	X	X	X	X	X	X	X				X		X
P1L-3	X		X		X		X	X	X		X	X	X		X	X	X		X	
P1L	X	X	X	X	X	X	X	X	X	X	X		X	X	X		X	X	X	
MV6P1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X
MV12P1	X	X	X		X	X	X	X	X		X		X		X		X			X
MV18P1	X		X		X		X	X	X				X							
MV24P1	X		X		X		X	X	X											X
RP1-24																X				
RP1-21				X				X												X
RP1-18		X				X		X		X				X				X		X
RP1-15				X												X				X
RP1-12								X				X								X
RP1-9								X								X				X
RP1-6		X				X						X								X
RP1-3								X								X				X
RP1				X		X		X		X				X				X		X
MV3RP1																X				X
MV6RP1		X		X		X		X		X		X		X				X		X
MV12RP1				X				X								X				X
MV24RP1		X				X				X		X		X		X		X		X
Total (54)	**24**	**18**	**10**	**21**	**20**	**21**	**9**	**45**	**16**	**18**	**5**	**15**	**11**	**21**	**4**	**16**	**7**	**18**	**3**	**37**

Models were developed at finer lead times of 6, 9, 12, 15 and 18 hours, which permitted exploration of the maximum warning time that can be anticipated. As can be seen in Table 5.30, model P selected input variables only from P1 and nothing from the upper gauging stations or the rain gauge. The only input determination techniques that included rain gauge data were models S and Pr. There was no correlation greater than 0.90 between the rain gauge and the P1 station. The S and Pr models reduced the input variables by between 60%-73% for the 9 and 18 hour lead times. Model Pr reduced the number of input variables by between 17% and 35% overall.

All input variables from P1 had correlations with the predicted levels of less than 0.90. Compared to Case Studies 1 and 2, the input data for selection were slightly narrower, i.e. 2/09 – 31/10/03 and 2/08 – 31/10/04.

5.6.2 Results

As with previous case studies, the model results are presented in terms of goodness-of-fit statistics and visual inspection of the hydrographs for each lead time. The full range of hydrographs can be found in Appendix D.

6 hour lead time

As can be seen in Figure 5.14, it is interesting that model P seems to be the best model for a 6 hour prediction at the peak and in terms of timing because model P selected only the input variables from P1 (removing 81% of the variables). However, this would not be acceptable in hydrological forecasting, i.e. the best method for selecting inputs should use information from the upper stations rather than the station itself. Model Pr (pruning algorithm) predicted a good rising time with only a 30 minute delay (Figure 5.14 and Table 5.31).

Table 5.31: The number of inputs selected and the goodness of fit measures for a lead time of 6 hours.

Input Variables	Input Determination Techniques			
	C	S	P	Pr
Number of inputs	24	18	10	21
PDIFF (m)	0.205	0.084	-0.265	0.120
MAE (m)	0.022	0.0206	0.0208	0.0208
RMSE (m)	0.042	0.037	0.042	0.041
CE	0.991	0.993	0.992	0.992
Time delay (hr)	3.5	1.5	-1	0.5

Model S selected inputs from each stage gauging station and the rainfall data, removing 66% of the total variables and giving a good forecasting performance overall, but it underpredicted the stage by 8 cm. For practical forecasting the model is telling

the forecasters when the River Ping will rise above bank stage, but where the forecasters want to warn about the extent of flooding, improving on the maximum stage predictions would be helpful. Model C, on the other hand, was the worst performing model similar to Case Studies 1 and 2.

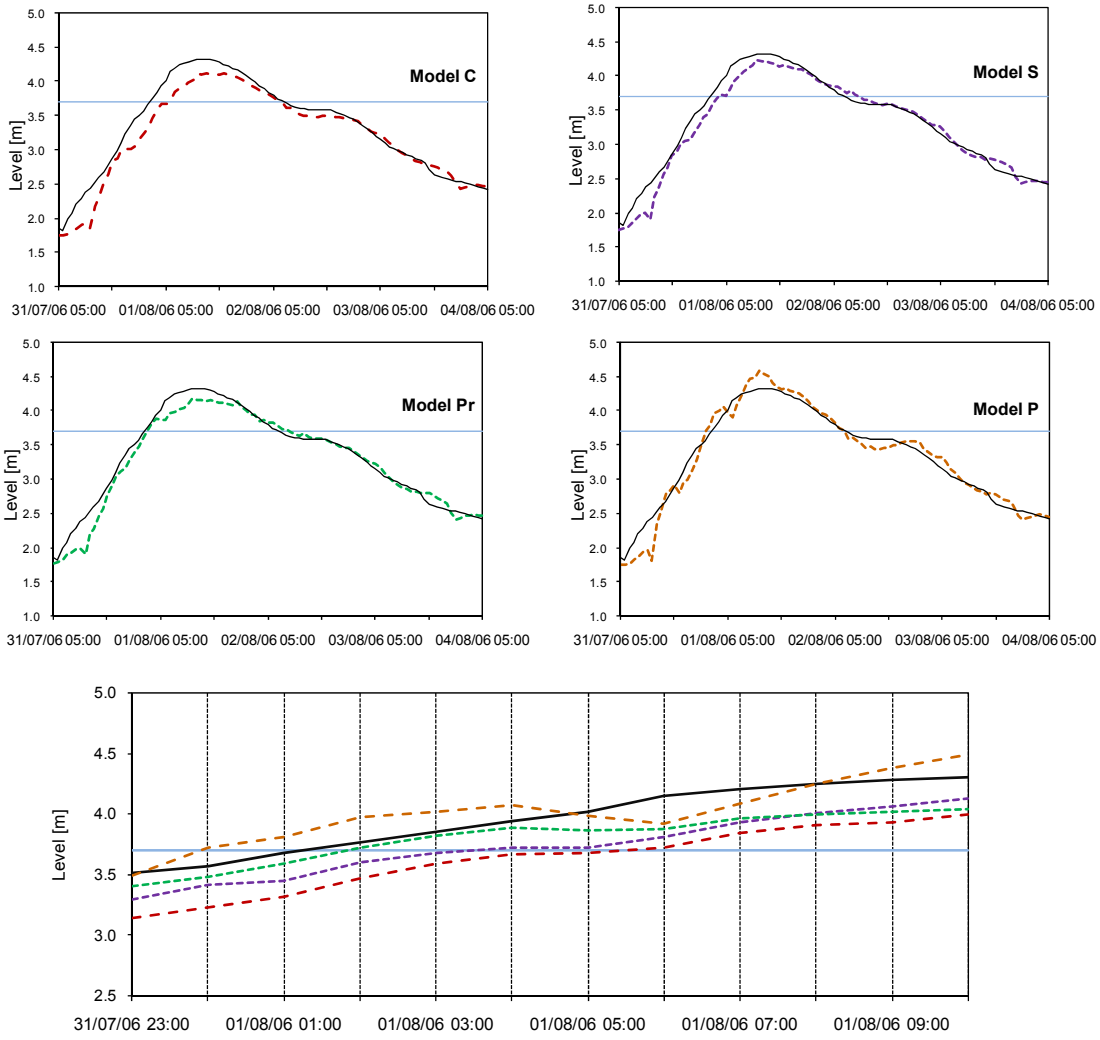


Figure 5.14: Comparison of the hydrographs for 4 models at t+6 hours (31 Jul – 4 Aug 2006).

9 hour lead time

As with the 6 hour forecast, model P is the best model for predicting the peak stage and time to peak (Figure 5.15). However, the goodness of fit statistics (Table 5.32) show that model Pr has the best performance at the peak. Moreover, model Pr selects almost all of the input variables and removes only 16%, while other models remove between 63% and 83%.

Table 5.32: The number of inputs selected and the goodness of fit measures for a lead time of 9 hours.

Input Variables	Input Determination Techniques			
	C	S	P	Pr
Number of Inputs	20	21	9	45
PDIFF (m)	0.277	0.215	-0.334	0.209
MAE (m)	0.032	0.031	0.033	0.030
RMSE (m)	0.063	0.057	0.065	0.056
CE	0.980	0.984	0.979	0.985
Time delay (hr)	6	2.5	-0.5	0

Once again, model C is the worst model at a 9 hour lead time, i.e. it has the highest delay in predicting the rising limb (6 hr) and a lower estimated peak of approximately 27 cm. Model P selects variables from only P1. The results suggest that the model Pr predicts the best results at this lead time.

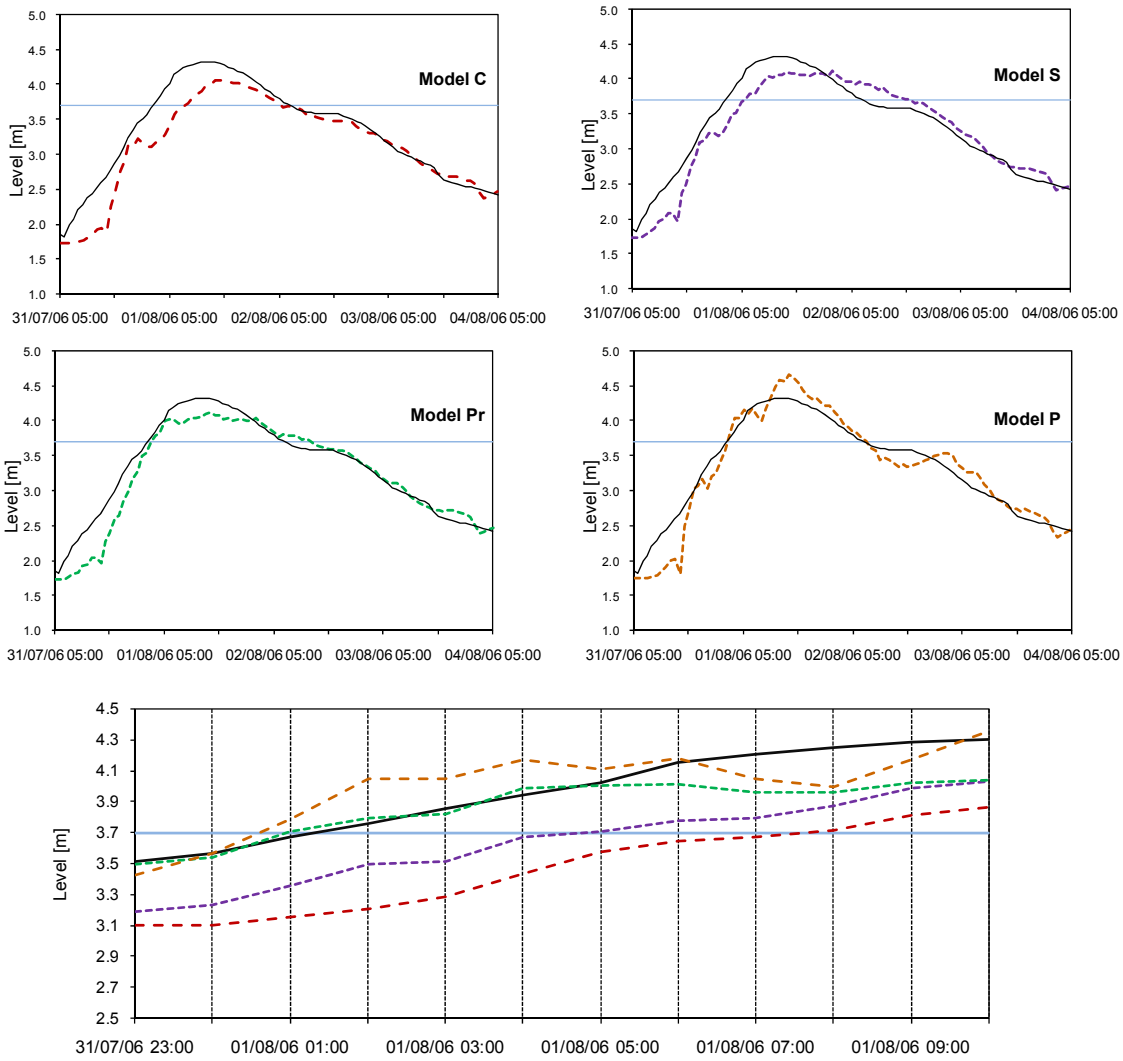


Figure 5.15: Comparison of the hydrographs for 4 models at t+9 hours (31 Jul – 4 Aug 2006).

12 hour lead time

Model P predicted the peak closer to the actual one and had the best rising limb prediction compared to other models, while model P only selected five variables from P1. As can be seen from Figure 5.16 for Model P, there are several points where the predicted and actual hydrographs cross each other. Only models C and S underestimate the peak stage, whereas model Pr and P have overestimated the peaks.

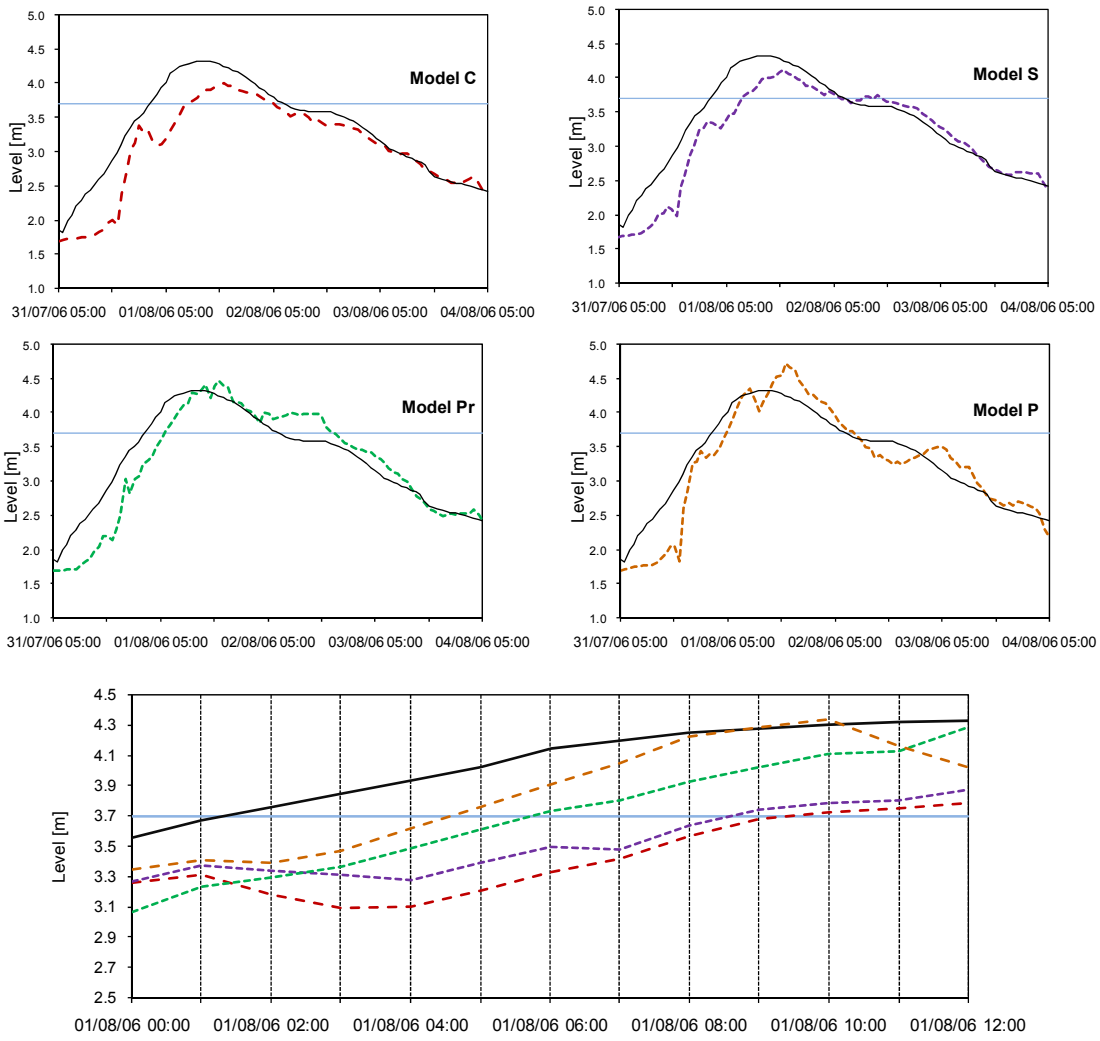


Figure 5.16: Comparison of the hydrographs for 4 models at t+12 hours (31 Jul – 4 Aug 2006).

Model S and model Pr are better than the other two models. Moreover, model Pr predicts the peak only 4 hours behind the actual time (Table 5.33). Model C is the worst performing model as it predicts with the most delay and the peak prediction is poor.

Table 5.33: The number of inputs selected and the goodness of fit measures for a lead time of 12 hours.

Input Variables	Input Determination Techniques			
	C	S	P	Pr
Number of Inputs	16	18	5	15
PDIFF (m)	0.325	0.137	-0.384	-0.121
MAE (m)	0.043	0.041	0.047	0.046
RMSE (m)	0.083	0.076	0.091	0.084
CE	0.966	0.972	0.959	0.965
Time delay (hr)	7.5	7	3	4

15 hour lead time

The forecast at this point was included to see whether this was more helpful than forecasts at lead times of 12 and 18 hours, as it is in the right travel time frame for the flood wave, and to see whether the causes of the noise on the forecasts could be identified.

The forecast stage hydrographs at t+15 hours ahead all show noise in the output as do t+12 and t+18. The best model for predicting the time to peak is model Pr followed by model P, C and S. At the highest point the model Pr is likely to be the best model as the predicted stage is the same as the actual line (Figure 5.17) but the statistics show that model S is better than other models as the CE is 0.957 and the RMSE is 0.093 (Table 5.34). This reconfirms the need to inspect the hydrographs and not to rely on the statistical tests alone.

Table 5.34: The number of inputs selected and the goodness of fit measures for a lead time of 15 hours.

Input Variables	Input Determination Techniques			
	C	S	P	Pr
Number of Inputs	11	21	4	16
PDIFF (m)	0.226	0.222	-0.391	0.111
MAE (m)	0.054	0.051	0.062	0.057
RMSE (m)	0.103	0.093	0.117	0.102
CE	0.948	0.957	0.932	0.949
Time delay (hr)	7.5	9	4.5	2.5

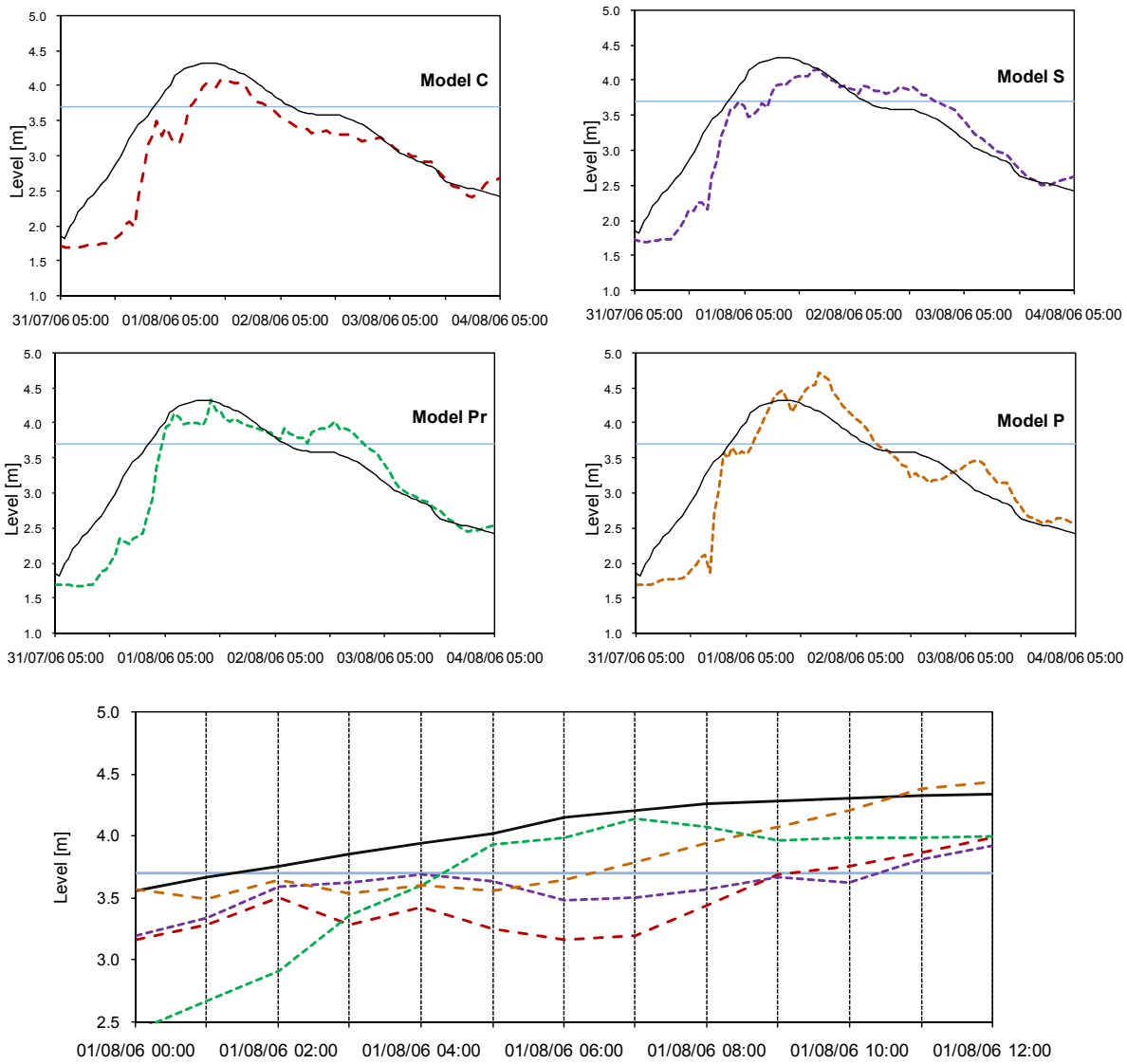


Figure 5.17: Comparison of the hydrographs for 4 models at t+15 hours (31 Jul – 4 Aug 2006).

18 hour lead time

Model P selects less input variables. It is the only model that selects variables from P1 only and overestimates the peak (Figure 5.18). There are also large errors of 1.5 m at the small peak (Figure D.5). Table 5.35 shows model S gave the best performance as it has the highest CE, although the model forecast underestimated the peak stage.

Table 5.35: The number of inputs selected and the goodness of fit measures for a lead time of 18 hours.

Input Variables	**Input Determination Techniques**			
	C	**S**	**P**	**Pr**
Number of Inputs	7	18	3	37
PDIFF (m)	0.074	0.233	-0.380	0.130
MAE (m)	0.068	0.061	0.075	0.063
RMSE (m)	0.126	0.111	0.141	0.113
CE	0.922	0.939	0.903	0.937
Time delay (hr)	3.5	3	1	4

Model C had poorer results at all previous lead times but is improved at a lead time of 18 hours. The model performance seems to be better at predicting the peak when compared with model S. However, model S has the best performance overall although P has the shortest time delay.

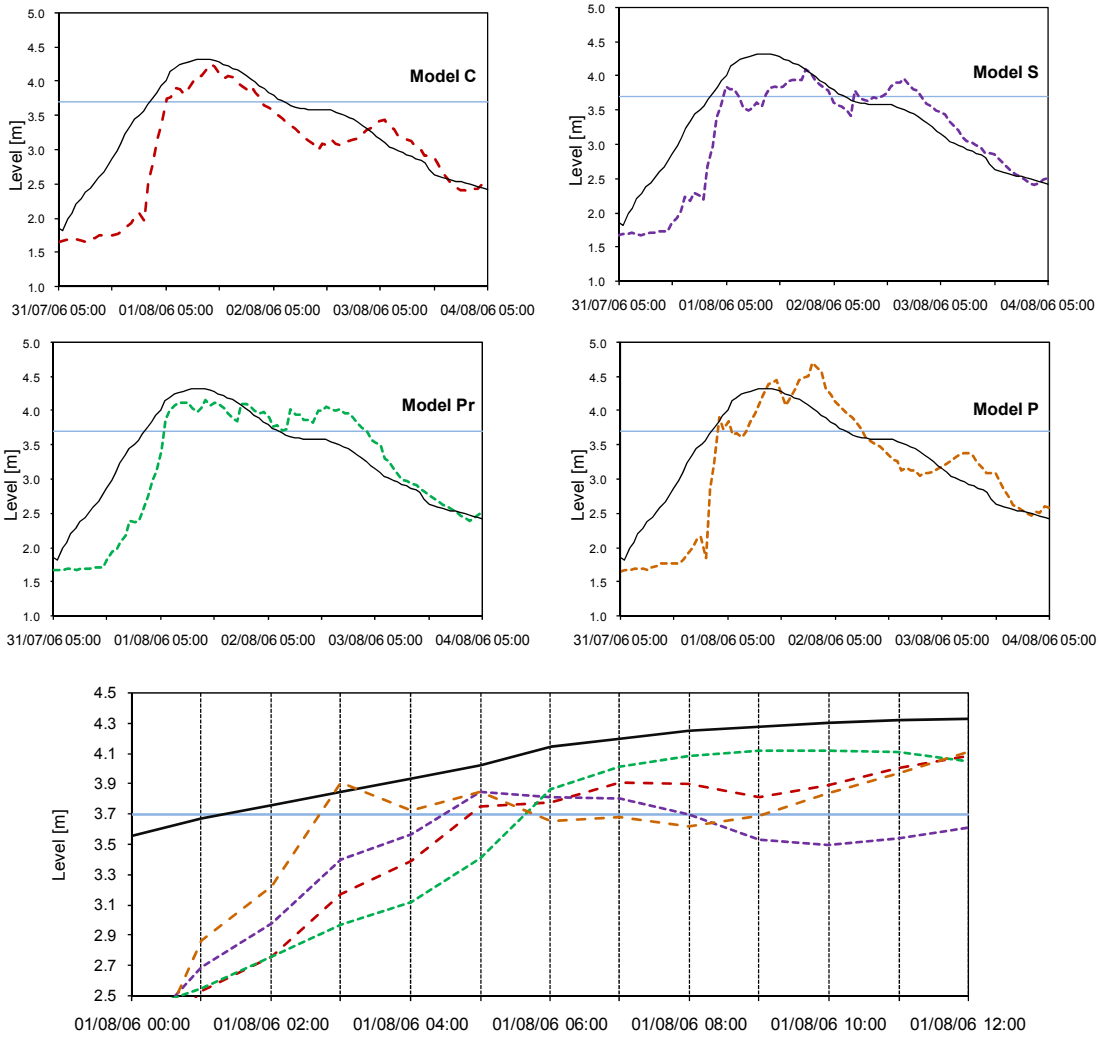


Figure 5.18: Comparison of the hydrographs for 4 models at t+18 hours (31 Jul – 4 Aug 2006).

5.6.3 Summary of case study 4

As discussed in Chapter 3, there is a scarcity of rainfall data for modelling, and the data from the Chiang Mai rain gauging station (R1) was made available relatively late in the research programme. In theory, the addition of a precipitation record should be helpful, but in a catchment that is this large, an individual station can never be representative of the whole catchment, and this station is very near to P1. There are other rain gauges available but they only record daily data. The Hydrology and Water Management Centre for Upper Northern Region, Thailand has now established more hourly rain gauge stations but they were only begun to be put in place after the large flood in 2005.

Table 5.36 provides a summary of the percentage of variables that were removed by each input determination method for the different lead times. The stepwise regression technique consistently removes the same number of input variables compared to Case Studies 1 and 2. PMI was introduced in this case study and removed 80-90% of the variables.

Table 5.36: The percentage of variables removed by the different input determination methods by lead time.

Model	**Lead time**				
	6 hours	**9 hours**	**12 hours**	**15 hours**	**18 hours**
C	55.5	62.9	88.8	79.6	87.0
S	66.6	61.1	66.6	61.6	66.6
P	81.4	83.3	90.7	92.5	94.4
Pr	61.0	16.6	72.2	70.3	31.4

Table 5.37 ranks the performance of different input determination methods. Based on the CE, RMSE and MAE, model S proved to be the best fit model and ranked second rank in terms of peak stage prediction, while model Pr was the best at peak stage prediction and ranked second in terms of time to peak prediction. Model P selected only P1 variables and was the best model to predict flood time and reduce the number of inputs. In contrast, the overall performance was the worst. The performance of model C did not show any significant results except that it ranked second in terms of removing the number of inputs.

Table 5.37: Ranking of model performance (1 is the best and 4 is the worst).

Statistic	**Lead time**	**C**	**S**	**P**	**Pr**	**Statistic**	**C**	**S**	**P**	**Pr**
Predict flood time	6 hours	12	15	20	17	PDIFF	13	19	7	17
	9 hours	9	14	19	18		6	11	4	12
	12 hours	7	8	13	11		5	14	2	16
	15 hours	7	6	10	14		9	10	1	18
	18 hours	12	13	16	11		20	8	3	15
	Total (**Rank**)	47 (**4**)	56 (**3**)	78 (**1**)	22 (**2**)		53 (**3**)	62 (**2**)	17 (**4**)	78 (**1**)
CE	6 hours	18	20	19	19	RMSE	18	20	18	19
	9 hours	15	16	14	17		15	16	14	17
	12 hours	12	13	10	11		12	13	10	11
	15 hours	7	9	4	8		7	9	4	8
	18 hours	3	6	2	5		3	6	2	5
	Total (**Rank**)	55 (**3**)	64 (**1**)	49 (**4**)	60 (**2**)		55 (**3**)	64 (**1**)	25 (**4**)	60 (**2**)
MAE	6 hours	18	20	19	19	Reduce input	8	11	15	9
	9 hours	15	16	14	17		10	9	17	6
	12 hours	12	13	11	10		12	11	18	13
	15 hours	8	9	5	7		14	9	19	12
	18 hours	3	6	2	4		16	11	20	7
	Total (**Rank**)	56 (**3**)	64 (**1**)	51 (**4**)	57 (**2**)		60 (**2**)	51 (**3**)	89 (**1**)	47 (**4**)

In conclusion, using rainfall data improved the forecasting performance for models S and Pr. It also confirmed that using data from P1 as an input variable is essential to predict water level at P1. For example, model P selected only P1 data for forecasting yet it predicted close to the actual water level. In contrast, model C, which did not select data from the rain gauge, was the least good forecasting model. Inspecting the hydrographs showed that model P gave the best performance at lead times of 12 and 18 hours, whereas model Pr was the most accurate for 9 hours ahead.

5.7 Conclusions

Eight input determination techniques were tested including: correlation > 0.90 (C), stepwise regression (S), correlation > 0.90 + stepwise regression (CS), M5 model tree (M), data mining (D), genetic algorithm (G), partial mutual information (P) and a pruning algorithm (Pr) across 4 case studies. Neural network models were developed on these inputs to predict river levels for the Upper Ping River at various lead times from 6 to 24 hours ahead.

The first case study used a standard feedforward network trained with the Levenberg-Marquardt (LM) algorithm. The results showed that neural network models could be developed for lead times of 6 and 12 hours but by 18 hours, the forecasts had declined in performance. Most of the input determination methods performed similarly although the CS and G methods generally had better performance statistics.

In Case Study 2, a Bayesian Regularization (BR) algorithm for neural network training was introduced that outperformed the LM algorithm. Inspection of the hydrographs and statistics showed that overall stepwise regression used with BR produced good results and reduced the large number of variables in the input datasets by approximately 60%, which is economical in computational terms for neural network development. Other research has suggested that Bayesian Regularization models work well in catchments with limited data (Ali *et al*., 2010; Chaipimonplin *et al*., 2008a, 2008b; Kunikazu *et al*., 2009; Youhang *et al*., 2008; Chaohong *et al*., 2008; van Hinsbergen *et al*., 2008). This research suggests that this is also true for the Upper Ping catchment. Bayesian Regularization does not require validation data, which is very helpful when data are limited. In this chapter each model was run 50 times for each experiment and the average forecast is presented after Anctil (2007). Chapter 7 examines the effect of this averaging and the variation across the model runs.

In Case Study 3, the number of hidden nodes in the pruning algorithm influences the processing time and model performance. A Neural network model with more hidden nodes requires more time to train and predicts a better result then a model with less hidden nodes for the complex catchment.

In Case Study 4, all 4 models at a lead time of 18 hrs produced predictions with an approximate 3 hr delay time in the peak and a 20 cm error at the peak. Moreover, it appears that training a model with a big storm in 2005 as well as using rainfall can improve the model performance when testing on smaller storms.

In terms of input determination methods, it was found that model M does not remove many variables while models D and P remove many input variables. However, model D gives a poor result on the rising limb. Model P never selected any input variables from the upper water level stations, which indicates it is not the best method. Model C and CS are not much different in terms of performance. However, the most significant finding is that the neural network model performance at lead times of 6 to 24 hours is simply not consistent. For example, model D has the highest CE at a 6 hour lead time but the lowest at 12 and 18 hours. Model G is best at lead times of 12 and 18 hours but had much poorer performances at lead times of 6 and 24 hours. In contract, model S and CS gave the most stable results even though model S does not remove that many input variables compared to model D or CS. In summary, the percentage of input variables removed by each technique is 90-95% for model D, 80-90% for model P, 70-80% for model CS, 60-70% for models Pr, G, C and S, and 30% for model M. Model CS and S predicted values with approximately a 3-4 hour time delay in the peak for an 18 hour lead time. Therefore, the best input determination technique emerging from these experiments is the combination between correlation and stepwise regression.

The results here suggest that the best forecast model so far is based on data from three water level stations and one rain gauge station. The travel times between P67 and P1 dictate that forecasting 18 hours ahead is practical. Including the rain gauge data made some small improvements but the geographical position of the gauge and the very large size of the catchment make this a limited step forward for the forecaster faced with a monsoonal storm event in real time. The research from this chapter is developed further in Chapter 6 with the inclusion radar data as a surrogate for rainfall data across the catchment. Radar data were used to enhance the data sets discussed here in further water stage forecasting experiments in order to see whether the lead time of the forecasts can be extended beyond 18 hours.

Chapter 6 Extending the Neural Network Forecast Lead Time with Radar Data

6.1 Introduction

Information from radar can be an important input to hydrological forecasting, whether through visualization of incoming storm events or to estimate rainfall. Radar data are generally calibrated using point rain gauge measurements. However, this is problematic in areas where sparse rain gauge networks exist, in particular in the Upper Ping catchment where there is only 1 hourly rain gauge near P1. Right now the maximum lead time for the neural network to forecast the water stage at P1 using 2 water stage stations in the upper catchment is approximately 18 hours (as shown in Chapter 5). An alternative approach explored in this chapter uses data from the raw radar image as an input to the neural network for extending the lead time. Moreover, this has not been tried before, since it is bound to be less useful than radar estimated precipitation values, but in this context it appears to be an option worth investigating.

The purpose of this chapter is to investigate whether the addition of radar data can improve these neural network hydrological models, and in particular to extend the lead time of the forecast. A brief overview of the use of radar data in hydrology is provided, followed by information on the radar configuration relevant to the Ping catchment in Thailand. A description is then provided of what radar data are available for the modelling experiments and how the radar data have been extracted and used as inputs to the neural network models. An initial set of experiments is undertaken whereby raw radar reflectivity values are added as intputs to neural network models for lead times of 6 to 24 hours. The focus of the chapter then shifts to developing models of much longer lead times, i.e. those beyond 24 hours. The radar reflectivity values are correlated with water levels at P1 to reveal that lead times as high as 48 hours are possible. Experiments with building neural network models of lead times between 24 and 48 hours are then presented. These experiments begin with using radar data only and then add rainfall and water level data to determine whether the model results can be further enchanced. A detailed schematic of the experiments is provided in section 6.6 to aid the reader in understanding the sequence of model development.

6.2 Radar Concepts

This section presents an overview of Doppler radar and how this is related to rainfall. Doppler radar transmits electromagnetic radiation through the atmosphere and the

receiver then detects the radiation scattered back from the target, measured in decibels. The radar reflectivity factor (Z) is correlated with the size and density of the precipitation, as shown in equation 6.1 and 6.2, so a high value of Z or dBZ represents a high density or potentially large amounts of precipitation (Collier, 1996, p 17):

$$Z = \sum_{V_c} \frac{D^6}{V_c} \tag{6.1}$$

$$dBZ = 10\log_{10}\left(\frac{Zmm^6m^{-3}}{1mm^6m^{-3}}\right) \tag{6.2}$$

where D is the drop diameter, V_c is the pulse volume and Z is the accumulated rain per unit volume of the diameter of the water drop. For the same pulse volume, precipitation of 64 units with raindrops of a 1 mm diameter would have the same value of Z with 1 unit of precipitation and raindrops with a 2 mm diameter (i.e. $2^6 = 64$).

To estimate the rainfall from the radar reflectivity, the radar values must be calibrated with data from rain gauges using a Z-R relationship (equation 6.3)

$$Z = aR^b \tag{6.3}$$

where a and b are dependent on the type of precipitation and R is the precipitation intensity (mm h^{-1}) (Collier, 1996). Jones (1956) has suggested suitable Z-R relationships for different types of precipitation. Other examples in the literature include Rosenfeld *et al*. (1993) applied $Z = 230R^{1.25}$ for the monsoon, Kalinga and Gan (2007), used $Z = 250R^{1.2}$ for a tropical regime and Chumchean *et al*. (2008) used $Z = 103R^{1.5}$ for convective clouds. Moreover, Joss *et al.* (1970) cited in Collier (1987) recommended the following values of a and b for thunder showers ($Z = 486R^{1.37}$), rain showers ($Z = 380R^{1.24}$), continuous rain ($Z = 313R^{1.25}$) and thunderstorms ($Z = 500R^{1.5}$).

There are several issues with using radar in this type of relationship. The first is ground clutter, which is defined as strong echoes from the terrain that could be misinterpreted as high amounts of rainfall (Figure 6.1). To minimize this problem, the radar site should be located on the top of a mountain. Error due to ground clutter can also be corrected by subtracting the signal received on days in which no rain fell from those in which it does (Hill and Robertson, 1987).

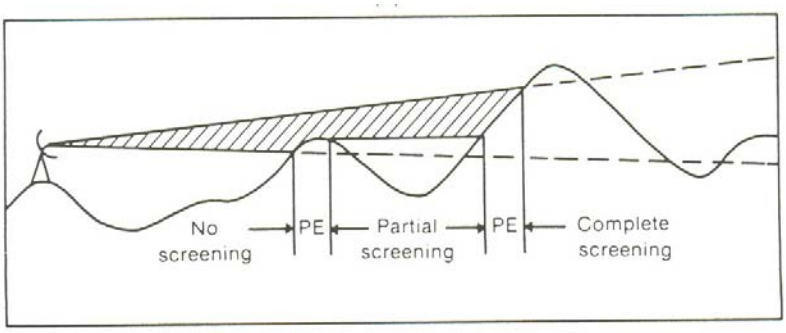


Figure 6.1: Ground clutter.
Taken from Collier (1996), p 43.

Another potential source of error is meteorological. In Figure 6.2, there are 6 possible errors in the radar measurement of surface rainfall intensity (Browning, 1987). The first error occurs at long ranges where the radar will overshoot the low level rainfall. The second and third errors in Figure 6.2 present examples where the radar is unable to detect low level rainfall. A fourth potential error may occur when ice or water melts, which increase the reflectivity and may lead to an overestimation of rainfall. On the other hand, underestimation of the rainfall intensity may occur when the radar beam does not sense the areas of large amounts of rainfall, as the fifth error, while a strong hydrolapse can bend the radar beam (sixth error).

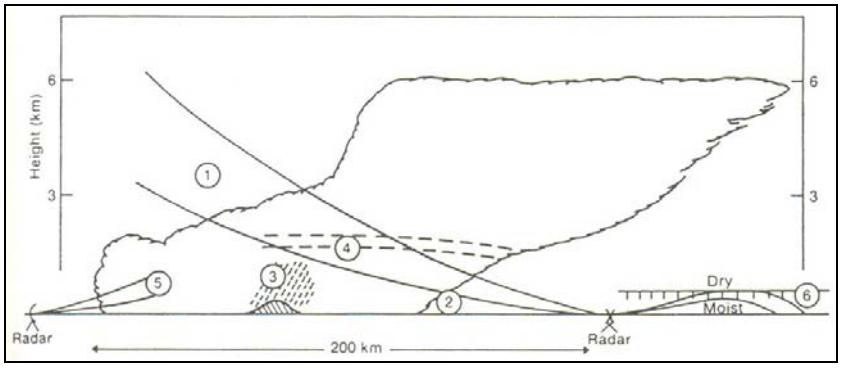


Figure 6.2: Meteorological errors.
Source: Browning (1987), p 242.

The final issue with the use of radar is selection of the wavelength. Figure 6.3 provides information aboutwavelength absorption, where a wavelength of 1.35 cm is highly absorbed by water, and for longer wavelengths, there is less attenuation by the atmosphere. To detect precipitation, a suitable wavelength of greater than 3 cm should be chosen.

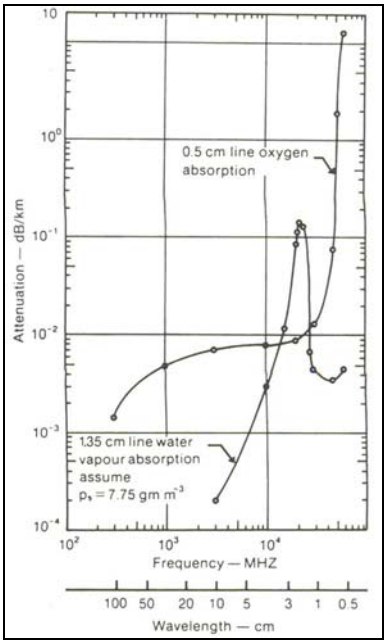


Figure 6.3: Atmospheric absorption by the 1.35 cm line of water vapour for a mean absolute humidity of 7.72g/m3, and by the 0.5 cm line of oxygen at a temperature of 20 C and pressure of 1 atm.
Source: Battan (1973) citied in Collier (1996) p 19.

Consequently, there are 3 different wavelengths: X band (λ= 3 cm); S band (λ = 10 cm) and C band (λ = 5 cm). The radar reflectivity from precipitation at a wavelength of 5 cm is 12 dB stronger than at a wavelength of 10 cm (Ulaby et al., 1981 cited in Collier, 1996). However, there is a great chance of ground clutter at the larger wavelength of 10 cm (Harrold, 1984 cited in Collier, 1996).

6.3 The Use of Radar in Hydrology

Doppler radar has been used for detecting rain drops, cloud precipitation, cloud movement or velocity (Lee *et al*., 2006). The advantage of radar is the high temporal and spatial resolution, i.e. images are recorded at every 5 to 15 minutes at a resolution of 1 km^2. Moreover, cloud indexing methods are used for integrating a satellite image's visible and infrared band. Cloud indexing can be used with both polar orbiting and geostationary satellites. This technique evaluates the cloud brightness from the visible and infrared bands, and it is best for convective clouds (Collier, 1996; Wardah *et al*., 2008):

$$R_v = a_o A_c + a_1 \frac{dA_c}{dt} \tag{6.4}$$

Where R_v is the volumetric rain rate, A_c is the area of the cloud, dA_c / dt is the rate of change in cloud area and a_o, a_1 are empirical coefficients. Collier (1996) also noted that cloud indexing provides the best accuracy for estimating rainfall from satellites when

compared with other techniques such as area averaging, life history, passive microwave and active microwave.

To predict downstream flow, rainfall and upstream flow are normally used. Moore (1987) suggested considering lateral inflow from the rainfall; however, lateral flow from rain gauges leads to errors in rainfall estimation due to the sparsity of rain gauge networks and low spatial variability. As a result, the rainfall estimated from radar would be an alternative way to add a higher spatial variability and provide better lateral flow information than data from a single point such as a rain gauge (Wilson *et al.*, 1979 cited in Collier, 1996). Collier (1987) investigated monthly average hourly rainfall data for 1 year from 4 different rain gauges situated at different distances away from the radar in the summer and winter in the UK. He found that to estimate rainfall from radar during the summer when the rainfall is produced by convective activity, the accuracy was high at all 4 stations. However, for the lower level winter rainfall, the values were underestimated at the far range and overestimated at the near range.

Using different rainfall data sources as the input to a flood forecasting model will influence the model results. To estimate the average rainfall over the catchment from an infinite number of points has less variance than estimation from a single point rain gauge (Moore, 1987). Moreover, radar data have been applied in hydrological modelling. Troutman (1983) found that calibrating the rainfall average from radar data with a rainfall runoff model for flood forecasting produced a model that under predicted heavy storms and over predicted small storms. In contrast, when calibrating the same model with rainfall from rain gauges, the heavy storms were over predicted and the small storms were under predicted. Moore (1987) applied an exponential density to estimate the average rainfall from the average of the pixel values of the radar image and he suggested that it would be sufficient to use a simple exponential density function to estimate the average rainfall. Borga (2002) found that a conceptual streamflow model performance improved by 30% when adjusted radar data with mean field bias were used. Smith *et al*. (2007) used a Z-R relationship to forecast flash floods. Cole and Moore (2008) investigated three different types of rainfall estimators as inputs to a lumped conceptual rainfall-runoff model, i.e. the rain gauge only; rainfall from the Z-R relationship, and rain gauges adjusted with radar data. Interestingly, the model performed best when using only rain gauge data as inputs to the model. In contrast, Filho and Santos (2006) succeeded in improving a neural network flood forecasting model by 40% when they supplemented stage with rainfall estimated using a Z-R

relationship. Chiang *et al*. (2007) applied neural networks to forecast precipitation from the Z-R relationship.

To deal with a large catchment, i.e. greater than 1000 km^2, Knowles (1987) divided the catchment into 3 areas: one covering the main river and the other two covering the tributaries. Each area was 260-460 km^2. He then calibrated the radar data and developed rainfall-runoff models for each area. In another study Morris (1980) subdivided the catchment and applied a lumped rainfall-runoff model to each sub-catchment with average rainfall from a 2 km radar grid.

6.4 Radar in Thailand

Radar in Thailand detects precipitation using the CAPPI (Constant Altitude Plan Position Indicator) technique. The CAPPI was developed from the PPI (Plan Position Indicator), which transmits electromagnetic radiation in only one angle, but the CAPPI can transmit in several angles. Therefore, with this technique, volume can be simulated. As Thailand is a tropical country, the precipitation is predominantly from the southwest and northeast monsoon (Okumura *et al*., 2003), with the majority of rainfall from convective clouds. As mentioned in Section 6.2, different areas or types of storm apply different Z-R relationships. Therefore, many researchers have investigated rainfall estimation from radar images in Thailand. For example, Rachaneewan (2006b) investigated a suitable Z-R relationship $Z=300R^{1.4}$ for the northern part of the Thailand catchment, Chumchean (2007) recommended $Z=128R^{1.5}$ for the Bangkok area, Mapiam and Sriwongsitanon (2008) investigated rainfall estimation from radar using a climatological Z-R relationship and Dejyothin (2008) developed an automatic value adjustment system for the Z-R relationship by rain gauge.

6.5 Analysis of Radar Data

After having reviewed the basics of radar data and their use in hydrology, this section will now outline how the data have actually been used in this research, in particular how the data have been extracted from the radar images for use as inputs to the neural network model to predict water stage at the station referred to as P1.

Radar images are only available for 2003, 2005 and 2006 (Table 3.2). Figure 6.4 presents all eight storms that occurred during this period:

- in 2003, labelled as
    - S1: 6-23 Sep, flood: 12-14 Sep, max level: 3.70 m;

- in 2005, labelled as
    - S2: 2-24 Aug, flood: 14-16 Aug, max level: 4.90 m;
    - S3: 25 Aug-17 Sep, flood: 12-13 Sep, max level: 3.79 m;
    - S4: 18-23 Sep, flood: 20-22 Sep, max level: 4.71 m;
    - S5: 28 Sep-11 Oct, flood: 29 Sep-2 Oct, max level: 4.93 m;
    - S6: 26 Oct-5 Nov, flood 1-2 Nov, max level: 3.79m;
- in 2006, labelled as
    - S7: 18-26 Sep, no flood, max level: 3.29 m; and
    - S8: 9-17 Oct, no flood, max level 3.53m.

Note that there are missing radar images for the first two storms in 2006 so these are not labelled in Figure 6.4.

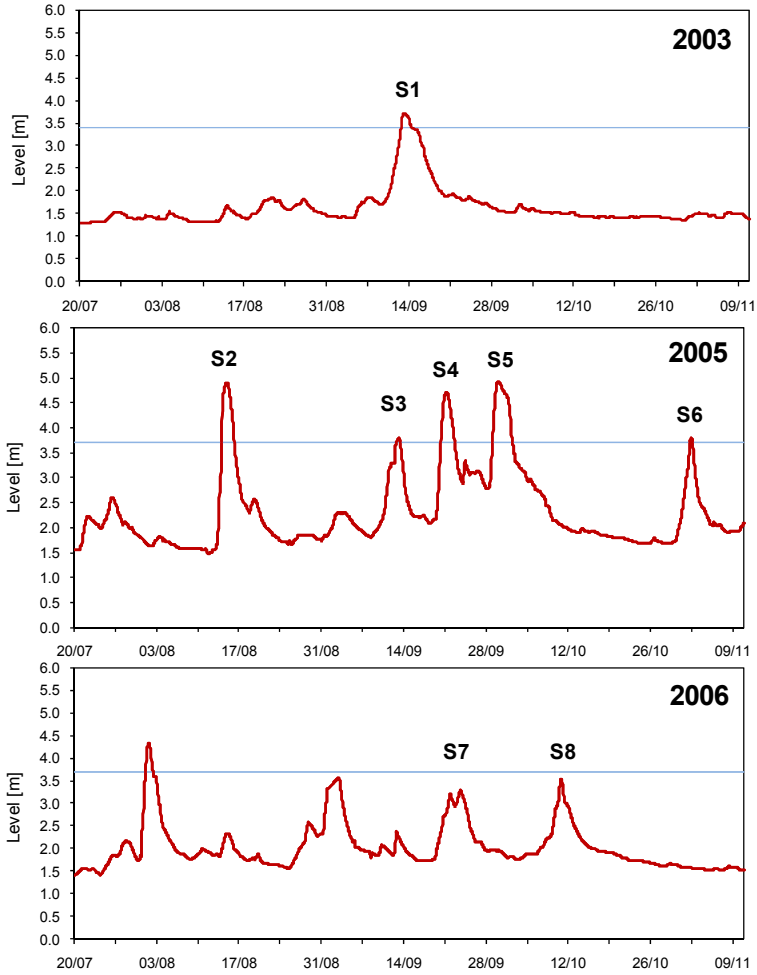


Figure 6.4: The storms that occurred in 2003, 2005 and 2006.

The radar images in Figure 6.5 illustrate that when the study area is covered with a big storm of more than 11 hr, this could lead to flooding in the area. For example, the first big storm 2005 (S2) passed the study area on 12 August and covered this area for 17 hours. Storm S4 approached on 18 September and covered the area for 11hr, storm,

S5 approached on 28 September and covered the area for 13 hr and storm S6 covered the area for 16 hr.

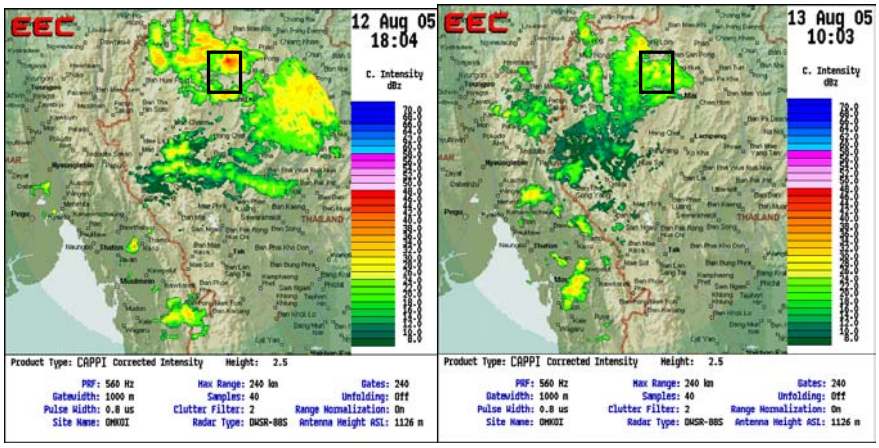


Figure 6.5: Example of radar images covering the study area during the S2 storm event.

The first step in working with the data was to extract a 30 x 50 km square north of Chiang Mai from each of the radar images, which are available at hourly intervals. As shown in Figure 6.6, this provides sufficient coverage of the river on both sides and also covers the P75, P67 and P1 water stage stations. The image was sampled at 12 points or pixels in a 4x3 matrix with a distance of 10 km between the points. The points are labelled as Z11, Z12, etc. to reflect the row and column position in the data matrix.

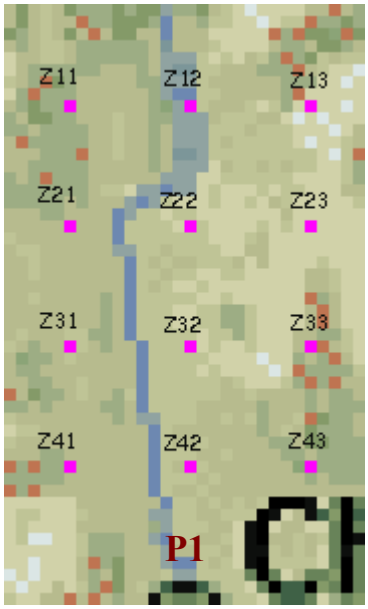


Figure 6.6: Twelve sample points extracted from a 30 x 50 km radar image covering the study area.

There are different methods of extracting the radar reflectivity at the sample points. In calibration of rainfall at rain gauges, normally the eight pixels surrounding the rain gauge point are used (Fisher *et al*., 2000; Georgrakakos *et al*., 2000). For this reason,

the surrounding 3x3 neighbours (9 values) were extracted from each of the 12 sample points and combined using four different methods as follows:

- Sum Step: sums all 9 pixel values at 6 hour intervals, e.g. t, t-6, t-12 up to t-24;
- Aver Step: same as Sum Step except the values are averaged instead of summed;
- Sum Accum: same as Sum Step except that cumulative values are calculated, e.g. the sum of t to t-5, t-6 to t-11, etc. up to t-24
- Aver Accum: same as Sum Accum except that the initial nine pixels are first averaged and then a cumulative value over a 6 hour interval is calculated.

The NN used in this section is the same as the model used in Case Study 4 (Section 5.6), which was trained with Bayesian Regularization and run 50 times. The first big storm in 2005 (S2) was used in the testing data set while the other 7 storms comprised the training data. The total number of input variables was 60 (12 sample points * 5 previous time steps at 6 hour intervals as explained in the previous paragraph). Models were developed to forecast water stage at 6, 12, 18 and 24 hours ahead.

Figure 6.7 presents the results for all four lead times using storm S2 in the testing data set. The first thing to note is that the results are very similar for ‘Sum Step’ and the ‘Aver Step’ and likewise for ‘Sum Accum’ and ‘Aver Accum’ at all lead times so it appears to make little difference as to whether a sum or average is used. The next observation is that the hydrographs appear to fluctuate more for the Accum methods compared to the Step methods. However, both Accum methods underpredicted the peak although this improved as the lead time increased to 24 hours (Figure 6.8).

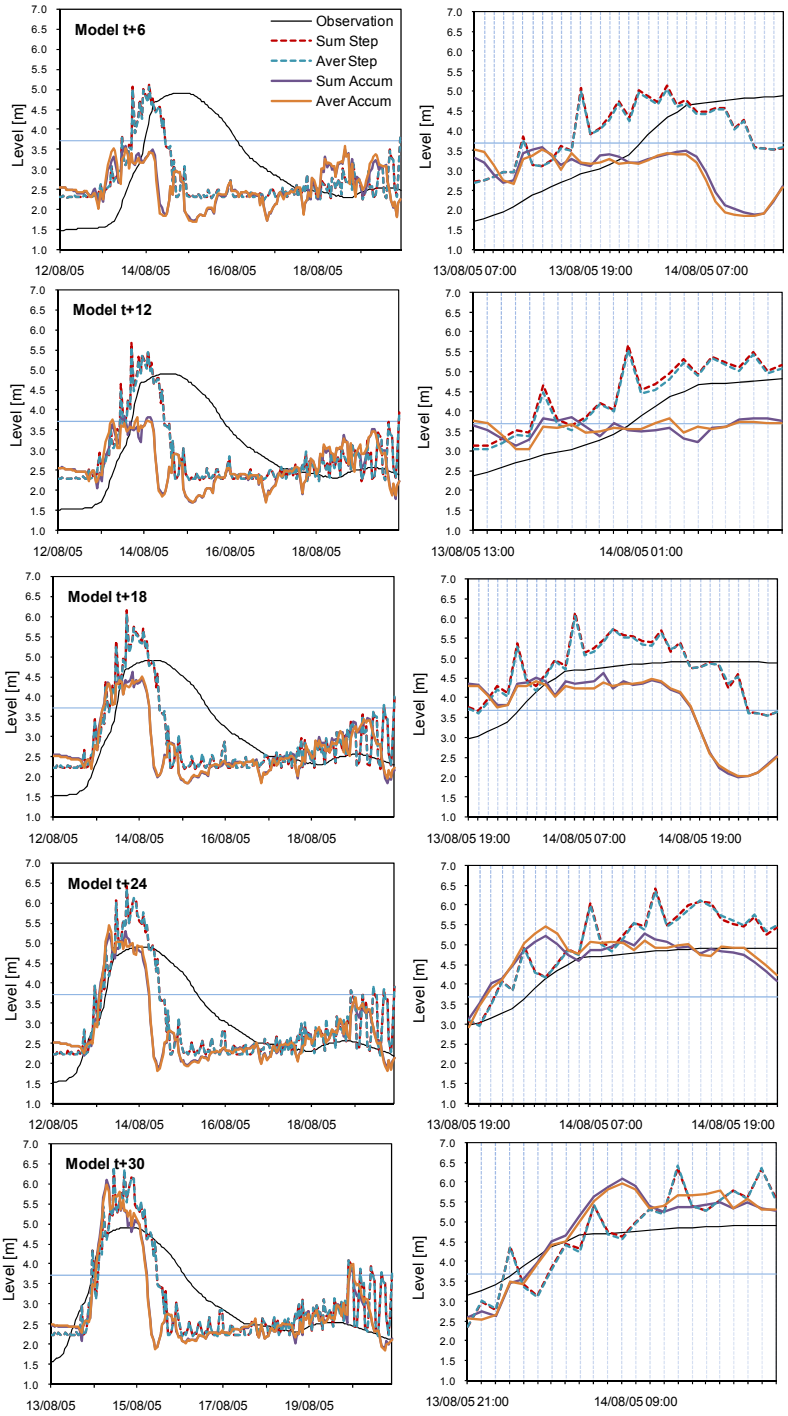


Figure 6.7: The results on storm S2 (12-20 Aug 2005) of applying four different methods of radar data extraction at lead times of 6, 12, 18 and 24 hours.

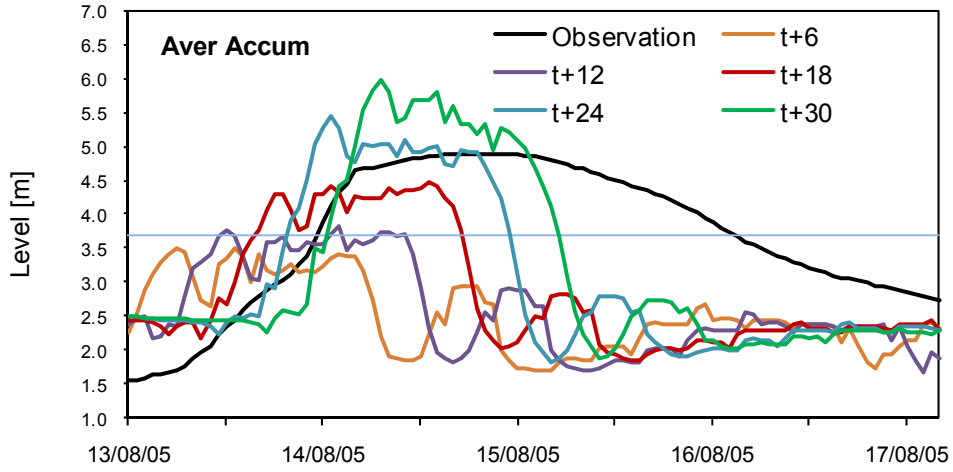


Figure 6.8: The results on storm S2 (12-20 Aug 2005) of applying one average method of radar data extraction at various lead times from 6 to 30 hours.

The best results appear to be achieved by averaging the nine radar reflectivity values and using cumulative values (i.e. the Aver Accum method). This method will, therefore, be applied in all experiments in this chapter. Figure 6.9 provides a schematic overview of how the rest of the chapter is organised in terms of different experiments using the extracted radar data. The first set of experiments are described in the next section (6.6) where the extracted radar data described in this section is used to develop NN models at lead times of 6 to 24 hours.

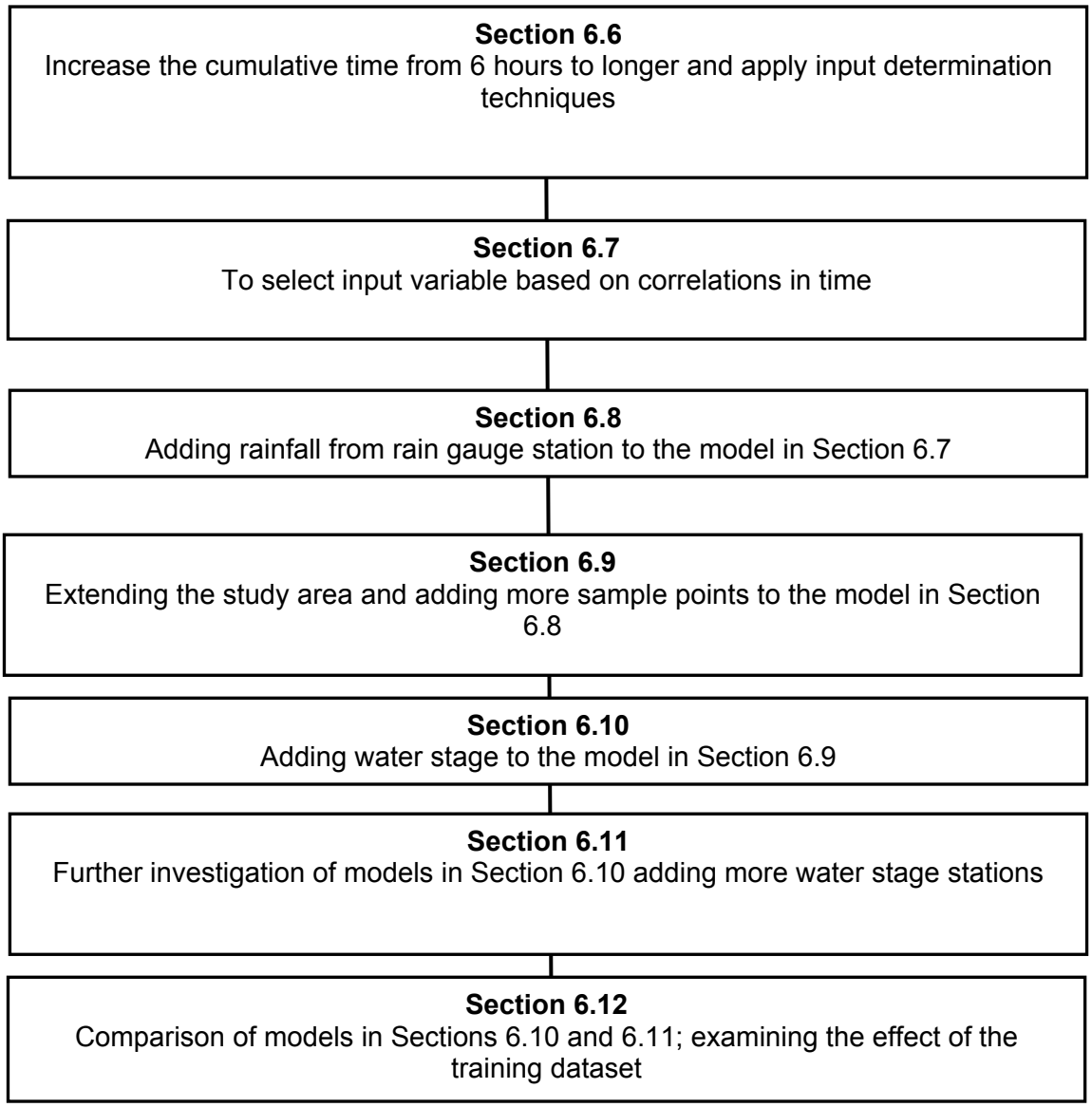


Figure 6.9: Flow diagram of the series of experiments in Chapter 6

6.6 Adding Radar Data to the Neural Network Models for Lead Times of 6 to 24 hours

This section provides the first set of initial experiments that use radar data as inputs to a neural network model for lead times of 6 to 24 hours. At this stage the point was not to extend the lead time but to simply determine if the addition of radar data had any positive effects on the resulting NN performance. From section 6.5, it was clear that the 'Aver Accum' method produced underestimations of the peaks at t+6 and t+12 hrs but improved at longer lead times. Therefore, this section will extend the cumulative interval to 12 hrs (instead of 6) and go beyond 24 hours up to t-168 hr. As a result the total number of input variables used in this set of experiments is 180 variables (12 samples*15). From Chapter 5 it was found that stepwise regression, the genetic

algorithm, correlation and the pruning algorithm were the best input determination methods for predicting the peak stage. However, the pruning algorithm was the worst method in terms of removing the number of inputs when compared with the other three methods. In addition, the genetic algorithm, correlation and the stepwise regression methods were the best in terms of overall performance based on RMSE, CE and MAE. The data mining and PMI methods were the best in terms of removing the most input variables. However, the PMI selected only the P1 variable, which did not make hydrological sense. Therefore, the input determination techniques used in this section are correlation (C), stepwise regression (S), data mining (D) and the genetic algorithm (G). All hydrographs, statistics and inputs selected can be found in Appendix E.

The experiments outlined in this section use different storms in the testing data sets to determine the model performance when using different training and testing data sets. As mentioned previously, there are 8 storms for which radar images are available. These 8 storms are divided for use in the test data set in 4 different experiments:

- Experiment 1 with S2 in the testing data set.
- Experiment 2 with S1 and S2 in the testing data set.
- Experiment 3 with S5 in the testing data set.
- Experiment 4 with S1 and S5 in the testing data set.

6.6.1 Results for Experiment 1 (Testing storm S2)

Since the main concern in flood forecasting is the time to peak and the peak prediction, only these two statistics are provided in Table 6.1. It is obvious that model D and C predicted the peak early as indicated by the negative values where both model D followed by model C chose the least number of inputs. Model C also provided the best performance in terms of peak prediction, i.e. less than a 5 cm error at lead times of 21 and 24 hours. Model S and G gave the worst model performance in this experiment, i.e. in terms of removing only 60 to 70% of input variables predicting a delay in the time of flooding.

Table 6.1: The PDIFF and flood time delay of four input determination methods, experiment 1.

Model	Statistic	Input Determination Techniques			
		C	S	G	D
t+6	PDIFF (m)	0.4997	0.2716	0.6707	0.5156
	Time delay (hr)	-13	8	10	-12
t+9	PDIFF (m)	0.1042	0.5500	0.3725	0.3799
	Time delay (hr)	-10	11	11	-9
t+12	PDIFF (m)	-0.223	0.3294	-0.1592	0.4858

	Time delay (hr)	-7.5	3	2	-6
t+15	PDIFF (m)	-0.1058	0.0521	0.3874	0.4212
	Time delay (hr)	-4.5	7	11	-4
t+18	PDIFF (m)	0.1279	0.4658	0.0414	-0.0661
	Time delay (hr)	-1	24	2.5	-2
t+21	PDIFF (m)	-0.0141	0.3426	0.0603	0.0895
	Time delay (hr)	1.5	11	14	0
t+24	PDIFF (m)	-0.0324	0.5367	0.4740	-0.3966
	Time delay (hr)	4	6	20	1

Another observation as to why only models D and C predicted the flood early is that inputs Z22 and Z42 were the most frequently selected, while variables from row Z1 were not selected and very few were selected from Z3. Models S and G, on the other hand, selected all 4 rows. All the inputs retained by the methods, the goodness of fit measures and the hydrographs on the test data can be found in Appendix E (Table E.1, E.2, E.3, Figure E.1 and E.2).

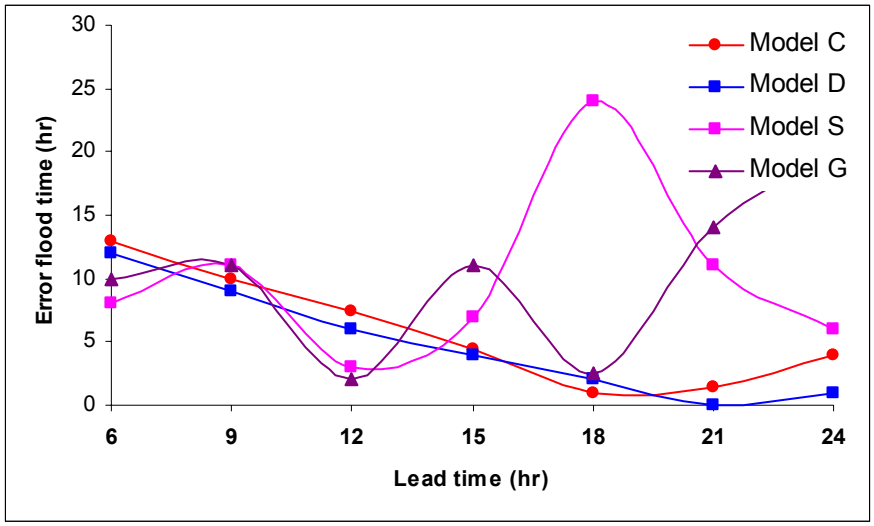


Figure 6.10: The timing error in the flood prediction on storm S2 (12-20 Aug 2005) for all four models for lead times of 6 to 24 hours, Experiment 1.

When plotting the time delay in flooding against the lead time of prediction in Figure 6.10, one can see that the error is reduced as the lead time increases. For Models C and D, this decrease in the error looks to be almost linear. Figure 6.11 shows that models C and D predict the rising limb early for lead times up to 12 hours. Both models also predicted similar patterns at the peak, which was a flat line underestimating the peak, except for model C at a lead time of 12 hours where an overestimation can be seen.

In summary, models D and C were the best input determination methods at maximum lead times of 21 and 19 hours respectively.

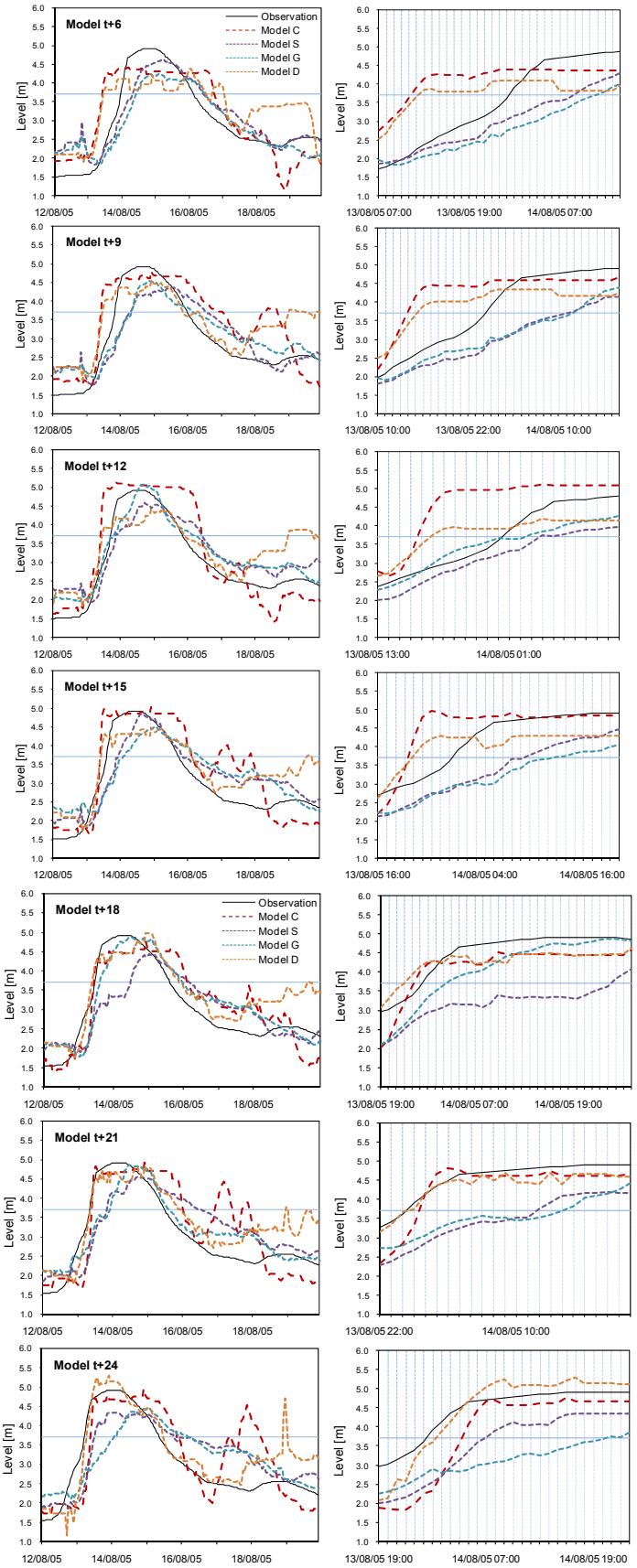


Figure 6.11: The results on storm S2 (12-20 Aug 2005) for lead times of 6 to 24 hours.

6.6.2 Results for Experiment 2 (Testing storms S1 and S2)

This second experiment considers the results on two storms in the test data set: S1 in 2003 and S2 in 2005. All models predicted an early time to rise. In particular, for the S1 storm, all models predicted values almost 24 hours early (Figure 6.12). Both models C and D predicted similar patterns and the best model for predicting storm S2 is model D, which had the best PDIFF of less than 5 cm at 12, 15 and 21 hour lead times (Table 6.2).

Figure 6.13 shows the same linear relationship when plotting the error in the time of prediction against the lead time. The model performance was improved when compared with Experiment 1, i.e. in terms of the time to peak prediction, the CE, RMSE and MAE (Table E.3 and E.6). The difference between Experiment 2 and 1 is that the 2003 storm was not included in the training dataset, which has clearly had some effect on the results.

Table 6.2: The PDIFF and flood time delay of four input determination methods, experiment 2.

Model	Statistic	Input Determination Techniques							
		S1				S2			
		C	S	G	D	C	S	G	D
t+6	PDIFF (m)	-0.4659	-0.2741	0.1164	-0.2116	0.1031	-0.4892	-0.5872	-0.1607
	Delay (hr)	/	/	/	/	-14.5	-1	-8	-14.5
t+9	PDIFF (m)	-0.4790	-0.2648	-0.4809	-0.2815	0.0847	-0.4981	-0.5497	0.0723
	Delay (hr)	/	/	/	/	-13.5	-3	-7	-13
t+12	PDIFF (m)	-0.4868	-0.0562	-0.5403	-0.5557	0.0957	-0.4309	-0.6133	0.002
	Delay (hr)	/	/	/	/	-10.5	-6	-10.5	-10.5
t+15	PDIFF (m)	-0.6731	-0.3422	-0.9349	-0.3698	0.1738	-0.7740	-0.9291	0.0133
	Delay (hr)	/	/	/	/	-8.5	-3.5	4	-8.5
t+18	PDIFF (m)	-0.7268	-0.7041	-0.8693	-0.8578	0.1809	-0.6016	-0.7213	-0.1147
	Delay (hr)	/	/	/	/	-6	-3.5	-2	-6
t+21	PDIFF (m)	-0.6724	-0.2842	-0.6856	-0.9272	0.0993	-0.2462	-1.7652	-0.0169
	Delay (hr)	/	/	/	/	-4	-3	-1	-4
t+24	PDIFF (m)	-0.7555	-0.2458	-0.6123	-1.1116	0.1155	-0.3401	-1.2584	-0.2702
	Delay (hr)	/	/	/	/	-2	-2.5	-2.5	-2

The amount of inputs removed was similar to Experiment 1, i.e. model D and C selected less input variables. However, models D and C selected more variables at Z1 and less at Z2 when compared with Experiment 1. All input variables retained by the models and the hydrographs can be found in Appendix E (Table E.4, E.5, Figure E.3 and E.4).

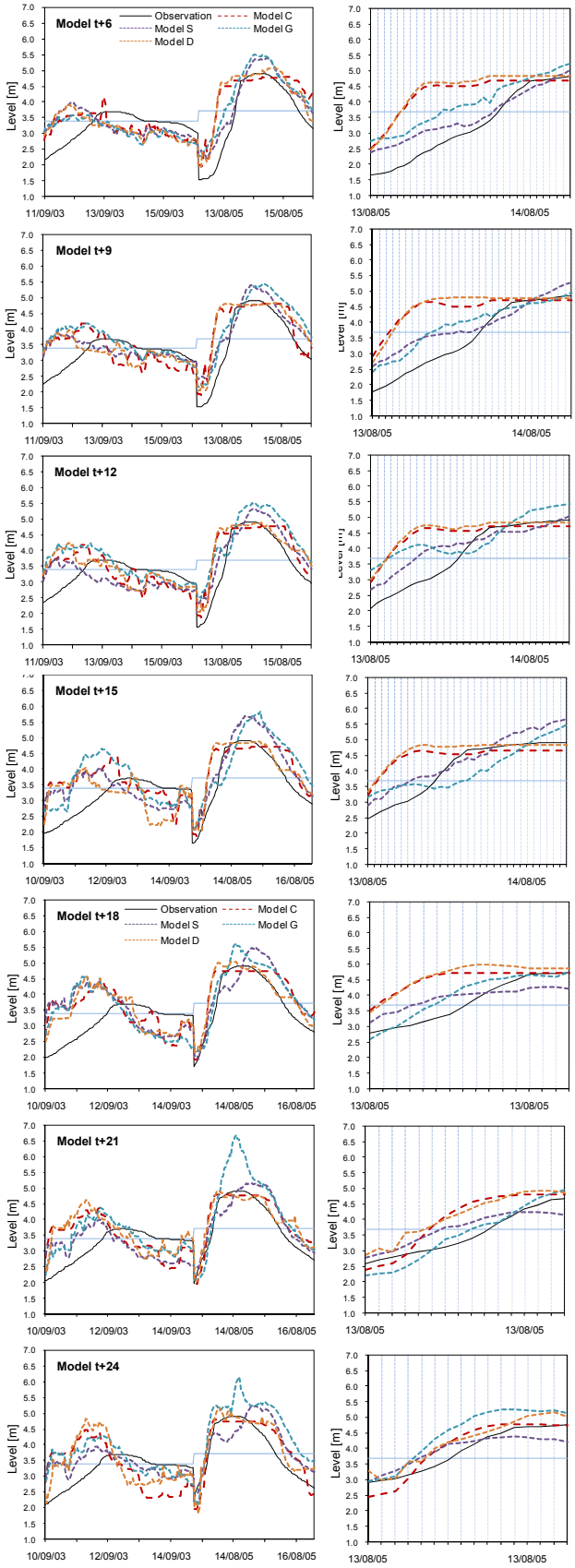


Figure 6.12: The results on storm S1 (10-15 Sep 2003) and S2 (12-20 Aug 2005) for lead times of 6 to 24 hours.

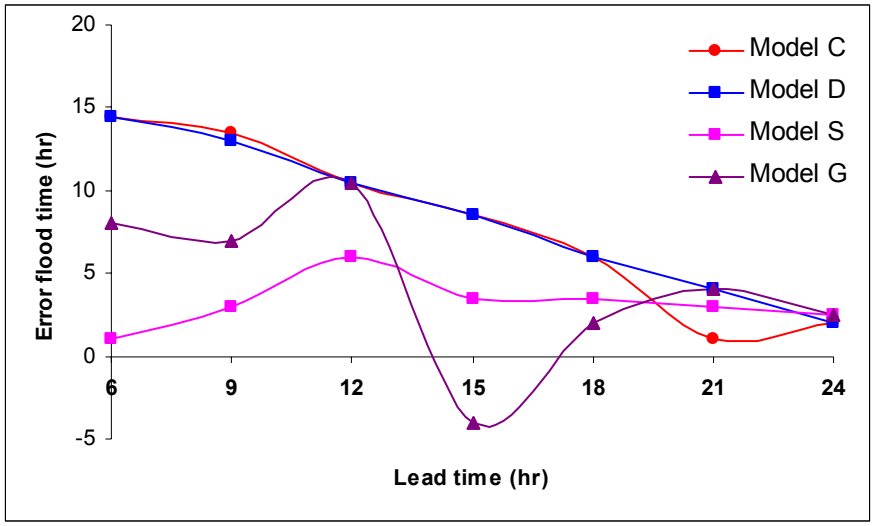


Figure 6.13: The timing error in the flood prediction of storm S2 (12-20 Aug 2005) for all four models for lead times of 6 to 24 hours, Experiment 2.

Overall it is clear that training without the 2003 storm improves the model performance when testing on the 2005 storm. The reasons could be that (1) the relationship (or rating curve) between water discharge and water stage at P1 station changed between 2003 and 2005 (Figure 3.6); (2) the flood level changed from 3.4 to 3.7 m in 2005 due to engineering works along the river; (3) the difference in moisture conditions in the catchment between the two storms. Unfortunately a NN is not able to capture the changing conditions in the catchment in the way that it has been set up here. Therefore, the neural network user could improve the model performance by selecting the right training data. This is, of course, a whole other area where research is needed, as set out in the recommendations by Maier et al. (2010).

6.6.3 Results for Experiment 3 (Testing storm S5)

The second largest storm in 2005 is tested in this experiment. All models were late in predicting the rising limb (Figure 6.14), while some models predicted an early rising limb for the first largest storm in 2005 (S2) in Experiment 1. This may be because, during storm S5, the Mae Ngat Dam (Figure 3.11) reached its maximum water capacity, so the water gate was opened and it increased the water stage and discharge in the river (Chatchawan, 2005). Another reason might be that storm S2 was the first storm of the year, so the catchment was still dry and could therefore absorb more water. In contrast, S5 is the fourth storm in that year and the catchment was mostly likely saturated, so the travel time of the water from the upper station to the lower station would be faster than in a dry catchment (S2 event). As a result, all models in Experiment 3 were late in predicting the actual rising limb.

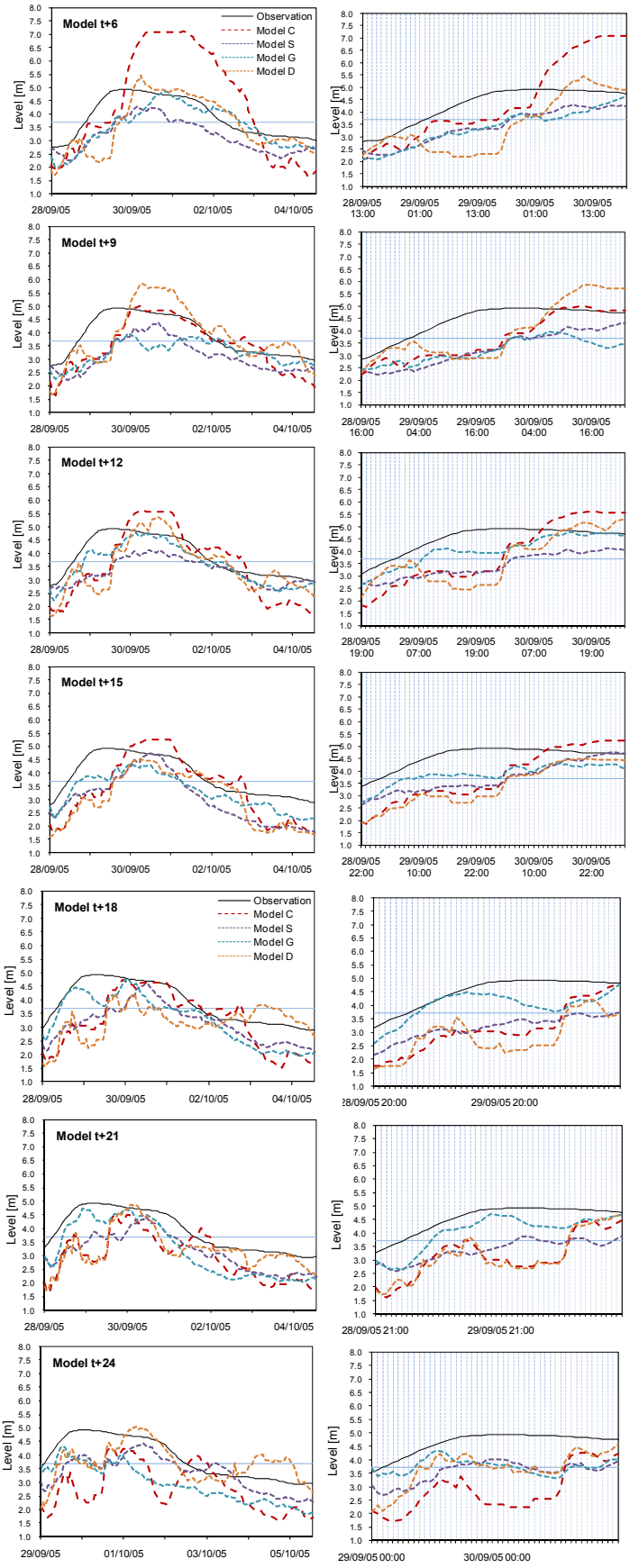


Figure 6.14: The results on storm S5 (28 Sep-4 Oct 2005) for lead times of 6 to 24 hours.

In addition, model G gave an outstanding performance in terms of predicting the rising limb while all other models performed worse (Table 6.3 and Figures 6.15). Model G had the best overall performance for PDIFF, CE, RMSE and MAE (Table E.9). Once again, when the training data included the 2003 storm, models C and D had similar patterns of variable selection similar to Experiment 1, i.e. no Z1 variables were selected and model D selected a few from Z3 (Table E.1, E.2, E.7 and E.8). All input variables retained by the models and hydrographs can be found in Appendix E (Table E.7, E.8, Figure E.5 and E.6).

Table 6.3: The PDIFF and flood time delay of four input determination methods, experiment 3.

Model	Statistic	Input Determination Techniques			
		C	S	G	D
t+6	PDIFF (m)	-2.1807	0.6316	0.1062	-0.5285
	Time delay (hr)	13	14	14	16
t+9	PDIFF (m)	-0.0814	0.5494	0.9419	-0.9162
	Time delay (hr)	18	19	19	18.5
t+12	PDIFF (m)	-0.6631	0.8142	0.1553	-0.4316
	Time delay (hr)	21	22	5	21.5
t+15	PDIFF (m)	-0.3179	0.1946	0.6263	0.4525
	Time delay (hr)	24	24	3	24
t+18	PDIFF (m)	0.1836	0.3134	0.1349	0.7101
	Time delay (hr)	27.5	28.5	2	28
t+21	PDIFF (m)	0.4521	0.4983	0.2096	0.0697
	Time delay (hr)	12	19	5	12
t+24	PDIFF (m)	0.6759	0.4947	0.6222	-0.1299
	Time delay (hr)	33	11	6	7

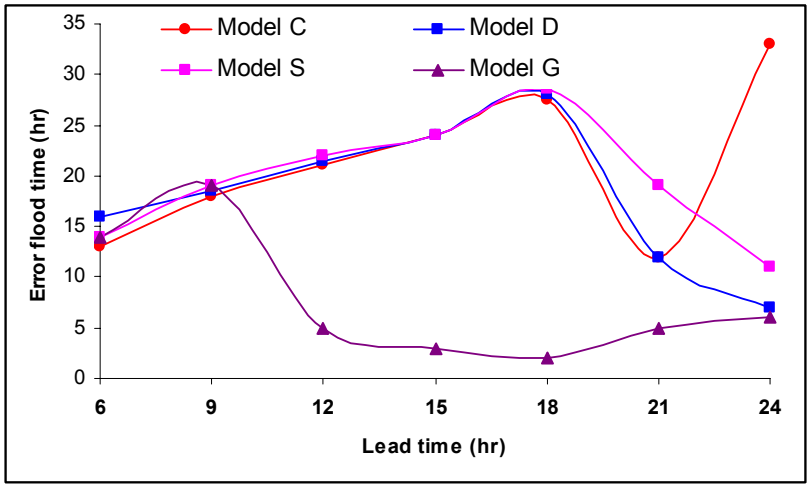


Figure 6.15: The tim timing error in the flood prediction on storm S2 (12-20 Aug 2005) for all four models for lead times of 6 to 24 hours, Experiment 3.

In summary, the catchment condition and the extra amount of water from the dam had an effect on the model performance as the model predicted a delay in the rising limb of the hydrograph. Therefore, to forecast water level in the Upper Ping catchment, the

opening times of the water gate of the dam need to be considered as part of the modelling.

6.6.4 Results for Experiment 4 (Testing storms S1 and S5)

Table 6.4 provides the performance measures when predicting the S1 and S5 storms. All models predicted the rising limb early and overestimated the peak of the S1 storm, while predicting a delay in the rising limb of the S5 storm. The inputs retained by models C and D were similar to all three previous experiments.

Table 6.4: The PDIFF and flood time delay of four input determination methods, experiment 4.

Model	Statistic	Input Determination Techniques							
		S1				S5			
		C	S	G	D	C	S	G	D
t+6	PDIFF (m)	-0.4713	-0.2113	-0.0052	-0.0961	-1.8859	0.4232	0.8615	-1.2742
	Time delay (hr)	/	/	/	/	19	8	14	14.5
t+9	PDIFF (m)	0.0564	-0.0312	-0.2589	-0.2693	-0.4528	0.6108	0.8077	-1.3528
	Time delay (hr)	/	/	/	/	24	3	7	24
t+12	PDIFF (m)	-0.1589	0.0348	-0.1823	-0.1397	0.4181	0.4789	0.2306	-0.0617
	Time delay (hr)	/	/	/	/	24	18	15	18
t+15	PDIFF (m)	-0.3278	-0.4969	-0.3295	-0.1471	0.6775	0.1734	0.4148	-0.2939
	Time delay (hr)	/	/	/	/	28	5	5	24
t+18	PDIFF (m)	-1.0656	-0.5283	0.0634	-0.5216	0.9394	0.1798	0.3887	-0.1226
	Time delay (hr)	/	/	/	/	4	4	4	28
t+21	PDIFF (m)	-0.7841	-0.3706	-0.0824	-0.5634	-0.1486	0.6263	-0.1157	0.1512
	Time delay (hr)	/	/	/	/	6	15	4	6
t+24	PDIFF (m)	-0.9861	0.128	-0.1944	-0.3879	-0.2157	-0.3681	0.7968	0.3342
	Time delay (hr)	/	/	/	/	8	-1	9	8.5

The timing errors in the rising limb are shown in Figure 6.16. There are no linear relationships like that found in the three previous experiments. However, model G and S gave the best performance especially in terms of the time to peak prediction (Table E.12). Therefore, models G and S are suitable for use in predicting floods in a wet catchment.

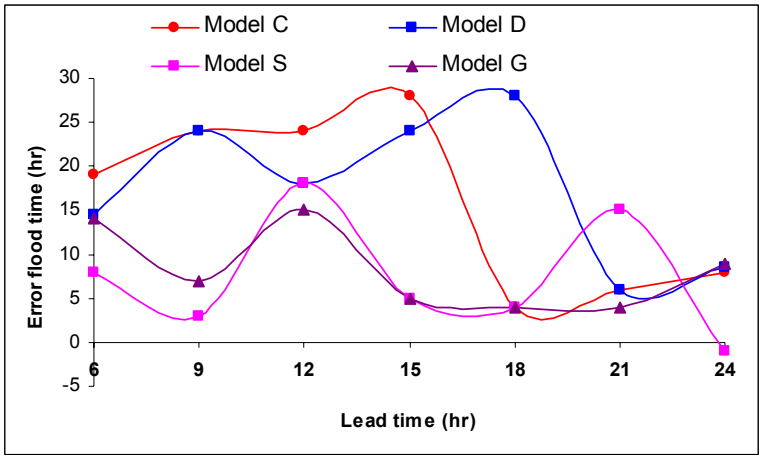


Figure 6.16: The timing error in the flood prediction on storm S2 (12-20 Aug 2005) for all four models for lead times of 6 to 24 hours, Experiment 4.

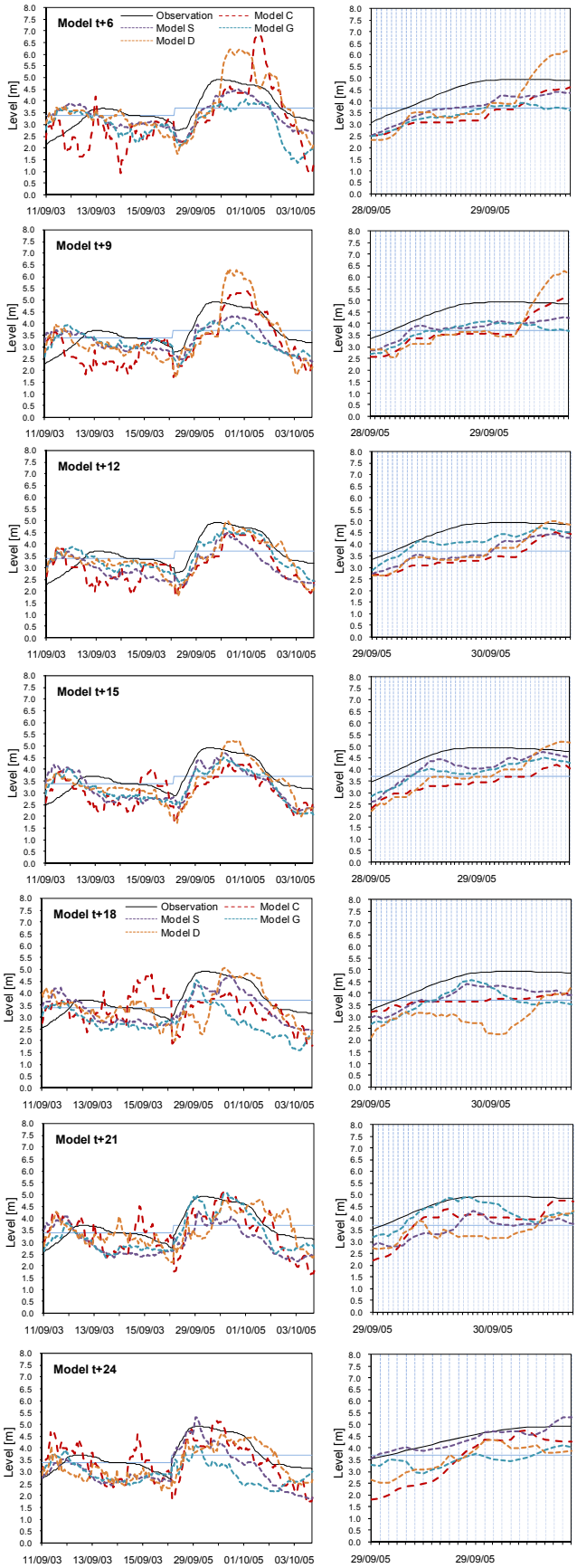


Figure 6.17: The results on storms S1 (11–16 Sep 2003) and S5 (28 Sep-4 Oct 2005) for lead times of 6 to 24 hours.

To compare the model performance in predicting the storm S1 (Experiments 2 and 4), all models predicted the rising limb while the overall performance, based on the goodness of fitness measures, was improved when training with the first storm of 2005 (S2) (Table E.6 and E.12). In contrast, when comparing the model performance when predicting storm S5 between Experiments 3 and 4, there was no significant change and all the models predicted a delay in the rising limb (Figures 6.17). The reasons were described in Experiment 3 (Section 6.6.3). All selected inputs, statistical results and hydrographs can be found Appendix E (Table E.10, E.11, E.12, Figure E.7 and E.8).

In summary, this section indicates that neural networks can employ raw radar reflectivity extracted from the radar images to predict the water level at P1 station. The maximum lead time of the forecast seems to be about 19-21 hours. However, the physical changes in the river, water from the dam and the wet/dry catchment condition influence the model performance. For example, to predict the first storm on a dry catchment, the model tends to predict the rising limb early, while the model predicts it late on a wet catchment. As different storms in the training dataset influence the model performance, the training dataset will be fixed with the same set of storms in the next section.

6.7 Extending the Lead Time of the Neural Network Forecasts

The previous section showed that radar reflectivity can be a useful input to the neural network to predict water level. However, the focus was on the development of models with lead times of between 6 and 24 hours. In this section and for the rest of the chapter, the focus shifts to developing models for longer lead times, i.e. beyond 24 hours. To do this, cumulative radar reflectivity going back longer in time was correlated with water level at P1. The next section describes how this was undertaken.

6.7.1 Calculating correlations between cumulative radar data and water level at P1

To find the correlation between the radar data and the water level at different lead times for station P1, the correlation between the time lag and the radar reflectivity values were calculated for each individual storm. The y-axis is the cumulative radar reflectivity value backwards in time and the x-axis is the time lag between the row (Z) and the P1 station at time t. Three days before and after each storm event were considered, and the correlation was examined for all 12 points. For a given row, the maximum correlation in that row was subsequently extracted, with the highest correlations occurring at the points in the middle of Figure 6.6, i.e. Z12, Z22, Z32 and

Z42. The lag time at which the maximum correlation occurs for storm S1 in 2003 can be found in Figure 6.18. For example at time t, the maximum correlation between the radar reflectivity and P1 occurred at 54 hours backwards in time for three of the points and 84 hours backwards in time for Z12. Thus the potential for extending the lead time of the forecast is clearly evident from this graph. The graph also shows a linear relationship between the time lag and the cumulative radar reflectivity value.

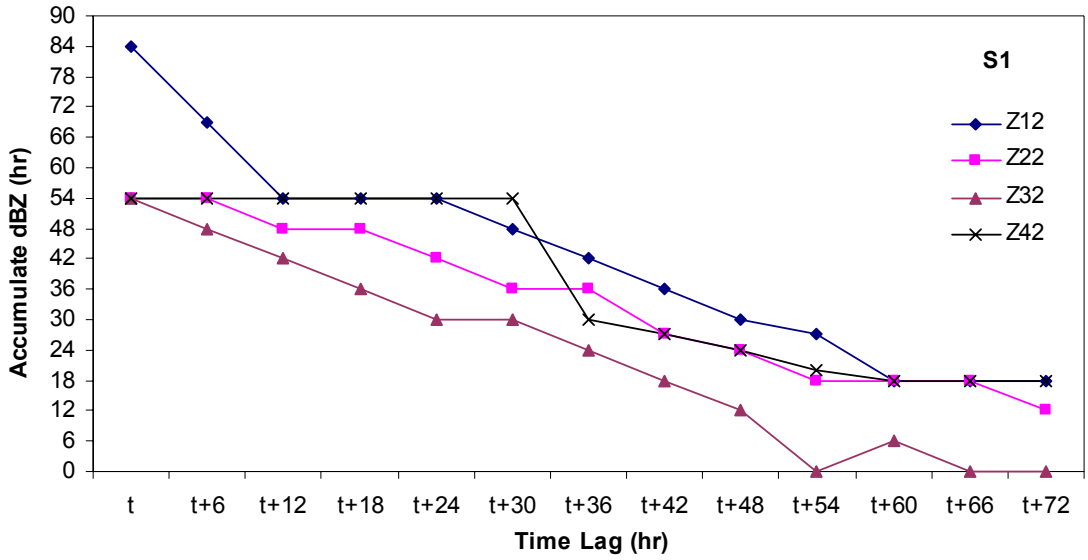


Figure 6.18: Time of maximum correlation between radar reflectivity and water level at P1 station and storm S1 in 2003.

The distance between sample points on the radar image to P1 station also influences when the maximum correlation occurs and also the magnitude of the correlation. Figure 6.19 provides the maximum time of correlation for the other storms. Storms S2, S4, S5 and S6 also appear to show the same linear relationship except for storm S3, which has a slightly longer lead time at Z12 but which is slightly shorter at Z22, Z32 and Z42 when compared with other storms.

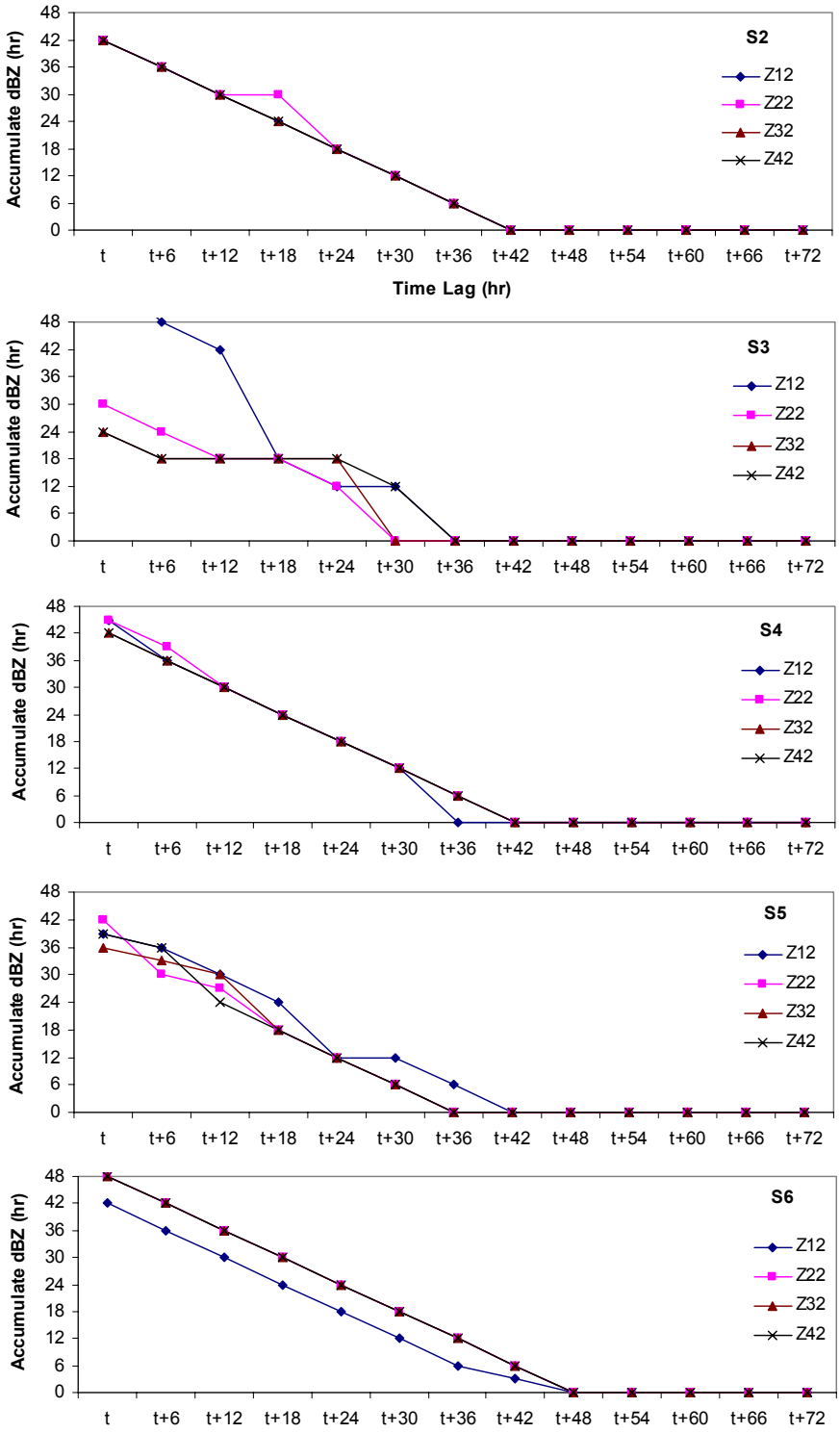


Figure 6.19: Time of maximum correlation between the radar reflectivity and the water level at Chiang Mai for storms S2 to S6 in 2005.

It is also clear from Figure 6.18 that the storm in 2003 shows different correlations to the storms in 2005 (Figure 6.19). This may be due to the fact that there were physical changes to the river (as described in Chapter 3) or that only one storm occurred in this year. The travel time of point Z12 is approximately 90 hours to P1, which may be a

result of very dry antecedent conditions. The P1 station is located at the bottom of the image so the travel time from row Z1 to P1 is the longest.

For storms in 2006, the maximum correlations occurred at points Z13, Z23, Z33 and Z43, so these are used to plot the correlations between the time lag at P1 and the cumulative radar reflectivity values. As can be seen from Figure 6.20, the year 2006 is again different than 2005. It seems that at P1 station, time t has a high correlation with 24 to 42 hours cumulative radar reflectivity values. Moreover, the correlations are not linear, which might be related to shorter, less intense storms covering the area.

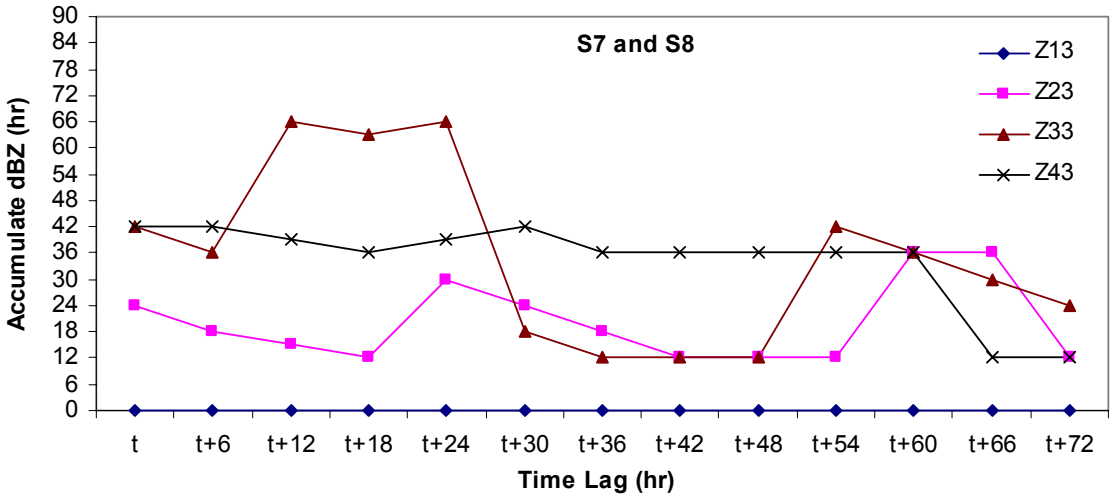


Figure 6.20: Time of maximum correlation between radar reflectivity and water level at Chiang Mai for storms S7 and S8 in 2006.

However, the results so far indicate that a 42 hr lead time could potentially be achieved with inputting radar data into a neural network, especially for the large storms in 2005. Moreover, the radar reflectivity values at all 12 points have similar correlations with water level at P1. Figure 6.19 also shows that the cumulative radar reflectivity of the radar image at 42 and 48 hr has a linear correlation with water level at time t at P1. Therefore, it should be possible to extend the forecast lead time of the neural network to 36 to 42 hours at P1 using the 6 hr cumulative radar values. The next section describes how these correlations are used to determine which data to extract, which are then used in the development of the neural network models.

6.7.2 Development of Neural Network Models based on the Correlated Inputs

In section 6.6 it was found that the use of different training datasets influences the model performance. Thus to avoid this effect, the training dataset was fixed from 25 August to 9 November 2005 (4 storms; S3-S6) for all models. The testing dataset comprised the following storms: the first large storm in 2005 (2 to 24 August 2005, S2),

the only storm in 2003 (6 to 23 September 2003, S1) and two small storms in 2006 (18 September to 20 October 2006, S7 and S8). In addition, the the correlations in Section 6.7.1 show that the neural network has the potential to predict the water stage at P1 at a lead time of greater than 24 hours. In this set of experiments, therefore, the lead times are extended to 24, 30, 36, 42 and 48 hours.

The input selection is based on the highest correlations between P1 and the radar reflectivity values at the 12 sample points. From Section 6.7.1, it was shown that P1 correlates with cumulative data from 24 to 72 hours for all 3 years of storms. Moreover, the points at Z11, Z12 and Z13 are less consistent so they are no longer used, leaving only 9 out of 12 points to be used as input variables. For each lead time, there were high correlations at different time periods in the past. For example, high correlations exist between cumulative radar 6 hours in the past and the level at P1 24 hours ahead. However, to simplify the modelling, only points at one time period in the past were used to develop the models. Thus for each lead time, multiple models were developed (where the 9 points were used at each time in the past to develop one model). Otherwise there would have been too many inputs to include for each model, many of which may have been redundant and therefore effectively have reduced the performance of the network. The list below shows each of these models by lead time:

Lead Time	**Different time periods with a high correlation between the cumulative radar data and P1**
24 hours	T-6 hours T-12 hours T-18 hours T-24 hours
30 hours	At time t T-6 hours T-12 hours T-18 hours
36 hours	At time t T-6 hours T-12 hours T-18 hours
42 hours	At time t T-6 hours T-12 hours T-18 hours
48 hours	At time t T-6 hours T-12 hours

The sub-sections that follow describe the modelling results on each of the storms in the testing data set.

6.7.2.1 Testing Storm S2

Figure 6.21 shows the results of testing storm S2 in 2005 for lead times of 24 to 48 hours based on correlations with different time periods. It is clear that the neural network model predicted the rising limb very accurately at lead times of 24 and 30 hours using inputs at time t or t-6. Predictions of the peak were one hour early and peak predictions were only 0.1 cm above the actual peak (for a 24 hour lead time) (Table E.13). However, when data from further back in time are used, the performance decreases and the hydrograph is effectively shifted, producing late predictions (the full range of hydrographs is given in Figure E.9).

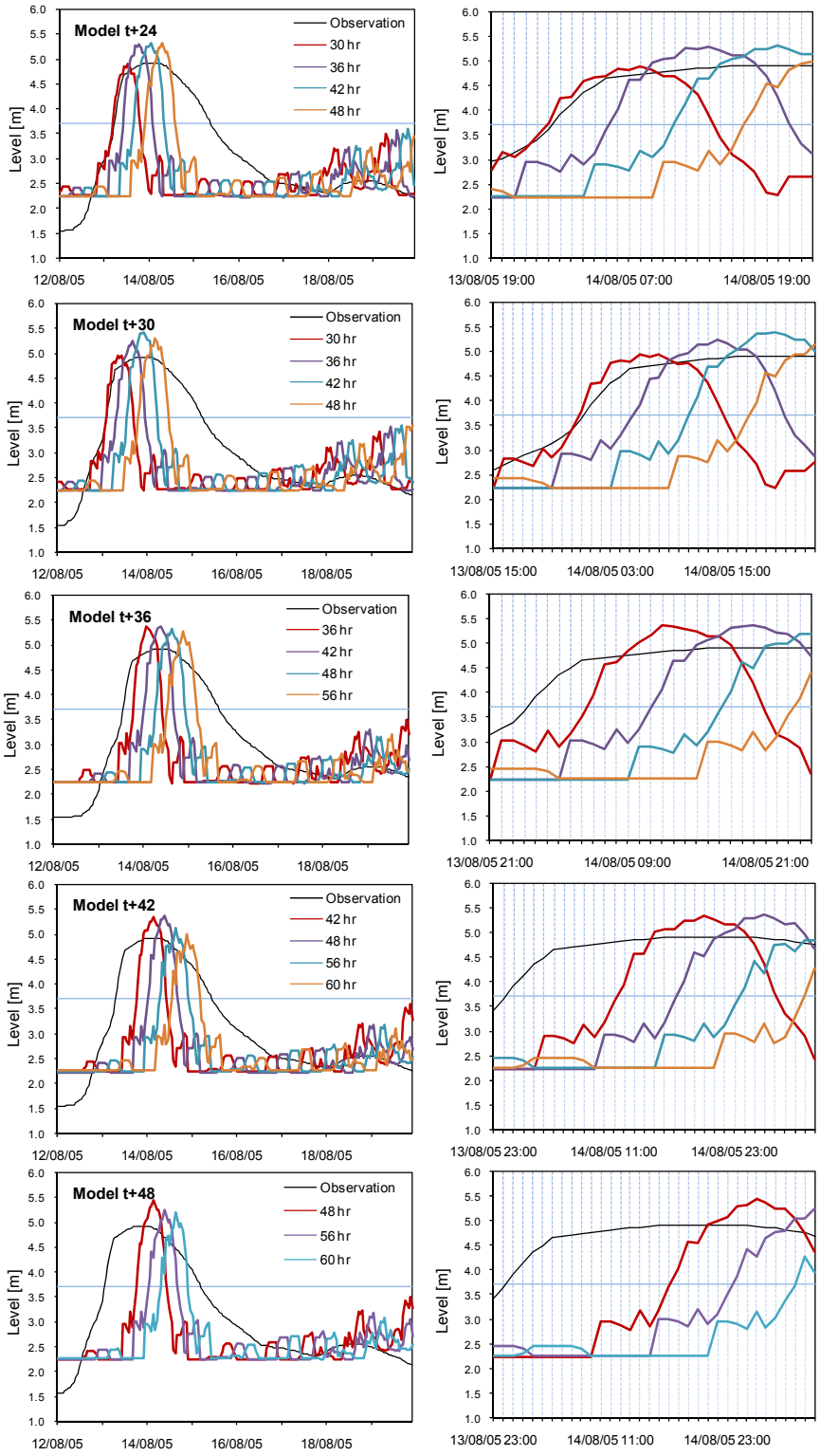


Figure 6.21:Comparison of the hydrograph predictions for different time correlations at lead times of 24 to 48 hours, testing on storm S2 (12 – 20 Aug 2005).

Figure 6.22 provides the hydrographs when using different input variables and training datasets. Some additional notation is added here to aid the reader in understanding the figures. Train 01 (which was the model developed in Section 6.7.2) used 9 input variables as this was based on the correlations at one time step only. The models

referred to as Train 02 and Train 03 selected input variables that were based on input determination techniques (which are related to Experiments 1 and 2, Section 6.6). Model Train 01 is the only model that predicted very close to the actual peak but did not perform well on the falling limb of the hydrograph. This is because the model referred to as Train 01 in this section is based on cumulative radar reflectivity for only 6 hours while the other models use cumulative values up to 168 hours (Table E.1, E.2, E.4 and E.5).

The bottom graph in Figure 6.22 shows more details at the peak with different training datasets and it is clear that all models trained without the 2003 storm (Train 01 and 02) predicted the rising limb much earlier than the actual time of occurrence (see Figure E.11 for the full range of hydrographs).

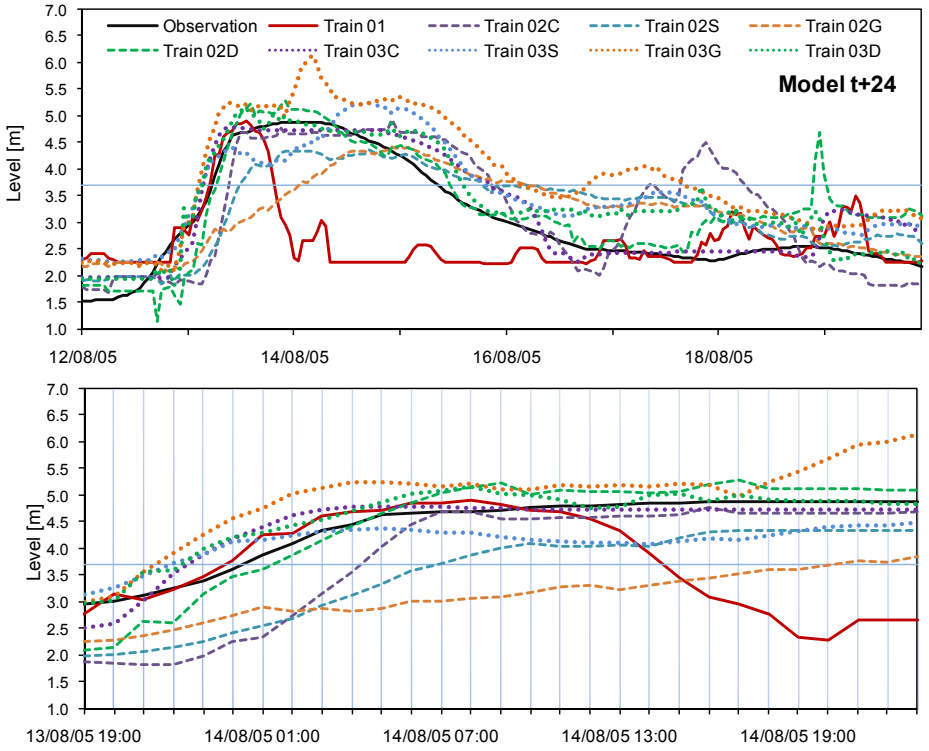


Figure 6.22: Compare of the hydrographs of different training data at 24 hour lead time, test S2 (12 – 20 Aug 2005).
Note: Train 01 (S3 – S6), train 02 (S1, S3 – S8) and train 03 (S3 - S8)

In summary, the neural network that predicted the best result at the peak and the rising limb was that using 9 input variables at t-6. In this case, other input determination techniques were not needed and the training time for the neural network was reduced.

6.7.2.2 Testing Storm S1

Figure 6.23 shows the hydrograph results of testing on storm S1. All the models predicted two peaks before and after the actual peak. However, the best model appears to be the one predicting at a lead time of 48 hours. Moreover, all models in this section and Section 6.6, which used only radar data as inputs could predict the water stage more than 24 hrs in advance while models using only water stage could not predict more than 24 hrs ahead. However, using radar data only as inputs meant that no models were able to produce good hydrograph shapes (see Figure E.10 for the full range of hydrographs).

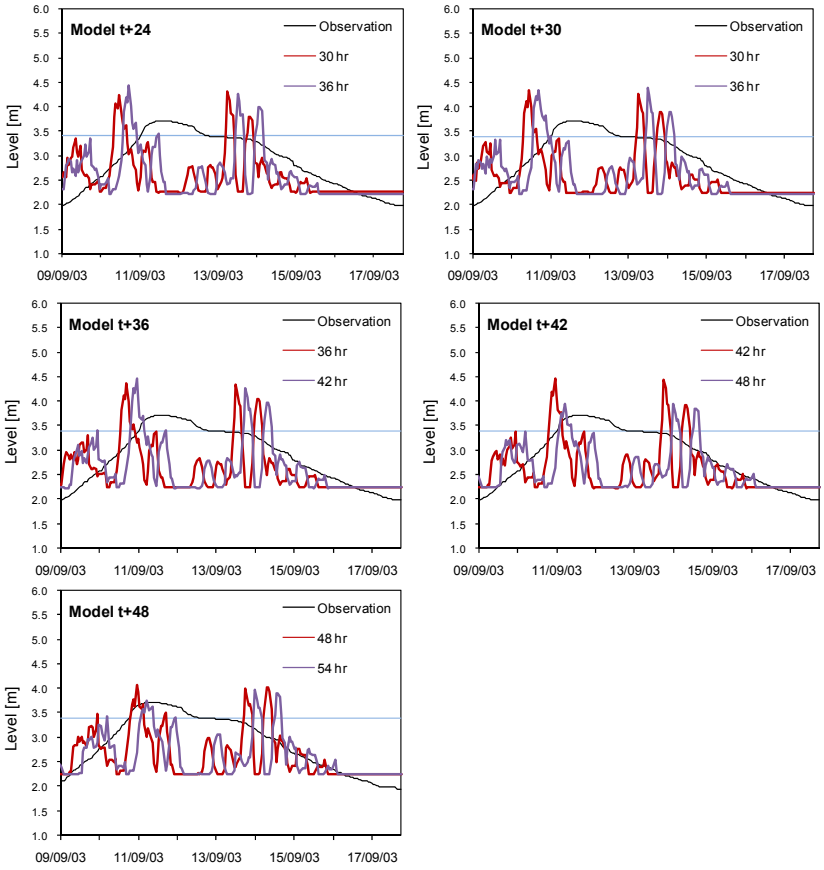


Figure 6.23: The results on storm S1 (9–18 Sep 2003) of different time correspondences at lead time of 24, 30, 36, 42 and 48 hours.

Figure 6.24 provides a comparison of the hydrographs between the three experiments; Train 01 is in this section, while Train 02 and 03 are Experiments 2 and 3 in Section 6.6. All models predicted the rising limb too early at a lead time of 24 hours. (See Figure E.12 for the full range of hydrographs).

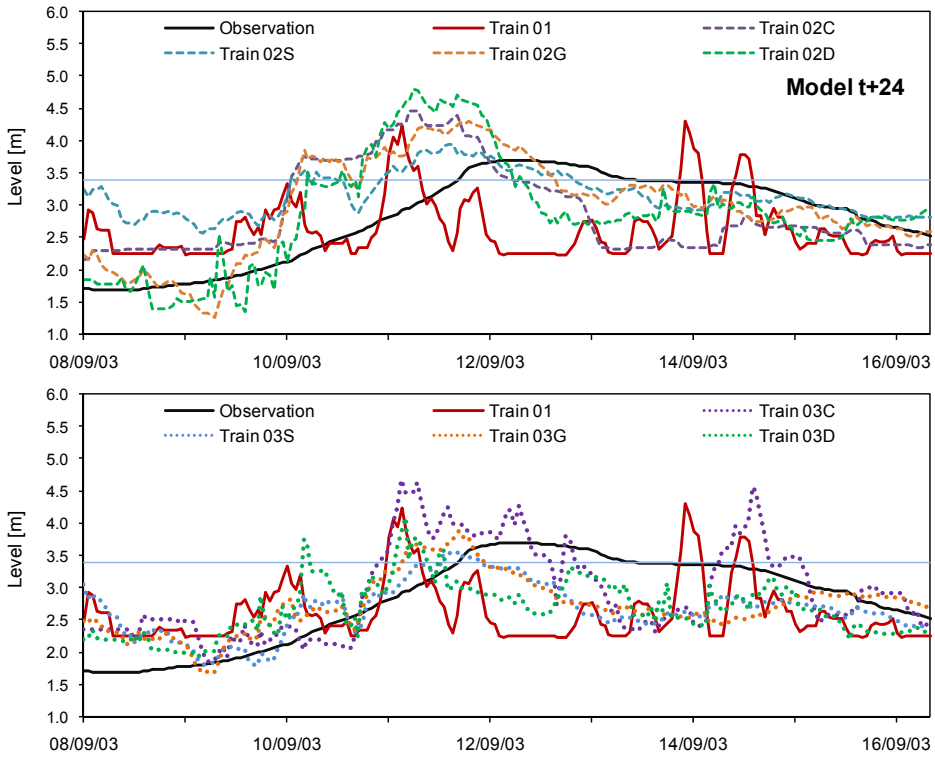


Figure 6.24: Compare of the hydrographs of different training data at 24 hour lead time, test S1 (2003), top compare train 01 and train 02, bottom compare train 01 and train 03. Note: Train 01 (S3 – S6), train 02 (S3 - S8) and train 03 (S2 - S4, S6 – S8)

In summary, when testing with the storm in 2003, the models predicted the rising limb too early, similar to previous experiments in Section 6.6. However, this section uses only 9 input variables based on simple correlations at time t and does not apply input determination techniques so the approach has advantages of simplicity.

6.7.2.3 Testing Storms S7 and S8

The results of testing two small storms in 2006 are shown in Figure 6.25. It seems that the model is unable to predict any peaks in the small storm. All the models were poor as they only predicted very small peaks. It may be that the radar image showed no rainfall during these events or over the sample points. However, there was heavy rain near Z4 and next to P1. This may point to one of the main disadvantages of predicting water level from radar images. Alternatively, the storm pattern may have been different, i.e. the storm may have come from a southwesterly direction instead of the more predominant one.

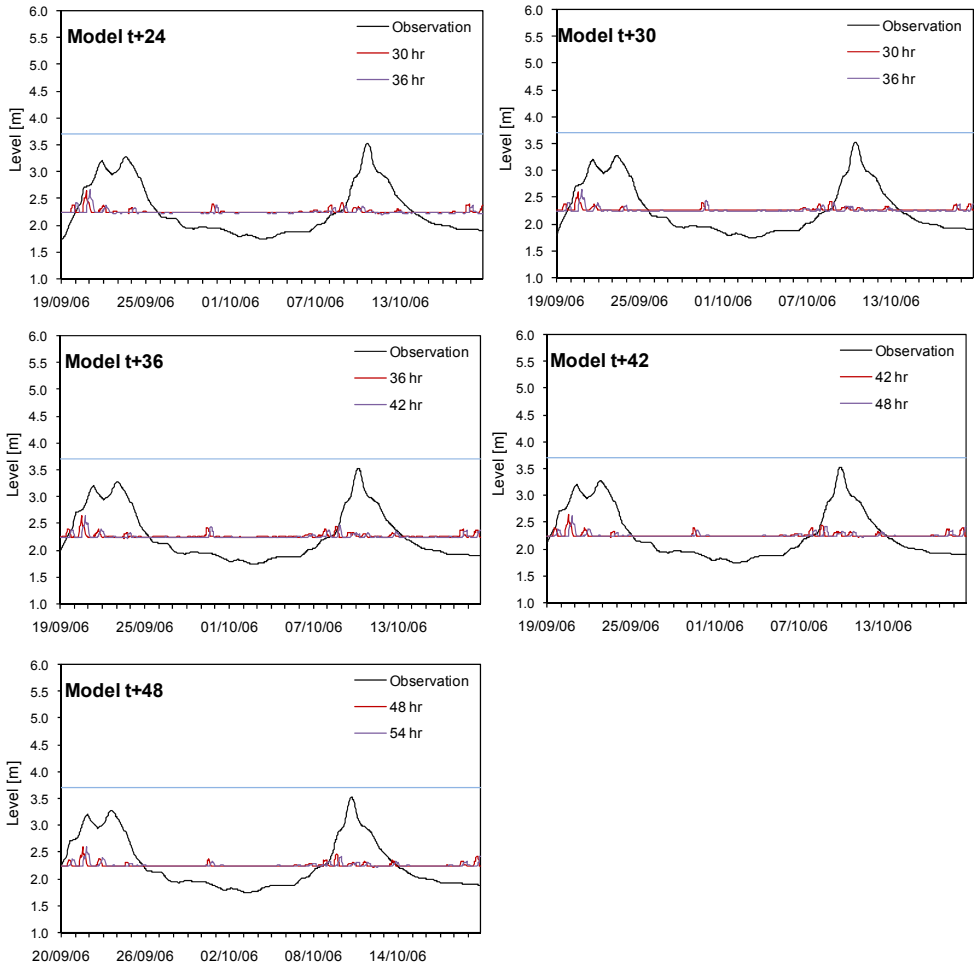


Figure 6.25: The results on storm S7 and S8 (19 Sep-20 Oct 2006) of different time correspondences at lead time of 24, 30, 36, 42 and 48 hours.

Since the neural network would not have learned a storm pattern like that seen in 2006, it might simply not have been able to predict such an unseen event before. For this reason, one of the storms in 2006 was provided to the neural network in the training dataset. Figure 6.26 shows that adding one 2006 storm resulted in some increases in level prediction but the performance is still poor.

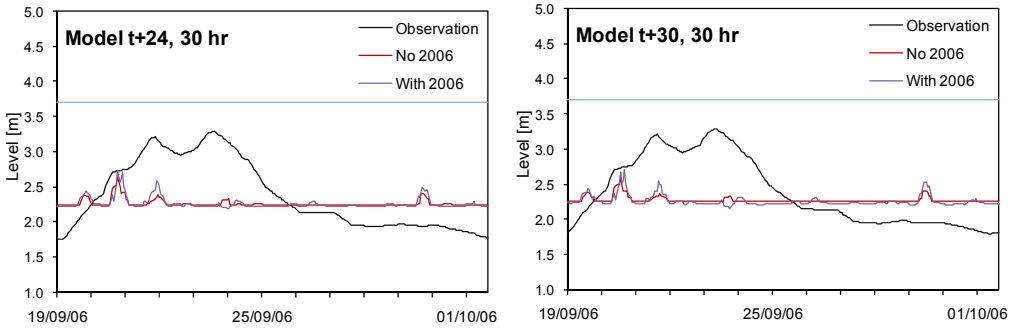


Figure 6.26: The results on storm S7 (19 Sep–2 Oct 2006) of different time correspondences at lead time of 24 and 30 hours.

The pattern of the storm movements and the location of the sample points may have an effect on model performance. However, the model performance might be improved if the storm movement pattern could be included in the training dataset. Also more sample points could be added (which is investigated later in Section 6.9).

In summary, the neural network can predict the water level at P1 by using only cumulative radar reflectivity over the study area with 9 spatially distributed input variables across the image. These examples used only simple correlation as the method for choosing the best inputs as many different combinations of inputs were tested. The maximum lead time for prediction is 48, 30 and 24 hours for storms in 2003, 2005 and 2006 storm, respectively. In addition, the model produces reasonable predictions if the storm type in the training and testing dataset are similar and if rain fell over the sample points (which would therefore have been picked up in the radar images). Otherwise the model performance was poor. The next section investigates the effect of adding rainfall data in addition to the radar data to the neural network models.

6.8 Adding Radar and Rainfall Data to the Neural Network Models

At this stage, only radar data have been used in developing the neural network models. This section considers whether the addition of rainfall data to the neural network models developed in Section 6.7 might improve the model performance. Neural network models were developed with nine radar inputs and data from one rain gauging station.

The results shown in Figure 6.27 for storm S2 indicate that adding rainfall to the models from the rain gauge near P1 did not improve the model performance, either from the hydrographs or through examination of the performance statistics (Table E.16). If the rain gauge was located higher up in the catchment, this input may potentially have had an effect. This once again draws attention to the sparse network of rainfall data in the catchment.

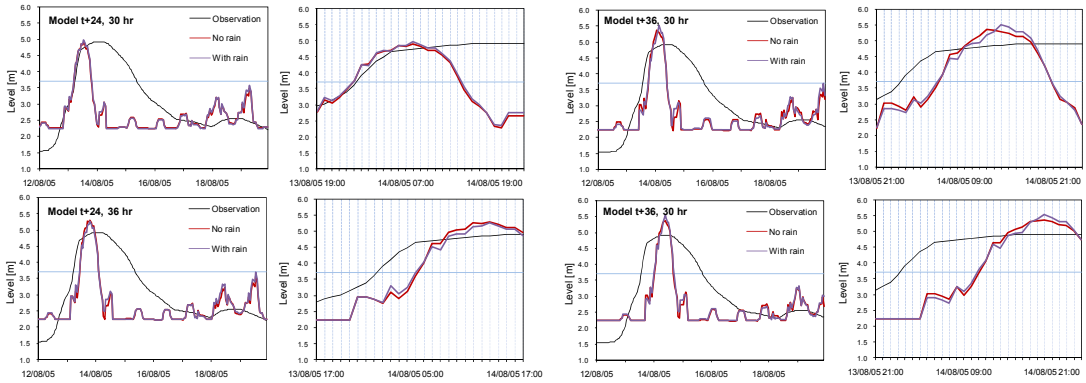


Figure 6.27: The results on storm S2 (12–20 Aug 2005) training with and without rainfall for lead times of 24 and 36 hours.

The next section considers what happens when the sample area is extended and if more sample points are added.

6.9 Extending the Sample Area and the Number of Sample Points

The initial experiments in Section 6.7 were based on sampling the radar image at a limited number of points. This is potentially helpful in limiting calculation times but may also limit accuracy. In this section the area of the radar image was expanded and a greater number of sample points was added, i.e. 25 as shown in Figure 6.28. However, the correlations between P1 and all 5 points in the first row (i.e. Z10 to Z14) are low, so these have not been used in any further experiments. This is probably because any rainfall here would be drained out of the river and the land use type at row Z1 is mountainous (Figure 3.10).

Figure 6.28: Extended area on the radar images with extra sample points.

The inputs to the neural network were rainfall from rain gauge R1 and 20 values extracted from the radar image. The results of the model are shown in Figure 6.29 for

storm S2. It is clear that the prediction of the time of the rising limb for a lead time of 24 hours has improved by 4 hours; the CE has also increased from 0.6409 to 0.6446.

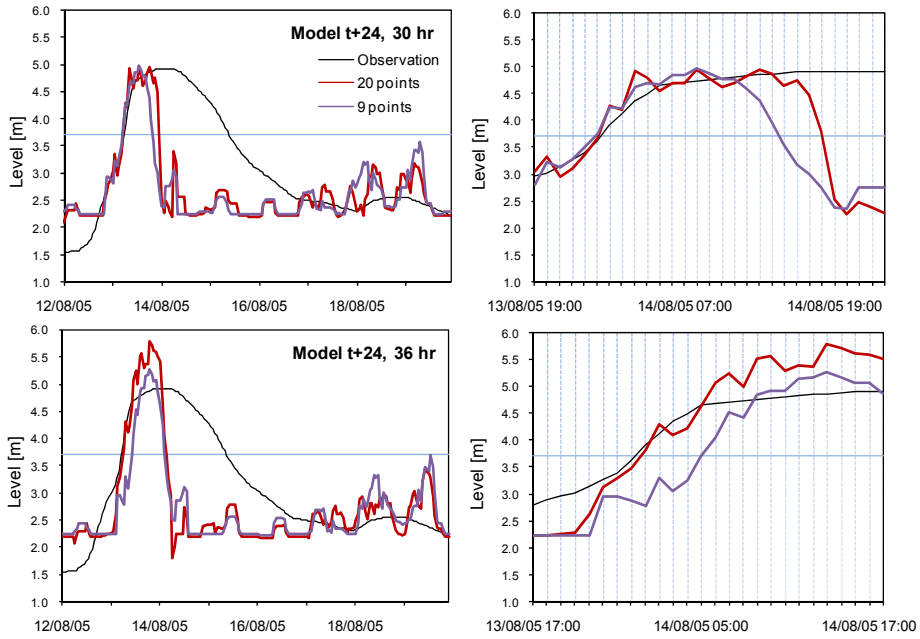


Figure 6.29: The results on storm S2 (12–20 Aug 2005) for 9 and 20 sample points at a lead time of 24 hours.

In summary, the addition of extra sample points has increased the model performance so this extra information has been used by the neural network for further experimentation in this chapter. The next section considers the effect of adding water level data from additional upstream stations in the catchment.

6.10 Addition of Water Stage from More Upstream Gauging Stations

The results so far show that adding radar data accumulated over a period of time in the past at points spatially distributed around the area of the river can improve the lead time of the neural network (Chaipimonplin *et al*., 2010). Due to physical changes in the channel in 2004 (Chatchawan, 2005), further experiments in this section will not use data before this period. Instead, a total of seven storms (2005; S2 – S6 and 2006; S7 – S8) will be used in different combinations of training and testing. Adding rainfall did not improve the model performance as shown in Section 6.8 but this section will further investigate the effect of adding both rainfall data and water level data to see whether these together improve the model performance. Two further water level stations (P21 and P4a) from the upper part of the catchment will be added. This section will, therefore, investigate the effect of adding more data to the neural network. Moreover, stepwise regression will also be used to select the most suitable input variables.

This section consists of three experiments. The first experiment uses data from the radar image, the rain gauge R1 and three main flow stations (P75, P67 and P1). The total number of inputs is 35: 20 radar points, 3 variables from 1 rain gauge (cumulative rainfall in 6 hour windows from time t) and 12 variables from 3 water level stations (t, t-6, t-12 and t-18). The second experiment is similar to the first experiment but 2 more water level stations (P21 and P4a) were added, increasing the total number of inputs to 41. In the third experiment, stepwise regression was used to select the input variables from the first experiment (Table 6.5).

Table 6.5: Input variables selected by Stepwise regression (denoted by x) for 18, 24 and 30 hour lead time.

Leadtime	18							24								30							
Storm	S2	S3	S4	S5	S6	S7	S8		S2	S3	S4	S5	S6	S7	S8		S2	S3	S4	S5	S6	S7	S8
P1t	x	x	x	x	x	x	x	P1t	x	x	x	x	x	x	x	P1t	x	x	x	x	x	x	x
P1t-6	x	x	x	x	x	x	x	P1t-6	x	x		x	x	x	x	P1t-6	x	x	x	x	x	x	x
P1t-12	x	x	x	x	x	x	x	P1t-12			x				x	P1t-12	x	x				x	
P1t-18			x	x	x			P1t-18	x							P1t-18		x					
P75t	x	x	x	x	x	x	x	P75t	x	x	x	x	x	x	x	P75t	x	x	x	x	x	x	x
P75t-6	x	x	x	x	x	x	x	P75t-6	x	x	x	x	x	x	x	P75t-6	x	x	x	x	x	x	x
P75t-12	x		x	x	x	x		P75t-12	x		x	x	x	x	x	P75t-12	x	x		x		x	
P67t-18	x		x					P67t-18		x	x				x	P67t-18						x	
P67t	x	x	x	x	x	x	x	P67t	x	x	x	x	x	x	x	P67t	x		x	x	x	x	x
P67t-6	x	x	x	x	x	x	x	P67t-6	x	x	x	x	x	x	x	P67t-6	x		x	x	x	x	x
P67t-12	x	x	x	x	x	x	x	P67t-12	x	x		x	x	x	x	P67t-12	x						x
P67t-18		x	x	x	x	x	x	P67t-18				x	x	x	x	P67t-18							
12Z20								6Z20		x	x	x	x	x		Z20		x	x	x	x	x	x
12Z21	x	x	x	x		x	x	6Z21	x	x						Z21	x	x	x	x	x	x	x
12Z22	x			x	x			6Z22		x	x	x	x	x	x	Z22		x	x	x	x	x	x
12Z23		x	x	x	x	x	x	6Z23		x	x	x	x	x	x	Z23		x	x	x	x	x	x
12Z24	x	x	x	x	x	x	x	6Z24	x		x		x	x	x	Z24	x		x	x	x	x	x
12Z30	x	x	x	x	x	x	x	6Z30	x		x			x	x	Z30	x	x	x	x	x	x	x
12Z31		x		x				6Z31			x		x			Z31	x	x	x	x	x	x	x
12Z32	x				x	x	x	6Z32	x			x	x	x	x	Z32	x	x	x		x	x	x
12Z33		x	x					6Z33			x					Z33	x	x			x	x	x
12Z34	x	x	x	x		x	x	6Z34			x					Z34		x	x	x	x	x	x
12Z40	x		x		x			6Z40		x	x					Z40	x		x	x		x	x
12Z41								6Z41		x	x	x	x	x	x	Z41	x			x		x	x
12Z42								6Z42	x						x	Z42	x	x		x			x
12Z43		x	x			x	x	6Z43		x	x			x	x	Z43	x	x	x	x	x	x	x
12Z44	x	x		x				6Z44	x							Z44	x	x	x	x		x	x
12Z50		x					x	6Z50	x	x			x	x	x	Z50	x	x	x		x	x	x
12Z51	x				x	x		6Z51	x							Z51	x						
12Z52	x	x		x	x	x	x	6Z52	x	x		x	x	x	x	Z52	x	x	x	x	x	x	x
12Z53	x	x	x	x	x	x	x	6Z53	x	x	x	x	x	x	x	Z53	x	x	x	x	x	x	x
12Z54	x			x	x			6Z54	x		x	x	x	x	x	Z54	x	x	x	x	x	x	x
R1	x	x	x	x	x	x	x	R1	x	x	x	x	x	x	x	R1	x						
6R1	x	x	x	x	x	x	x	6R1		x	x		x	x	x	6R1	x	x	x	x	x	x	x
12R1	x			x			x	12R1	x		x	x	x	x	x	12R1	x	x	x	x	x	x	x
Total (35)	25	23	23	25	23	22	22		22	20	24	19	23	24	26		28	25	24	25	23	29	28

6.10.1 Testing Storm S2

The results of testing the first storm 2005 (S2) are shown in Figure 6.30. It is clear that adding more water level stations improved the overall performance especially in the falling limb of the hydrograph. For example, when compared with the model using only radar (Section 6.7.2) at lead times of 24 and 30 hours, CE increased from 0.1543 to

0.7573 and 0.1775 to 0.5332, respectively (Table E.13 and E.17). However, the predicted lead time was reduced from 30 to 17 hours (Figure 6.21). The reasons for this could be the limitation of travel time from the upper water level stations to P1. In addition, combining water level with radar data improves the prediction compared to using only water level alone (see Case Study 2, Chapter 5).

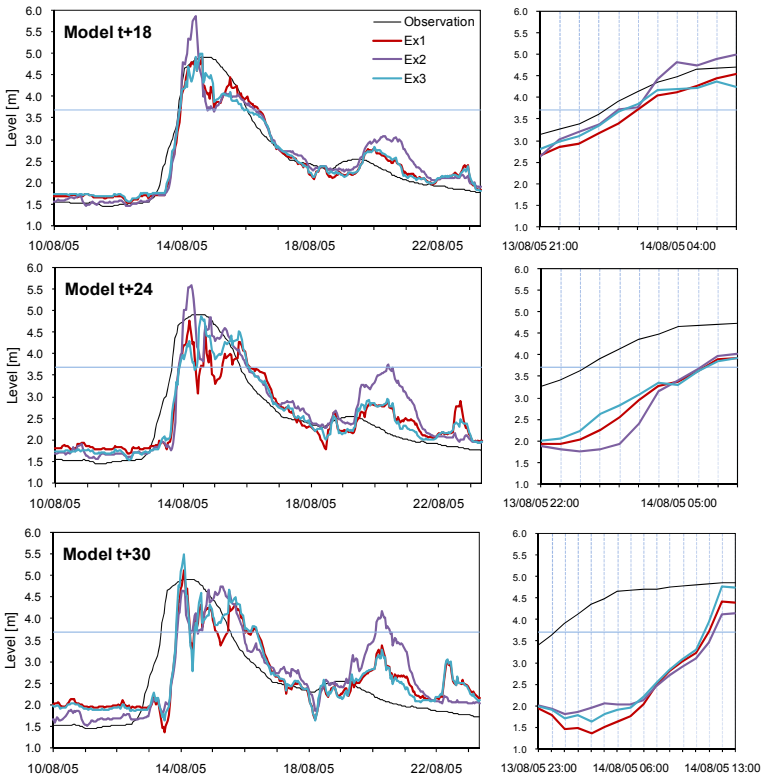


Figure 6.30: The results for storm S2 (10–23 Aug 2005) when using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

It can also be seen that the models appear to predict the events quite well when compared with using only radar data as inputs to the model. The results show that all three experiments predicted similar times to peak but only the model in the second experiment predicted higher than the actual peak at lead times of 18 and 24 hours, which might be the effect from the two extra water level stations. In addition, the prediction at a lead time of 18 hours was the most accurate with less than a 50 cm error in the peak prediction and a 1 to 3 hr delay in the timing. The third experiment (where the input variables were selected with stepwise regression) also gave the best accuracy in terms of the timing of the peak although there was still some delay. The error in peak prediction and the performance statistics are provided in Table E.17.

By the standards for forecasting in European rivers with many stage gauges and good rainfall data, these results are bad. However, in the Upper Ping the data are very

limited and improvements in forecasting are relative. Using the radar image data alone gives a better lead time prediction for a large event but the falling limb was poorly predicted. In contrast, using only water level gives a better falling limb but a reduced lead time. Moreover, integrating water level and radar data gave a better result at t+18 than using water level alone but at a longer lead time, the results for predicting the rising limb were worse. However, adding in stage data for two more water level stations, which are not located on the main river, did not to improve the model performance.

6.10.2 Testing Storms S3 and S6

S3 and S6 are the small storms in 2005 and all the models showed similar patterns, including some noisy predictions, and early and over predictions at the peaks (Figure 6.31, 6.32). This may be due to the fact that there were clouds over the sample points but these did not result in precipitation that fell to the ground.

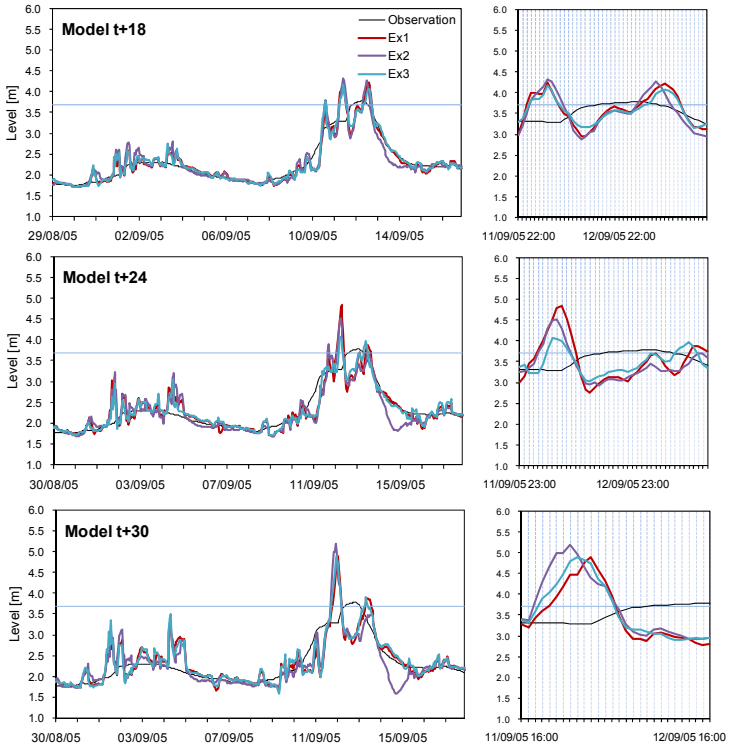


Figure 6.31: The results for storm S3 (29 Aug–19 Sep 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

However, according to the goodness of fit statistics for testing storm S3 in Table E.17, it can be said that the second experiment showed the worst results at lead times of 18, 24 and 30 hours in terms of peak prediction while the best performance was for the third experiment. Again, it shows that adding two extra water level stations does not improve the model performance, but using stepwise regression to reduce the number

of input variables could improve the performance. The goodness of fit statistics are provided in Table E.19.

In addition, Experiment 2 testing S3 and S6 showed a higher overestimation than the other two experiments. This might be an effect of adding stations P21 or P4a. However, when training without P21 and P4a, the first and third experiment predicted very similar overpredictions at lead times of 24 and 30 hours.

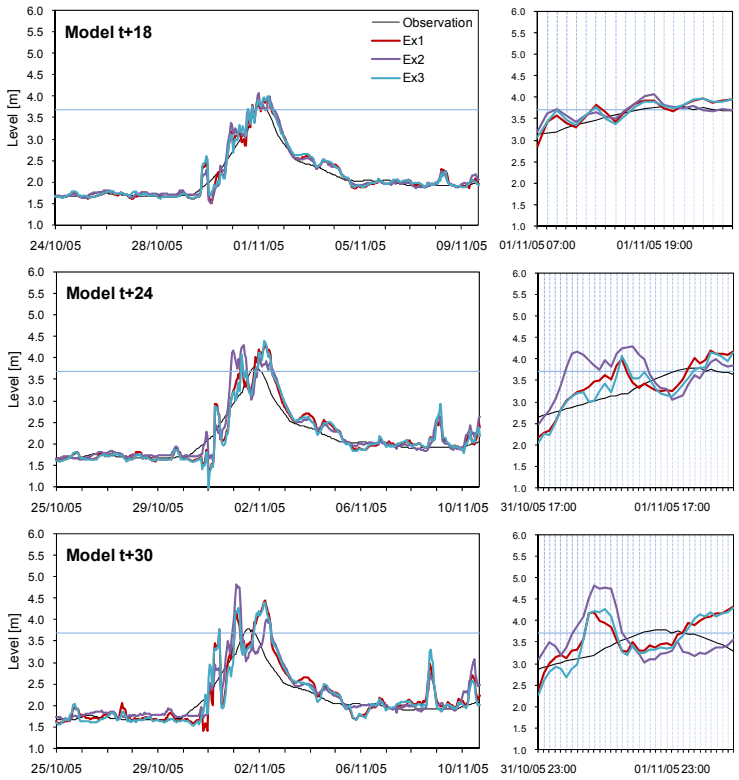


Figure 6.32: The results for storm S6 (24 Oct–11 Nov 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

6.10.3 Testing Storms S4 and S5

The first and third experiments (testing S4 and S5) predicted similar patterns again using similar inputs. However, the models in the second experiment underestimated the peak and had a large delay in the time to rise (Figure 6.33, 6.34). The goodness of fit statistics are provided in Table E.18. The results from the second experiment showed an overprediction at the peak on a small storm and an underprediction of the peak for a big storm. This may be a result of the difference in the input variables, which is the addition of the data from P21 and P4a.

It is clear that all models predicted a delay of approximately 10-15 hours because the catchment was wet. As storms S4 and S5 happened when the entire catchment was

saturated, it produced more runoff than a dry catchment. This may be a disadvantage of the use of radar data to predict water level as it appears to work better in terms of the rising limb when the catchment is not fully saturated.

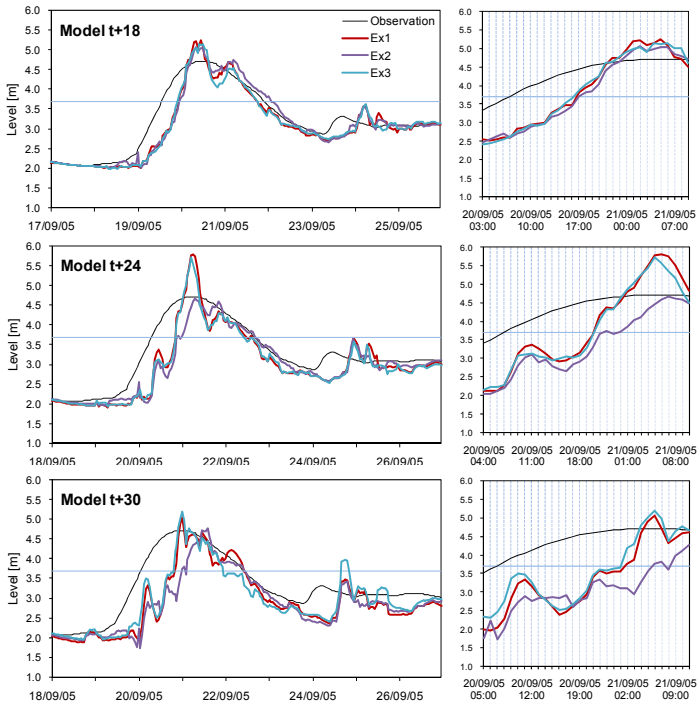


Figure 6.33: The results on storm S4 (17–27 Sept 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

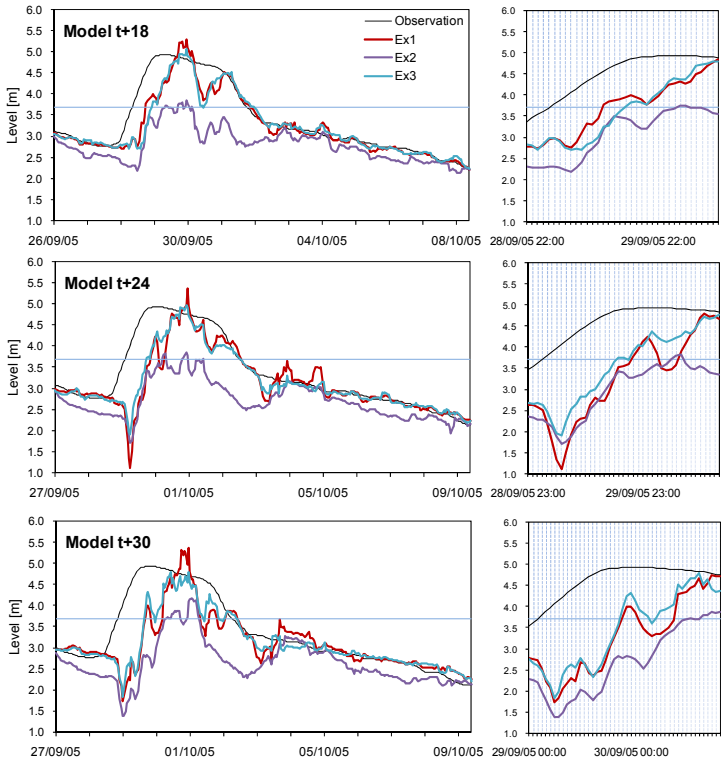


Figure 6.34: The results for storm S5 (26–9 Oct 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

6.10.4 Testing Storms S7 and S8

The hydrographs of testing the storm in 2006 by learning a storm in 2005 produced similar results but once again the second experiment predicted a different pattern than other two experiments, and still overpredicted at a lead time of 24 hours. Moreover, feeding the neural network model with radar and water level data does improve the results (Table E.19).

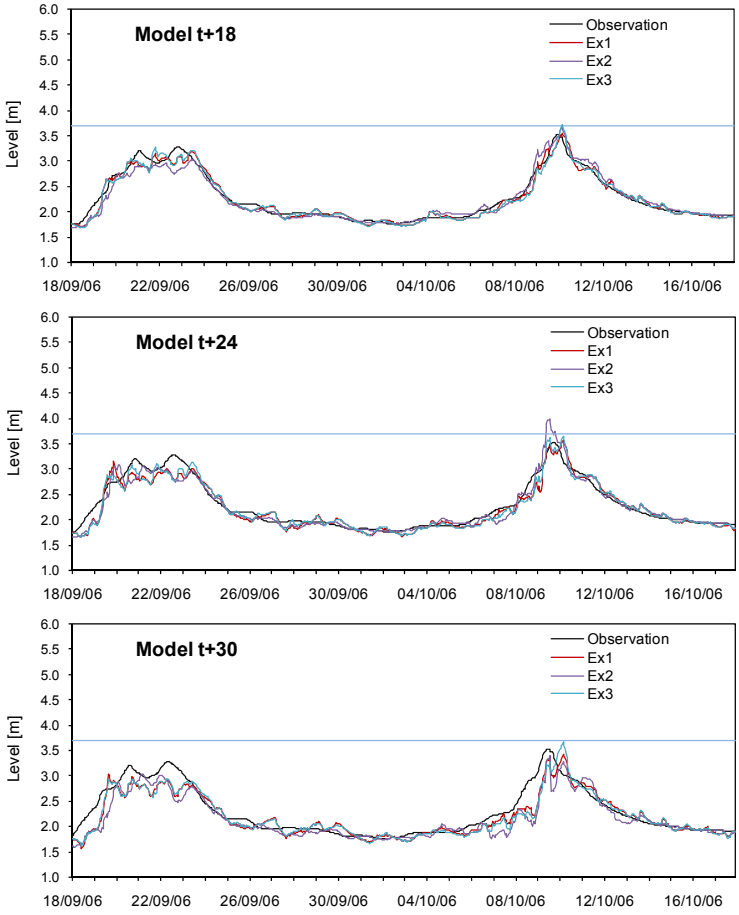


Figure 6.35: The results for storms S7 and S8 (18 Sep–17 Oct 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

In summary, all the models predicted good performance at a lead time of 18 hours. For a lead time of 24 hours, there were delays in the hydrographs. However, the use of stepwise linear regression to select the input variables had little effect on the results. The majority of the results for the rising time to peak were similar for the three experiments but adding P21 and P4a as inputs to the model does influence the peak prediction. It is unclear whether the peak overprediction is caused by the addition of P21 and P4a, and as a result, the effect of additional water stage stations will be investigated.

6.11 Experiments with Additional Stage Stations

In the previous section, it was clear that using water stage from different stations influenced model performance particularly at the peak, i.e. P21 and P4a. Therefore, in this section further experiments will explore the effect of adding P21 and P4a data. Moreover, Figure 6.36 shows the hydrograph of all 5 stations for 7 storms (S2-S8). It can be pointed out that the water level at P21 is the lowest when compared to the other four stations especially for the S5 storm (28/9-11/10). As a result, when testing S5, the second experiment underpredicted the peak (Figure 6.34). This may be a result of water released from the Mae Ngat Dam or that no heavy rainfall occurred over the P21 station.

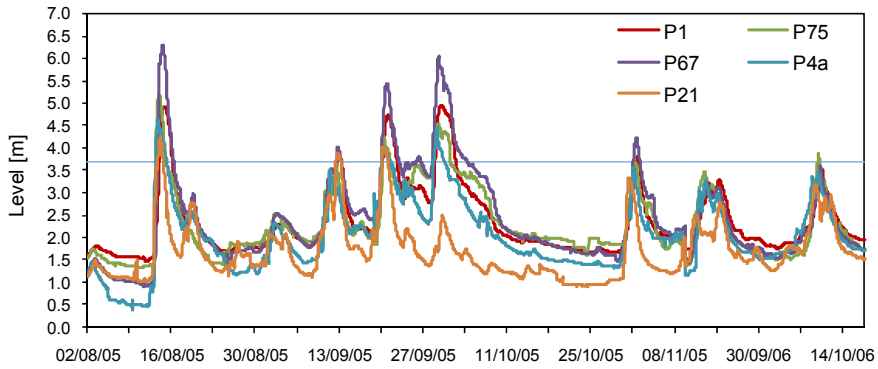


Figure 6.36: The hydrographs of storms in 2005 and 2006 with 5 water stage stations.

Therefore, three experiments are undertaken using different numbers of water stage stations. The first uses five stations (P21, P4a, P75, P67 and P1), the second 4 stations (P4a, P75, P67 and P1) and the third uses 3 (P75, P67 and P1). This section uses only 2005 data for training and testing.

6.11.1 Testing Storm S2

Based on Figure 6.37, it is clear that to predict 18 and 30 hours ahead for storm S2, the best model for predicting the rising limb is the one with four stations although the peak is overestimated. Models with three and five stations predicted the same rising time towards the peak. According to the performance statistics in Table E.20, the model with four stations is the best model at lead times of 24 and 30 hours and three stations at a lead time of 18 hours. The model with five stations gave the worst performance for all three lead times.

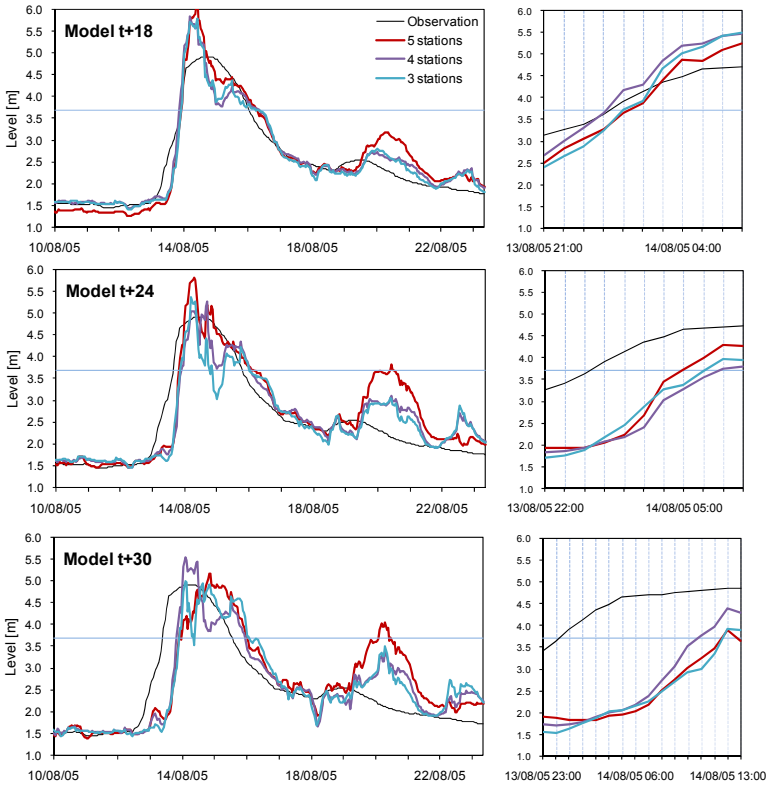


Figure 6.37: The results on storm S2 (10–24 Aug 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

Therefore, including the water stage from P21 reduces the model performance while adding data from P4a improved the model performance, i.e. the timing on the rising limb.

6.11.2 Testing storms S3 and S6

The hydrographs for two small storms are shown in Figures 6.38 and 6.39. All models predicted an early rising time. However, the model with five stations produced the highest error in the peak compared with the other models, while the model with three and four stations predicted similar levels. This could be due to the addition of data from P21. However, the use of three, four or five stations produced similar results. The goodness of fit statistics are provided in Table E.20 and E.21. From this it is possible to conclude that model performance in testing S3 is the same as for S2 and that the model with four stations gave the best performance at lead times of 24 and 30 hours.

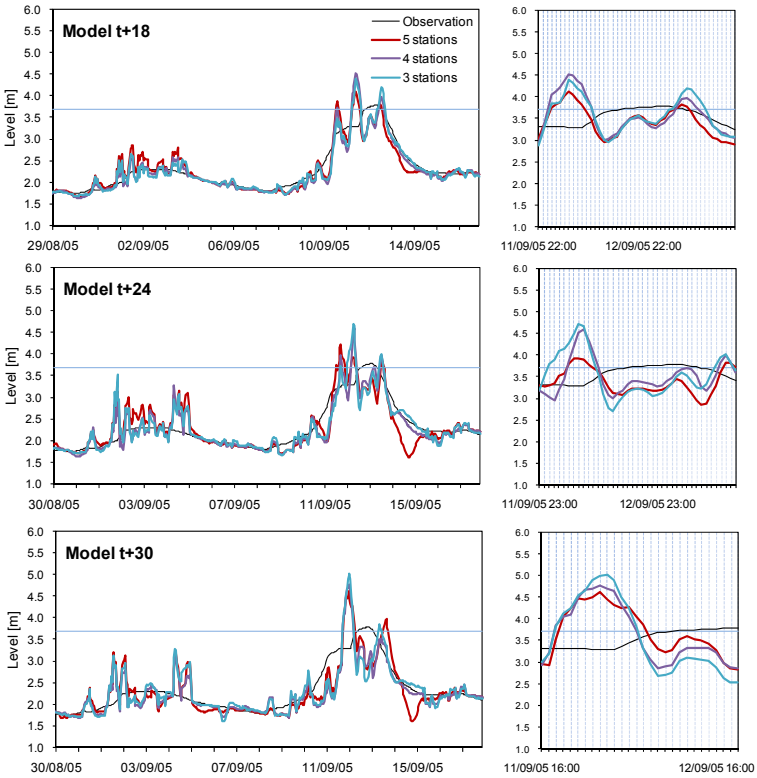


Figure 6.38: The results for storm S3 (29 Aug-18 Sep 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

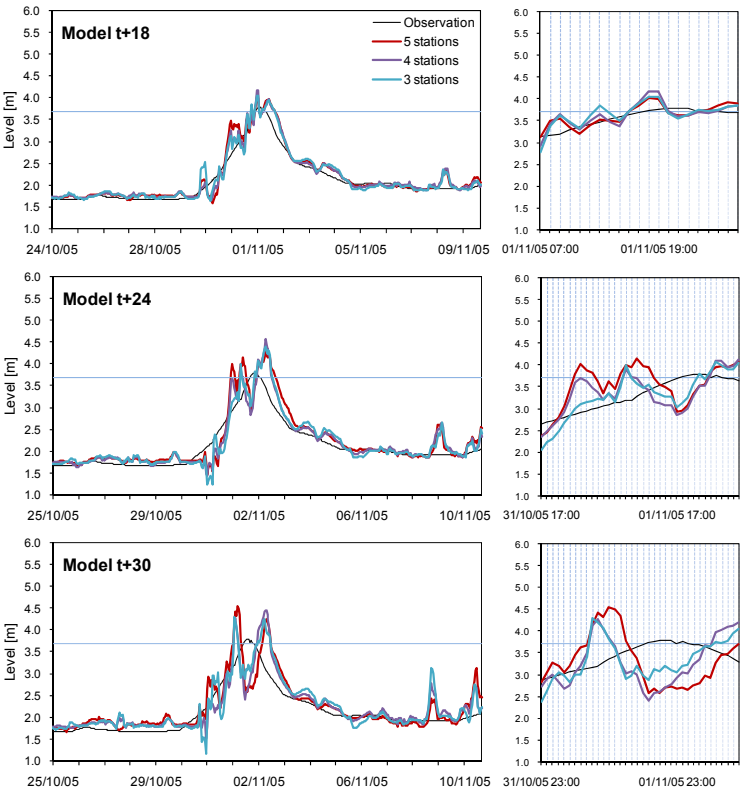


Figure 6.39: The results for storm S6 (24 Oct-11 Nov 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

6.11.3 Testing storms S4 and S5

The model with five stations produced the worst performance in terms of underestimation of the peak and a delay in the rising limb. However, the model with three and four stations predicted similar results in terms of the rising limb but different results for the peak prediction (Figure 6.40, 6.41). In addition, the model with three flow stations seemed to be the best model at all lead times for both storms S4 and S5 except for predicting at storm S4 for an 18 hour lead time. Overall the model with four stations gave the best performance (Table E.22).

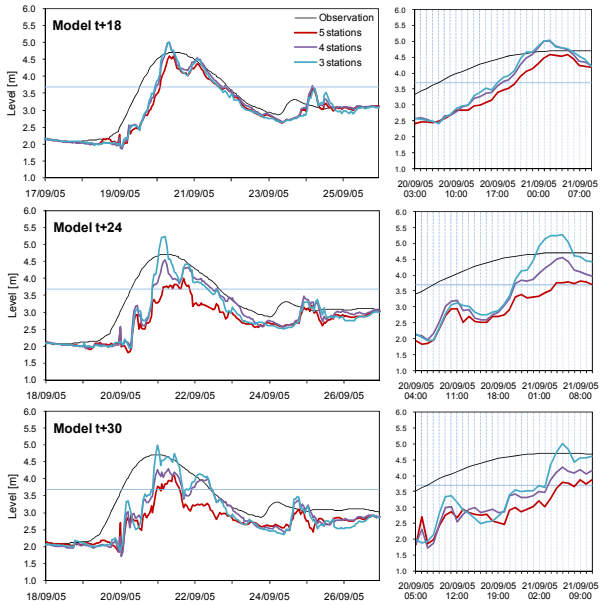


Figure 6.40: The results for storm S4 (17–27 Sep 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

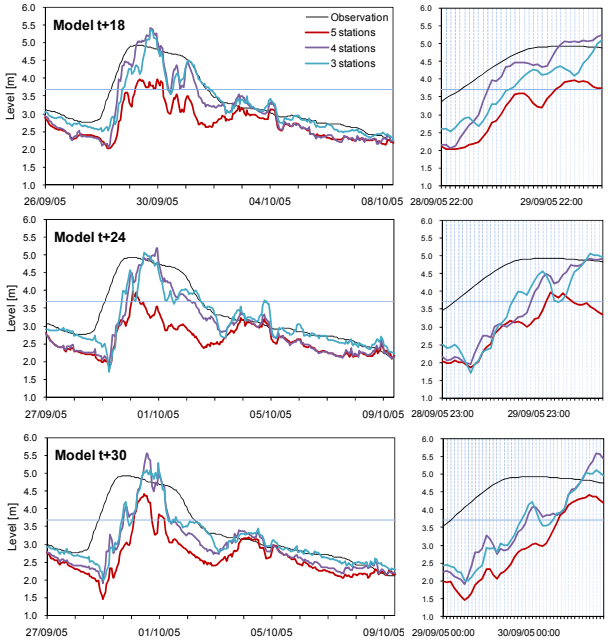


Figure 6.41: The results for storm S5 (26 Sep–9 Oct 2005) of using differing numbers of water stage stations as inputs for lead times of 18, 24 and 30 hour lead times.

In summary, for most of the storms in 2005, there are some relationships with P4a, P21, P75, P67 and P1. Using three, four or five flow stations does not result in significant differences in model performance under normal storm conditions. In contrast, the water level from the P21 station might have no effect as a result of the water released from the Mae Ngat Dam, while the other 3 stations had a direct effect. This may be due to the fact that P21 is not located on the main river and there might not be much rainfall over the P21 area (from the radar images). Therefore, the recommended stage stations for flood forecasting at P1 would be P1, P75 and P67. The final section of this chapter considers the effect of different storm patterns on the performance of the neural network models.

6.12 Experiments with Events from Different Rainfall Patterns

As mentioned previously, the pattern of storms in 2005 differs to that of 2006, i.e. one storm track moves from the northeast and one from the southwest. Therefore, this section will compare models with different training data sets. The first model was trained with storms in both 2005 and 2006 (S3-S8, Experiment 1 in Section 6.10), while the second model (3 stations) was trained with storms in 2005 only (S3-S6, model with 3 water stage stations in Section 6.11). Both models will be tested with the storm in 2005 (S2). This experiment would ideally be completed with more historical data, but the storm data are sparse and the issue is worth addressing in practice.

6.12.1 Results

Figure 6.42 shows the results for both neural network models, which produced similar results on the timing of the rising limb for a lead time of 24 hours. Based on the overall performance in Table 6.6, the performance statistics show that the model trained with only storms in 2005 gave a better result than the model trained with storms in both 2005 and 2006. For the other testing storms such as S3, S4, S5 and S6, all the hydrographs and the performance results can be found in Figure E.13 – E.16 and Table E.23, E.24. Overall, the model trained using both 2005 and 2006 gave the best results based on the errors in the rising limb, PDIFF, CE, RMSE and MAE.

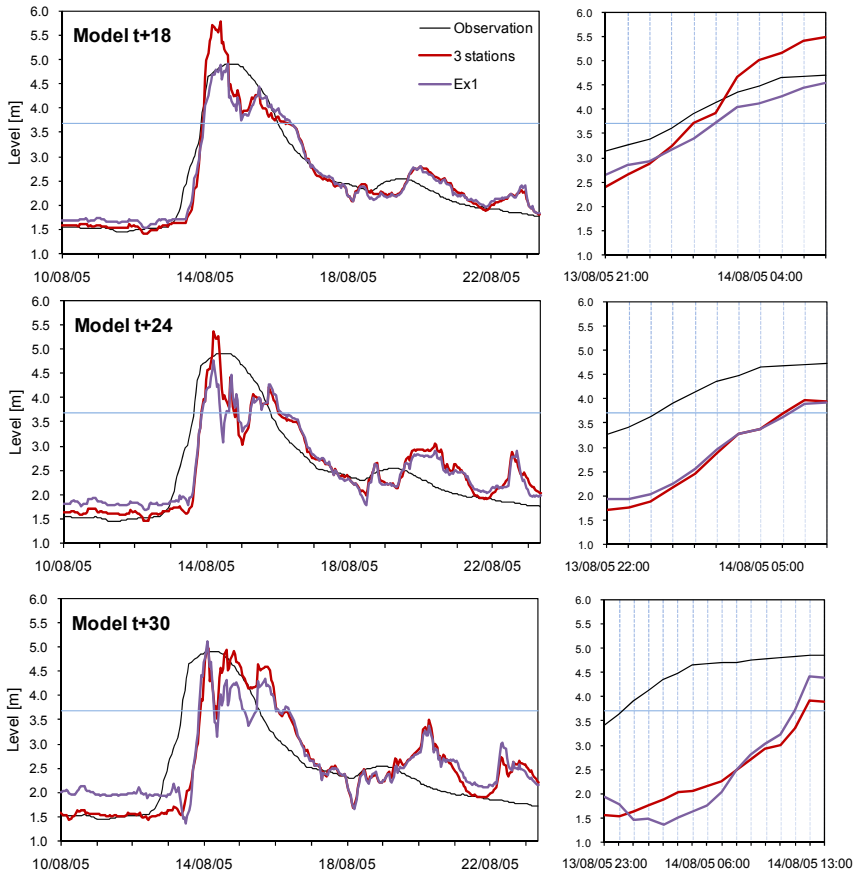


Figure 6.42: The results on storm S2 (10-23 Aug 2005) of applying different training data sets for lead times of 18, 24 and 30 hour lead times.

Table 6.6: Goodness of fit statistics testing S2 for the two experiments

Model	Statistic	Trained with 2005 and 2006	Trained with only 2005
t+18	PDIFF (m)	-0.0300	-0.8977
	MAE (m)	0.2014	0.1678
	RMSE (m)	0.2669	0.2607
	CE	0.9106	0.9147
	Time delay (hr)	2	1
t+24	PDIFF (m)	0.1400	-0.4634
	MAE (m)	0.3387	0.2765
	RMSE (m)	0.4399	0.4203
	CE	0.7573	0.7785
	Time delay (hr)	6.2	6
t+30	PDIFF (m)	-0.2300	-0.0943
	MAE (m)	0.4644	0.3157
	RMSE (m)	0.6104	0.5362
	CE	0.5332	0.6399
	Time delay (hr)	11.5	11.5

In summary, training the model with water stage and radar data that includes storms with different directions did appear to improve the results. The best model performance is still achieved when training with only radar data when considering long lead times

(Figure 6.26). However, including water stage data can reduce the error when storms of different patterns are included.

6.13 Conclusions

The results of the experiments in this chapter show that using only radar information as the inputs to a neural network can successfully predict the rising limb of the hydrograph for an extended lead time. Raw radar reflectivity was extracted from sample points across the image to create a series of input variables. Four input determination techniques from Chapter 5 (correlation, stepwise regression, data mining and a genetic algorithm) were chosen to select input variables from the radar data. The best model for testing the first large storm in 2005 was model D with an error of 8.95 cm at the peak and a 21 hr lead time. However, selecting inputs based on the correlation between cumulative radar reflectivity and the water level at P1 resulted in neural networks that could forecast at improved lead times of 24 to 30 hrs with an error in the peak stage of only 0.1-4.4 cm. However, the performance on the falling limb was poor. In addition, when water stage data were added to the radar data, the performance in predicting the falling limb improved but this reduced the maximum lead time to 18 hrs with a 2 hr delay and a 3 cm error at the peak. Compared with using only water level data in the neural network modesl (testing storm S2 in Case Study 2, Chapter 5), the maximum lead time was also 18 hrs with a 3.5 hr delay and 1.9 cm error at the peak.

There were also problems when the sample points were not in the storm tracks, which then led to a flood event, or if the storms were of quite different types. The best results were achieved when combining radar with stage data and can result in models capable of predicting 18 hours ahead. However, using different upper stage stations leads to different model performance and the major influence is the upper stage stations of P75 and P67. Therefore, the results from this chapter show that (1) selecting input variables based on simple correlation produces the best results with the very large storm from a northeastern monsoon and (2) for small storms or a wet catchment, the neural network model may require water stage data and radar data for the best timing of the rising limb or only water stage data (Chapter 5) for the best peak stage predictions.

As there are two main directions in which a storm can move, these limited results show that providing the neural network with data from similar storm tracks can improve the results. While further research is needed with many more events, it tentatively suggests that, in this catchment, a neural network could be trained on both types of storm where the rainfall may come from two distinct directions, i.e. two different neural networks

would need to be trained. It is clear from an examination of the hydrograph that rainfall adds little to improve the neural network model. If a greater network of gauging stations was available, this might have improved the model results further. However, the results show the potential of using raw radar data to improve the lead time of neural network forecasts, especially when rainfall gauging stations are not available for calibration. In the next chapter experiments with neural network parameters and architecture will be undertaken.

Chapter 7 Further Experimentation to Explore Neural Network Parameters and Performance

7.1 Introduction

Chapter 5 presented some initial experiments with the development of neural networks in order to predict the flood level at station P1 near Chiang Mai. In particular, the focus was on input determination techniques, as this is one area where there is little guidance in the literature. Chapter 6 then went on to experiment with the idea of adding radar information as a direct input to the neural network, which showed that it is possible to improve the lead time of the forecast with this additional input. This chapter experiments with other issues related to neural network development where guidance is lacking, or if present, appears in the form of heuristics and empirical evidence in the literature. The problem with such heuristics is that they may be related specifically to the catchment on which the model was developed rather than generalisable patterns that are applicable to all neural network hydrological models. Three areas will be investigated: (i) the application of the BR algorithm, in particular the use of 50 runs to produce an average forecast; (ii) the number of hidden nodes in the neural network; and (iii) the effect of the pre-processing or normalization of the input data before training the neural network.

7.2 Experiments with the Bayesian Regularization (BR) Algorithm

The Bayesian regularization algorithm (BR) used in previous modelling experiments has resulted in the development of good performing models as demonstrated in Chapters 5 and 6. However, there are only a few examples of studies in which the BR algorithm has been used within the water domain. For example, Coulibaly *et al*. (2001) used BR to train neural networks where the inputs had been selected using a peak and low flow criterion in order to find a tradeoff model that could predict both peak and low flows adequately. Anctil and Lauzon (2004) compared BR with four other generalisation approaches; stop training, stacking, bagging and boosting and found that all approaches provide improvement compared to neural networks that do not use these techniques, while Anctil *et al.* (2004a) used the BR algorithm to develop neural network models of rainfall-runoff using additional inputs including potential evapotranspiration, the antecedent precipitation index and a soil moisture index. The results of the latter study showed that only the addition of the soil moisture index resulted in better model performance. Aqil *et al*. (2007) examined the performance of neural networks trained with LM and BR as well as a neuro-fuzzy model, where the latter outperformed the

other neural networks. Rai and Mathur (2008) developed neural networks with the BR algorithm for sediment modelling. Compared to a linear transfer function, the neural networks performed better in terms of both computation of runoff hydrographs and sedimentographs for the two catchments compared. Krishna *et al.* (2008) compared BR to LM and an RBF network for the modelling of groundwater levels in India. They found that models developed using the LM algorithm were better than those developed with the BR algorithm, in contrast to the findings in this research, although all models performed well in terms of the overall performance statistics. More recently, Yonaba *et al.* (2010) used neural networks trained with BR to examine the effect of different transfer functions on network performance for stream flow forecasting. The authors found that the tangent sigmoid was the best performing function. No other algorithms were used for neural network development.

In utilizing this algorithm in previous modelling experiments, 50 neural network models were developed and the results were averaged. This was based on a heuristic supplied by Anctil (2007), who suggested this approach in order to smooth out the variations caused by different initializations of the network. The examples of research provided above do not refer to the use of multiple runs nor do they mention any variations that could result from different initialisations of the network. However, it was noticed from the experiments run in previous chapters than when the 50 individual runs were plotted for a storm in the testing data set, there was a wide variation in the predictions of the different runs. Figure 7.1 illustrates this point clearly where just 5 runs have been randomly selected and plotted against the actual values as well as the average of the 50 runs. The results clearly show that the model runs vary quite significantly between one another, especially on the prediction of the peak. Some models under predict the peak, others over predict the peak and yet others are late in peak prediction. However, after averaging all the individual runs, the result is improved. This section will investigate the behaviour of the model performance based on the number of runs used in the model prediction. This has implications computationally as each additional training run takes time to complete.

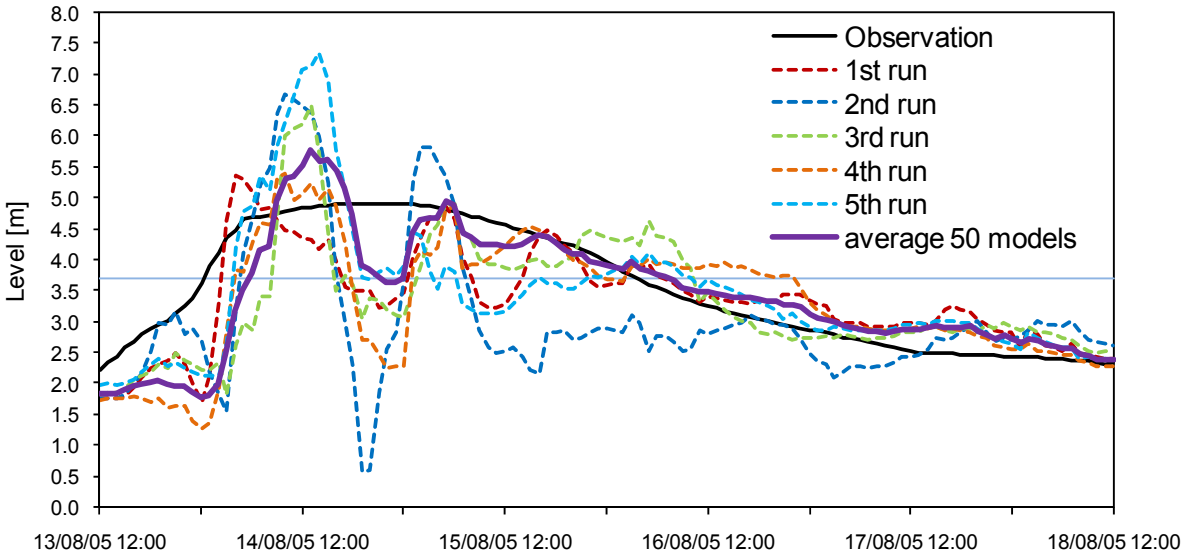


Figure 7.1: Illustration of the variation between runs for a storm event and lead time of t+18.

7.2.1 Models Used in this Experiment

Eight models were selected for further experimentation in this section of the thesis. These models are listed in Table 7.1. The best performing models were selected for lead times of 12 and 18 hours when using only 3 water level stations and rainfall data. Two models were chosen for a lead time of 24 hours using only radar data. Choosing a range of different models allows us to determine if the choice of the number of runs to average is different between lead times and models built with different inputs. Model 1 – 6 test 2006-2007 storms and train 2001-2005 storms (taken from Case study 4, Chapter 5). While model 7 and 8 test S2 and train S1, S3 – S8 (taken from Section 6.7.2, Chapter 6). All results of selected input variables can be seen in Table F.1.

Table 7.1: Models used in this experiment.

Model	Lead Time (hours)	Inputs	Input Determination Method	Number of Model Inputs
1	12	Water level + rainfall	Stepwise	18
2	12	Water level + rainfall	Correlation + Stepwise	10
3	12	Water level + rainfall	Pruning	15
4	18	Water level + rainfall	Stepwise	18
5	18	Water level + rainfall	Correlation + Stepwise	6
6	18	Water level + rainfall	Pruning	37
7	24	Radar data	Chosen based on 30 hr time correspondence with P1 (Section 6.7)	9
8	24	Radar data	Stepwise	3

The 50 runs for each model were first processed using Hydrotest (Dawson *et al.*, 2007) to calculate the RMSE, CE and MAE. These performance measures were then used to

rank the runs from best to worst performance based on the aforementioned three measures. The runs with the 5 highest values of each performance statistic were then averaged to produce a forecast. This was repeated for the 10 highest values, 15 highest values and so on at intervals of 5 runs until 50 runs was reached. Performance statistics for each of these forecasts were then calculated and are reported in the next section along with hydrographs taken from the testing data set.

7.2.2 Results

Tables 7.2 and 7.3 contain the goodness of fit statistics for the results when selecting different numbers of model runs for RMSE/CE (which resulted in no difference in their ranking of model runs) and MAE for Models 1 to 8. The bold figures shaded in the tables indicate the number of runs that produced the best values for each model. It is clear that when examining the MAE, RMSE and CE, the best results are produced by taking all 50 runs into the average. Model 8 is the only exception where 45 runs produced the best overall results for runs chosen based on best CE/RMSE while 5 runs was the best based on MAE.

Table 7.2: Goodness of fit statistics for numbers of runs selected using the best CE/RMSE for Models 1 to 8.

Model	Goodness of fit	Number of runs selected and averaged									
		5	10	15	20	25	30	35	40	45	50
1	PDIFF	**0.0226**	0.1159	0.0982	0.1401	0.1417	0.1655	0.1749	0.1821	0.1809	0.1722
	MAE	0.0514	0.0500	0.0491	0.0490	0.0490	0.0490	0.0488	0.0488	0.0486	**0.0485**
	RMSE	0.0971	0.0939	0.0915	0.0911	0.0910	0.0913	0.0910	0.0911	0.0906	**0.0902**
	CE	0.9667	0.9688	0.9705	0.9707	0.9708	0.9705	0.9707	0.9707	0.9710	**0.9713**
2	PDIFF	**0.0022**	0.1728	0.2025	0.2834	0.2738	0.3060	0.3044	0.2922	0.2850	0.2722
	MAE	0.0540	0.0544	0.0544	0.0541	0.0535	0.0534	0.0529	0.0524	0.0519	**0.0512**
	RMSE	0.1137	0.1109	0.1112	0.1099	0.1071	0.1068	0.1051	0.1033	0.1015	**0.0995**
	CE	0.9543	0.9566	0.9564	0.9573	0.9595	0.9597	0.9610	0.9623	0.9636	**0.9650**
3	PDIFF	-0.3922	-0.2986	-0.2518	-0.2479	-0.2505	-0.2230	-0.1921	-0.1688	-0.1487	**-0.1331**
	MAE	0.0603	0.0606	0.0596	0.0596	0.0592	0.0587	0.0578	0.0576	0.0571	**0.0565**
	RMSE	0.1095	0.1103	0.1082	0.1084	0.1070	0.1060	0.1045	0.1041	0.1033	**0.1021**
	CE	0.9577	0.9570	0.9587	0.9585	0.9596	0.9603	0.9614	0.9618	0.9624	**0.9632**
4	PDIFF	0.2059	**0.0517**	0.0876	0.1532	0.1897	0.2153	0.2420	0.2392	0.2351	0.2442
	MAE	0.0785	0.0779	0.0769	0.0762	0.0758	0.0755	0.0752	0.0751	0.0749	**0.0745**
	RMSE	0.1446	0.1396	0.1390	0.1364	0.1361	0.1361	0.1361	0.1360	0.1357	**0.1352**
	CE	0.9260	0.9311	0.9316	0.9342	0.9345	0.9345	0.9345	0.9346	0.9348	**0.9353**
5	PDIFF	-0.2822	-0.0696	-0.0590	-0.1193	-0.0599	-0.0547	-0.0469	-0.0357	-0.0300	**-0.0288**
	MAE	0.0878	0.0895	0.0888	0.0887	0.0886	0.0882	0.0875	0.0867	0.0859	**0.0851**
	RMSE	0.1732	0.1782	0.1696	0.1664	0.1660	0.1634	0.1609	0.1587	0.1569	**0.1553**
	CE	0.8938	0.8877	0.8982	0.9020	0.9025	0.9056	0.9085	0.9109	0.9130	**0.9146**
6	PDIFF	-0.2354	**-0.0056**	0.0180	0.0485	0.0549	0.0752	0.0768	0.1110	0.1254	0.1353
	MAE	0.0790	0.0773	0.0770	0.0776	0.0777	0.0773	0.0771	0.0766	0.0765	**0.0749**
	RMSE	0.1426	0.1388	0.1382	0.1398	0.1405	0.1394	0.1392	0.1386	0.1385	**0.1356**
	CE	0.9281	0.9318	0.9324	0.9308	0.9302	0.9312	0.9314	0.9320	0.9321	**0.9350**
7	PDIFF	-0.5284	-0.2151	**-0.0458**	-0.0724	-0.0889	-0.1094	-0.1286	-0.1072	-0.0924	-0.0837
	MAE	0.7040	0.6864	0.6717	0.6683	0.6688	0.6695	0.6691	0.6673	0.6661	**0.6652**
	RMSE	0.8795	0.8579	0.8404	0.8338	0.8313	0.8305	0.8287	0.8262	0.8241	**0.8225**
	CE	0.0300	0.0771	0.1145	0.1282	0.1334	0.1351	0.1389	0.1441	0.1484	**0.1516**
8	PDIFF	0.1656	0.0586	-0.0233	-0.0496	-0.0204	-0.0344	-0.0302	**0.0035**	0.0367	0.0403
	MAE	0.6670	0.6667	0.6660	0.6655	0.6653	0.6645	0.6630	0.6627	**0.6626**	0.6630
	RMSE	0.8148	0.8155	0.8151	0.8149	0.8146	0.8140	0.8130	0.8128	**0.8125**	0.8127
	CE	0.1676	0.1661	0.1668	0.1673	0.1680	0.1692	0.1711	0.1716	**0.1723**	0.1719

Note that PDIFF is based on the first storm.

However, if PDIFF is the main measure of concern, then the pattern is much less clear cut. For this error measure, using a smaller number of runs generally produced a better result. For example, Models 1 and 2 for runs selected by CE/RMSE and Model 2 selected by MAE produced the best PDIFF when averaging over only 5 runs. Other models produced the best PDIFF when averaging over 10, 15, 20, 30 and 40 runs. As there is no pattern regarding the best PDIFF, it seems clear that using 50 model runs generally gives the best overall model results regardless of lead time or input determination method.

Table 7.3: Goodness of fit statistics for numbers of runs selected using the best MAE for Models 1 to 8.

Model	Goodness of fit	Number of runs selected and averaged									
		5	10	15	20	25	30	35	40	45	50
1	PDIFF	0.2054	0.1445	0.1384	**0.1295**	0.1482	0.1681	0.1749	0.1867	0.1809	0.1722
	MAE	0.0501	0.0498	0.0498	0.0494	0.0492	0.0489	0.0488	0.0487	0.0486	**0.0485**
	RMSE	0.0945	0.0955	0.0938	0.0921	0.0920	0.0912	0.0910	0.0909	0.0906	**0.0902**
	CE	0.9685	0.9678	0.9689	0.9700	0.9701	0.9706	0.9707	0.9708	0.9710	**0.9713**
2	PDIFF	**-0.0219**	0.1083	0.2025	0.2525	0.2883	0.2896	0.3017	0.2903	0.2868	0.2722
	MAE	0.0534	0.0539	0.0546	0.0541	0.0536	0.0533	0.0530	0.0525	0.0519	**0.0512**
	RMSE	0.1079	0.1095	0.1115	0.1104	0.1085	0.1070	0.1054	0.1033	0.1014	**0.0995**
	CE	0.9589	0.9577	0.9561	0.9570	0.9585	0.9596	0.9607	0.9623	0.9637	**0.9650**
3	PDIFF	-0.3559	-0.2274	-0.2598	-0.2479	-0.2290	-0.1903	-0.1922	-0.1744	-0.1487	**-0.1331**
	MAE	0.0599	0.0585	0.0596	0.0596	0.0591	0.0584	0.0580	0.0576	0.0571	**0.0565**
	RMSE	0.1087	0.1061	0.1084	0.1084	0.1070	0.1056	0.1050	0.1042	0.1033	**0.1021**
	CE	0.9583	0.9602	0.9585	0.9585	0.9596	0.9606	0.9611	0.9616	0.9624	**0.9632**
4	PDIFF	0.2958	0.1969	**0.1747**	0.2138	0.2009	0.2209	0.2342	0.2310	0.2374	0.2442
	MAE	0.0798	0.0774	0.0770	0.0762	0.0757	0.0754	0.0753	0.0751	0.0749	**0.0745**
	RMSE	0.1459	0.1386	0.1386	0.1374	0.1362	0.1362	0.1362	0.1360	0.1357	**0.1352**
	CE	0.9247	0.932	0.9321	0.9333	0.9344	0.9344	0.9344	0.9346	0.9348	**0.9353**
5	PDIFF	-0.4265	0.0266	-0.0331	**-0.0110**	-0.0592	-0.0499	-0.0507	-0.0378	-0.0285	-0.0288
	MAE	0.0911	0.0902	0.0898	0.0890	0.0883	0.0881	0.0876	0.0869	0.0860	**0.0851**
	RMSE	0.1979	0.1756	0.1726	0.1696	0.1655	0.1631	0.1608	0.1589	0.1570	**0.1553**
	CE	0.8615	0.8910	0.8947	0.8982	0.9031	0.9058	0.9085	0.9107	0.9129	**0.9146**
6	PDIFF	-0.3256	**-0.0002**	0.0170	0.0474	0.0472	0.0728	0.0894	0.112	0.1172	0.1353
	MAE	0.0789	0.0777	0.0775	0.0774	0.0776	0.0776	0.0772	0.0767	0.0765	**0.0749**
	RMSE	0.1423	0.1400	0.1399	0.1397	0.1404	0.1404	0.1395	0.1385	0.1384	**0.1356**
	CE	0.9284	0.9307	0.9308	0.9310	0.9303	0.9303	0.9311	0.9321	0.9322	**0.9350**
7	PDIFF	-0.5284	-0.2231	-0.0346	**-0.0325**	-0.0830	-0.0963	-0.1097	-0.1159	-0.1025	-0.0837
	MAE	0.7040	0.6803	0.6696	0.6662	0.6669	0.6672	0.6675	0.6668	0.6653	**0.6652**
	RMSE	0.8795	0.8554	0.8400	0.8331	0.8303	0.8287	0.8273	0.8257	0.8235	**0.8225**
	CE	0.0300	0.0824	0.1152	0.1297	0.1354	0.1388	0.1417	0.1450	0.1497	**0.1516**
8	PDIFF	0.1561	0.1280	0.1066	0.0633	0.0490	**0.0096**	-0.0155	0.0164	0.0419	0.0403
	MAE	**0.6447**	0.6536	0.6562	0.6578	0.6596	0.6602	0.6616	0.6625	0.6625	0.6630
	RMSE	**0.8016**	0.8065	0.8084	0.8096	0.8106	0.8110	0.8122	0.8125	0.8124	0.8127
	CE	**0.1943**	0.1844	0.1805	0.1781	0.176	0.1752	0.1728	0.1722	0.1725	0.1719

Note that PDIFF is based on the first storm.

Figures 7.2 and 7.3 show the results for a hydrograph in the testing dataset for Models 3 and 6. The dashed red line is the forecast after averaging over 5 runs and it is clearly higher than the forecasts from averaging a higher number of runs. This explains why using a smaller number of runs often produces the best PDIFF value.

Models 3 and 6 are for different lead times so the results are similar regardless of lead time. The graphs for Models 1, 2, 4 and 5 showed similar patterns and are therefore not included here (see Appendix F).

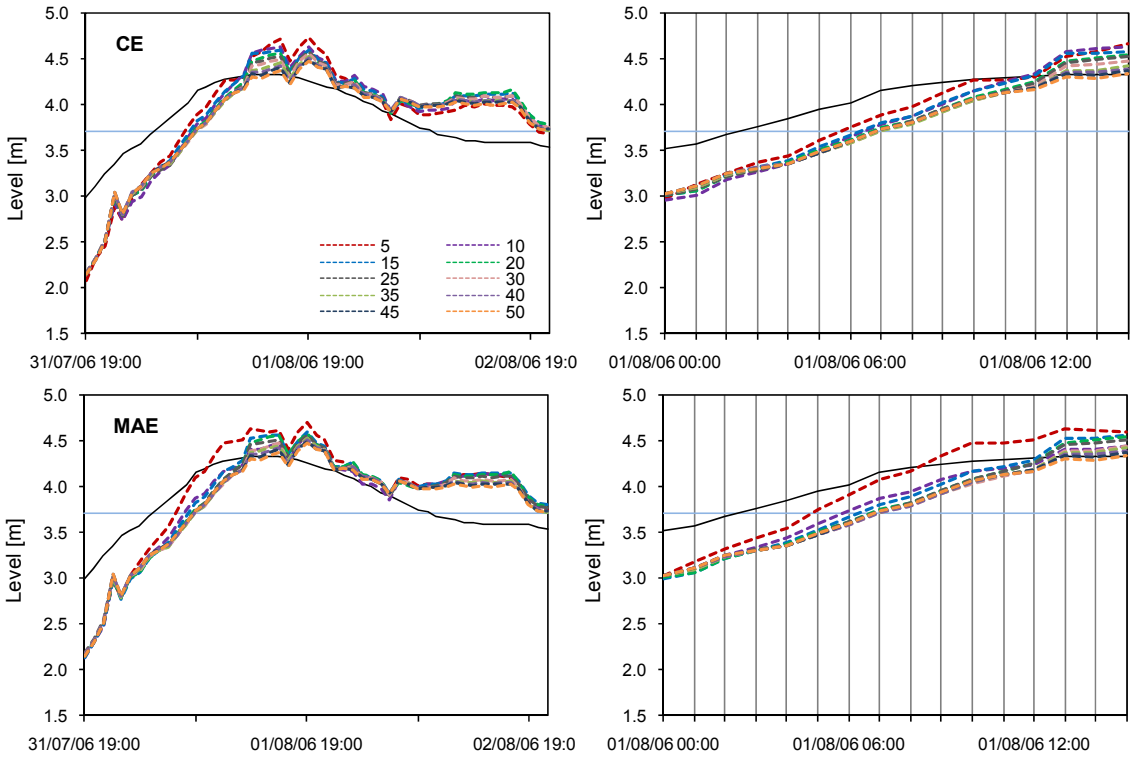


Figure 7.2: Results for a storm in 2006 for increasing numbers of runs averaged together, chosen by ranking CE/RMSE and MAE for Model 3.

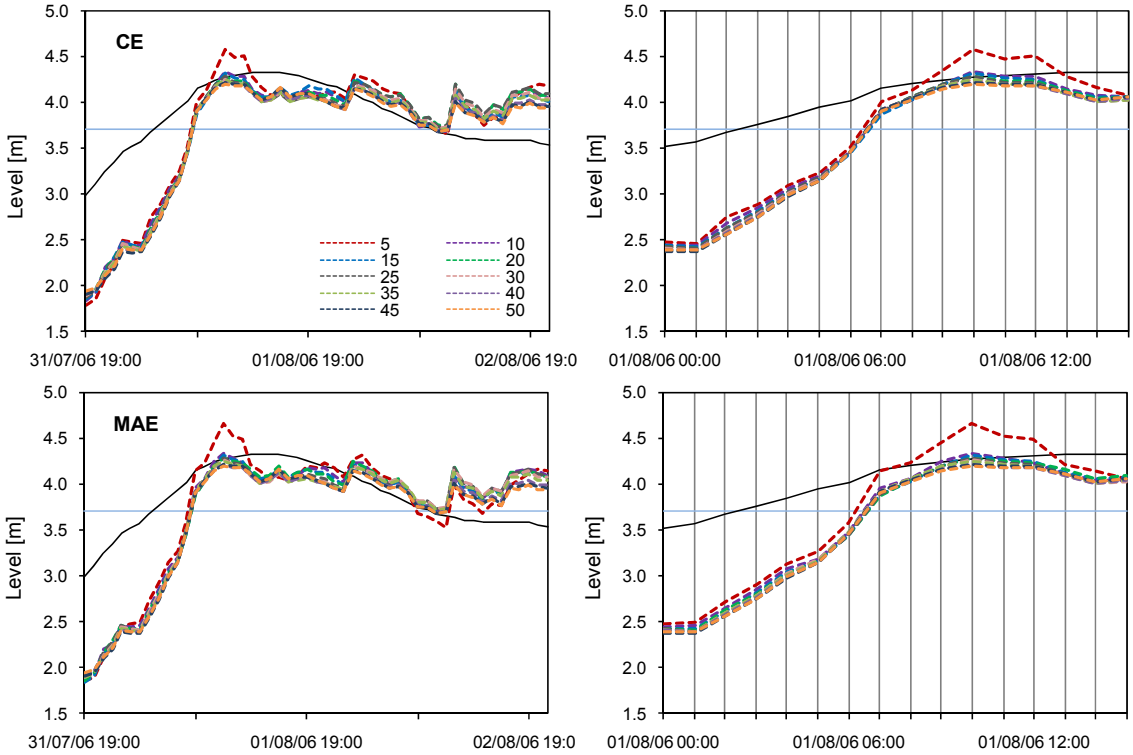


Figure 7.3: Results for a storm in 2006 for increasing numbers of runs averaged together, chosen by ranking CE/RMSE and MAE for Model 6.

Figures 7.4 and 7.5 are for Models 7 and 8 using radar data as inputs. The same patterns can again be seen for these models although Model 8 shows very little

difference when using different numbers of model runs. This may explain the differing result in Table 7.3.

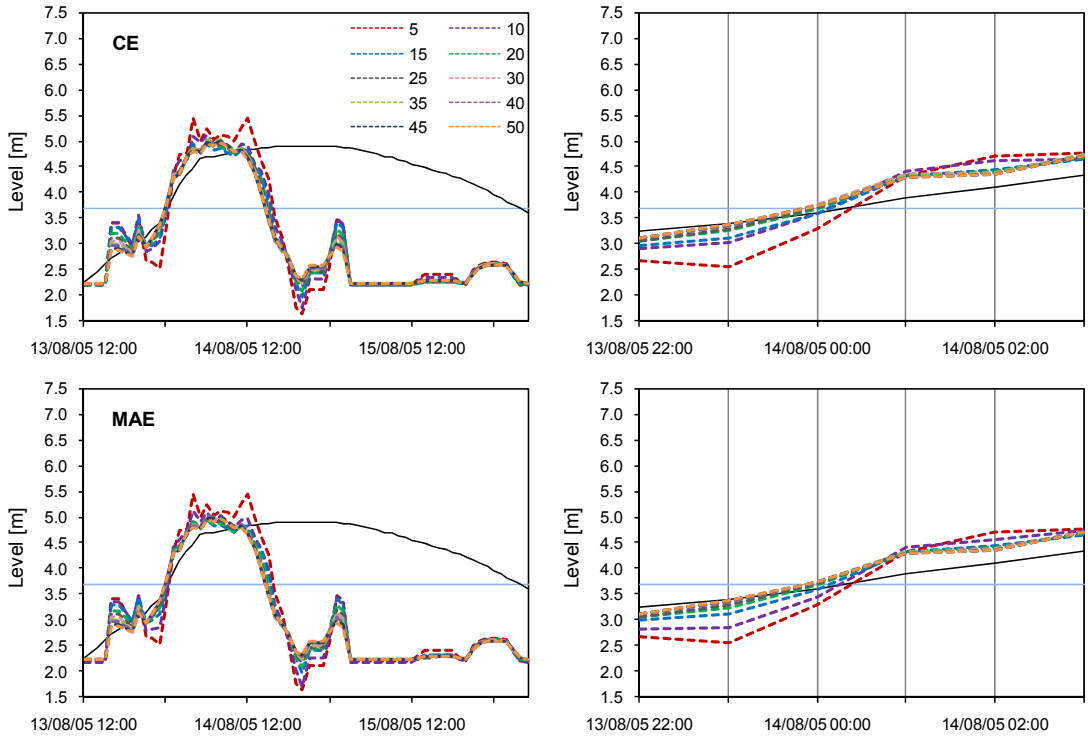


Figure 7.4: Results for a storm in 2005 for increasing numbers of runs averaged together, chosen by ranking CE/RMSE and MAE for Model 7

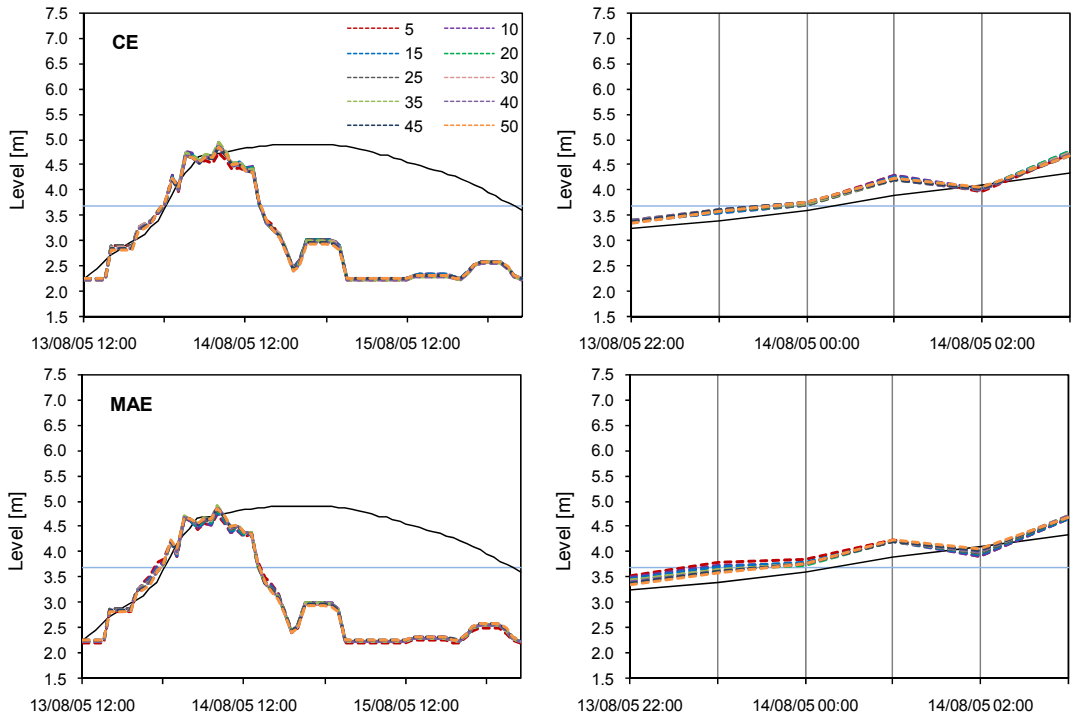


Figure 7.5: Results for a storm in 2005 for increasing numbers of runs averaged together, chosen by ranking CE/RMSE and MAE for Model 8.

One further experiment was added based on the results found here. Since 50 runs seemed to generally produce the best result overall, the models were rerun with 100 runs to see whether the results would keep improving as more runs were added. Table 7.4 shows the results in terms of goodness of fit statistics. The first thing to note is that there is very little difference in the statistics between 50 and 100 runs. For the majority of models, 50 runs provides the best results. However, even when models with 100 runs provide better results, the difference is very small. Therefore, the use of 50 runs seems to be a good number when using the BR algorithm (Figure F.5, F.6, F.7).

Table 7.4: Goodness of fit statistics for 50 and 100 runs averaged for Models 1 to 8.

Goodness of fit	Model 1		Model 2		Model 3	
	50 runs	100 runs	50 runs	100 runs	50 runs	100 runs
PDIFF	**0.1722**	0.1938	**0.2722**	0.2917	-0.1331	**-0.1195**
MAE	**0.0485**	**0.0485**	0.0512	**0.0507**	**0.0565**	0.0568
RMSE	**0.0902**	0.0904	**0.0995**	**0.0995**	**0.1021**	0.1027
CE	**0.9713**	0.9712	**0.9650**	0.9651	**0.9632**	0.9628
Goodness of fit	**Model 4**		**Model 5**		**Model 6**	
	50 runs	**100 runs**	**50 runs**	**100 runs**	**50 runs**	**100 runs**
PDIFF	0.2442	**0.2136**	**-0.0288**	0.0858	**0.1353**	0.1701
MAE	0.0745	**0.0742**	0.0851	**0.0835**	**0.0749**	0.0772
RMSE	**0.1352**	**0.1352**	0.1553	**0.1527**	**0.1356**	0.1397
CE	0.9353	**0.9354**	**0.9146**	0.9175	**0.9350**	0.9309
Goodness of fit	**Model 7**		**Model 8**			
	50 runs	**100 runs**	**50 runs**	**100 runs**		
PDIFF	**-0.0837**	-0.0929	**0.0403**	0.0417		
MAE	**0.1256**	0.1259	**0.1252**	0.1255		
RMSE	**0.3574**	**0.3574**	**0.3532**	0.3536		
CE	**0.8526**	0.8527	**0.8561**	0.8558		

7.3 Varying the Number of Hidden Nodes

Experiments in Chapters 5 and 6 used hidden nodes of 10 and 20 in the development of the neural networks. The number 20 was originally chosen based on a heuristic supplied by Dawson and Wilby (1999) and Demuth *et al*. (2009). Further experimentation used 10 nodes as this did not appear to affect the performance too much while reducing the computational time considerably so was adopted for subsequent experiments. However, the optimal number of hidden nodes to use remains an area where little guidance is provided in the literature. Many studies have approached this problem using a trial and error procedure. Based on the literature review in Chapter 2, three different heuristics will be used: the number of nodes in the hidden layer should be half the number of nodes in the input layer (Minns and Hall, 1996); the number of hidden nodes should be 75% of the number of input nodes (Lenard *et al*., 1995; Jain and Nag, 1995; Walczak and Cerpa, 1999); and the number of hidden nodes should be twice the number of input nodes plus 1 (Patuwo *et al*., 1993; Caudill, 1991). The 8 models used in the previous experiments are used again here in

this experiment. Table 7.5 lists the models and the number of hidden nodes used based on the three heuristics above.

Table 7.5: The number of hidden nodes used in each model experiment based on the three heuristics.

Models	Number of input nodes	Number of hidden nodes		
		50%	75%	2*input nodes + 1
1	18	9	13	37
2	10	5	7	21
3	15	7	11	31
4	18	9	13	37
5	6	3	4	13
6	37	18	27	75
7	9	3	6	19
8	3	1	2	7

7.3.1 Results t+12 hr

Table 7.6 provides the goodness of fit statistics for Models 1 to 3 for differing numbers of hidden nodes. The shaded, bold numbers denote the best performance overall. The best performers in this case are the most parsimonious models, with no example of where the heuristic 2n +1 produces the best result. The PDIFF and MAE were best when half of the input nodes were chosen as the number of hidden nodes while RMSE and CE were best when 75% were chosen. Thus there is no conclusive result, other than that the hidden nodes should be less than the number of input nodes.

Table 7.6: Goodness of fit statistics for different numbers of hidden nodes for Models 1 to 3. The best results overall are shaded and in bold.

Goodness of fit	Model 1			Model 2			Model 3		
	50%	75%	2n+1	50%	75%	2n+1	50%	75%	2n+1
PDIFF	0.1629	0.1618	0.1105	0.2532	0.2112	0.1437	**-0.0637**	-0.1674	-0.2839
MAE	**0.0482**	0.0484	0.0489	0.0490	0.0489	0.0530	0.0561	0.0580	0.0588
RMSE	0.0901	**0.0895**	0.0897	0.0933	0.0932	0.1022	0.1024	0.1048	0.1074
CE	0.9713	**0.9717**	0.9716	0.9693	0.9693	0.9631	0.9630	0.9612	0.9593

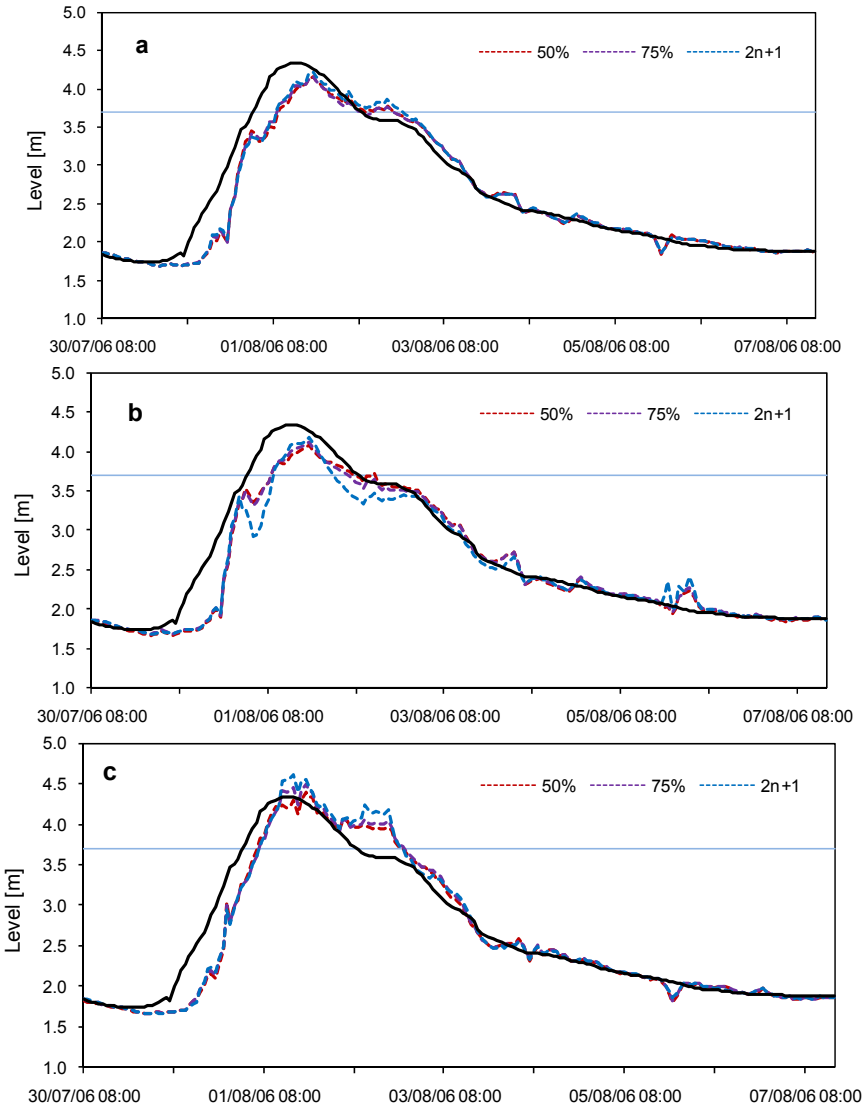


Figure 7.6: Results for a storm in 2006 at 12 hour lead time for (a) Model 1; (b) Model 2; and (c) Model 3 with different numbers of hidden nodes.

Figure 7.6 shows the model performance visually for a hydrograph in 2006 in the testing data set. There is very little difference for Model 1. However, both Models 2 and 3 show some differences for the model with 2n+1 hidden nodes. For Model 2, there is strange decrease in river level during the rising limb of the hydrograph. For Model 3, the overestimation of the peak is higher. Both of these results would suggest that taking a smaller number of hidden nodes than input nodes is a good strategy. Figure 7.7 shows a more detailed view of the hydrograph around the peak. These behaviours discussed above can be seen even more clearly.

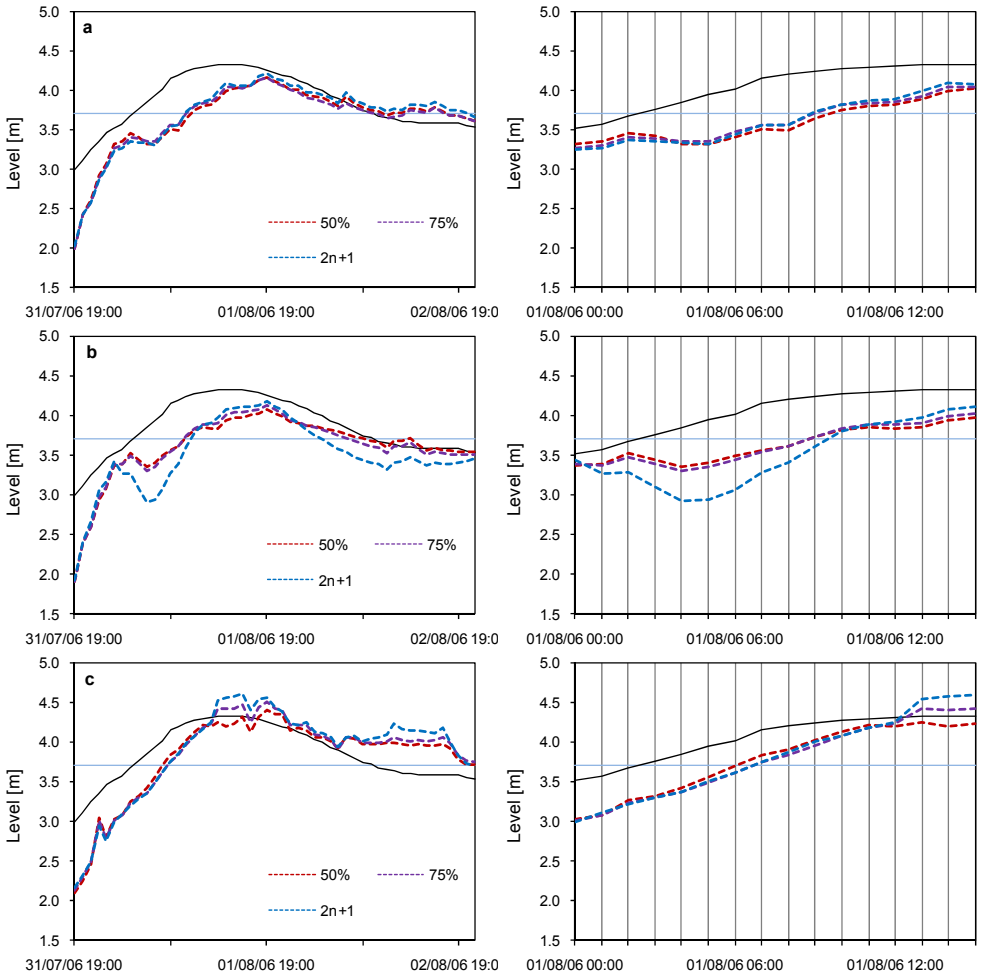


Figure 7.7: Results for a storm in 2006 at 12 hour lead time for (a) Model 1; (b) Model 2; and (c) Model 3 with different numbers of hidden nodes.

7.3.2 Results t+18 hr

Table 7.7 shows the goodness of fit statistics for Models 4 to 6 where the difference between these and the previous results is a longer lead time. For RMSE and CE, Model 4 produced the best results when taking 75% of the number of input nodes as hidden nodes. However, both PDIFF and MAE are better when using 2n+1. A lead time of 18 hours is more difficult to predict than 12 hours so this more complex problem may need more hidden nodes to learn the function properly.

Table 7.7: Goodness of fit statistics for different numbers of hidden nodes for Models 4 to 6. The best results overall are shaded and in bold.

Goodness of fit	Model 4			Model 5			Model 6		
	50%	75%	2n+1	50%	75%	2n+1	50%	75%	2n+1
PDIFF	0.3396	0.2592	0.3245	0.1507	0.1875	-0.0185	0.0269	0.0180	**-0.0086**
MAE	0.2155	0.2041	0.2230	0.2097	0.2161	0.2300	0.2134	0.2075	**0.2028**
RMSE	0.3607	**0.3417**	0.3635	0.3768	0.3774	0.3887	0.3620	0.3563	0.3465
CE	0.8085	**0.8280**	0.8054	0.7909	0.7903	0.7775	0.8071	0.8131	0.8232

Figures 7.8 and 7.9 provide visual examination of the model performance. This time there is very little difference to note between Models 4 and 6 when using stepwise linear regression and pruning to select the inputs. Only in Model 5, where stepwise regression and correlation together were used to choose the inputs, is there any noticeable difference, i.e. the peak is predicted better while the falling limb is poorly predicted. Since there is actually little difference in general between the models, the most parsimonious one should always be chosen.

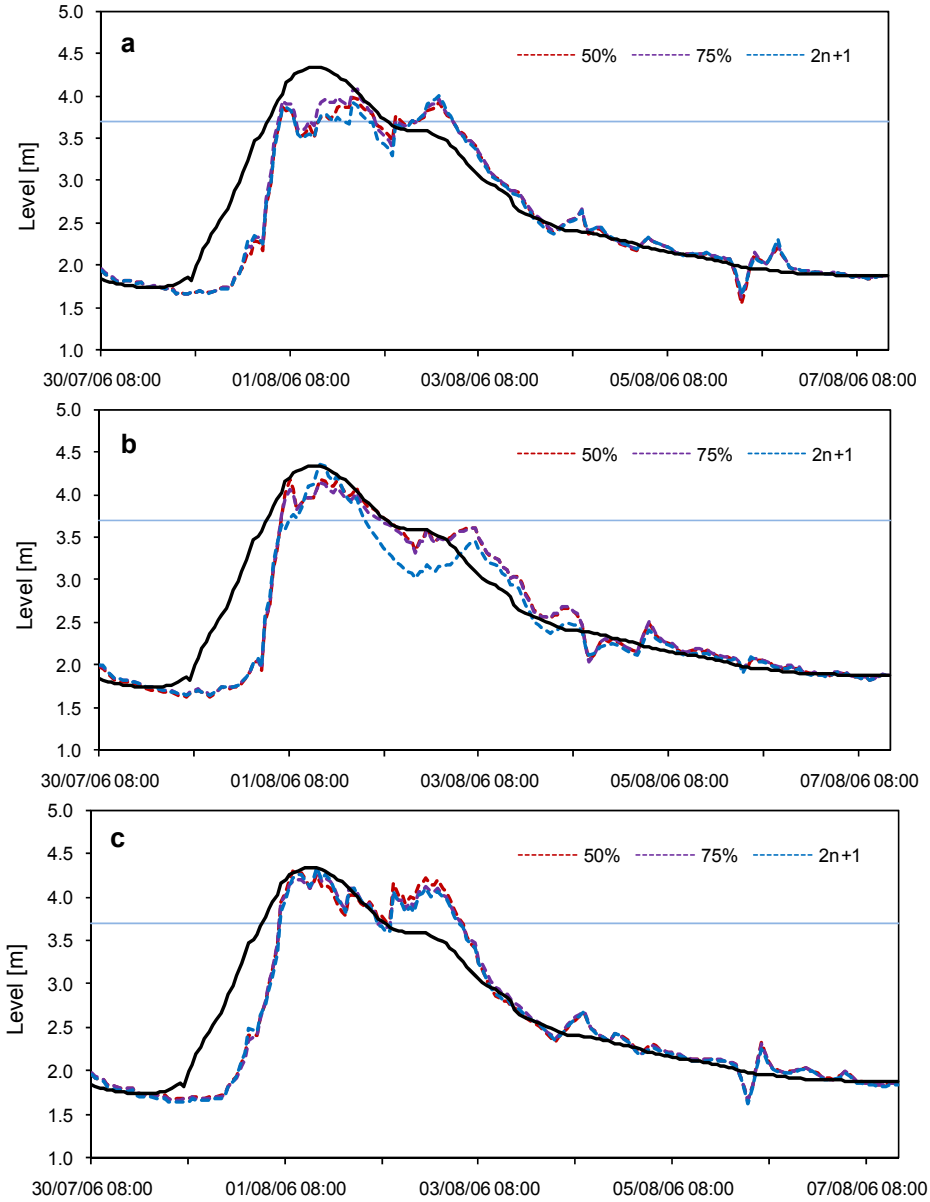


Figure 7.8: Results for one storm in 2006 at 18 hour lead time for (a) Model 4; (b) Model 5; and (c) Model 6 with different numbers of hidden nodes.

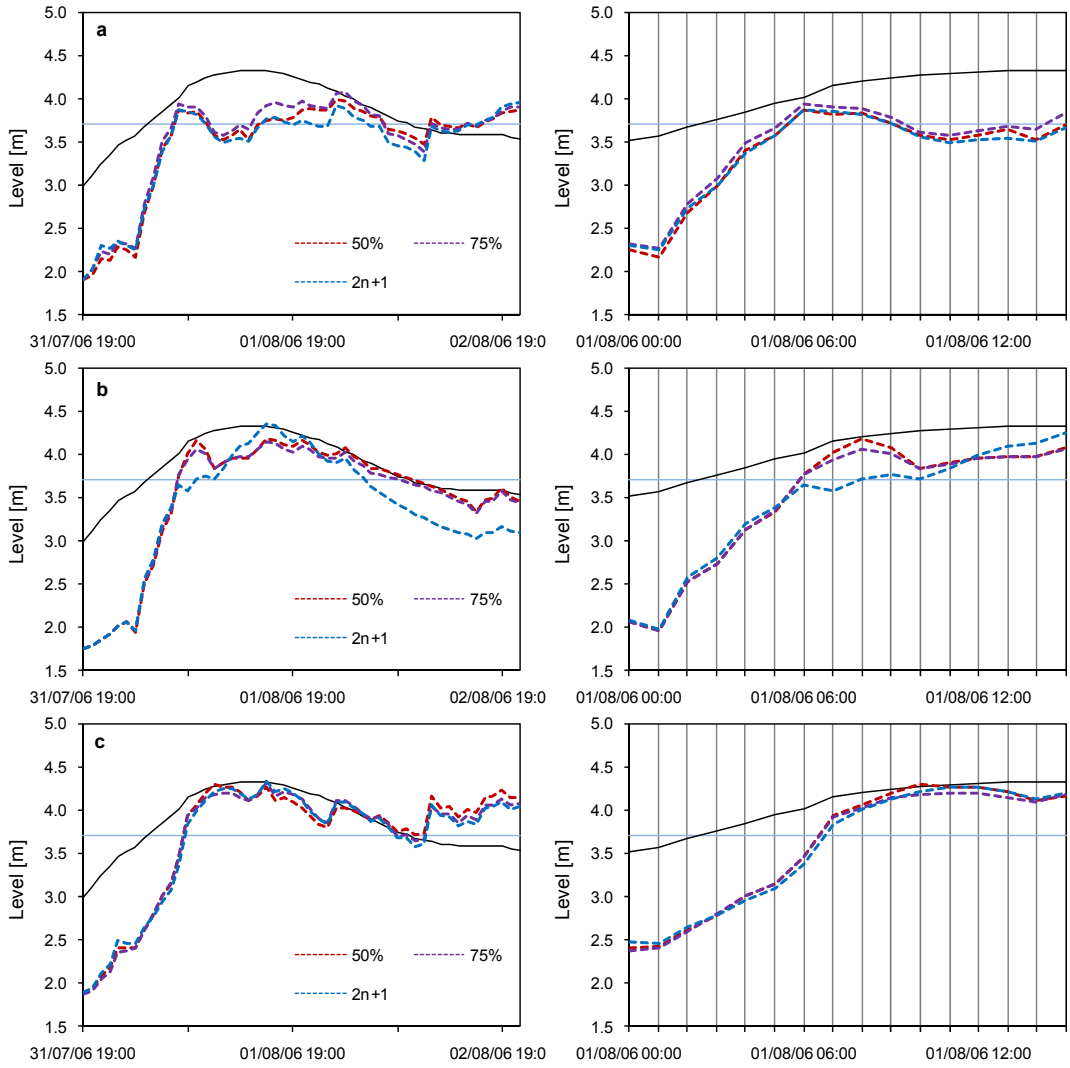


Figure 7.9: Results for a storm in 2006 at 18 hour lead time for (a) Model 4; (b) Model 5; and (c) Model 6 with different numbers of hidden nodes.

7.3.3 Results t+24 hr

Table 7.8 contains the goodness of fit statistics for Models 7 and 8 for different numbers of hidden nodes based on the three heuristics. The best performing model overall for each statistic is shaded and in bold. The result clearly shows that Model 8 (which uses stepwise regression to reduce the number of input variables) with 75% of the input variables as the number of hidden nodes, produces the best results overall. This is very parsimonious model, with only 3 inputs and 2 hidden nodes.

Table 7.8: Goodness of fit statistics for different numbers of hidden nodes for Models 7 and 8. The best results overall are shaded and in bold.

Goodness of fit	Model 7			Model 8		
	50%	75%	2n+1	50%	75%	2n+1
PDIFF	0.0625	-0.2522	0.5223	0.4610	**0.0450**	0.0625
MAE	0.6668	0.6741	0.6691	0.6681	**0.6622**	0.6668
RMSE	0.8248	0.8283	0.8194	0.8150	**0.8120**	0.8248

CE	0.1470	0.1397	0.1582	0.1670	**0.1732**	0.1470

Figure 7.10 shows the results visually for a storm in 2005 in the testing data set for Models 7 and 8 while Figure 7.11 shows the same storm but in more detail. The figures show that all three models predict the majority of the rising limb well. However, the model that uses 2n+1 hidden nodes, i.e. the least parsimonious model, is closer to predicting the peak than the other two models, especially in the detailed view of Model 8.

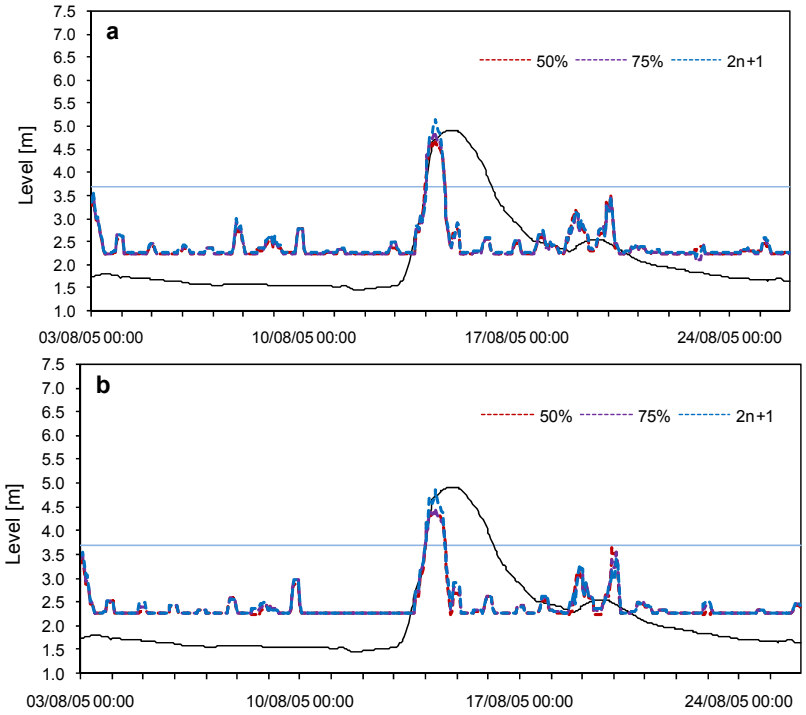


Figure 7.10: Results for a storm in 2005 at 24 hour lead time for (a) Model 7 and (b) Model 8 with different numbers of hidden nodes.

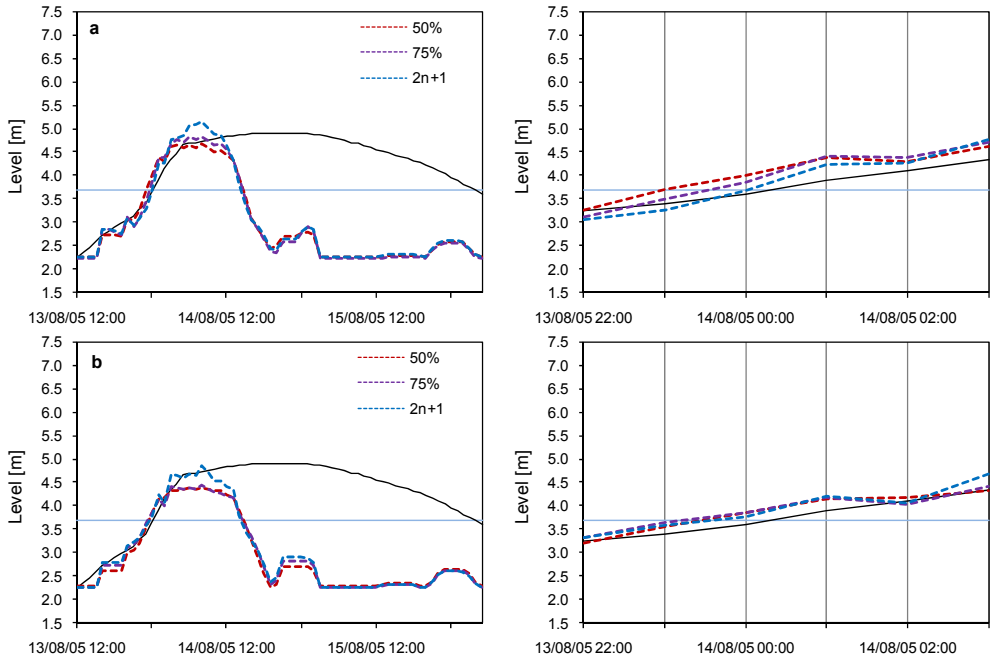


Figure 7.11: More detailed results for a storm in 2005 at 24 hour lead time for (a) Model 7 and (b) Model 8 with different numbers of hidden nodes.

In summary, there is no conclusive evidence to suggest that choosing one of the heuristics over the other produces the best result. Examination of the goodness of fit statistics and hydrographs suggests similarities between them, in which case it would be prudent to suggest that the most parsimonious models should always be chosen, i.e. select models with hidden nodes that are half the number of input variables. However, on occasion, choosing 2n+1 produced a better peak prediction. It also, however, produced poorer behaviour in other places. It would seem logical that as the complexity of the problem increases, so too should the number of hidden nodes. However, until better guidance is provided as to the choice of the number of hidden nodes, it is suggested that trial and error continues to be the best method available.

7.4 Experimenting with Normalisation of the Input Data

As mentioned in section 2.5.2, the data are normalized prior to being input to the neural network. However, there is no consensus over what range the data should be normalized or what effect this might have on the performance of the model. For example, Dawson and Wilby (1998) used [0, 1], Braddock *et al.* (1998) used [-0.9, 0.9], Shamseldin (1997) and Dawson *et al*. (2006a) used [0.1, 0.9] while Dawson *et al*. (2006b) used [0.2, 0.8]. Therefore, this section explores the use of different ranges of normalization of the input data prior to training. Three ranges of normalization were investigated including two of the most commonly used ranges, i.e. [-1 to 1] and [0.1 to 0.9] as well as a more narrow range [0.3, 0.7] after Varoonchotikul (2003).

Models 1 and 8 were used in the experiments so that models using different inputs could be compared. The minimum and maximum were calculated from the training dataset and the training and testing data were then normalized using the three ranges listed above. The minimum and maximum values for Model 1 are listed in Table 7.9.

Table 7.9: The minimum and maximum values in the training and testing datasets for Model 1 and across all historical records.

Input variables	Minimum			Maximum		
	Training	Testing	Across all historical records	Training	Testing	Across all historical records
R	0	0	0	122.8	96	122.8
P75	0.59	0.90	0.59	5.19	3.90	5.19
P67	0.43	0.18	0.18	6.28	3.90	6.28
P1	1.66	1.32	1.46	4.93	4.33	4.93

The maximum value in the training data set at P1 is 4.93 m, which is slightly higher than the maximum in the testing data set of 4.33 m. The model is therefore not required to extrapolate. A second normalization was also applied. This time the minimum and maximum values were calculated across the entire historical record. These are listed in the final columns of Table 7.9. The training and testing data were then normalized using these minimum and maximum values leading to six different combinations:

- Normalized using min/max from training data set over range [-1, 1]
- Normalized using min/max from training data set over range [0.1, 0.9]
- Normalized using min/max from training data set over range [0.3, 0.7]
- Normalized using min/max from entire historical data set over range [-1, 1]
- Normalized using min/max from entire historical data set over range [0.1, 0.9]
- Normalized using min/max from entire historical data set over range [0.3, 0.7]

7.4.1 Results

Table 7.10 shows the performance statistics for Model 1 for these six combinations listed above. The bold, shaded numbers are the best results overall. The first observation is that using the minimum and maximum values across the entire historical record is better than choosing the values from the training dataset. The exception is for PDIFF which is better when normalizing using values from just the training dataset. The second observation is that the range [-1, 1] provides the best result overall although there are very small differences between the three ranges of normalization.

Table 7.10: Goodness of fit statistics for Model 1 using the min/max from the training dataset for normalization compared to using values across the entire historical record.

Goodness of fit	Normalization using min/max from the training dataset (Norm1)			Normalization using min/max from the entire historical record (Norm2)		
	[-1, 1]	[0.1, 0.9]	[0.3, 0.7]	[-1, 1]	[0.1, 0.9]	[0.3, 0.7]
PDIFF	0.1116	0.0856	**0.075**	0.1022	0.1649	0.1209
MAE	0.2914	0.2913	0.2913	**0.1380**	0.1394	0.1393
RMSE	0.3018	0.3016	0.3016	**0.1513**	0.1527	0.1526
CE	0.6783	0.6788	0.6788	**0.9192**	0.9177	0.9178

(Bold and shading denotes the best performer.)

Figure 7.12 provides a detailed look at the top of a hydrograph for a storm in 2006 in the testing data set. The predictions are similar although it is clear that when normalizing across a wider range, the peak predictions, although later, are better. The figure also shows a slightly better performance when normalizing using minimum and maximum values from the entire historical record. This may be due to the fact that the testing data has lower minimum values than the training data set.

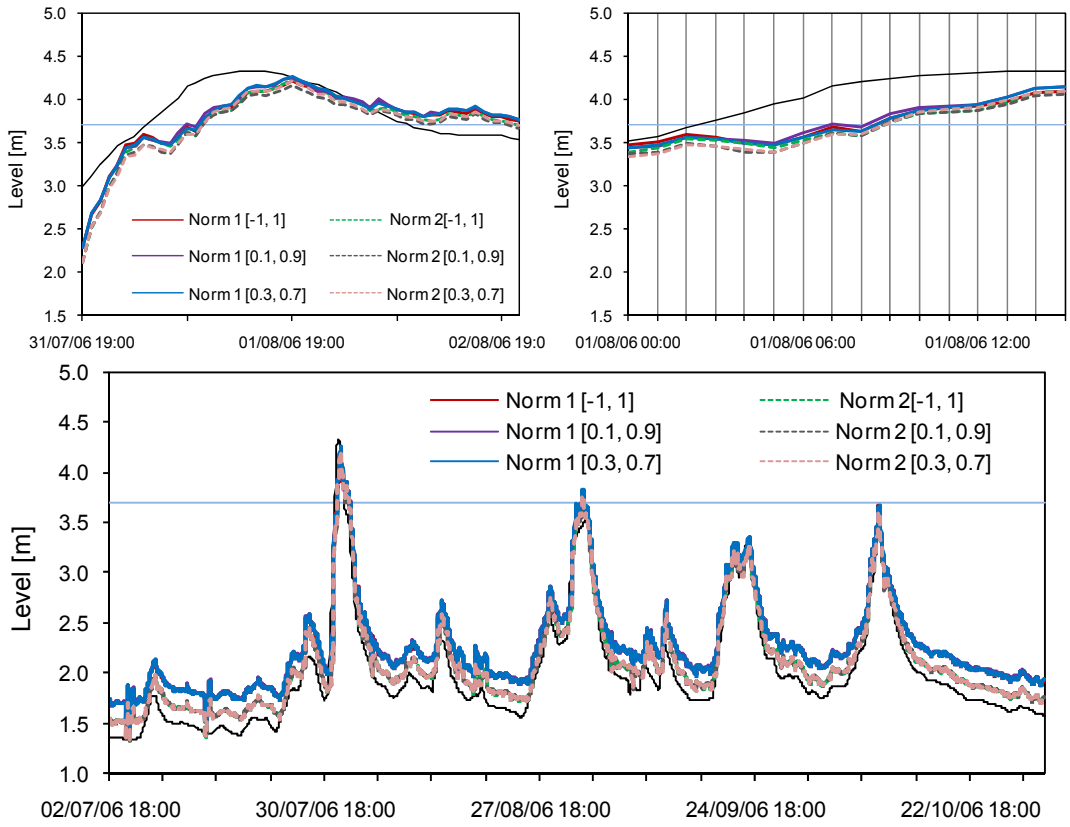


Figure 7.12: Results for a storm in 2006 comparing different ranges of normalization for Model 1.

Table 7.11 shows the minimum and maximum values for Model 8. This time, instead of taking all values across the historical data record, an artificial maximum above the real maximum has been used. This is to allow more room at the top end of the prediction since it was clear that normalizing using the actual minimum and maximum resulted in an under prediction of the peak.

Table 7.11: The minimum and maximum values in the training and testing datasets for Model 8 including an artificial maximum.

Input variables	Minimum				Maximum			
	Training	Testing	Across all historical records	Artificial	Training	Testing	Across all historical records	Artificial
Z2	0	0	0	0	186.00	196.22	196.22	250
Z3	0	0	0	0	178.44	185.56	185.56	250
Z4	0	0	0	0	171.56	205.78	205.78	250
P1	1.66	1.46	1.66	1.46	4.93	4.90	4.93	6.00

Table 7.12 provides performance statistics using the different normalization ranges and the artificial maximum value when normalized between [-1 and 1]. Once again the best results are obtained using the minimum and maximum across the entire historical record for a narrower range [0.1, 0.9] although, as before, there are little differences in the performance measures between the models.

Table 7.12: Goodness of fit statistics for Model 8 using the min/max from the training dataset for normalization compared to using artificial maximum values.

Goodness of fit	Normalization using min/max from the training dataset (Norm1)			Normalization using min/max from across the entire historical record (Norm2)			Using artificial values (Norm3)
	[-1, 1]	[0.1, 0.9]	[0.3, 0.7]	[-1, 1]	[0.1, 0.9]	[0.3, 0.7]	[-1, 1]
PDIFF	**0.0082**	-0.0555	0.0611	0.1272	0.1136	0.183	0.0171
MAE	0.6622	0.6661	0.6641	0.5776	**0.5746**	0.5774	0.5764
RMSE	0.8129	0.8145	0.8135	0.7734	**0.7728**	0.7735	0.7738
CE	0.1715	0.1681	0.1701	0.2499	**0.2512**	0.2498	0.2491

(Bold and shading denotes the best performer.)

Figure 7.13 provides a detailed look at a storm event in 2005 in the testing data set. The results show little difference between using the minimum and maximum from training or an artificial maximum although slight overestimations in peak predictions can be observed.

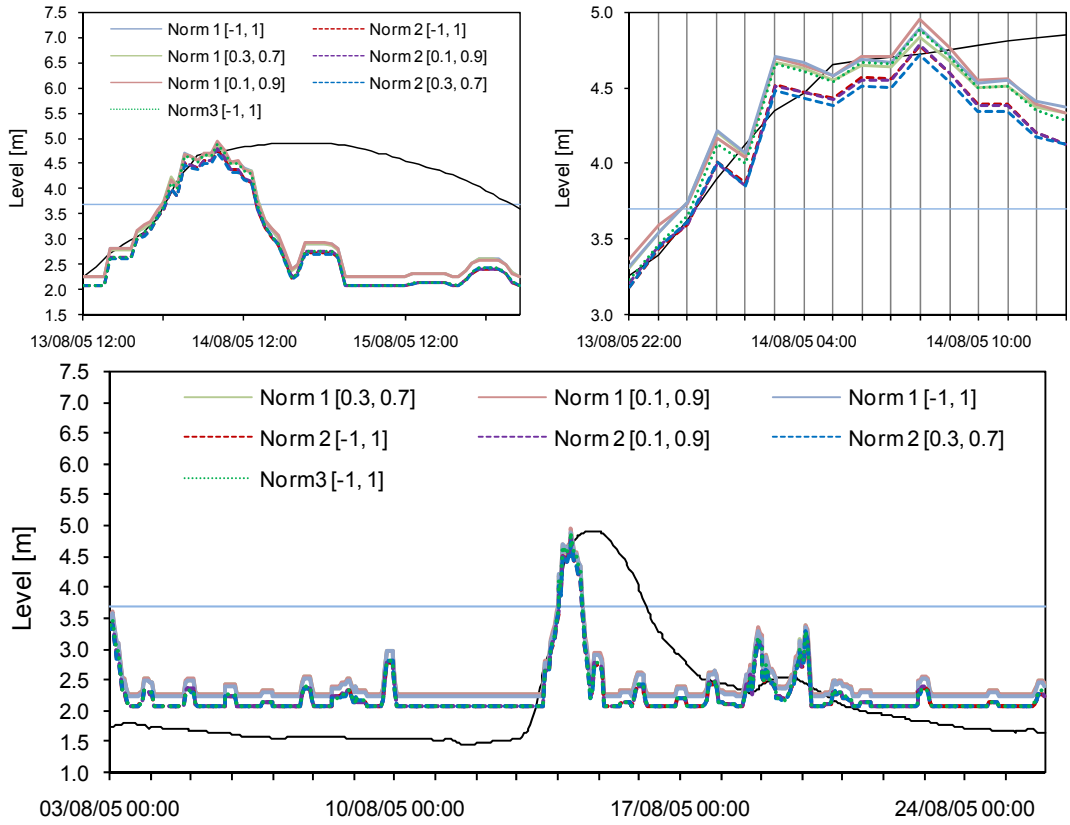


Figure 7.13: Results for a storm in 2005 comparing normalization using the maximum in the training dataset vs normalization using an artificial maximum value for Model 8.

In summary, when using the minimum and maximum from the entire historical record for normalization, the MAE, RMSE and CE are better than taking either the minimum and maximum from the training dataset or using an artificial maximum. Regardless of which normalization range was used, the error measures were quite similar to one another. The range of normalization therefore appears to have less effect than the values of the minimum and maximum used for normalization. This is clearly an area that still requires further experimentation.

7.5 Conclusions

This chapter investigated three aspects of neural network modelling: the number of runs needed in the BR algorithm; the number of hidden nodes to choose; and what range of normalization to select. The best models were selected from the previous chapters for further experimentation. Different numbers of runs were used in the averaging process from 5 to 50 at intervals of 5. The results clearly showed that 50 runs is the best number to choose. It is also evident that a sufficient number of runs is needed to average out the variation between individual runs, which is a function of weight initialization. However, this is rarely reported in the literature. Most papers

appear to use a single run and are unaware of the consequences of the effect of weight initialization and ultimately the uncertainty in the model predictions.

The experimentation with hidden nodes was generally inconclusive. Using less hidden nodes than the number of inputs (whether 50% or 75%) generally produced better performance statistics. However, as the lead time of the model increased, there were situations where 2n+1 hidden nodes produced a better result. Inspection of the hydrographs showed that the least parsimonious model sometimes produced a better peak prediction but with other consequences, i.e. poorer behaviour on other parts of the hydrograph. Therefore trial and error remains the best method for determining the number of hidden nodes until further guidance appears in the literature.

Experiments with normalization of the data provided no conclusive evidence regarding which range to choose, i.e. a wide range such as [-1, 1] or a narrower one such as [0.3, 0.7]. However, it did show that choosing a larger minimum and maximum (either from the entire historical record or using an artificial maximum higher than the maximum in the training data set) was better than choosing the minimum and the maximum from the training data set, which is common practice, especially if prediction of the peak is the main concern.

Chapter 8 Conclusions

8.1 Introduction

This chapter summarizes the research findings of this study in terms of the original aims and objectives to demonstrate how these have been achieved and to highlight what the main findings of the research are. This is followed by a discussion of the main limitations encountered during the research. This final section forms a set of recommendations for future work that should be carried out by researchers in the area of neural network flood forecasting.

8.2 Summary of the Research Findings

The thesis has made several significant contributions to the knowledge of neural network flood forecasting. These contributions will be highlighted in this section. The overall aim of this thesis was to determine the most effective neural network approach to forecasting stage in a large monsoon-fed river system through multiple experiments with different inputs and neural network parameterisations. The main findings of the research are placed within the context of the original objectives of the research (as listed in Chapter 1) to demonstrate how these objectives have been achieved and the significance of the research findings.

Objective 1: To review the relevant literature on neural network modelling in hydrology

A comprehensive review of the literature on neural network modelling was undertaken in Chapter 2. Neural networks were first placed within the general framework of approaches to hydrological modelling. An overview of neural networks was then provided along with different applications in hydrological modelling that have appeared in the literature over the last two decades. During this review, some important issues regarding the lack of guidance on neural network model development were raised. One of these issues was regarding how to choose the inputs to the model. A more detailed review of this issue was undertaken as part of Objective 2 and was the subject of subsequent experimentation in Objective 3. In addition to input determination, the lack of guidance with respect to other model development decisions was also highlighted, including how to choose the number of hidden nodes and the range of normalization. Further experimentation with these issues was undertaken as part of Objective 5.

The literature review then summarised the use of neural networks in hydrology and water resource management through the use of a search engine. It was clear from this exercise that neural networks are being used with increasing momentum in this field. Finally, hydrological modelling (both physical/conceptual and data-driven) was reviewed for the study area. Although some applications do exist, many are not relevant in terms of the temporal resolution of the forecasts or the results were not presented in sufficient detail to make any reasonable conclusions about model performance. Moreover, the lead times of these models were generally short, i.e. 12 hours or less. Attempts at extending the lead time of the forecasts using radar data has been tackled further in Objective 4.

Objective 2: To review and critically evaluate existing input determination techniques

A critical evaluation of input determination techniques was undertaken and the results were presented in Chapter 4. Of the different methods available, a subset was chosen for subsequent experimentation (i.e. Objective 3) based on two criteria: the ability to automate the method and ease of implementation. The final techniques chosen for experimentation included: correlation; stepwise regression; a combination of correlation and stepwise regression, a genetic algorithm; M5 model trees, a data mining search algorithm; a pruning algorithm and PMI (partial mutual information). No such comprehensive review has yet been published apart from the section on input determination methods by Maier et al. (2010). Thus, this chapter could be turned into a valuable review paper for researchers to build upon.

Objective 3: To experiment with a range of different input determination techniques

Building on the review undertaken in Objective 2, the eight methods were evaluated in a series of different experiments involving the development of neural networks for the Upper Ping catchment. The results of the neural network models were evaluated using a series of commonly employed goodness-of-fit statistics and visual inspection of the hydrographs. The set up and results of these experiments were provided in Chapter 5.

Using all the inputs did not produce the best model so it is clear that some form of input determination technique is required. The use of stepwise regression, genetic algorithms and pruning algorithms resulted in a reduction of input variables by 60-70%, while the combination correlation and stepwise regression reduced input variable of approximately 70-80%. Other input determination methods such as the data mining

search method and PMI reduced the number of inputs by approximately 80-90%. The M5 model tree resulted in a reduction of approximately 30%. The PMI method did not produce a satisfactory set of variables since only P1 was selected, there were no upstream stations. The data mining search method produced the best peak prediction but had the greatest delay in the rising limb of the hydrograph. In contrast, the genetic algorithm and stepwise regression predicted a reasonable peak and a better rising time than the data mining search method although the performance of the stepwise regression method was better than that of the genetic algorithm. In contrast, the combination of correlation and stepwise regression showed the strongest performance on the rising limb of the hydrograph, and produced a better overall performance than either correlation or stepwise regression used on their own. Therefore, the most successful technique emerging from this study is the combination of correlation and stepwise regression. This method has not appeared before in the literature and represents a simple yet innovative approach to input determination. Moreover, such a comprehensive experimentation with input determination methods has not been published and therefore represents a potential contribution to the scientific literature.

In addition, the overall neural network performance in predicting water stage at P1 station using three water stage stations as input variables can forecast at a maximum lead time of approximately 18 hrs with a 3-8 hr delay in the rising limb and a 0.1-10 cm error in peak prediction. However, the model performance depends on the wet/dry conditions in the catchment and physical changes in the river.

Objective 4: To experiment with radar data as an input to the neural network models as a way of improving the model accuracy and extending the lead time of the forecasts

This objective was achieved by adding raw radar reflectivity data to the neural network models in a series of different experiments. The setup of these experiments and the results were presented in Chapter 6. The results clearly showed that it is possible to increase the lead time of the forecast by a considerable amount, i.e. 24-30 hr lead time with 0.1 cm error at the peak but unfortunately, the model could not predict the full hydrograph, especially the falling limb. Using both radar and three water stage stations as input variables does improve the falling limb but the lead time of prediction drops back to 18 hours with a 3 cm error in pea prediction. Therefore separate NN models should be used for different lead times.

In addition, increasing the number of sample points across the catchment resulted in a slight improvement in the model performance. It was shown that the pattern of storm movement affects the model performance with the best results achieved for the large storm in 2005. Unfortunately, the model performance was degraded for storms in other years, i.e. 2003 and 2006.

It was hypothesised that storm movement patterns and wet/dry conditions in the catchment will influence the ability of the neural networks to accurately predict the flood using radar data. The results here can only be considered to be indicative. With such a small number of storm events for training and testing, and missing radar data for the first two storms in 2006, the model could not be calibrated separately for different rainfall patterns with enough confidence to draw conclusions. It was not possible to make vast improvements to the forecasts for these years. However, the potential of using radar reflectivity data in this way was clearly demonstrated. Moreover, radar data have not been used in this way before in NN flood forecasting and therefore represents a significant scientific contribution. This has already been proven through the recent publication by Chaipimonplin et al. (2010).

Objective 5: To investigate model improvements through experimentation with the training algorithm and internal neural network parameters.

This objective was tackled through a series of experiments with the BR training algorithm, the number of hidden nodes and the range of normalisation of the input data. It was shown in Chapter 5 that Bayesian Regularization performed better than using the Levenberg-Marquardt algorithm on its own. However, the neural network must be trained many times and the results averaged to produce a good result in order to average out the variation between individual runs. Experiments showed that 50 runs was a good number to use in creating an average prediction and it confirms the number used by Anctil (2007).

Experiments were undertaken whereby the number of hidden nodes was set based on taking 50% of the number of inputs, 75% of the number of inputs and 2n+1 hidden nodes where n is the number of inputs. Using less hidden nodes than the number of inputs (whether 50% or 75%) generally produced better performance statistics. However, as the lead time of the model increased, there were situations where 2n+1 hidden nodes produced a better result, perhaps because the modelling problem at longer lead times becomes more complex. Inspection of the hydrographs showed that the least parsimonious model sometimes produced a better peak prediction but with

other consequences, i.e. poorer behaviour on other parts of the hydrograph. Therefore trial and error remains the best method for determining the number of hidden nodes until further guidance appears in the literature.

Experiments were also undertaken with different ranges of normalization ([-1, 1], [0.1, 0.9] and [0.3, 0.7]) using the minimum and maximum from (a) the training data set (the most common practice); (b) across the entire historical record; and (c) using an artificial maximum value. The hydrographs revealed that there is not much different between the three ranges. However, it did show that choosing a larger minimum and maximum (either from the entire historical record or using an artificial maximum higher than the maximum in the training data set) was better than choosing the minimum and the maximum from the training data set itself.

The research undertaken as part of this objective has added a scientific contribution to the growing literature on what parameters and architecture to use.

Objective 6: To highlight the limitations of the study and to make recommendations for areas for further research.

The limitations and recommendations for further research are outlined in the next section, which comprises a short research agenda for the future. This section also concludes the thesis.

Overall the results have therefore shown that, despite all the data limitations, neural networks can be applied for effective flood prediction in the Upper Ping, which suggests that they should also be considered for use in similar large monsoon-fed catchments. The most viable results were achieved for a lead time of approximately 15 hours with a 1.9 cm error at the peak. However, when data from radar images was added, this lead time could be increased to 24 to 30 hours in the future with only 0.1 cm error at the peak (Table 8.1). This is very helpful when defence and evacuation plans have to be put into operation in a large and complex city like Chiang Mai.

Table 8.1: Summary the best model by PDIFF and error in the time of the rising limb at t+18 hr.

Section		2001	S1	S2	S4	S5	2006
5.3	Time (hr)	7.5	/	/	/	/	/
	PDIFF (m)	0	/	/	/	/	/
5.4	Time (hr)	/	/	3.5	9	5.5	/
	PDIFF (m)	/	/	-0.019	-0.011	0.083	/
5.5	Time (hr)	/	/	3	8.5	6	/
	PDIFF (m)	/	/	-0.447	0.0788	-0.018	/
5.6	Time (hr)	/	/	/	/	/	3.5
	PDIFF (m)	/	/	/	/	/	0.074
6.6	Time (hr)	/	-24	-2	/	2	/
	PDIFF (m)	/	-0.063	-0.066	/	0.134	/
6.7*	Time (hr)	/	-17	-1	/	/	/
	PDIFF (m)	/	-0.610	-0.001	/	/	/
6.10	Time (hr)	/	/	2	10	9	/
	PDIFF (m)	/	/	-0.030	-0.360	-0.353	/

*At t+24 hr

Although some modelling (both conceptual/physical and data-driven) has been undertaken in the Upper Ping catchment (as detailed in Chapter 2), it is difficult to compare the results found here directly with any of these studies. However, an attempt to compile the results is shown below in Table 8.2.

Table 8.2: Comparison of results from this research and previous studies.

Goodness of fit	Testing 2005 at t+12 hr		Test 2006	
	This research (1 Aug-4 Nov)	Chidthong *et al.* (2009) (1-31Aug/1-30 Sep)	This research t+18 hr (2 Aug-31Oct)	Ninprom and Chumchean (2009) t+24 hr (29 Aug-20 Sep)
RMSE (m)	0.125	0.100/0.137	0.111	0.52

The study by Chidthong *et al.* (2009) produces results that are comparable to those found here although the period of testing (and therefore the period over which the statistics were calculated) is different. The study by Ninprom and Chumchean (2009) was undertaken for a longer lead time and a shorter testing period so the higher RMSE may be partly artificial. This table also illustrates the lack of previous work in this area using hourly data.

8.3 Limitations of the Research

The first major limitation of this study is the amount of data available for the Upper Ping catchment. There was only one hourly rain gauge available and that was very close to Chiang Mai so coverage of this large catchment was not good. Therefore, more hourly rain gauges are needed higher up in the catchment. Moreover, there were only three water stage stations available that recorded hourly stage data, i.e. P75, P67 and P1,

where the travel time between P1 and P75 is less than 24 hours. Adding water level stations higher up in the catchment may improve the model performance. There were no hourly data available from station P20, which is further up in the catchment, and has been used in previous forecasting daily models (Patsinghasanee *et al*., 2004; Chidthong and Supharatid, 2007; Chidthong *et al*., 2009). This fact also limited the opportunity to compare the performance of these models with the previous studies. Further attempts should be made to obtain data for this station in the future. Secondly, there are problems with the locations of other water stage stations, i.e. P21, P4a and P75. For example, P21 is not located on the main river and P75 is located downstream of a dam, so the water stage at P75 is affected by releases from the dam. Future models could potentially be improved if the timing records of water release from the dam could be obtained and integrated. Finally, as the Ping catchment is a large and complex catchment, increasing the number of hidden layers from one to two might improve the model performance (Bodri and Cermak, 2000).

The second limitation involves the use of radar image data and information about the direction in which the storm is moving. Clearly the reflectivity imagery data used here are a limited surrogate for precipitation data. In theory the easiest floods to predict will be for those events where storms move across the catchment from the north-east, covering the upstream stations and then P1. When the storms come from a southwesterly direction, the storm will hit the lower catchment first and then move north. As a consequence, the neural network model may fail to predict the floods at P1 in this situation. More data are needed, in particular more radar data (capturing a greater variety of storm types) as data representative of all situations should be included in the training data (Maier *et al*., 2010). The study was limited by the amount radar data that was available, i.e. only images from 2005 and 2006. More images will be obtained in the future in order to try to pick out storm patterns, possibly training different neural networks to handle different types of storms.

8.4 Recommendations for Futher Research

Maier et al. (2010) make a series of six recommendations for future work in the area of NN modelling in hydrology. The research undertaken in this thesis specifically addresses three of these recommendations, which clearly shows the relevance of the research questions and the subsequent findings. These recommendations and their relationship to this research are summarised below:

1. Work should continue on the development of hybrid models – although hybrid models are generally considered to be the integration of more than one model, e.g. ensemble NNs or neuro-fuzzy, the research in this thesis has paved the way for hybrid modelling research in the future, in particular with respect to capturing different storm patterns. The research has clearly shown that some storms were predicted better than others, where it was hypothesised that this may be a result of different monsoonal storm patterns. A recommendation for further research is therefore to build different NN models to capture different storm patterns and to then integrate these individual NN models into a single hybrid solution.

2. More work should be undertaken on investigating methods of input determination – a huge portion of this research has dealt with exactly this recommendation as it was recognised by the author as an important area of work. This research represents the first systematic attempt at testing a range of automated methods for NN modelling. The results showed that simple methods such as correlation or stepwise linear regression and correlation together (which is something tried out in this research for the first time) represent good methods. However, when the problem becomes more complex, then other methods such as PMI, a genetic algorithm and data mining may produce better solutions. It is clear, however, that much work needs to continue in this area to see whether patterns or guidelines emerge that can help NN modellers make the best choice of input determination method.

3. More work should continue on determining the optimal NN structure and architecture – Chapter 7 dealt with the recommendation by undertaking a series of experiments with varying model parameters. It may well be that no optimal solution actually exists but that there are many good possible solutions. The findings in chapter 7 showed that modellers using Bayesian Regularisation should never use a single model instance but develop several and then integrate the results either through averaging or some other ensemble method. Here is definitely an area where further research could be undertaken in terms of how to best develop a NN using the BR algorithm. In terms of the number of hidden nodes, the research showed that trial and error may still be the best way to determine this parameter until some type of more automated selection process appears. Finally, in terms of the ranges of normalisation, the significant finding here was that the range should be wider than the actual range of the training data set. Thus the research has added valuable knowledge to this area of determining an optimal NN structure and

architecture but it is clear that much more work of the type that was undertaken in Chapter 7 is recommended for further research.

Another absolutely crucial area for further investigation is the use of radar data in neural network modelling. The potential for using raw reflectivitiy data as an input to NN models for flood forecasting represents one of the biggest innovations of this research. Further experiments should be undertaken to assess how useful the data could be in other catchments and other types of river systems. More specifically, further research should involve increasing the number of sample points over the upper catchment and increasing the area of sampling to see what effect this has on the accuracy of the forecasts. The problem with the prediction of the falling limb of the hydrograph also needs to be resolved. Chumchean (2007) investigated the Z-R relationship in Thailand and she suggested that the radar reflectivity less than 15 dBZ and greater than 53 dBZ should be ignored to prevent errors due to noise or from hail. Thus, the input data should be preprocessed to remove these values before they are input to the NN. This may already lead to a better performance at lower water levels where the predictions are currently poor when using radar data alone. This is clearly an area of great potential where many new research questions can be asked in the future.

References

Abbott, M.B., Bathurst, J.C., Cunge, J.A., Connell, P.E.O. and Rasmussen, J. (1986a) An introduction to the European Hydrological System-System Hydrologique Europeen, "SHE", 1: history and philosophy of a physically-based distributed modeling system. *Journal of Hydrology,* 87, 45-59.

Abbott, M.B., Bathurst, J.C., Cunge, J.A., Connell, P.E.O. and Rasmussen, J. (1986b) An introduction to the European Hydrological System-System Hydrologique Europeen, "SHE", 2: structure of a physically-based distributed modeling system. *Journal of Hydrology,* 87, 61-77.

Abrahart, R.J. and See, L. (2000) Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological Processes,* 14, 2157-2172.

Abrahart, R.J., See, L.M. and Heppenstall, A.J. (2007) Neuroevolution applied to river level forecasting under winter flood and drought conditions. *Journal of Intelligent Systems*, 16(4), 373-386.

Abrahart, R.J., See, L. and Kneale, P.E. (1999) Using pruning algorithms and genetic algorithms to optimise network architectures and forecasting inputs in a neural network rainfall-runoff model. *Journal of Hydroinformatics,* 1, 103-114.

Abrahart, R.J., See, L. and Kneale, P.E. (2001) Investigating the role of saliency analysis with a neural network rainfall-runoff model. *Computers & Geosciences,* 27, 921-928.

Abrahart, R.J., See, L.M., Dawson, C.W., Shamseldin, A.Y. and Wilby, R.L. (2010) Nearly Two Decades of Neural Network Hydrological Modeling. In: Sivakumar, B. and Berndtsson, R. (Eds.) *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting*. World Scientific.

Achela, D., Fernando, K. and. Shamseldin, A.Y. (2009) Investigation of Internal Functioning of the Radial-Basis-Function Neural Network River Flow Forecasting Models, *Journal of Hydrologic Engineering*, 14(3), 286-292.

Ali, Y.A., Ramlan, M. and Rahman, R.A. (2010) Dynamic Bayesian networks and variable length genetic algorithm for designing Cue-based model for dialogue act recognition. *Computer Speech and Language,* 24, 190-218.

Allen, G. and Le Marshall, J.F. (1994) An evaluation of neural networks and discriminate analysis methods for application in operational rain forecasting. *Australian Meteorological Magazine,* 43, 17-28.

ASCE (2000a) Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering,* 5, 115-123.

ASCE (2000b) Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering,* 5, 124-137.

Anctil, F. (2007) Tools for the assessment of hydrological ensemble forecasts. In: Proceedings of the International Workshop on Advances in Hydroinformatics 2007, Coulibaly, P. (Eds.), 4-7 June, Niagara Falls, Canada.

Anctil, F. and Lauzon, N. (2004) Generalisation for neural networks through data sampling and training procedures, with applications to streamflow predictions. *Hydrology and Earth System Sciences,* 8, 940-958.

Anctil, F., Lauzon, N., Andreassian, V., Oudin, L. and Perrin, C. (2006) Improvement of rainfall-runoff forecasts through mean areal rainfall optimization. *Journal of Hydrology,* 328, 717-725.

Anctil, F., Michel, C., Perrin, C. and Andreassian, V. (2004a) A soil moisture index as an auxiliary ANN input for stream flow forecasting. *Journal of Hydrology,* 286, 155-167.

Anctil, F., Perrin, C. and Andreassian, V. (2004b) Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models. *Environmental Modeling & Software,* 19, 357-368.

Anderson, M.G. and Burt, T.P. (1985) Modeling strategies. IN Anderson, M. G. and Burt, T. P. (Eds.) *Hydrological Forecasting.* Suffolk, John Wiley & Sons Ltd.

Aqil, M., Kita, I., Yano, A. and Nishiyama, S. (2007) A comparative study of artificial neural networks and Neuro-Fuzzy in continuous modeling of the daily and hourly behaviour of runoff. *Journal of Hydrology,* 337, 22-34.

Battan, L.J. (1973) *Radar Observation of the Atmosphere,* Chicago, University of Chicago Press.

BBC (2001) *Thailand floods kill 70* [Online], [Accessed 19 July 2010], Available at http://news.bbc.co.uk/1/hi/world/asia-pacific/1487394.stm

BBC (2002) *Flood devastation in northern Thailand* [Online], [Accessed 19 July 2010], Available at http://news.bbc.co.uk/1/hi/world/asia-pacific/2241130.stm

BBC (2007) *Fatal flash flooding in Thailand by Nina Ridge* [Online], [Accessed 19 July 2010], Available at http://www.bbc.co.uk/weather/world/news/12092007news.shtml

Beven, K.J. and Kirby, M.J. (1979) A physically based variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin,* 24, 43-69.

Bhattacharya, B. and Solomatine, D.P. (2005) Neural networks and M5 model trees in modeling water level-discharge relationship. *Neurocomputing,* 63, 381-396.

Blackie, J.R. and Eeles, C.W.O. (1985) Lumped catchment models. IN Anderson, M.G. and Burt, T.P. (Eds.) *Hydrological Forecasting.* Suffolk, John Wiley & Sons. Ltd.

Bodri, L. and Cermak, V. (2000) Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Movavia. *Advances in Engineering Software,* 31, 311-321.

Boochabun, K., Tych, W., Chappell, N.A., Carling, P.A., Lorsirirat, K. and Pa-Obsaeng, S. (2004) Statistical modeling of rainfall and river flow in Thailand. *Journal of the Geological Society of India,* 64, 503-515.

Borga, M. (2002) Accuracy of radar rainfall estimates for streamflow simulation. *Journal of Hydrology,* 267, 26-39.

Bowden, G.J., Dandy, G.C. and Maier, H.R. (2005a) Input determination for neural network models in water resources applications. Part 1-background and methodology. *Journal of Hydrology,* 301, 75-92.

Bowden, G.J., Maier, H.R. and Dandy, G.C. (2005b) Input determination for neural network models in water resources applications. Part 2. case study: forecasting salinity in a river. *Journal of Hydrology,* 301, 93-107.

Bowden, G.J., Nixon, J.B., Dandy, G.C., Maier, H.R. and Holmes, M. (2006) Forecasting chlorine residuals in a water distribution system using a general regression neural network. *Mathematical and Computer Modeling,* 44, 469-484.

Brace, N., Kemp, R. and Snelgar, R. (2006) *SPSS for Psychologists A Guide to Data Analysis Using SPSS for Windows (Version 12 and 13)*, Palgrave Macmillan.

Braddock, R.D., Kremmer, M.L. and Sanzogni, L. (1998) Feed-forward artificial neural network model. *Water Resources Journal,* 35, 1191-1197.

Breierova, L. and Choudhari, M. (2001) *An introduction to sensitivity analysis* [Online], [Accessed 12 September 2009], Available at http://sysdyn.clexchange.org/sdep/Roadmaps/RM8/D-4526-2.pdf

Browning, K.A. (1987) Towards the more effective use of radar and satellite imagery in weather forecasting. IN Collinge, V.K. and Kirby, C. (Eds.) *Weather Radar and Flood Forecasting.* John Wiley & Sons Ltd.

Campolo, M., Andreussi, P. and Soldati, A. (1999) River flood forecasting with a neural network model. *Water Resources Research,* 35, 1191-1197.

Cannas, B., Fanni, A., See, L. and Sias, G. (2006) Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Physics and Chemistry of The Earth,* 31, 1164-1171.

Carcano, E.C., Bartolini, P., Muselli, M. and Piroddi, L. (2008) Jordan recurrent neural network versus IHACRES in modeling daily streamflows. *Journal of Hydrology*, 362(3-4), 291-307.

Caudill, M. (1991) Neural network training tips and techniques. *AI Expert*, 6(1), 56-61.

Chaipimonplin, T., See, L.M. and Kneale, P.E. (2010) Using radar data to extend the lead time of neural network forecasting on the River Ping. *Disaster Advances*, 3(3), 35-43.

Chaipimonplin, T., See, L.M. and Kneale, P.E. (2008a) Neural network prediction of flooding in Chiang Mai, Thailand: comparison of input determination techniques. EGU, Vienna, Austria, 13-18 April 2008.

Chaipimonplin, T., See, L.M. and Kneale, P.E. (2008b) Use of neural network to predict flooding in Chiang Mai, Thailand: comparison of input determination techniques. AOGS, Pusan, South Korea, 17-19 June 2008.

Chang, F.J. and Hwang, Y.Y. (1999) A self-organization algorithm for real-time flood forecast. *Hydrological Processes,* 13, 123-138.

Chang, F.J., Chang, L.C. and Huang, H.L. (2002) Real-time recurrent learning neural network for stream-flow forecasting. *Hydrological Processes*, 16(13), 2577-2588.

Chang, L-C., Chang, F-J. and Chiang, Y-M. (2004) A two-step-ahead recurrent neural network for stream-flow forecasting. *Hydrological Processes*, 18(1), 81-92.

Chang, F.-J., Chang, L.-C. and Wang, Y.-S. (2007) Enforced self-organizing map neural networks for river flood forecasting. *Hydrological Processes,* 21, 741-749.

Chaohong, S., Qiang, L. and Feng, S. (2008) A Bayesian dynamic forecast model based on neural network, *2008 International Symposium on Intelligent Information Technology Application Workshop.* IEEE Computer Soc.

Chatchawan, S. (2005) The report of flood's situation and resolving at Chiang Mai City 2005 (published in Thai). *2005 Annual Conference.* Chonburi.

Chatfield, C. (2001) *Time-Series Forecasting,* Florida, Chapman & Hall/CRC.

Chen, S.H., Lin, Y.H., Chang, L.C. and 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. and Yu, P.S. (2007) Pruning of support vector networks on flood forecasting. *Journal of Hydrology,* 347, 67-78.

Chiang, Y.M., Chang, F.J., Jou, B.J.-D. and Lin, P.F. (2007) Dynamic ANN for precipitation estimation and forecasting from radar observations. *Journal of Hydrology,* 334, 250-261.

Chidthong, Y. and Supharatid, S. (2007) Developing a hybrid multi model for peak flood forecasting (Case study: Chiang Mai 2005 flood), *The 12th National Convention on Civil Engineering (published in Thai)* Phitsanulok, Thailand.

Chidthong, Y., Tanaka, H. and Supharatid, S. (2009) Developing a hybrid multi-model for peak flood for forecasting. *Hydrological Processes,* 23, 1725-1738.

Chumchean, S. (2007) Investigation of climatological Z-R relationship for Pasicharoen radar, *The Second National Convention on Water Resources Engineering (published in Thai).* Kasetsart University, Bangkok, Thailand: The Engineering Institute of Thailand under H.M. The King's Patronage.

Chumchean, S., Seed, A. and Sharma, A. (2008) An operational approach for classifying storms in real-time radar rainfall estimation. *Journal of Hydrology,* 363, 1-17.

Cigizoglu, H.K. (2005) Generalized regression neural network in monthly flow forecasting. *Civil Engineering and Environmental Systems,* 22, 71-84.

Cole, S.J. and Moore, R.J. (2008) Hydrological modeling using raingauge and radar based estimators of areal rainfall. *Journal of Hydrology,* 358, 159-181.

Collier, C.G. (1987) Accuracy of real-time radar measurements. IN Collinge, V.K. and Kirby, C. (Eds.) *Weather Radar and Flood Forecasting.* John Wiley & Sons.

Collier, C.G. (1996) *Applications of Weather Radar Systems,* Chichester, West Sussex, John Wiley & Sons Ltd.

Corani, G. and Guariso, G. (2005) An application of pruning in the design of neural networks for real time flood forecasting. *Neural Computing and Applications,* 14, 66-77.

Corzo, G.A., Solomatine, D.P., Hidayat, Wit, M.D., Werner, M., Uhlenbrook, S. and Price, R.K. (2009) Combining semi-distributed process-based and data-driven models in flow simulation: a case study of the Meuse river basin. *Hydrology and Earth System Sciences,* 13, 1619-1634.

Crespo, J.L. and Mora, E. (1993) Drought estimation with neural networks. *Advances in Engineering Software,* 18, 167-170.

Coulibaly, P., Bobee, B. and Anctil, F. (2001) Improving extreme hydrologic events forecasting using a new criterion for artificial neural network selection. *Hydrological Processes*, 15(8), 1533-1536.

Danh, N.T., Phien, H.N. and Gupta, A.D. (1999) Neural network models for river flow forecasting. *Water SA,* 25, 33-39.

Dawson, C.W., Abrahart, R.J. and See, L.M. (2007) HydroTest: a web-based toolbox of statistical measures for the standardised assessment of hydrological forecasts. *Environmental Modeling & Software,* 27, 1034-1052.

Dawson, C.W., Abrahart, R.J., Shamseldin, A.Y. and Wilby, R.L. (2006a) Flood estimation at ungauged sites using artificial neural networks. *Journal of Hydrology,* 319, 391-409.

Dawson, C.W., Brown, M.R. and Wilby, R.L. (2000) Inductive learning approaches to rainfall-runoff modeling. *International Journal of Neural Systems,* 10, 43-57.

Dawson, C.W., See, L.M., Abrahart, R.J. and Heppenstall, A.J. (2006b) Symbiotic adaptive neuron-evolution applied to rainfall-runoff modeling in Northern England. *Neural Networks,* 19, 236-247.

Dawson, C.W. and Wilby, R. (1998) An artificial neural network approach to rainfall runoff modeling. *Hydrological Sciences Journal,* 43, 47-66.

Dawson, C.W. and Wilby, R.L. (1999) A comparison of artificial neural networks used for river flow forecasting. *Hydrology and Earth System Sciences,* 3, 529-540.

Dawson, C.W. and Wilby, R.L. (2000) 'An evaluation of artificial neural network techniques for flow forecasting in the River Yangtze', European Geophysical Society General Assembly, Nice.

Dawson, C.W. and Wilby, R.L. (2001) Hydrological modeling using artificial neural networks. *Progress in Physical Geography,* 25, 80-108.

Dejyothin, C. (2008) *Automatic Value Adjustment Systems for the Relationship between Radar Reflectivity and Rain Volume Measured by Rain-Gauge* [Online], [Accessed 12 October 2008], Available at http://www.royalrainmaking.thaigov.net/MainITRoyalrain.php?p=13

Demuth, H., Beale, M. and Hogan, M. (2009) *Neural Network Toolbox TM 6: User's Guide* [Online], [Accessed 1 November 2009], Available at http://www.mathworks.com

Department of Water Resources (2007) *Ping catchment* [Online], [Accessed 16 November 2008], Available at http://mekhala.dwr.go.th/mekhala/Basin_North.asp

De Vos, N.J. and Rientjes, T.H.M. (2007) Multi-objective performance comparison of an artificial neural network and a conceptual rainfall-runoff model. Hydrological Sciences Journal, 52(3), 397-413.

Draper, N.R. and Smith, H. (1966) *Applied Regression Analysis,* London, John Wiley & Sons.

Eiamkarn, N., Chanyotha, S. and Thaksa-udom, C. (2009) Effects of Ping morphology changes on flood flow, *The 14th National Convention on Civil Engineering (published in Thai)* Nakhonratchasima, Thailand.

Ei-Shafie, A., Abdin, A.E., Noureldin, A. and Taha, M.R. (2009) Enhancing inflow forecasting model at Aswan high dam utilizing radial basis neural network and upstream monitoring stations measurements. *Water Resources Management,* 23, 2289-2315.

Fahlman, S.E. and Lebiere, C. (1990) The cascade-correlation learning architecture. IN Touretzky, D.S. (Eds.) *Advance in Neural Information Processing Systems 2.* Los Altos, CA, Morgan Kaufmann Publishers.

Fernando, T.M.K.G., Maier, H.R. and Dandy, G.C. (2009) Selection of input variables for data driven models: an average shifted histogram partial mutual information estimator approach. *Journal of Hydrology,* 367, 165-176.

Filho, A.J.P. and Santos, C.C.D. (2006) Modeling a densely urbanized watershed with an artificial neural network, weather radar and telemetric data. *Journal of Hydrology,* 317, 31-48.

Firat, M. and Turan, M.E. (2010) Monthly river flow forecasting by an adaptive neuro-fuzzy inference system. *Water and Environment Journal*, 24(2), 116-125.

Fisher, B., Wolff, D.B. and Amitai, E. (2000) Analytical software for validation time synchronous radar and raingauge data, *Remote Sensing and Hydrology 2000.* Santa Fe, New Mexico, USA: IAHS.

Fletcher, D. and Goss, E. (1993) Applications forecasting with neural networks; An application using bankruptcy data. *Information and Management*, 24, 159-167.

Foresee, F.D. and Hagan, M.T. (1997) Gauss-Newton approximation to Bayesian learning, *Proceedings of the 1997 International Joint Conference on Neural Network*, 1930-1935.

Franchini, M. (1996) Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models. *Hydrological Sciences Journal,* 41, 21-39.

Fun, M.-H. and Hagan, M.T. (1996) Levenberg-marquardt training for modular networks, *Proceedings of the 1996 International Conference on Neural Network*, 1, 468-473.

Georgrakakos, K.P., Tsintikidis, D., Attia, B. and Roskar, J. (2000) Estimation of pixel-scale daily rainfall over Nile River catchments using multi-spectral METEOSAT data, *Remote Sensing and Hydrology 2000.* Santa Fe, New Mexico, USA: IAHS Publication.

Giustolisi, O. and Laucelli, D. (2005) Improving generalization of artificial neural networks in rainfall-runoff modeling. *Hydrological Sciences Journal,* 50, 439-457.

Goldberg, D.E. (2003) *Genetic Algorithms in Search, Optimization, and Machine Learning,* USA, Addison Wesley Longman Inc.

Golob, R., Stokelj, T. and Grgic, D. (1998) Neural-network-based water inflow forecasting. *Control Engineering Practice,* 6, 593-600.

Haigh, M. (1997) Headwater control: despatches from the research front, Haigh, M. and Krecek, J. (Eds.) *NATO Advanced Research Workshop on Environmental Reconstruction in Headwater Areas.* Prague, Czech Republic: Springer.

Harrold, T.W. (1984) Ground clutter observed in the Dee weather radar project. *Meteorological Magazine,* 103, 140-141.

Haykin, S. (1999) Self-organizing maps. *Neural Networks - A Comprehensive Foundation.* 2nd ed. New Jersey, Prentice-Hall.

Hejazi, M.I. and Cai, X. (2009) Input variable selection for water resources systems using a modified Minimum Redundancy Maximum Relevance (mMRMR) algorithm. *Advances in Water Resources,* 32, 582-593.

Heo, G.-S. and Oh, I.-S. (2008) Simultaneous node pruning of input and hidden layers using genetic algorithms, *Proceedings of 2008 International Conference on Machine Learning and Cybernetics.* China: IEEE.

Hill, G. and Robertson, R. B. (1987) The establishment and operation of an unmanned weather radar. IN Collinge, V. K. and Kirby, C. (Eds.) *Weather Radar and Flood Forecasting.* John Wiley & Sons Ltd.

Hitch, T.J. (1991) Assessment of the adjustment of hourly radar rainfall fields by the use of daily raingauge totals. IN Collier, C.G. and Cluckie, I.D. (Eds.) *Hydrological Applications of Weather Radar.* Ellis Horwood limited.

Holland, J. (1975) *Adaptation in Natural and Artificial Systems*, Ann Arbor, University of Michigan Press.

Hornik, K. (1993) Some new results on neural network approximation, *Neural Networks*, 6, 1069-1072.

Hsu, K.L., Gao, X., Sorooshian, S. and Gupta, H.V. (1997) Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of Applied Meteorology,* 36, 1176-1190.

Hsu, K.L., Gupta, H.V., Gao, X. and Sorooshian, S. (2000) Rainfall estimation from satellite imagery. IN Govindaraju, R. S. and Rao, A. R. (Eds.) *Water Science and Technology Library: Artificial Neural Networks in Hydrology.* Dordrecht, The Netherlands, Kluwer Academic Publishers.

Hsu, K.-L., Gupta, H. V., Gao, X., Sorooshian, S. and Imam, B. (2002) Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis. *Water Resources Research,* 38, 1302-1319.

Hsu, K.-L., Gupta, H.V. and Sorooshian, S. (1995) Artificial neural network modeling of the rainfall-runoff process. *Water resources research,* 31, 2517-2530.

Huang, Y., Cai, J., Yin, H. and Cai, M. (2009) Correlation of precipitation to temperature variation in the Huanghe River (Yellow River) basin during 1957-2006. *Journal of Hydrology,* 372, 1-8.

Hung, N.Q., Babel, M.S., Weesakul, S. and Tripathi, N.K. (2009) An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *Hydrology and Earth System Sciences,* 13, 1413-1425.

Hydrology and Water Management Center for Upper Northern Region (2005) Report of flooding in Chiang Mai city: 14-16 August 2005 (published in Thai). Chiang Mai.

Hydrology and Water Management Centre for Upper Northern Region (2007a) *Ping Report* [Online], [Accessed 16 November 2007], Available at http://www.hydro-1.net/DATASHOW/dspic/ping-report.html

Hydrology and Water Management Centre for Upper Northern Region (2007b) *Upper Northern Region Flood Warning Brochures* [Online], [Accessed 16 November 2007], Available at http://www.hydro-1.net/

IPCC (2007) Climate Change 2007: *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (Eds.), Cambridge and New York, Cambridge University Press.

Jain, A. and Srinivasulu, S. (2004) Development of effective and efficient rainfall-runoff models using integration of deterministic, real-coded genetic algorithms and artificial neural network techniques. *Water Resources Research,* 40, 1-12.

Jain, B.A. and Nag, B.N. (1995) Artificial neural network model for pricing initial public offerings. *Decision Sciences*, 26(3), 283-302.

Jankad, P. (2007) *Royal Weather Radar* [Online], [Accessed 1 November 2008], Available at http://www.royalrainmaking.thaigov.net

Jia, Y., Hongli, Z., Cunwen, N., Yunzhong, J., Hong, G., Zhi, X., Xueli, Z. and Zhixin, Z. (2009) A WebGIS-based system for rainfall-runoff prediction and real-time water resources assessment for Beijing. *Computers & Geosciences,* 35, 1517-1528.

Jones, D.M.A. (1956) *Raindrop Size Distribution and Radar Reflectivity*. Urbana, Illinois, Illinois State Water Survey Meteorologic Laboratory.

Joss, J.K., Schran, J.C., Thams, J.C. and Waldvogel, A. (1970) On the quantitative determination of precipitation by radar. *Wissenschaftliche Mitfeilung,* 63, 33.

Kalinga, O.A. and Gan, T.Y. (2007) Small storm precipitation estimation using merged radar and gauging data. *International Journal of Remote Sensing,* 28, 1101-1112.

Kalra, A. and Ahmad, S. (2009) Using oceanic-atmospheric oscillations for long lead time streamflow. *Water Resources Research,* 45, 1-18.

Kamp, R.G. and Savenije, H.H.G. (2006) Optimising training data for ANNs with Genetic Algorithms. *Hydrology and Earth System Sciences,* 10, 603-608.

Karunanithi, N., Grenney, W.J., Member, ASCE, Whitley, D. and Bovee, K. (1994) Neural networks for river flow prediction. *Journal of Computing in Civil Engineering,* 8, 201-220.

Karunasinghe, D.S.K., Liong, S.Y. (2006) Chaotic time series prediction with a global model: artificial neural network. *Journal of Hydrology*, 323(1-4), 92-105.

Kasabov, N. (1996) *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*, Cambridge MA, The MIT Press.

Kavetski, D., Kuczera, G. and Franks, S.W. (2006) Calibration of conceptual hydrological models revisited: 1. overcoming numerical artefacts. *Journal of Hydrology,* 320, 173-186.

Kerh, T. and Lee, C.S. (2006) Neural networks forecasting of flood discharge at an unmeasured station using river upstream information. *Advances in Engineering Software,* 37, 533-543.

Khu, S.T. and Madsen, H. (2005) Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration. *Water Resources Research,* 41, 1-14.

Kim, S., Kim, J.-H. and Park, K.-B. (2009) Neural networks models for the flood forecasting and disaster prevention systems in the small catchment. *Disaster Advances,* 2, 51-63.

Kim, T.-W. and Ahn, H. (2009) Spatial rainfall model using a pattern classifier for estimating missing daily rainfall data. *Stochastic Environmental Research and Risk Assessment,* 23, 367-376.

Kisi, O. (2006) Generalized regression neural networks for evapotraspiration modeling. *Hydrological Sciences Journal,* 51, 1092-1105.

Kisi, O. (2007) Streamflow Forecasting Using Different Artificial Neural Network Algorithms, *Journal of Hydrologic Engineering*, 12, 5, 532-539.

Kneale, P.E., See, L. and Evans, A. (1999) Developing a Neural Network for Flood Forecasting in the Northumbria Area of the North East Region, Environment Agency: Final Report.

Knowles, J.M. (1987) Flood forecasting hydrology in North West water. IN Collinge, V. K. and Kirby, C. (Eds.) *Weather Radar and Flood Forecasting.* John Wiley & Sons.

Kohonen, T. (1984) *The Self-Organizing Map,* Heidelberg, Germany, Springer-Verlag.

Krishna, B., Rao, Y.R.S. and Vijaya, T. (2008) Modeling groundwater levels in an urban coastal aquifer using artificial neural networks. *Hydrological Processes*, 22(8), 1180-1188.

Kuchment, L.S. and Gelfan, A.N. (2009) A study of effectiveness of the ensemble long-term forecasts of spring floods issued with physically based models of the river runoff formation. *Russian Meteorology and Hydrology,* 34, 100-109.

Kumar, A.R.S., Sudheer, K.P., Jain, S.K. and Agarwal, P.K. (2005) Rainfall-runoff modeling using artificial neural networks: comparison of network types. *Hydrological Processes,* 19, 1277-1291.

Kunikazu, K., Masanao, O. and Takashi, K. (2009) A Bayesian local linear wavelet neural network. *Advances in Neuro-Information Processing,* 5507, 147-154.

Leahy, P., Kiely, G. and Corcoran, G. (2008) Structural optimisation and input selection of an artificial neural network for river level prediction. *Journal of Hydrology,* 355, 192-201.

Lee, J.L., Lee, W.C. and Macdonald, A.E. (2006) Estimating vertical velocity and radial flow from Doppler radar observations of tropical cyclones. *Quarterly Journal of The Royal Meteorological Society,* 132, 125-145.

Lee, K.T. and Chang, C.H. (2005) Incorporating subsurface-flow mechanism into geomorphology-based IUH modeling. *Journal of Hydrology,* 311, 91-105.

Lenard, M.J., Alam, P. and Madey, G.R. (1995) The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decision Sciences*, 26(2), 209-227.

Li, Z., Deng, P. and Dong, J. (2009) Application of artificial neural network in rainfall-runoff model, Taniguchi, M., Burnett, W.C., Fukushima, Y., Haigh, M. and Umezawa, Y. (Eds.) *International Conference on Hydrological Changes and Management from Headwaters to the Ocean.* Kyoto, Japan: CRC Press Taylor & Francis Group.

Lin, G.-F. and Chen, G.-R. (2008) A systematic approach to the input determination for neural network rainfall -runoff models. *Hydrological Processes,* 22, 2524-2530.

Liong, S.Y., Lim, W.H. and Paudyal, G. (2000) River stage forecasting in Bangladesh: neural network approach. *Journal of Computing in Civil Engineering*. 14(1), 1-8.

Liu, M., Kenji, Y., Tadaharu, I. and Kentaro, K. (2009) New approach for estimation of pollutant load by using artificial neural network, *16th Asia and Pacific Division Congress of the International Association of Hydraulic Engineering and Research.* China.

Lorrai, M. and Sechi, G.M. (1995) Neural net for modeling rainfall-runoff transformations. *Water Resources Management,* 9, 299-313.

Mackay, D.J.C. (1992) Bayesian interpolation. *Neural Computation,* 4, 415-447.

Maier, H.R. and Dandy, G.C. (1996) The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research,* 32, 1013-1022.

Maier, H.R. and Dandy, G.C. (1997) Determining inputs for neural network models of multivariate time series. *Microcomputers in Civil Engineering,* 12, 353-368.

Maier, H.R. and Dandy, G.C. (1998) The effect of internal parameters and geometry on the performance of back-propagation neural networks: An empirical study. *Environmental Modeling & Software,* 13, 193-209.

Maier, H.R. and Dandy, G.C. (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environmental Modeling & Software,* 15, 101-124.

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

Mapiam, P.P. and Sriwongsitanon, N. (2008) Climatological Z-R relationship for radar rainfall estimation in the Upper Ping basin. *ScienceAsia,* 34, 215-222.

Mapiam, P.P. and Sriwongsitanon, N. (2009) Estimation of the URBS model parameters for flood estimation of ungauged catchments in the upper Ping river basin, Thailand. *SciencesAsia,* 35, 49-56.

Maren, A., Harston, C. and Pap, R. (1990) *Handbook of Neural Computing Applications,* San Diego, California, Academic Press Inc.

Martinec, J. (1975) Snowmelt - runoff model for stream flow forecasts. *Nordic Hydrology,* 6, 145-154.

May, D.B. and Sivakumar, M. (2009) Prediction of urban stormwater quality using artificial neural networks. *Environmental Modeling & Software,* 24, 296-302.

May, R.J., Dandy, G.C., Maier, H.R. and Nixon, J.B. (2008a) Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. *Environmental Modeling & Software,* 23, 1289-1299.

May, R.J., Maier, H.R., Dandy, G.C. and Fernando, T.M.K.G. (2008b) Non-linear variable selection for artificial neural networks using partial mutual information. *Environmental Modeling & Software,* 23, 1312-1326.

Minns, A.W. and Hall, M.J. (1996) Artificial neural networks as rainfall-runoff models. *Hydrological Sciences Journal,* 41, 399-417.

Montano, J.J. and Palmer, A. (2003) Numeric sensitivity analysis applied to feedforward neural networks. *Neural Computing & Applications,* 12, 119-125.

Moore, R.J. (1987) Towards more effective use of radar data for flood forecasting. IN Collinge, V.K. and Kirby, C. (Eds.) *Weather Radar and Flood Forecasting.* John Wiley & Sons.

Moradkhani, H., Hsu, K.-L., Gupta, H.V. and Sorooshian, S. (2004) Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. *journal of Hydrology,* 295, 246-262.

Morris, E.M. (1980) Forecasting flood flows in grassy and forested basins using a deterministic distributed mathematical model. *Hydrological Forecasting,* 129, 247-255.

Mukerji, A., Chatterjee, C. and Raghuwanshi, N.S. (2009) Flood forecasting using ANN, Neuro-Fuzzy and Neuro-GA models. *Journal of Hydrologic Engineering,* 14, 647-652.

Murre, J.M. and Sturdy, D.P. (1995) The connectivity of the brain: multi-level quantitative analysis. *Biological Cybernetics*, 73(6), 529–545.

Nash, J.E. and Sutcliffe, J.V. (1970) River flow forecasting through conceptual models 1: a discussion of principles. *Journal of Hydrology,* 10, 282-290.

Natural Disasters Research Unit (2007a) *Water stage network in Upper Ping catchment* [Online], [Accessed 16 November 2007], Available at http://www.cendru.eng.cmu.ac.th/flooding/?name=/chapter1/cp1_2/artical2

Natural Disasters Research Unit (2007b) *Flood Forecasting System* [Online], [Accessed 16 November 2007], Available at http://www.cendru.eng.cmu.ac.th/engflood/

Natural Disasters Research Unit (2007c) *General Information of Upper-Ping Basin* [Online], [Accessed 16 November 2007], Available at http://www.cendru.eng.cmu.ac.th/flooding/?name=/chapter1/cp1_4/artical4

Nemec, J. (1986) *Hydrological Forecasting*, Dordrecht, Holland, D. Reidel Publishing Company.

Neural Ware Inc (1991) *Neural computing, neural Works professional II/plus and neural works explorer*.

NineNews (2006) *Flooding in Thailand Kills 32* [Online], [Accessed 19 July 2010], Available at http://news.ninemsn.com.au/article.aspx?id=74707

Ninprom, S. and Chumchean, S. (2009) Effectiveness of bias adjustment technique to reduce errors in MM5 rainfall forecasting on flood forecasting results of the Upper Ping river basin, *The 14th National Convention on Civil Engineering (published in Thai)* Suranaree University of Technology, Nakhon Ratchasima:

Northern Meteorological Center (2007) Climatological data for the period 30 years (1971-2000). Thai Meteorological Department, Chiang Mai.

Noruani, V. and Mano, A. (2007) Semi-distributed flood runoff model at the sub continental scale for southwestern Iran. *Hydrological Processes,* 21, 3173-3180.

O'connor, K. (2002) River flow forecasting, *EU Advanced Study Course on River Basin Modeling.* University of Birmingham.

Okumura, K., Satomura, T., Oki, T. and Khantiyanan, W. (2003) Diurnal variation of precipitation by moving mesoscale systems: radar observations in Northern Thailand. *Geophysical Research Letters,* 30.

Ozesmi Sland Ozesmi U (1999) An artificial neural network approach to spatial habitat modeling with interspecific interaction. *Ecological Modeling,* 116, 15-31.

Pai, P.F. (2006) System reliability forecasting by support vector machines with genetic algorithms. *Mathematical and Computer Modeling,* 43, 262-274.

Partal, T. (2009) River flow forecasting using different artificial neural network algorithms and wavelet transform. *Canadian Journal of Civil Engineering,* 36, 26-39.

Patsinghasanee, S. (2004) Performance comparison of artificial neural network and hydrodynamic models on flood forecasting in the Upper Ping River Basin. Master thesis, *Department of Water Resources Engineering.* Bangkok, Kasetsart University.

Patsinghasanee, S., Lipiwattanakarn, S. and Sriwongsitanon, N. (2004) An optimized back propagation neural network for flood forecasting in The Ping River, *The 9th National Convention on Civil Engineering (published in Thai)* Thailand.

Patuwo, E., Hu, M.Y. and Hung, M.S. (1993) Two group classification using neural network. *Decision Sciences*, 24(4), 825-845.

Phien, H.N. and Nda, K. (2003) Flood forecasting for the Upper Reach of the Red River basin, North Vietnam. *Water SA,* 29, 267-272.

Pramanik, N. and Panda, R.K. (2009) Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction. *Hydrological Sciences Journal*, 54(2), 247-260.

Quinlan, J.R. (1992) Learning with continuous classes, *5th Australian Joint Conference on Artificial Intelligence.* Singapore: World Scientific.

Rachaneewan, T. (2006a) Correcting radar rainfall mean field bias. IN Bureau of Royal Rainmaking and Agricultural Aviation (Eds.), Ministry of Agriculture and Cooperatives.

Rachaneewan, T. (2006b) Radar rainfall estimates. IN Bureau of Royal Rainmaking and Agricultural Aviation (Eds.) Bangkok, Ministry of Agriculture and Cooperatives.

Rai, R.K. and Mathur, B.S. (2008) Event-based sediment yield modeling using artificial neural network. *Water Resources Management*, 22(4), 423-441.

Rajurkar, M., Kothyari, U. and Chaube, U. (2004) Modeling of the daily rainfall-runoff relationship with artificial neural network. *Journal of Hydrology,* 285, 96-113.

Raman, H. and Sunilkumar, N. (1995) Multivariate modeling of water resources time series using artificial neural networks. *Hydrological Sciences Journal,* 40, 145-163.

Ravn, O. (2003) *The NNSYSID Toolbox-for Use with MATLAB* [Online], [Accessed 29 November 2007], Available at *http://www.iau.dtu.dk/research/control/nnsysid.html*

Regional Centre of Geo-Informatics and Space Technology (2006), Faculty of Social Sciences, Chiang Mai University, Thailand.

Remesan, R., Shamim, M.A., Han, D. and Mathew, J. (2009) Runoff prediction using an integrated hybrid modeling scheme. *Journal of Hydrology,* 372, 48-60.

Rodratana, P. and Piamsa-Nga, N. (2008) 'The study of flood alleviation of Chiang Mai city area', *The 13th National Convention on Civil Engineering (published in Thai)* Phataya, Thailand.

Rosenfeld, D., Wolff, D.B. and Atlas, D. (1993) General probability-matched relations between radar reflectivity and rain rate. *Journal of Applied Meteorology,* 32, 50-72.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning internal representations by error propagations. In: Rumelhart, D.E. and McClelland, J.L. (Eds.) *Parallel Distributed Processing: Explorations in the Microstructures of Cognition Vol. 1* (MIT Press, Cambridge, Massachusetts, USA), 318-362.

Sahoo, G.B. and Ray, C. (2006) Flow forecasting for a Hawaii stream using rating curves and neural networks. *Journal of Hydrology,* 317, 63-80.

Sahoo, G.B., Ray, C. and Carlo, E.H.D. (2006) Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii. *Journal of Hydrology,* 327, 525-538.

Sardinas, J.G. and Pedreira, M.D.J. (2003) Calibration of rainfall-runoff models using global optimization. *Ingenieria Hidraulica en Mexico,* 18, 55-73.

Schreider, S.Y., Jakeman, A.J., Gallant, J. and Merritt, W.S. (2002) Prediction of monthly discharge in ungauged catchments under agricultural land use in the Upper Ping basin, northern Thailand. *Mathematics and Computers in Simulation,* 59, 19-33.

Schultz, G.A. (1987) Flood forecasting based on rainfall radar measurement and stochastic rainfall forecasting in the Federal Republic of Germany. IN Collinge, V. K. and Kirby, C. (Eds.) *Weather Radar and Flood Forecasting.* John Wiley & Sons.

See, L. and Openshaw, S. (1999) Applying soft computing approaches to river level forecasting. *Hydrological Sciences Journal,* 44, 763-778.

See, L.M. (2008) Data fusion methods for integrating data-driven hydrological models, *Studies in Computational Intelligence (SCI)*, 79, 1-18.

Seidou, O. and Ouarda, T.B.M.J. (2007) Recursion-based multiple change point detection in multiple linear regression and application to river streamflows. *Water Resources Research,* 43.

Shamseldin, A.Y. (1997) Application of a neural network technique to rainfall-runoff modeling. *Journal of Hydrology,* 199, 272-294.

Shamseldin, A.Y. and O'connor, K.M. (2001) A non-linear neural network technique for updating of river flow forecasts. *Hydrology and Earth System Sciences*, 5, 577-597.

Shamseldin, A.Y., O'connor, K.M. and Nasr, A.E. (2007) A comparative study of three neural network forecast combination methods for simulated river flows of different rainfall-runoff models. *Hydrological Sciences Journal,* 52, 896-916.

Sharma, A. (2000) Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 A strategy for system predictor identification. *Journal of Hydrology,* 239, 232-239.

Sharma, A., Luk, K.C., Cordery, I. and Lall, U. (2000) Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: part 2-predictor identification of quarterly rainfall using ocean-atmosphere information. *Journal of Hydrology,* 239, 240-248.

Sharma, D., Gupta, A.D. and Babel, M.S. (2007) Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand. *Hydrology and Earth System Sciences Discussions,* 4, 35-74.

Shevnina, E.V. (2009) Methods of long-range forecasting of dates of the spring flood beginning and peak flow in the Estuary sections of the Ob and Yenisei Rivers. *Russian Meteorology and Hydrology,* 34, 51-57.

Shrestha, R.R., Theobald, S. and Nestmann, F. (2005) Simulation of flood flow in a river system using artificial neural networks. *Hydrology and Earth System Sciences,* 9, 313-321.

Smith, J. and Eli, R.N. (1995) Neural-network models of rainfall-runoff process. *Journal of Water Resources Planning and Management,* 121, 499-508.

Smith, J.A., Baeck, M.L., Meierdiercks, K.L., Miller, A.J. and Krajewski, W.F. (2007) Radar rainfall estimation for flash flood forecasting in small urban watersheds. *Advances in Water Resources,* 30, 2087-2097.

Solomatine, D.P. and Xue, Y. (2004) M5 model trees and neural networks: application to flood forecasting in the Upper Reach of the Huai River in China. *Journal of Hydrologic Engineering,* 9, 491-501.

Solomatine, D., See, L. and Abrahart, R.J. (2008) Data-driven modeling: concepts, approaches and experiences. In: Abrahart, R.J., See, L. and Solomatine, D.P. (Eds) *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications*. Heidelberg, Springer-Verlag, 17-30.

Sophhonphattanakul, S., Wangwongwiroj, N. and Israngkura, U. (2009) Effect on change in land uses on stream flow in the Upper Ping River basin, *The 14th National Convention on Civil Engineering (published in Thai)* Nakhonratchasima, Thailand.

Stephenson, D. (1979) Direct optimization of Muskingum routing coefficients. *Journal of Hydrology*, 41(1-2), 161-165.

Sudheer, K.P. (2005) Knowledge extraction from trained neural network river flow models. *Journal of Hydrologic Engineering,* 10, 264-269.

Sukka, P. (2005) Forecasting of daily inflow to large reservoir in Upper Ping River basin using artificial neural network. Master thesis, *Civil Engineering.* Chiang Mai, Chiang Mai University.

Supharatid, S. (2003) Application of neural network model in establishing a stage-discharge relationship for a tidal river. *Hydrological Processes*, 17(15), 3085-3099.

Sureerattanan, S. and Phien, H.N. (1997) Back-propagation networks for daily streamflow forecasting. *Water Resources Journal December*, 1-7.

Taesombat, W. and Sriwongsitanon, N. (2006) An evaluation of the effectiveness of hydrodynamic models application for flood routing investigation in the Upper Ping River Basin. *Engineering Journal Kasetsart,* 20, 74-82.

Taesombat, W. and Sriwongsitanon, N. (2010) Flood investigation in the Upper Ping River Basin using mathematical models. *Kasetsart Journal Natural Science,* 44, 152-166.

Tang, Z., Dealmeida, C. and Fishwick, P.A. (1991) Time series forecasting using neural networks VS Box-Jenkins methodology. *Simulation,* 57, 303-310.

Thaisawasdi, O., Sriwongsitanon, N. and Lipiwattanakarn, S. (2007) Daily flow estimation for small ungauged basins using artificial neural network models, *The 12th National Convention on Civil Engineering (published in Thai)* Phitsanulok, Thailand.

The World Bank (2006) Final report: developing watershed management organizations in pilot sub-basins.

Thompson Reuters ISI database (2010) Web of Knowledge [Online], [Accessed 19 July 2010], Available at http://www.wok.mimas.co.uk

Tingsanchali, T. and Gautam, M. R. (2000) Application of tank, NAM, ARMA and neural network, models to flood forecasting. *Journal of Hydrological Processes,* 14, 2473-2487.

Troutman, B.M. (1983) Runoff prediction errors and bias in parameter estimation induced by spatial variability of precipitation. *Water Resources Research,* 19, 791-810.

Ulaby, F.T., Moore, R.K. and Fung, A.K. (1981) *Microwave Remote Sensing,* Reading, Massachusetts, Addison-Wesley.

UPI (2008) *At Least 14 Killed by Flooding in Thailand* [Online], [Accessed 19 July 2010], Available at http://www.upi.com/Top_News/2008/09/20/At-least-14-killed-by-flooding-in-Thailand/UPI-10991221955481/

van Hinsbergen, C., van Lint, H. and van Zuylen H.J. (2008) Neural network committee to predict travel times: comparison of Bayesian evidence approach to the use of a validation set, *11th IEEE International Conference on Intelligent Transportation Systems.* China: IEEE.

Vargas, X., Gonzalez, R., Valck, P. and Nunez, C. (2003) A flood warning system using meteorological predictions coupled to a neural network and transfer function hydrological model, *2nd International Conference on Water Resources Management.* SPAIN.

Varoonchotikul, P. (2003) *Flood Forecasting Using Artificial Neural Networks,* Lisse, The Netherlands, A.A Balkema.

Vongtanaboon, S., Lim, H.S. and Richards, K. (2008) Data-based mechanistic rainfall-runoff modeling for a large monsoon dominated catchment in Thailand. *Silpakorn University Sciences and Technology Journal,* 2, 14-28.

Walczak, S. and Cerpa, N. (1999) Heuristic principles for the design of artificial neural networks. *Information and Software Technology*, 41, 107-117.

Wang, Q.J. (1991) The Genetic algorithms and its application to calibrating conceptual rainfall-runoff models. *Water Resources Research,* 27, 2467-2471.

Wang, W.-C., Chau, K.-W., Cheng, C.-T. and Qiu, L. (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology,* 374, 294-306.

Wardah, T., Bakar, S.H.A., Bardossy, A. and Maznorizan, M. (2008) Use of geostationary meteorological satellite images in convective rain estimation for flash-flood forecasting. *Journal of Hydrology,* 356, 283-298.

Wei, W.W.S. (2006) *Time Series Analysis: Univariate and Multivariate Methods,* USA, Pearson Education, Inc.

Wheater, H.S., Jakeman, A.J. and Beven, K.J. (1993) Progress and directions in rainfall-runoff modeling. IN Jakeman, A.J., Beck, M.B. and Mcaleer, M.J. (Eds.) *Modeling Change in Environmental Systems.* Chichester, Wiley, 101-132.

Whitley, D., Starkweather, T. and Bogart, C. (1990) Genetic algorithms and neural networks: Optimizing connections and connectivity. *Parallel Computing,* 14, 347-361.

Widrow, B. (1989) Current and past applications of neural networks to adaptive control problems. West Virginia University, College of Engineering (unpublished work).

Wilby, R.L. (1997) Contemporary Hydrology: Towards Holistic Environmental Science. IN Watts, G. (Eds.) *Hydrological Modeling in Practice.* Chichester, John Wiley & Sons.

Wilson, C.B., Valdes, J.B. and Rodriguez-Iturbe, I. (1979) Storm runoff from small lowland catchments in South-West England. *Journal of Hydrology,* 55, 321-337.

Witten, I.H. and Frank, E. (2005) *Data Mining Practical Machine Learning Tools and Techniques* USA, Morgan raufmann publishers.

Wongprayoun, T. and Sangsrijang, V. (2005) Upper Ping sub catchment geology. Bangkok, Department of Mineral Resources, Thailand.

Wood, E.F. and Connell, P.E.O. (1985) Real-time forecasting. IN Anderson, M.G. and Burt, T.P. (Eds.) *Hydrological Forecasting.* Suffolk, John Wiley & Sons.

Wu, C.L. and Chau, K.W. (2006) A flood forecasting neural network model with genetic algorithm. *International Journal of Environment and Pollution,* 28, 261-273.

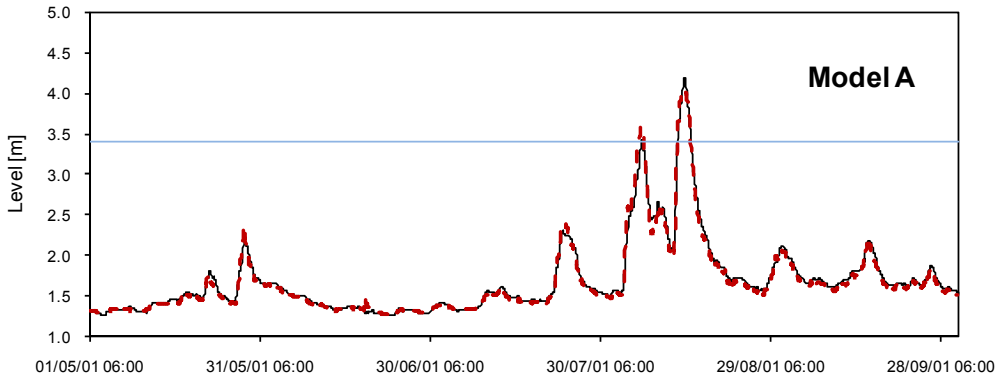
Yang, C.-C. and Chen, C.-S. (2009) Application of integrated back-propagation network and self organizing map for flood forecasting. *Hydrological Processes,* 23, 1313-1323.

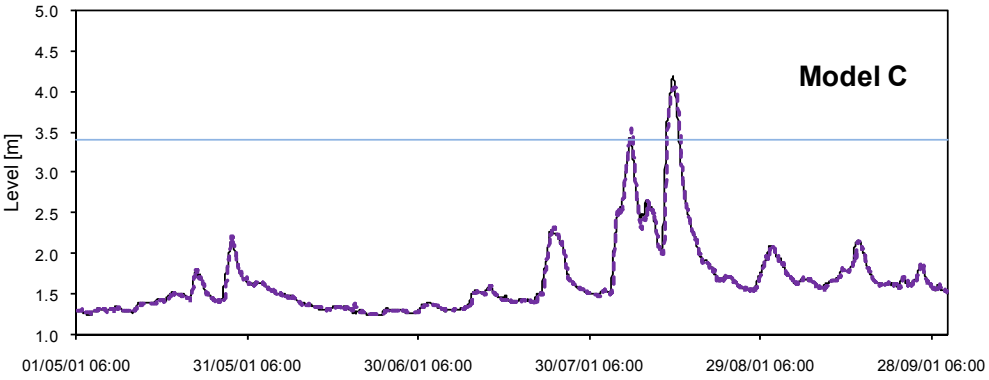
Yonaba, H., Anctil, F. and Fortin, V. (2010) Comparing sigmoid transfer functions for neural network multistep ahead streamflow forecasting. *Journal of Hydrologic Engineering,* 15, 275-283.

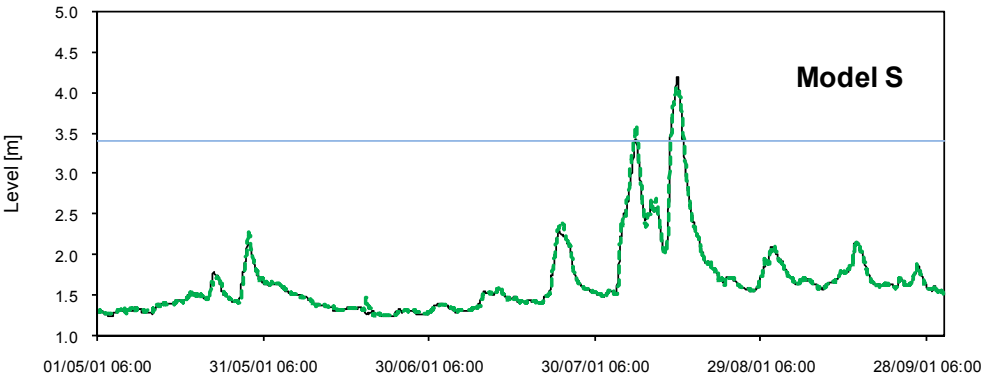
Youhang, Z., Wenzhuang, T. and Jianxun, Z. (2008) 'Arithmetic for multi-joint redundant robot inverse kinematics based on the Bayesian BP neural network', *International Conference on Intelligent Computation Technology and Automation* China: IEEE.

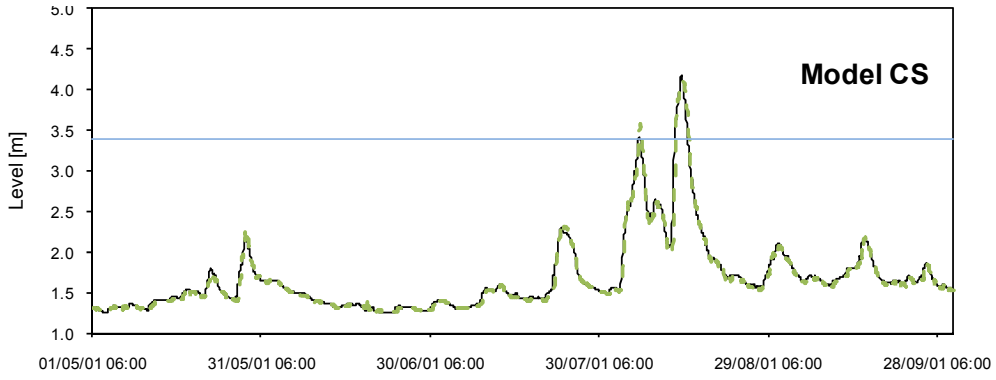
Zhang, B. and Govindaraju, R. S. (2000) Prediction of watershed runoff using Bayesian concepts and modular neural networks. *Water Resources Research,* 36, 753-762.

Appendix A: Hydrographs of Case Study 1









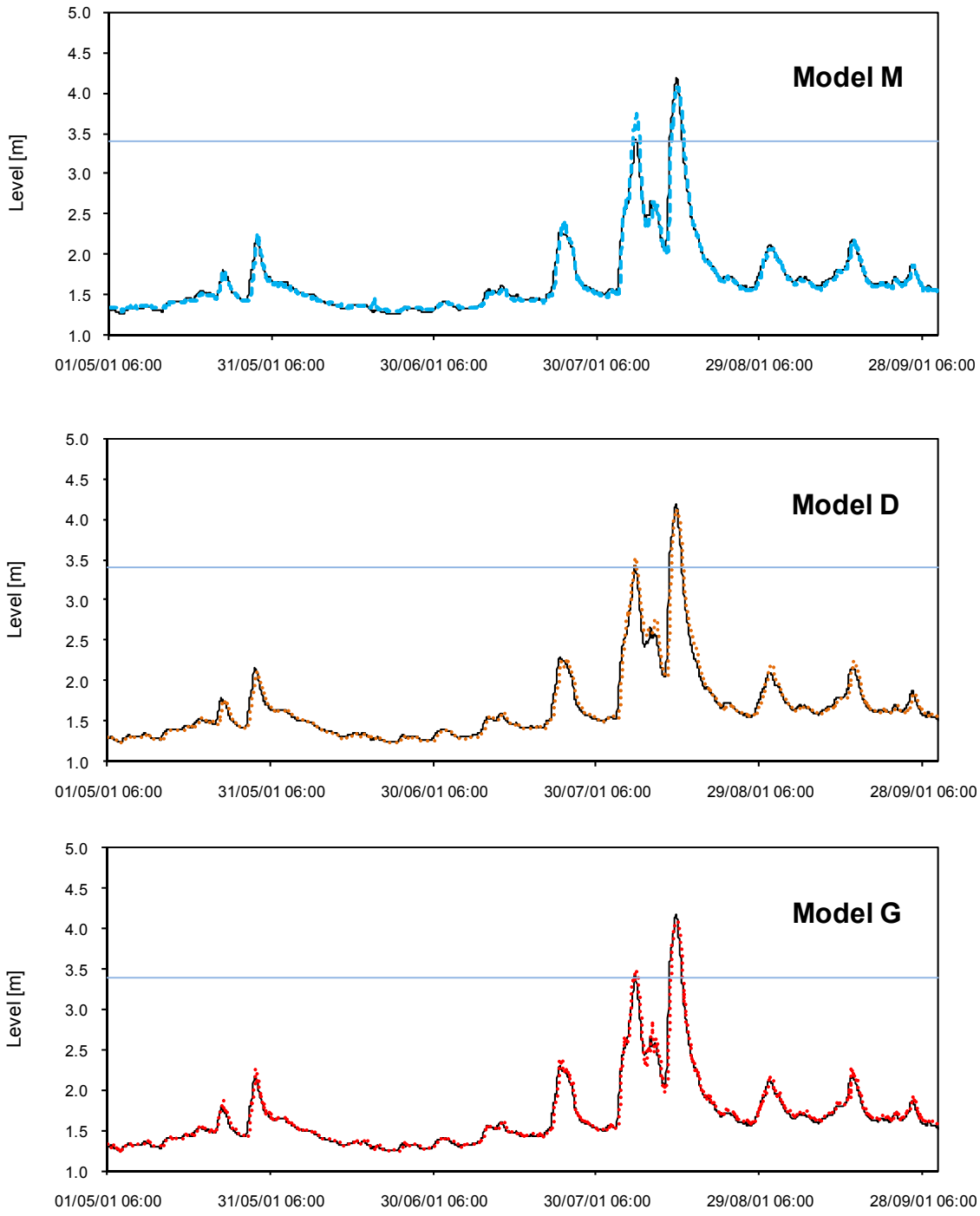
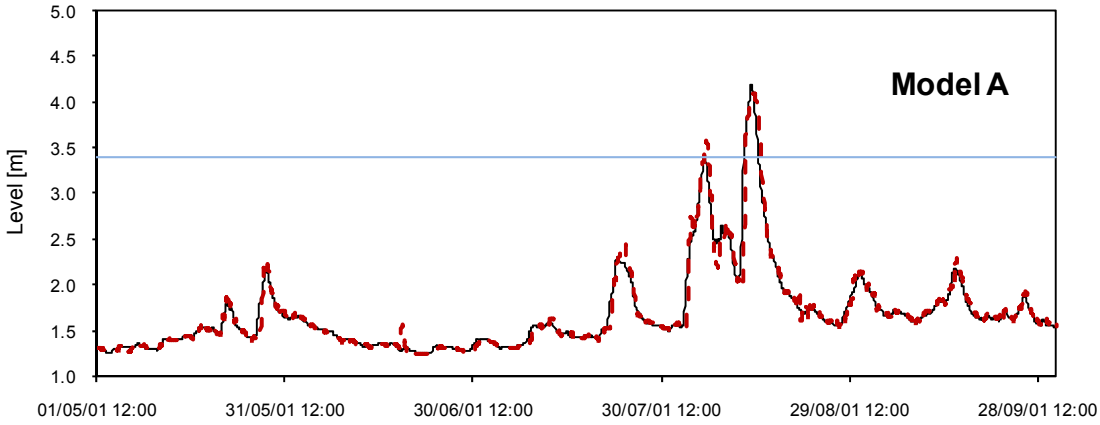
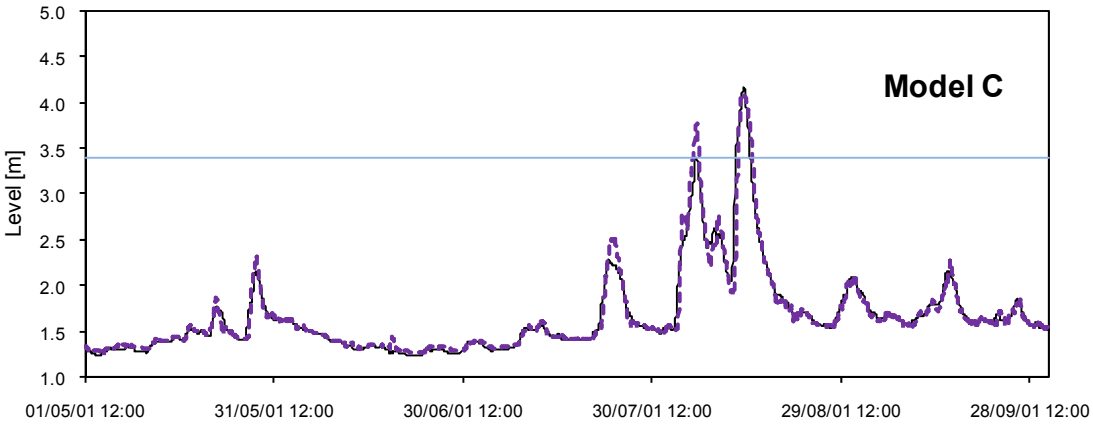
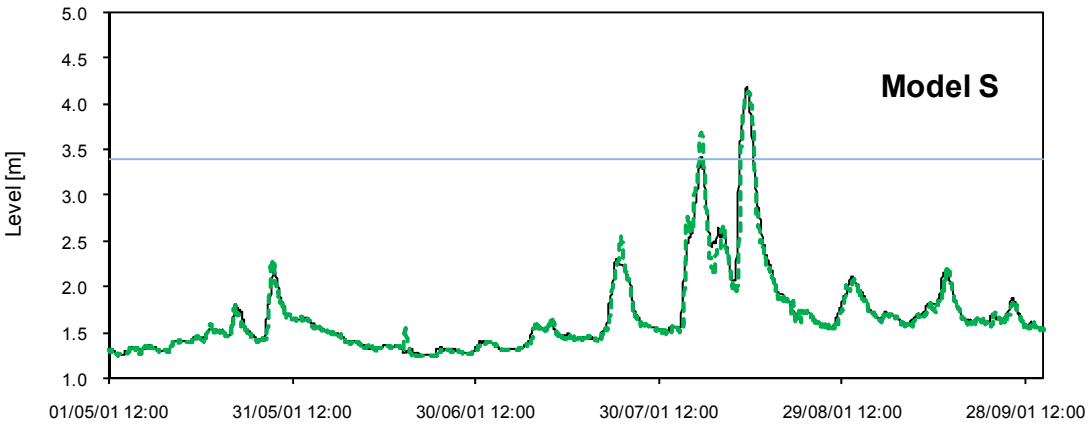
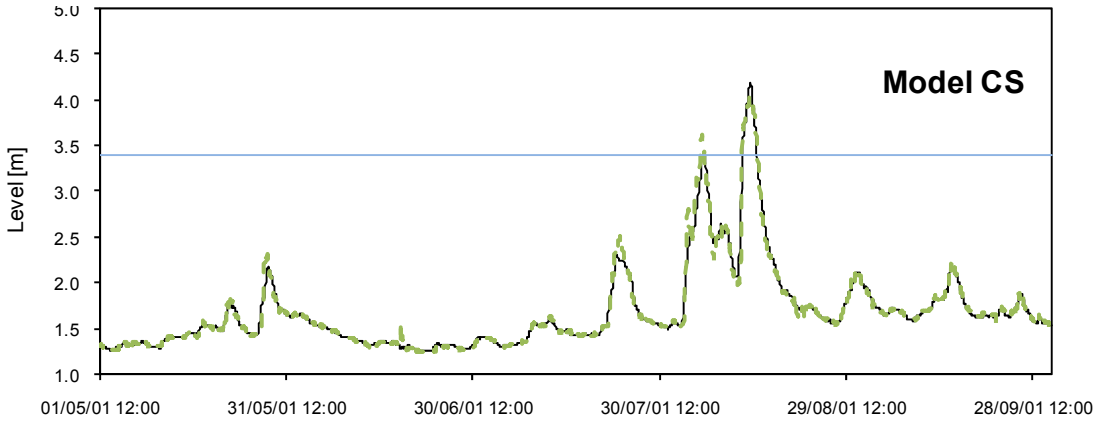


Figure A.1: Observed and predicted water level at P1 station, 1 May – 30 Sep 2001 (testing data) for lead time of 6 hours.


Model A
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01/05/01 12:00
31/05/01 12:00
30/06/01 12:00
30/07/01 12:00
29/08/01 12:00
28/09/01 12:00


Model C
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01/05/01 12:00
31/05/01 12:00
30/06/01 12:00
30/07/01 12:00
29/08/01 12:00
28/09/01 12:00


Model S
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01/05/01 12:00
31/05/01 12:00
30/06/01 12:00
30/07/01 12:00
29/08/01 12:00
28/09/01 12:00


Model CS
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01/05/01 12:00
31/05/01 12:00
30/06/01 12:00
30/07/01 12:00
29/08/01 12:00
28/09/01 12:00

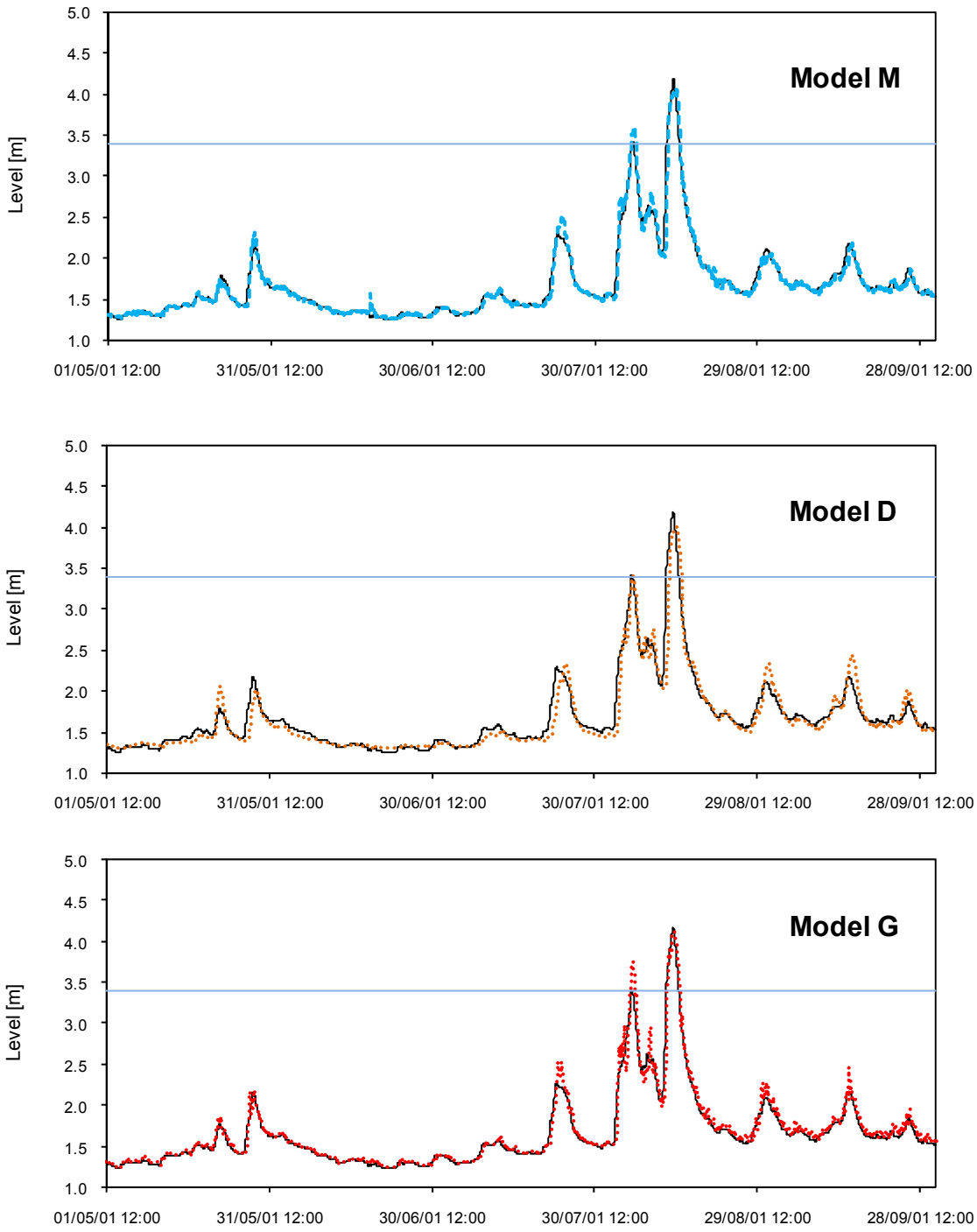
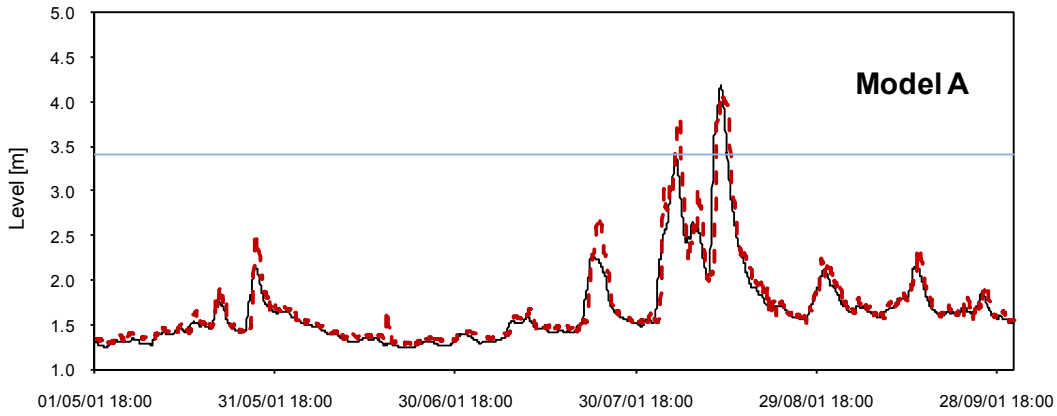
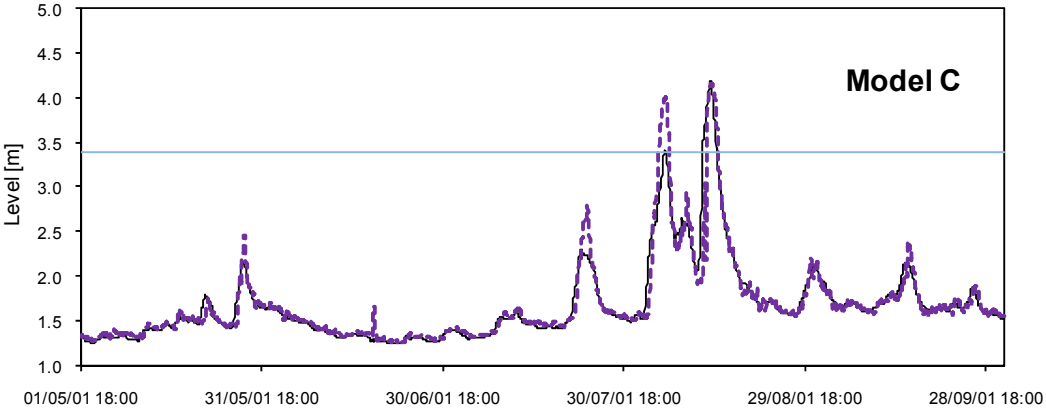
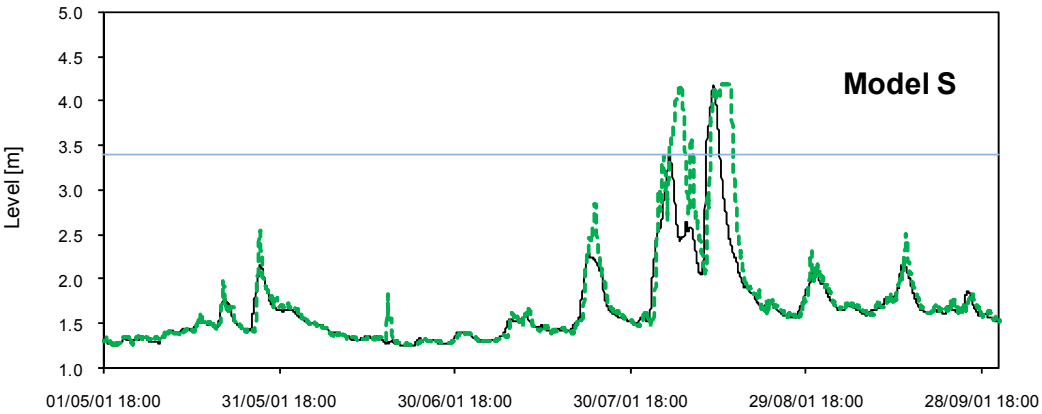
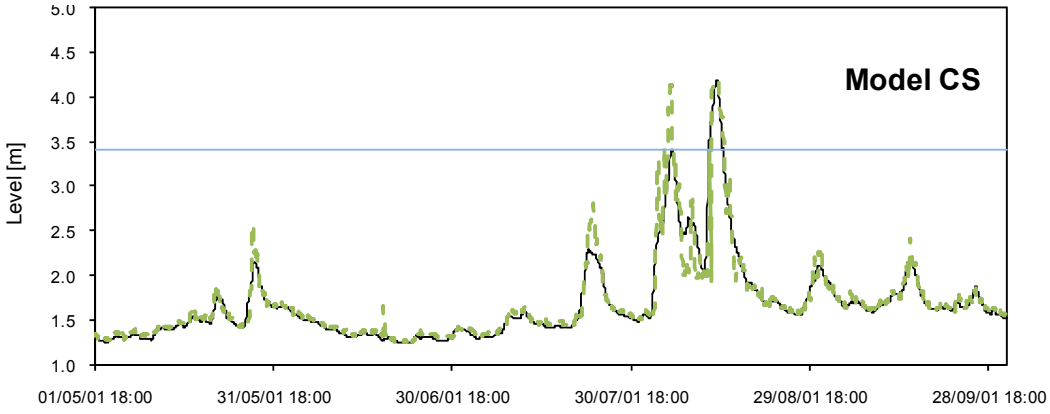


Figure A.2: Observed and predicted water level at P1 station, 1 May – 30 Sep 2001 (testing data) for lead time of 12 hours.


Model A
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01/05/01 18:00
31/05/01 18:00
30/06/01 18:00
30/07/01 18:00
29/08/01 18:00
28/09/01 18:00


Model C
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01/05/01 18:00
31/05/01 18:00
30/06/01 18:00
30/07/01 18:00
29/08/01 18:00
28/09/01 18:00


Model S
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01/05/01 18:00
31/05/01 18:00
30/06/01 18:00
30/07/01 18:00
29/08/01 18:00
28/09/01 18:00


Model CS
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
01/05/01 18:00
31/05/01 18:00
30/06/01 18:00
30/07/01 18:00
29/08/01 18:00
28/09/01 18:00

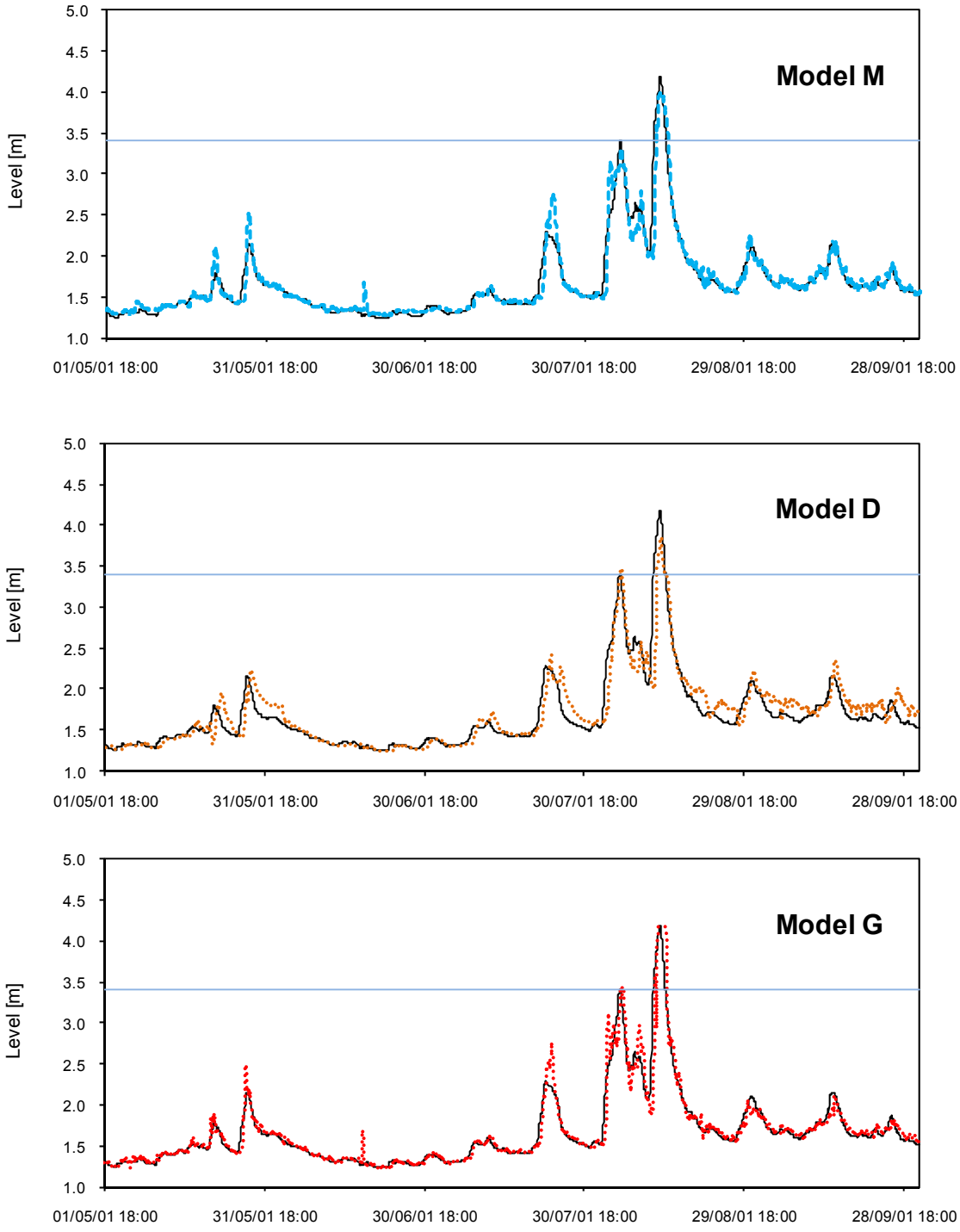
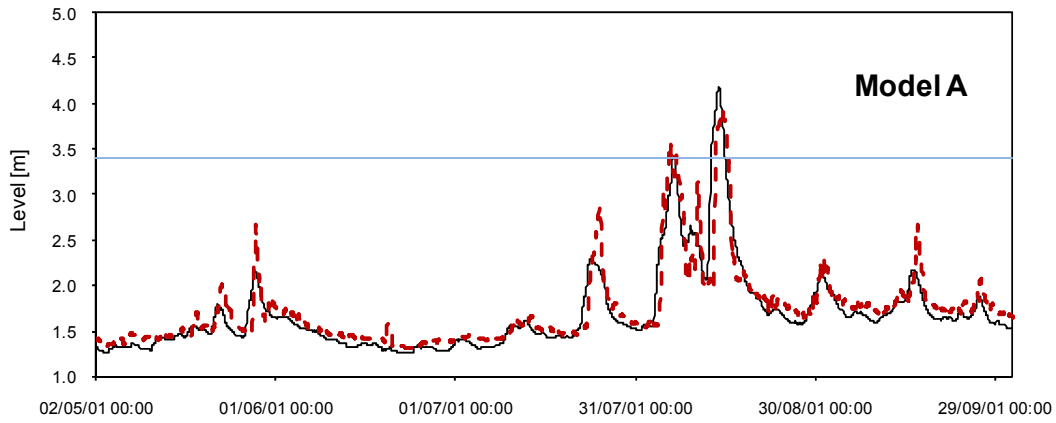
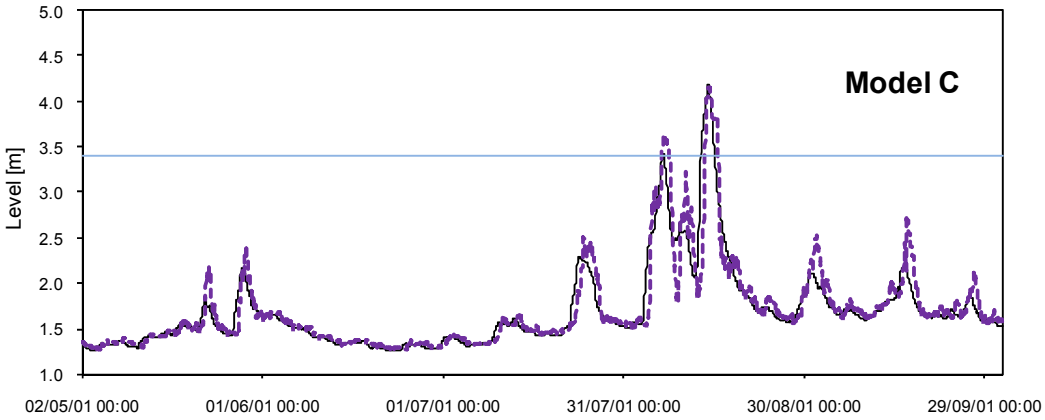
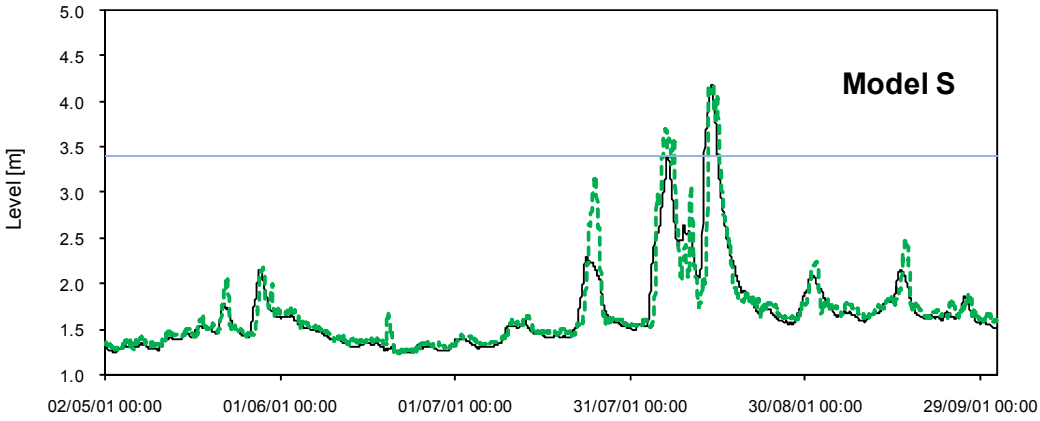
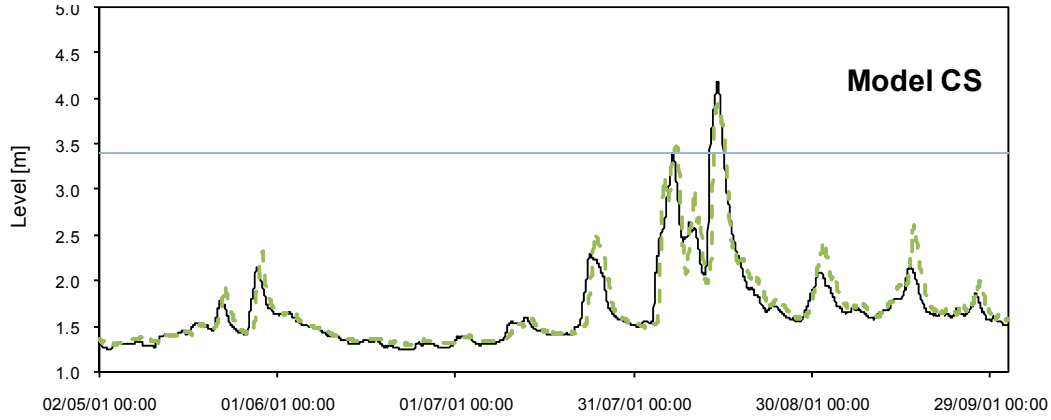


Figure A.3: Observed and predicted water level at P1 station, 1 May – 30 Sep 2001 (testing data) for lead time of 18 hours.


Model A
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
02/05/01 00:00
01/06/01 00:00
01/07/01 00:00
31/07/01 00:00
30/08/01 00:00
29/09/01 00:00


Model C
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
02/05/01 00:00
01/06/01 00:00
01/07/01 00:00
31/07/01 00:00
30/08/01 00:00
29/09/01 00:00


Model S
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
02/05/01 00:00
01/06/01 00:00
01/07/01 00:00
31/07/01 00:00
30/08/01 00:00
29/09/01 00:00


Model CS
Level [m]
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
02/05/01 00:00
01/06/01 00:00
01/07/01 00:00
31/07/01 00:00
30/08/01 00:00
29/09/01 00:00

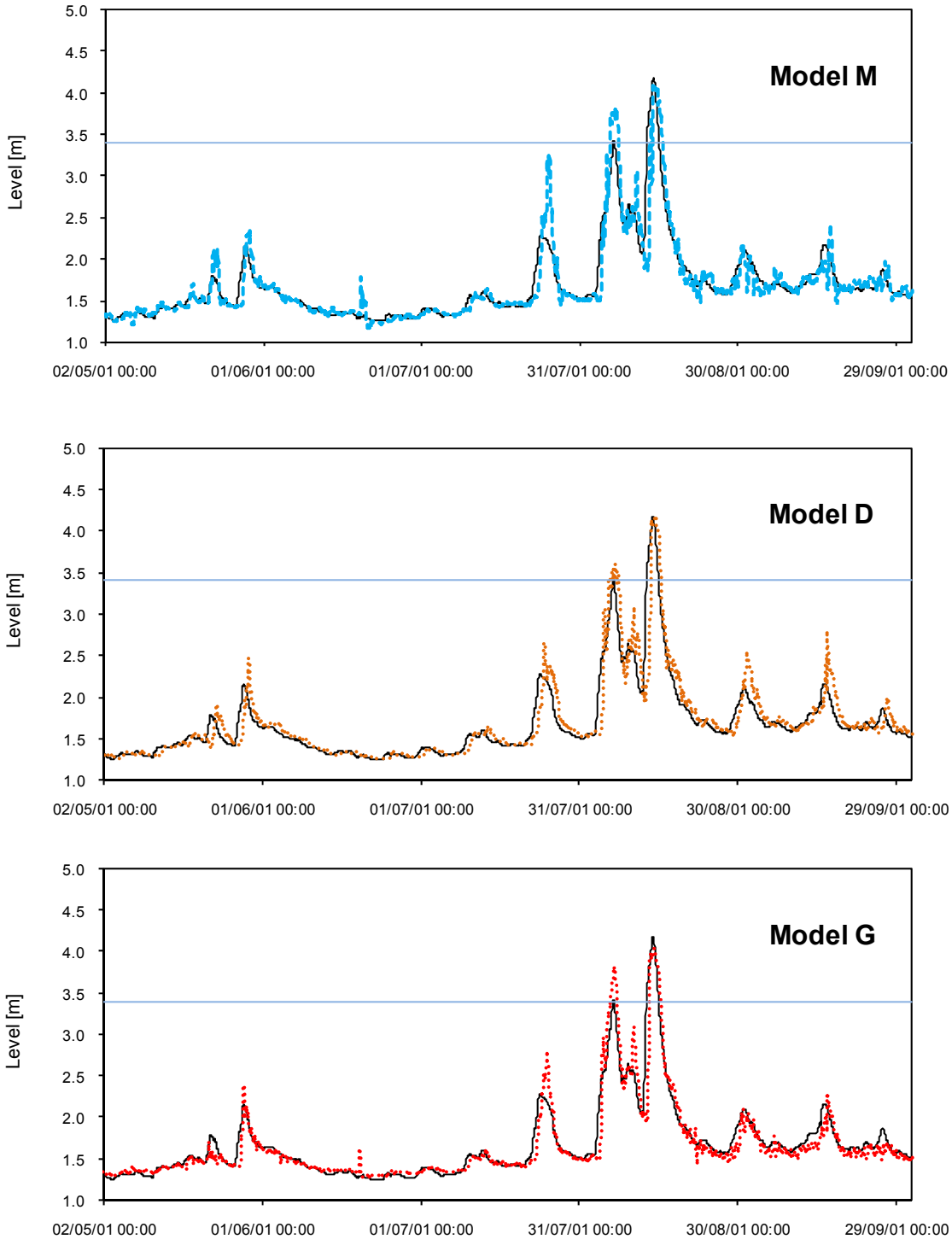


Figure A.4: Observed and predicted water level at P1 station, 1 May – 30 Sep 2001 (testing data) for lead time of 24 hours.

Note

A: All input, C: Correlation > 0.9, S: Stepwise regression, CS: Correlation > 0.9 + Stepwise regression, M: M5 model tree, D: Data mining, G: Genetic algorithms

Appendix B: Hydrographs of Case Study 2

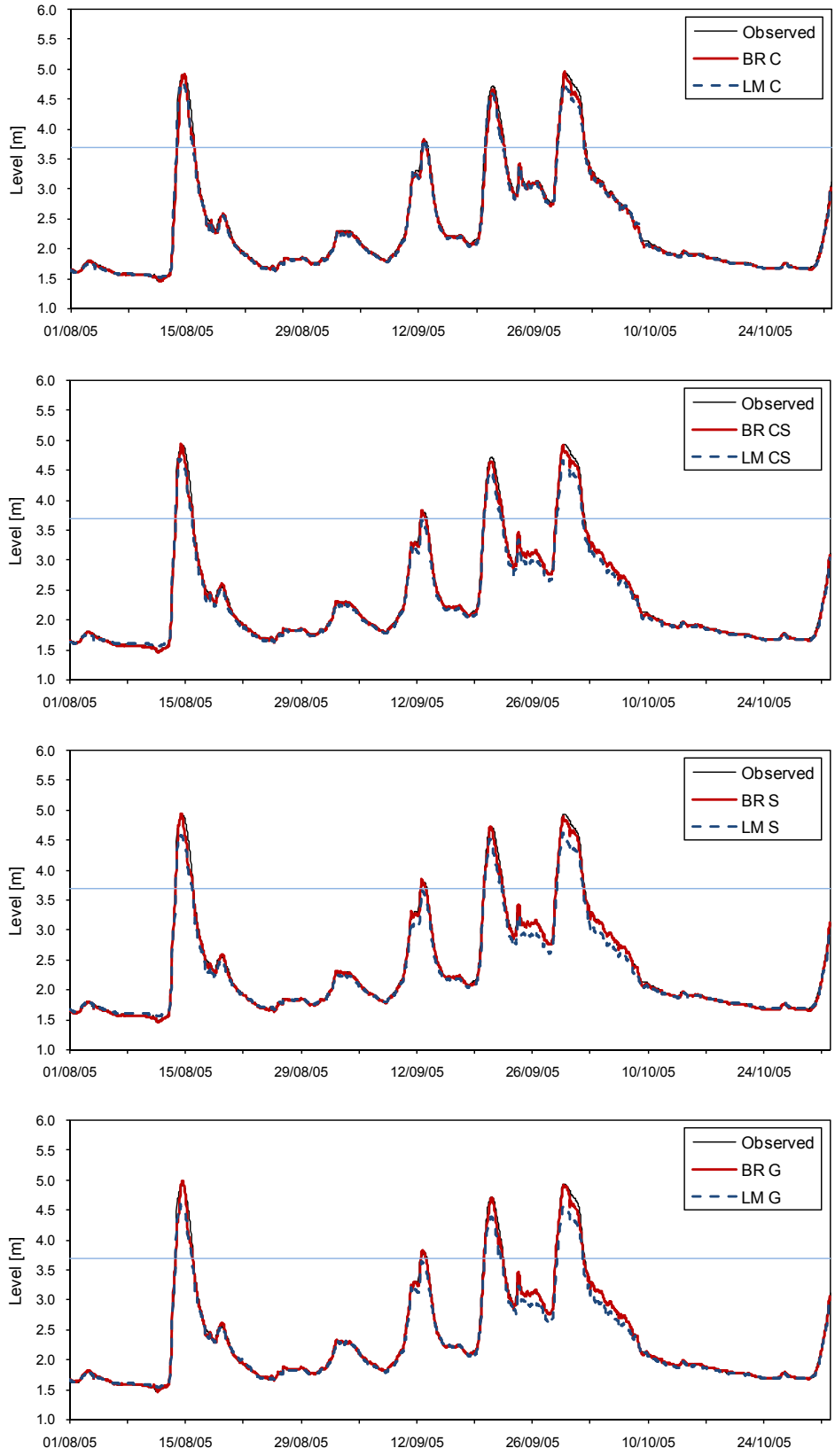


Figure B.1: Dataset one, compare of the BR and LM model at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 6 hours.

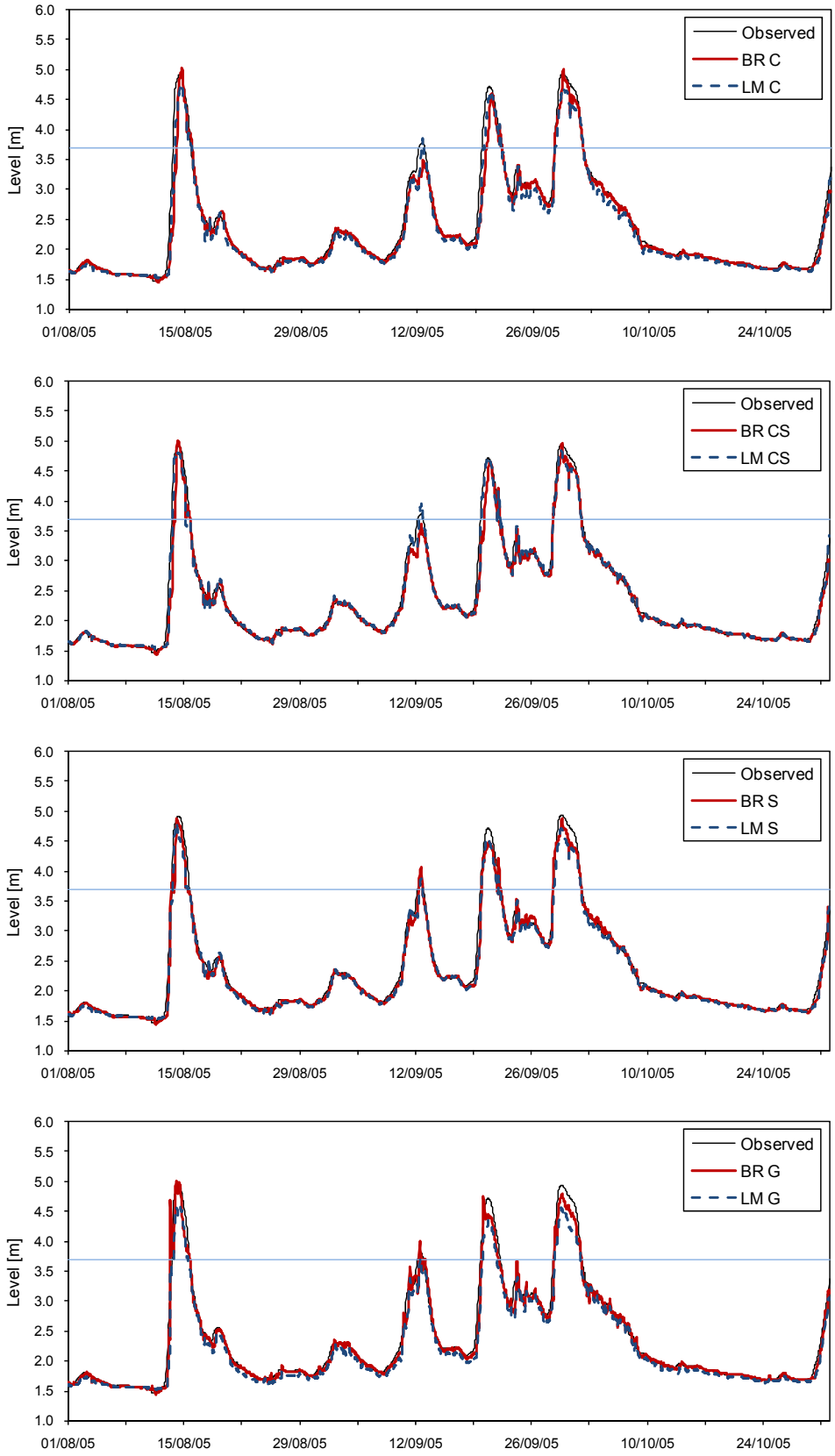


Figure B.2: Dataset one, compare of the BR and LM model at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 12 hours.

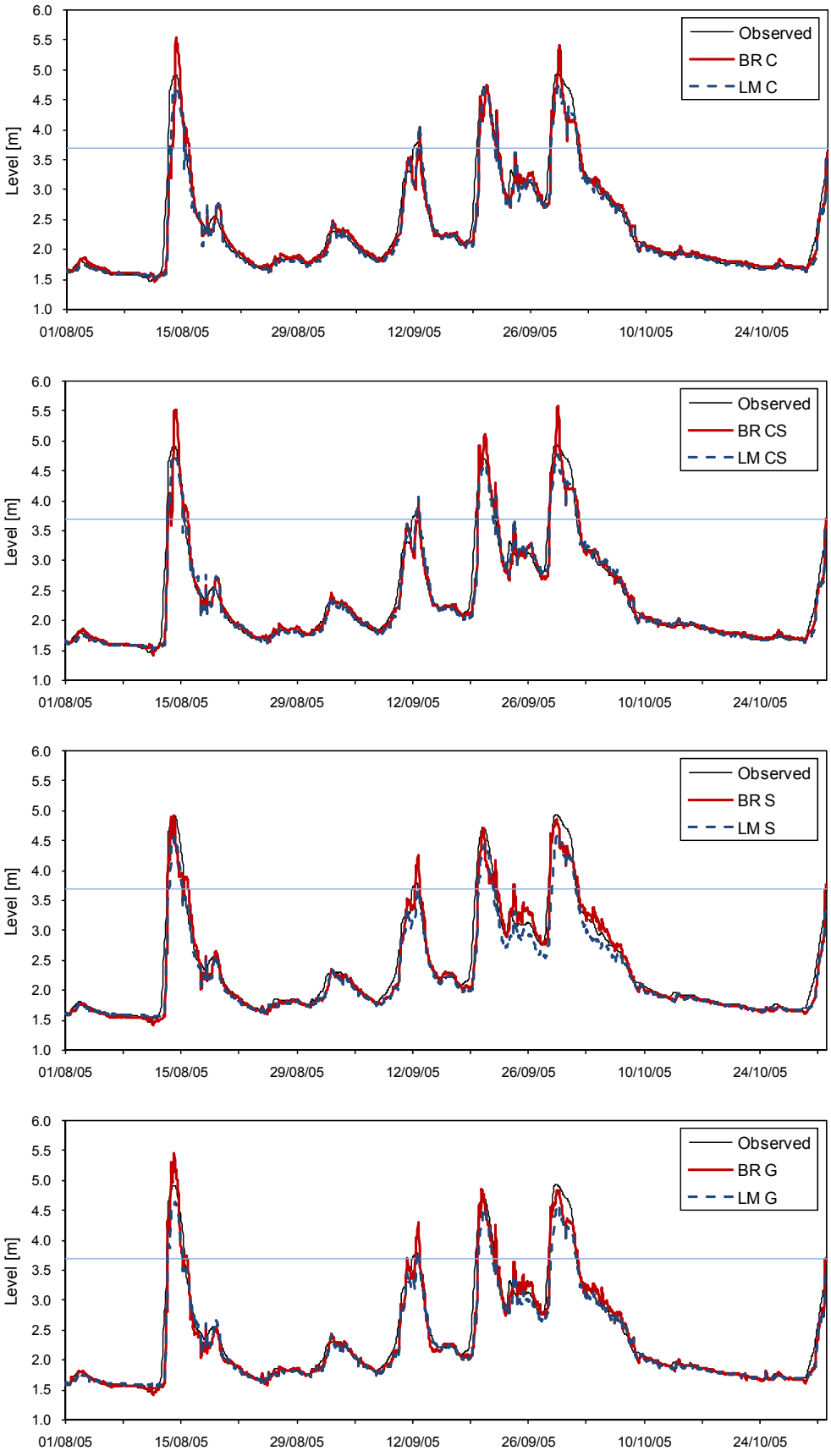


Figure B.3: Dataset one, compare of the BR and LM model at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 18 hours.

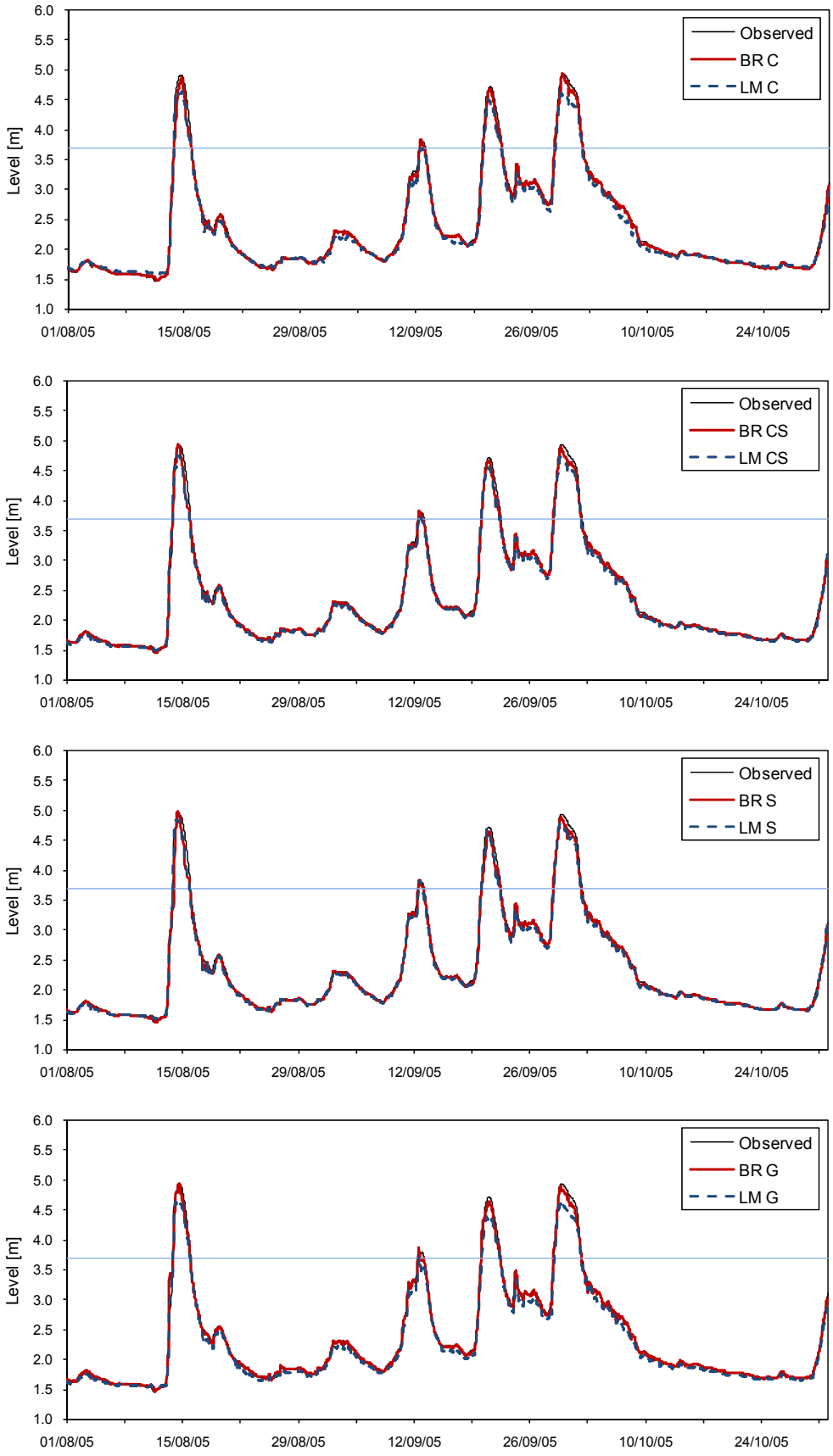


Figure B.4: Dataset two, compare of the BR and LM model at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 6 hours.

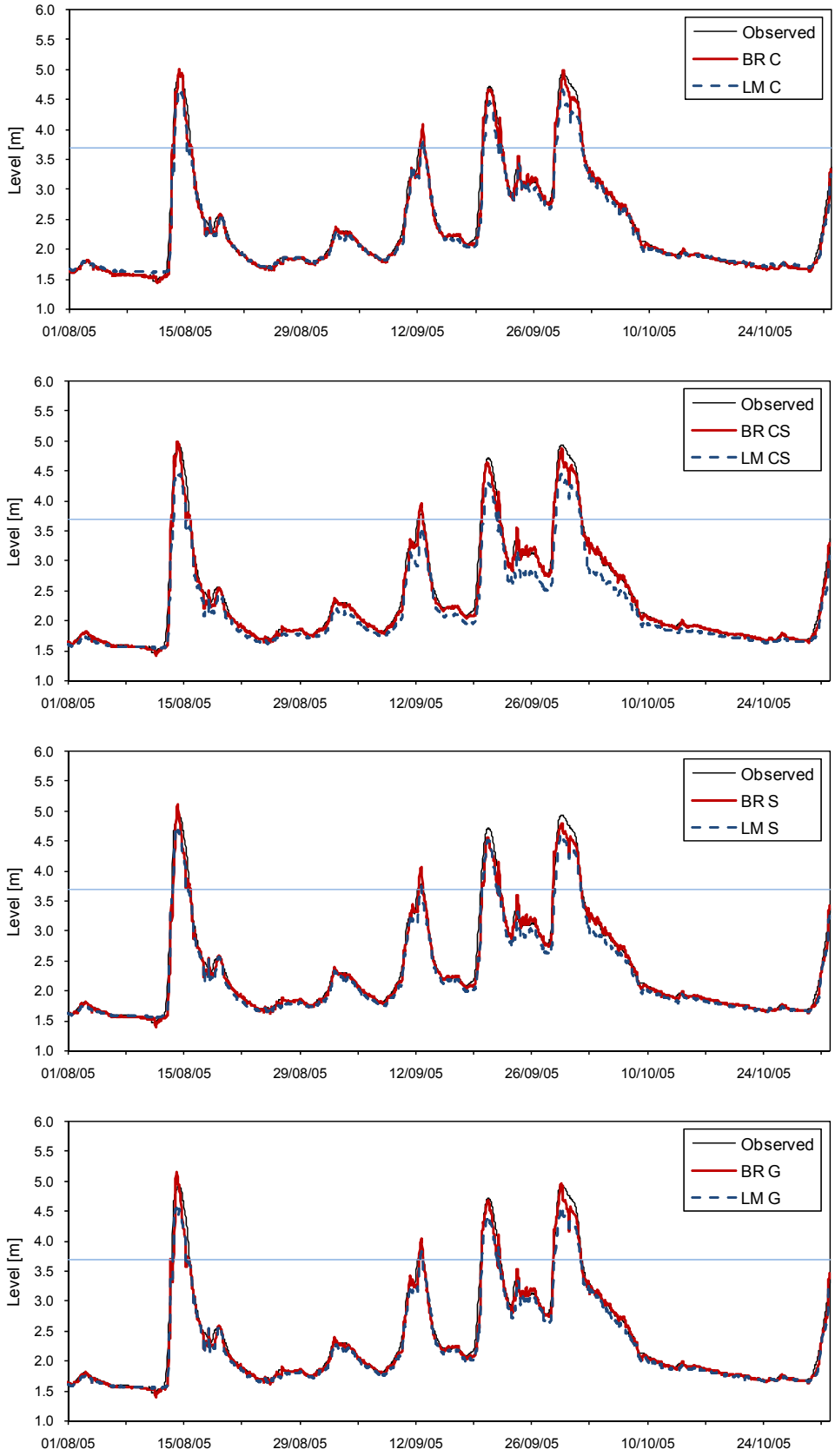


Figure B.5: Dataset two, compare of the BR and LM model at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 12 hours.

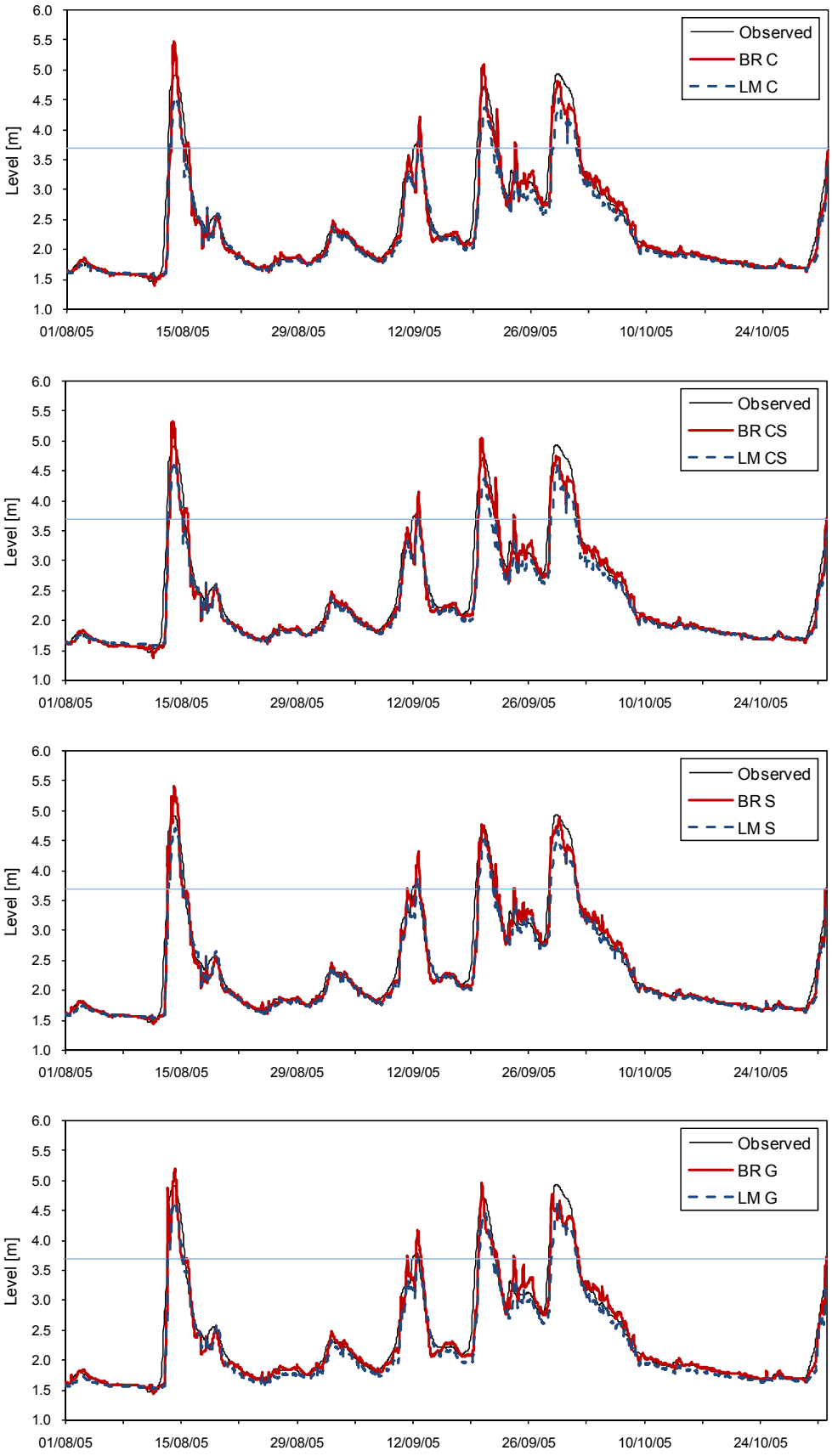


Figure B.6: Dataset two, compare of the BR and LM model at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 18 hours.

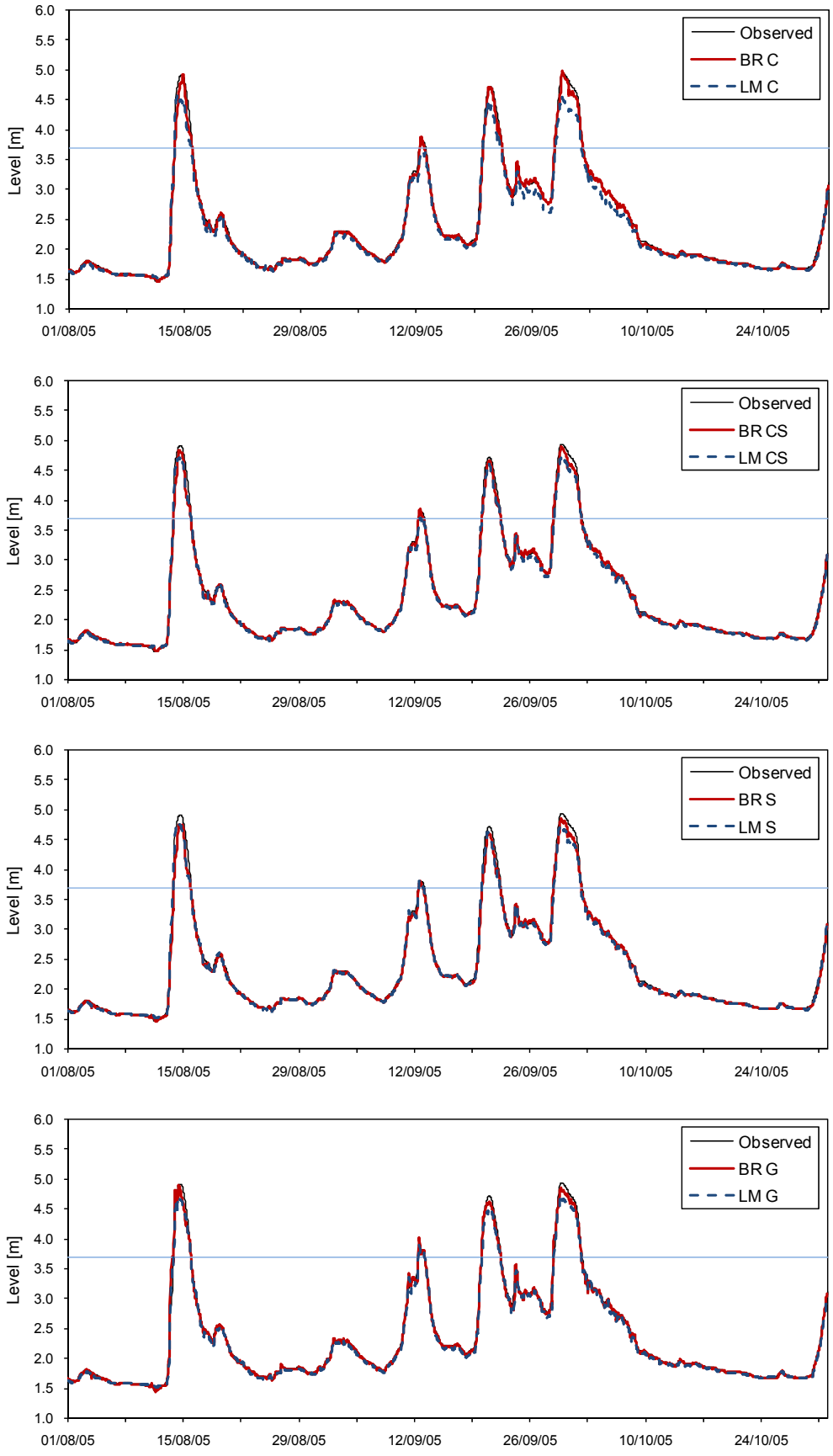


Figure B.7: Dataset three, compare of the BR and LM model at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 6 hours.

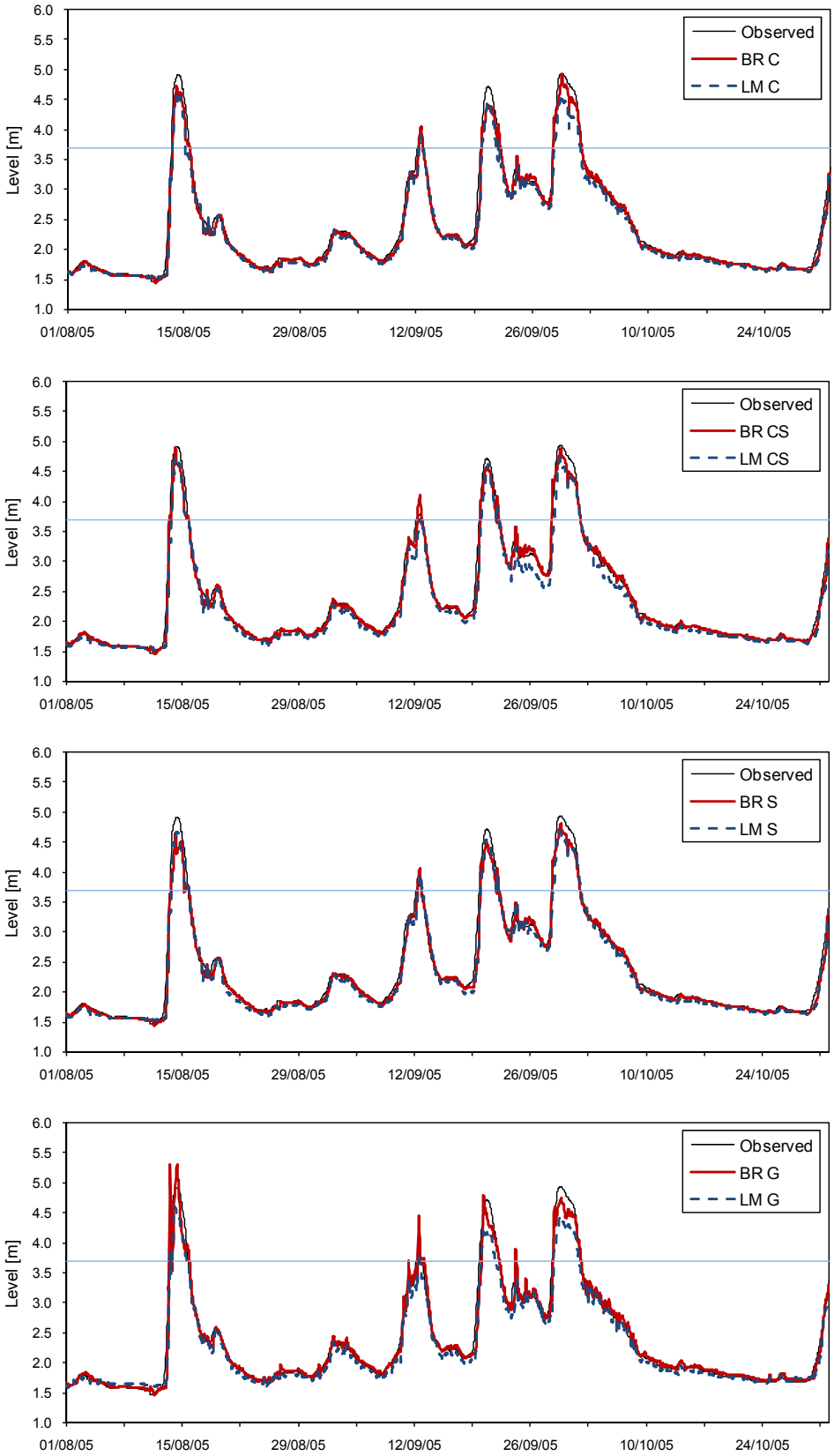


Figure B.8: Dataset three, compare of the BR and LM model at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 12 hours.

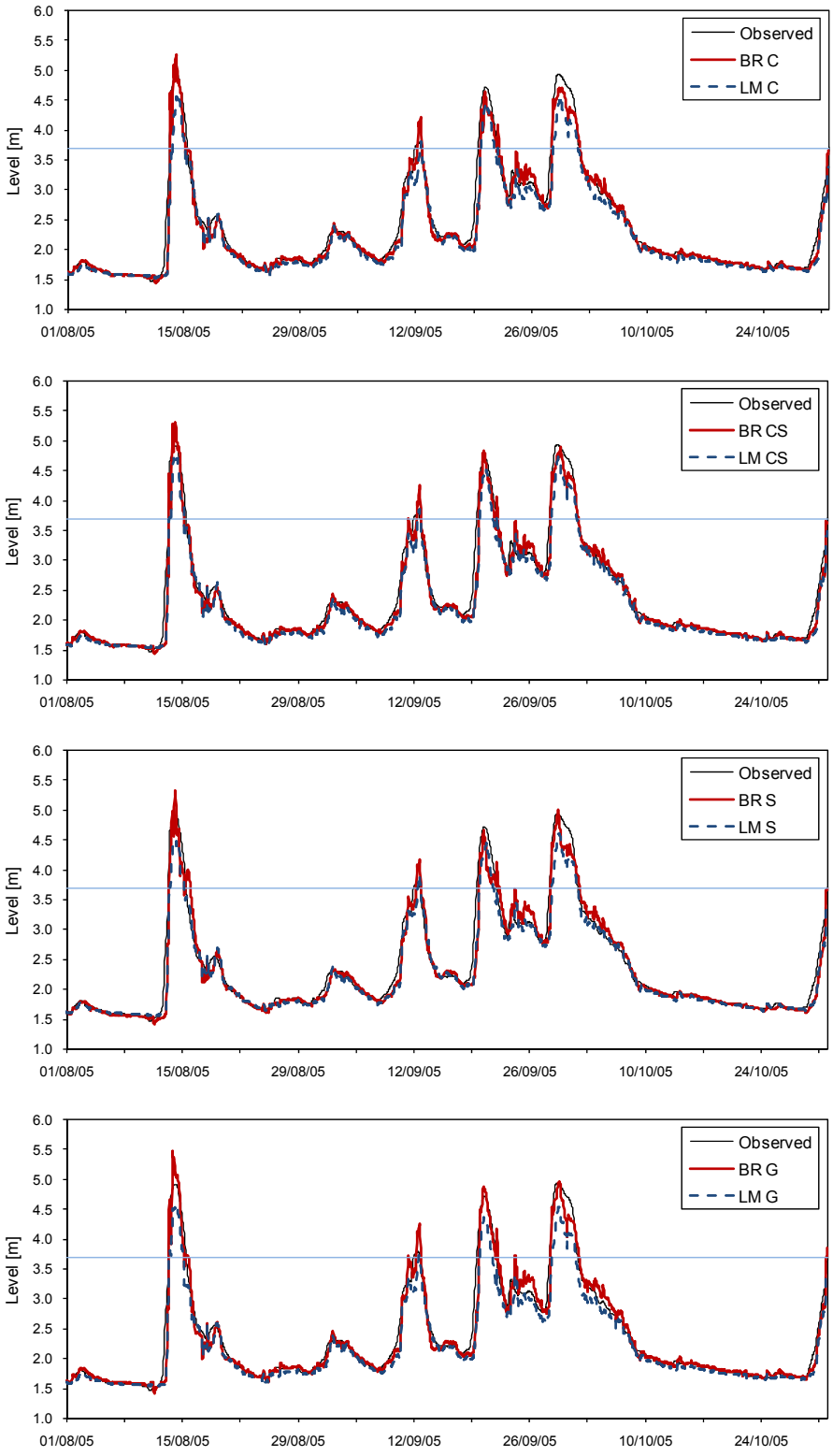


Figure B.9: Dataset three, compare of the BR and LM model at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 18 hours.

Note

C: Correlation > 0.9, S: Stepwise regression, CS: Correlation > 0.9 + Stepwise regression, G: Genetic algorithms

BR: Bayesian regularization, LM: Levenberg-Marquardt

Appendix C: Hydrographs of Case Study 3

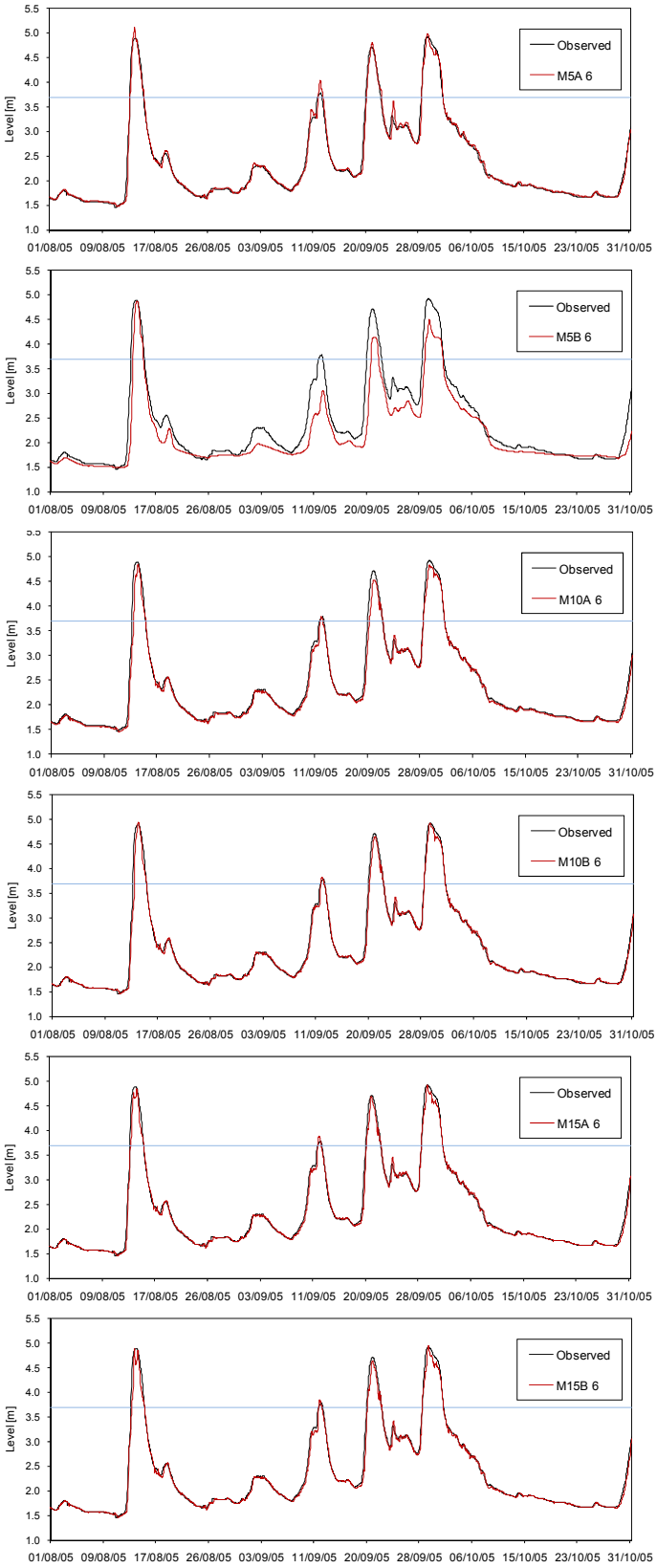


Figure C.1: Observed and predicted water level at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 6 hours.

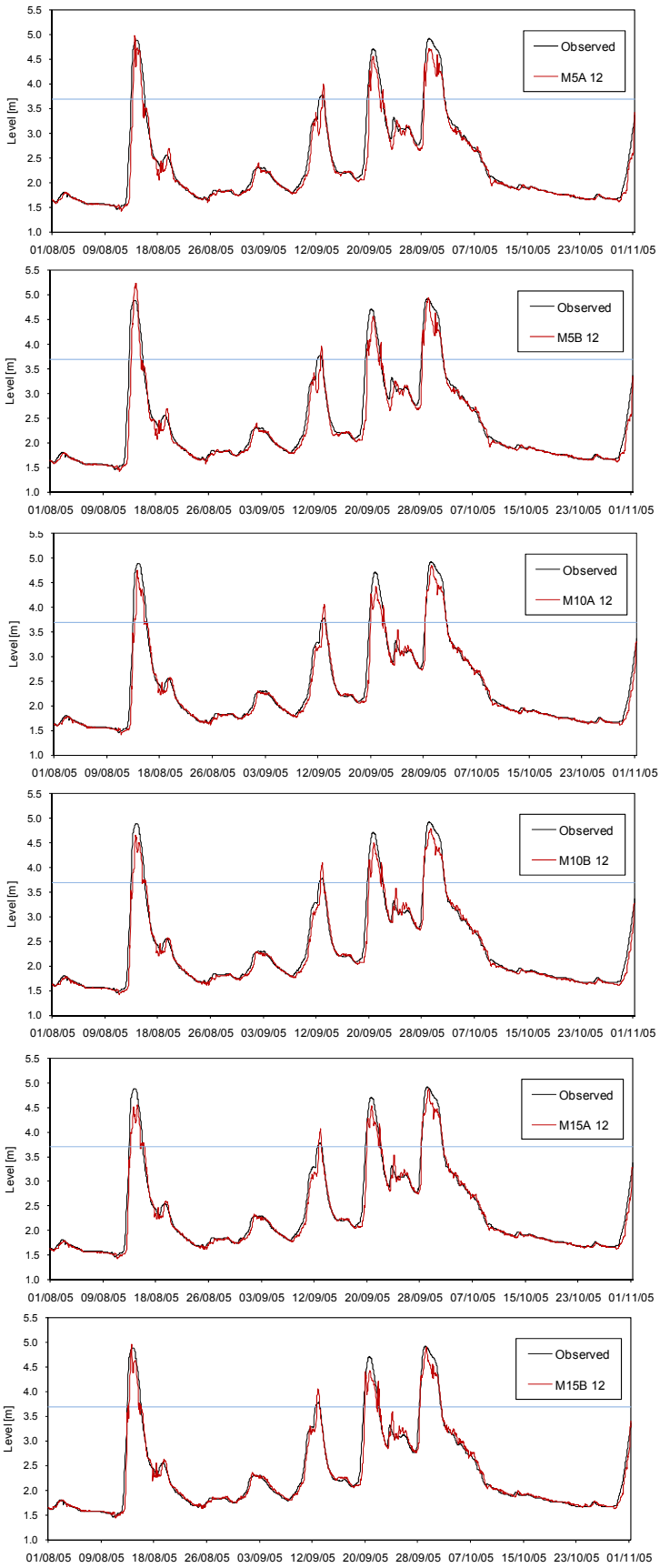


Figure C.2: Observed and predicted water level at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 12 hours.

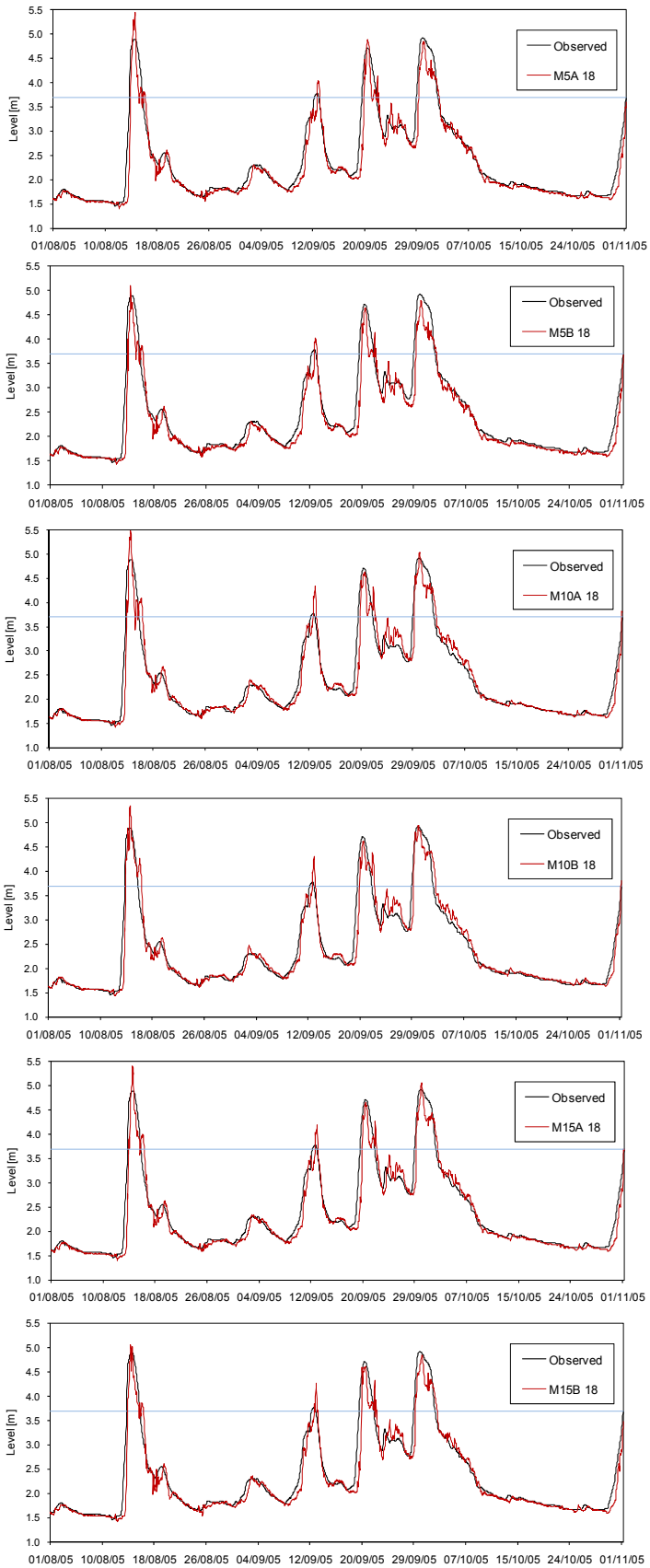


Figure C.3: Observed and predicted water level at P1 station, 1 Aug – 31 Oct 2005 (testing data) for lead time of 18 hours.

Appendix D: Hydrographs of Case Study 4

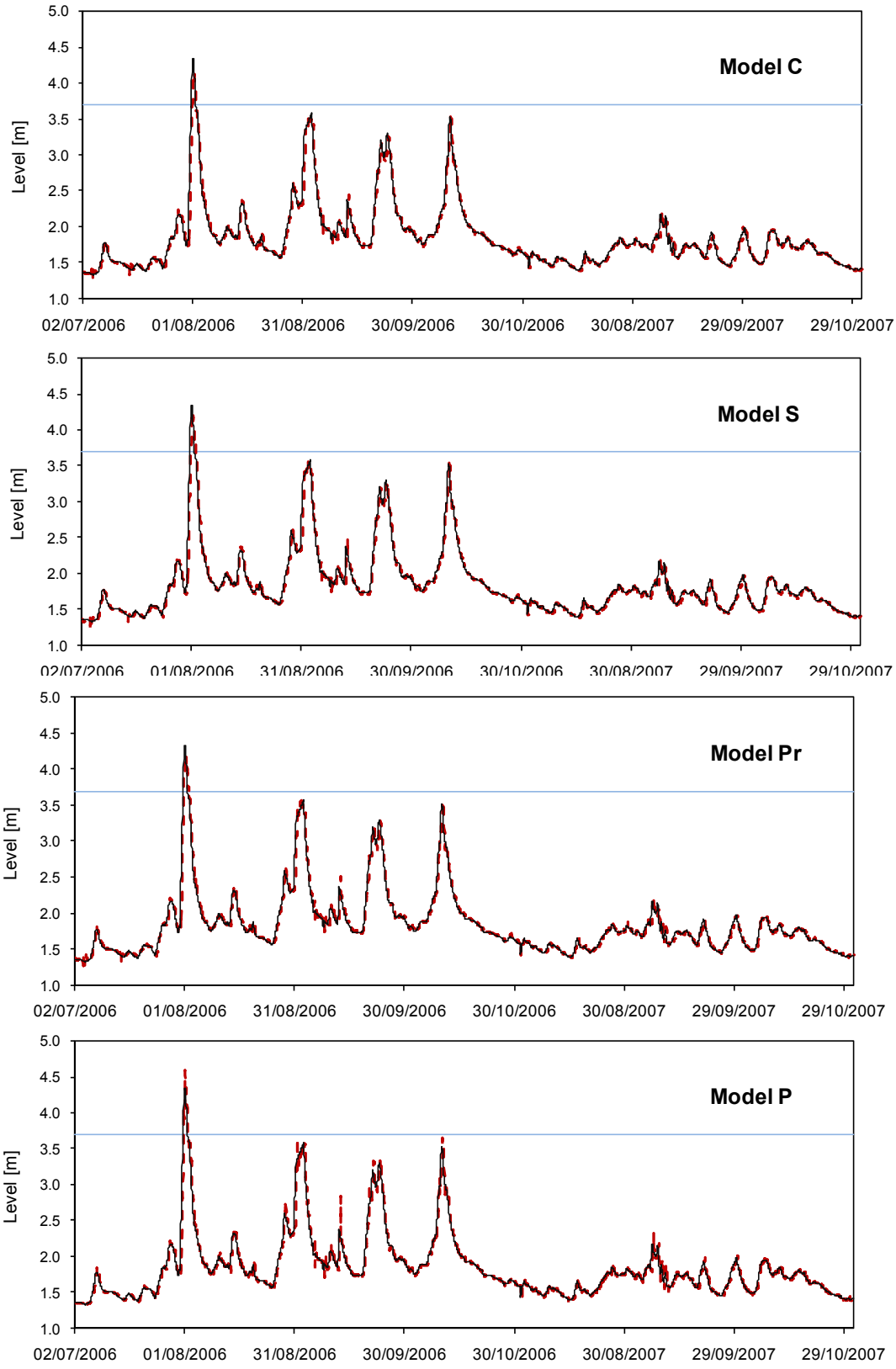


Figure D.1: Observed and predicted water level at P1 station, 2 Jul – 31 Oct 2006 and 2 Aug – 31 Oct 2007 (testing data) for lead time of 6 hours.

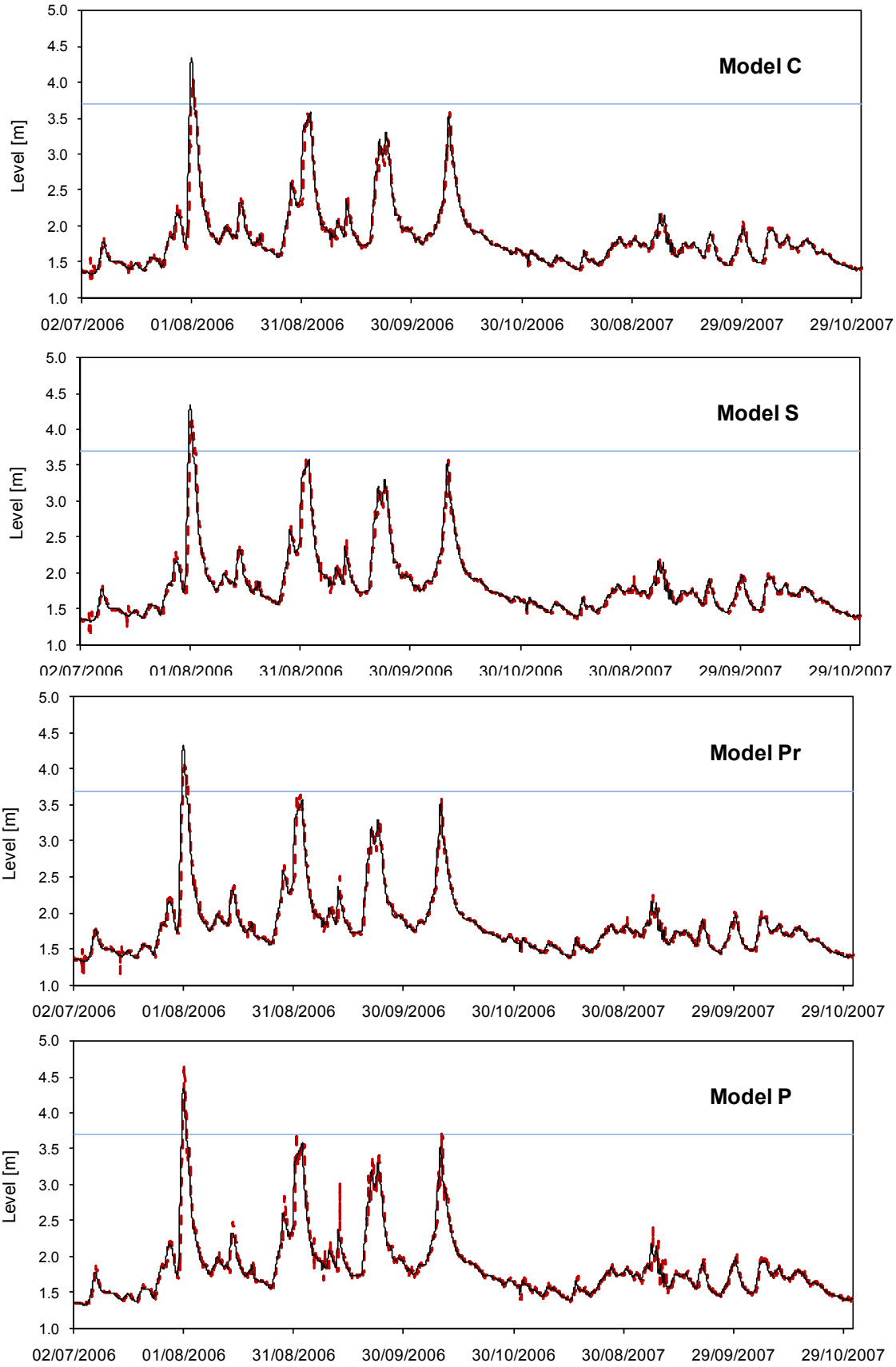


Figure D.2: Observed and predicted water level at P1 station, 2 Jul – 31 Oct 2006 and 2 Aug – 31 Oct 2007 (testing data) for lead time of 9 hours.

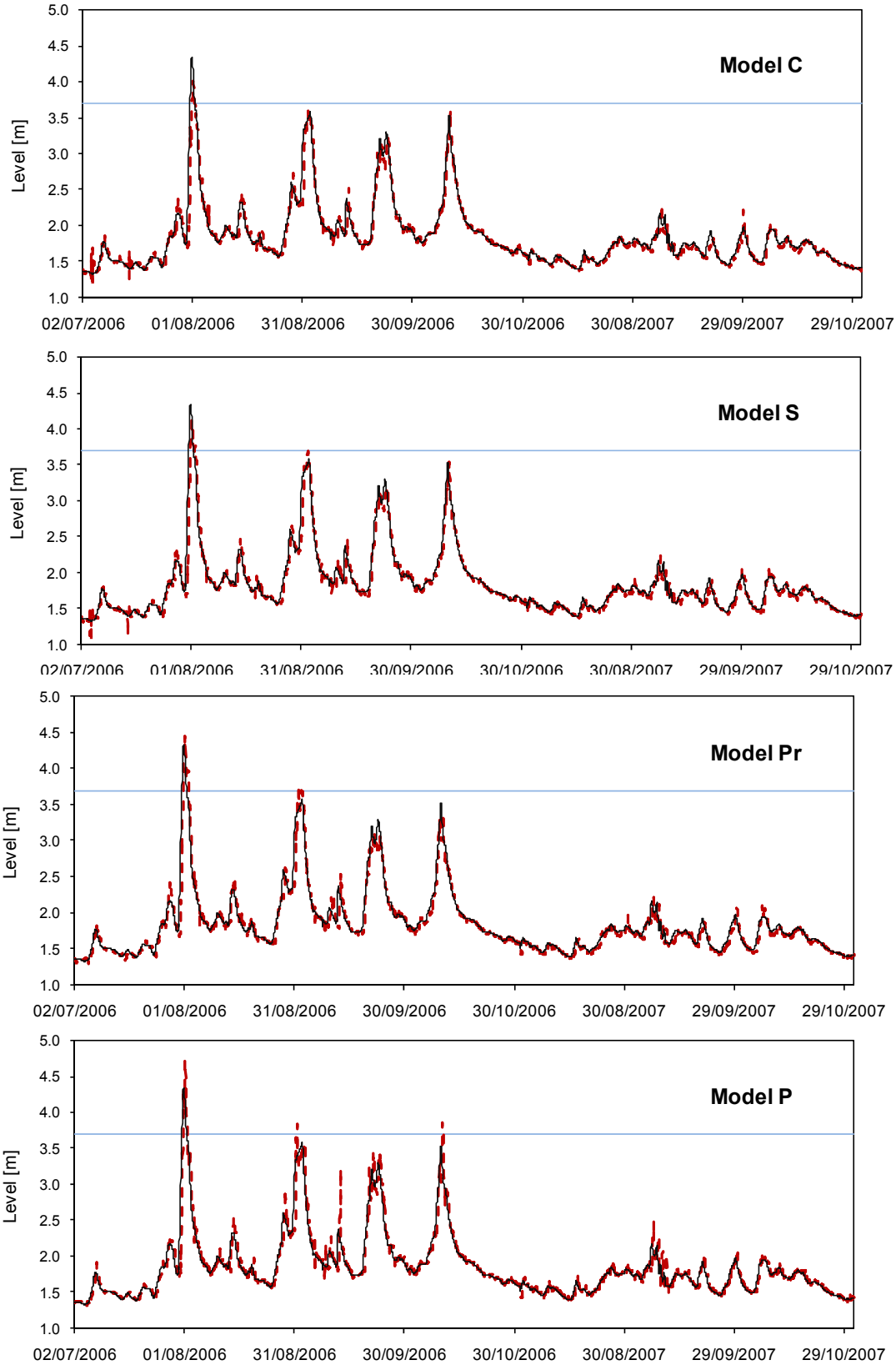


Figure D.3: Observed and predicted water level at P1 station, 2 Jul – 31 Oct 2006 and 2 Aug – 31 Oct 2007 (testing data) for lead time of 12 hours.

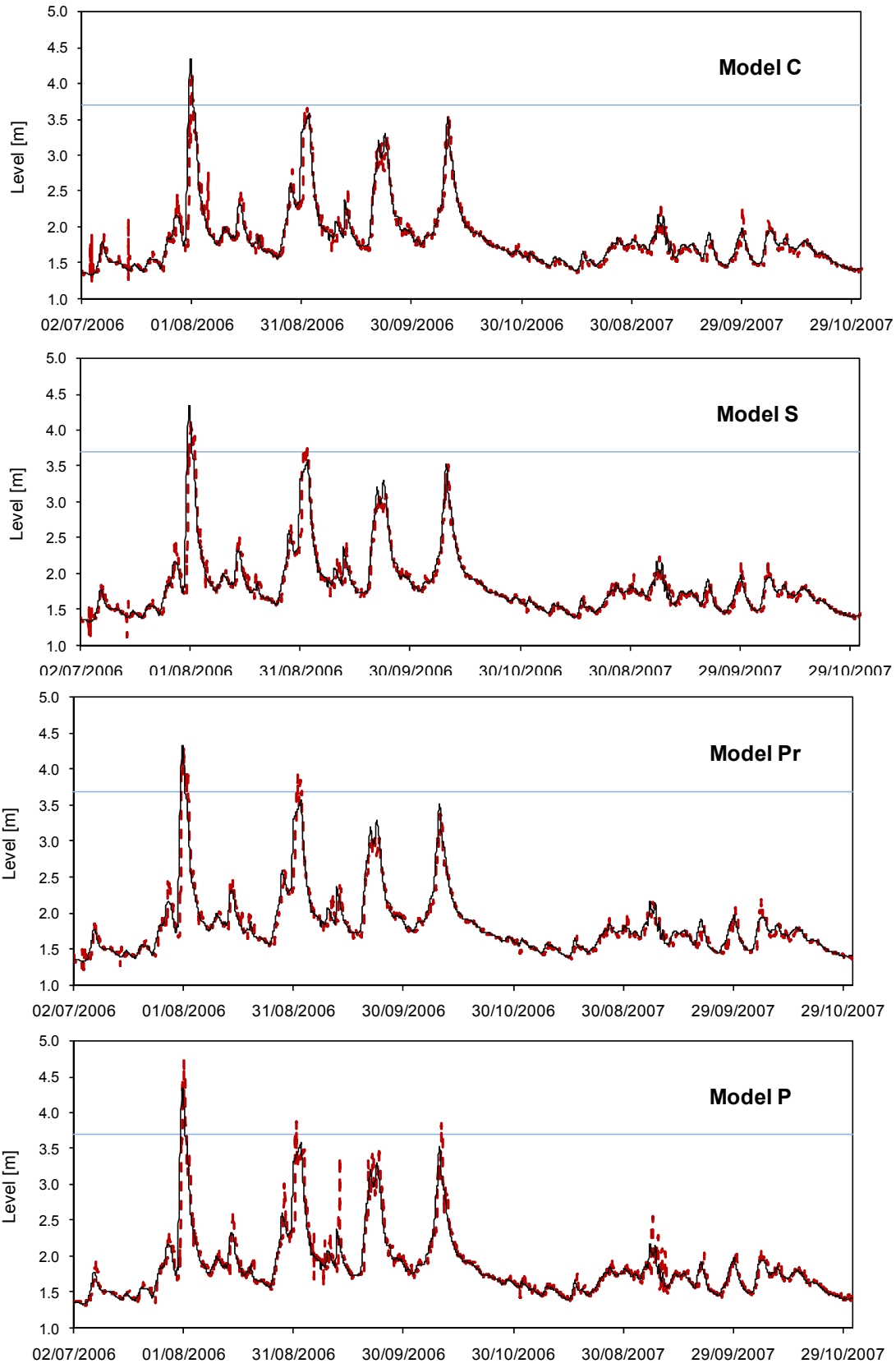


Figure D.4: Observed and predicted water level at P1 station, 2 Jul – 31 Oct 2006 and 2 Aug – 31 Oct 2007 (testing data) for lead time of 15 hours.

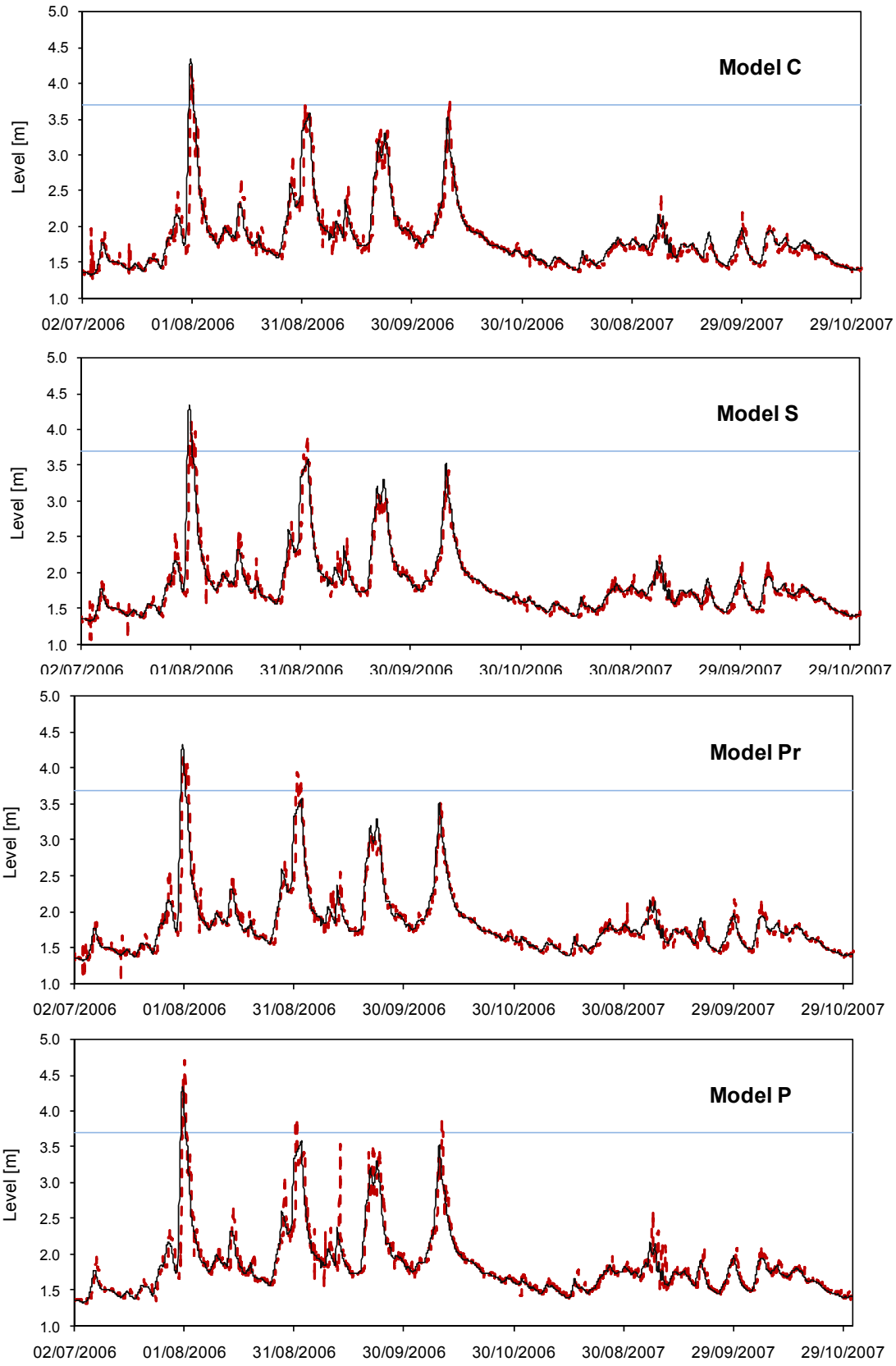


Figure D.5: Observed and predicted water level at P1 station, 2 Jul – 31 Oct 2006 and 2 Aug – 31 Oct 2007 (testing data) for lead time of 18 hours.

Appendix E

Table E.1: Input variables selected by four input determination techniques (denoted by x) for 6, 9, 12, 15, 18, 21 and 24 hour lead time of Experiment 1.

E1	Correlation							Stepwise Regression							Data Mining							Genetic Algorithm						
	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24
Z11																								x		x		x
Z12																								x			x	
Z13																							x	x		x	x	x
12Z11										x	x													x				x
12Z12																												x
12Z13									x	x	x	x	x													x		
24Z11									x			x												x	x			
24Z12																							x			x		
24Z13								x	x		x														x	x	x	
36Z11								x					x	x											x		x	x
36Z12																										x	x	
36Z13														x								x						x
48Z11																								x	x		x	x
48Z12																									x	x		
48Z13																									x	x		
60Z11																							x					
60Z12																						x	x	x	x	x		
60Z13																							x	x		x	x	
72Z11																								x	x		x	x
72Z12																							x				x	x
72Z13																							x		x	x		
84Z11													x	x									x		x			
84Z12										x				x									x	x		x	x	
84Z13								x	x	x	x	x	x	x								x				x		
96Z11																							x					
96Z12								x	x		x	x	x															x
96Z13																												
108Z11																							x	x	x		x	x
108Z12																						x			x			x
108Z13																						x		x		x		
120Z11																										x		x
120Z12												x																
120Z13												x																
132Z11																						x	x		x			
132Z12									x		x		x	x								x	x					
132Z13									x			x	x	x														
144Z11									x		x											x						
144Z12								x		x													x		x			
144Z13								x	x	x	x			x										x				
156Z11								x		x		x	x									x			x	x		
156Z12											x	x																
156Z13												x											x					
168Z11														x														
168Z12																										x		x
168Z13																							x	x	x	x	x	
Z21																										x		x
Z22																				x	x		x			x	x	x
Z23																				x	x		x				x	x
12Z21								x	x				x	x									x					x
12Z22																							x	x	x		x	x
12Z23								x														x			x		x	
24Z21								x		x	x		x													x	x	x
24Z22											x		x									x	x	x		x	x	x
24Z23																								x			x	x
36Z21											x	x		x													x	
36Z22												x	x	x					x	x	x				x	x	x	x
36Z23																										x		
48Z21								x	x	x	x	x	x											x	x	x	x	x
48Z22								x									x	x	x	x	x	x	x	x		x		x
48Z23								x	x	x	x	x	x	x											x	x		x
60Z21								x														x	x			x	x	x
60Z22							x				x					x	x	x	x	x	x	x	x	x	x	x	x	
60Z23											x														x	x		x	
72Z21									x														x		x			x
72Z22					x	x	x		x						x	x		x	x	x			x		x	x	x	
72Z23																								x	x		x	
84Z21								x														x				x		x
84Z22					x	x	x															x		x	x		x	
84Z23								x		x		x	x	x											x		x	x
96Z21									x	x	x	x										x		x		x		
96Z22	x	x	x	x	x																	x	x	x	x	x	x	x
96Z23										x	x		x	x														x
108Z21									x													x	x	x		x		
108Z22	x	x			x																		x	x	x		x	x
108Z23								x	x			x										x					x	
120Z21								x		x		x	x	x												x	x	
120Z22																										x		x
120Z23													x												x		x	x
132Z21								x	x	x	x												x				x	x
132Z22														x									x	x	x	x		
132Z23														x									x			x		x
144Z21																						x	x	x	x	x	x	x
144Z22												x	x										x	x	x			
144Z23																							x	x				
156Z21								x				x	x											x	x	x		x	
156Z22								x				x	x									x	x	x	x	x		x
156Z23												x	x									x	x		x		x	x
168Z21								x	x	x	x	x	x	x								x	x	x			x	x
168Z22									x	x	x			x								x	x	x		x		x
168Z23								x	x	x	x	x	x	x											x	x		x	
Total	2	2	1	1	4	2	3	23	21	18	22	24	24	21	1	2	2	3	4	6	5	25	40	39	40	42	37	39

Table E.2: Input variables selected by four input determination techniques (denoted by x) for 6, 9, 12, 15, 18, 21 and 24 hour lead time of Experiment 1 (continue).

E1	Correlation							Stepwise Regression							Data Mining							Genetic Algorithm						
	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24
Z31																				x	x	x			x	x	x	
Z32																				x	x					x		
Z33																						x	x	x	x	x	x	
12Z31								x	x				x	x										x			x	x
12Z32																										x	x	x
12Z33								x			x																x	x
24Z31								x		x	x	x	x											x		x	x	
24Z32													x														x	x
24Z33																											x	
36Z31														x										x		x	x	
36Z32														x								x						
36Z33								x			x	x	x	x												x		x
48Z31									x	x	x	x	x												x	x		x
48Z32									x	x	x														x		x	x
48Z33																							x	x	x		x	x
60Z31								x	x															x	x		x	x
60Z32								x				x		x									x	x		x	x	x
60Z33										x		x	x											x			x	
72Z31																						x	x	x				
72Z32								x	x													x	x	x		x		
72Z33														x										x	x			
84Z31								x					x									x				x		
84Z32										x				x									x	x	x		x	
84Z33																							x	x	x			
96Z31									x					x											x			
96Z32										x	x		x	x											x	x		
96Z33										x				x										x				x
108Z31																												
108Z32								x	x			x										x		x				
108Z33								x	x																x	x		x
120Z31																									x			
120Z32																						x				x		x
120Z33														x								x	x				x	
132Z31														x														
132Z32																								x			x	
132Z33																												
144Z31									x															x		x		
144Z32																						x		x				
144Z33																						x		x				
156Z31																											x	
156Z32										x													x				x	
156Z33										x				x									x					x
168Z31								x	x	x	x	x	x	x									x		x			x
168Z32								x	x		x	x	x	x								x		x	x			
168Z33								x	x	x	x	x	x	x											x		x	
Z41																			x		x				x	x		
Z42																												x
Z43																				x	x			x				x
12Z41										x	x		x	x														x
12Z42										x			x	x								x		x		x	x	x
12Z43																							x	x				
24Z41								x	x															x	x	x	x	
24Z42																						x		x			x	x
24Z43																										x		x
36Z41												x														x		x
36Z42																								x	x		x	x
36Z43																							x					
48Z41														x										x		x	x	x
48Z42									x	x	x	x	x											x	x			
48Z43									x			x											x			x		
60Z41									x													x		x		x	x	
60Z42								x						x									x	x	x	x	x	x
60Z43											x	x	x											x	x			x
72Z41														x								x	x	x		x		x
72Z42				x	x	x	x																		x		x	
72Z43									x													x	x		x	x		x
84Z41																							x	x	x			
84Z42	x	x	x	x	x	x	x			x				x				x	x	x	x		x	x		x	x	x
84Z43								x				x	x											x				x
96Z41										x	x	x	x									x						
96Z42	x	x	x	x	x	x	x			x	x		x	x	x	x	x					x	x	x	x	x		x
96Z43											x																x	
108Z41								x						x										x	x	x		
108Z42	x	x	x	x	x	x	x	x	x			x		x										x		x		x
108Z43								x	x			x		x											x			
120Z41																								x			x	
120Z42	x	x	x	x	x	x	x	x		x		x	x										x		x	x	x	
120Z43										x		x	x	x														
132Z41									x														x		x		x	
132Z42	x	x	x		x		x																	x	x		x	
132Z43																										x		
144Z41														x									x	x		x		x	x
144Z42	x	x			x								x						x	x	x		x	x	x	x		x	x
144Z43													x												x		x		x
156Z41													x													x	x		
156Z42								x				x		x	x	x	x						x	x	x	x	x	x	x
156Z43												x	x	x	x								x	x			x		
168Z41								x	x	x	x	x	x	x									x	x		x	x	x	x
168Z42								x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
168Z43								x	x	x													x			x	x	x	x
Total	6	6	5	5	7	5	6	24	23	22	19	25	25	31	2	3	3	3	4	6	6	28	30	45	37	38	39	38

Table E.3: The goodness of fit measures for a lead time 6, 9, 12, 15, 18, 21 and 24 hours of Experiment 1.

Model	Statistic	Input Determination Techniques			
		C	S	G	D
t+6	Reduce input	8 (95.5%)	47 (73.8%)	53 (70.5%)	3 (98.3%)
	PDIFF (m)	0.4997	0.2716	0.6707	0.5156
	MAE (m)	0.3782	0.2550	0.3034	0.4191
	RMSE (m)	0.4869	0.3748	0.3937	0.5209
	CE	0.7026	0.8238	0.8056	0.6597
	Time delay (hr)	-13	8	10	-12
t+9	Reduce input	8 (95.5%)	44 (75.5%)	70 (61.1%)	5 (97.2%)
	PDIFF (m)	0.1042	0.5500	0.3725	0.3799
	MAE (m)	0.3753	0.3062	0.2721	0.3569
	RMSE (m)	0.5021	0.4106	0.3640	0.4810
	CE	0.6837	0.7885	0.8337	0.7097
	Time delay (hr)	-10	11	11	-9
t+12	Reduce input	6 (96.6%)	40 (77.7%)	84 (53.3%)	5 (97.2%)
	PDIFF (m)	-0.223	0.3294	-0.1592	0.4858
	MAE (m)	0.3692	0.3395	0.2608	0.3426
	RMSE (m)	0.4758	0.4141	0.3214	0.4735
	CE	0.7159	0.7848	0.8704	0.7187
	Time delay (hr)	-7.5	3	2	-6
t+15	Reduce input	6 (96.6%)	41 (77.2%)	77 (57.2%)	6 (96.9%)
	PDIFF (m)	-0.1058	0.0521	0.3874	0.4212
	MAE (m)	0.3706	0.2913	0.3560	0.3198
	RMSE (m)	0.4763	0.3775	0.4741	0.4145
	CE	0.7152	0.8212	0.7179	0.7844
	Time delay (hr)	-4.5	7	11	-4
t+18	Reduce input	11 (93.8%)	49 (72.7%)	80 (55.5%)	8 (95.5%)
	PDIFF (m)	0.1279	0.4658	0.0414	-0.0661
	MAE (m)	0.3713	0.3902	0.2989	0.3045
	RMSE (m)	0.4519	0.4946	0.3613	0.3941
	CE	0.7437	0.6930	0.8362	0.8051
	Time delay (hr)	-1	24	2.5	-2
t+21	Reduce input	7 (96.1%)	49 (72.7%)	76 (57.7%)	12 (93.3%)
	PDIFF (m)	-0.0141	0.3426	0.0603	0.0895
	MAE (m)	0.4015	0.3401	0.3063	0.3051
	RMSE (m)	0.5175	0.4607	0.4075	0.3881
	CE	0.6641	0.7337	0.7917	0.8110
	Time delay (hr)	1.5	11	14	0
t+24	Reduce input	9 (95%)	52 (71.1%)	77 (57.2%)	11 (93.8%)
	PDIFF (m)	-0.0324	0.5367	0.4740	-0.3966
	MAE (m)	0.3992	0.3584	0.3956	0.2885
	RMSE (m)	0.5255	0.4710	0.5271	0.3847
	CE	0.6537	0.7219	0.6516	0.8145
	Time delay (hr)	4	6	20	1

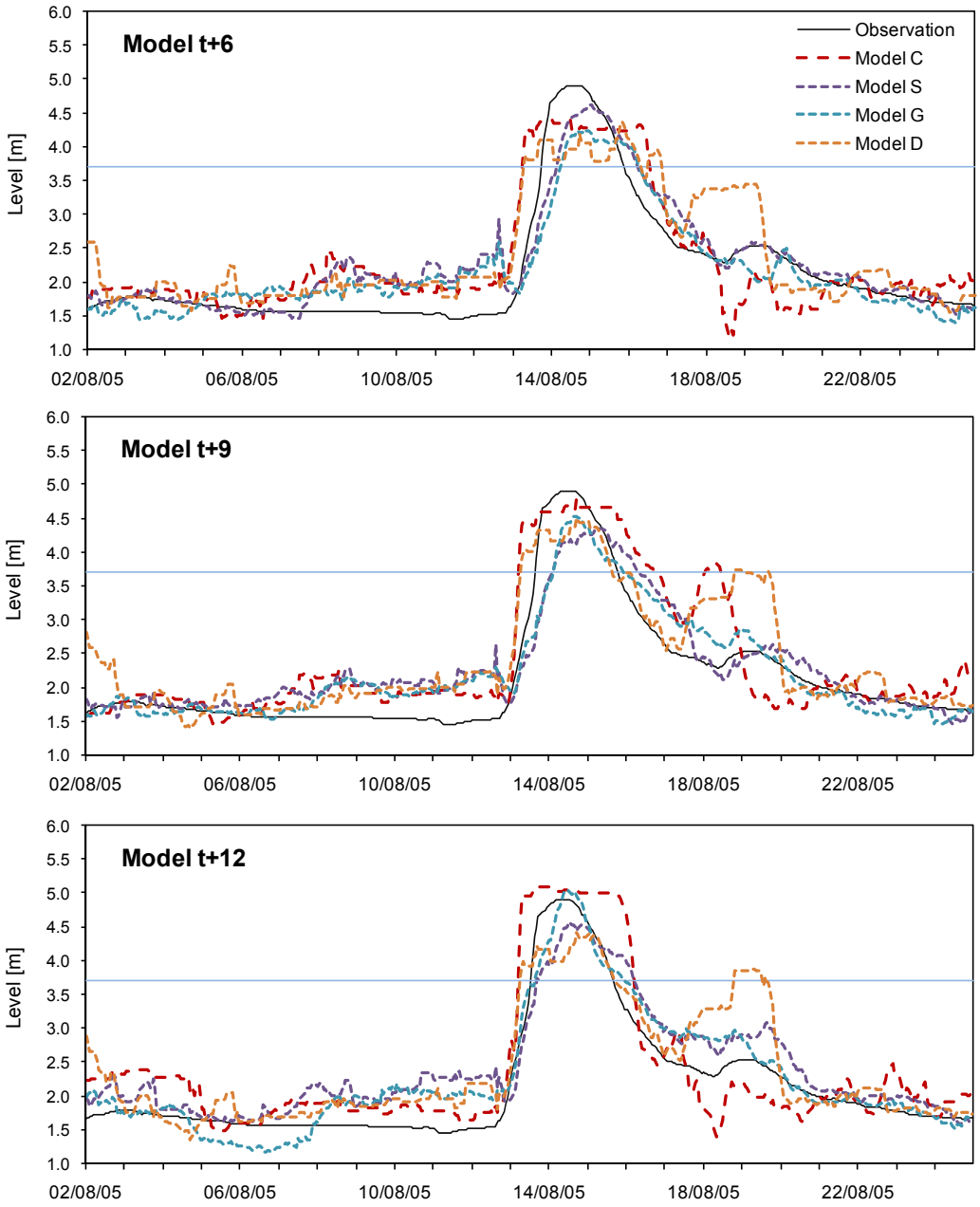


Figure E.1: Comparison of the hydrographs for 4 models at 6, 9 and 12 hour lead time, testing S2 (2 – 24 Aug 2005) of Experiment 1.

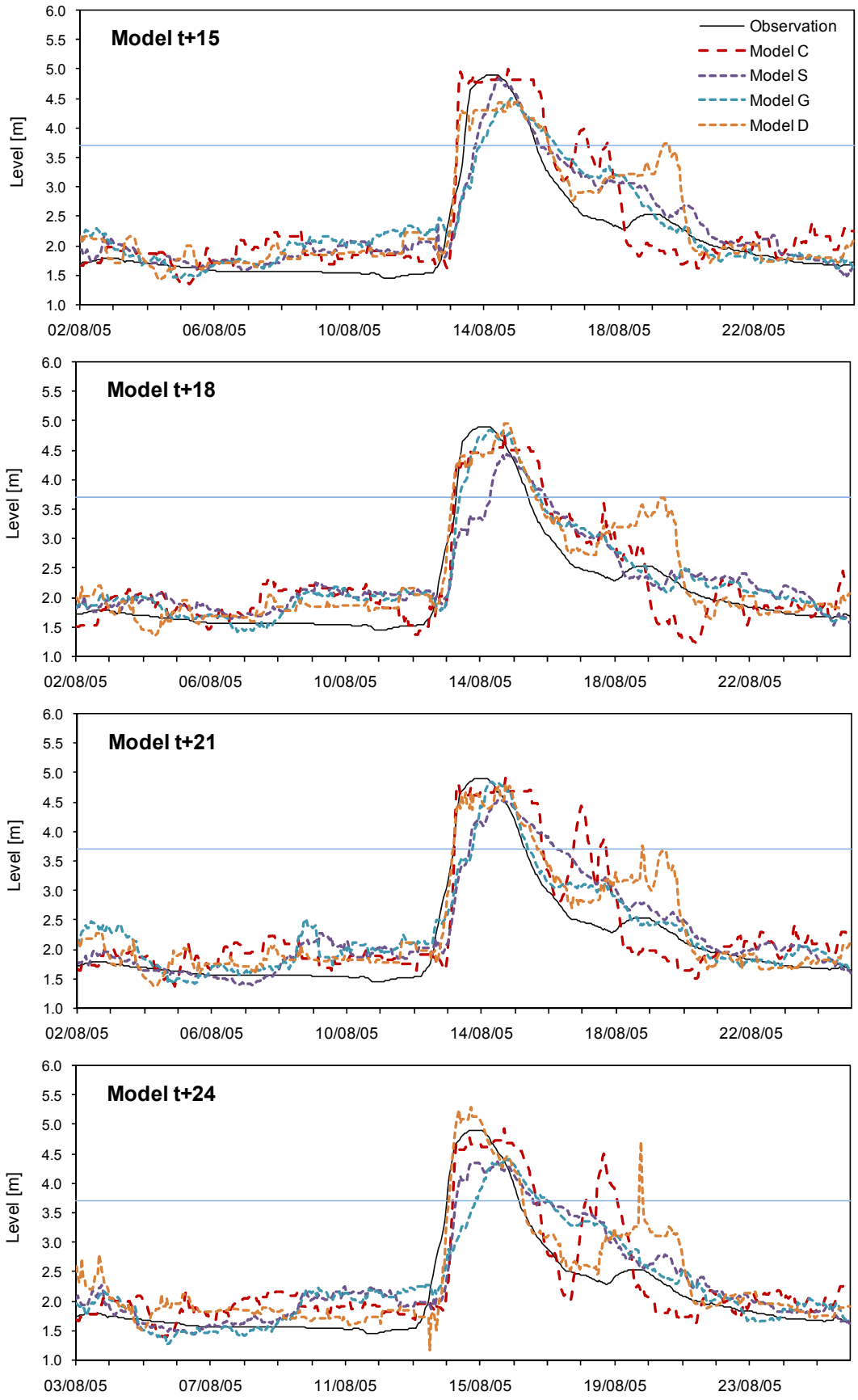


Figure E.2: Comparison of the hydrographs for 4 models at 15, 18, 21 and 24 hour lead time, testing S2 (2 – 24 Aug 2005) of Experiment 1.

Table E.4: Input variables selected by four input determination techniques (denoted by x) for 6, 9, 12, 15, 18, 21 and 24 hour lead time of Experiment 2.

E2	Correlation							Stepwise Regression							Data Mining							Genetic Algorithm						
	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24
Z11								x	x	x	x	x	x	x							x	x	x	x			x	x
Z12										x	x									x	x	x	x					
Z13																						x	x					x
12Z11																							x	x	x	x		x
12Z12											x	x	x	x									x	x		x		
12Z13								x	x	x	x	x	x	x								x		x	x		x	x
24Z11									x				x	x											x	x	x	x
24Z12																							x					x
24Z13								x	x	x	x	x	x	x											x	x	x	x
36Z11																							x		x		x	x
36Z12								x	x	x	x	x							x		x					x	x	
36Z13								x	x	x	x	x	x	x								x						x
48Z11								x				x	x	x							x		x	x	x		x	x
48Z12						x	x	x												x	x		x		x	x		
48Z13									x	x	x		x	x											x	x		
60Z11													x	x									x		x			
60Z12					x	x	x		x												x	x	x	x	x	x		
60Z13								x	x			x											x	x		x	x	
72Z11																								x	x		x	x
72Z12		x	x	x	x	x	x												x	x			x				x	x
72Z13									x														x		x	x		
84Z11								x	x	x	x	x	x	x									x		x			
84Z12	x	x	x	x	x	x								x	x	x	x						x	x		x	x	
84Z13									x													x				x		
96Z11								x		x	x	x	x	x									x					
96Z12	x	x	x	x											x													x
96Z13								x	x																			
108Z11									x														x	x	x		x	x
108Z12																						x			x			x
108Z13																						x		x		x		
120Z11										x	x	x	x	x												x		x
120Z12								x	x	x											x							
120Z13										x			x	x														
132Z11										x	x	x	x	x								x	x		x			
132Z12								x	x	x	x	x	x	x								x	x					
132Z13									x	x	x	x	x	x														
144Z11																						x						
144Z12																				x			x		x			
144Z13												x												x				
156Z11									x													x		x	x			
156Z12																x			x									
156Z13								x	x	x	x		x	x									x					
168Z11								x	x	x	x	x	x	x														
168Z12								x	x	x	x	x	x	x	x		x	x			x					x		x
168Z13								x	x	x	x	x	x	x									x	x	x	x	x	
Z21																									x			x
Z22									x											x	x		x		x	x		x
Z23																							x			x	x	
12Z21								x															x				x	
12Z22																							x	x	x		x	x
12Z23								x	x	x	x	x	x	x								x			x		x	
24Z21										x																x	x	x
24Z22								x	x		x	x	x	x								x	x	x		x	x	x
24Z23																								x			x	x
36Z21								x	x		x	x	x	x													x	
36Z22										x										x	x				x	x	x	x
36Z23								x	x			x	x	x												x		
48Z21									x		x		x	x										x	x	x	x	x
48Z22											x	x	x	x			x	x	x			x	x	x		x		x
48Z23									x	x	x	x	x	x											x	x		x
60Z21																						x	x			x	x	x
60Z22																						x	x	x	x	x	x	
60Z23								x					x	x										x	x		x	
72Z21								x		x	x		x	x									x		x			x
72Z22								x			x												x		x	x	x	
72Z23																								x	x		x	
84Z21								x	x	x	x	x	x	x								x				x		x
84Z22										x			x	x								x		x	x		x	
84Z23									x	x	x	x	x	x										x		x		x
96Z21								x	x	x	x	x	x	x								x		x		x		
96Z22														x								x	x	x	x	x	x	x
96Z23								x	x	x	x	x	x	x														x
108Z21																						x	x	x		x		
108Z22								x		x	x	x		x									x	x	x		x	x
108Z23								x	x	x	x	x		x								x					x	
120Z21								x		x															x	x		
120Z22								x	x	x	x	x	x	x											x			x
120Z23								x	x	x	x	x	x	x										x		x	x	
132Z21												x	x	x									x			x		x
132Z22																							x	x	x	x		
132Z23																							x			x		x
144Z21								x	x	x	x	x	x	x								x	x	x	x	x	x	x
144Z22											x	x	x	x									x	x	x			
144Z23											x												x	x				
156Z21								x	x	x													x	x	x		x	
156Z22								x	x	x	x		x	x								x	x	x	x	x		x
156Z23																						x	x		x		x	x
168Z21								x	x	x	x	x	x	x								x	x	x			x	x
168Z22								x	x	x	x	x	x	x								x	x	x		x		x
168Z23								x	x	x	x	x	x	x										x	x		x	
Total	2	3	3	3	3	4	3	39	42	40	41	38	43	47	3	2	3	2	4	6	10	29	47	38	43	41	38	42

Table E.5: Input variables selected by four input determination techniques (denoted by x) for 6, 9, 12, 15, 18, 21 and 24 hour lead time of Experiment 2 (continue).

E2	Correlation							Stepwise Regression							Data Mining							Genetic Algorithm						
	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24
Z31																					x						x	x
Z32																												
Z33																						x	x			x		x
12Z31																								x		x	x	x
12Z32									x														x				x	
12Z33									x																			x
24Z31								x		x												x		x		x	x	
24Z32									x		x	x		x									x					x
24Z33											x	x	x	x									x		x			x
36Z31									x														x	x	x		x	x
36Z32								x	x	x		x	x										x			x	x	
36Z33								x		x	x	x	x	x														
48Z31								x	x	x	x	x													x			x
48Z32									x																x			x
48Z33								x	x													x	x		x			x
60Z31								x	x	x												x					x	
60Z32											x											x	x					
60Z33								x	x	x	x																	
72Z31								x															x			x		x
72Z32											x															x	x	x
72Z33								x		x													x	x			x	x
84Z31											x												x					
84Z32												x														x		
84Z33																										x	x	x
96Z31																												
96Z32												x	x	x								x	x					
96Z33													x	x														
108Z31																												
108Z32									x			x																
108Z33								x			x																	
120Z31								x	x					x								x						
120Z32									x			x	x	x														
120Z33									x													x						
132Z31								x		x																		
132Z32																												
132Z33													x	x														
144Z31								x	x	x	x	x	x	x														
144Z32													x	x														
144Z33								x				x		x														
156Z31																						x		x				
156Z32									x													x	x	x				
156Z33									x	x	x		x	x								x						
168Z31									x			x	x										x					
168Z32								x	x	x	x	x	x	x														
168Z33								x	x	x	x	x	x	x														
Z41									x	x										x	x	x					x	
Z42																									x			
Z43																					x	x		x		x	x	x
12Z41										x		x	x	x											x			x
12Z42								x	x	x	x	x	x	x														
12Z43											x	x		x								x				x		
24Z41																								x		x	x	x
24Z42												x	x														x	
24Z43									x	x																		x
36Z41								x	x		x			x							x	x	x	x			x	
36Z42								x	x		x	x	x	x										x		x		x
36Z43								x			x	x		x								x					x	x
48Z41									x		x	x	x	x								x			x		x	
48Z42								x	x	x	x	x	x	x				x	x	x	x	x	x	x				
48Z43								x	x	x	x	x	x	x								x	x	x	x	x	x	x
60Z41								x		x		x			x	x						x		x	x	x		x
60Z42								x		x		x					x					x	x	x				
60Z43																							x					
72Z41								x	x	x	x	x	x	x								x	x	x				
72Z42			x	x											x								x		x	x		
72Z43																						x	x		x		x	x
84Z41										x		x	x	x								x		x				
84Z42	x	x	x							x	x	x	x	x										x		x	x	x
84Z43								x			x												x					
96Z41									x	x			x	x										x				
96Z42								x	x	x	x											x	x	x				x
96Z43																								x				
108Z41																						x						
108Z42										x													x		x	x	x	x
108Z43														x										x				
120Z41																									x		x	
120Z42								x	x		x	x	x	x								x	x	x				
120Z43																									x		x	x
132Z41											x	x		x										x				
132Z42								x	x	x	x	x	x	x										x				
132Z43									x	x	x	x	x	x														
144Z41												x																
144Z42								x	x	x	x	x	x	x									x	x		x		
144Z43												x																
156Z41								x	x	x	x		x															
156Z42								x	x	x	x		x	x									x	x				
156Z43																						x	x					
168Z41								x	x	x	x	x	x	x								x		x				x
168Z42								x	x	x	x	x	x	x								x	x	x	x			
168Z43								x			x																	
Total	1	1	2	1	-	-	-	35	39	33	35	37	32	36	2	1	1	1	1	2	5	30	30	30	14	18	23	28

Table E.6: The goodness of fit measures for a lead time 6, 9, 12, 15, 18, 21 and 24 hours of Experiment 2.

Model	Statistic	Input Determination Techniques							
		S1				S2			
		C	S	G	D	C	S	G	D
t+6	Reduce inputs	3	74	59	5	(98.3%)	(58.8%)	(67.2%)	(97.2%)
	PDIFF (m)	-0.4659	-0.2741	0.1164	-0.2116	0.1031	-0.4892	-0.5872	-0.1607
	MAE (m)	0.4611	0.4370	0.4370	0.3629	0.3438	0.2722	0.3122	0.3327
	RMSE (m)	0.5380	0.5086	0.5155	0.4438	0.5685	0.4058	0.4664	0.5334
	CE	0.3220	0.3941	0.3775	0.5387	0.7979	0.8970	0.8639	0.8220
	Time delay (hr)	/	/	/	/	-14.5	-1	-8	-14.5
t+9	Reduce inputs	4	81	77	3	(97.7%)	(55.0%)	(57.2%)	(98.3%)
	PDIFF (m)	-0.4790	-0.2648	-0.4809	-0.2815	0.0847	-0.4981	-0.5497	0.0723
	MAE (m)	0.4634	0.4929	0.3577	0.3722	0.3346	0.3404	0.3136	0.3631
	RMSE (m)	0.5564	0.5789	0.4661	0.4515	0.5658	0.5038	0.4776	0.5535
	CE	0.2748	0.2149	0.4910	0.5225	0.7998	0.8412	0.8573	0.8083
	Time delay (hr)	/	/	/	/	-13.5	-3	-7	-13
t+12	Reduce inputs	5	73	68	4	(97.2%)	(59.4%)	(62.2%)	(97.7%)
	PDIFF (m)	-0.4868	-0.0562	-0.5403	-0.5557	0.0957	-0.4309	-0.6133	0.002
	MAE (m)	0.4506	0.4402	0.4813	0.3510	0.3260	0.2758	0.3295	0.3681
	RMSE (m)	0.5397	0.5006	0.5850	0.4378	0.5373	0.4153	0.4999	0.5700
	CE	0.3176	0.4129	0.1982	0.5510	0.8194	0.8921	0.8437	0.7968
	Time delay (hr)	/	/	/	/	-10.5	-6	-10.5	-10.5
t+15	Reduce inputs	4	76	61	3	(97.7%)	(57.7%)	(66.1%)	(98.3%)
	PDIFF (m)	-0.6731	-0.3422	-0.9349	-0.3698	0.1738	-0.7740	-0.9291	0.0133
	MAE (m)	0.4558	0.5184	0.3674	0.3592	0.3175	0.2835	0.3101	0.3309
	RMSE (m)	0.5528	0.5793	0.4665	0.4566	0.5140	0.4236	0.4900	0.5507
	CE	0.2837	0.2135	0.4900	0.5115	0.8348	0.8878	0.8498	0.8103
	Time delay (hr)	/	/	/	/	-8.5	-3.5	4	-8.5
t+18	Reduce inputs	3	75	59	5	(98.3%)	(58.3%)	(67.2%)	(97.2%)
	PDIFF (m)	-0.7268	-0.7041	-0.8693	-0.8578	0.1809	-0.6016	-0.7213	-0.1147
	MAE (m)	0.4927	0.5120	0.4905	0.3730	0.3423	0.302	0.3018	0.3417
	RMSE (m)	0.5670	0.5998	0.5919	0.4714	0.607	0.4461	0.5105	0.5509
	CE	0.2463	0.1565	0.1787	0.479	0.7696	0.8755	0.8370	0.8102
	Time delay (hr)	/	/	/	/	-6	-3.5	-2	-6
t+21	Reduce inputs	4	75	61	8	(97.7%)	(58.3%)	(66.1%)	(95.5%)
	PDIFF (m)	-0.6724	-0.2842	-0.6856	-0.9272	0.0993	-0.2462	-1.7652	-0.0169
	MAE (m)	0.4794	0.6200	0.3146	0.3887	0.3367	0.3256	0.3727	0.3488
	RMSE (m)	0.5529	0.6896	0.4057	0.4962	0.6024	0.5034	0.5724	0.5702
	CE	0.2832	-0.1153	0.6140	0.4227	0.7730	0.8415	0.7950	0.7966
	Time delay (hr)	/	/	/	/	-4	-3	-1	-4
t+24	Reduce inputs	3	83	70	15	(98.3%)	(53.8%)	(61.1%)	(91.6%)
	PDIFF (m)	-0.7555	-0.2458	-0.6123	-1.1116	0.1155	-0.3401	-1.2584	-0.2702
	MAE (m)	0.5065	0.6449	0.3640	0.3728	0.3306	0.4045	0.4239	0.3084
	RMSE (m)	0.5850	0.7286	0.4690	0.4593	0.5975	0.6350	0.6355	0.4819
	CE	0.1972	-0.2454	0.4841	0.5052	0.7767	0.7477	0.7474	0.8547
	Time delay (hr)	/	/	/	/	-2	-2.5	-2.5	-2

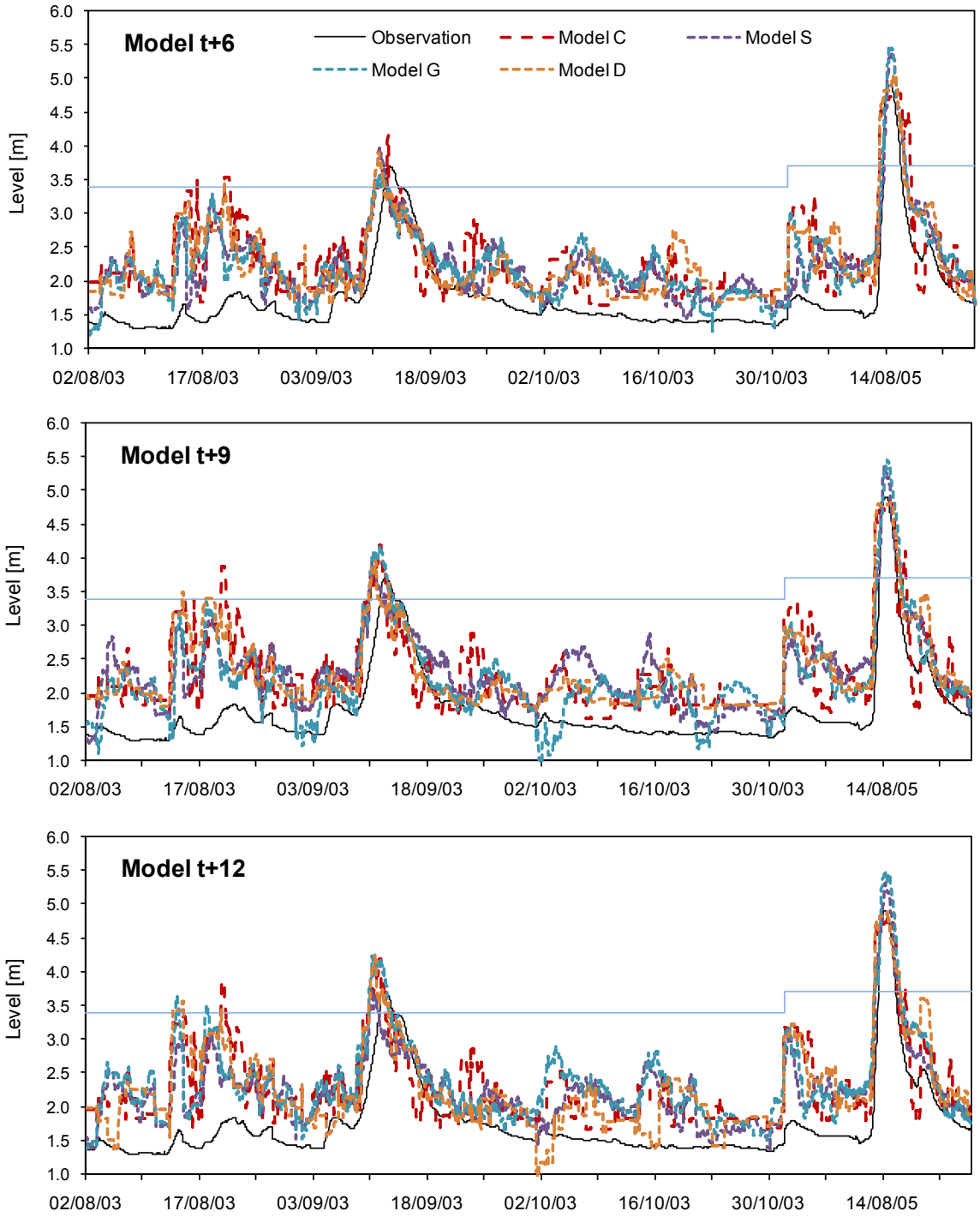


Figure E.3: Comparison of the hydrographs for 4 models at 6, 9 and 12 hour lead time, testing S1 and S2 (2 Aug – 31 Oct 2003 and 2 – 24 Aug 2005) of Experiment 2.

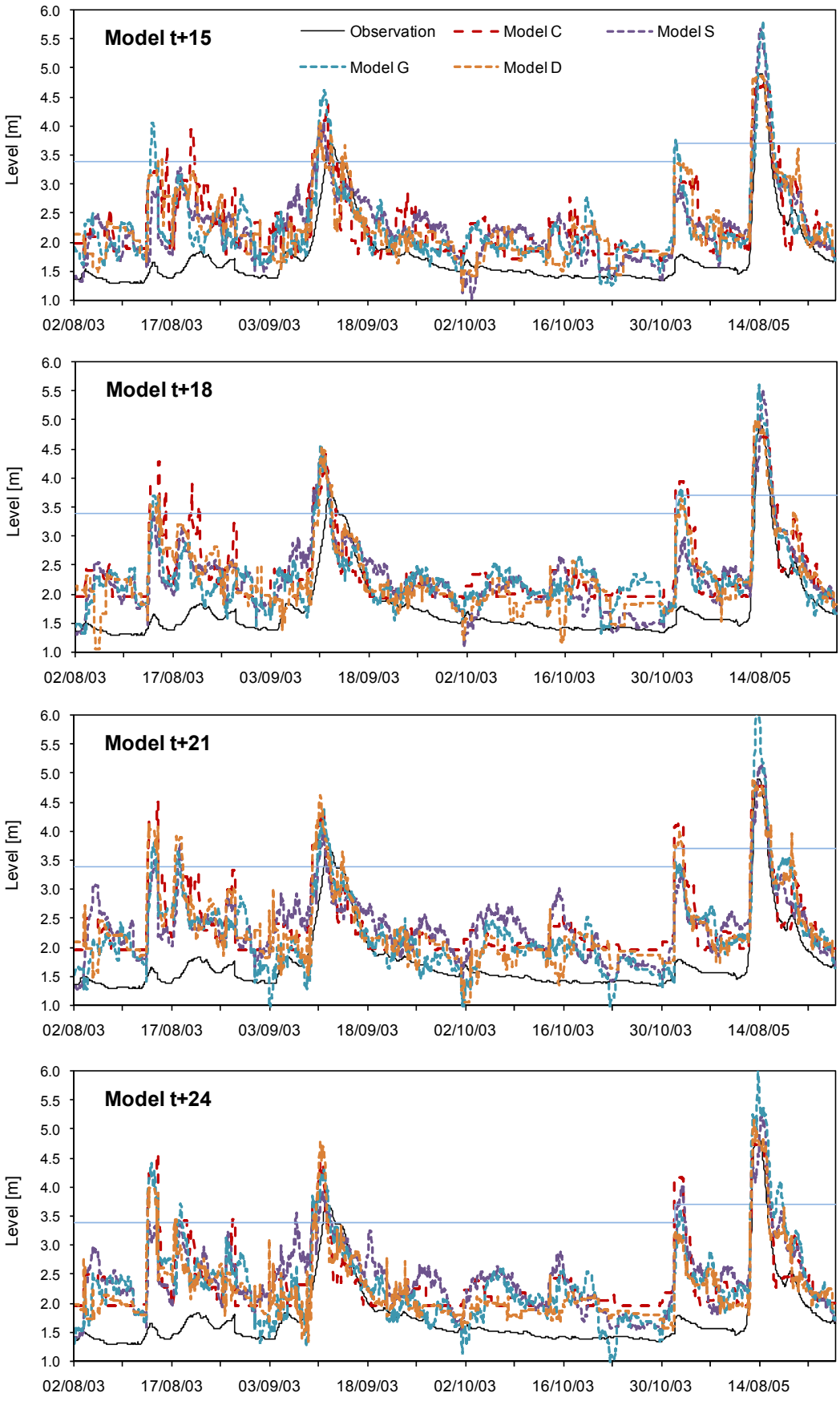


Figure E.4: Comparison of the hydrographs for 4 models at 15, 18, 21 and 24 hour lead time, testing S1 and S2 (2 Aug – 31 Oct 2003 and 2 – 24 Aug 2005) of Experiment 2.

Table E.7: Input variables selected by four input determination techniques (denoted by x) for 6, 9, 12, 15, 18, 21 and 24 hour lead time of Experiment 3.

E3	Correlation							Stepwise Regression							Data Mining							Genetic Algorithm						
	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24
Z11																						x			x			
Z12									x	x	x																x	
Z13																												
12Z11									x																			x
12Z12								x																		x		x
12Z13																												
24Z11								x		x															x			x
24Z12																												x
24Z13																									x			
36Z11																									x		x	x
36Z12																									x			x
36Z13													x	x											x			
48Z11																										x		x
48Z12																										x	x	x
48Z13																												
60Z11																							x			x	x	x
60Z12																						x	x	x		x	x	x
60Z13								x	x																x			x
72Z11																							x			x	x	
72Z12																									x			x
72Z13																							x		x		x	
84Z11																						x			x		x	x
84Z12								x	x	x	x			x									x	x				
84Z13									x	x	x	x	x	x														x
96Z11																												
96Z12								x	x			x	x	x														x
96Z13								x	x				x	x														x
108Z11																							x	x	x			x
108Z12										x																	x	x
108Z13																								x				
120Z11								x	x	x												x	x	x	x			x
120Z12											x	x																
120Z13													x	x														
132Z11																						x	x	x				
132Z12														x								x			x			
132Z13									x	x												x						
144Z11																						x	x					
144Z12													x	x												x		
144Z13								x															x					
156Z11										x																	x	x
156Z12								x		x	x	x																
156Z13																												
168Z11								x	x																			
168Z12								x	x	x	x	x	x	x														
168Z13								x	x	x			x	x														
Z21																					x		x	x				
Z22								x												x	x						x	
Z23																					x			x	x	x	x	
12Z21								x	x	x	x	x	x	x								x						x
12Z22													x									x		x				x
12Z23											x	x	x	x										x				
24Z21																									x	x	x	
24Z22																							x		x	x		x
24Z23								x		x																		x
36Z21																								x		x	x	x
36Z22												x													x	x		x
36Z23																										x		
48Z21										x	x	x									x			x	x		x	x
48Z22																				x	x				x	x	x	x
48Z23																						x		x	x			x
60Z21						x	x	x	x										x	x	x	x		x	x	x	x	x
60Z22											x						x		x	x	x	x	x	x			x	x
60Z23													x	x								x	x			x		x
72Z21				x	x	x	x							x		x			x		x	x		x	x	x	x	x
72Z22															x	x						x	x	x	x	x	x	x
72Z23									x			x		x											x	x		
75Z21		x	x	x	x	x	x						x				x			x				x	x			x
75Z22																									x		x	
75Z23																						x						
78Z21		x	x	x	x	x	x		x						x				x			x	x	x		x	x	x
78Z22																x						x			x	x	x	x
78Z23																								x				x
84Z21		x	x	x	x	x	x				x											x	x	x	x		x	
84Z22																						x	x		x			
84Z23								x		x	x		x										x		x	x		
96Z21	x	x	x	x	x	x	x								x						x	x	x	x	x	x	x	x
96Z22																						x	x	x	x		x	x
96Z23								x	x	x	x	x	x	x									x	x			x	x
108Z21	x	x	x	x	x	x		x							x	x	x	x	x	x			x		x			
108Z22														x								x		x	x	x		
108Z23								x	x	x															x			
120Z21		x						x	x	x	x	x			x	x	x				x	x		x	x		x	x
120Z22													x										x			x		
120Z23																							x					x
132Z21								x															x		x			x
132Z22																										x	x	x
132Z23									x	x				x												x		
144Z21									x			x	x									x	x					
144Z22																						x	x	x		x	x	x
144Z23													x	x														
156Z21											x											x		x		x		x
156Z22																									x	x	x	
156Z23								x		x	x	x																x
168Z21								x	x	x	x	x	x	x								x		x	x			x
168Z22								x	x	x	x	x													x			
168Z23								x	x	x	x		x															
Total	2	4	3	4	4	5	4	26	23	23	19	16	19	20	4	4	3	1	4	5	10	26	27	28	35	28	30	43

Table E.8: Input variables selected by four input determination techniques (denoted by x) for 6, 9, 12, 15, 18, 21 and 24 hour lead time of Experiment 3 (continue).

E3	Correlation							Stepwise Regression							Data Mining							Genetic Algorithm						
	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24
Z31														x							x		x	x	x	x		x
Z32																				x	x							x
Z33																						x		x				
12Z31								x														x						
12Z32									x																x			
12Z33																												
24Z31																									x			
24Z32																										x		x
24Z33																											x	
36Z31								x														x	x	x	x			
36Z32								x																			x	x
36Z33																												
48Z31														x											x			
48Z32								x		x	x	x	x	x								x	x		x	x	x	x
48Z33									x														x	x				x
60Z31												x	x										x	x		x	x	x
60Z32								x	x	x		x	x									x		x		x		
60Z33											x		x	x													x	x
72Z31														x									x				x	
72Z32								x	x																			
72Z33										x	x	x		x										x	x	x	x	
84Z31								x	x	x	x			x											x			x
84Z32								x														x	x		x	x		
84Z33											x		x														x	
96Z31								x														x	x	x				
96Z32											x											x		x	x	x		
96Z33								x	x	x	x	x	x	x														
108Z31																						x			x		x	
108Z32														x								x						x
108Z33														x									x			x		
120Z31																							x					x
120Z32																						x	x		x			
120Z33									x		x	x	x	x											x			
132Z31																									x	x		
132Z32										x															x			
132Z33									x	x				x													x	
144Z31																							x					
144Z32																								x	x			x
144Z33								x					x	x								x						
156Z31																						x		x				
156Z32																						x	x	x				
156Z33										x	x	x										x						
168Z31								x	x	x	x	x	x	x								x						x
168Z32								x	x	x	x	x	x	x										x	x			
168Z33								x	x	x	x	x	x	x														x
Z41																				x	x	x		x			x	x
Z42																					x	x		x	x	x		
Z43																					x	x				x	x	x
12Z41									x		x	x	x	x												x		x
12Z42									x		x	x	x	x											x			x
12Z43								x	x	x															x	x	x	
24Z41								x		x											x			x				x
24Z42								x	x	x		x													x			
24Z43																									x			x
36Z41																						x			x	x		x
36Z42																									x		x	x
36Z43																												x
48Z41																				x		x				x	x	x
48Z42								x		x	x	x												x	x	x		x
48Z43																								x		x		x
60Z41									x			x				x	x	x	x	x	x				x	x		
60Z42									x													x			x	x	x	
60Z43																							x					x
72Z41															x	x	x						x		x	x		
72Z42		x	x	x	x	x	x															x	x	x	x			
72Z43																							x	x			x	
75Z41															x									x	x		x	
75Z42		x	x	x	x	x	x															x	x	x	x			x
75Z43																						x			x			
78Z41																						x			x	x		x
78Z42	x	x	x	x	x	x	x						x									x	x	x	x	x	x	
78Z43																						x			x			
84Z41								x		x	x		x		x	x						x	x	x	x	x		x
84Z42	x	x	x	x	x	x	x	x	x	x	x	x		x			x	x						x	x			
84Z43																								x				
96Z41		x																							x	x		x
96Z42	x	x	x	x	x	x	x								x	x						x	x	x	x	x		x
96Z43																							x		x			x
108Z41								x	x	x												x	x	x		x		
108Z42	x	x																			x	x		x	x	x	x	x
108Z43																							x			x		x
120Z41																							x	x	x	x	x	
120Z42										x	x			x					x	x		x	x	x	x		x	x
120Z43											x	x	x	x												x	x	
132Z41								x	x	x															x	x	x	
132Z42																							x	x		x		
132Z43									x	x													x		x			
144Z41														x	x	x				x		x			x	x		x
144Z42														x			x		x		x	x	x					x
144Z43											x		x														x	
156Z41																						x	x	x	x		x	
156Z42													x		x	x						x	x	x	x	x	x	
156Z43									x	x		x																
168Z41																						x	x	x	x	x	x	
168Z42								x	x	x	x	x	x	x								x	x	x	x		x	
168Z43																											x	
Total	3	5	3	3	3	3	3	23	23	23	21	19	18	24	5	6	4	2	4	5	9	34	33	36	44	33	27	36

Table E.9: The goodness of fit measures for a lead time 6, 9, 12, 15, 18, 21 and 24 hours of Experiment 3.

Model	Statistic	Input Determination Techniques			
		C	S	G	D
t+6	Reduce input	5 (97.2%)	49 (72.7%)	60 (66.6%)	9 (95%)
	PDIFF (m)	-2.1807	0.6316	0.1062	-0.5285
	MAE (m)	0.1228	0.0876	0.071	0.0807
	RMSE (m)	0.4152	0.2677	0.2309	0.2829
	CE	0.8567	0.9405	0.9557	0.9335
	Time delay (hr)	13	14	14	16
t+9	Reduce input	9 (95%)	46 (74.4%)	60 (66.6%)	10 (94.4%)
	PDIFF (m)	-0.0814	0.5494	0.9419	-0.9162
	MAE (m)	0.0728	0.0895	0.0852	0.0822
	RMSE (m)	0.2456	0.2804	0.2791	0.2689
	CE	0.9497	0.9344	0.9351	0.9397
	Time delay (hr)	18	19	19	18.5
t+12	Reduce input	6 (96.6%)	46 (74.4%)	64 (64.4%)	7 (96.1%)
	PDIFF (m)	-0.6631	0.8142	0.1553	-0.4316
	MAE (m)	0.0901	0.0784	0.0575	0.0790
	RMSE (m)	0.2852	0.2601	0.1882	0.2759
	CE	0.9320	0.9434	0.9704	0.9363
	Time delay (hr)	21	22	5	21.5
t+15	Reduce input	7 (96.1%)	40 (77.7%)	3 (98.3%)	79 (56.1%)
	PDIFF (m)	-0.3179	0.1946	0.6263	0.4525
	MAE (m)	0.0866	0.0903	0.0767	0.0974
	RMSE (m)	0.2854	0.2842	0.2297	0.3206
	CE	0.9316	0.9322	0.9557	0.9137
	Time delay (hr)	24	24	3	24
t+18	Reduce input	7 (96.1%)	35 (80.5%)	8 (95.5%)	61 (66.1%)
	PDIFF (m)	0.1836	0.3134	0.1349	0.7101
	MAE (m)	0.0858	0.0914	0.0760	0.1004
	RMSE (m)	0.3015	0.2891	0.2406	0.3451
	CE	0.9234	0.9296	0.9512	0.8997
	Time delay (hr)	27.5	28.5	2	28
t+21	Reduce input	8 (95.5%)	37 (79.4%)	10 (94.4%)	57 (68.3%)
	PDIFF (m)	0.4521	0.4983	0.2096	0.0697
	MAE (m)	0.1054	0.0782	0.0831	0.0828
	RMSE (m)	0.3379	0.2527	0.2502	0.2988
	CE	0.9034	0.9460	0.9471	0.9244
	Time delay (hr)	12	19	5	12
t+24	Reduce input	7 (96.1%)	44 (75.5%)	19 (89.4%)	79 (56.1%)
	PDIFF (m)	0.6759	0.4947	0.6222	-0.1299
	MAE (m)	0.1075	0.0706	0.0935	0.0788
	RMSE (m)	0.3741	0.2307	0.2903	0.2616
	CE	0.8808	0.9547	0.9282	0.9418
	Time delay (hr)	33	11	6	7

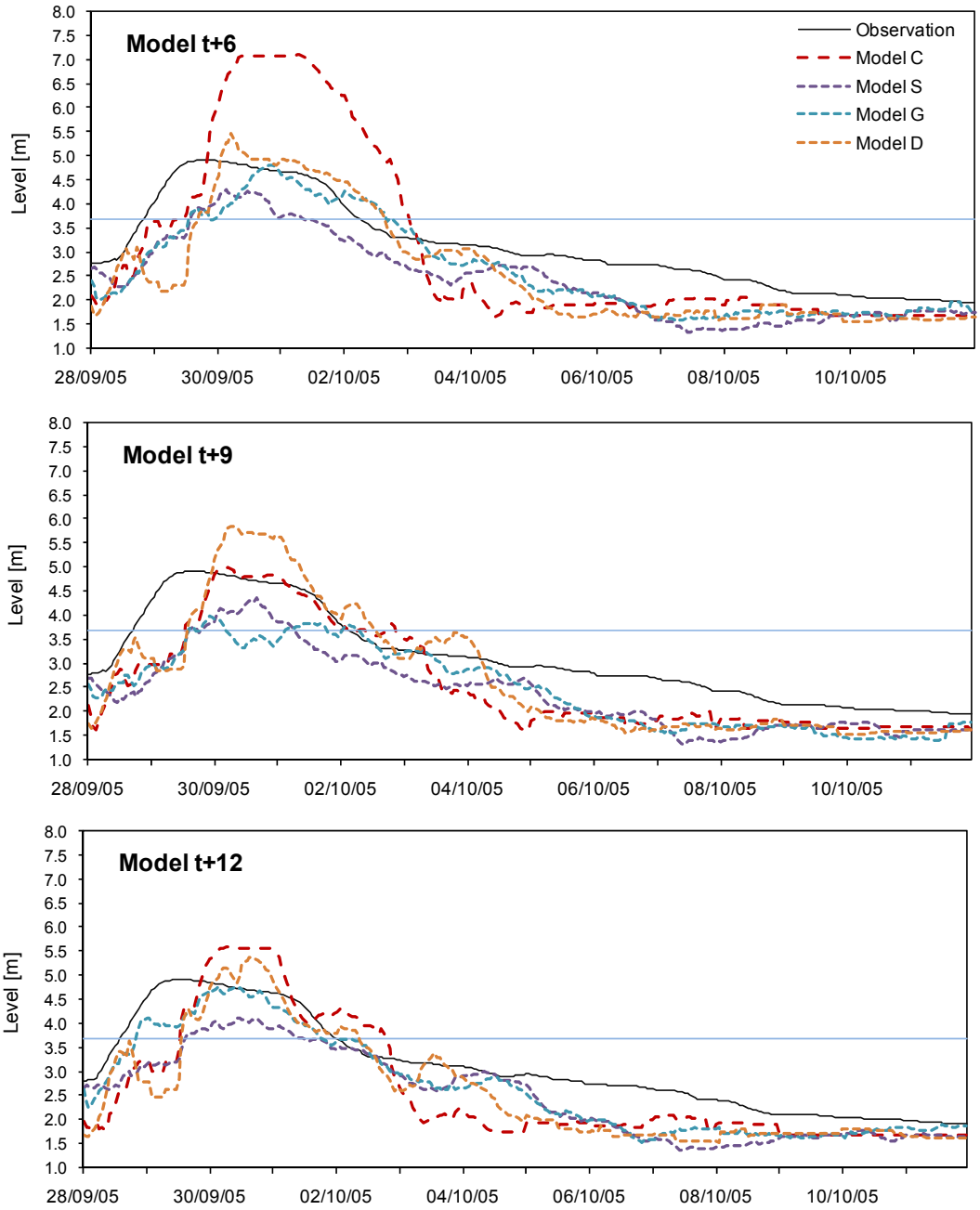


Figure E.5: Comparison of the hydrographs for 4 models at 6, 9 and 12 hour lead time, testing S5 (28 Sep – 11 Oct 2005) of Experiment 3.

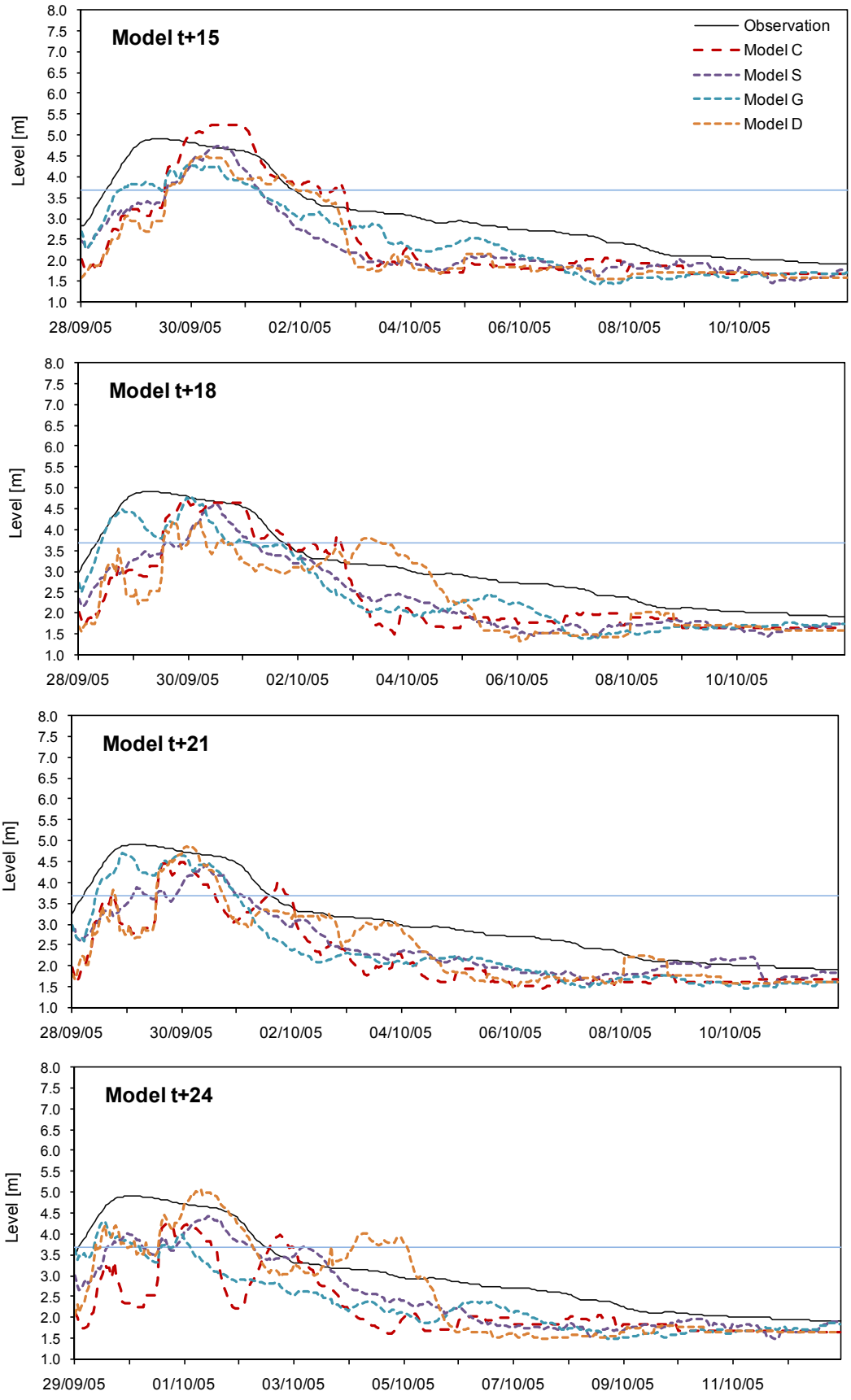


Figure E.6: Comparison of the hydrographs for 4 models at 15, 18, 21 and 24 hour lead time, testing S5 (28 Sep – 11 Oct 2005) of Experiment 3.

Table E.10: Input variables selected by four input determination techniques (denoted by x) for 6, 9, 12, 15, 18, 21 and 24 hour lead time of Experiment 4.

E4	Correlation							Stepwise Regression							Data Mining							Genetic Algorithm						
	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24
Z11									x		x	x										x				x		
Z12									x											x	x						x	
Z13																							x					
12Z11								x															x	x		x	x	
12Z12									x												x							
12Z13											x													x			x	
24Z11											x	x	x	x										x		x	x	
24Z12									x			x	x	x										x				
24Z13								x	x	x		x	x	x									x	x		x	x	
36Z11								x	x	x													x		x	x	x	x
36Z12								x	x	x	x								x				x	x		x	x	x
36Z13								x	x		x	x											x	x		x		
48Z11										x												x						
48Z12										x				x											x	x	x	
48Z13										x			x	x									x	x	x	x	x	x
60Z11								x		x	x											x	x	x	x	x	x	x
60Z12								x			x											x	x		x			
60Z13								x			x			x														x
72Z11																								x			x	
72Z12									x	x		x	x									x		x	x	x		
72Z13								x	x	x	x	x	x									x	x	x				
84Z11								x	x		x		x	x								x	x		x	x		
84Z12																								x			x	
84Z13																							x	x				
96Z11								x		x	x	x	x	x									x			x	x	
96Z12																						x						
96Z13													x	x									x					
108Z11								x	x	x				x									x	x	x	x	x	x
108Z12								x		x			x	x									x	x		x	x	
108Z13								x		x	x											x						
120Z11																												
120Z12													x	x								x						
120Z13												x	x	x								x						
132Z11																												
132Z12								x		x	x											x	x			x	x	
132Z13								x		x	x											x						
144Z11																									x	x		x
144Z12												x	x	x								x						
144Z13												x	x	x														
156Z11																												
156Z12								x	x	x	x		x															
156Z13								x	x	x			x															
168Z11								x	x	x	x	x	x	x														
168Z12											x			x											x		x	
168Z13								x	x		x			x														
Z21								x		x	x		x									x						
Z22																				x	x					x	x	
Z23														x								x		x				
12Z21								x	x																		x	
12Z22									x					x													x	
12Z23								x	x		x	x	x	x								x	x					
24Z21										x																x	x	
24Z22										x	x		x	x								x		x		x	x	x
24Z23										x		x																
36Z21								x	x	x	x	x	x	x										x	x	x		x
36Z22								x														x					x	x
36Z23																						x		x		x	x	
48Z21							x			x			x	x					x	x	x	x				x	x	x
48Z22										x	x		x	x										x	x	x		x
48Z23													x	x								x		x		x		
60Z21					x	x	x	x		x	x			x											x		x	x	x
60Z22						x	x	x		x	x			x								x	x	x	x		x	
60Z23								x		x	x														x		x	
72Z21		x	x	x	x	x	x	x														x	x					
72Z22									x	x		x	x	x								x				x	x	
72Z23																						x		x				x
75Z21		x	x	x	x	x	x																x	x	x	x	x	
75Z22								x			x											x	x	x	x			
75Z23																								x				
78Z21	x	x	x	x	x	x														x		x	x	x		x	x	
78Z22																										x		
78Z23																						x						
84Z21									x		x		x	x								x		x				
84Z22								x	x	x		x											x				x	
84Z23								x	x	x	x	x	x	x								x	x		x		x	x
96Z21								x		x	x	x											x					
96Z22																						x				x		
96Z23								x		x	x	x	x	x												x		
108Z21													x	x	x	x	x	x	x					x		x		
108Z22														x								x		x				
108Z23								x	x	x	x		x															
120Z21													x			x						x		x	x			
120Z22									x				x	x												x		
120Z23									x			x																
132Z21																						x				x		x
132Z22								x		x	x														x	x		
132Z23								x		x				x														
144Z21								x		x	x	x	x	x											x		x	
144Z22									x			x										x	x		x			x
144Z23																							x	x				
156Z21								x	x	x	x											x		x			x	
156Z22																							x		x			
156Z23								x	x	x	x	x	x															x
168Z21												x	x	x								x		x				x
168Z22								x	x	x	x	x	x	x													x	
168Z23																												
Total	-	1	1	1	2	3	4	40	33	39	35	27	36	39	1	2	1	1	3	3	4	35	27	36	18	33	32	19

Table E.11: Input variables selected by four input determination techniques (denoted by x) for 6, 9, 12, 15, 18, 21 and 24 hour lead time of Experiment 4.

E4	Correlation							Stepwise Regression							Data Mining							Genetic Algorithm						
	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24	6	9	12	15	18	21	24
Z31																									x			x
Z32																											x	
Z33																						x	x					
12Z31								x			x												x		x	x		x
12Z32																							x			x	x	x
12Z33								x	x	x	x	x	x	x								x						x
24Z31										x			x												x	x		
24Z32													x															
24Z33													x	x														
36Z31									x					x								x			x		x	x
36Z32														x								x				x		
36Z33								x	x	x	x	x	x	x											x		x	x
48Z31																									x	x		x
48Z32																									x			
48Z33								x	x	x	x	x		x								x					x	
60Z31											x		x	x												x	x	
60Z32														x								x					x	
60Z33									x				x	x								x						
72Z31								x	x	x	x	x											x		x			x
72Z32									x	x			x	x										x	x			
72Z33								x	x	x	x	x	x									x		x		x		
84Z31								x			x		x															
84Z32								x		x													x	x				
84Z33								x	x	x	x	x	x	x								x	x		x	x	x	x
96Z31								x		x		x	x													x		
96Z32												x																
96Z33								x			x		x	x														
108Z31									x																		x	
108Z32									x					x									x					
108Z33										x				x								x						
120Z31																								x			x	
120Z32									x					x														
120Z33										x	x	x	x	x														
132Z31														x												x	x	x
132Z32								x		x	x																	
132Z33										x	x																	
144Z31											x	x	x	x										x				
144Z32									x					x														
144Z33								x	x			x	x	x											x			
156Z31																						x						
156Z32												x	x									x						
156Z33								x		x	x											x						
168Z31								x	x	x	x	x	x	x														
168Z32								x	x	x	x	x	x	x														
168Z33								x	x	x	x	x	x	x														
Z41								x	x											x	x	x			x	x		x
Z42																									x			
Z43								x															x		x	x	x	x
12Z41												x											x				x	
12Z42								x				x	x	x										x		x		
12Z43										x																x	x	
24Z41																									x	x	x	
24Z42									x	x	x	x	x	x													x	
24Z43								x				x		x														
36Z41														x						x	x		x	x			x	
36Z42								x	x	x	x	x	x	x											x	x	x	
36Z43																											x	
48Z41									x	x	x	x				x	x	x	x	x		x	x				x	x
48Z42						x	x		x	x	x	x	x									x					x	x
48Z43													x									x	x	x	x		x	x
60Z41					x	x	x		x				x	x	x	x	x	x	x			x			x	x		x
60Z42				x	x	x	x													x	x	x				x	x	x
60Z43								x		x	x	x		x									x		x		x	
72Z41	x	x	x	x	x			x	x	x	x	x			x	x	x					x	x	x	x	x		x
72Z42	x	x	x	x	x	x	x					x						x	x			x	x	x				x
72Z43														x										x			x	
75Z41	x	x	x	x											x	x						x	x		x	x		
75Z42	x	x	x	x	x	x	x										x					x		x		x	x	
75Z43										x													x				x	
78Z41	x	x	x	x												x						x	x	x	x		x	
78Z42	x	x	x	x	x	x		x	x					x		x						x			x			
78Z43																						x	x		x	x	x	
84Z41	x	x	x																			x	x	x	x	x		x
84Z42	x	x	x	x							x		x	x	x				x						x	x		
84Z43								x		x	x		x	x									x				x	
96Z41	x																						x	x		x	x	x
96Z42	x							x		x	x	x	x		x	x	x				x	x	x					
96Z43								x														x	x					
108Z41								x	x	x	x		x									x	x	x	x			
108Z42									x			x		x						x		x	x	x	x	x		x
108Z43										x		x	x										x					
120Z41														x								x	x				x	
120Z42									x		x													x				
120Z43										x		x	x									x		x	x	x		x
132Z41																							x					
132Z42								x		x				x														
132Z43																												
144Z41									x													x		x				
144Z42								x	x		x	x	x	x										x			x	
144Z43																												
156Z41																						x	x					
156Z42													x										x			x	x	x
156Z43								x	x	x	x	x																
168Z41								x	x	x	x	x	x	x								x			x			
168Z42								x	x	x	x	x	x	x								x		x				
168Z43										x	x	x																
Total	6	4	4	4	4	4	4	33	32	35	35	33	39	36	4	4	4	3	4	5	4	32	27	24	25	28	29	22

Table E.12: The goodness of fit measures for a lead time 6, 9, 12, 15, 18, 21 and 24 hours of Experiment 4.

Model	Statistic	Input Determination Techniques							
		S1				S5			
		C	S	G	D	C	S	G	D
t+6	Reduce inputs	6	73	67	5	(96.6%)	(59.4%)	(46.1%)	(97.2%)
	PDIFF (m)	-0.4713	-0.2113	-0.0052	-0.0961	-1.8859	0.4232	0.8615	-1.2742
	MAE (m)	0.4448	0.3045	0.3961	0.4553	0.2377	0.1824	0.2586	0.2373
	RMSE (m)	0.5872	0.3937	0.4645	0.6105	0.5061	0.3363	0.5116	0.4665
	CE	0.1923	0.6369	0.4946	0.1271	0.8969	0.9545	0.8946	0.9124
	Time delay (hr)	/	/	/	/	19	8	14	14.5
t+9	Reduce inputs	5	65	54	6	(97.2%)	(63.8%)	(70%)	(96.6%)
	PDIFF (m)	0.0564	-0.0312	-0.2589	-0.2693	-0.4528	0.6108	0.8077	-1.3528
	MAE (m)	0.4298	0.2900	0.3571	0.4758	0.1983	0.2048	0.2136	0.2222
	RMSE (m)	0.5725	0.3811	0.4222	0.6077	0.4019	0.3851	0.4222	0.4531
	CE	0.2323	0.6598	0.5825	0.1349	0.9348	0.9401	0.928	0.9171
	Time delay (hr)	/	/	/	/	24	3	7	24
t+12	Reduce inputs	51	74	60	5	(71.6%)	(58.8%)	(66.6%)	(97.2%)
	PDIFF (m)	-0.1589	0.0348	-0.1823	-0.1397	0.4181	0.4789	0.2306	-0.0617
	MAE (m)	0.4123	0.3423	0.2643	0.4564	0.2146	0.2182	0.1567	0.1920
	RMSE (m)	0.5512	0.4312	0.3352	0.6073	0.4270	0.4189	0.3035	0.3977
	CE	0.2881	0.5644	0.7368	0.1357	0.9264	0.9292	0.9628	0.9362
	Time delay (hr)	/	/	/	/	24	18	15	18
t+15	Reduce inputs	5	70	43	4	(97.2%)	(61.1%)	(76.1%)	(97.7%)
	PDIFF (m)	-0.3278	-0.4969	-0.3295	-0.1471	0.6775	0.1734	0.4148	-0.2939
	MAE (m)	0.4199	0.3472	0.3088	0.3956	0.222	0.1689	0.2024	0.1987
	RMSE (m)	0.5406	0.4976	0.3798	0.5244	0.4401	0.3271	0.3981	0.4001
	CE	0.3151	0.4197	0.6620	0.3554	0.9214	0.9566	0.9357	0.9350
	Time delay (hr)	/	/	/	/	28	5	5	24
t+18	Reduce inputs	6	60	61	7	(96.6%)	(66.6%)	(66.1%)	(96.1%)
	PDIFF (m)	-1.0656	-0.5283	0.0634	-0.5216	0.9394	0.1798	0.3887	-0.1226
	MAE (m)	0.4509	0.4193	0.3205	0.4065	0.2269	0.1449	0.2743	0.1896
	RMSE (m)	0.5548	0.5557	0.3973	0.5654	0.4551	0.2869	0.5619	0.4376
	CE	0.2785	0.2761	0.6300	0.2506	0.9157	0.9665	0.8714	0.9220
	Time delay (hr)	/	/	/	/	4	4	4	28
t+21	Reduce inputs	7	75	61	8	(96.1%)	(58.3%)	(66.1%)	(95.5%)
	PDIFF (m)	-0.7841	-0.3706	-0.0824	-0.5634	-0.1486	0.6263	-0.1157	0.1512
	MAE (m)	0.3383	0.4114	0.4524	0.3732	0.2032	0.2247	0.1177	0.1758
	RMSE (m)	0.4414	0.5399	0.5470	0.4717	0.4086	0.4313	0.2411	0.3702
	CE	0.5431	0.3165	0.2983	0.4783	0.9317	0.9239	0.9762	0.944
	Time delay (hr)	/	/	/	/	6	15	4	6
t+24	Reduce inputs	8	75	41	8	(95.5%)	(58.3%)	(77.2%)	(95.5%)
	PDIFF (m)	-0.9861	0.128	-0.1944	-0.3879	-0.2157	-0.3681	0.7968	0.3342
	MAE (m)	0.3688	0.3402	0.3132	0.3725	0.1973	0.2009	0.2167	0.1573
	RMSE (m)	0.4841	0.4245	0.3950	0.4362	0.3971	0.4096	0.4913	0.3084
	CE	0.4502	0.5774	0.6341	0.5537	0.9351	0.9310	0.9007	0.9609
	Time delay (hr)	/	/	/	/	8	-1	9	8.5

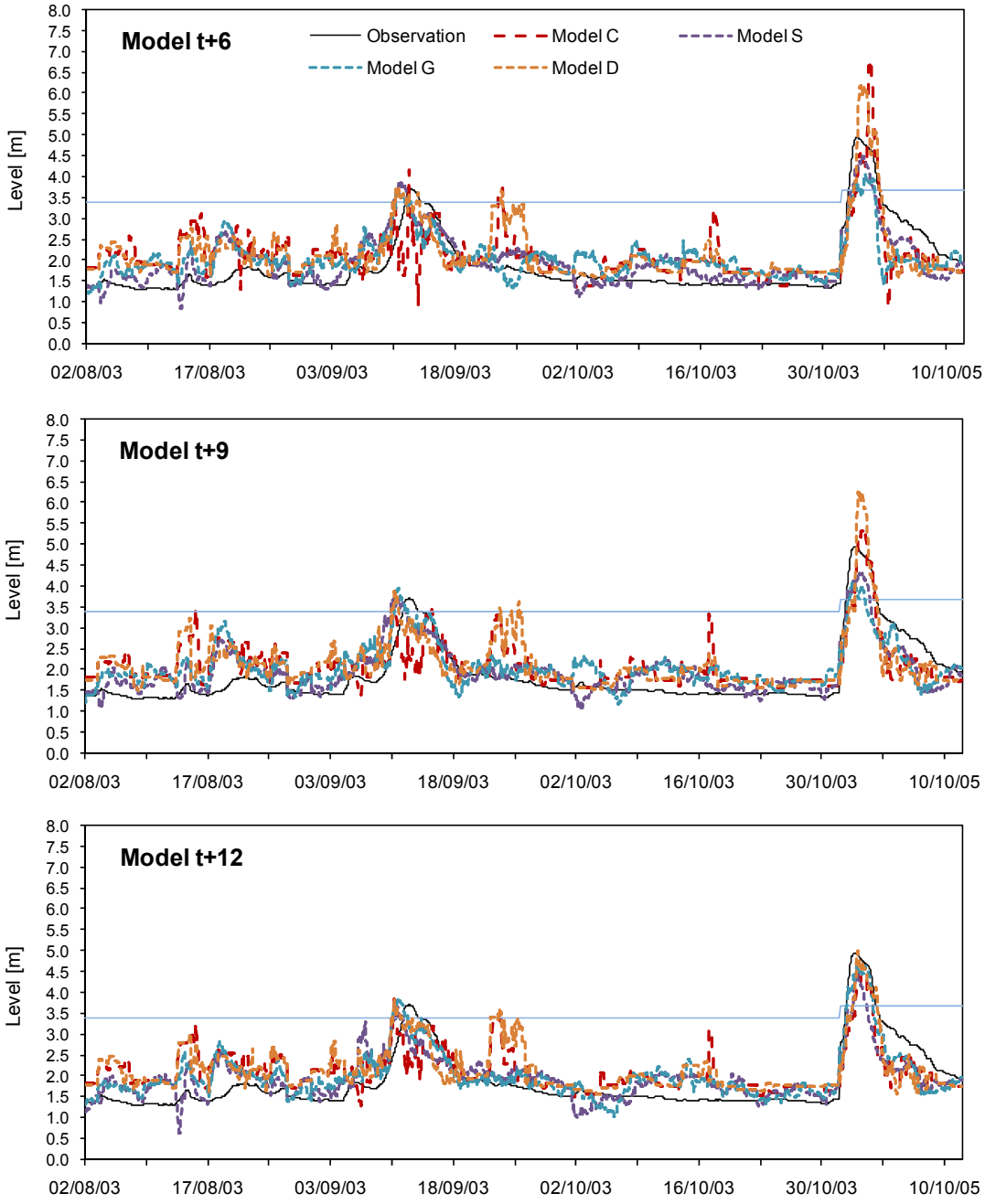


Figure E.7: Comparison of the hydrographs for 4 models at 6, 9 and 12 hour lead time, testing S1 and S5 (2 Aug – 31 Oct 2003 and 28 Sep – 11 Oct 2005) of Experiment 4.

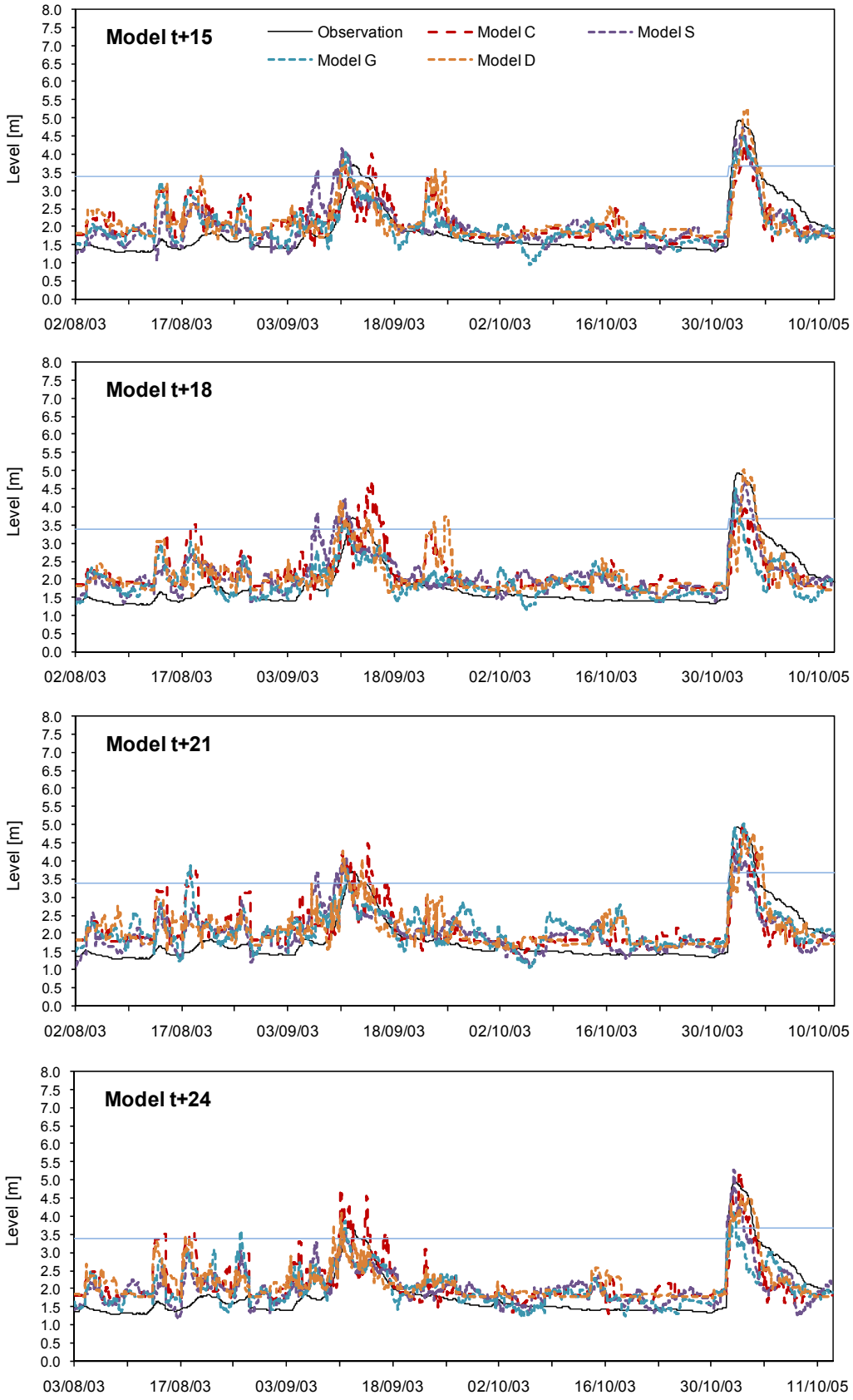


Figure E.8: Comparison of the hydrographs for 4 models at 15, 18, 21 and 24 hour lead time, testing S1 and S5 (2 Aug – 31 Oct 2003 and 28 Sep – 11 Oct 2005) of Experiment 4.

Table E.13: The goodness of fit measures for a lead time 24, 30, 36, 42 and 48 hours of different time correspondences, test S2.

Model	Statistic	Time Correspondences (hr)					
		30	**36**	**42**	**48**	**54**	**60**
t+24	PDIFF (m)	-0.0010	-0.3980	-0.4170	-0.4239	/	/
	MAE (m)	0.6653	0.6707	0.6853	0.6874	/	/
	RMSE (m)	0.8212	0.8113	0.8221	0.8257	/	/
	CE	0.1543	0.1747	0.1526	0.1451	/	/
	Time delay (hr)	-1	5	10	16	/	/
t+30	PDIFF (m)	-0.0444	-0.3397	-0.4992	-0.3956	/	/
	MAE (m)	0.6556	0.6555	0.6683	0.6843	/	/
	RMSE (m)	0.8104	0.7848	0.7965	0.8218	/	/
	CE	0.1775	0.2285	0.2053	0.1540	/	/
	Time delay (hr)	-1	5	10.5	16.5	/	/
t+36	PDIFF (m)	/	-0.4610	-0.4520	-0.4020	-0.3560	/
	MAE (m)	/	0.6523	0.6515	0.6701	0.7005	/
	RMSE (m)	/	0.7819	0.7699	0.8009	0.8415	/
	CE	/	0.2346	0.2580	0.1971	0.1136	/
	Time delay (hr)	/	5	10.5	16.5	23	/
t+42	PDIFF (m)	/	/	-0.4370	-0.4740	-0.2200	-0.0852
	MAE (m)	/	/	0.6486	0.6575	0.6841	0.7080
	RMSE (m)	/	/	0.7663	0.7803	0.8227	0.8594
	CE	/	/	0.2651	0.2381	0.153	0.0757
	Time delay (hr)	/	/	10.5	16.5	23	29.5
t+48	PDIFF (m)	/	/	/	-0.5290	-0.337	-0.3009
	MAE (m)	/	/	/	0.6495	0.6697	0.6955
	RMSE (m)	/	/	/	0.7687	0.7983	0.8433
	CE	/	/	/	0.2602	0.2022	0.1098
	Time delay (hr)	/	/	/	16.5	23	29.5

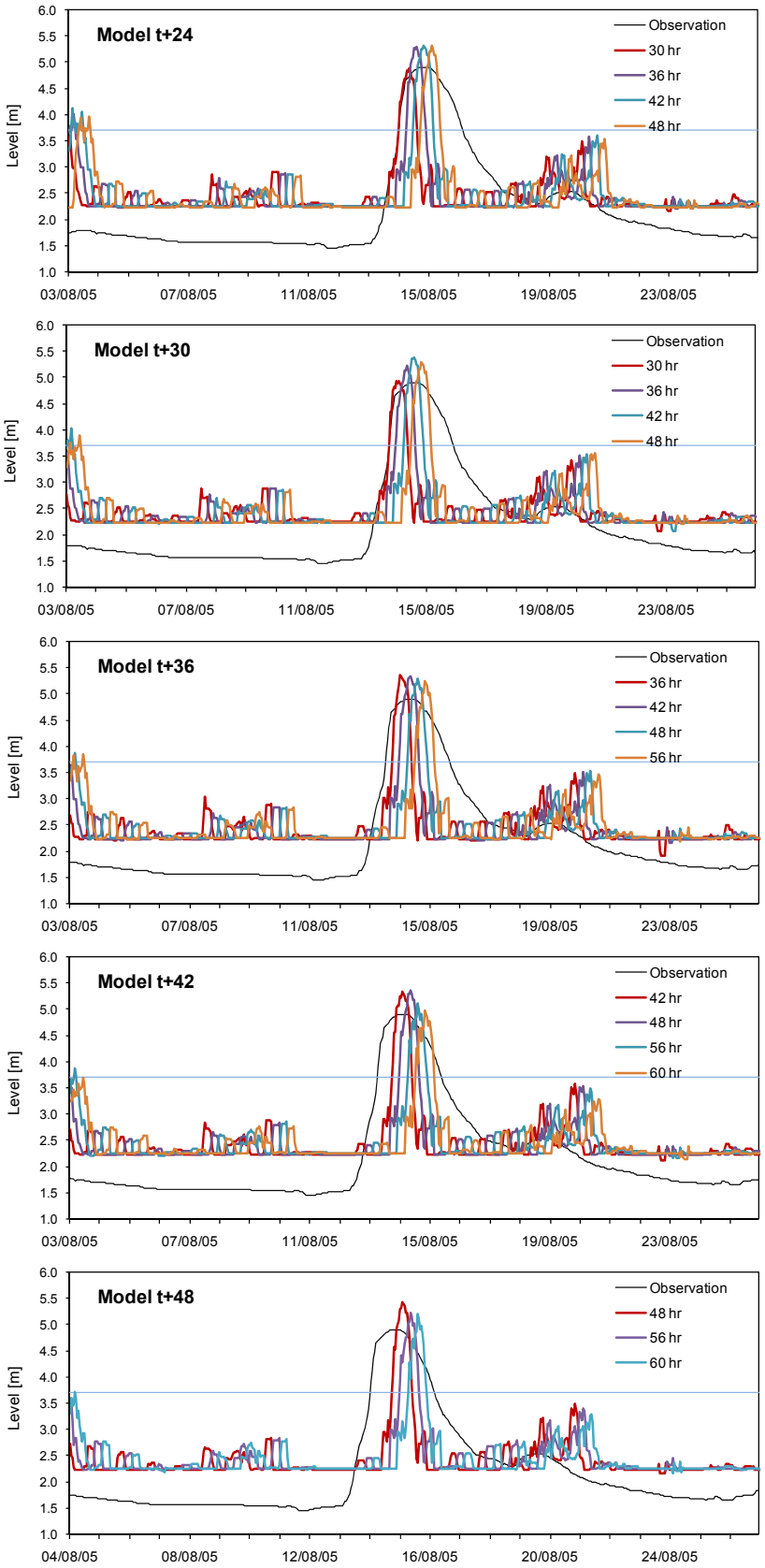


Figure E.9: Compare of the hydrographs of different training data at 24, 30, 36, 42 and 48 hour lead time, test S2 (3 – 27 Aug 2005).

Table E.14: The goodness of fit measures for a lead time 24, 30, 36, 42 and 48 hours of different time correspondences, test S1.

Model	Statistic	Time Correspondences (hr)					
		30	**36**	**42**	**48**	**54**	**60**
t+24	PDIFF (m)	-0.6100	-0.7240	-0.6430	-0.4930	/	/
	MAE (m)	0.4403	0.4233	0.4051	0.3876	/	/
	RMSE (m)	0.5834	0.5622	0.5432	0.5214	/	/
	CE	0.7304	0.7496	0.7662	0.7847	/	/
	Time delay (hr)	-17	-11	-5	0.5	/	/
t+30	PDIFF (m)	-0.6320	-0.686	-0.5640	-0.3640	/	/
	MAE (m)	0.4420	0.4310	0.4105	0.3878	/	/
	RMSE (m)	0.5843	0.5736	0.5474	0.5206	/	/
	CE	0.7295	0.7394	0.7627	0.7853	/	/
	Time delay (hr)	-17	-10	-5	0.5	/	/
t+36	PDIFF (m)	/	-0.6600	-0.7650	-0.5460	-0.3470	/
	MAE (m)	/	0.4369	0.4205	0.402	0.3924	/
	RMSE (m)	/	0.5772	0.5595	0.5339	0.5216	/
	CE	/	0.7361	0.7521	0.7742	0.7846	/
	Time delay (hr)	/	-10	-5	0	7	/
t+42	PDIFF (m)	/	/	-0.7720	-0.2540	-0.3090	-0.2009
	MAE (m)	/	/	0.4228	0.4024	0.3971	0.3956
	RMSE (m)	/	/	0.5600	0.5357	0.5254	0.5251
	CE	/	/	0.7516	0.7727	0.7814	0.7816
	Time delay (hr)	/	/	-4	1	6.5	12.5
t+48	PDIFF (m)	/	/	/	-0.3570	-0.2750	-0.2918
	MAE (m)	/	/	/	0.4052	0.4012	0.4008
	RMSE (m)	/	/	/	0.5418	0.5335	0.5301
	CE	/	/	/	0.7675	0.7746	0.7775
	Time delay (hr)	/	/	/	1	7	12

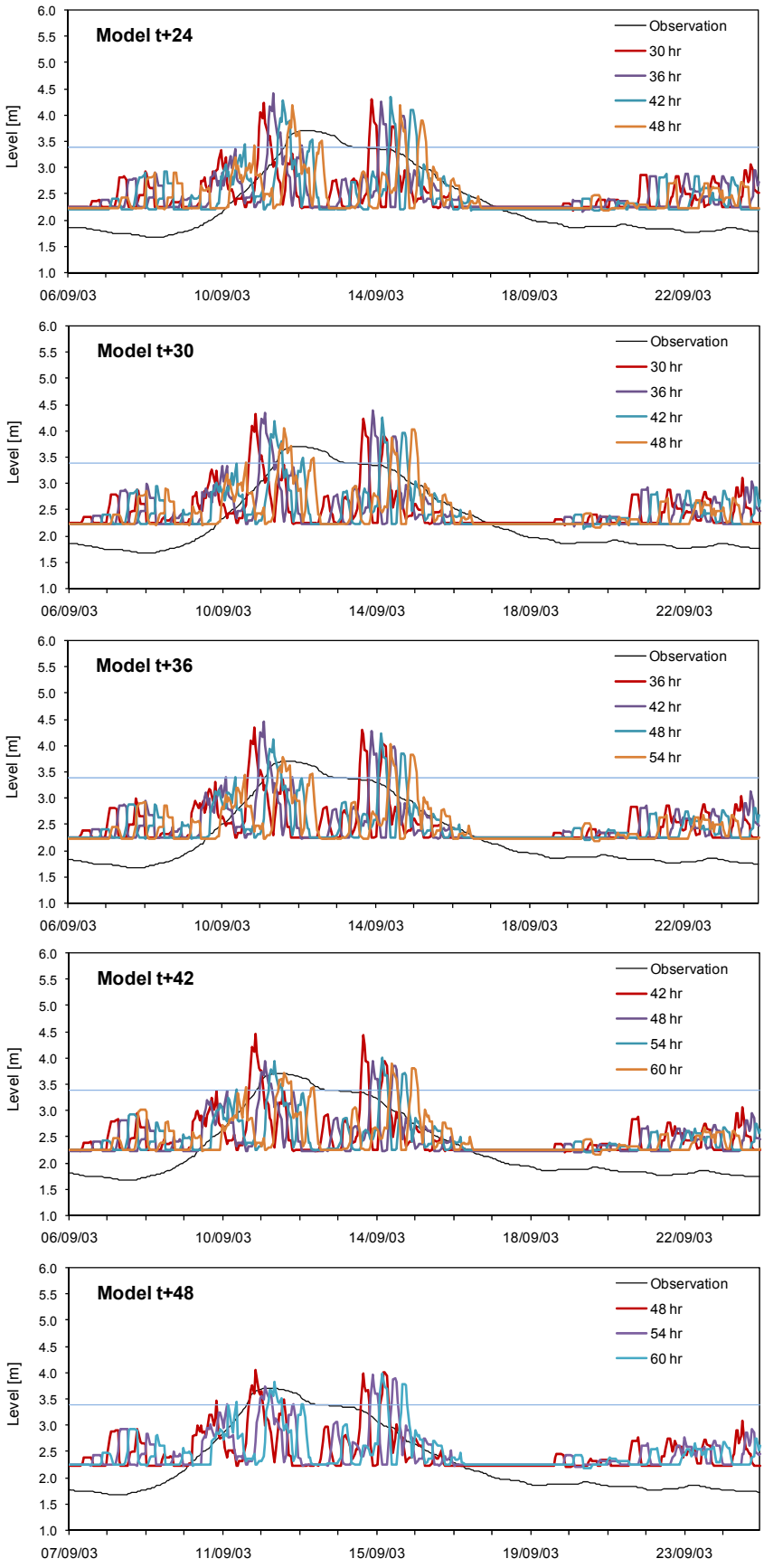


Figure E.10: Compare of the hydrographs of different training data at 24, 30, 36, 42 and 48 hour lead time, test S1 (6 – 23 Sep 2003).

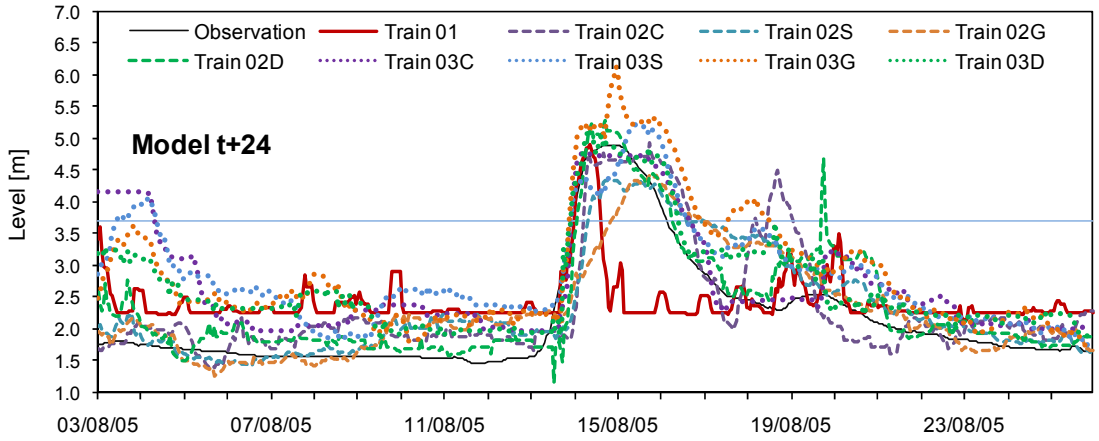


Figure E.11: Compare of the hydrographs of different training data at 24 hour lead time, test S2 (12 – 20 Aug 2005) Note: Train 01 (S1, S3 – S8), train 02 (S3 – S8) and train 03 (S3-S5).

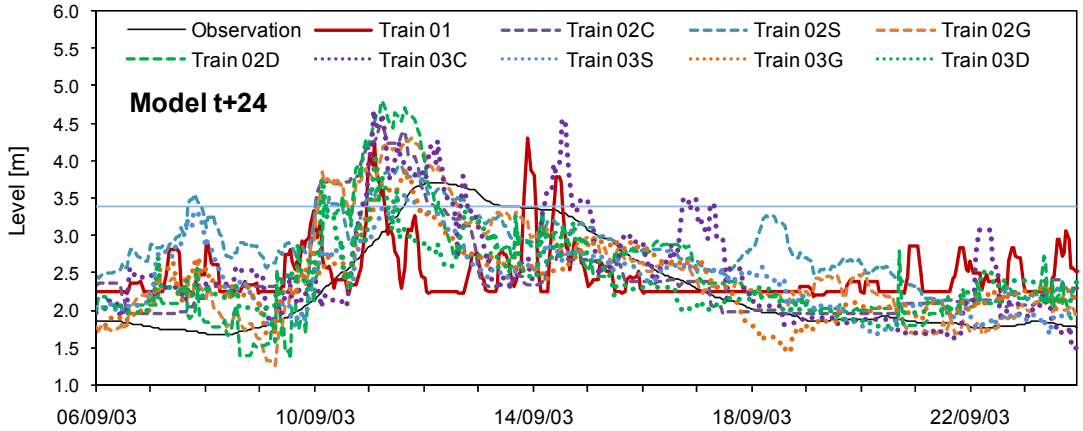


Figure E.12: Compare of the hydrographs of different training data at 24 hour lead time, test S1 (2003) Note: Train 01 (S3 – S8), train 02 (S2 - S4, S6 – S8) and train 03 (S3-S5).

Table E.15: The goodness of fit measures for a lead time 24, 30, 36, 42 and 48 hours of different time correspondences, test S7 and S8.

Model	Statistic	Time Correspondences (hr)					
		30	36	42	48	54	60
t+24	PDIFF (m)	0.8905	0.8665	0.8343	0.8435	/	/
	MAE (m)	0.3928	0.3835	0.3771	0.3779	/	/
	RMSE (m)	0.4691	0.4671	0.4650	0.4639	/	/
	CE	0.0206	0.0289	0.0373	0.0423	/	/
t+30	PDIFF (m)	0.9305	0.8911	0.8924	0.8408	/	/
	MAE (m)	0.3953	0.3854	0.3804	0.3793	/	/
	RMSE (m)	0.4689	0.4668	0.4645	0.4614	/	/
	CE	0.0167	0.0256	0.0353	0.0482	/	/
t+36	PDIFF (m)	/	0.8858	0.8899	0.8795	0.8675	/
	MAE (m)	/	0.3886	0.3823	0.3811	0.3803	/
	RMSE (m)	/	0.4679	0.4655	0.4620	0.4604	/
	CE	/	0.0198	0.0298	0.0444	0.0510	/
t+42	PDIFF (m)	/	/	0.8893	0.9019	0.9087	0.9356
	MAE (m)	/	/	0.3844	0.3835	0.3844	0.3891
	RMSE (m)	/	/	0.4660	0.4642	0.4621	0.4638
	CE	/	/	0.0306	0.038	0.0470	0.0397
t+48	PDIFF (m)	/	/	/	0.9323	0.9292	0.8830
	MAE (m)	/	/	/	0.3899	0.3879	0.3870
	RMSE (m)	/	/	/	0.4662	0.4646	0.4616
	CE	/	/	/	0.0347	0.0416	0.0536

Table E.16: The goodness of fit measures for a lead time 24 and 36 hours with radar only and radar and rain gauging station, test S2.

Model	Statistic	30 hr		36 hr		42 hr	
		No rain	Rain	No rain	Rain	No rain	Rain
t+24	PDIFF (m)	-0.0014	-0.0734	-0.3981	-0.3619	/	/
	MAE (m)	0.5115	0.5159	0.5156	0.5226	/	/
	RMSE (m)	0.7201	0.7202	0.7113	0.7186	/	/
	CE	0.6410	0.6409	0.6497	0.6425	/	/
t+36	PDIFF (m)	/	/	-0.4607	-0.6024	-0.4519	-0.6342
	MAE (m)	/	/	0.5015	0.5085	0.5009	0.5152
	RMSE (m)	/	/	0.6856	0.6865	0.6751	0.6872
	CE	/	/	0.6745	0.6737	0.6845	0.6730

Table E.17: The goodness of fit measures for a lead time 18, 24 and 30 hours with three experiments, test S2 and S3.

Model	Statistic	S2			S3		
		Ex1	Ex2	Ex3	Ex1	Ex2	Ex3
t+18	PDIFF (m)	-0.0300	-0.9760	-0.1030	-0.4490	-0.5240	-0.400
	MAE (m)	0.2014	0.1911	0.2001	0.1058	0.1372	0.100
	RMSE (m)	0.2669	0.3128	0.2522	0.1708	0.2119	0.1700
	CE	0.9106	0.8772	0.9202	0.8947	0.8379	0.900
	Time delay (hr)	2	1	1	-17	-16	-16
t+24	PDIFF (m)	0.1400	-0.6880	0.0365	-1.049	-0.7360	-0.2740
	MAE (m)	0.3387	0.2833	0.2709	0.1694	0.1944	0.1290
	RMSE (m)	0.4399	0.4503	0.3600	0.2776	0.3079	0.1960
	CE	0.7573	0.7457	0.8375	0.7187	0.6540	0.8590
	Time delay (hr)	6.2	6	6.3	-14.5	-14	-12
t+30	PDIFF (m)	-0.2300	0.1613	-0.5892	-1.102	-1.396	-1.1127
	MAE (m)	0.4644	0.3714	0.4354	0.2256	0.2420	0.2279
	RMSE (m)	0.6104	0.5981	0.5788	0.3519	0.3960	0.3505
	CE	0.5332	0.5519	0.5804	0.5429	0.4214	0.5467
	Time delay (hr)	11.5	12	11	-13	-15	-14.5

Table E.18: The goodness of fit measures for a lead time 18, 24 and 30 hours with three experiments, test S4 and S5.

Model	Statistic	S4			S5		
		Ex1	Ex2	Ex3	Ex1	Ex2	Ex3
t+18	PDIFF (m)	-0.5310	-0.3600	-0.400	-0.3530	1.0728	-0.100
	MAE (m)	0.0917	0.0832	0.0900	0.1161	0.3074	0.1200
	RMSE (m)	0.2163	0.2199	0.2200	0.2501	0.5611	0.2800
	CE	0.9829	0.9824	0.9800	0.9277	0.6362	0.9100
	Time delay (hr)	9.5	10	9	9	21	13
t+24	PDIFF (m)	-1.092	0.0429	-1.000	-0.4510	1.0731	-0.0490
	MAE (m)	0.1335	0.1182	0.1309	0.1764	0.3414	0.1475
	RMSE (m)	0.2986	0.3108	0.2946	0.4305	0.6100	0.3381
	CE	0.9678	0.9651	0.9687	0.7874	0.5731	0.8688
	Time delay (hr)	13	14	13	17	24	15
t+30	PDIFF (m)	-0.3450	-0.067	-0.4833	-0.4370	0.7581	0.1452
	MAE (m)	0.1606	0.1727	0.1555	0.2175	0.3749	0.1804
	RMSE (m)	0.3670	0.4079	0.3429	0.4758	0.7098	0.4149
	CE	0.9518	0.9405	0.9579	0.7421	0.4261	0.8039
	Time delay (hr)	18.5	22	18	18.5	33	18

Table E.19: The goodness of fit measures for a lead time 18, 24 and 30 hours with three experiments, test S6 and S7, S8.

Model	Statistic	S6			S7, S8		
		Ex1	Ex2	Ex3	Ex1	Ex2	Ex3
t+18	PDIFF (m)	-0.1700	-0.2880	-0.200	-0.0290	-0.1450	-0.200
	MAE (m)	0.0776	0.0772	0.0800	0.0740	0.0827	0.0700
	RMSE (m)	0.1282	0.1336	0.1300	0.0985	0.1206	0.1000
	CE	0.9756	0.9735	0.9800	0.9571	0.9356	0.9600
	Time delay (hr)	-3	-1	-3	-	-	-
t+24	PDIFF (m)	-0.477	-0.5014	-0.608	-0.052	-0.4550	-0.1200
	MAE (m)	0.1261	0.1252	0.1274	0.1092	0.0956	0.1041
	RMSE (m)	0.2136	0.2332	0.2231	0.1472	0.1462	0.1412
	CE	0.9323	0.9193	0.9261	0.9035	0.9049	0.9113
	Time delay (hr)	-10.5	-19	-10	-	-	-
t+30	PDIFF (m)	-0.6520	-1.031	-0.6208	0.1110	0.1331	-0.1319
	MAE (m)	0.1594	0.1525	0.1639	0.1486	0.1555	0.1534
	RMSE (m)	0.2778	0.2733	0.2911	0.2093	0.2470	0.2206
	CE	0.8856	0.8893	0.8744	0.8042	0.7272	0.7824
	Time delay (hr)	-10	-12	-10	-	-	-

Note

Experiment 1: using radar, three flow stations (P1, P67 and P75) and rain gauging station
Experiment 2: using radar, five flow stations (P1, P67, P75, P21 and P4a) and rain gauging station
Experiment 3: using same input as experiment 1 but selected with Stepwise regression

Table E.20: The goodness of fit measures for a lead time 18, 24 and 30 hours with different number of flow stations, test S2 and S3.

Model	Statistic	S2			S3		
		5	4	3	5	4	3
t+18	PDIFF (m)	-1.3953	-0.9297	-0.8977	-0.3211	-0.7278	-0.5990
	MAE (m)	0.2122	0.1773	0.1678	0.1383	0.1261	0.1160
	RMSE (m)	0.3041	0.2671	0.2607	0.2136	0.2086	0.1937
	CE	0.8839	0.9105	0.9147	0.8353	0.8430	0.8646
	Time delay (hr)	1	0	1	-17	-18	-17
t+24	PDIFF (m)	-0.9252	-0.367	-0.4634	-0.4247	-0.8097	-0.9151
	MAE (m)	0.2619	0.2558	0.2765	0.2257	0.1761	0.1888
	RMSE (m)	0.4545	0.3877	0.4203	0.3467	0.2799	0.3093
	CE	0.7410	0.8115	0.7785	0.5611	0.7139	0.6507
	Time delay (hr)	4	7	6	-12	-12	-15
t+30	PDIFF (m)	-0.2652	-0.6505	-0.0943	-0.8363	-0.9710	-1.2312
	MAE (m)	0.3568	0.3014	0.3157	0.2414	0.2259	0.2415
	RMSE (m)	0.5937	0.4891	0.5362	0.3948	0.3738	0.3966
	CE	0.5585	0.7003	0.6399	0.4248	0.4843	0.4196
	Time delay (hr)	11.5	10	11.5	-15.5	-16	-16

Table E.21: The goodness of fit measures for a lead time 18, 24 and 30 hours with different number of flow stations, test S6.

Model	Statistic	S6		
		5	4	3
t+18	PDIFF (m)	-0.2321	-0.3829	-0.2615
	MAE (m)	0.0861	0.0844	0.0873
	RMSE (m)	0.1450	0.1297	0.1356
	CE	0.9687	0.9750	0.9727
	Time delay (hr)	-1	-1	5
t+24	PDIFF (m)	-0.4945	-0.7896	-0.5976
	MAE (m)	0.1465	0.1329	0.1314
	RMSE (m)	0.2478	0.2241	0.2197
	CE	0.9089	0.9254	0.9283
	Time delay (hr)	-20	-9	-9
t+30	PDIFF (m)	-0.7614	-0.6547	-0.4991
	MAE (m)	0.1706	0.1696	0.1688
	RMSE (m)	0.2920	0.2803	0.2696
	CE	0.8736	0.8835	0.8922
	Time delay (hr)	-10.5	-10	-10

Table E.22: The goodness of fit measures for a lead time 18, 24 and 30 hours with different number of flow stations, test S4 and S5.

Model	Statistic	S4			S5		
		5	4	3	5	4	3
t+18	PDIFF (m)	1.2743	1.2624	1.2851	0.9533	-0.4952	-0.4294
	MAE (m)	0.1193	-0.2872	-0.3153	0.3318	0.1896	0.1405
	RMSE (m)	0.2641	0.2328	0.2363	0.5607	0.3095	0.2753
	CE	0.9746	0.9802	0.9796	0.6366	0.8893	0.9124
	Time delay (hr)	13	11	10	19	6	10
t+24	PDIFF (m)	1.8620	1.7719	1.7313	0.9483	-0.2637	-0.1430
	MAE (m)	0.7087	0.1548	-0.5462	0.3929	0.2459	0.1832
	RMSE (m)	0.4361	0.3290	0.3282	0.6660	0.4329	0.3918
	CE	0.9313	0.9609	0.9611	0.4911	0.7849	0.8239
	Time delay (hr)	19	14	13.5	22	18	14
t+30	PDIFF (m)	2.0978	1.984	1.9224	0.5132	-0.6447	-0.3749
	MAE (m)	0.6014	0.4215	-0.2955	0.3827	0.2515	0.1995
	RMSE (m)	0.4578	0.3865	0.3785	0.6837	0.4640	0.4286
	CE	0.9250	0.9466	0.9488	0.4674	0.7547	0.7907
	Time delay (hr)	22	20	19	28	20	19

Note

5: using radar, rain gauging station and five flow stations (P1, P67, P75, P4a and P21)
4: using radar, rain gauging station and four flow stations (P1, P67, P75 and P4a)
3: using radar, rain gauging station and three flow stations (P1, P67 and P75)

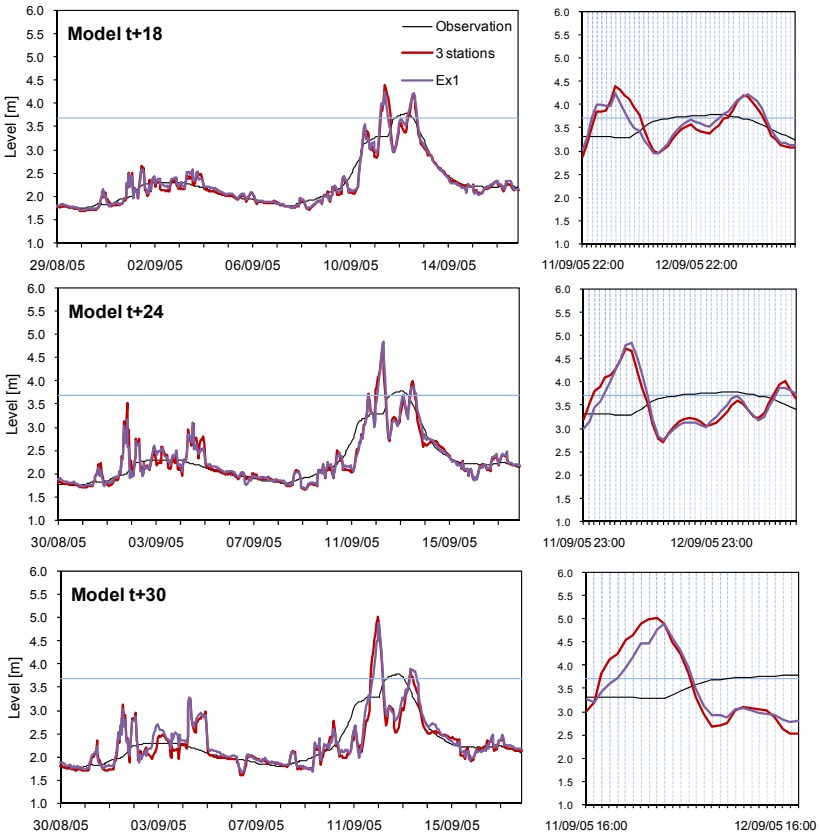


Figure E.13: Compare of the hydrographs of different training data (only 2005 and 2005-06) at 18, 24 and 30 hour lead time, test S3 (29 Aug – 18 Sep 2005).

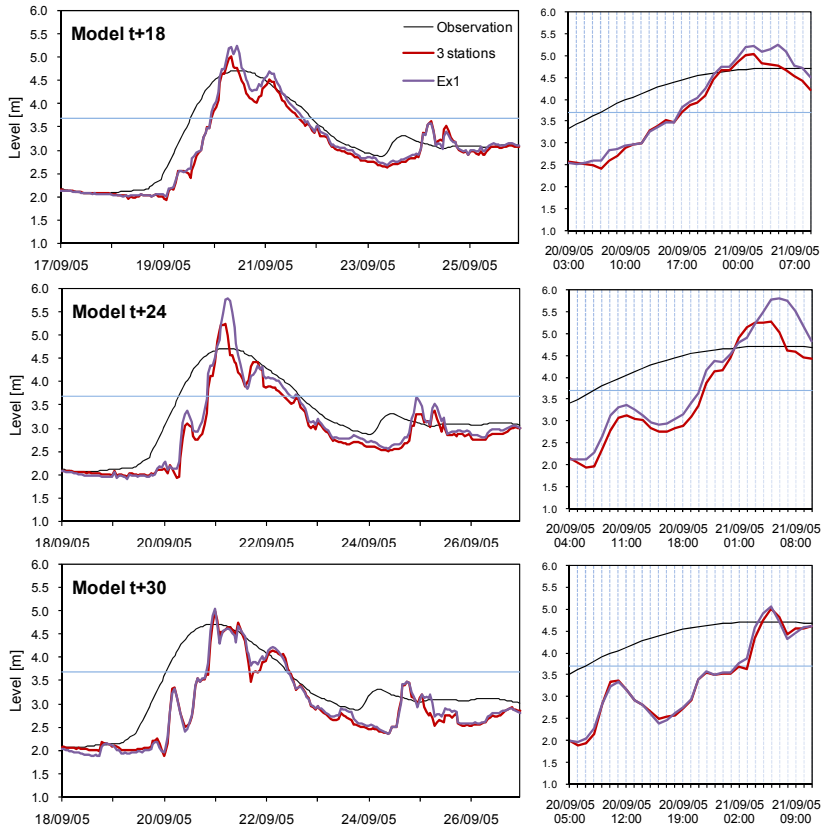


Figure E.14: Compare of the hydrographs of different training data (only 2005 and 2005-06) at 18, 24 and 30 hour lead time, test S4 (17 – 27 Sep 2005).

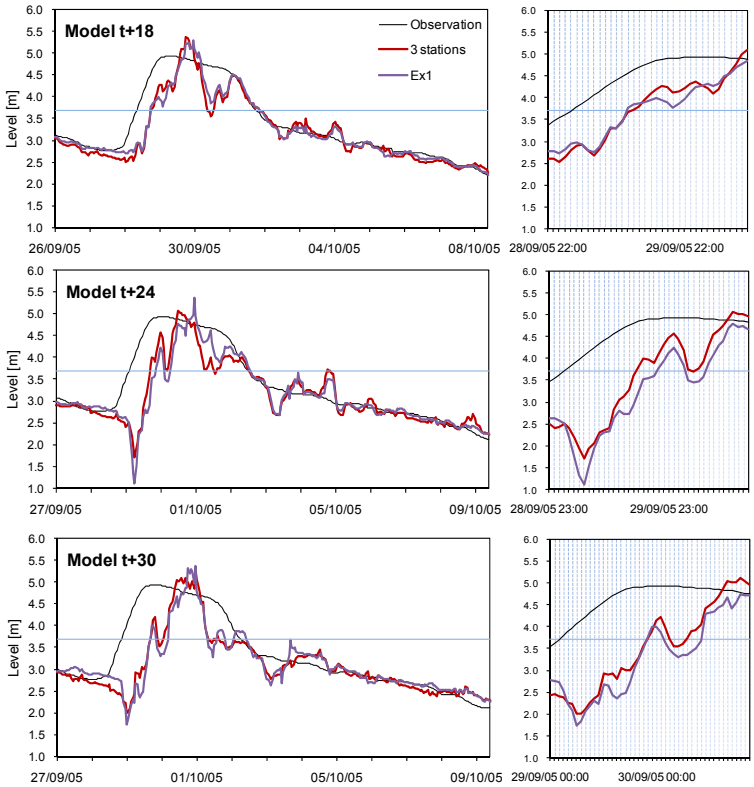


Figure E.15: Compare of the hydrographs of different training data (only 2005 and 2005-06) at 18, 24 and 30 hour lead time, test S5 (26 Sep – 10 Oct 2005).

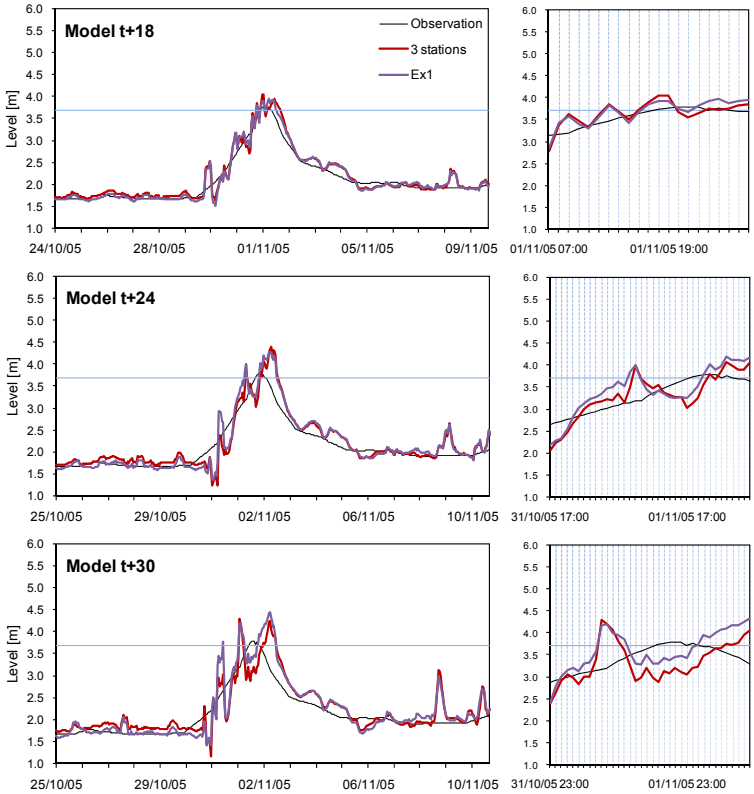


Figure E.16: Compare of the hydrographs of different training data (only 2005 and 2005-06) at 18, 24 and 30 hour lead time, test S6 (26 Sep – 10 Oct 2005).

Table E.23: The goodness of fit statistics testing S3 and S4 for Ex1 and 3 stations at a lead time 18, 24 and 30 hours.

Model	Statistic	S3		S4	
		Ex1	3 stations	Ex1	3 stations
t+18	PDIFF (m)	-0.4490	-0.5990	-0.5310	1.2851
	MAE (m)	0.1058	0.1160	0.0917	-0.3153
	RMSE (m)	0.1708	0.1937	0.2163	0.2363
	CE	0.8947	0.8646	0.9829	0.9796
	Time delay (hr)	-17	-17	9.5	10
t+24	PDIFF (m)	-1.049	-0.9151	-1.092	1.7313
	MAE (m)	0.1694	0.1888	0.1335	-0.5462
	RMSE (m)	0.2776	0.3093	0.2986	0.3282
	CE	0.7187	0.6507	0.9678	0.9611
	Time delay (hr)	-14.5	-15	13	13.5
t+30	PDIFF (m)	-1.102	-1.2312	-0.3450	1.9224
	MAE (m)	0.2256	0.2415	0.1606	-0.2955
	RMSE (m)	0.3519	0.3966	0.3670	0.3785
	CE	0.5429	0.4196	0.9518	0.9488
	Time delay (hr)	-13	-16	18.5	19

Table E.23: The goodness of fit statistics testing S5 and S6 for Ex1 and 3 stations at a lead time 18, 24 and 30 hours.

Model	Statistic	S5		S6	
		Ex1	3 stations	Ex1	3 stations
t+18	PDIFF (m)	-0.3530	-0.4294	-0.1700	-0.2615
	MAE (m)	0.1161	0.1405	0.0776	0.0873
	RMSE (m)	0.2501	0.2753	0.1282	0.1356
	CE	0.9277	0.9124	0.9756	0.9727
	Time delay (hr)	9	10	-3	5
t+24	PDIFF (m)	-0.4510	-0.1430	-0.477	-0.5976
	MAE (m)	0.1764	0.1832	0.1261	0.1314
	RMSE (m)	0.4305	0.3918	0.2136	0.2197
	CE	0.7874	0.8239	0.9323	0.9283
	Time delay (hr)	17	14	-10.5	-9
t+30	PDIFF (m)	-0.4370	-0.3749	-0.6520	-0.4991
	MAE (m)	0.2175	0.1995	0.1594	0.1688
	RMSE (m)	0.4758	0.4286	0.2778	0.2696
	CE	0.7421	0.7907	0.8856	0.8922
	Time delay (hr)	18.5	19	-10	-10

Appendix F

Table F.1: Input variables selected by different input determination techniques (denote by an X) for a 12, 18 and 24 hour lead time.

Input Variables	Models								Input Variables
	1	2	3	4	5	6	7	8	
P75L-24			X			X	X		6Z21
P75L-21	X		X			X	X	X	6Z22
P75L-18							X		6Z23
P75L-15							X	X	6Z31
P75L-12							X		6Z32
P75L-9	X			X			X		6Z33
P75L-6		X					X		6Z41
P75L-3	X	X		X		X	X	X	6Z42
P75L	X	X	X	X	X	X	X		6Z43
MV6P75	X		X			X			/
MV12P75									/
MV18P75									/
MV24P75									/
MV30P75						X			/
MV36P75						X			/
P67L-24						X			/
P67L-21				X		X			/
P67L-18				X		X			/
P67L-15						X			/
P67L-12	X					X			/
P67L-9	X	X		X					/
P67L-6	X			X					/
P67L-3	X	X		X	X	X			/
P67L	X	X		X	X	X			/
MV6P67	X	X		X		X			/
MV12P67						X			/
MV18P67			X	X					/
MV24P67			X			X			/
P1L-24			X						/
P1L-21									/
P1L-18						X			/
P1L-15			X			X			/
P1L-12						X			/
P1L-9						X			/
P1L-6	X	X	X	X		X			/
P1L-3			X						/
P1L	X	X		X	X				/
MV6P1	X	X	X	X	X	X			/
MV12P1					X	X			/
MV18P1									/
MV24P1						X			/
RP1-24									/
RP1-21						X			/
RP1-18	X			X		X			/
RP1-15						X			/
RP1-12			X			X			/
RP1-9						X			/
RP1-6			X			X			/
RP1-3						X			/
RP1	X			X		X			/
MV3RP1						X			/
MV6RP1	X		X	X		X			/
MV12RP1						X			/
MV24RP1	X		X	X		X			/
Total (54)	**18**	**10**	**15**	**18**	**6**	**37**	**9**	**3**	**9**

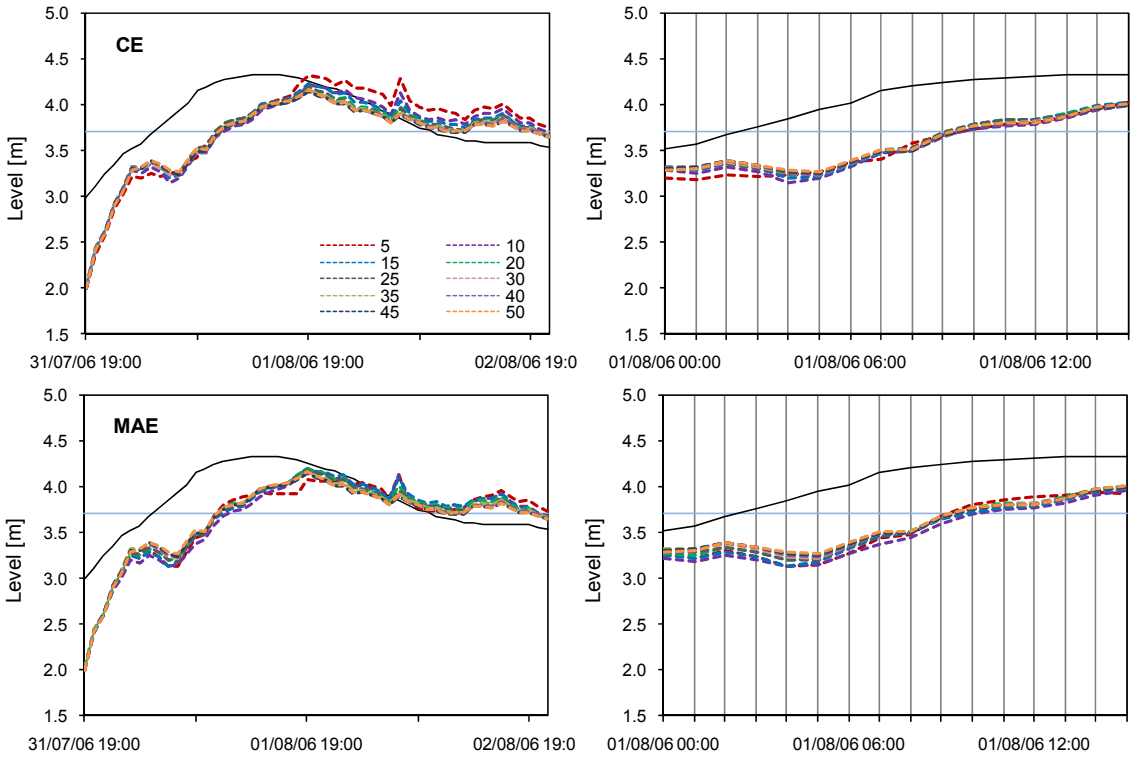


Figure F.1: Results for a storm in 2006 for increasing numbers of runs averaged together, chosen by ranking CE/RMSE and MAE for Model 1.

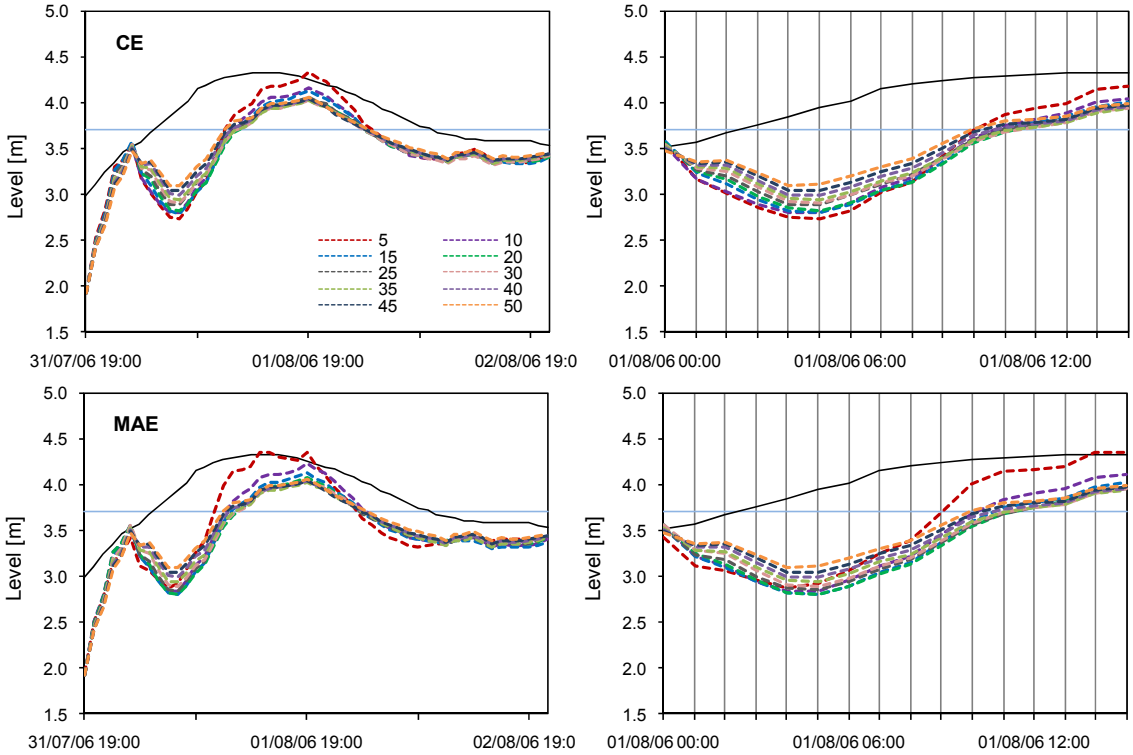


Figure F.2: Results for a storm in 2006 for increasing numbers of runs averaged together, chosen by ranking CE/RMSE and MAE for Model 2.

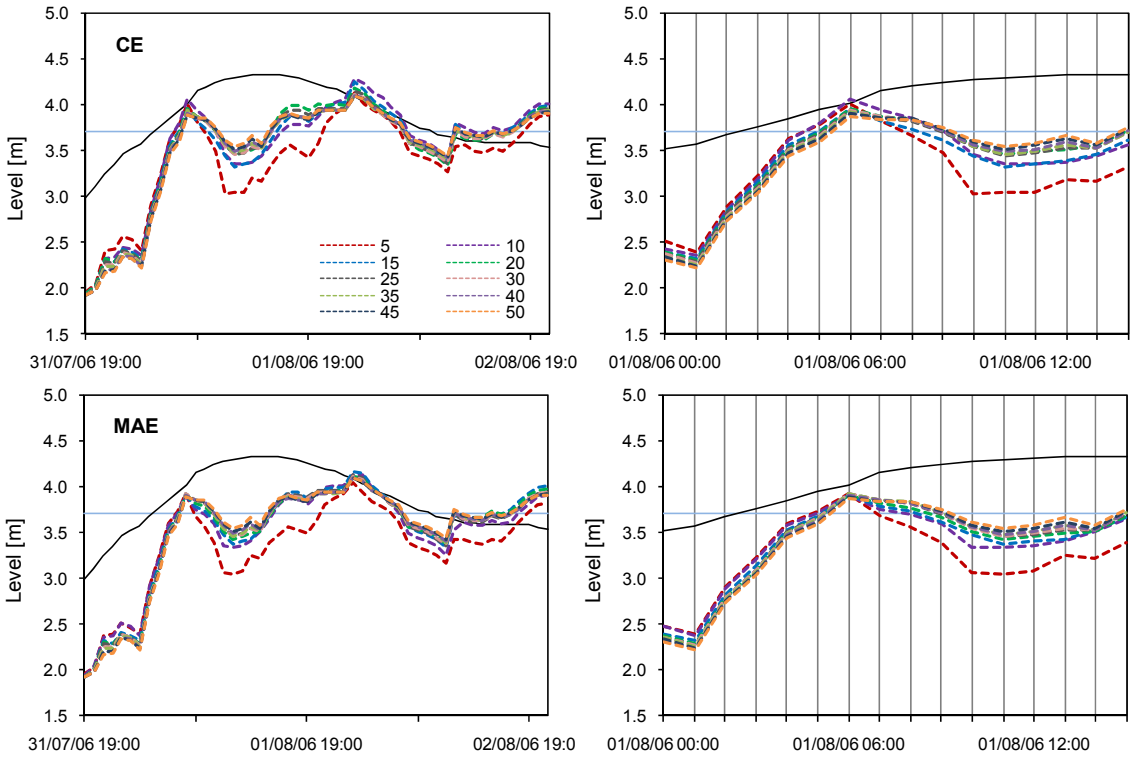


Figure F.3: Results for a storm in 2006 for increasing numbers of runs averaged together, chosen by ranking CE/RMSE and MAE for Model 4.

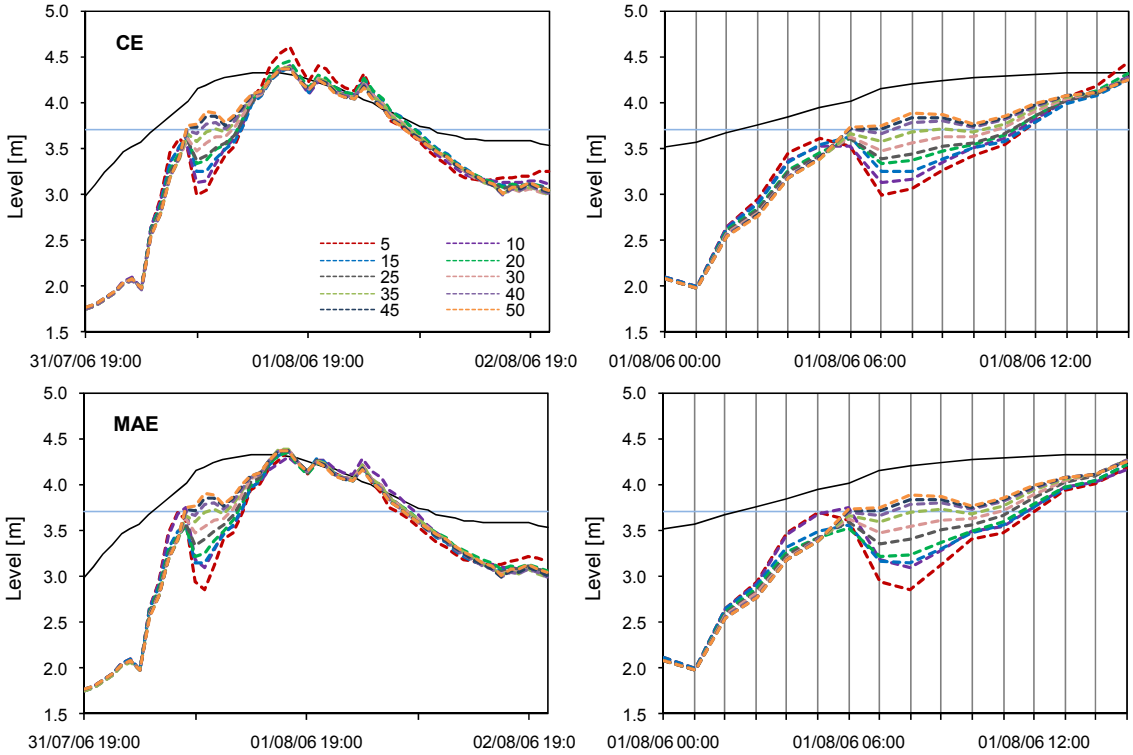


Figure F.4: Results for a storm in 2006 for increasing numbers of runs averaged together, chosen by ranking CE/RMSE and MAE for Model 5.

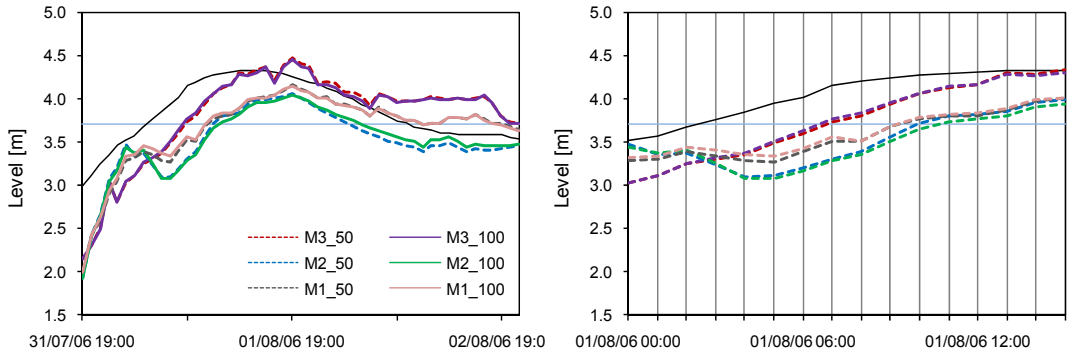


Figure F.5: Compare of the hydrographs between average 50 and 100 runs at 12 hour lead time, test 2006

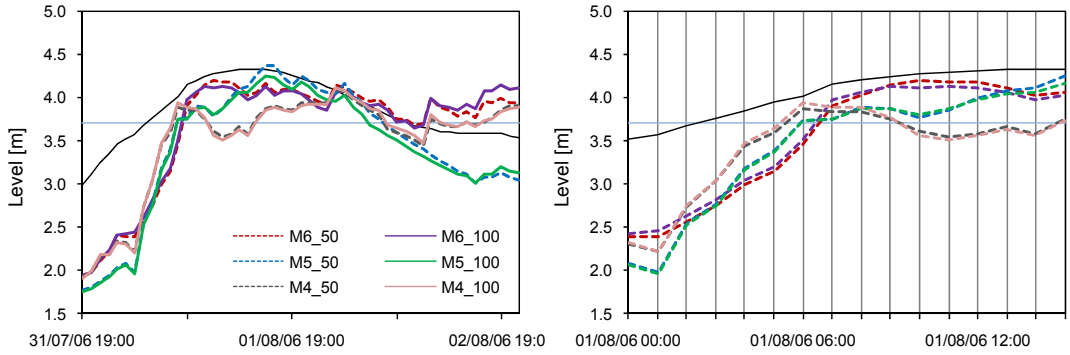


Figure F.6: Compare of the hydrographs between average 50 and 100 runs at 18 hour lead time, test 2006,

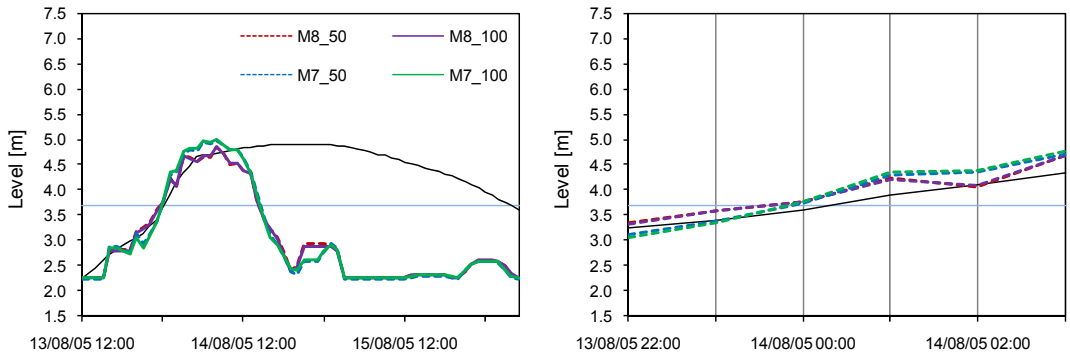


Figure F.7: Compare of the hydrographs between average 50 and 100 runs at 24 hour lead time, test S2, 2005