University of Southampton

USING ARTIFICIAL NEURAL NETWORKS TO PREDICT STORM SURGE IN THE NORTH SEA AND THE THAMES ESTUARY

Daniel Bruce Prouty

Thesis submitted for the degree of Doctor of Philosophy

School of Civil Engineering and the Environment

June 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE
SCHOOL OF CIVIL ENGINEERING AND THE ENVIRONMENT

Doctor of Philosophy

USING ARTIFICIAL NEURAL NETWORKS TO PREDICT STORM SURGE PROPAGATION IN THE NORTH SEA AND THE THAMES ESTUARY

By Daniel Bruce Prouty

An artificial neural network (ANN) was developed to predict storm surge magnitudes and arrival times at selected locations in the North Sea. The model predicts storm surges based solely on past measured water level residuals at one or more tidal stations. The research focuses on the performance of the model at the Sheerness tide station near the entrance of the River Thames in the UK. To take advantage of the specificity of surge propagation in the North Sea, the ANN uses input from both the target station and an additional station located where the peak of the storm surge has just passed. The ANN is trained to relate surge at the primary station from measured surge at a secondary station. The optimal secondary location is correlated to the forecast interval and the storm surge's propagation time between the secondary and primary station.

ANN performance is analyzed on an annual basis and on a 72-hour window centred on individual storm events which focuses the evaluation on a time when it is most critical. Performances are also compared at the times of both maximum surge and maximum water elevation during the passage of individual storm events. The simplest ANNs developed uses data from Sheerness only and predict surges with an absolute average error of 0.11 m for 3-hour predictions when analyzed on an annual basis. Models were systematically made more complex in an attempt to increase model performance by changing the both the size of the models, and the number of inputs used to train the ANN. A new ANN modelling method using input from several possible secondary stations was developed, decreasing the error to 0.08 m. This ANN model was compared to the continental shelf model (CS3) for 1, 3, and 5-hour predictions. The ANN model performed better than the CS3 model on an annual basis, but results were mixed when evaluating performance over the shorter 72-hour storm intervals.

This research further explores new forecasting methods using ANN ensembles to reduce variance and minimize error. The ensemble forecasting method averages results from multiple ANN models trained based on different model initializations. The use of ensemble forecasting with ANNs was found to significantly reduce variance when analyzed over a 72-hour storm window, but not model accuracy. The average absolute error for an ensemble ANN using 5 repetitions had 50% of the variance of a single ANN model. An ensemble model using 50 repetitions had 5% of the variance of a single ANN model.

A significant result of this research is the ANN's ability to accurately predict maximum water elevations. A single ANN model had a 4-hour forecast error of 0.017 m, while a simple [1,1] ensemble model using 20 repetitions performed better with an average 4-hour forecast error of 0.008 m. When over-training is included to reduce the model bias, the error is further reduced to 0.004 m. ANN ensemble model performances for predicting maximum storm surge were however less impressive. Best results were obtained for ensembles of [30,1] models with an average 4-hour forecast error of 0.68 m.

List of Contents

List of Figures

List of Tables

Chapter 3

List of Appendices

Acronyms

AAE	Average Absolute Error
ADCIRC	ADvanced CIRCulation Model
ANFIS	Adaptive Neuro Fuzzy Interference System
ANN	Artificial Neural Network
AR	AutoRegressive
ARIMA	AutoRegressive Integrated Moving Average
ARMA	AutoRegressive Moving Average
ARMAX	AutoRegressive Moving Average with eXogenous inputs
ARX	Auto-Regressive eXogenous
BODC	British Oceanographic Data Centre
BPTT	Back Propagation Through Time
CF	Central Frequency
CG	Conjugate Gradient
CS3	Continental Shelf Model
DEFRA	Department for Environment, Food and Rural Affairs
DF	Data Fusion
DRPE	Decoupled Recursive Predictive Error
EA	Environment Agency
FFNN	Feed Forward Neural Network
GRNN	Generalized Regression Neural Network
GRPE	Global Recursive Predictive Error
GWCE	Generalized Wave Continuity Equation
HNN	Hybrid Neural Network model
LAF-MLN	Local Activation Feedback - Multi Layer Network
MA	Moving Average model
MAE	Mean Absolute Error
MLP	MultiLayer Perceptron
MLP-IIR	MultiLayer Perceptron using Infinite Impulse Response filter
NLM-RPE	Neuron Linear Model Recursive Predictive Error
NMSE	Normalized Mean Square Error
NNM-RPE	Neuron Non-linear Model Recursive Predictive Error
NOAA	National Oceanic and Atmospheric Administration
NOS	National Ocean Service
NTSLF	National Tidal and Sea Level Facility
POLA	Port Of London Authority
POM	Princeton Ocean Model
R	Correlation Coefficient
RBF	Radial Basis Function
RBFL	Rule Based Fuzzy Logic model
RBP	Resilient Back Propagation
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RPE	Recursive Predictive Error
RTRL	Real Time Recurrent Learning
SLOSH	Sea Lake and Overland Surges from Hurricane Model
SOM	Self Organizing Map
STFS	Storm Tide Forecasting Service
SWAN	Simulating WAves Near-shore model
TCOON	Texas Coastal Ocean Observation Network
WAM	WAve Model

Notation

A	Coefficient of horizontal diffusion
A_o	Mean Height
a1	Weight 1
b1	Bias 1
Cd	Drag coefficient
D	Total water depth
ε_i	Phase of constituent
$e_{i,t}$	Equilibrium arguments for the constant i at time t
f	Coriolis parameter
$F_{i,t}$	Node factor for constituent I at time t
g	Acceleration due to gravity
h	Rotation of the Earth about the Sun
h_i	Amplitude of constituent
k	Unit vector in the vertical
K1	Lunisolar diurnal constituent
K2	Lunisolar semi-diurnal constituent
Ki	Phase of constituent i
L2	Smaller lunar semi-diurnal constituent
M2	Principle Lunar constituent
Mm	Lunar monthly constituent
N	Precession of the plane of the Moon's orbit
N2	Larger lunar semi-diurnal constituent
NU2	Larger Lunar elliptic semi-diurnal constituent
O1	Lunar diurnal constituent
p	Density of sea water
p	Precession of the Moon's perigee
pa	Atmospheric pressure at sea surface
pa	Density of air
q	Depth mean current
q_n	Component of mean current normal to boundary
s	Rotation of the Moon about the Earth
S2	Principle solar semi-diurnal constituent
Sa	Solar semi-diurnal constituent
T	Rotation of the Earth on its axis with respect to the Sun
t	Time
t_b	Bottom stress
t_s	Stress on sea surface
W	Wind velocity
xi	Surge value i
ζ	Elevation of sea surface
ω_i	Speed of constituent

Acknowledgements

I would like to thank the following who played important parts in my "English" education:

Heidi Shaw, who first inspired me to continue my education and to travel to England; My family, past and present, that helped make it possible for me to pursue a degree later in my career; Dr. Philippe Tissot provided the initial idea for this project and was instrumental in helping me achieve this goal. Dr. Tissot always provided a positive outlook for the project, no matter how big the obstacle; Dr. Arif Anwar provided essential guidance and helped me develop proper scientific methodology for research. Dr Anwar always had the time to discuss the project despite his busy schedule; Luke Blunden was very helpful in his knowledge all things of science, but particularly tidal hydrodynamics; Ivan Haigh provided much needed information on tide surge interaction and numerical approaches to surge modelling; Sally Brown was always willing to help with proof reading or scientific writing tips; Marc de Ruyter, who helped motivate the office, and provided professional engineering advice as required.

1 Introduction and Dissertation Overview

1.1 Introduction

The United Kingdom has experienced its share of devastation from storms in the past, most notably the great storm of 1953. During this event, flood defences were breached, and 307 lives were lost in the U.K. Over 150 km^2 of London was flooded, but Central London was spared. With the increased threat of global warming and the possible associated increase in storms, coastal regions of the United Kingdom are under an ever increasing threat from the sea. Several methods are available today to help predict storm surge heights and arrival times to help protect and aid in evacuation of low-lying areas. Numerical models are used today to predict tidal elevation and currents for large areas at a time, relying in part on input from additional large scale meteorological models for forcing. While the predictions are of great help, the models are based on initial conditions and broad scale forcings that are difficult to know precisely and are very computationally expensive. This research explores the use of artificial neural networks for forecasting storm surge elevations. Artificial neural networks are computationally inexpensive, requiring only historical sea-level information for training and are a sub-second process once trained.

Although a relatively new science, artificial neural networks (ANNs) have been successfully used for the last 20 years for numerous applications including storm surge prediction. Artificial neural network storm surge models have been implemented in various locations around the world, but none have the advantage of the unique spatial configuration of sequential tide gauges along the eastern coast of the United Kingdom. For this location, a typical storm path would enter the North Sea from the North Atlantic Ocean; the earth's rotation would force the surge against the coast which would act as a wave guide, as the surge moves southward it would pass each of the tide station sequentially, offering a "pre-view" of potential storm surge for each subsequent location to the south. This favourable regional setting promises a unique opportunity to take advantage of a multiple station artificial neural network for early warning and prediction of storm surge events.

1.2 Hypothesis

Decimetre level prediction of storm surge in the North Sea and Thames Estuary can be made using artificial neural networks (ANNs). The use of multiple potential secondary station locations will enable an artificial neural network to predict accurate storm surge elevations for different forecast intervals, with longer term forecasts aided using secondary stations progressively further away.

1.2.1 Basis for hypothesis

Artificial neural networks use historical and current sea level data, as a basis on which to develop reasonable expectations about the future sea levels. Historical sea level data are used to train artificial neural network's, and then tested against other data sets. Results are compared to actual measured sea levels to determine model validity and performance. There exists need for an inexpensive (resource wise) method to predict sea levels that can be used stand-alone or to complement or test existing models.

1.3 Research Aims

The aim of this research is to prove that an artificial neural network can predict accurate storm surge elevations with warning times from 1 to 24 hours in the North Sea and Thames estuary. Ideally this model should:

- Use minimal data resources (using only historical tidal records and current water-level measurements).
- Minimize computational time. (short training times < 24 hours)
- Be easy to implement after training. (Final model can easily be implemented in any programming language or format)
- Accurate (compete with similar level of accuracy) with existing numerical models)

Research programme:

- Develop a artificial neural network storm surge Forecast Model for use in the North Sea
- Evaluate artificial neural networks storm surge prediction accuracy at the Sheerness Tide station
    i. Optimize secondary station location.
    ii. Optimize amount of data used for training.
    iii. Optimize size and structure of network.

- Evaluate artificial neural networks performance vs. Storm Tide Forecasting Service (STFS) model.
- Evaluate artificial neural networks performance with ensemble forecasting
- Evaluate artificial neural networks performance at the Thames Barrier

This research will develop several Artificial Neural Network models to achieve objectives. Numerous programs have been written to process/import data sets.

1.4 Research Objectives

This research will:

- Construct a database for UK tide gauges in the North Sea.
- Build artificial neural network models for surge prediction at Sheerness.
- Optimize the artificial neural network size and structure.
- Optimize the secondary station selection based on forecast interval.
- Develop an artificial neural network using Ensemble forecasting methods.
- Compare the artificial neural network model to Proudman Oceanographic Laboratory's Continental Shelf Model (CS3).
- Use Port of London Authority tide data to judge accuracy of artificial neural network models near the Thames Barrier.

1.5 Research Scope

(Limitations)

- Use sea-level data only for input variables to the artificial neural network. (To better understand North Sea artificial neural network basics)
- Limit the artificial neural network analysis to UK tide gauges locations only. (To simplify artificial neural network training process)
- Limit the locations of the North Sea tide gauges to be used as input to the artificial neural network to Sheerness, Immingham, North Shields, and Wick. (To simplify artificial neural network training process)
- Restrict the total number of tide gauge stations to be used as input to the artificial neural network to two stations per model.

1.6 Dissertation Organization

The Dissertation is organized into 8 Chapters. The following is a short description of the chapters and their contents:

1. Introduction and Dissertation Overview – An introduction to the research, Hypothesis, Aims, Objectives, and scope (This chapter)
2. Background – essential scientific background information concerning storm surges and neural networks.
3. Literature Review – A general review of literature of water level forecasting methods, artificial neural networks in environmental forecasting, and using artificial neural networks for water-level forecasts.
4. Materials – Description of data sources, hardware and software used in the research.
5. Methods – General discussion of methodology used for the experiments.
6. Results and Analysis – Results and analysis from experiments discussed in Chapter 5.
7. Discussion - General discussion on the research and results from each experiment.
8. Conclusions – overall conclusions, future work.

2 Background

2.1 Regional setting

The North Sea is a relatively shallow body of water located on the European continental shelf between Great Britain, Norway, Sweden, Denmark, Germany, the Netherlands, Belgium and France (Figure 2-1). It has a surface area of 750,000 km², and spans approximately 960 km north-south, and 580 km east-west. The depth varies from about 200 m in the north to about 30 m in the south, averaging around 95 m. The Norwegian trench traverses the north-western portion of the North Sea with a maximum depth of 700 m. The maximum tidal range is about 8 m. The North Sea is one of the most frequently traversed sea areas of the world and contains two of the world's largest ports, Rotterdam, and Hamburg.

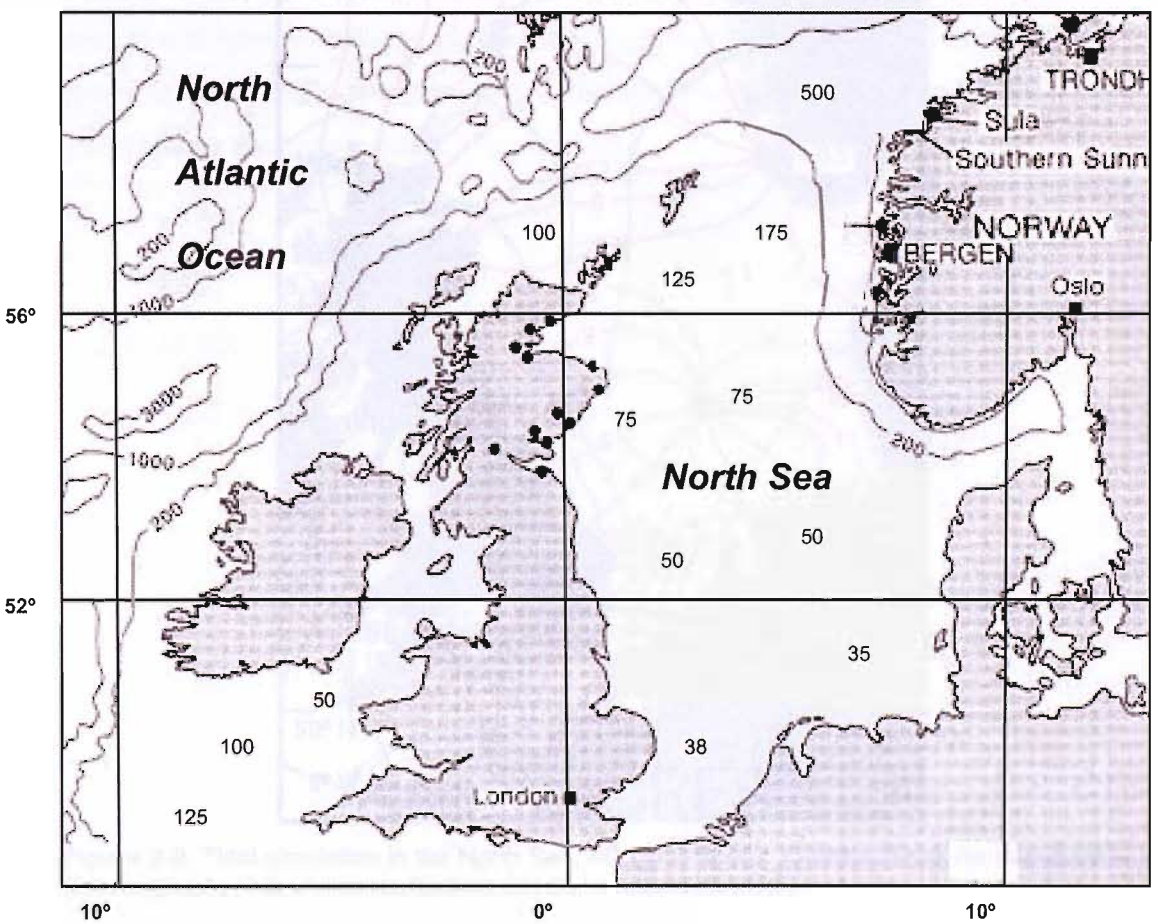


Figure 2-1 North Sea and regional bathymetry (Depths in Meters)

2.2 Tidal circulation in the North Sea

The primary tidal circulation in the North Sea rotates counter-clockwise around an amphidromic point located in the centre of the southern half of the North Sea

(Figure 2-2). This circulation causes tides to propagate from north to south along the East coast of the United Kingdom. The circulation is shown for the largest tidal constituent (M2) in Figure 2-2, where Co-tidal lines (radiating outward from amphidrome) connect points experiencing the same phase of the tide. Co-range lines (circling amphidrome) connect points of equal tidal range. Figure 2-2 shows that by counting the number of co-tidal lines it takes approximately 15 hours for a tide to travel from Wick to Sheerness. The other tidal constituents are smaller and have different circulation patterns. Cross-correlation analysis of sea level records show that a typical storm surge averages 14 hours to traverse this distance, and indicates that tidal and storm surges propagate at approximately the same speed, allowing several hours interaction between the two phenomena.

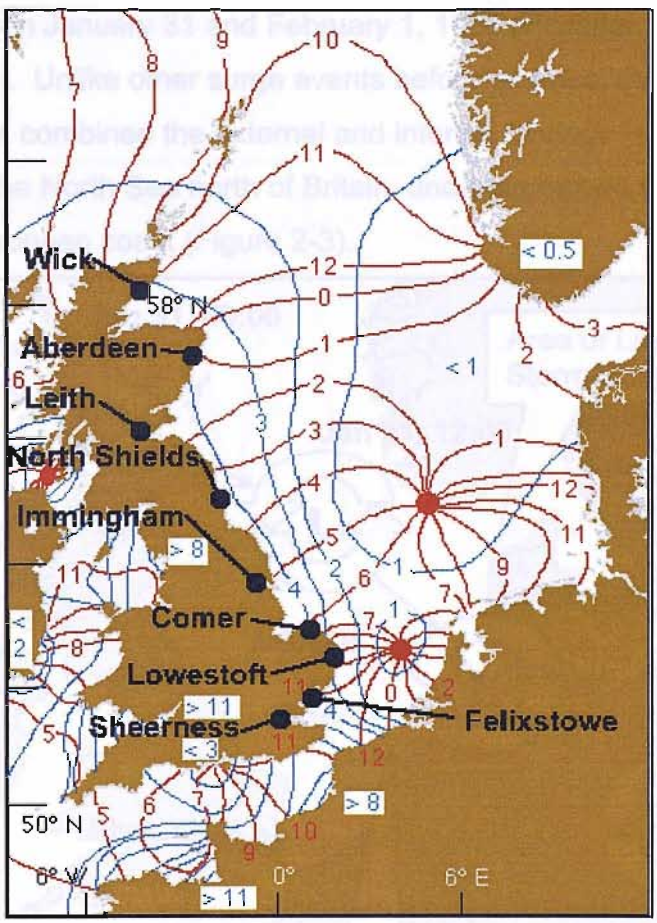


Figure 2-2 Tidal circulation in the North Sea, M2 Constituent. (Adapted from An Introduction to Oceanography, http://www.es.flinders.edu.au/~mattom/IntroOc/)

2.3 North Sea storm surges

Storm surge events are meteorologically induced water level changes, and are defined as the difference between measured water level and tidally predicted water levels. These surge events are caused by regional differences in barometric pressure and associated wind stress on the water surface (Pugh 1987).

There are two types of surge events that can occur in the North Sea. Internal surge events are surges generated within the North Sea Basin itself, and external surge events created in the North Atlantic over deep waters that then move onto the shallow continental shelf regions (Pugh 1987). The North Sea is open to the North Atlantic Ocean at its northern end, but is essentially a closed basin at its southern end. Surges that enter the North Sea from the North Atlantic are affected by Coriolis forces and rapidly decreasing depths. These forces cause the surge to increase in height and move southward along the eastern coastline of Britain. Internal surges occur less frequently than external surges, but generally produce more severe surges (Pratt 1995).

One of the most intense surge events ever recorded in the North Sea occurred between January 31 and February 1, 1953 (Rossiter 1954) and (Royal Society A 2005). Unlike other surge events before or since, the 1953 storm appears to have combined the external and internal forcings (Ishiguro 1976). This storm entered the North Sea north of Britain, and then moved southeast towards the western European coast (Figure 2-3).

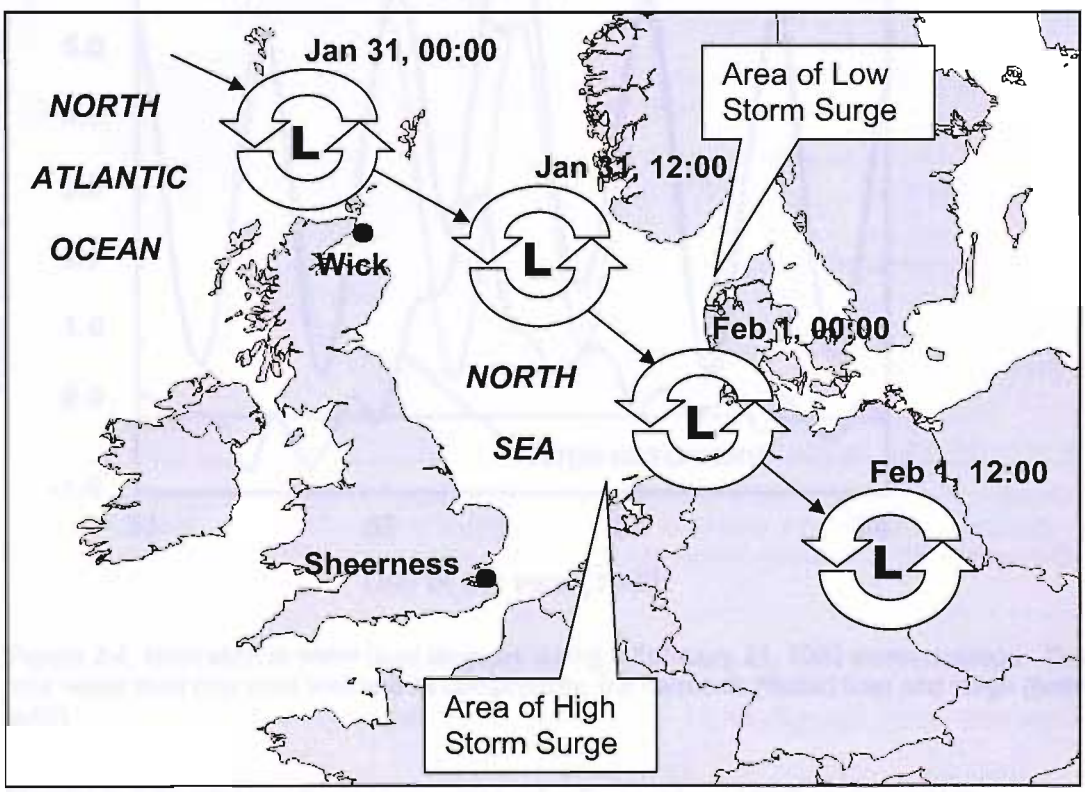


Figure 2-3 Path of the 1953 storm in the North Sea. The Large L's indicate the location of the centre of circulation at the date and time given. The large arrows show the storms anti-clockwise rotation and the small arrows indicate direction of the storms travel.

The surge event itself has two generating forces, (1) regional barometric pressure differences which can cause an increase in sea-surface elevation near the low

pressure centre, and (2) wind stress along the surface of the water generating surface waves which can cause additional set-up of the sea surface.

A comparison of harmonically predicted water levels, and measured water levels are presented in Figure 2-4 for a 1993 storm event recorded at the Sheerness tidal station. The predicted or harmonic component of water level is indicated by the dotted line, and the surge component is indicated by the solid line at the bottom of the figure. The total measured water level is therefore the combination of the two components and is indicated by a solid line at the top of the figure. During this storm, a maximum storm surge elevation of 3.0 m occurred approximately 3-hours before high tide, and then decreased to 1.2 m at the time of high tide, resulting in a maximum measured water level of 6.5 m above Ordinance Datum (which approximates mean sea level).

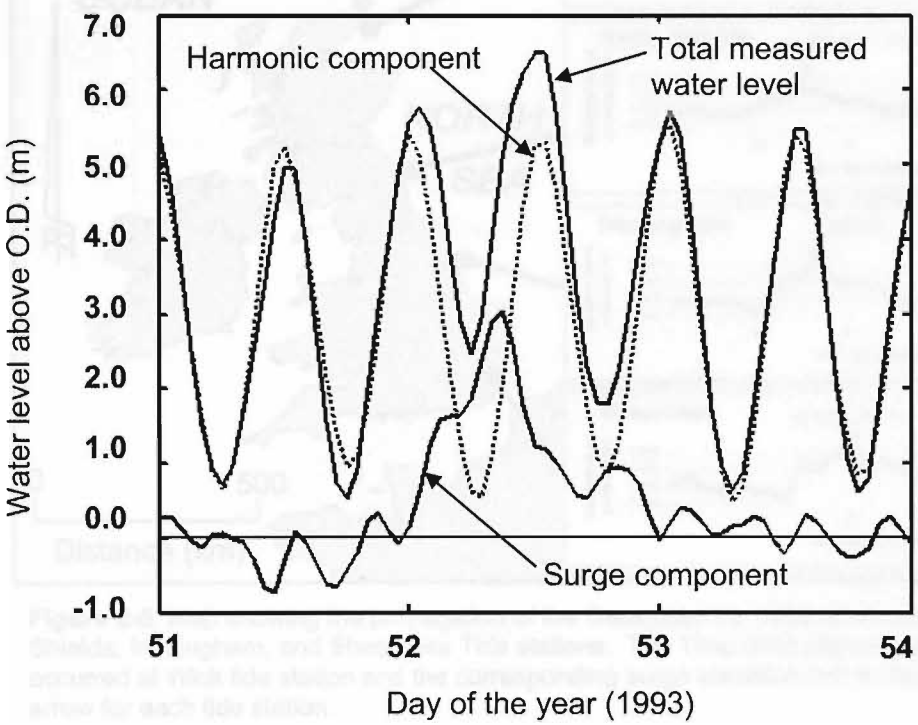


Figure 2-4 Illustration of water level changes during a February 21, 1993 storm passage. The total water level (top solid line) and its components, the harmonic (dotted line) and surge (bottom solid).

Figure 2-5 Illustrates the progression of a December 1990 storm and its associated storm surge recorded at four tide stations: Wick, North Shields, Immingham and Sheerness. When the storm surge passed Wick, the most northerly station, the storm surge peak was 0.3 m above predicted harmonic water level. As the storm surge moved southward, the maximum surge levels at North

Shields and Immingham occurred respectively 6 hours and 8 hours later, and peaked at 1.2 m and 1.7 m above the harmonic predicted water level. The storm surge reached Sheerness approximately 16 hours later, with a peak elevation of 2.1 m above predicted harmonic water level. The growing size of the storm surge can be explained by a combination of the North Sea bathymetry relative to the storm path, and meteorological influences. When moving southward, the relative depth of the North Sea shallows and its width (measured East -West) narrows. As a consequence, the storm surge height increases to compensate for the decreased volume of the basin (this can also be exacerbated by a deepening low pressure from a strengthening storm).

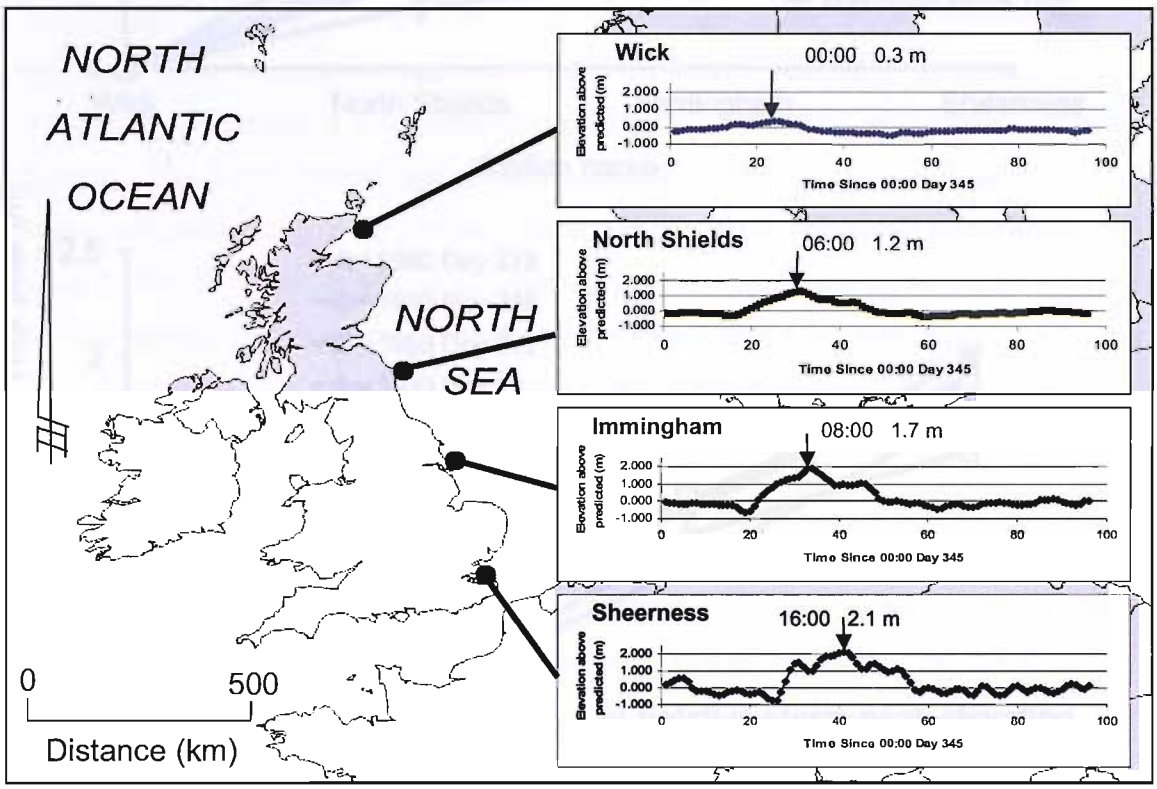


Figure 2-5 Map showing the propagation of the December 12, 1990 storm surge for Wick, North Shields, Immingham, and Sheerness Tide stations. The Time (hrs) elapsed since peak water level occurred at Wick tide station and the corresponding surge elevation (m) is displayed above an arrow for each tide station.

Storm surge propagation along the coast is illustrated for five of the largest storms found in the 1990 – 2004 data sets (Figure 2-6). These figures were created by determining the peak surge elevation at Wick, then subtracting it from each station's peak surge elevation to give a relative change in height as the surge moved down the coast. As can be seen from Figure 2-6a, the time lag in storm surge, as well as the sizes of the storm surges at each location (Figure 2-6b), are relatively consistent between storms. The development of the ANN models used in this project are in large part premised on taking advantage of the relatively

consistent storm characteristics and time delay between the storm surges at the northern stations as compared to the surge at the targeted Sheerness station.

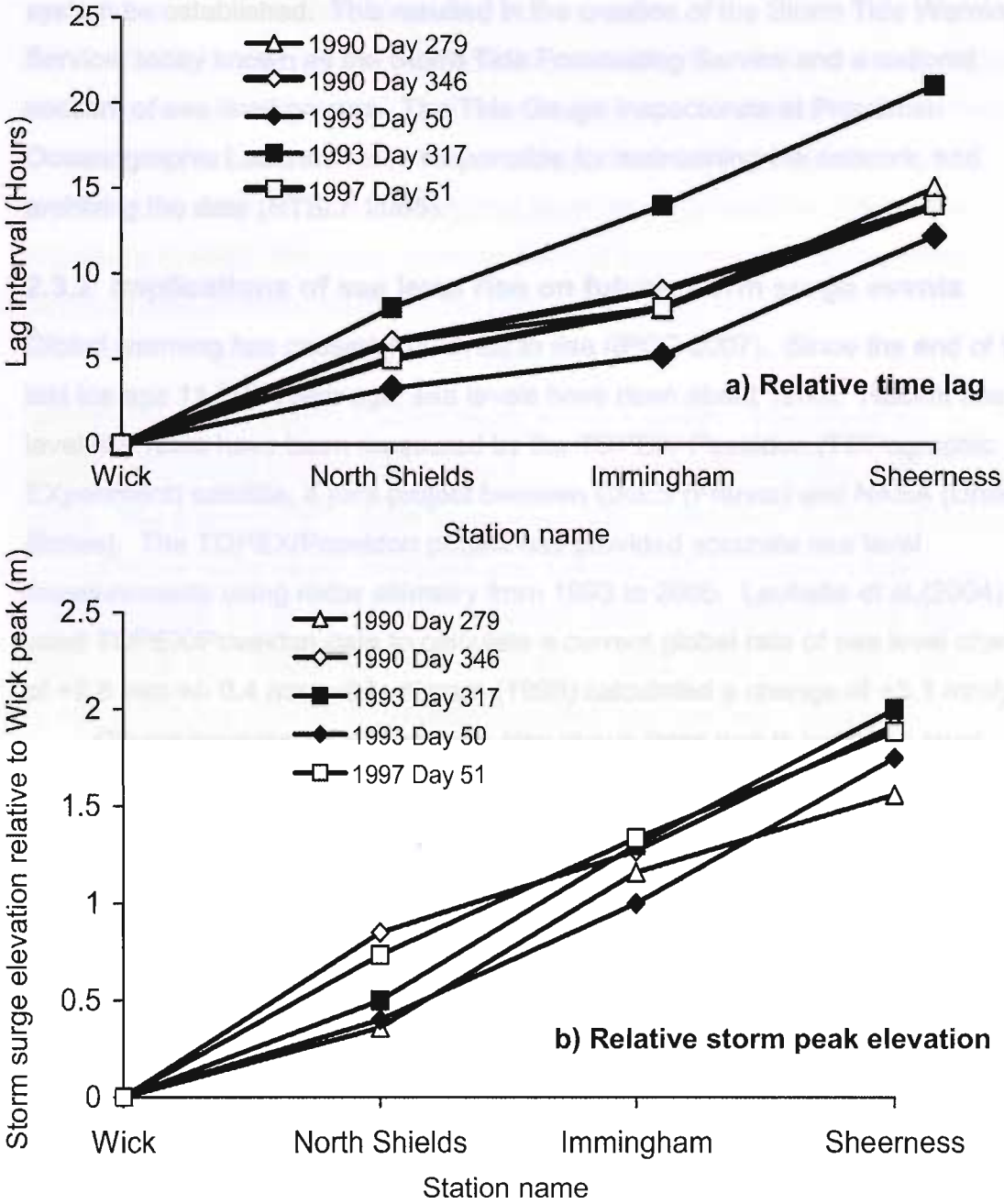


Figure 2-6 Dynamics of the 5 largest storm surges between Wick and Sheerness (years 1990 – 2004). a) Shows the time lag in hours since the maximum water level was recorded at Wick. b) Shows the change in maximum residual for each station relative to the maximum elevation recorded at Wick.

2.3.1 Storm surge impact on Europe and the United Kingdom.

Historically storm conditions on the North Sea have produced large surge events (Flather et al. 1998). Every year storm surges propagate southward along the UK's North Sea coastline. These events are meteorologically driven and are not predicted by harmonic models. After the floods of 1953, the UK government set

up a committee, led by Lord Waverly to investigate and report on the event (Wordie et al. 1953). The Committee recommended that a national flood warning system be established. This resulted in the creation of the Storm Tide Warning Service, today known as the Storm Tide Forecasting Service and a national network of sea level gauges. The Tide Gauge Inspectorate at Proudman Oceanographic Laboratories is responsible for maintaining the network, and archiving the data (NTSLF 2005).

2.3.2 Implications of sea level rise on future storm surge events

Global warming has caused sea levels to rise (IPCC 2007). Since the end of the last ice age 11,000 years ago, sea levels have risen about 120m. Recent sea level rise rates have been measured by the TOPEX/ Poseidon (TOPographic EXperiment) satellite, a joint project between CNES (France) and NASA (United States). The TOPEX/Poseidon project has provided accurate sea level measurements using radar altimetry from 1993 to 2005. Leuliette et al.(2004) used TOPEX/Poseidon data to calculate a current global rate of sea level change of +2.8 mm +/- 0.4 mm, while Nerem (1999) calculated a change of +3.1 mm/year.

Glacial isostatic adjustment can also play a large part in local sea level changes. By the end of the last ice age large portion of North America and northern Europe were covered by ice sheets up to 3 km thick. In Great Britain, the ice covered all of Scotland, Wales, and all of England north of London. The weight of this ice depressed the lithosphere and caused it to sink into the asthenosphere. This caused the southern portion of England to rise in response. As the ice retreats, the load on the lithosphere and asthenosphere is reduced and they rebound back towards their equilibrium levels. This rebound can cause a local increase of sea level of up to 0.9mm/yr for the London area (Shennan and Horton, 2002). When this local subsidence rate is combined with the current global sea level rise rate (Nerem 1999) of 3.1 mm, changes of up to +4 mm per year can occur, which can cause substantial increase in flood risk in the future. While long term sea level rises are of great importance to the impact of storm surges, the present work is focused on the prediction of short term events.

2.4 Tide-Surge Interaction

Tide-surge interaction can also influence the magnitude and arrival times of large surge events in the North Sea. The tidal range of the North Sea and its relatively shallow depth causes the speed of propagation of both the surge and tide to

interact. Rossiter (1961) suggested, and more recently confirmed by Horsburgh and Wilson (2007), that a key mechanism of interaction between tide and surge is one of mutual phase alteration.

Storm surges can be interpreted as coastally trapped long waves (Tang, Holloway, and Grimshaw 1997). These waves move counter clockwise along coasts in the northern hemisphere, and using the coastline as a waveguide. Traditionally storm surge modelling has been based on shallow water wave equations in which the speed of propagation is independent of their period, and depends only on water depth, in the form

$$c = \sqrt{gh} \tag{2.1}$$

where c = wave speed, h = water depth, and g = gravity. The speed of propagation of the storm surge varies as the water depth changes during the tidal cycle. The presence of a storm surge alters the propagation speed of the tide by changing the water depth (As-Salek and Yasuda 2001). This interaction can cause the high tide to accelerate and arrive before its predicted arrival time (Pugh 1987). This early arrival time of the tide will be interpreted as "surge", since surge is defined as the difference between observed and predicted water levels. Superimposed on (or in addition to) this accelerated tide, can be the meteorologically induced, water level changes themselves. Prandle and Wolf (1978) investigated the tide-surge interaction in the North Sea and Thames Estuary. They and found that positive surge peaks are not coincident with times of high water, and usually occur on the rising tide. Horsburgh and Wilson (2007) found that surge generation is modulated by the state of the tide. Their work indicated that enhanced surge is generated during low tides, and this effect is less pronounced during times of high tides. Horsburgh and Wilson (2007) have also shown that for all realistic situations, a delay between the time of high surge and high tide will always occur. This project will not attempt to distinguish what proportions of the surge are attributable to acceleration of the tide or due to meteorological influences, and will treat all differences between observed and predicted as storm surge. A detailed analysis of tide and surge interaction at Sheerness is given in Appendix 1.

2.5 Methods of prediction water levels.

In the United Kingdom, several methods for predicting water levels are being used today. Tidal analysis is computed by harmonic analysis of previous water level

records and is therefore primarily based on periodic astronomical forcings (Schureman, 1958). The method works well for large portions of the year but meteorological effects can significantly influence water levels and can introduce substantial errors in tidal predictions. These additional major forcings on water levels are fundamentally different in their variability with tidal influences being periodic, while weather forcings are fast changing, and mostly aperiodic. To include these additional forcings Proudman Oceanographic Laboratories have developed finite difference models (NTSLF, 2005). These model predictions provide substantial improvements over harmonic forecasts but also require large amounts of meteorological and oceanographic data, computer time and are updated only 4 times per day.

2.5.1 Using Artificial Neural Networks to help predict Storm surge.

This research investigates the potential of an alternate methodology to the numerical methods based on artificial neural networks to predict water levels and storm surges at the Sheerness Tide Station located at the entrance of the River Thames Estuary. Surge events are to be predicted at the Sheerness station, because of its importance on the decision to open or close the Thames Barrier. The model development for this project takes advantage of the large set of observations archived by the National Tidal and Sea Level Facility (NTSLF) for stations along the UK coastline, and storm propagation characteristics in the North Sea. Measurements from these tide stations provide information on the advancing storm. The model uses the non linear modelling capability of ANNs (Rumelhart et al. 1995) to predict future water levels at the target station. This thesis also reports on the use of more complex ANNs and in particular ensemble ANN models to improve forecasting during the largest storms. The performance and robustness of these larger models are studied as the size of the ANN hidden layer and the number of ensemble members is increased. The continuing development of accurate predictive models is important for the safety of growing coastal communities and navigation. This thesis investigates the use of Artificial Neural Networks (ANNs) as a tool to predict storm surge propagation in the North Sea.

2.6 Artificial Neural Network Methods

An Artificial Neural Network (ANN) is an information processing method that is inspired by the way biological nervous systems, such as the human brain, process information (Rumelhart et al. 1995, Hecht-Nielsen 1989). The key element of this method is the structure of the information processing system. It is composed of a

number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN can be configured for a specific application, such as pattern recognition or data classification. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. Artificial neural networks work in a similar way with the ANN parameters, weights and biases, being adjusted as part of a learning process supervised or unsupervised (Hagan et al. 1996). This modeling methodology does not attempt to understand explicitly the underlying physics. The physics of the problem is captured in the choice of the input data sets and the choice of the ANN structure (with larger number of hidden neurons enabling more non-linear relationships). If the input data sets do not contain the physical forcing parameters associated with the event being forecasted, the ANN will not be able to establish a relationship between inputs and outputs.

2.7 ANN Structure

The artificial neural network model used in this study is a feed-forward back-propagation model. The ANN structure consists of an Input layer consisting of previous water level residuals from a primary and/or a secondary station, a hidden layer using 1 to as many as 60 neurons, and an output layer using only one neuron (Figure 2-7). Square bracket notation will be used for designating the size of ANN structures during this research. For example [2,1] indicates the model uses 2 hidden neurons, and 1 output neuron.

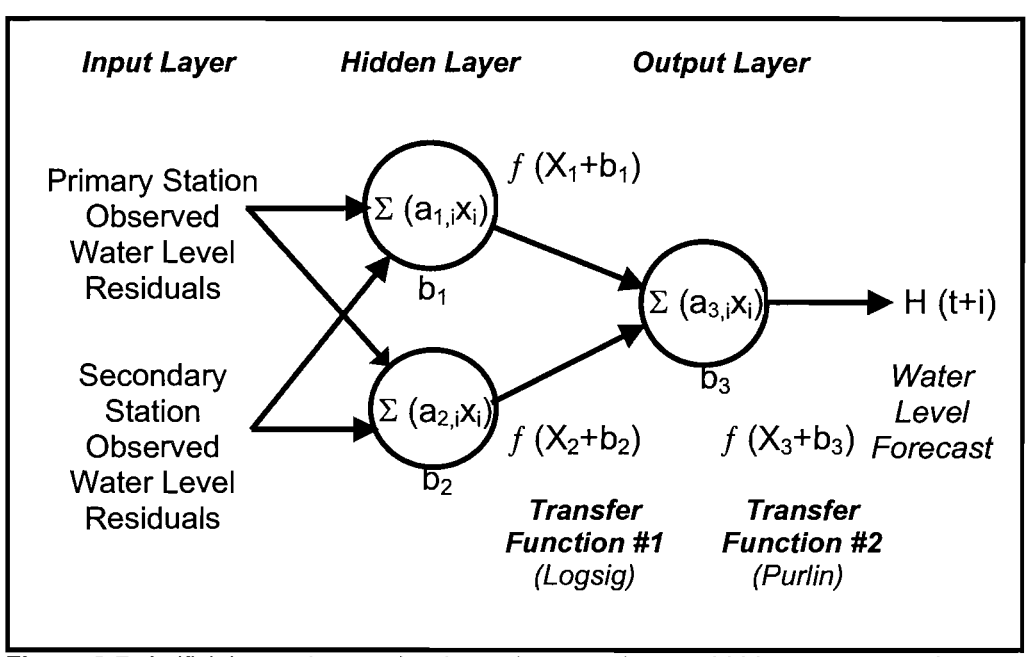


Figure 2-7 Artificial neural network schematic – showing two hidden neurons and one output neuron.

2.7.1 Input Layer

The input layer consists of inputs from a primary station (the location where the water levels will be predicted) and for most experiments, a secondary station. These inputs are time series of water level residuals from 1 to 48 hours of past sea level data. A water level residual is defined as predicted water level minus observed water level. Information from the input layer is used as input for the hidden layer. Each component of the ANN is described in more detail below.

2.7.2 Hidden Layer

Within the hidden layer, are the hidden neuron(s). Each hidden neuron is connected to all inputs. For example when using a single station ANN with 24-hours of previous water level measurements (for each forecast), each of the 24 inputs are multiplied by a weight ($a_{1,i}$) (initially randomly generated number), summed, and finally a bias b_1 is added to the previous sum.

$$X_1 = \Sigma (a_{1,i} x_i) + b_1 \tag{2.2}$$

This requires data storage of 24 weights and one bias, one for each of the 24 previous water level measurements used for each forecast. Output from the hidden layer is used for input for the transfer function.

2.7.3 Transfer function #1

Next, a bias (b_1) is added, and a transfer function is applied to the result from each hidden neuron (this requires data storage of 1 bias to be used for each forecast).

$$y = f(X_1 + b_1) \tag{2.3}$$

For this project logsig transfer functions were used for the hidden layer transfer function (Figure 2-8). The result from transfer function #1 is used as input for the output layer. The use of a sigmoid type transfer function for the hidden layer allows for the non-linear capability of the ANN.

$$y = \frac{1}{1 + e^{-x}} \tag{2.4}$$

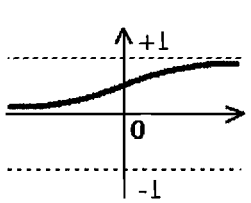


Figure 2-8 Logsig transfer function

2.7.4 Output Layer

The output layer consists of a single output neuron forecasting a single water level forecast value. The output neuron uses the results from transfer function as input. These inputs are multiplied by a weight ($a_{2,i}$) (initially an arbitrary number), and then summed.

$$X_2 = \Sigma (a_{2,i} x_i) \tag{2.5}$$

For a single station ANN with one hidden neuron, this operation requires data storage of one weight for the output layer.

2.7.5 Transfer function #2

Results from the output layer are used as input for the transfer function #2, where a bias (b_2) is added, and a transfer function is applied to the result from each neuron

$$y = f(X_2 + b_2) \tag{2.6}$$

(this requires data storage of 1 bias to be used for each forecast). For this project purlin transfer functions were used for the output transfer functions (Figure 2-9). After applying transfer function #2, the result is the forecasted water level once the ANN is trained.

$$y = x \tag{2.7}$$

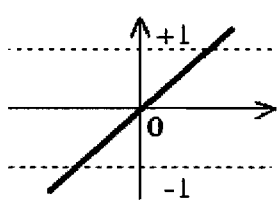


Figure 2-9 Purlin transfer function

2.8 Training and Application of the Artificial Neural Network Model

There are two phases in artificial neural network processing. They are the training phase and the testing phase. In the training phase, a training data set is used to (determine) find an optimum set of the weights and biases to be used later in the testing phase to predict water levels using additional data sets. It is important to note that the results of the optimization are not unique and that for each training run a slightly different result can be expected.

2.8.1 Training Phase

All ANN models are trained using the Levenberg-Marquardt algorithm as implemented within MATLAB unless otherwise indicated. Training times varied between a few minutes and several hours depending on the size of the ANN. It is important to note that although training times can be lengthy, generating water level forecasts is a sub-second process once the models are trained. Once the ANN models are trained, they are ideally suited for streamed forecasting (an automatically generated, real-time forecast based on streaming data).

When initializing an ANN, all weights and biases are assigned random values. During the training phase, the weight and biases are adjusted after each iteration until a pre-set goal is reached. During the first iteration, the network produces an output which has no relationship to the target data. This output is then compared to the target data, and an error is calculated for each set of inputs. The error is then back propagated through the hidden layer, and the weights are updated in a direction that will decrease the overall error. After several iterations, the network begins to adapt to the given model target, and typically gives outputs close to that which was expected.

Training – When to stop

During training the network might over-learn the training data set. This problem is referred to as over-fitting, and is of special concern for ANNs using a large number of hidden layers or neurons. This is analogous to over fitting a set of data points when performing a polynomial non-linear regression. Given a polynomial with a high enough order, any data set can be fitted almost exactly. However excessive noise will usually appear when applying the model to other related data sets. In the same manner, large structure ANNs have the tendency to over-train more easily than small structure ANNs. If not stopped during training, they will over train and start forcing a fit to the training data set. At this point, the ANN starts to memorize the training data set rather than capturing the underlying relationship between input and output. When an ANN is over-trained it can predict very accurately the original training data set, but performs poorly when using "other" data sets. Numerous methods are used stop training and prevent over-fitting, they include:

1. **Manually** - Visually watch root mean square error (RMSE) levels and

manually stop when error level "flattens out". This method was not used, it was found to produce inconsistent results especially when training networks with large number of neurons.

2. **Training Parameter Goal** - Training is automatically stopped by the program when the error level (RMSE) meets a pre-defined target goal.
3. **Maximum number of epochs** - Training is stopped when a set number of training epochs has been reached.
4. **Cross-validation** - An independent data set is used to stop training. As training proceeds, error values are tracked for both the training data set and the validation data set. Training is stopped when the validation set error starts to increase. This is an independent data set, usually a different year than was used for training or testing. If training continued, error levels for the training data set would continue to decrease, however the ANN would likely become over-trained and not perform as well on data sets other than the training year. For this project over-training was avoided using the cross-validation technique.

2.8.2 Testing phase (ANN Implementation)

After the network is trained, the weights and biases are fixed; the ANN can then be used to compute water level predictions for other data sets as a simple matrix operation. When using the MATLAB neural network toolbox this can be performed automatically after the training is finished. The trained network is applied to the new data with the following command: y = sim(net,ANNInput). The ANN is then applied to the testing set and statistics are computed to assess and compare the performance of the ANNs.

3 Literature Review

This literature review has three parts. First, the literature review starts by examining the different modelling strategies, past and present, developed for the prediction of water levels with a particular emphasis on their relevance for the present work. Second, the review focuses on one of these methods, artificial neural networks, and their broad use and application for forecasting environmental time series. Finally the review will concentrates specifically on the use of artificial neural networks for modelling of storm surges.

3.1 Water level forecasting methods

Several methods/models are currently being used to predict water levels, tides, and surges, these include:

- Harmonic analysis and prediction of tidal water levels
- Persistence model
- Statistical regression models (linear)
- Chaos Theory models (non-linear)
- Numerical models
- Artificial neural network Models (non-linear)
- Self organizing feature maps (artificial neural network)
- Neuro-Fuzzy Inference Systems
- Data fusion
- Ensemble averaging

These models are discussed in more detail below.

3.1.1 Harmonic model

One of the earliest uses of harmonic analysis to predict tides is reported by Thomson (1867) and expanded by Darwin (1883), Harris (1897), and Doodson (1921). The tidal signal is primarily the result of influences from the Sun and the Moon, combined with the response of the basin. These influences can be divided into separate harmonic constituents. A tidal constituent is composed of three parts: amplitude, phase and angular speed or frequency. The amplitude and the phase for each constituent can be determined from the water level records using least squares. Each constituent has a different angular speed, which is comprised

of integral multiples of the five basic astronomical speeds shown in column three of Table 3-1.

Table 3-1 Basic astronomic speeds

Symbol (1)	Description (2)	Speed (degrees/hour) (3)
T	Rotation of the Earth on its axis, with respect to the Sun	15.0
h	Rotation of the Earth about the sun	0.04106864
s	Rotation of the Moon about the Earth	0.54901653
p	Precession of the Moon's perigee	0.00464183
N	Precession of the plane of the Moon's orbit	0.00220641

After the phase and amplitude of each constituent is calculated, a synthetic tidal level can be calculated as the sum of harmonic terms:

$$Y(t) = A_0 + \sum_{i=1}^{N} h_i \cos(\omega_i t + \varepsilon_i) \tag{3.1}$$

where $Y(t)$ = water level at time t; A_0 = mean height; N = total number of constituents; h_i= amplitude of the constituent; ω_i = speed of the constituent; and ε_i = phase of the constituent. The angular speed is given in degrees/hour, and the phase in degrees. The number of constituents needed varies, depending on location and application. The National Oceanic and Atmospheric Administration (USA) typically use up to 37 constituents (Schureman, 1958), and Proudman Oceanographic Laboratories (UK) uses 26 for its Storm surge model (Flather and Williams, 2000). Ten of the largest constituents and their angular speed (degrees/hour) are given in Table 3-2 column 4. The amplitude and phase of each constituent (Table 3-2, columns 5 and 6) vary depending on the location.

Table 3-2 Major tidal constituents

Symbol	Description	Formula	Speed degrees/ hour	Amplitude	Phase (referenced from a fixed date/time)
(1)	(2)	(3)	(4)	(5)	(6)
M2	Principle Lunar	2T-2s+2h	28.984	Varies for Each Location	Varies for Each Location
N2	Larger lunar semi-diurnal	2T-3s+2h+p	28.439		
S2	Principle Solar semi-diurnal	2T	30.000		
K1	Lunisolar diurnal	T+h	15.041	Determined by Harmonic Analysis	Determined by Harmonic Analysis
L2	Smaller lunar semi-diurnal	2T-s+2h-p	29.528		
O1	Lunar diurnal	T-2s+h	13.943		
Sa	Solar Semi-annual	h	0.041		
NU2	Larger Lunar elliptic semi-diurnal	2T-3s+4h-p	28.512		
K2	Lunisolar semi diurnal	2T+2h	30.082		
Mm	Lunar Monthly	s-p	0.544		

Harmonic analysis can be used to reconstruct or "predict" water levels at any location to a fairly high degree of accuracy, but does not take into account meteorological influences, and is therefore ineffective for the prediction of storm surges. Ideally 18.6 years of tidal data is needed to include all of the major astronomic effects of tidal variation (The precession of the axis of the Moon's elliptical orbit around the Earth takes 18.6 years).

3.1.2 Persistence model

The persistence model is a very simple model that assumes the difference between the latest measured water level and the tidally forecasted water level will persist, or remain unchanged. Tissot et al.(2004) compared forecasts using the persistence model with the harmonic model and found that it can improve considerably upon harmonic forecasts in some locations. The persistence model is predictable in its shortcomings: always lagging meteorologically driven water level changes. The persistence model will always under-predict a rising storm surge and over-predict a falling surge event. The persistence model is included as a reference because of its excellent performance in low tidal range/ inland bay systems, such as the Texas coast studied by Tissot et al. (2004). The regions small tidal range leads to little tide/surge interaction, a significantly different setting than the North Sea.

3.1.3 Statistical regression models (Linear)

The mechanisms affecting water level changes include gravity, and meteorological forcings. Weather forcings such as wind are non-linearly related to water level changes. It is difficult to model such non-linear systems, but acceptable results can sometimes be obtained with linear approximations. Of all the papers reviewed, the most common linear forecast model was the AutoRegressive Integrated Moving Average (ARIMA) model (Sfetsos (2002), More and Deo (2003), Solomatine et al. (2000), and Bazartseren et al. (2003)). The ARIMA model is a generalization of an AutoRegressive Moving Average or (ARMA) model. The ARMA model consists of two separate models, an AutoRegressive (AR) model and a Moving Average (MA) model. The AutoRegressive model is essentially an Infinite Impulse Response filter (a fundamental element in digital signal processing). The impulse response is "infinite" because there is feedback in the filter. For example: if the input to the filter was a single impulse (a single "1" sample followed by many "0" samples), an infinite number of non-zero values will be output (the output signal will gradually decay and approach zero). The Moving Average (MA) part is essentially a Finite Impulse Response filter (also a fundamental element in digital signal processing). The impulse response is "finite" because there is no feedback in the filter. For example: if the input to the filter was a single impulse (that is, a single "1" sample followed by many "0" samples), zeroes will eventually come out after the "1" sample has made its way in the delay line past all the coefficients.

The following is a list of papers using statistical linear regression models (ARIMA, ARMA, or AR) in their comparison to artificial neural network models. The papers are reviewed individually in their appropriate sections and are shown here only to compare ARIMA methods with artificial neural networks.

- Sfetsos (2002) wind speed forecasts ARIMA comparison with artificial neural networks (see section 3.2.2).
- More and Deo (2003) wind speed forecasts ARIMA comparison with artificial neural networks (See section 3.2.2).
- Deo and Naidu (1999) wave height forecasts AR comparison with artificial neural networks (See section 3.2.3).
- Solomatine et al. (2000) Water level forecast ARIMA comparison with Chaos (See section 3.1.4).

- Bazartseren et al. (2003) Long-term (15-hour) water level forecast ARMA comparison with artificial neural networks and ANFIS. (See section 3.1.8).
- Bazartseren et al. (2003) Short-term (15-min) water level forecast ARMA comparison with artificial neural networks and ANFIS. (See section 3.1.8).

Linear models were outperformed by non-linear models in all of the above examples with exception of the (Bazartseren et al. 2003) where short-term (15-min) linear water level forecasts outperformed artificial neural networks. Perhaps this is because in very short-term forecasts the behaviour of the system is essentially linear.

3.1.4 Chaos theory models (non-linear)

Solomatine et al. (2000) were interested in predicting surge water levels in the coastal waters of the Netherlands near the entrance to the port of Rotterdam. Chaotic features were identified in the surge data and methods of chaos theory applied. The average mutual information function was used to estimate the time delay and the method of false nearest neighbour was used to find the extent of the input time series. Results showed that for short-term forecasts, the chaos theory models had a root mean square error of 2.3 cm for a 20-min to 1-hour forecasts, and 6.1 cm for a 3-hour forecast. Solomatine et al. (2000) concluded that non-linear chaos theory models performed much better than linear (autocorrelation and ARIMA) models, although they did not publish the linear results for comparison. The authors did not produce any forecasts intervals longer than 3 hours for comparison with other models, or show results for the other methods, making it difficult to fully assess their results.

3.1.5 Hydrodynamic Numerical models

Hydrodynamic numerical models are usually based on integration of several different models, such as atmospheric and hydrodynamic models. These models have to solve non-linear shallow water equations, bottom stress, boundary conditions, and bathymetry must be accurately known. In addition, current atmospheric pressure and wind fields need to be known to accurately predict storm behaviour. Numerical solutions have several advantages over analytical solutions. First, the equations are much more intuitive than statistical methods, and can solve for multiple locations at the same time. Unfortunately, numerical models require considerable computational resources, take long times to run,

require additional data for driving or forcing the model and detailed bathymetric models of the domain. Examples of hydrodynamic numerical models and there primary uses are shown in Table 3-3.

Table 3-3 Hydrodynamic numerical models and their primary uses.

Model Name / reference	Type of Model	Uses
United Kingdom Continental Shelf Model (CS3) (Flather and Proctor 1983)	Finite difference	Tidal circulation/currents and elevations
Dutch Continental Shelf Model (DCSM) (Verlaan et al. 2005)	Finite difference	Sediment transport / hydrodynamics 2D/3D /Water levels / Storm surges
Princeton Ocean Model (POM) (Blumberg and Mellor 1987)	Finite difference	Sediment transport / hydrodynamics 2D/3D /Water levels / Storm surges
Sea Lake an Overland Surges from Hurricanes model (SLOSH) (Jelesnianski et al.1984)	Finite difference	Define flood prone areas for hurricane evacuation
ADvanced CIRCulation (ADCIRC) (Gica and Teng 2002)	Finite Element	Model for Coastal Oceans, Inlets, Rivers and Floodplains

The continental shelf model (CS3 model) is currently used by the Storm Tide Forecasting Service in United Kingdom to predict tidal water levels and currents around the British Isles (Flather and Williams 2004). The model is a finite difference model based on the conservation of volume, also known as the continuity equation, and the law of conservation of momentum. Flather and Proctor (1983) presented these two equations respectively where Eq. 3.2 is the equation of continuity which represents conservation of volume resulting in a change of surface elevation, and Eq. 3.3 equates the acceleration of water (left side of the equation) to the forces acting on it (right side of equation). These forces are wind stress, and horizontal forces of surface atmospheric pressure.

$$\frac{\partial \zeta}{\partial t} + \nabla.(D\vec{q}) = 0 \tag{3.2}$$

$$\frac{\partial \vec{q}}{\partial t} + \vec{q}.\nabla\vec{q} - f\vec{k}x\vec{q} = -g\nabla\zeta - \frac{1}{\rho}\nabla p_a + \frac{1}{\rho D}(\vec{\tau}_s - \vec{\tau}_b) + A\nabla^2\vec{q} \tag{3.3}$$

where: t = time; ζ = elevation of the sea surface; $\vec{q}$ = depth mean current; $\vec{\tau}_s$ = stress on the sea surface; $\vec{\tau}_b$ = bottom stress; p_a = atmospheric pressure on the sea surface; D = total water depth; ρ = density of sea water; g = acceleration due to gravity; f = Coriolis parameter; $\vec{k}$ = unit vector in the vertical; and, A = coefficient of horizontal diffusion.

The wind stress in Eq. 3.4 is divided by water depth, indicating that wind stress becomes increasingly important in shallow water. Wind stress is parameterized for the mean wind velocity at 10 m above the surface

$$\vec{\tau}_s = C_D \rho_a \underline{W}|\underline{W}| \tag{3.4}$$

where W = wind velocity; C_D = drag coefficient; ρ_a = density of air. The current at mean depth is related to bottom stress by the following formula

$$\vec{\tau}_b = \kappa \rho \vec{q}|\vec{q}| \tag{3.5}$$

where κ = a friction parameter, assumed constant at 0.0025.
Initial boundary conditions are defined as a coastal boundary given by

$$q_n = 0 \tag{3.6}$$

where q_n is the component of depth mean current normal to the boundary or, open sea boundary where a radiation condition is used:

$$q_n = q_n^M + q_n^T + \frac{c}{h}(\zeta - \zeta^M - \zeta^T) \qquad \text{where} \quad c = \sqrt{gh} \tag{3.7}$$

Where q_n^M, ζ^M and q_n^T, ζ^T describe the input of meteorological (M) and tidal (T) origin. This allows perturbations generated within the model to propagate freely across the boundary.

The equations are solved by finite differences method using a time-stepping procedure within the area of interest, with p_a (atmospheric pressure at the sea surface) and W (surface wind velocity) defined as functions of time and position, and ζ (sea surface height) as a function of time and position along the open sea boundaries.

The present operational model is composed of a continental shelf ocean model (12 km grid with coverage extending from 48°N to 63°N and from 12°W to 13°E) forced by wind and pressure data from the Met Office's 50 km grid Mesoscale model of wind speed and direction at the 10 m level. The model is run four times a day for the Storm Tide Forecasting Service (STFS) to predict storm surge along the UK coastline.

The surge component of the water level is obtained by running the CS3 model twice. Initially the model is run with open boundary surge input, and meteorological and tidal forcing to produce a total solution (tide + surge). The model is then run a second time with tides only. The difference between the two runs ((tide+surge)-(tide only)) gives the surge estimate. Table 3-4 shows the

model results indicating the number of surge events over 1.0 m and error levels for years 2002, 2003, and a 13-year average at Sheerness (STFS 2004).

Table 3-4 CS3 Model results at Sheerness for years 2002 – 2003. Source: STFS (2004).

Year	No. of Surge Events > 1m	Mean Error (m)	Max. Positive Error (m)	Max. Negative Error (m)	RMSE (m)	SD of Errors (m)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2003/4	22	-0.13	0.15	-0.47	0.21	0.17
2002/3	14	-0.04	0.26	-0.43	0.19	0.20
13-Year Mean	24	-0.10	0.20	-0.51	0.20	0.18

The Princeton Ocean Model (POM) is a three-dimensional primitive-equation model that uses sigma coordinates in the vertical (sigma coordinate systems conform to natural terrain, which allows the model to better define boundary-layer processes), and curvilinear orthogonal coordinates and an "Arakawa C" grid scheme in the horizontal. The Arakawa C grid is a staggered grid, which give a much better dispersion of gravity waves which are a closer approximation to the real solution. One of the models features is its use of an embedded turbulent closure sub-model which yields realistic Ekman surface and bottom layers (flow direction rotates as one moves away from the boundary). The governing equations along with the boundary conditions are solved with finite difference techniques. The model integrates into its equations the three components of velocity, temperature, salinity, and turbulence (Blumberg and Mellor 1987).

Several independent groups have utilized the POM model for use in storm surge prediction. Liu et al. (2006) coupled the POM model with the Simulating Waves Near-shore (SWAN) model to produce a three-dimensional wave-current model that can be used for storm-surge and inundation prediction during hurricanes. Xie et al. (2003) utilized a mass conserving inundation and draining scheme into a three-dimensional hydrodynamic model (POM). Bargagli et al. (2002) used a high resolution pressure and surface wind model to serve as input to a wave model (WAM) and a shallow water model (POM-2D) for prediction of storm surge in the Northern Adriatic Sea. The numerical experiments that were performed cannot be considered as a forecast, but rather as a hindcast simulation. Results are shown in Table 3-5 for 2 domains; the first called "Arc2" is a polar, non-uniform grid covering the Adriatic Sea and a portion of the Mediterranean Sea, and the second is called "MED" a rectangular latitude–longitude grid with

non-uniform spacing encompassing the entire Mediterranean Sea. The model output was divided into a High frequency (>18h) and low frequency (<18h) components to separate the first Adriatic mode characterized by a period of 22 hours from modes of shorter periods. The models low-frequency (> 18h) output performed better than the high-frequency (< 18h) output, which was expected because the storm events were in the high frequency part of the water level spectrum. The Broader "MED" Domain model (which extends the integration domain over the entire Mediterranean basin) performed best in both the High and low frequency models. This could be due to the better representation of the barotropic oscillation of the Mediterranean Sea caused by pressure forcing.

Table 3-5 Princeton ocean model results for Venice, Italy, October 1998. Source: Liu et al. (2006).

Domain Type	Driving Met. Data	High Frequency >18hrs			Low Frequency <18Hrs		
		RMSE	Correlation Coefficient	Amp Ratio	RMSE	Correlation Coefficient	Amp Ratio
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Arc2	Bologna Limited Area Model	7.32	-0.12	0.19	16.49	0.61	0.81
MED	Bologna Limited Area Model	7.83	-0.16	0.20	12.12	0.91	0.80
MED	European Centre for Medium-Range Weather Forecasts Model	4.53	0.24	0.18	19.17	0.81	0.82

Arc2 – Domain covering the Adriatic basin and the central part of the Mediterranean Sea.
MED – Domain covering the entire Mediterranean Sea.

The Sea, Lake and Overland Surges from Hurricanes (SLOSH) model is a two-dimensional numerical model where the computed surge is designed to reproduce a long-period gravity wave. This model uses generalized shallow sea dynamic equations in curvilinear coordinates. The model utilizes a moving boundary (wet/dry) which corresponds to changes in tidal elevation and its principle use would be for shallow water flow in wide tidal-flat areas and/or areas with large tidal ranges (which is ideal for hurricane storm surge of low lying coastal areas).

The SLOSH model must be tailored to a geographical area before it can be run (terrain height, water-depth, water barriers, and river channels). Imbedded within SLOSH is a hurricane wind model. The user must input the storms central

pressure and radius of maximum winds. Finally the user must input the storm's latitude and longitude at 6-hour intervals from 48 hours before landfall until 24-hours after landfall.

SLOSH models have been used by several research groups studying storm surge flooding. Shi et al (1998) compared SLOSH model results to a fixed boundary model with 1/12 degree lat/long spacing. Both models were used to simulate the storm event of April 7, 1994 in Huanghe Delta in China. Results showed that the fixed boundary model over-predicted the water levels presumably because the un-realistic piling up of waters at the fixed vertical side-wall boundary, while the SLOSH model allowed the water to spill over to adjacent areas. Jelenianski et al. (1984) compared SLOSH models for nine storms in eight basins. A total of 542 tide gauge, staff gauge, and high water mark observations were compared to SLOSH forecasts, and found the SLOSH model was within +/- 20% for all significant surge heights.

The ADvanced CIRCulation (ADCIRC) model was developed by Luettich et al. (1992). ADCIRC can be used in a 3D or 2DDI (two dimensional Depth Integrated) mode. Elevation of the sea surface is obtained from the solution of the depth-integrated continuity equation in Generalized Wave-Continuity Equation (GWCE) form. Velocity is obtained from the solution of either the 2DDI or 3D momentum equations. ADCIRC can use conservative and non-conservative momentum equations. ADCIRC has been used to calculate a combined tide and storm surge forecast for sites along the US East and Gulf coasts for large storm surge events such as hurricane landfall (Luettich et al. 1996) and (Blain 1997).

Gica and Teng (2002) used an ADCIRC model to simulate a storm surge generated by Hurricane Iwa in Hawaii on November 24, 1982. Two different Wind models, the modified Rankine and the Holland models were applied to generate wind and pressure fields. The maximum recorded surge height at Nawiliwili was 0.42 m. The model simulated a maximum surge height for the same location of 0.29 m for both the modified Rankine and Holland models. The difference was most likely due to the wave set-up that was not included in the shallow water equations of ADCIRC. It is interesting to note that by using the barometric pressure recorded at Nawiliwili and the recorded pressure at the same time period a distance of 74km away, Gica and Teng (2002) calculated a static storm surge height of 0.32m. By comparing this value to the simulated surge height it is evident that the storm surge generated by Iwa was mainly due to the barometric

pressure drop. This is also probably due too the very steep bathymetric slopes around the Hawaiian Islands.

3.1.6 Artificial Neural Network Models

Several features make artificial neural networks attractive and valuable for forecasting applications. Artificial neural networks are data-driven, self-adaptive, and learn from examples. They also can capture subtle functional relationships even if the underlying relationships are unknown. Artificial neural networks can generalize, even if the sample data contains noisy information. Artificial neural networks are universal approximators, and can approximate any function to any desired accuracy. Artificial neural networks are non-linear, an advantage for long-term forecasting when the underlying mechanism is non-linear.

3.1.7 Self-organizing Feature Maps

Self organizing feature maps are a data visualization technique developed by Professor Teuvo Kohonen which reduces the dimensions of data through the use of self-organizing neural networks (Ultsch and Röske 2002). Kohonen networks consist of only two layers (input and output) with the major difference from conventional artificial neural networks, that the neurons in the output layer are connected. This allows Neurons with similar tasks to communicate over short pathways. This corresponds to an abstracting capability which suppresses unimportant details and maps the most important details along the map dimensions (Ultsch and Röske 2002).

3.1.8 Neuro-Fuzzy Inference Models

Neural networks can learn from data, but cannot be interpreted i.e. they are essentially black boxes to the user. Fuzzy Systems consist of interpretable linguistic rules, but they cannot learn. Combinations of the two systems can take advantage of the learning ability of the neural networks to create fuzzy systems from data. The learning algorithms can learn both fuzzy sets, and fuzzy rules, and can also use prior knowledge. This was one of the methods used by Bazartseren et al. (2003) in their study comparing artificial neural network, neuro-fuzzy, and statistical methods for forecasting short term water levels. Bazartseren et al. (2003) predicted water levels at two successive gauging stations up-stream on the Oder River. For the artificial neural network portion of the project, a three layer

artificial neural network was used with one input neuron, one hidden neuron and one output neuron. Training utilized the Levenberg-Marquardt algorithm, and validation vectors were used for early stopping to prevent overtraining. The neuro-fuzzy portion of the project utilized the Adaptive Neuro Fuzzy Interference System (ANFIS) model. The Auto Regressive Moving Average (ARMA) model and Auto-regressive exogenous (ARX) models were used as the linear models for comparison Table 3-6 shows the Neuro-Fuzzy model performances for different forecasting intervals. The ARMA model performed best for short forecast intervals (15 min and 2-hour) with a root mean square error of 0.38 cm. For longer forecast intervals (5-10 hours) non-linear models (artificial neural network and ANFIS) performed better than the linear models, with a root mean square error of 4.5cm.

Table 3-6 Neuro-Fuzzy Performances at Frankfurt (Oder River) root mean square error (cm). Source: Bazartseren et al. (2003).

Model	Forecast interval			
	15 min	2-hour	5-hour	10-hour
(1)	(2)	(3)	(4)	(5)
ANN	1.39	1.52	2.37	4.58
ANFIS	1.61	1.75	2.66	5.09
ARMA	0.38	1.08	2.81	9.25
ARX	1.27	1.61	2.59	6.05

ANN – Artificial neural network model; ANFIS-Adaptive neuro fuzzy interference system model; ARMA-AutoRegressive integrated moving average model; ARX-AutoRegressive eXogenous model

3.1.9 Data fusion

Data fusion combines forecasts from numerous models to provide a better solution than could be achieved from the use of a single source alone. The concept is analogous to the way that humans use a combination of senses and the ability to reason to improve their chance of survival.

See and Abrahart (2001) used data fusion to model river levels; combining results from four independent models (neural networks, fuzzy logic, statistical, and persistence models) to produce a single forecast output. This type of approach is effective when individual model residuals follow a consistent pattern of over and under prediction. The four individual model outputs are used as input to a single hidden layer of a feed forward network trained to produce a single final forecast. Four individual forecasting models were used.

- Hybrid neural network (HNN) model – uses a fuzzy logic model to combine two different types of neural networks: first a self-organizing map (SOM) and second a multi-layer perceptron (MLP).

- Rule based fuzzy logic (RBFL) model – uses simple linguistic rule-based fuzzy logic. Used 5 inputs with 3 fuzzy sets for each, producing 125 rules defining changes in water levels.
- Autoregressive moving average (ARMA) model – Box and Jenkins (1970) was fitted to the first 60% of the data and validated with the remaining 40%.
- Persistence model – the current value is used as the forecast value.

Four Data fusion strategies were constructed from the results of the individual models. Two statistical approaches were used:

- DF1 – the mean of the four individual forecasts was used as the final forecast.
- DF2 – the median of the four individual forecasts was used as the final forecast.

Two neural network solutions were trained to learn the optimal solution and then produce a weighted fusion.

- DF3 – Absolute values of each forecast was used as inputs to a feed forward network trained to forecast future river levels at t+6.
- DF4- the difference between the original forecasts of each model and the current value are used as inputs trained to forecast future river levels at t+6.

Results are shown in Table 3-7 for training and validation data sets. The Hybrid Neural Network did a very good individual performance with a correlation coefficient of 0.997. Only the data fusion models DF3 and DF4 were able to better this performance with a correlation coefficient of 0.999 each. Although data fusion technique used in this example did improve on individual models, the individual models would be much easier to implement, and the small amount of improvement would have to be weighed against the much larger amount of work.

Table 3-7 Correlation Coefficient statistics for individual models and data fusion solutions. Results are for both training (T) and validation (V) data sets. Source: See and Abrahart (2001).

Specific Method	Type of Solution	River Ouse	
(1)	(2)	T 60% (3)	V 40% (4)
HNN	Individual Model	0.997	0.995
RBFL	"	0.988	0.979
ARMA	"	0.992	0.989
Persistence	"	0.956	0.972
DF1	Data Fusion	0.992	0.987
DF2	"	0.992	0.987
DF3	"	0.999	0.999
DF4	"	0.999	0.999

HNN-Hybrid neural network model; RBFL-Rule based fuzzy logic model; ARMA AutoRegressive moving average model; DF(x); Data fusion model(s)

3.2 Artificial Neural Network Applications

An Artificial Neural Network (ANN) is an information processing method that is inspired by the way biological nervous systems, such as the human brain, process information. The concept of artificial neurons was first introduced by McCulloch and Pitts (1943), but little research was done in the field until the back-propagation algorithm was introduced (Rumelhart et al. 1986). The key element of artificial neural networks is the structure of the information processing system. It is composed of a number of highly interconnected processing elements called neurons, working in unison to solve specific problems. Artificial neural networks, like people, learn by example. An artificial neural network can be configured for a specific application, such as pattern recognition or data classification. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. Artificial neural networks work in a similar way with the artificial neural network parameters, weights and biases, being adjusted as part of the learning process, supervised or unsupervised. The use of neural networks for time series forecasting has been studied and proven successful for a number of cases. Chakraborty et al. (1992) used non-linear modelling of multivariate time series to predict future prices which consistently outperformed statistical models. Their work was shown to be quite useful in solving other problems in the fields of dynamical system modelling, recognition, prediction and control.

3.2.1 Environmental Systems

Neural networks have been applied successfully to a number of coastal and riverine cases, such as the forecasting of physical or water quality parameters.

Brion and Lingireddy (1999) used artificial neural networks with bacterial concentrations and weather data as inputs to predict the source (type) of water contamination of a drinking water reservoir for Lexington Kentucky. Fecal Coliforms/fecal streptococci ratios were measured at seven locations around the reservoir for a total of 106 days. The seven sampling sites were identified as urban, agricultural or a mixture of the two. Meteorological conditions of each sampling day were also recorded for each sample consisting of 0 or 1 for rain or non-rain. These data were used as input for the model. The identification of the source location was identified though case study and used as the desired output. If the source of contamination was urban runoff it was assigned the output value of 1.000. If the source of contamination was predominantly agricultural it was assigned the output value of 0.000. If the source was a mixture of urban and agricultural it was assigned the output value of 0.500. Table 3-8 shows an example of random selected data points as they were used for input into the artificial neural network model during training. 85 of the 106 data points were used for training and the remainder was used to verify the efficacy of the trained neural network model. The output of the predicted model is shown in Table 3-9. There are small differences between the predicted and expected values, which can be attributed to the small training set and the crude classification of rainfall.

Table 3-8 Example of training data used as input to the artificial neural network model. Source: Lingireddy (1999).

Rain	Log Fecal Coliforms	Log Fecal Streptococci	Known Output Type	Sampling Site Location
(1)	(2)	(3)	(4)	(5)
1.0	2.415	2.857	0.500	Overflow
1.0	2.602	3.017	0.000	Horse
0.0	2.602	2.892	0.000	Horse
0.0	2.000	1.678	1.000	Andover
1.0	1.740	1.752	0.500	Boat Dock
0.0	3.720	3.623	0.000	Jacobson

Table 3-9 Model predictions at all sites. Agricultural = 0; Urban = 1; Mixed = 0.5. Source: Lingireddy (1999).

Site Name	Source of Contamination	Expected Output	Model Prediction
(1)	(2)	(3)	(4)
Jacobson	Agricultural	0.000	0.090
Boat Dock	Mixed	0.500	0.660
Cattle pond	Agricultural	0.000	0.140
Horse	Agricultural	0.000	0.000
Andover	Urban	1.000	1.000
Squires	Urban	1.000	0.990

Recknagel et al. (1997) modelled coastal algal blooms using up to 11 water quality parameters as inputs nodes. The output layer consisted of cell numbers of dominating algal species. The artificial neural network used was a feed-forward with back propagation during training. The results showed that artificial neural network's can fit the complexity and non-linearity of ecological phenomena to a high degree. It is unclear in the reporting what the model's prediction method was (fore-cast, now-cast, or hind-cast)

Moatar et al. (1999) used river discharge and solar radiation level as inputs in an artificial neural network to estimate the daily pH levels of rivers. Under acidic conditions pH is directly related to flow. Under alkaline conditions the pH is principally related to photosynthesis. The authors needed a method for detecting instrumentation errors for pH measurement. A significant difference between the modelled pH and measured pH could be used for error detection (abnormal values, discontinuities, and recording drifts) of instrumentation used to monitor river conditions down-stream from nuclear power plants.

Moatar et al. (1999) used the classical Multilayer Perceptron Model. In a multilayer perceptron model, extra hidden layers, are used in addition to input and output layers the Levenberg-Marquardt algorithm was used for training, and validation as a method to prevent over-training. The artificial neural network model was compared to a linear Auto-Regressive Moving Average with eXogenous inputs (ARMAX) model. A comparison was made of the performance of the models using different individual variables as inputs. This was performed to see which variables had the best relationship to the predicted pH value. The results are shown in Table 3-10, and indicate that the artificial neural network performed the best using the discharge as input with a standard deviation of 0.34 compared to the ARMAX model which had a standard deviation of 0.50. Discharge also performed best in terms of the efficiency criterion. Next a comparison was made of the performance using different combinations of multiple variables. The model that performed best used discharge and solar radiation as a combination of inputs and are shown in Table 3-11. The artificial neural network model performed best with a standard deviation of 0.30 vs. 0.41 for the ARMAX model.

Table 3-10 Comparison of ARMAX and ANN using single input variables. Source: Moatar et al. (1999).

	ARMAX		ANN	
	Efficiency Criterion	Standard Deviation (pH units)	Efficiency Criterion	Standard Deviation (pH units)
(1)	(2)	(3)	(4)	(5)
Discharge	0.45	0.50	0.72	0.34
Discharge Log	0.69	0.39	0.71	0.34
Solar Radiation	0.37	0.53	0.42	0.53
Temperature	0.33	0.54	0.26	0.55

ARMAX-AutoRegressive moving average with eXogenous inputs; ANN-Artificial neural network

Table 3-11 Comparison of ARMAX and ANN using multiple input variables. Source: Moatar et al. (1999).

	ARMAX		ANN	
	Efficiency Criterion	Standard Deviation (pH units)	Efficiency Criterion	Standard Deviation (pH units)
(1)	(2)	(3)	(4)	(5)
Discharge, Solar	0.60	0.41	0.77	0.30
Discharge, Temp	0.61	0.41	0.73	0.34
Log(Discharge), Temp	0.73	0.31	0.73	0.33
Log(Discharge), Temp	0.74	0.33	0.77	0.31
Discharge, Solar, Temp	0.62	0.41	0.71	0.35
Log(Discharge), Solar, Temp	0.76	0.32	0.74	0.33

ARMAX-AutoRegressive moving average with eXogenous inputs; ANN-Artificial neural network

The artificial neural network model performed slightly better than the ARMAX time series approach. Mortar et al. (1999) found the best results were obtained when using an artificial neural network structure with two input neurons, three hidden neurons with one output neuron.

3.2.2 Wind

Sfetsos (2002) used artificial neural networks to forecast mean hourly wind speed data. The author compared auto-regressive integrated moving average (ARIMA) models with artificial neural networks. The Levenberg-Marquardt algorithm was used for training, and validation vectors were used as a early stopping method to prevent over-training. The artificial neural network structure which resulted in the least error used 2 input neurons, 3 hidden neurons, and 1 output neuron. This model used a structure identical to Moatar et al. (1999) and the same training algorithm (Levenberg-Marquardt). The comparison of performances showed that artificial neural networks had the smallest root mean square error of 0.72 m/s compared to 0.73 m/s for the ARIMA and 0.74 m/s for the persistence model.

More and Deo (2003) used artificial neural networks to forecast daily, weekly and monthly wind speed at two locations. The authors used a feed-forward neural network (FFNN) with two training methods, back propagation and cascade correlation. Also recurrent neural networks (RNN) were used using the Jordan Elman training algorithm. Recurrent neural networks (RNN's) are dynamic in structure and are fully interconnected. For every new input, the state of the previous network is used as a term in the current network. This makes the output a function of everything the network has seen so far. In other words the network behaviour is based on its own history. Results for the different training methods are shown in Table 3-12. The model with the best performance was the Recurrent Neural Network with a 4.3 % error. The feed forward neural network using cascade correlation training finished second with a 4.5 % error, and feed forward neural network using back propagation finished third with a 4.7% error. The linear ARIMA model finished last with a 5.9% error.

Table 3-12 Wind speed forecasting schemes and their performances. Source: More and Deo (2003).

Scheme Used	Network Type	Training Algorithm	Topology used	Percent Error (Predicted vs. Observed)
(1)	(2)	(3)	(4)	(5)
Neural Network	Feed Forward	Back Propagation	(2,3,12,3,1)	4.7%
Neural Network	Feed Forward	Cascade Correlation	(3,1)	4.5%
Neural Network	Recurrent	Jordan Elman	(3,6,7,6)	4.3%
Time series	ARIMA	ARIMA	(2,3665,2)	5.9%

ARIMA-AutoRegressive integrated moving average

It is interesting that the authors found optimum neural structures used 3 hidden layers and up to 12 hidden neurons. In contrast Tissot et al. (2001) found that 1 or 2 hidden neurons performed best for predicting water levels. This could be because of more complex driving parameters for wind forecasting than for water level forecasts.

Barbounis and Theocharis (2006) used locally recurrent neural networks for long-term (2-3 days or more) wind speed and power prediction in Eastern Crete, Greece using Greek National Meteorology Service forecasts as training. Three different architectures were used:

1. Multi-Layer Perceptron (MLP) using Infinite Impulse Response (IIR) filters (MLP-IIR),

2. Local Activation Feedback – Multi-Layer Network (LAF-MLN) and
3. Output Feedback Multi-Layer Network (OF-MLN).

Six different training methods were used; the first two are based on the gradient decent method of error minimization.

1. Back Propagation Through Time (BPTT), and
2. Real Time Recurrent Learning (RTRL).

The authors noted that these two training methods required long convergence times because of the small learning rates required, and often become trapped in local minima of the error function.

A second set of training methods were based on Recursive Prediction Error algorithm (RPE), and consist of four main algorithms.

1. Global Recursive Predictive Error algorithm (GRPE) where all weights are handled simultaneously
2. Decoupled Recursive Predictive Error algorithm (DRPE), a simplification of the global into a set of decoupled algorithms.
3. Neuron Linear Model RPE (NLM-RPE)
4. Neuron Non-Linear Model RPE (NNM-RPE)

The results are shown in Table 3-13 where the mean absolute error (MAE) and normalized mean square error (NMSE) are displayed in columns 3, 4, 5 and 6. The MLP-IIR Model performed best in all learning approaches. The best learning approach was GRPE, which had the lowest normalized mean square error of 0.21 m/s.

Table 3-13 Model performances forecasting wind power and speed, evaluated using Mean Absolute Error (MAE) and Normalized Mean Squared Error (NMSE), using 3 model types and four training methods. Source: Barbounis and Theocharis (2005)

Model Type (1)	Learning Approach (2)	Wind Power (MW)		Wind Speed (m s^{-1})	
		MAE (3)	NMSE (4)	MAE (5)	NMSE (6)
MLP-IIR	GRPE	1.12	0.21	1.95	0.25
	DRPE	1.22	0.26	2.09	0.28
	NNM-RPE	1.19	0.23	2.05	0.27
	NLM-RPE	1.27	0.28	2.15	0.30
	RTRL	1.40	0.31	2.37	0.35
	BPTT	1.39	0.32	2.40	0.36
LAF-MLN	GRPE	1.21	0.23	2.14	0.30
	DRPE	1.27	0.27	2.28	0.34
	NNM-RPE	1.28	0.28	2.25	0.34
	NLM-RPE	1.31	0.30	2.41	0.36
	RTRL	1.40	0.33	2.45	0.40
	BPTT	1.45	0.34	2.46	0.39
OF-MLN	GRPE	1.20	0.23	2.10	0.31
	DRPE	1.26	0.26	2.22	0.34
	NNM-RPE	1.27	0.27	2.27	0.35
	NLM-RPE	1.32	0.28	2.31	0.36
	RTRL	1.46	0.33	2.49	0.40
	BPTT	1.43	0.32	2.47	0.39

MLP-IIR-Multilayer perceptron using infinite impulse response filter; GRPE-Global recursive predictive error; DRPE-Decoupled recursive predictive error; NNM-RPE-neuron non-linear model recursive predictive error; NLM-RPE-Neuron linear model recursive predictive error; RTRL-Real time recurrent learning; BPTT-Back propagation through time; LAF-MLN-Local activation Feedback-multi layer network; OF-MLN-Output feedback-multilayer network

Their results showed that artificial neural networks trained using IIR-Recurrent neural networks are more difficult to train than traditional artificial neural networks with hidden layers using back propagation. When training Recurrent neural networks the training algorithm could become unstable; the error between the target and the output of the RNN may not be monotonically decreasing; the gradient computation is more complicated; there may be long-range dependencies and the convergence times may be long.

3.2.3 Waves

Deo and Naidu (1999) presented a technique to make real-time forecasts of wave heights directly from observed waves. The network they used is a feed forward back propagation. Input values varied from 3, 12, and 24-hours of previous measurements, forecast varied from 3 to 24 hours. back propagation, Conjugate gradient and Cascade correction algorithms were used for training and were configured as follows: Back propagation - 1 input neuron, 5 hidden neurons, and 1 output neuron; Cascade correction - 1 input neuron 0 hidden neurons and 1 output neuron; Conjugate gradient - 1 input neuron 2 hidden neurons and 1 output

neuron. One of the main foci of the paper was determination of training times for the different training algorithms. This may have been important at the time the paper was written (algorithms were tested on a PC-AT 486). The results showed that cascade correlation was the quickest training algorithm with a time of 4 seconds, and back propagation was the slowest with a performance of 62,050 seconds (17.2 hours). Results for model performance comparing the artificial neural network method and the auto-regressive method are shown in Table 3-14. The artificial neural network model performed better than the auto-regressive model in all three forecast intervals. The authors noted that the correlation coefficient did not vary significantly between the training algorithms, but the individual results for the different training algorithms were not published.

Table 3-14 Model performances forecasting wave heights. Comparison of artificial neural network vs. auto-regressive models using correlation coefficient. Source: Deo and Naidu (1999).

Model	Forecast Interval		
	3-hour	12-hour	24-hour
(1)	(2)	(3)	(4)
ANN	0.81	0.78	0.71
AR	0.78	0.72	0.70

ANN-Artificial neural network; AR-AutoRegressive

Deo et al. (2001) used wind speed and direction to forecast wave height and period. The experiment was conducted at three locations, the first used an offshore wave buoy and a land based wind recording station, the second used wave rider buoy with on-board wind recording, the third utilized TOPEX satellite radar altimeter derived wave heights. The artificial neural network used was a three layer feed forward architecture. Best results were found when using 2 inputs nodes (current wind speed and previous time step wind speed); the number of hidden neurons was fixed at 4. Two output neurons were used for forecasted values of significant wave height and wave period. The three different algorithms were used for training were back propagation, cascade correction, and conjugate gradient. Individual results for the training methods were not published. The final results are given in Table 3-15. The TOPEX satellite performed best with a correlation coefficient of 0.77. The wave rider buoy performed second best with a correlation coefficient of 0.68. The land based anemometer performed the worst with a correlation coefficient of 0.52. The results are a little misleading because the TOPEX model was forecasting a weekly mean instead of a 3-hour forecast.

Table 3-15 Model performance for forecasting wave heights using correlation coefficient. Source: Deo et al. (2001).

Wind measurement location (1)	Wave Height (Hs) (2)	Wave Period (Tz) (3)
Land based anemometer and wave buoy	0.52	0.55
Wave rider buoy with wind speed and direction	0.68	NA
TOPEX satellite (Weekly Mean)	0.77	NA

Makarynskyy (2004) used two approaches to predict wave heights and periods. The first approach used eight separate neural networks to forecast wave parameters, varying the amount of input neurons from 2 to 4 then 8 neurons, a single forecast interval at a time. The second approach used a fixed artificial neural network structure two neural networks to simultaneously forecast 4 intervals (1, 6, 12, and 24-hours). Table 3-16 shows the results of the first approach. For wave-heights using the 8 previous hours of measurements resulted in best short-term forecasts (3 and 6-hours) but for long-term forecasts (12 and 24-hours) 4 previous hours of measurements performed best, with forecasts using only 2 previous hours of measurements performing worst. For wave periods using only the last 2 hours of previous measurements performed best in all forecast intervals, with networks using the last 8 hours of measurements performing worst.

Table 3-16 Model performances for forecasting wave heights and periods while varying network structure. Source: Makarynskyy (2004).

Forecast Period (1)	Network Structure (2)	Wave Height RMSE(m) (3)	Wave Height R (4)	Wave Period RMSE(s) (5)	Wave Period R (6)
+3	2 x 5 x 1	0.15	0.97	0.82	0.91
+6	2 x 5 x 1	0.21	0.94	1.11	0.83
+12	2 x 5 x 1	0.34	0.84	1.39	0.72
+24	2 x 5 x 1	0.53	0.69	1.69	0.54
+3	4 x 9 x 1	0.15	0.97	0.83	0.91
+6	4 x 9 x 1	0.21	0.94	1.11	0.83
+12	4 x 9 x 1	0.31	0.89	1.45	0.70
+24	4 x 9 x 1	0.53	0.74	1.71	0.55
+3	8 x 17 x 1	0.14	0.97	0.88	0.89
+6	8 x 17 x 1	0.20	0.94	1.13	0.81
+12	8 x 17 x 1	0.37	0.82	1.56	0.64
+24	8 x 17 x 1	0.77	0.40	1.93	0.41

In a second approach Makarynskyy (2004) used several different artificial neural network structures, varying the number of inputs and the number of nodes in the hidden layer. The optimum network structures were 6 input neurons 17 hidden neurons and 4 output neurons (wave height) and 12 input neurons, 21 hidden neurons and 4 output neurons for wave period. The results are shown in Table 3-17. This model produced 4 simultaneous forecasts from a single run of

the network. The network was trained using three different training algorithms, the resilient back-propagation (RBP), conjugate gradient (CG), and the Levenberg-Marquardt (LM) training algorithms. Comparison of the results for wave heights show the conjugate gradient method worked slightly better for all forecast intervals (RMSE and R). For wave periods the conjugate gradient method worked slightly better for 3, 6 and 24-hour forecasts.

Table 3-17 Model performances for forecasting wave heights and periods for 4 simultaneous intervals (Wave heights use (6x17x4) network, Wave periods use (12x21x4) network.). Source: Makarynskyy (2004).

Training Method (1)	Forecast Period (hours) (2)	Wave Height		Wave Period	
		RMSE(m) (3)	R (4)	RMSE(s) (5)	R (6)
Resilient Back-	+3	0.20	0.95	1.03	0.87
Propagation	+6	0.23	0.93	1.30	0.79
	+12	0.66	0.64	2.01	0.49
	+24	0.91	0.38	2.33	0.23
Conjugate	+3	0.17	0.96	0.89	0.89
Gradient	+6	0.23	0.93	1.30	0.75
	+12	0.46	0.79	2.16	0.35
	+24	0.75	0.53	2.31	0.23
Levenberg-	+3	0.21	0.95	1.08	0.85
Marquardt	+6	0.29	0.90	1.50	0.73
	+12	0.59	0.72	2.84	0.31
	+24	0.91	0.50	3.51	0.17

Examination of the results for the two tests indicates that, an artificial neural network with one output neuron forecasting one interval will perform better than an artificial neural network with four output neurons forecasting 4 simultaneous intervals.

3.2.4 River flood forecasting

Thirumalaiah and Deo (1998) used artificial neural networks to forecast river stage at a specific target station. Two separate artificial neural networks were used, the first used data from the target station, and the second used data from a station up-stream of the target station to forecast water levels at the target station. The network structure used 1 input neuron, 3 hidden neurons and 2 output neurons. The two output neurons produced a 24 and 48-hour forecast. Three different training algorithms were used, Back Propagation, Conjugate gradient, and Cascade correlation. Table 3-18 shows that when a station used its own input for training, that Cascade correlation provided the best training method, with a value of 0.973 for a 24-hour forecast, and 0.885 for a 48-hour forecast. Better results were found when using a station upstream from the target station for training, with

a correlation coefficient of 0.995 for a 24-hour forecast, and 0.980 for a 48-hour forecast when using the Back-propagation training algorithm.

Table 3-18 Model performances for forecasting river stage using correlation coefficient using different training algorithms. Source: Thirumalaiah and Deo (1998)

Training Algorithm used	ANN using data from Target station		ANN using data from station Up-stream from Target	
	24-hour forecast	48-hour forecast	24-hour forecast	48-hour forecast
(1)	(2)	(3)	(4)	(5)
Back propagation	0.875	0.950	0.995	0.980
Conjugate gradient	0.957	0.895	0.954	0.981
Cascade correlation	0.973	0.885	0.956	0.885

Data sets up-stream from the target station provided a better view of future conditions because they measured actual "future water" coming towards the target station instead of a forecast derived from measurements of water that has already flowed past the target station. This introduces spatial information into the model, using knowledge from one location to influence results at a second. The numbers in the tables are slightly counter intuitive because several 48-hour forecasts have better results than 24-hour forecasts. The authors also reported different numbers in the text as compared to the tables.

Kim and Barros (2001) used neural networks to forecast flooding along rivers. This model uses rain-gauge, stream-flow, radiosonde, and satellite data. Several pre-configured artificial neural network models were used. The artificial neural network selection depended on the current weather classification. The model constructed uses 3 different modules. The first module (#1) uses a network of 160 rain-gauges to monitor for rain. Next a classification and decision module (#2) is used to classify current weather conditions using satellite or radiosonde data. A second rain/no rain forecast is issued based of the likelihood of rainfall occurring within the region during the forecast interval. If positive a specific neural network is chosen depending upon regional weather conditions. Finally a forecast module (#3) selects one of four different neural networks, each configured for a specific weather classification as shown in Table 3-19.

Table 3-19 Criteria used for selection of neural network model. Source: Kim and Barros (2001).

Model (1)	Current weather Conditions and rain gauge state (2)
C4	Convective weather systems are present and 100% of predictor rain gauges are wet.
C3	Convective weather systems are present and 75% of predictor rain gauges are wet.
R4	Convective weather systems are not present and 100% of predictor rain gauges are wet.
R3	Convective weather systems are not present and 75% of predictor rain gauges are wet

Five quantitative measures of forecast skill were used to evaluate the model performance.

- Skill Score (SS) defined as percentage reduction in mean-squared error with respect to the persistence forecasting method.
- The correlation coefficient (CC) between forecasts and observations.
- The bias defined as the degree of correspondence between the mean forecast and the mean observation.
- The root mean square error (RMSE), defined as the sum of the square of the differences of the forecasts and the observations.
- The threat score, a categorical verification measure equal to the number of correct event forecasts divided by the total number of forecasts.

Table 3-20 shows values for forecast skill at Williamsburg, three other locations were included in the study. The results show large percentage reduction in mean-squared error in respect to the persistence model (from 96% with the C4 model to 32% with the R3 model). This corresponds to higher reduction when rain is likely (C4 classification – convection systems present and 100% predictor gauges are wet) less reduction when rain chances are less (R3 classification – convection systems not present 75% of predictor gauges are wet). This project shows that expert systems such as neural networks can lead to significant gains in forecasting rainfall and flood events.

Table 3-20 Statistical performance measures for forecast skills at Williamsburg. Source: Kim and Barros (2001).

Name	Model	RMSE	SS	CC	Bias	Threat scores				
						25%	20%	15%	10%	5%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Williams-burg 18-hour Forecast	C4	11	0.96	0.97	3	0.79	0.94	1.00	1.00	0.71
	R4	33	0.62	0.62	-1	0.69	0.63	0.60	0.47	0.50
	R3	23	0.32	0.32	5	0.71	0.69	0.64	0.58	0.38
	Com-bination	23	0.54	0.54	3	0.72	0.71	0.67	0.60	0.44
Williams-burg 24-hour Forecast	C4	28	0.48	0.65	3	0.85	0.90	0.89	0.84	0.57
	R4	20	0.17	0.71	5	0.75	0.69	0.64	0.73	0.40
	R3	22	0.25	0.68	3	0.67	0.59	0.54	0.59	0.38
	Com-bination	24	0.36	0.65	4	0.71	0.67	0.62	0.64	0.39

3.3 Use of Artificial Neural Networks in Sea Level Forecasts

For water level modelling, Tsai and Lee (1999) used neural networks to predict hourly tidal levels up to one year using only fifteen days of training at Tanshui Harbour in Northern Taiwan. The authors claim that "hourly tidal levels over a long duration can be efficiently predicted using only a very short-term hourly tidal record". Results are interesting but the interval for each forecast is only one hour and has limited applicability for forecasting purposes. A major shortcoming in their paper is that the authors say they are predicting tidal levels over these long intervals, when in fact they are only predicting a 1-hour forecast. For example, the authors state they used one day (24 hours of data) of training data to predict water levels for one month (720 hours), but for each new prediction interval, their model makes requires one additional hour of information provided by the observation data. Hence, the model only predicts water levels one hour in the future, and never actually predicts water levels 30 or 365 days in the future as stated by the author.

Tissot et al. (2001) modelled water levels along the Texas Gulf Coast region. Tidal and weather induced water level components were modelled separately. Since the tidal components for a specific location can be modelled quite accurately, they were removed from the model, which then only required a simple artificial neural network structure for accurate water level forecasts. This simplification allowed a more direct relationship between short-term forcing and changes in water levels. This model compares quite differently with models used by Makarynskyy et al. (2004), who used total water levels (combining tide and

surge components) which required large artificial neural network structures (using 25-145 hidden neurons) to predict total water levels in Western Australia. The model by Tissot et al. (2001) compared very similarly to those used by Sztobryn (2003), who predicted surge levels on the tide-less portion of the Baltic Sea. Sztobryn (2003) found that the optimum model used 3 layers with 2-6 hidden neurons.

Tissot et al. (2001) used several variables as inputs for training the artificial neural networks including: water levels, wind stress, barometric pressures tidal forecasts, and wind forecasts. The optimized model used for the study was found to use: 1 hidden neuron, 1 output neuron, included 5-hours of wind speed, wind direction, water level measurements, 20-hours of barometric pressures and a wind forecast at the time of the forecast.

The authors found that although Artificial neural networks using water levels only (no winds or barometric pressure) initially performed substantially better then Harmonic forecast water levels (for forecasts from 1 to 24-hours), they approached similar error levels after increasing the forecast time to 48-hours. Substantial improvement in the artificial neural network model was then found after including wind speed and direction as an additional input at the forecast time, even when including forecast intervals of up to 48-hours.

In a related application along the south shore of Long Island, New York Huang et al. (2003) developed a regional neural network to predict water levels at a temporary location based on water levels measured at permanent National Oceanic and Atmospheric Administration (NOAA) tidal stations located about 60 km-100 km away from the inlets. The model was developed to re-construct long-term historic water levels using remote temporary sea level measurement stations and provided very good results for both tidal and non-tidal historical water levels.

Sztobryn (2003) was looking to improve upon existing numerical models for the forecasting of sea levels in the Baltic Sea. Tides are virtually non existent in the region, enabling storm-surge models to be applied directly to total water levels. While their existing models performed well during average conditions, performance was poor during storm events. Four different artificial neural network models were tested and are described below:

- Radial Basis Function (RBF) (22 neurons in hidden layer);

- Generalized Regression (GRNN) (Number of neuron varies);
- Multilayer Perceptrons with 3 layers (MLP3) (2-6 neurons in hidden layer);
- Multilayer Perceptrons with 4 layers (MLP4) (8 neurons in first hidden layer, 7 in second).

The testing data was divided into 3 sets used for training, validation, and testing. Results from 3 alternative methods (Method 1, Method 2 and Method 3) were given, but descriptions were not included in the paper.

Two experiments were run. The first experiment used a continuous data series. This experiment used the following variables for the input data vector: the daily mean sea level value from the previous day, and the 6-hour forecast of wind-speed and direction generating sea level changes in the western part of the Polish coastal waters. The results are shown in Table 3-21 and indicate the neural network methods resulted in correlation coefficients ranging from 0.80 to 0.82, while the previous alternative methods gave values between 0.67 and 0.86. RMSE errors for the previous forecasting methods were from 17 – 42 cm while ANN forecasts did not exceed 14 cm.

Table 3-21 Model performances for forecasting sea level heights using continuous data series, evaluated using correlation coefficient (R) and root mean square error (RMSE). Source: Sztobryn (2003).

Training Algorithm Used	Sea Level Values			Correlation Coefficient R (m^2)	Root mean square error RMSE (cm)
	Max (cm)	Min (cm)	Mean (cm)		
(1)	(2)	(3)	(4)	(5)	(6)
Observed	562	432	497	1	-
Method 1	558	460	504	0.67	35
Method 2	573	467	500	0.73	17
Method 3	571	476	523	0.86	42
MLP3	542	467	496	0.82	13
RBF	551	458	493	0.80	13
MLP4	547	461	494	0.81	14

MLP-Multilayer perceptron; RBF-Radial basis function

The second experiment used data from 150 storm events. These events were selected by (1) applying the standard Holland genetic algorithm, (2) calculating the correlation coefficients, and (3) by checking the sensitivity of the model to successive input data introduction, by first checking the quality of the performance the model with the examined data series, then without. The results are shown in Table 3-22 were the correlation coefficient for ANN methods vary from 0.27 to 0.58, while the previous alternative methods values range from 0.22

to 0.48. The RMS values for the ANN ranged from 13 to 34 cm, and 10 to 39 cm for the previous alternative methods. The model that gave the best performance was the multilayer perceptron model with 3 layers and 2-6 hidden neurons. This is the same optimum structure found by Cox et al. (2003) for neural network structures predicting water levels using the surge component only.

Table 3-22 Model performances for forecasting sea level heights using data from 150 storm events, evaluated using correlation coefficient (R) and root mean square error (RMSE). Source: Sztobryn (2003).

Training Algorithm Used	Sea Level Values			Correlation Coefficient R (m^2)	Root mean square error RMSE (cm)
	Max (cm)	Min (cm)	Mean (cm)		
(1)	(2)	(3)	(4)	(5)	(6)
Observed	600	565	580	1	
Method 1	575	571	573	0.22	103
Method 3	542	541	542	0.48	39.3
MLP3	581	566	571	0.58	12.7
RBF	620	548	578	0.27	23.4
MLP4	650	544	586	0.31	34.1

MLP-Multilayer perceptron; RBF-Radial basis function

Tissot et al. (2003) used a three step procedure to optimize an artificial neural network used to predict storm-surge levels at Bob Hall Pier on the Texas Coast near Corpus Christi. The first step was to determine the optimum number of previous water level differences used to forecast storm surge. They found that as the forecasting time increases, the importance of previous water levels decreases. The second step in the optimization process was the inclusion of past wind measurements. For 3-hour forecasts, including the last 6 hours of wind records leads to the best performance. For forecasts longer than 3 hours, only 3 hours of past wind records are needed. The third step of the optimization process was to add wind forecasts. Wind forecasts were extracted from the National Centre for Environmental Predictions (NCEP) Eta-12 model, where they are provided in 3-hour increments. Significant improvements were found when including forecast winds when the forecast interval is 12 hours or greater; however the overall computational efficiency of the model is significantly impacted by including atmospheric model predictions. For 24-hour forecasts, the model performance as measured by average absolute error improves from 6.4cm to 6.0 cm, or a 6% improvement. Finally the number of neurons in the hidden layer was varied from 1 to 3. The increase in the number of hidden layer did not lead to significant

increases in performance and in fact the performances decreased in all cases when testing ANN's with 3 neurons in the hidden layer.

Makarynskyy et al. (2004) computed total water levels including both tidal and meteorological components as output from a single artificial neural network. A saliency analysis was performed to determine the optimum number of neurons to be used. This technique allows an estimation of the relative importance of the input and processing nodes of the network. This resulted in the various size networks as shown in Table 3-23. Because this technique is predicting total water levels (a combination of astronomical and meteorological signals) a relatively large artificial neural network structures were required just to interpret the tidal portion of the signal. This is unlike artificial neural network structures used by Cox et al. (2002), who found that very simple structures (one or two input neurons and one hidden neuron) were the optimal size needed to accurately forecast storm surge levels. This is because the Cox et al. (2002) separated the harmonic and surge components which greatly simplified the signal, and therefore required a much simpler artificial neural network to interpret it. Makarynskyy et al. tested artificial neural networks varying sizes, using up to 72 inputs, 145 hidden and 24 output neurons. The models used log-sigmoid for the transfer function in the hidden layer, and linear transfer function in the output layer. The results are shown in Table 3-23 where R = the correlation coefficient, RMSE = root mean square error, and SI = scatter index. The best model used 24 input neurons, 49 Hidden neurons and had a RMSE of 111mm. For this study it was determined that an overcomplicated network can degrade the quality of the forecasts.

Table 3-23 Model performances for forecasting sea levels while varying number of input and hidden neurons. Source: Makarynskyy et al. (2004).

Input Neurons (1)	Hidden Neurons (2)	Averaged R (3)	Averaged RMSE(mm) (4)	Averaged SI (5)
72	145	0.816	139	0.195
60	121	0.830	130	0.182
48	97	0.859	118	0.165
36	73	0.859	119	0.166
24	49	0.875	111	0.155
12	25	0.823	122	0.172

The method of using artificial neural networks for forecasting total water levels may be desirable for cases when a harmonic analysis has not been performed. Because of their large size these models can be time consuming to

train, and are susceptible to poor generalization, especially when training relies on relatively small data sets.

Tissot et al. (2004) used tidal measuring stations along the Texas coast to predict storm surge events for inland bay regions. Artificial neural networks were used to predict surge events for two Texas bay systems, Corpus Christi Bay and Galveston Bay. Bob Hall Pier tide gauge is located in the Gulf of Mexico on a barrier island protecting Corpus Christi Bay. Any storm surge event occurring in the Gulf of Mexico will be measured or detected at the Bob Hall Pier gauge first, before moving into Corpus Christi Bay (Location of the Packery Channel gauge) or similarly at the Pleasure pier at Galveston before moving into Galveston Bay. The tidal lag period for The Bob Hall Pier to Corpus Christi Bay (Packery Channel gauge) is approximately 4 hours and The Galveston Bay time lag is approximately 7 hours.

One of the principle skill assessments used by the National Ocean Service (NOS) to assess the adequacy of water level forecasts is the Central Frequency of 15cm (CF15). For a model to be considered operational it must perform equal to or greater than 90% of the time within +/-15cm. Table 3-24 shows the central frequency for stations used in this study. Initially, none of the stations satisfied the NOS's 90% central frequency criteria using sea level data alone.

Table 3-24 Surge level forecast central Frequency (Percentage of time forecasts values are within 0.15 m of observed values) for selected Texas tide stations. Source: Tissot et al. (2004).

Bay System (1)	Station (2)	Station Description (3)	Central Frequency (%) (4)
Corpus Christi	Bob Hall Pier	Gulf of Mexico Station	84.2
	Port Aransas	Ship Channel Station	83.7
	Packery Channel	In-bay Station	85.1
Galveston Bay	Pleasure Pier	Gulf of Mexico Station	72.8
	Morgans Point	In-bay Station	67.3

To improve the models Tissot et al. (2004) included secondary water level measuring stations and winds to be used as additional input variables, depending on the station selected. The results of this optimization are shown in Table 3-25. For example, when using Bob Hall as a primary station Galveston Pleasure Pier was added to help with frontal boundaries arriving from the North. When using Packery Channel as a primary station Bob Hall Pier is added because of its strong correlation and long lag period with Corpus Christi Bay waters. When using

Morgan's point as a primary station, Pleasure Pier is added because of its strong correlation and long lag period with Galveston Bay waters.

Table 3-25 Secondary station location and optimum input time series used for each primary station location. Source: Tissot et al. (2004).

Primary Station (1)	Forecast Interval (2)	Secondary Station (3)	Input Time Series (4)
Bob Hall Pier	24	Pleasure Pier	Previous water levels and winds for both stations (48 hrs)
Bob Hall Pier	48	Pleasure Pier	Previous water levels and winds for primary station (past 48 hrs)
Port Aransas	24	None	Previous water levels (48 hrs)
Port Aransas	48	None	Same as 24 hr forecasts
Packery Channel	24	Bob Hall Pier	Previous water levels at primary station (4 hrs) and previous water levels and winds at secondary station (24 hrs)
Packery Channel	48	Bob Hall Pier	Same as 24 hr forecasts
Pleasure Pier	24	None	Previous water levels (48 hrs)
Pleasure Pier	48	None	Same as 24 hr forecasts
Morgans Point	24	Pleasure Pier	Previous water levels at primary station (24 hrs) and previous water levels and winds at secondary station (3 hrs)
Morgans Point	48	Pleasure Pier	Same as 24 hr forecasts

After inclusion of secondary stations and wind as inputs to the artificial neural network, significant improvements in Central Frequency 15cm is evident. ANN model performances are compared to harmonic, Persistence, and Linear Regression models for each of the stations and are shown in Table 3-26. Stations in the Corpus Christi Bay system show the best improvement with all ANN models performing above 90 % for Central Frequency. Stations in the Galveston Bay area also show significant improvement with the models for both stations now reporting a Central Frequency (15cm) of over 80%.

Table 3-26 Comparison of model performances after optimization (ANN model performances are shaded grey) for selected Texas locations. Source: Tissot et al. (2004).

Station / Model (1)	Root mean square error (RMSE) (2)	Central Frequency (15cm) % (3)
Bob Hall Pier		
Harmonic	0.114	84.2
Persistence (24 hr)	0.086	92.0
Linear Regression (24 hr)	0.224	93.2
ANN (24 hr)	0.075	94.6
Persistence (48 hr)	0.114	84.3
Linear Regression (48 hr)	0.188	76.6
ANN (48 hr)	0.102	88.2
Port Aransas		
Harmonic	0.112	83.7
Persistence (24 hr)	0.075	94.2
Linear Regression (24 hr)	0.172	94.8
ANN (24 hr)	0.070	95.7
Persistence (48 hr)	0.103	87.0
Linear Regression (48 hr)	0.208	89.8
ANN (48 hr)	0.093	89.9
Packery Channel		
Harmonic	0.108	85.1
Persistence (24 hr)	0.055	97.5
Linear Regression (24 hr)	0.101	97.3
ANN (24 hr)	0.044	99.2
Persistence (48 hr)	0.078	93.0
Linear Regression (48 hr)	0.124	93.3
ANN (48 hr)	0.068	96.0
Pleasure Pier		
Harmonic	0.149	72.8
Persistence (24 hr)	0.146	79.6
Linear Regression (24 hr)	0.149	83.7
ANN (24 hr)	0.123	84.6
Persistence (48 hr)	0.172	71.2
Linear Regression (48 hr)	0.137	71.5
ANN (48 hr)	0.140	79.0
Morgans Point		
Harmonic	0.174	67.3
Persistence (24 hr)	0.178	71.2
Linear Regression (24 hr)	0.110	55.6
ANN (24 hr)	0.142	80.4
Persistence (48 hr)	0.219	61.6
Linear Regression (48 hr)	0.173	64.5
ANN (48 hr)	0.175	71.3

3.4 Ensemble averaging

A major method for improving ANNs is the use of ensemble averaging. Ensemble forecasting utilizes the results from multiple training runs to create different sets of weights and biases. After the training period, prediction is deterministic. Naftaly et al. (1997) demonstrated that averaging the results from multiple training runs can have a significant effect in the reduction of variance of a forecast. They found that ensemble averaging requires a special training methodology and can be even

more effective when not combined with training constraints such as weight decay and early stopping.

It was found that typically the bias decreases and the variance increases as the number of training epochs is increased. When training neural networks, the variance arises from two terms. The first comes from inherent data randomness, and the second comes from the "non-identifiably" of the model, namely the fact that for a given set of training data, there may be several local minima of the error surface. During the training period one aims to stop when the sum of their error reaches a minimum. Since we are able to reduce the variance via ensemble averaging, one should therefore search for a point with a smaller bias (longer training time) as the optimal trade off as the ensemble predictor.

Naftaly et al. (1997) tested their models using sunspot statistic data gathered since 1700. Their training set consisted of the data from the years 1701 – 1920 and the test data was the years 1921 to 1955. The network used 12 inputs, 1 sigmoidal hidden layer with 4 neurons, and one linear output layer. This network was enlarged to form a simple recurrent network in which the input layer is increased by adding the previous point in the time series. Back propagation was used as the training algorithm. Average relative variance (ARV) was used as an error indicator and was defined as the Mean Square Error of the data set divided by its variance.

The training procedure of each ANN starts out with a given set of connection weights. The authors considered an ensemble of ANNs that differ only by these connection values. Since the number of possible networks is quite large, they developed a technique that allowed approximate averaging over the whole space. Their technique consisted of constructing groups of fixed number of networks, Q. Several groups were chosen with the same size Q, and then averaged, defining a finite size average. The authors then estimated the limit $Q \rightarrow \infty$. They regarded the specific choice of initial conditions to be equivalent to a random error added to the predictor, therefore expected the error to decrease as $1/Q$. A simple regression in 1/Q was then used to obtain this limit. The results are shown in Figure 3-1. The highest curve corresponds to $Q=1$ or the case of a single network. Below are the curves of the larger size ensemble networks, with the bottom curve extrapolated to $Q \rightarrow \infty$. The extrapolation method used is shown in Figure 3-2.

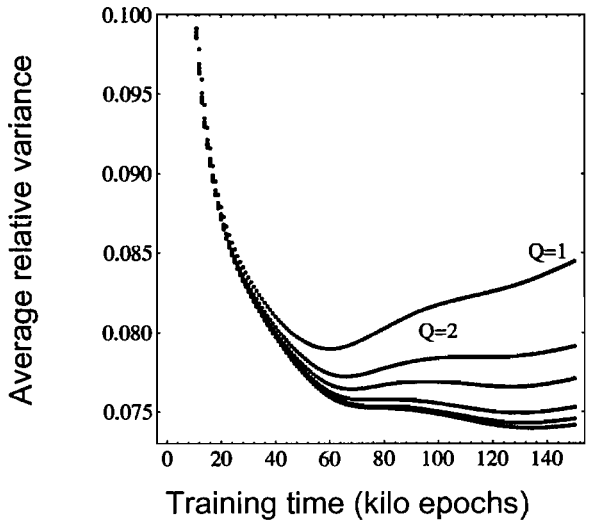


Figure 3-1 Average relative variance of test data set. Results show a shallow Q→∞ curve, and the two minima. The curves are shown for different size ensemble groups: Q = 1, 2, 4, 10, 20, and the lowest curve is the extrapolation to Q→∞. Figure adapted from (Naftaly et al. 1997).

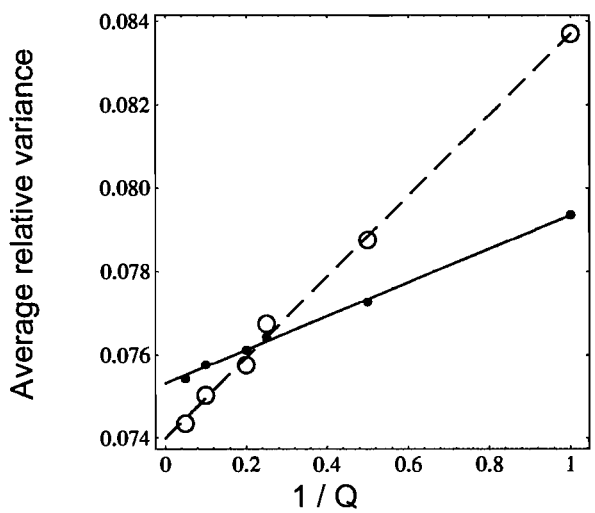


Figure 3-2 Extrapolation method used for extracting the Q→∞ prediction. The results for two different Q at two different training periods are shown. t = 70 kilo epochs (open Circles) and t = 140 kilo epochs (closed circles). Extrapolated values were obtained where each line crosses 1/Q = 0 on the horizontal axis. Figure adapted from (Naftaly et al. 1997).

The authors concluded that when a portion of the variance from the initial conditions is large, ensemble averaging is very effective, and by using large ensembles it can be eliminated altogether. The variance of the ensemble is inversely proportional to its size. The authors also concluded that early stopping may not be useful to the ensemble. In fact, that ensemble performance may improve if single ANN models are over-trained. Over-training reduces the bias portion of the error, while paying the price of higher variance for individual

networks. Later the variance portion of the error is reduced by the ensemble averaging, with no effect on the bias itself.

3.5 Summary

The review of the existing literature provided a guide for the direction and methods that were to be used for the project. The following are the primary guidelines developed after review of the literature.

- Artificial neural networks were selected as the modelling method because:
    - While numerical models provide simultaneous spatial data at any point within the model's domain, they require excessive information apart from historical observations and are complex and tedious to apply when point-forecasts at specific locations are needed.
    - Artificial neural networks can model non-linear phenomena.
    - Artificial neural network's can work with limited data sets (one or two years of historical data) (although ideally data should be broken up into training, validation, and testing data sets).
    - Artificial neural network's can accurately forecast surge levels using only historical and current water level data from a single station. There is no need for extensive hydrodynamic data sets to drive model.
    - Artificial neural network's can model data where bathymetry and meteorological data is poor or unknown. No need for bathymetric data sets describing the regional environment or wind forcing data sets. Although use additional data sets can significantly increase accuracy, if they include data that has a relationship to the parameter being modelled.
    - Spatial relationships for data inputs are not needed. The model does not need to know the distances between data stations, or if the numbers represent bacteria counts or storm water levels. The artificial neural network is able to establish a relationship between the input data sets (including tidal lag periods) without knowledge of position.
    - Artificial neural network's ability to model data where underlying relationships are unknown.
    - Artificial neural network's can model noisy data.

  - Computational efficiency when the method is utilized to provide real time predictions (once trained computing predictions is virtually instantaneous once the input is available)
- This project will model weather-induced water level changes only, using the same methodology of Tissot et al. (2001) and others. Some researchers forecast total water levels, requiring larger and more complex artificial neural networks to predict the harmonic components in addition to the meteorological component.
- For this project the artificial neural network will use a feed forward back propagation structure. The majority of the research projects reviewed used this type of artificial neural network structure. Its greatest strength is in non-linear solutions to ill-defined problems. The typical back-propagation network has an input layer, an output layer, and at least one hidden layer.
- Three layer artificial neural network was chosen because most artificial neural network's modelling similar environmental data sets had best results using 3 layers (1 input layer, 1 hidden layer, and 1 output layer)
  - This project will also investigate the forecast accuracy when varying number of neurons used in the hidden layer as previously investigated by Tissot et al. (2004) and Rajasekaran et al. (2005)
- Levenberg-Marquardt (LM) Algorithm for training. The LM algorithm was the primary training algorithm used for most of the artificial neural networks investigated, and the fastest for artificial neural network's trained with moderate number of network parameters.
- Two-station artificial neural network will be used (primary and secondary station) utilizing information from a source "up-stream" from the primary station to help forecast. This method of using a secondary station was investigated by Thirumalaiah and Deo (1998), Tissot et al. (2004), and Huang et al. (2003). This method takes advantage of storm propagation in the North Sea, since most storms track from north to south, allowing northern stations to "see" the effects of a storm before southern stations.
- Validation used as an early stopping method to prevent overtraining.
- Ensemble averaging can be used to significantly reduce variance in a forecast.

4 Materials

4.1 Data sources

Three different data sets are used in this project. The first is a sea level data set consisting of 9 tide gauge locations in the United Kingdom along the North Sea coastline. The second is also a sea level data set consisting of 4 tide gauge locations along the Thames Estuary. The last data set is composed of data extracted from a numerical model data set for the Sheerness tide station location from the CS3 model output. For an overview of the data sources used within the project organization see Figure 4-1.

4.1.1 North Sea Data - National Tidal and Sea Level Facility

The primary data set used for this project comes from the National Tidal and Sea Level Facility (NTSLF) through the British Oceanographic Data Centre (BODC). This data set contains a selected group of stations from the UK Tide Gauge Network. The network, run by the Tide Gauge Inspectorate, includes 45 gauges as illustrated in Figure 4-2. These gauges are tied together through a national leveling network to Ordnance Datum Newlyn for vertical reference (NTSLF, 2006). Tide data are collected, processed and archived by BODC. Tidal data sets were downloaded from the Internet via the BODC web site found at the following location: http://www.bodc.ac.uk/data/online_delivery/ntslf/ . The data is archived in 1-year data files at 1 hour intervals. The Cycle number, Date, Time, ASVBG02 (measured water-elevation), and residual (POL calculated harmonic water elevation – measured water elevation) values are logged at 10 minute intervals. Format for the National Tidal and Sea Level data sets are shown in Figure 4.

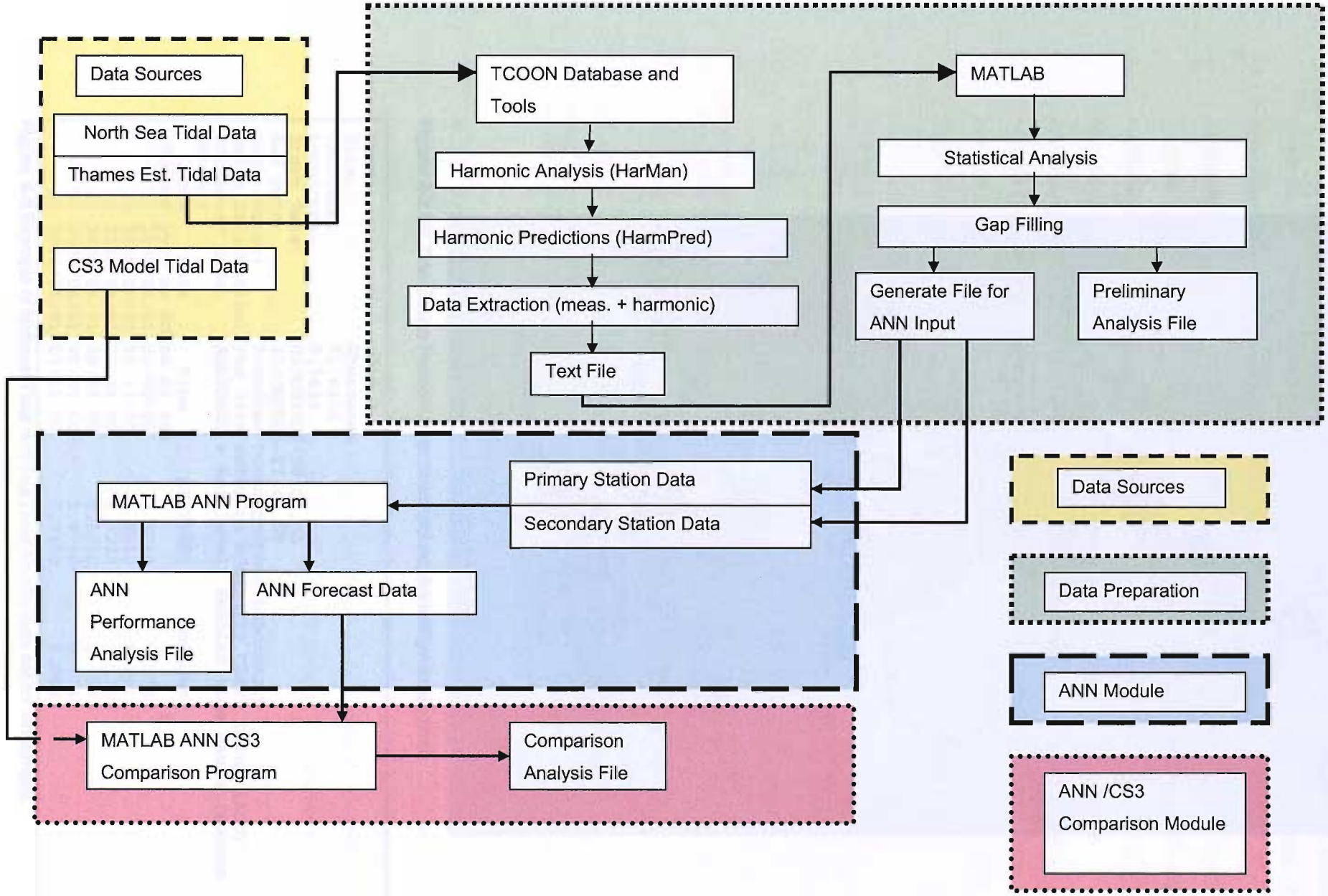


Figure 4-1 Research organization chart

Figure 4-2 UK Tide Gauge Network (http://www.pol.ac.uk/ntslf/tgi/ukmap.html)

```
Port:                P015
Site:                Sheerness
Latitude:            51.4456
Longitude:           0.7434
Start Date:          01JUN2006-00.00.00
End Date:            30JUN2006-23.45.00
Contributor:         Proudman Oceanographic Laboratory
Datum information:   The data refer to Admiralty Chart Datum (ACD)
Parameter code:      ASLVBG02 = Sea level, Bubbler tide gauge (second
sensor)
  Cycle     Date        Time       ASLVBG02      Residual
 Number yyyy mm dd hh mi ssf            f             f
      1) 2006/06/01 00:00:00       2.5760        0.2163
      2) 2006/06/01 00:15:00       2.8010        0.2011
      3) 2006/06/01 00:30:00       3.0230        0.1762
      4) 2006/06/01 00:45:00       3.2510        0.1521
      5) 2006/06/01 01:00:00       3.4790        0.1246
      6) 2006/06/01 01:15:00       3.7100        0.0988
```

Figure 4-3 Example of National Tidal and Sea Level Facility data set for Sheerness.

Data sets were obtained for the following stations: Wick, Aberdeen, Leith, North Shields, Immingham, Commer, Lowestoft, Felixstowe, and Sheerness. The completeness of the tide data sets is shown in Table . The percentage of complete data for each station is shown below the indicated year. The percent complete for each station for the 1990-2003 interval is shown in column 15, with each station's rank in column 16. Immingham had the highest level of % of complete data for the period 1990-2003 with 98.2% of data recorded. The percentage of complete data for all stations for each year is indicated on the bottom (indicated by "Year %"). The years are also ranked in the order of the most complete. The year with the most complete data was 1992 with 99.1% of data recorded.

Table 4-1 Completeness of data for the North Sea tide stations

Station	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	Total%	Rank
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Wick	99	100	98	100	99	77	100	99	97	98	98	96	99	100	97.1	4
Aberdeen	100	100	100	88	85	100	100	100	99	96	98	99	96	98	97.1	4
Leith	100	77	94	100	100	100	100	100	100	98	76	100	70	100	93.5	7
North Shields	95	92	100	82	93	99	100	100	100	100	100	100	92	99	96.6	5
Immingham	100	99	100	100	100	100	87	99	96	98	96	100	100	100	98.2	1
Commer	85	100	100	100	100	100	95	100	100	85	92	99	99	96	96.5	6
Lowestoft	100	99	100	100	100	98	80	94	90	98	94	100	99	99	96.5	6
Felixstowe	100	99	100	100	99	100	98	100	100	96	93	90	100	99	98.1	2
Sheerness	100	100	100	100	99	100	98	98	100	98	95	97	83	99	97.6	3
Year %	97.4	96.2	99.1	96.7	97.2	97.1	95.3	98.9	98	96.3	93.6	97.9	93.1	98.9		
Rank	6	10	1	5	7	8	11	2	3	9	12	4	13	2		

4.1.2 Thames Estuary Data – Port of London Authority

The Port of London Authority (PLA) is a public trust established in 1908 to 'Administer, preserve and improve the Port of London'. PLA archives tide gauge information at their office at the London River House in Gravesend. Station name, date, time, water-elevation, and surge residual are logged in 10 minute intervals. The format is shown in Figure 4-4.

```
SILVERTOWN,01/01/2004 00:10:00,1.733602,-0.3848932
SILVERTOWN,01/01/2004 00:20:00,1.609902,-0.4100389
SILVERTOWN,01/01/2004 00:30:00,1.496614,-0.4379637
SILVERTOWN,01/01/2004 00:40:00,1.395898,-0.4667445
SILVERTOWN,01/01/2004 00:50:00,1.31036,-0.4940104
SILVERTOWN,01/01/2004 01:00:00,1.242621,-0.517379
SILVERTOWN,01/01/2004 01:10:00,1.193689,-0.5360842
```

Figure 4-4 Example of Port of London Authority data set for Silvertown.

Port of London Authority data sets were obtained for the following stations: Herne Bay, Silvertown, Southend, and Walton. For locations of these stations see Figure 4-5. The completeness of the tide data sets is shown in Table 4.2. The percentage of complete data for each station is shown below the indicated year. The percent complete for each station for the 2000-2004 interval is shown in column 7, with each station's rank in column 8. Southend had highest level of % of complete data for the period 2000 - 2004 with 99.8% of data recorded.

Figure 4-5 Port of London Authority tide gauge locations.

Table 4-2 Completeness of data for Thames Estuary tide stations

Station (1)	2000 (2)	2001 (3)	2002 (4)	2003 (5)	2004 (6)	Total% (7)	Rank (8)
Herne Bay	84.0	99.3	100	99.4	100	96.5	4
Silvertown	100	99.8	99.7	99.4	96.4	99.1	3
Southend	100	99.4	100	99.4	100	99.8	1
Walton	100	99.2	100	99.4	100	99.7	2

4.1.3 Continental Shelf Model (CS3) Data – Proudman Oceanographic Laboratories

The Continental Shelf Model (CS3) was developed at Proudman Oceanographic laboratories and is used to model the storm surge and tidal currents (Flather and Proctor 1983). Proudman Oceanographic Laboratories also archive the CS3 model data. This CS3 model is used by the Storm Tide Forecasting Service (STFS) to predict storm tides and currents around the British Isles. The STFS is operated by the United Kingdom's Meteorological Office (Met Office) on behalf of

the Department for Environment, Food and Rural Affairs (DEFRA), and supplies storm surge forecasts to the Environment Agency (EA) four times per day.

The CS3 model's forcing data is derived from a suite of atmospheric and ocean models that are run as part of an automated schedule. The primary atmospheric model used by the Met office is the Unified Model which has several components; the first is the Global Model with a resolution of about 60 km; the second is the Mesoscale model which has a resolution of about 12 km (Meteorological Office, 2006). The Domain of Mesoscale model covers all of the British Isles, North Sea and English Channel.

The CS3 model has approximately the same domain as the Mesoscale model and is driven by the wind and atmospheric pressure data from the Mesoscale model. The model is run 4 times a day and is comprised of hourly hindcasts from T-06 to T+00, and hourly forecasts from T+00 to T+36 hours. Post processing extracts data from the models at specific grid points to make up standard surge residual port tables. Data from the CS3 model are archived for all of the "A class" tidal stations in the United Kingdom at Proudman Oceanographic Laboratories. Data for the T+00 thru T+05 hour forecasts are archived at 0:00, 06:00, 12:00, and 18:00 hours. This format archives 6 different forecast intervals from T+00 to T+05 hours, 4 times a day. An example of the archived surge forecast data is shown in Figure 4-5. Each line of data consists of 12 hourly measurements, requiring 2 lines for each day. Forecast intervals (T+00, T+01 ...) and times are highlighted in grey and are shown for the first 24-hours for purposes of clarity only. Unfortunately, the CS3 model's archived format limits comparison of the model to a maximum of T+05 hours.

```
T+00   T+01  T+02  T+03  T+04  T+05  T+00  T+01  T+02  T+03  T+04  T+05
 00:00 01:00 02:00 03:00 04:00 05:00 06:00 07:00 08:00 09:00 10:00 11:00
 0.05  0.04  0.03  0.03  0.02  0.03L 0.04  0.04  0.06  0.07  0.08  0.06H
 T+00  T+01  T+02  T+03  T+04  T+05  T+00  T+01  T+02  T+03  T+04  T+05
 12:00 13:00 14:00 15:00 16:00 17:00 18:00 19:00 20:00 21:00 22:00 23:00
 0.03  0.02  0.03  0.05  0.06  0.05L 0.02 -0.01  0.00  0.04  0.06  0.01H
-0.06 -0.07 -0.05 -0.01  0.02  0.01L-0.05 -0.08 -0.09 -0.05  0.01  0.02H
-0.03 -0.06 -0.05 -0.01  0.02  0.06  0.07L 0.09  0.10  0.08  0.07  0.07H
 0.08  0.10  0.11  0.10  0.09  0.08  0.08L 0.16  0.23  0.22  0.17  0.13
 0.12H 0.11  0.14  0.17  0.17  0.16  0.19L 0.26  0.37  0.39  0.34  0.23
```

Figure 4-6 Example of archived CS3 model data for Sheerness.

Archived CS3 Model data for the Sheerness Class A tide station was obtained for the years 1993 – 2003. Table 4-3 shows the months the data was archived marked by an 'X'. 1999 and 2002 were the only complete years where all the months were archived.

Table 4-3 Data Completeness - CS3 Model

	Month											
	J	F	M	A	M	J	J	A	S	O	N	D
1993									X	X	X	X
1994	X	X	X	X					X	X	X	X
1995	X	X	X	X					X	X	X	X
1996	X	X	X	X					X	X	X	X
1997	X	X	X	X					X	X	X	X
1998	X	X	X	X					X	X	X	X
1999	X	X	X	X	X	X	X	X	X	X	X	X
2000	X	X	X	X					X	X	X	X
2001	X	X	X						X	X	X	X
2002	X	X	X	X	X	X	X	X	X	X	X	X
2003	X											

4.2 Computers Used

The following computers were used for this project:

- 1 - Dell Dimension 5000 PC with a 3.0 GHz Pentium 4 processor, 1 GB RAM and Microsoft Windows XP Professional operating system. (Primary computer used during the project)
- 1 - Dell Precision 380 PC with a 3.2 GHz Pentium D processor with 2 GB RAM Microsoft Windows X64 operating system (Used during the ensemble forecasting portion of the project).
- 4 - Dell Dimension 2400 PCs with a 2.66 GHz Pentium 4 processor with 512mb of RAM and Microsoft XP Professional operating system (Used during the ensemble forecasting portion of the project).

4.3 Software Used

The Artificial Neural network used for this project was implemented using MATLAB Version 7.2.0.232, Release 2006a, utilizing the neural network toolbox.

5 Methods

The methodology for this project has been divided up into six sections:

1. **Data Preparation Methods** - This section describes the methodology used to process the raw data before its use for input to the artificial neural network models. The preprocessing includes separation of the water level time series into its tidal and surge components. See Section 5.1.
2. **Artificial Neural Network Methods** - This section describes the choice of structure for the artificial neural networks and the methodology used to train and test the models. See Section 5.2.
3. **Model Performance Assessment** - This section describes the methods used to assess model performance. See Section 5.3.
4. **ANN Optimization Methods** - This section describes the optimization procedures involving structure, design, and selection of inputs used to minimize forecasting errors produced by the artificial neural networks. See Section 5.4. (Experiments 1, 2, and 3)
5. **Model Comparison Methods** – This section compares the results of the optimized ANN models with the CS3 numerical model. See Section 5.5. (Experiment 4)
6. **Ensemble Forecasting Methods** – This section uses Ensemble models to reduce variance, or the instability of the neural network. See Section 5.6. (Experiment 5)
7. **Engineering Application Methods** – This section applies lessons learned from previous experiments to a new location, and develops an optimal ANN model. See Section 5.7. (Experiment 6)

5.1 Data Preparation Methods

Before raw water level data can be used in the artificial neural network, it must be processed. This processing is performed in several steps and has been divided up into two main sections. The first section uses tools developed as part of the Texas Coastal Ocean Observation Network (TCOON) Database (Tissot et al. 2005). These tools enable a harmonic analysis of the data to be performed and the data to be separated into its tidal and surge components (See Section 5.1.2). The second section uses MATLAB programs that are used to analyze, linearly interpolate missing data, and format the data for use by the ANN (See Section

5.1.2). For an overview of the data preparation module within the overall project organization, see Figure 4-1 (page 53).

5.1.1 Texas Coastal Ocean Observation Network Database

The Texas Coastal Ocean Observation Network (TCOON) is a network of 35 tide gauges located on the Texas gulf coast (Tissot et al. 2005). TCOON has several built-in features used to help maintain and process tidal records. Web based tools were developed as part of TCOON to help during the preliminary processing of tidal data. This section describes how data is entered into the TCOON database, how harmonic analysis is performed, and how a synthetic harmonic tide and measured tide file is generated. The primary feature used for this project is HarmAN, a harmonic analysis program and HarmPred, a program which generates a synthetic tidal prediction. For more general information about the TCOON network see Appendix 2.

Database import

To be able to use the TCOON tools the project data had to be first stored in the database. To do so, a location data key was constructed for each project water level data series. The data key for this project is shown in Figure 5.1. Each water level measuring station used in this project is assigned a unique TCOON station number, and the latitude and longitude was entered for later use during the harmonic analysis.

```
950 1990001+0000 2004168+0000 name Wick, Scotland ##( dprouty)
950 1990001+0000 2004168+0000 abbr WIC    ##( dprouty)
950 1990001+0000 2004168+0000 loc 58.4413,-3.0849 ##( dprouty)
951 1990001+0000 2004168+0000 name Aberdeen, Scotland ##( dprouty)
951 1990001+0000 2004168+0000 abbr ABE ##( dprouty)
951 1990001+0000 2004168+0000 loc 57.1441,-2.0787 ##( dprouty)
952 1990001+0000 2004168+0000 name Leith, Scotland ##( dprouty)
952 1990001+0000 2004168+0000 abbr LEI ##( dprouty)
952 1990001+0000 2004168+0000 loc 55.9898,-3.1803 ##( dprouty)
953 1990001+0000 2004168+0000 name North Shields, England ##( dprouty)
953 1990001+0000 2004168+0000 abbr NSH ##( dprouty)
953 1990001+0000 2004168+0000 loc 55.0073,-1.4383 ##( dprouty)
954 1990001+0000 2004168+0000 name Immingham, England ##( dprouty)
954 1990001+0000 2004168+0000 abbr IMM ##( dprouty)
954 1990001+0000 2004168+0000 loc 53.6330,-0.1869 ##( dprouty)
```

Figure 5-1 Data-key for entry into the TCOON database.

Harmonic analysis

Harmonic analysis is performed for two primary reasons. The first reason is to aid in the calculation of the storm residual. An accurate harmonic analysis is needed, because any harmonic constituent that is not included in the analysis will then be

considered part of the surge component, and will cause errors later for the ANN. The second reason is to fill gaps in the data record with approximate water levels for use during training of ANNs.

This project uses water level data from two different sources, the British Oceanographic Data Centre (BODC) and the Port of London Authority (PLA). Both data sets provide measured water levels and surge residuals and small gaps are flagged and filled. Unfortunately, large gaps can still be found in these records due to gauge maintenance or other reasons. The use of artificial neural networks for gap filling has been used for data recovery in other applications (Gorban and Rossiev 2002), but were not used here, because of the selection method. Testing years were selected because they contained only small gaps (less than 6 hours), and storms were selected for analysis only if they contained no gaps. The synthetic harmonic data derived at this stage are used to interpolate missing data which is performed later and is described in Section 5.1.2. To do this, a new harmonic analysis is performed for all data sets and a new set of surge residuals are calculated.

The Harman program is implemented as a web based program and was designed by Mostella et al. (2002). Harman is a Perl/PDL program developed to determine harmonic constants from previously collected data sets. Harman uses three files as input; a file of constituent names to be used in the analysis; a file of constituent names speeds, node factors and equilibrium arguments; and a file containing water level observations. To use HarmAn, the user logs into the web site:

http://lighthouse.tamucc.edu:4001/harman

where a form appears (See Figure 5.2). The user then enters the station number and the date range to be used in the analysis (in the "New Analysis" area at the bottom of form), then selects the "Analyze" button at the bottom of the form. It is important to note that at least one year of hourly data is required, preferably with less than 2% missing, however up to 10% missing may be adequate. After the analysis is performed, the user is presented with the same form but with the Amplitude and Phase components displayed for each of the constituents (See Figure 5.2). The user can then automatically save these constituents to the TCOON database by filling in the name field at the top of the form with "hcset" and selecting the "Post" button at the bottom of the form. If the text "hcset" is used as the name, then those constituents will be used in the future to generate harmonic

tidal predictions. As seen in Figure 5-2, various size data sets for each station were analyzed. Ideally, 17.6 years of data are needed to encompass most of the astronomical influences seen in the tidal record, but acceptable results can be obtained with as little as one year of measured water level records. Tests performed comparing 1, 5, and 10 year harmonic analysis show little differences between them. For this project, 10 years of data is used for all BODC data sets, and 5 years of data was used for all POL data sets.

Name	Sheerness		1year		5year		10-year		Hcset	
Description	2000		1990		1990-1994		1990-1999		1990-1999	
Time Offset	0		0		0		0		0	
H0	3065.5		3008.76		2976.15		2989.17		2989.17	
J1	4.23	-91.36	10.46	81.13	6.35	81.10	8.63	97.59	8.63	97.59
K1	105.66	11.21	116.57	18.52	114.33	11.14	114.26	10.96	114.26	10.96
K2	171.83	54.62	171.76	49.58	168.30	50.98	167.67	51.22	167.67	51.22
L2	124.89	7.39	117.36	-5.76	129.45	5.00	133.82	8.15	133.82	8.15
M1	10.09	125.87	4.73	107.99	5.06	-169.64	2.49	135.41	2.49	135.41
N2	336.03	-27.88	348.19	-27.96	339.57	-29.58	338.77	-30.38	338.77	-30.38
2N2	78.56	-61.92	80.20	-99.40	46.30	-63.43	42.85	-57.01	42.85	-57.01
	Post		Edit Graph		Edit Graph		Edit Graph		Edit Graph	

New Analysis:

Station: 958

Dates: 2000

Analyze

Figure 5-2 Example of input form for HarmAn, a web-based Harmonic analysis program (Note: only a first 7 constituents are displayed in this example)

When the user requests harmonic water levels for a specific station from the TCOON database, a harmonic tidal prediction program (HarmPred) is run automatically. HarmPred is a Perl/PDL program used to build a synthetic tide from

the tidal constituents generated when using the HarmAn program. HarmPred uses three inputs: first a file of constituent names, amplitudes, and phases for the target station, second, a file of constituent speeds, equilibrium arguments, and node factors, and finally, the desired time of prediction. A water level prediction can be made by using the following formula:

$$h_t = \sum H_i f_{i,t} \cos(a_i t + e_{i,t} - k_i) \quad (5.1)$$

where h_t = water level at time t; a_i = speed of constituent *I;* H_i = amplitude; and k_i = phase for constituent *i* from the harmonic constants, $f_{i,t}$ = node factor for constituent *i* at time *t*; and $e_{i,t}$ = equilibrium arguments for the constituent *i* at time *t*.

Data extraction

After water level data has been entered into the TCOON data base, and a harmonic analysis is performed, the data is extracted in preparation for the statistical analysis and gap filling programs that follow. The data is extracted via the web using the following http command:

http://lighthouse.tamucc.edu/pd/stnlist=950&serlist=pwl,res,offswl,harmwl&when=1990?-action=c&na=9999&interval=3600

Where stnlist = station id number; serlist = list of output parameters requested; pwl = primary water level; res = residual water level; offswl = offset water level; harmwl = harmonic water level; when = time range for output (year of 1990); action = type of output format; c = column output format; na = no data available; flag designator for output; interval = interval in seconds between output values. An example of the output is shown in Figure 5-3.

```
# DISCLAIMER: The data described below have been collected by automated
# equipment and are furnished "as is". DNR makes no warranties (including
# no warranties as to merchantability or fitness) either expressed or implied
# with respect to the data or their fitness for any specific application.
#-------------------------------------------------------------------------------
#printcols: start
# 950-pwl: generated Tue Jul 25 15:17:37 2006 UTC
# 950-pwl: Wick, Scotland (58.4413,-3.0849)
# 950-pwl: Primary Water Level (mm)
# 950-pwl: Elevations above Station Datum (stnd)
# -------------------------------------------------------------------------------
# 950-res: generated Tue Jul 25 15:17:37 2006 UTC
# 950-res: Wick, Scotland (58.4413,-3.0849)
# 950-res: Residual Water Level (mm)
# -------------------------------------------------------------------------------
# 950-offswl: generated Tue Jul 25 15:17:37 2006 UTC
# 950-offswl: Wick, Scotland (58.4413,-3.0849)
# 950-offswl: Offset from Predicted Water Level (mm)
# 950-offswl: hcset=hcset
# 950-offswl: Elevations above Station Datum (stnd)
# -------------------------------------------------------------------------------
# 950-harmwl: generated Tue Jul 25 15:17:37 2006 UTC
# 950-harmwl: Wick, Scotland (58.4413,-3.0849)
# 950-harmwl: Harmonic Predicted Water Level (mm)
# 950-harmwl: hcset=hcset
# 950-harmwl: Elevations above Station Datum (stnd)
# -------------------------------------------------------------------------------
#  date+time 950-pwl 950-res 950-offswl 950-harmwl
1990274+0000    1781      12         54       1727
1990274+0100    1411      23         58       1353
1990274+0200    1162      26         30       1132
1990274+0300    1084      10        -13       1097
1990274+0400    1225     -22        -23       1248
1990274+0500    1549     -45        -18       1567
1990274+0600    1992     -40        -12       2004
1990274+0700    2455     -20         -8       2463
1990274+0800    2821      14          7       2814
```

Figure 5-3 Example TCOON database output file

Output from the screen is then saved to a text file and named in the following format:

1990_A_WIC.txt

Where 1990 = year of data set, _A_= text to indicate data was processed, WIC = Wick or 3-letter station identifier, .txt = file extension.

5.1.2 Data Interpolation, Formatting and Preliminary Analysis

As in most environmental data sets, gaps are found in these time series. As continuous time series are desirable for input into the neural networks the gaps are filled by linearly interpolating the surge component of the water level time series harmonic data before it is used by the artificial neural network. Gaps of one month or smaller were found to have relatively small effects on harmonic predictions (Mostella et. al 2002). A MATLAB program was written to process the data exported from the Texas Coastal Ocean Observatory (TCOON) database and format, analyze and condition the data before its input into the Neural Network. The actual MATLAB code used to do the preliminary processing and gap filling is

called Convert_DNR_NN and is shown in Appendix 3. This program formats and calculates the following:

- Harmonic water level (calculated using TCOON HCSET values)
- Observed water level (from original BODC or PLA values + Gaps filled)
- Surge water level (Calculated from Observed water level - Texas Coastal Ocean Observatory (TCOON) harmonic) (Contains no gaps)
- Residual water level (Values from input file (calculated by the BODC (With gaps)) this value is not used.

The program identifies: the number of records in each data series; their starting and ending times; and the number, size, and location of gaps. The program fills in the gaps found in the observed water levels by linearly fitting the missing data from the surge component and adding this to the harmonic component (since Observed water levels = Harmonic water levels + Surge water levels). The preliminary analysis and gap filling program produces two output files, an Analysis file and a NN file. The Analysis file reports the following statistics:

- Data input file name
- Number of records processed
- Title of data series read
- Number of points per series
- Total days of data in series
- Starting day and time of series
- Ending day and time of series
- The number of missing data for each series and percent of total
- Longest gap for each time series (before being filled)
- All Skill assessment statistics for the harmonic model

The following Skill assessment statistics (with the error being defined as the difference between model results, harmonic predictions and the measured water levels):

- Average Error: $E_{avg} = (1/N) \Sigma e_i$
- Average Absolute Error:
- Standard Deviation of the Error:
- Root Mean Square of the signal:
- Root Mean Square Error: $E_{rms} = ((1/N) \Sigma e_i^2)^{1/2}$
- Normalized RMS Error:

- Central Frequency (X=15 cm): Central Frequency or percentage of forecasts within X cm of the actual measurement
- Positive Outlier Frequency (X=30 cm): Positive Outlier Frequency or percentage of the forecasts X cm or more above the actual measurement
- Negative Outlier Frequency (X=30 cm): Negative Outlier Frequency or percentage of the forecasts X cm or more below the actual measurement
- Maximum Duration of Positive Outlier (X=30 cm):
- Maximum Duration of Negative Outlier (X=30 cm):
- Worst Case Outlier Frequency (X=15 cm):
- Worst Case Outlier Frequency (X=30 cm):

An example of an Analysis file is shown in Figure 5-4.

```
Data originating from BODC raw data file: 1990_A_wic
Number of data series read from the raw data file including date and hour time
series: 8760
Name of data series read from the raw data file:
  time
  pwl (primary water level)
  res (residual water level)
  offswl (offset water level)
  harmwl (harmonic water level)
Number of data points per series: 8760
Representing 365 days of data
Starting time of the data series: Day 1990001 @ 0
Ending time of the data series: Day 1990365 @ 1990001
Ending time of the data series: Day 2300 @ 0
Missing data for time series "pwl         ": 63 data points or 0.72 percent
Missing data for time series "res         ": 63 data points or 0.72 percent
Missing data for time series "offswl    ": 0 data points or    0 percent
Missing data for time series "harmwl ": 0 data points or    0 percent
Longest gap for time series "pwl         ": 44 data points
          Starting at data point 8089 and ending at data point 8132
Longest gap for time series "res          ": 44 data points
          Starting at data point 8089 and ending at data point 8132
Longest gap for time series "offswl     ": 0 data points
          Starting at data point 0 and ending at data point 0
Longest gap for time series "harmwl   ": 0 data points
          Starting at data point 0 and ending at data point 0
Skills for Harmonic model for year ****************************************
Skill                                          Tide Tables
Average Error:                                -64.6144
Average Absolute Error:                       153.9295
Standard Deviation of the Error:                 190.8561
Root Mean Square of the signal:                 2242.8246
Root Mean Square Error:                          201.4867
Normalized RMS Error:                              0.0898
Central Frequency (X=15 cm):                      59.8858
Positive Outlier Frequency (X=30 cm):           1.7466
Negative Outlier Frequency (X=30 cm):          10.2055
Maximum Duration of Positive Outlier (X=30 cm):    63.0000
Maximum Duration of Negative Outlier (X=30 cm):    64.0000
Worst Case Outlier Frequency (X=15 cm):          0.0000
```

Figure 5-4 Example of Preliminary Data Analysis File.

The second output file created by the preliminary analysis and gap filling program is the NN file. The NN file is automatically named by the program and in the following format:

1990_A_WICNN.txt

where 1990 = year of data set; _A_= text to indicate data was pre-processed; WIC = Wick or any 3-letter station identifier; NN = processed and ready for entry into Neural Network; .txt = file extension. An example of the preliminary processed data output file is shown in Figure 5-5. This file is in the proper format for use as an input file for the artificial neural network.

```
#Date      Time   Pwl    Hwl    Res
1990001    0      2683   2793   -49
1990001    100    3071   3145   -42
1990001    200    3174   3222   -36
1990001    300    2928   2979   -29
1990001    400    2459   2488   -15
1990001    500    1917   1917   11
1990001    600    1448   1433   36
1990001    700    1195   1163   64
```

Figure 5-5 Example of a NN data file

5.2 Artificial Neural Network Methods

For all experiments, the ANN models were developed, trained, and tested within the MATLAB R13 computational environment and the related Neural Network Toolbox (The MathWorks, Inc., 2002). For an overview of the ANN module within the overall project organization, see Figure 4-1 (page 53). The primary computer used was a 3.0 GHz Pentium 4 PC running Windows XP. The models were trained using the Levenberg-Marquardt back-propagation algorithm as implemented within the MATLAB Neural Network Toolbox.

5.2.1 Application of the Artificial Neural Network Model

Several different training algorithms were tested, and the results are shown in Table 5-1. The Levenberg-Marquardt had the lowest error level of all the algorithms tested with an average absolute error of 0.125 m for the one-year test. The next best performing algorithm tested was the gradient decent method with adaptive learning rate with an average absolute error of 0.146 m. The Levenberg-Marquardt algorithm was selected as the training algorithm to be used for this project because it had the lowest average absolute error.

Table 5-1 Training methods tests

Training Algorithm	Description	Training Times (Sec)	Average Absolute Error (m)
Trainlm	Levenberg-Marquardt back-propagation	197	0.125
traingda	Gradient descent with adaptive learning rate back-propagation	185	0.146
traingdx	Gradient descent with momentum & adaptive learning rate back-propagation	183	0.157
traingdm	Gradient descent with momentum back-propagation	356	0.148

All ANN models will initially be trained using the Levenberg-Marquardt algorithm as implemented within MATLAB. Training times vary between a few minutes and several hours depending on the size of the ANN. It is important to note that although training times can be lengthy, generating water level forecasts is a sub-second process once the models are trained. Once the ANN models are trained, they are ideally suited for streamed forecasting (an automatically generated, real-time forecast based on streaming data).

Training algorithm used

Levenberg-Marquardt was selected for the training algorithm in this project. Tissot et al. (2001) and Cox et al. (2002) both selected the Levenberg-Marquardt algorithm because it was found to be the fastest when processing large data sets, and gave the best results. The Levenberg-Marquardt algorithm processes the whole matrix (96x8760 (48 hours for each station x 8760 hours per year)) at a time (one epoch or iteration), during which it calculates new weights and biases for each column of data inputs. The Levenberg-Marquardt algorithm improves the convergence speed of the standard back-propagation algorithm by modifying the rate changes are made to the search direction and size of the step, with initial step sizes being large and using smaller size steps when approaching a minimum. It is important to note that the Levenberg-Marquardt algorithm is implemented in batch mode, which processes all weights and biases for all cases at one time. A simplified example for training a single station ANN using 48 hours of previous water level data with a 6-hour forecast is shown next.

Training Example

The entire training year surge residuals are loaded into a single vector.

A second matrix for use by the Levenberg-Marquardt algorithm is formed; each row is incremented by one hour and contains the following data:

- 48 water level residuals or surge heights (from T-48 hours to T-00 hours)
- 48 weights, and 1 bias for the hidden layer (initially randomly chosen)
- 1 weight and 1 bias for the output layer (initially randomly chosen)
- Predicted water level (at T+6 hours given by ANN algorithm)
- Measured water level (at T+6 hours, for a 6-hour forecast.)
- Error (Predicted - Measured)

This step is repeated until all 8760 data points (one year) are loaded (requiring 8760 Rows). The Levenberg-Marquardt algorithm as implemented in MATLAB is done in batch mode i.e. for all weights and biases and for an entire year at the one time. This algorithm utilizes a combination of methods for minimizing RMSE.

During the first part of the process the steepest descent decent method was used to determine the direction and size of the next step to look for the minimum RMSE. The values of the derivatives of the error are computed individually for each weight then the error is back propagated through the network until it reaches each weight or bias. Note: This step is modified as the training progresses; large steps are taken at the beginning and getting smaller at the end.

During the second part of the process the step values are computed through a Newton chord technique and become smaller as the ANN converges towards a minimum error level. At the end of each iteration step, a value for root mean square error is calculated. The root mean square error (defined as the square root of the sum of the errors squared) is calculated by comparison of the predicted water levels (ANN output) and the measured water levels (ANN target) for the entire year. This completes the first iteration.

The Levenberg-Marquardt algorithm calculates a direction and step size and back propagates the error using the inverse of each transfer function to calculate new weights and biases. This process is repeated until the number of iterations reaches a preset limit or a pre-defined RMSE is reached. At this point the training is over and the current weights and biases are saved, and can be used to process data.

5.3 Model Performance Assessment

The performances of the models in this work are assessed based on criteria used by NOAA for the development and implementation of operational nowcast and forecast systems (NOAA, 1999). A single forecasting error is defined as the difference between the predicted value and the observed value. The models are assessed by averaging the individual errors over the full data sets, often one year of water level residuals or surge heights and forecasts. The statistical parameters used to evaluate the models performance for this paper are the Average Absolute Error (AAE) between predictions and measurements and the Central Frequency (CF) of 150 mm. The Central Frequency (CF) is the percentage of predictions that are within 150 mm of the measured water levels. The 150 mm selected for the Central Frequency (CF) measure is the requirement typically used by NOAA and is based on NOAA's estimates of pilots' needs for under keel clearance value (NOAA 1999). In the UK the Storm Tide Forecasting Service (STFS) model performance is measured by a similar method, but is called a "skill measure". This is similar to the CF described by NOAA, but uses a less stringent value of 200 mm. This paper will use the AAE and CF (150 mm) for performance analysis. Several other skill assessment statistics are tracked during the project and are described below:

- Average Error: $E_{avg} = (1/N) \Sigma e_i$
- Average Absolute Error:
- Standard Deviation of the Error:
- Root Mean Square of the signal:
- Root Mean Square Error: $E_{rms} = ((1/N) \Sigma e_i^2)^{1/2}$
- Normalized RMS Error:
- Central Frequency (X=15 cm): Central Frequency or percentage of forecasts within X cm of the actual measurement
- Positive Outlier Frequency (X=30 cm): Positive Outlier Frequency or percentage of the forecasts X cm or more above the actual measurement
- Negative Outlier Frequency (X=30 cm): Negative Outlier Frequency or percentage of the forecasts X cm or more below the actual measurement
- Maximum Duration of Positive Outlier (X=30 cm):
- Maximum Duration of Negative Outlier (X=30 cm):
- Worst Case Outlier Frequency (X=15 cm):

- Worst Case Outlier Frequency (X=30 cm):

These skill assessment statistics are produced for every artificial neural network model that is produced, and enables an accurate assessment of the models performance.

5.4 Artificial Neural Network Optimization Methods

The general optimization scheme used for this project is to start with the simplest configuration, then optimized after each new parameter is added. The initial artificial neural network model for this project used only one station. The final configuration uses a primary station and a secondary station, varies the amount of data used for each, depending on the forecast interval, varies the size of the ANN structure, and varies the number of individual forecasts used in an ensemble forecast. Each of the optimization steps is run as a separate experiment. Each experiment uses the optimal number of parameters or structure configuration set-up found during the previous experiment. To better understand the behavior and potential of ANN models for storm surge predictions, six suites or groups of related experiments were conducted. The following is a short description of each suite.

Suite #1 - Training data selection

1. Determination of optimum size of training data set. A single station artificial neural network model is run varying the size of the training data set from 1 to 10 years. This test determines how performance varied using larger training data sets and determined if performance gains were worth the longer training times. This test is called Experiment 1.1
2. Selection of training year. A single station artificial neural network model is trained on each year individually, and then tested on every other year. This experiment is conducted to see if any particular year is better than others for use as a training year. Individual results are automatically saved in output files then tabulated and compared. This test was called Experiment 1.2.

Suite #2 – Optimize number of inputs and secondary station location

1. Determining the optimum number of inputs for a single station ANN. The model was trained varying the number of previous water level residuals or surge heights used for input for a single station ANN. Individual results were automatically saved in output files then tabulated and compared. This test is called Experiment 2.1.

2. Determining the optimum number of inputs for a dual station ANN. The model was trained varying the number of previous water levels used for the primary and secondary stations. Individual results were automatically saved in output files then tabulated and compared. This test is called Experiment 2.2.
3. Determining the optimum location for a secondary station was performed by varying the location of the secondary station for each forecast interval. This test was done to determine the optimum secondary station location for each forecast interval. Individual results were automatically saved in output files then tabulated and compared. This test is called Experiment 2.3.

Suite #3 – ANN topology and performance

1. Varying structure size. This test is called Experiment 3.1.
2. ANN performance variability. This test is called Experiment 3.2.

The training is conducted by selecting at least one full year of water levels and by assembling input vectors, each consisting of a time series of previous storm surge levels from one or more tide stations. Weights and biases are randomly assigned at the beginning of each training session and their values were updated during each iteration such that the error between ANN output and target (or observed) is progressively minimized. Training times vary between a few minutes and several hours depending on the size of the ANN. Although training times can be lengthy, it should be emphasized that for real-time applications, generating water level forecasts is a sub-second process. Once the ANN models are trained, they are ideally suited for streamed forecasting (an automatically generated, real-time forecast based on streaming data).

5.4.1 Experiment Suite #1: Training data selection

The first part of Experiment Suite 1, Experiment 1.1, concentrates on how the training year is selected and on the impact of the training set length on ANN performance. In Experiment 1.2, input to the ANN models was constructed by using data from a single primary station, while varying the number of previous water level residuals or surge heights used for each forecast from 1 to 48 hours. For Experiment 1.3 additional previous water level residuals or surge heights from a secondary station were added. The numbers of previous water level residuals or surge heights used at both the primary and secondary station were varied from 1

to 48 hours. The focus of Experiment 1.4 was to explore the changes in model performance while selecting secondary stations at different locations along the UK North Sea coastline. A full year was selected as the minimum data set length for the training of the ANNs to include seasonal variations of water levels. The early stopping method used to prevent over-training in Experiment Suite 1 is to stop training when a preset error level is reached (1 mm).

Experiment 1.1 Varying training data set lengths
A schematic of a typical ANN model used for this study is presented in Figure 2-7 (page 14). The structure used for the first suite of experiments is a two layer ANN, using one output neuron, one hidden neuron. The optimum ANN topology including number of hidden neurons will be discussed as part of the second suite of experiments. The lengths of the training sets were varied from 1 year (1990), 3 years (1990-1992), 5 years (1990-1994), and 10 years (1990-1999). All ANNs were trained using 48 hours of previous water level residuals or surge heights for each forecast. Models were then tested on the 2001 data set. A substantial disadvantage of using a large training set is that computational time increases from 20 minutes (using a one-year data set) to 20 hours (using a 10-year data set) for the computer used for this study.

Experiment 1.2 Selection of training year
Testing was performed to evaluate the importance of the selection of a particular training year including the possible impact of year to year variability in the frequency and magnitude of storms. Testing was also performed to assess potential improvements when including longer training periods. A series of basic models making 3-hour predictions were successively trained, using each year of water levels at the Sheerness station from 1990 to 1999 with an input consisting of 48 hours of previous water levels. For each of the training years the model was then tested on all other years.

5.4.2 Experiment Suite #2: Optimize number of inputs and secondary location

Experiment 2.1 Single-Station ANN
The ANN was first trained using Sheerness as the only (or primary) station. The forecast time periods used were: 3, 6, 12 and 24 hours. The number of previous water levels included during training varied, using 1, 3, 6, 12, 24, 36, and 48 hours

for each forecast period. No secondary station was used in this experiment. The models were trained on the 1997 data set, and tested on the 1993 data set.

Experiment 2.2 Two-Station ANN
For this experiment, additional data from a second water level measuring station is included in the ANN training set. This secondary station, is located north of the primary station (a direction towards the approaching coastally trapped wave/surge), and provides the ANN with additional information of a surge's presence before it can be measured at the primary station. The ANNs were trained on the 1997 data set, tested using the 1993 data set. In this experiment the number of previous residuals or surge heights used for the primary station (Sheerness) is maintained constant at 24-hours, while the number of previous water level residuals or surge heights from the secondary station, Immingham is incremented from 1 to 24 hours.

Experiment 2.3 Two-station ANN, varying secondary location
Before the selections of a secondary station are made, a cross-correlation analysis of the surge data from each of the potential secondary station is performed. This analysis will determine the approximate time lags needed for the surge component to travel to Sheerness from each secondary station location. The cross-correlation analysis program was written in MATLAB and the source code is included in Appendix 6.3

The performance of the ANN is analyzed after varying the location of the secondary station. Three secondary station locations Immingham, North Shields and Wick are selected for each test. The stations are respectively 337 km, 510 km and 945 km north of Sheerness. For each test, 24 hours of secondary station data was used, and the number of previous water level residuals or surge heights used from the primary station varied from 1 to 48 hours.

5.4.3 Experiment Suite #3: ANN topology and performance

This suite of experiments we investigates how model performance is affected by changes in ANN topology. While the performance is tracked for both yearly averages and during storms, the discussions will focus primarily on the storm performance because of its importance for safety and commerce. Experiment 3.1 varies the ANN structure size, by changing the number of hidden neurons used. Experiment 3.2 tests the robustness (or repeatability) of the ANN model. For all experiments in Suite 3, Sheerness is used as the primary station, and Immingham

as the secondary station. The training year used is 1997, testing year 1993, and validation year 1999. 24-hours of previous water level residuals or surge heights are used for training for each station. During initial runs of Experiment suite #3, models that used more than 5 hidden neurons experienced overtraining problems, causing large variations in predictions. To avoid over-training, a verification data set is used for all Suite 3 experiments as a early stopping method.

Experiment 3.1: ANN performance varying structure size
In this experiment, the variability of ANN models with different number of hidden neurons was determined. 3-hour ANN forecasts were calculated for each of the following size models: 1,2,5,10,20, and 50 hidden neurons.

Experiment 3.2: ANN performance variability.
The tests performed for experiment 3.1 were repeated 20 times with random starting values for the weights and biases. The variability arises from MATLAB's Levenberg-Marquardt algorithm, which assigns random values when initializing weights and biases during training. This variability was found to only be significant when using larger ANNs (greater than 5 hidden neurons).

5.5 Model comparison methods

This test compares yearly performances of the artificial neural network with the Continental Shelf Model (CS3) designed by the Proudman Oceanographic Laboratories for use by the Storm Tide Forecasting Service for the Met Office. The numerical model used by the STFS is called the continental shelf model or CS3. The CS3 model was designed by Proudman Oceanographic Laboratories (POL) to forecast tidal elevations and currents around the British Isles. The model is run four times daily, producing hourly forecasts from T-12 to T+36. The short-term forecasts for T+00 to T+05 are archived for each run by appending these values to a monthly file that is saved at POL. Data was archived for the storm season only (September – March) until 1999, after which time it was archived for the entire year. For this project only 2 years of complete data were received (years 1999 and 2002) for the CS3 model. For an overview of the ANN / CS3 Comparison module within the overall project organization, see Figure 4-1 (page 53).

The artificial neural network model used for this comparison uses Sheerness as the primary station and Immingham as the secondary station. 24-hours of previous water level residuals or surge heights were used for each

station. The network was trained using year 1997 data, validated using year 2001 data and tested on years 1999 and 2002 (Testing and validation years had to be changed for this experiment because of the limited archived data was available from Proudman oceanographic laboratories). The ANN model was run using 3 different forecasting intervals, a 1-hour, 3-hour and 5-hour. The ANN hourly forecast data was saved to a file for each forecast interval for both test years (1999 and 2002).

5.5.1 Experiment Suite #4: ANN CS3 Comparisons

Experiment 4.1: Model comparison using yearly performance
A MATLAB program was written to parse the archived CS3 data sets into 6 different forecast files from T+00 to T+05. Each file contained forecasts at six-hour intervals, and is offset one hour from the previous file. The program then compares the selected forecast interval values from the CS3 model and ANN model to the actual measured data from the tide station. The root mean square error (RMSE), central frequency (CF(150 mm)%), and Average Absolute Error (AAE) are calculated for the CS3 and ANN models for each test year. The source code for the yearly forecast comparison program can be seen in Appendix 5.

5.5.2 Experiment 4.2: Model comparison using storm performance

The program also selected from each 6-hour data-sets (T+00 to T+05), all data points with surge elevations greater than 1.0 m above the predicted harmonic water elevation. 36 hours of data or 6, 6-hour data points, before and after the peak values are used to define a 72-hour "storm window". The program then calculates a RMSE, CF (150 mm)%, and a AAE for the ANN and CS3 models, restricting the data used to a 72-hour window for each storm found in the yearly data set. The source code for the 1-hour forecast comparison program can be seen in Appendix 6.

5.6 Ensemble Forecasting Methods

Previous experiments hinted that larger structures gave better peak water level predictions than small structures. One of the problems found is that the variance or instability of the neural network increases substantially as the complexity of the ANN structure increases. The error surface of neural network training is full of local minima. These error surfaces have two primary causes:

- Trainings with different initial weights are usually trapped in different local minima. This is a result of the random initialization of the weights and biases when the ANN is started.
- Noise in the actual data itself. The data is sampled once an hour, and is only a single point, but represents an entire hour of changing water levels. During this hour the tide may be falling, rising, or reaching a maximum, or minimum. To reduce this noise the data sampling interval would have to be decreased.

A solution for this problem is to use ensemble forecasting. Ensemble averaging lets the noise portion of the solution cancel itself out, thus allowing a more accurate fitting of the data. Ensemble forecasting involves running the ANN model several times and averaging the results. Before each iteration, of the ANN, the weights and biases are initialized to random numbers. This causes the model to compute a slightly different solution each time. The idea behind ensemble forecasting is that forecast errors will average out and approach the true value after several iterations.

5.6.1 Experiment Suite #5: Ensemble Forecasting

This experiment is divided into 3 different parts:

- Experiment 5.1 Tests the model performance when varying the number of repetitions used for each ensemble forecast.
- Experiment 5.2 Tests the model performance when holding the number of runs used and varying the number of hidden neurons used in the hidden layer.
- Experiment 5.3 Tests to see if better peak water and surge levels can be forecasted using ensemble models with more complex structure.
- Experiment 5.4 Tests the amount of variance reduction obtained when allowing/not allowing the network to over-train.

For all the models tested, Sheerness was used as the primary station, and Immingham was used as the secondary station. All models are trained on year 1997 data, verified on year 2001 data, and tested using year 1999 data.

Experiment 5.1: Ensemble forecasts, varying the number of repetitions

During this test the number of repetitions an individual ensemble used was varied from 1 to 50. For each size of ensemble tested, 20 individual ensembles were run. The work load for this test was distributed to 6 individual PC's.

- Main program – ANN code that trains and tests a single ANN model.
- Ensemble Program – A looping program that runs the Main program several times and generates a "Averaged" forecast value for every data point. The program sums up all individual forecast values for every hour of the year and divides by the number of forecasts. This is then saved as the ensemble forecast.
- Multi-Ensemble Program – A looping program that runs multiple copies of the Ensemble program.

Each computer was assigned to test a 1, 5, 10, 20, or 50 repetition ensemble model. The models with a larger number of repetitions (20 and 50 repetition ensembles) were assigned to the 3.0 GHz PC's and models with few repetitions (1, 5, and 10 repetition ensembles) were assigned to the 1.0 GHz PC's. Each PC ran its own copy of the Multi-Ensemble program to generate 20 individual ensemble models. The duration of each test varied from approximately 1 to 10 days depending on the test size and speed of the processor. PCs that finished early were re-assigned to help finish tests still in progress. During this test, the statistical results are restricted to a pre-defined 72-hour storm window.

Experiment 5.2: Ensemble forecasts, varying the structure size

During this test structure size of individual ensembles used was varied to use from 1 to 30 neurons. Each ensemble forecast used 20 repetitions for each run. The work load for this test was also distributed to 6 individual PC's.

- Main program – ANN code that trains and tests a single ANN model.
- Ensemble Program – A looping program that runs the Main program several times and generates a "Averaged" forecast value for every data point. The program sums up all individual forecast values for every hour of the year and divides by the number of forecasts. This is then saved as the ensemble forecast.
- Multi-Ensemble Program – A looping program that runs multiple copies of the Ensemble program.

Each computer was assigned to test a [1,1], [5,1], [10,1], [15,1], or [30,1] ensemble model. The larger structure ANN models ([15,1] and [30,1] ensembles) were assigned to the 3.0 GHz PC's and the smaller structure models (1, 5, and 10 run ensembles) were assigned to the 1.0 GHz PC's. Each PC ran its own copy of the Multi-Ensemble program to generate 20 individual ensemble models. The

duration of each test varied from approximately 1 to 10 days depending on the test size and speed of the processor. PCs that finished early were re-assigned to help finish tests still in progress. During this test, the statistical results are restricted to a pre-defined 72-hour storm window.

Experiment 5.3: Ensemble forecasts, effect of structure size on model accuracy during periods of maximum water and surge elevation

This test is used to determine the effect of ANN structure size on the accuracy during maximum water level and surge events. This test focuses on the two most critical periods during a storm; the maximum water level (which usually occurs near the time of high tide) and maximum surge-level (Figure 5-6). Several different sized ANN structures were tested to see how the size of the ANN structure affects the accuracy of peak water and surge prediction. All ensemble forecasts used during this test used 20 repetitions per ensemble. Individual or single ANN forecasts were included as reference.

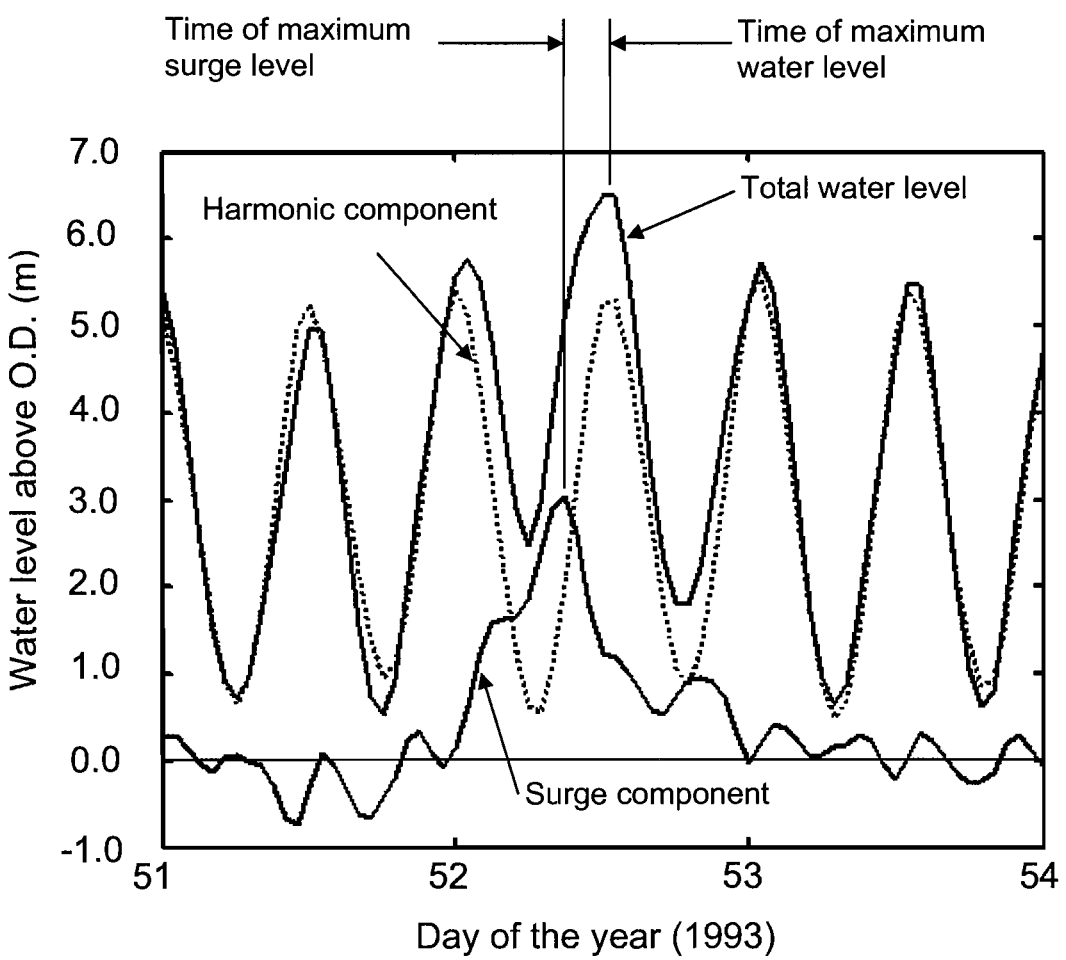


Figure 5-6 Maximum surge level, water level and their components.

Experiment 5.4: Ensemble forecasts, effect of overtraining on forecast variance.

Naftaly et al. (1997) demonstrated the effect of variance reduction for ensembles of networks. They found that ensemble averaging can be more effective when "not" combined with training constraints such as early stopping. During this experiment, two ensemble models were set up identical to those used in Experiment 5.3, with one exception. The first model used the validation method developed for experiment 3 for early stopping to prevent over-training. The second model was set-up to stop training only after 3000 epochs were reached. The first model was run using a 20 repetition ensemble model. The second model was run using 20, 40, and 100 repetition ensemble models. The models that were using no early stopping methods took on average ~2 hours to complete a single network and ~40 hours to complete a single 20 repetition ensemble. The motivation for the second model is that over-training reduces the bias portion of the error, at the cost of increasing the variance. Later when the ensemble forecasts are averaged, the variance portion of the error is reduced with no effect on the bias itself.

5.7 Engineering Application

Experiment #6: Engineering Application

This experiment applies the lessons learned from the previous experiments to demonstrate an engineering example at a new location. Optimal configurations and model parameters found during previous experiments are used to forecast surge levels for this new location.

5.7.1 Experiment Suite #6: Appling the model at a new location

The location selected for this example is Silvertown, a water level measuring station located on the River Thames approximately 0.5 km downstream from the Thames Barrier. Sea level records from this station were obtained for the years 2000 – 2004. Historical records of when the barrier was closed for water level control purposes are shown in Appendix 6.

Experiment 6.1 Engineering application

From the results found during Experiment #1, it was determined that selection of a training year is not critical, and that large multiyear training data sets are not

necessary for accurate forecasts. Because of this, the validation, training, and testing data sets for this experiment are somewhat arbitrarily selected. The validation data set uses year 2000 data, the training data set uses year 2001 data and the testing data set uses year 2002 data.

The results from Experiment #2 show that using a secondary station significantly reduces forecasting error. ANN forecasts using more than 24-hours of previous water level did not provide significant improvement in forecasting accuracy. Experiment #2 also showed that the optimal selection of a secondary station depended on the forecast interval. For these reasons, this experiment will use a secondary station with 24-hours of previous water level data from both stations for each forecast, and the selection of the secondary station will depend on the forecasting interval.

The results from Experiment #3 show that for single run ANN models, simple [1,1] structures have much less variability than more complex structures. On average, simple [1,1] ANN structures have lower error levels than complex ANN structures, but complex ANN models were more accurate when examining results from an individual model run. The results from Experiment #5 show that ensemble forecasting can greatly reduce the variance problems found when using complex ANN structures, and that by using an ensemble model the variance can be significantly less than single run ANN models. Experiment #5 also showed that by using ensemble forecasting with complex ANN structures, the model error at the time of maximum surge elevation can be reduced. These results guide the selection of the optimal structure used for this experiment. This experiment will use an ensemble ANN model because of the significant reduction in variance over the single ANN model. This ensemble model will use a complex [10,1] ANN structure, to take advantage of the complex ANN structure's ability to better predict maximum surge elevations.

The station selected for Experiment #6 is Silvertown. Its location is 56 km from the Sheerness tide station used in Experiments 1-5. Cross-correlation analysis of the surge components from tidal records for the years 2000 – 2004 show a lag period of 1-hour for surge events to reach Silvertown after being recorded at Sheerness, and a 2-hour lag period for the tide. These lag periods are added to those found in the results for experiment 2.3 to help determine optimum secondary station location for use at Silvertown for each forecast interval.

The final model will be tested using two conditions, when the barrier is open and when the barrier is closed. The closure events are tested only to see how significant the model results are affected. This experiment will give an indication of how well an artificial neural network using only water levels as input performs in a complex hydrodynamic environment near the Thames Barrier.

6 Results and Analysis

To better understand the behaviour and potential of ANN models for storm surge predictions, six suites of experiments are conducted. The first suite of experiments is concerned with the selection of a training data set. Tests are run to determine the effect on performance of an ANN when changing the size and selection of data sets used for training. The second suite determines the optimal design of a neural network for predicting water levels and the performance of the resulting model as compared to other standard models. The third suite of experiments focuses on how varying the structure of an ANN impact upon the robustness and storm performance of the model. The fourth suite of experiments compares performances of the ANN with the Continental Shelf Model (CS3) designed by the Proudman Oceanographic Laboratories for use by the Storm Tide Forecasting Service for the Met Office. The fifth suite of experiments examines the performance of artificial Neural Networks utilizing ensemble forecasting. Finally an engineering application is presented, demonstrating an ANN application.

6.1 Experiment Suite #1: Model performance varying training data

6.1.1 Experiment 1.1 Varying training data set lengths

The lengths of the training data sets are varied from 1 year (1990), 3 years (1990-1992), 5 years (1990-1994), and 10 years (1990-1999) for the Sheerness tide station. One year is considered as the shortest data set length to be used for the training of the ANNs so that seasonal variations of water levels are included. All ANNs are trained using 48 hours of previous water level residuals used for each prediction. The models are then tested on the 2001 data set (a year selected because it was excluded from the training data sets). The results of this experiment are shown in Figure 6-1. Although the average absolute error decreased from 0.153 m when 1 year of training is used to 0.146 m when 10 years of training is used, an average absolute error of 0.146 m for year 2001 can also be obtained by training on the year 1997 data set alone. A substantial disadvantage of using a large training set is that computational time increases from 20 minutes for a 1-year training data set to 20 hours for a 10-year training data set for the computer used for this study. Based on these results, training is conducted using

a 1 year data set (1997), for the remainder of the study, determining which year to use as the training data set is the objective of the next experiment.

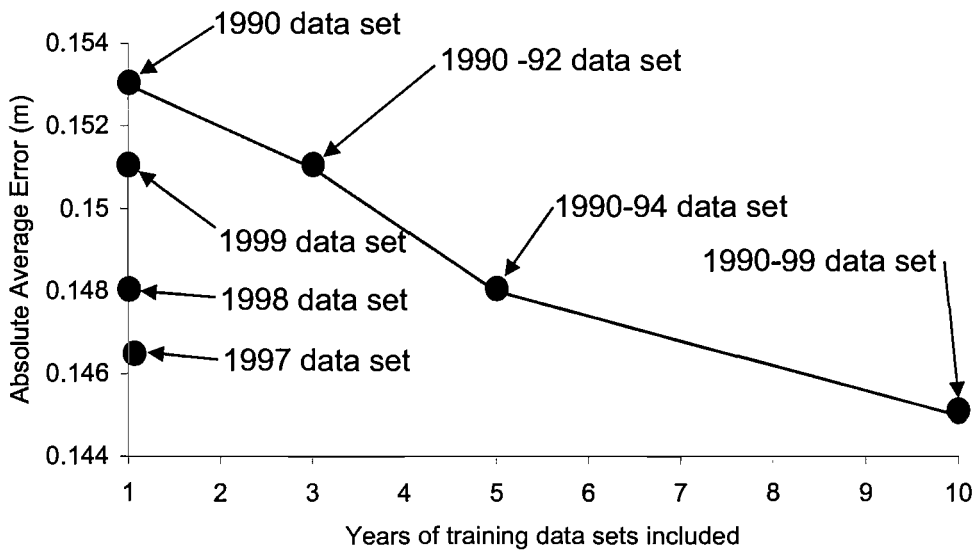


Figure 6-1 Multiple-year training tests. (12-hour forecast using 48 hours of previous data).

6.1.2 Experiment 1.2 Varying the training year selected

A series of basic ANN models using one input neuron and one output neuron are successively trained to predict a 3-hour forecast of water levels at Sheerness. The models are trained using historical water level data sets for years 1990 to 1999. For this model, each 3-hour forecast uses 48 hours of previous water level residuals for each prediction. This is done sequentially for every hour for the entire 1 year data set.

For each training year selected, the model was tested on all other years. The results are displayed in Table 6-1. For each training year, the average absolute error for each testing year was computed. Individual yearly results varied only 0.01 m, with the smallest average absolute error measured using year 1997 data (0.15 m) and the largest was measured using year 1991 data (0.16 m). For all test cases the 1997 training year lead to the lowest average absolute error when applying the model to the test years.

Table 6-1 Training Year Test (3-hour forecast using 48 hours of previous data at Sheerness)

Test Year	Training Year Absolute Average Error (m)									
(1)	1990 (2)	1991 (3)	1992 (4)	1993 (5)	1994 (6)	1995 (7)	1996 (8)	1997 (9)	1998 (10)	1999 (11)
1990		0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
1991	0.16		0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.17
1992	0.16	0.15		0.15	0.15	0.15	0.15	0.15	0.16	0.16
1993	0.16	0.16	0.16		0.15	0.16	0.16	0.15	0.16	0.17
1994	0.16	0.15	0.15	0.15		0.15	0.15	0.15	0.16	0.16
1995	0.16	0.16	0.16	0.15	0.15		0.15	0.15	0.16	0.16
1996	0.15	0.14	0.14	0.14	0.14	0.14		0.14	0.15	0.15
1997	0.14	0.14	0.14	0.14	0.14	0.14	0.13		0.14	0.14
1998	0.16	0.16	0.16	0.15	0.15	0.15	0.15	0.15		0.15
1999	0.16	0.17	0.16	0.16	0.16	0.16	0.16	0.15	0.15	
2000	0.17	0.17	0.17	0.16	0.16	0.16	0.16	0.16	0.15	0.15
Average	0.16	0.16	0.16	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Rank	8	10	9	6	3	4	2	1	5	7

6.2 Experiment Suite #2: Model performance varying input parameters

6.2.1 Experiment 2.1 Single-Station ANN

The ANN is first trained using Sheerness as the primary station; no secondary station is used in this experiment. The forecast time periods used for each model vary from 3, 6, 12 to 24 hours. The number of previous water level residuals to be used as the input set for each model varies from 1, 3, 6, 12, 24, 36, to 48 hours. The models are trained on the 1997 data set, and tested on the 1993 data set. Figure 6-2 shows changes in ANN yearly performance for a [1,1] ANN trained for various forecasting times, and number of previous water levels used for each forecast. Each data point represents an individual ANN model's average absolute error for an entire year. For clarity points with the same forecast intervals are connected.

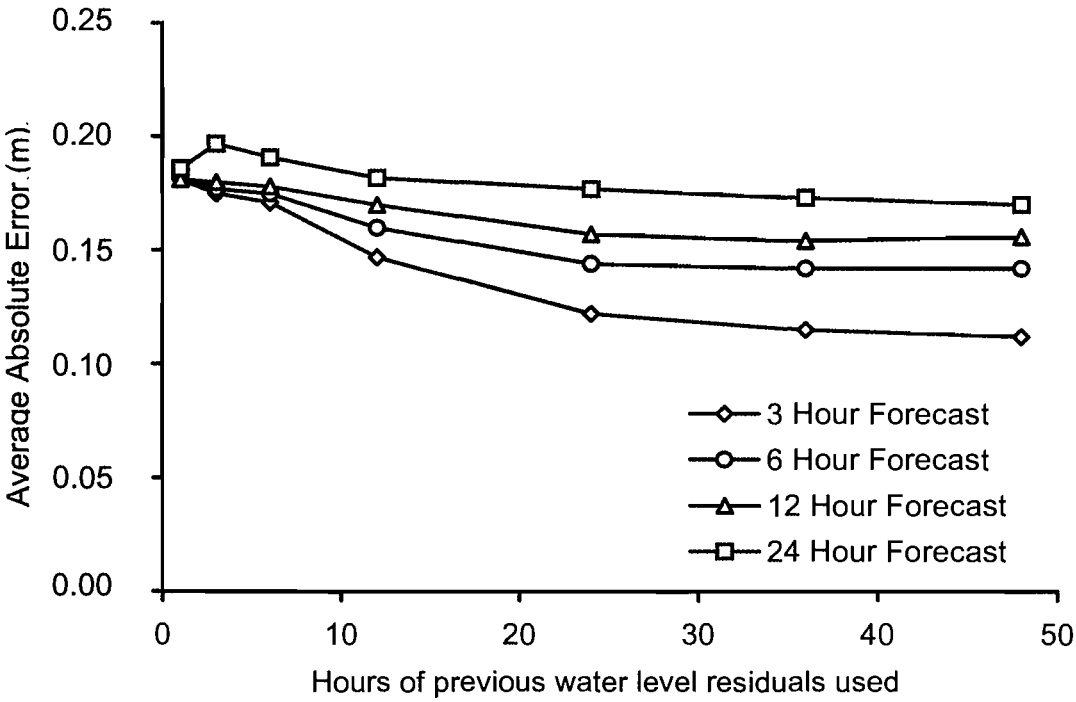


Figure 6-2 Performance of single station ANN model. Each data point shows the average absolute error of a single ANN model.

Model improvement is the largest for the 3-hour forecasts. When increasing the number of previous water level residuals in the input from 1 to 24 hours, the average absolute error decreases from 0.18 m to 0.12 m, an improvement of 0.06 m. The improvements are modest for 6 and 12-hour forecasts with a performance improvement of 0.04 m and 0.03 m respectively. For the 24-hour forecasts, only 0.01 m of improvement is observed. In all cases, very little improvement is found when including more than 24 hours of previous water level residuals from the primary station. It is important to note that the longer term forecasts are less influenced by local history. This is because any significant water level changes that occur locally travel away from the site during the forecast interval. Important future water level information is not found locally, but in stations of increasing distance as the forecast time increases. This concept is covered in the next section.

6.2.2 Experiment 2.2 Two-Station ANN

For this experiment, additional data from a second water level measuring station is included for training and testing the ANN. Immingham is used as the secondary water level measuring station. This secondary station, is located north of the primary station (a direction of travel of a coastally trapped wave), and provides the

ANN with additional information of a storm's presence before it can be detected at the primary station. The ANNs were trained on the 1997 data set and tested using the 1993 data set. The number of previous water level residuals used for each forecast at the primary station at Sheerness is fixed at 24-hours, while the number of previous water level residuals used for each forecast from the secondary station (Immingham), and is changed for each model from 1 to 24 hours. The results for this experiment are presented in Figure 6-3, where each data point represents an individual ANN model's average absolute error for an entire year. For clarity, points with the same forecast intervals are connected. For short-term forecasts (both the 3 and 6-hour) the average absolute error decreases significantly when increasing the number of previous water level residuals of the secondary station. When comparing a single-station ANN (from Experiment 2.1) to a two-station ANN, the average absolute error decreases from 0.10 m to 0.08 m for a 3-hour forecast, and from 0.10 m to 0.09 m for a 6-hour forecast.

The 12-hour and 24-hour forecasts show little improvement when including information from Immingham as a secondary station. This result is not surprising given that storm surge propagation time from the Immingham to Sheerness is approximately 5 to 7 hours. For both the 12-hour and 24-hour forecasts, the storm surge has yet to reach the Immingham station and therefore the data from Immingham is of little help for the forecast of a storm surge at the Sheerness station.

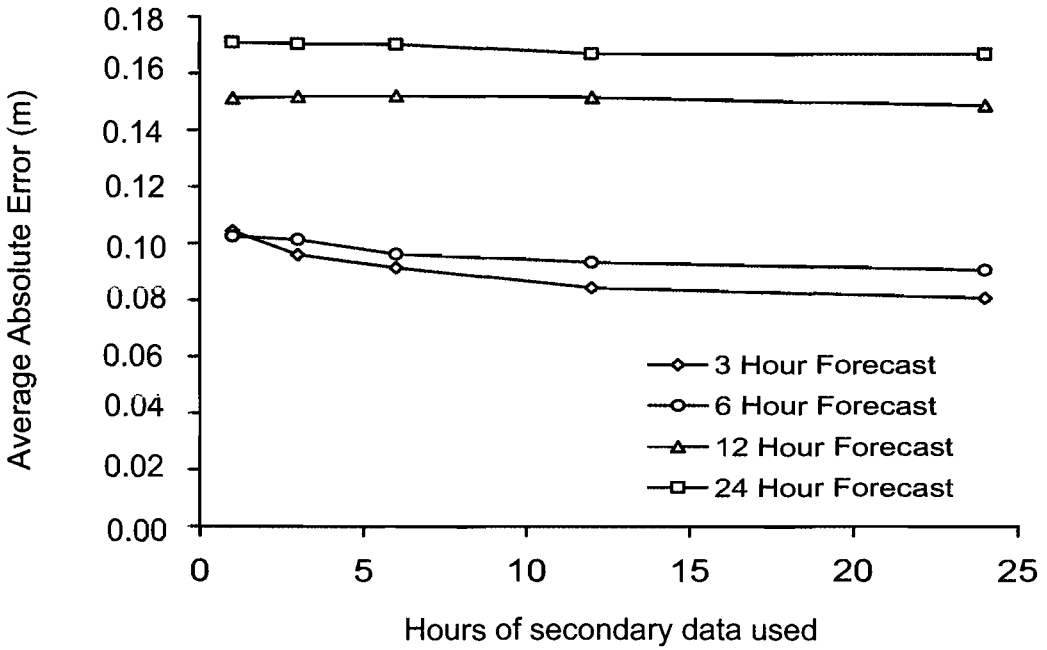


Figure 6-3 Yearly Performance of a two-station ANN model - varying the secondary station input data. Each data point shows the average absolute error of a single ANN model. Note the significant improvement of the 3 and 6-hour forecasts.

6.3 Experiment 2.3 Two-station ANN, varying secondary location

The performance of the ANN is analyzed by varying the location of the secondary station. Three secondary stations Immingham, North Shields and Wick are selected. The stations are respectively 337 km, 510 km and 945 km north of Sheerness. For each test, 24 hours of previous water level residuals from the secondary station data was used, and the number previous water level residuals used from the primary station varied from 1 to 48 hours. The results are shown in Figure 6-4 where each data point represents an individual ANN model's average absolute error for an entire year. For clarity, points with the same forecast intervals are connected. From Figure 6-4a and 6-4b it can be seen that for a 3-hour or 6-hour forecast at Sheerness, Immingham which is located 337 km north of Sheerness is the best choice for a secondary station as it produced the lowest average absolute error. From Figure 6-4c it can be seen that for a 12-hour forecast at Sheerness, North Shields, which is located 510 km North of Sheerness is the best choice for a secondary station. From Figure 6-4d it can be seen that for a 24-hour forecast as Sheerness, Wick, which is located 945 km north of

Sheerness, performed the best, although only marginally better than using no secondary station.

It is important to note that during each experiment, the inclusion of a secondary station always improves model performance. The amount the performance is improved is related to the position of the secondary station relative to the storm at the time of the forecast. The station that is located closest to the surge peak at the time of the forecast was found to have the best performance. For the 3 and 6-hour forecasts, the closest station to the surge peak at the time of the forecasts was Immingham, which had the best performance (Figures 6-4a and 6-4b). For the 12-hour forecast, the station closest to the storm surge peak at the time of the forecasts was North Shields, which performed best (Figure 6-4c). Finally, for the 24-hour forecasts, the station closest to the peak of the storm surge at the time of the forecast was Wick, which had the best performance (Figure 6-4d). In this case, although Wick had the best performance, it was only marginally better the other secondary station locations. A surge located 24 hours away from the primary station will have not peaked yet at any of the secondary station locations at the time of the forecast, but Wick performed best because it had the best information on the changing surge of the approaching storms.

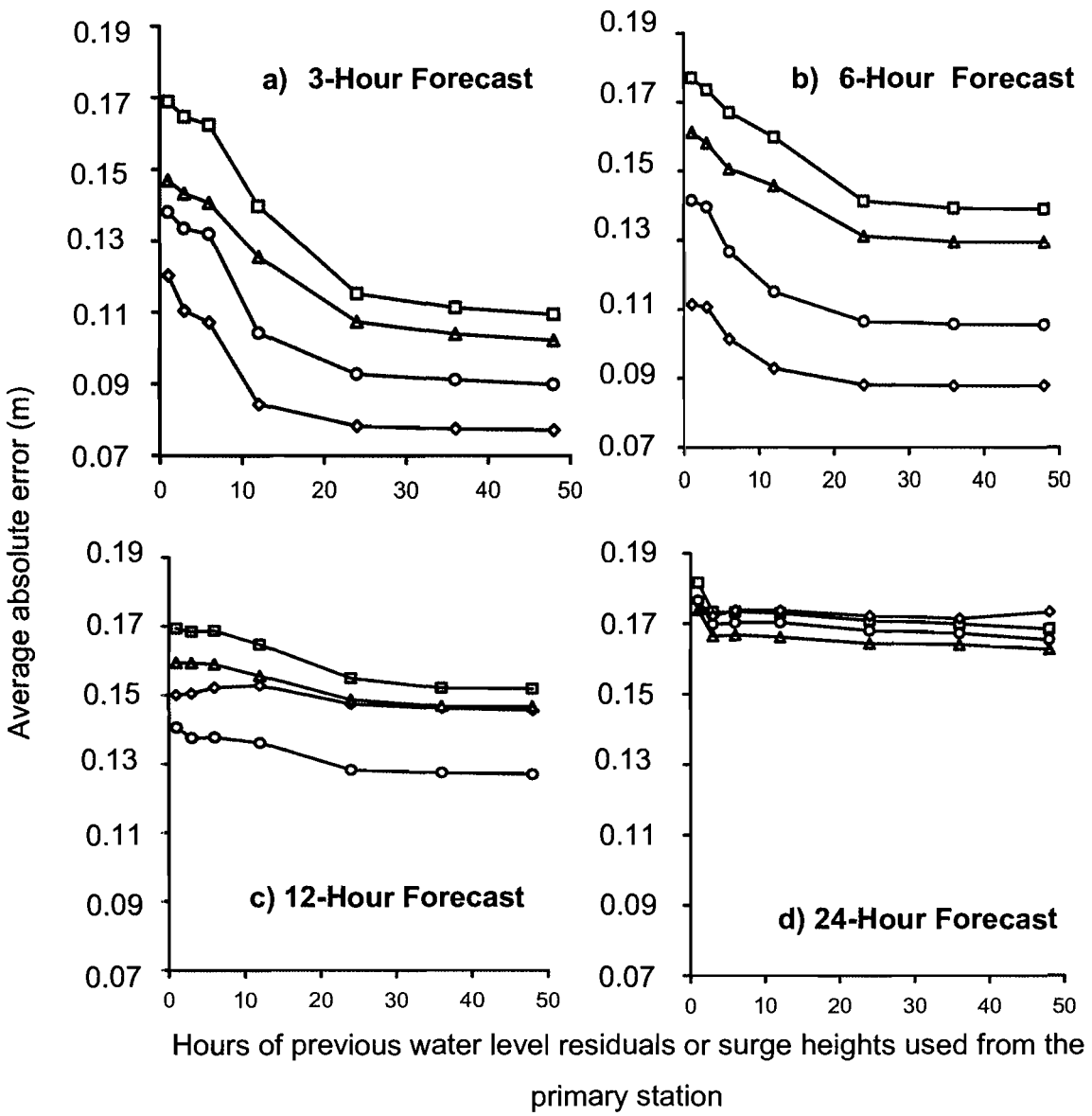


Figure 6-4 Performance of a two-station ANN model - varying the secondary location. Square – No secondary station, Diamond – Immingham, Circle – North Shields, Triangle – Wick. Each data point shows the average absolute error of a single ANN model.

The effectiveness of a secondary station location is determined by its proximity to the primary station. The selection of the secondary station should be based on the forecast interval and its distance from the primary station. This effective range varies with the forecast interval. For example, in Figure 6-4c when using Immingham for a 12-hour forecast, the average absolute error is 0.15 m; this error drops to 0.13 m when switching the secondary station to North Shields.

Cross-correlation comparison

To help explain the results found using the ANN model a cross-correlation analysis test was run to determine typical lag periods between stations for surge and tide events. The tidal and surge lag periods were determined for each station from the tide records archived by the National tidal and sea level facility. Both the tide and the surge behave as long waves and their propagation times can be found by a simple cross correlation analysis tests. The tests were run for the years 1990 – 2002, and the results are shown in Table 6-2. The tidal lag and surge lag periods indicate the best correlated time-lag found during the test interval. This amount of time was different for each indicating that the surge component travels at a faster speed than the tidal component of the change in water level. Also as expected the lag period increases as the distance to Sheerness increases.

Table 6-2 Cross-correlation analysis test for Sheerness station (Average for Years 1990 – 2002).

(1)	Tidal Lag Period(hrs) (2)	Surge Lag Period(hrs) (3)
Immingham	8	6
North Shields	10	9
Wick	15	14

Figure 6-5 shows how the correlation coefficient (vertical axis) varies for each station while changing the time offset from Sheerness (horizontal axis). These offset values indicate why a specific secondary station location worked best for a particular forecast interval. It should be noted that the station closest to Sheerness has the best correlation coefficient with a value of 0.79 compared to Wick with a value of only 0.02. This suggests that the closer a secondary station is to the primary station, that the ANN will derive a stronger relationship and use this to better predict water levels. It is also important to note that cross-correlations can be misleading in part because linear relationships are implied when using them and this project has chosen to use ANNs because the relationships involved are non-linear.

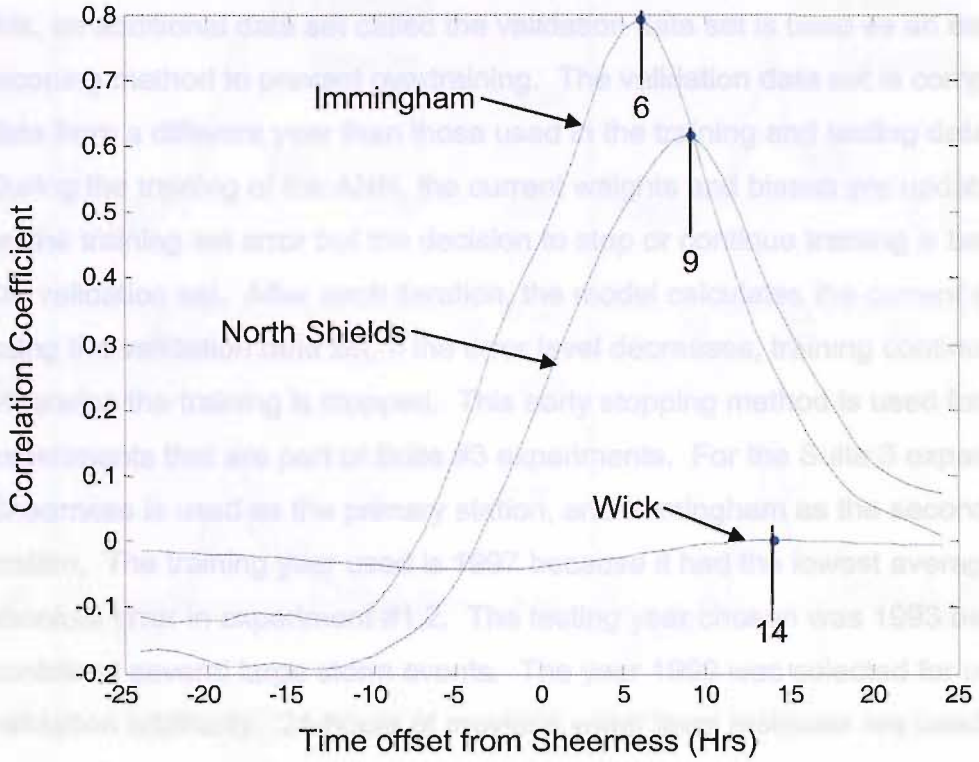


Figure 6-5 Cross-correlation analysis test for Sheerness and the selected tide stations comparing surge elevations (1992 data).

In conclusion, for a secondary station to be effective in predicting a storm surge event, it must provide some information about the storm surge's existence at the time of the forecast. With this additional information represented by an increase in surge height at the secondary station, the secondary data set can then be used by the ANN to more accurately predict the surge height and arrival time at the primary station. The optimal secondary station location is one where the distance between primary and secondary stations is similar to the distance a typical storm surge would take to travel between the stations.

6.4 Experiment Suite #3: ANN topology and performance

This suite of experiments investigates how model performance is affected by changes in ANN topology. While the performance is tracked for both yearly averages and during storms, the discussions will focus primarily on the storm performance because of its importance for safety and commerce. Experiment 3.1 varies the ANN structure size, by changing the number of hidden neurons used. Experiment 3.2 tests the robustness of the ANN model. During initial runs of Experiment 3, models that used more than 5 hidden neurons experienced

overtraining problems, causing very large variations in predictions. Because of this, an additional data set called the validation data set is used as an early stopping method to prevent overtraining. The validation data set is composed of data from a different year than those used in the training and testing data sets. During the training of the ANN, the current weights and biases are updated based on the training set error but the decision to stop or continue training is based on the validation set. After each iteration, the model calculates the current error level using the validation data set, if the error level decreases, training continues, otherwise the training is stopped. This early stopping method is used for all experiments that are part of Suite #3 experiments. For the Suite 3 experiments, Sheerness is used as the primary station, and Immingham as the secondary station. The training year used is 1997 because it had the lowest average absolute error in experiment #1.2. The testing year chosen was 1993 because it contained several large storm events. The year 1999 was selected for use as a validation arbitrarily. 24-hours of previous water level residuals are used for training for each station.

6.4.1 Experiment 3.1: ANN performance varying structure size

In this experiment, the performance of ANN models with different numbers of hidden neurons is investigated. 3-hour ANN forecasts were calculated for each of the following size models: 1,2,5,10,20, and 50 hidden neurons. Figure 6-6 shows the results for different size ANN models for the February 19, 1993 storm. The increase in the number of hidden neurons leads to a better maximum storm surge prediction than a [1,1] ANN structure (1 hidden neuron, 1 output neuron). However the use of 10 or more hidden neurons also leads to increased noise and variability in the predictions, primarily during storm events.

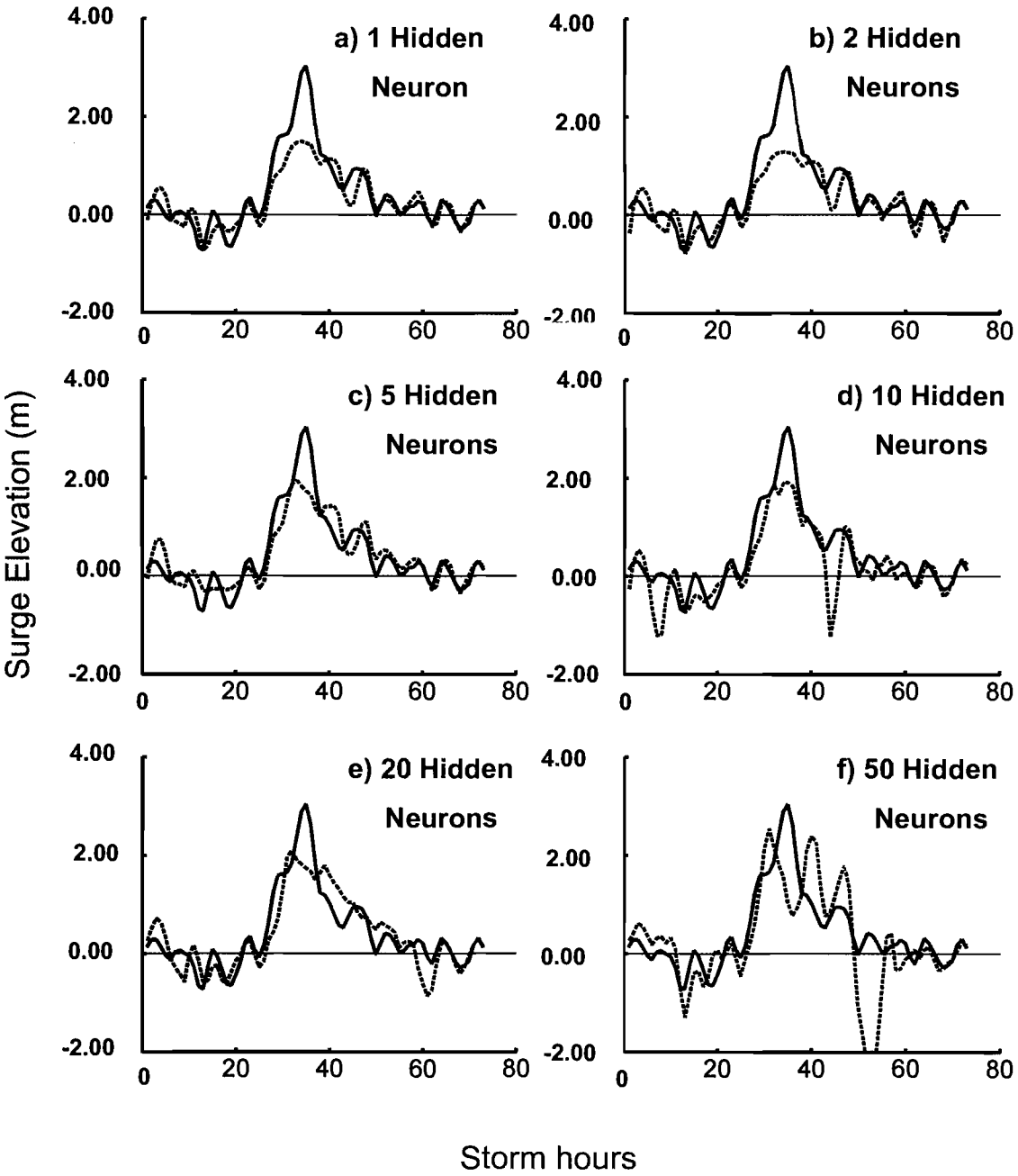
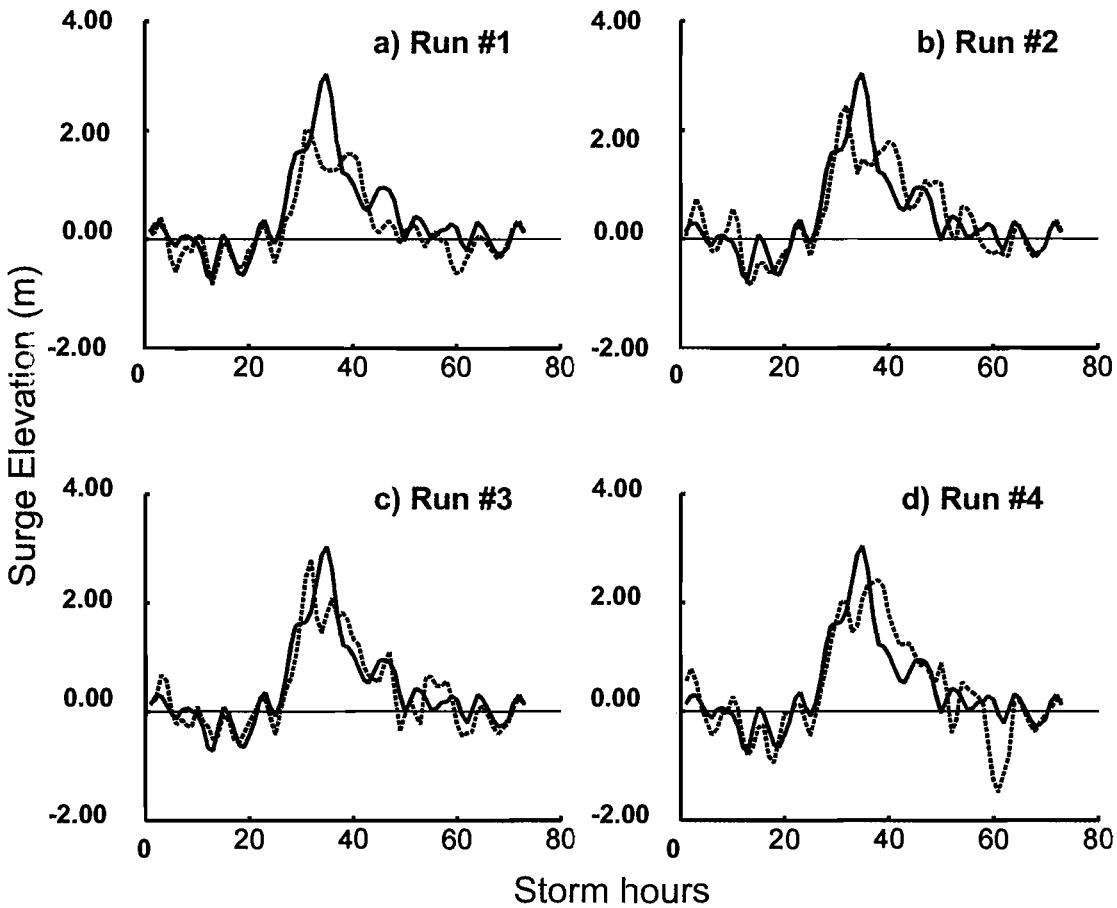


Figure 6-6 Performance of ANN for an example storm - varying the number of hidden neurons. Dashed line - ANN Model, Solid line - Measured Surge elevation. (72-hour window centred on the February 19, 1993 storm event)

6.4.2 Experiment 3.2: ANN performance variability.

The tests performed for experiment 3.1 were repeated 20 times with random starting values for the weights and biases. Figure 6-7 shows the results of four ANN test runs using the same initial starting parameters, and illustrates visually, the variability of individual ANN forecasts. The variability arises from the Levenberg Marquardt algorithm, which assigns random values when initializing weights and biases during training (such is the case for most ANN training

algorithms). This variability was found to only be significant when using larger ANNs (greater than 5 hidden neurons).



Fgure 6-7 Performance of ANN during multiple runs using the same ANN structure (20 hidden Neurons). Dashed line – ANN Model, Solid line – Measured Surge Elevation. (72-hour window centered on February 19, 1993 storm event). Differences observed between each run are due to random initialization of weights and biases during the training period.

The results of these experiments show that larger ANN topologies can have a considerable impact on the variability of an ANN when measured during a storm. Quantitatively very little change is observed for the yearly averages, but significant differences are observed for the short-term evaluation of storm events. The varying model performance with an increasing number of hidden neurons is further illustrated in Figure 6-8. While small ANNs have the best average performance, the overall best performance is always reached by an individual instance of a large ANN. However the repeatability and variability of large ANN predictions during storms is a concern. The error bars in the Figure 6-8 illustrate the range of average absolute error and central frequency (15cm) % obtained for the 20 runs for each case, using 1, 2, 5, 10, 20, and 50 hidden neurons. The average

performance decreases (increase in the average absolute error and decrease in the average central frequency (15cm) %) when increasing the number of hidden neurons. The variance of the predictions also increases significantly with the number of hidden neurons. Based on this experiment, selecting smaller ANNs and in particular [1,1] ANNs leads to better average performance during the selected storm. However the best performance based on both average absolute error and central frequency (15cm) % was obtained for one of the [10,1] models and at least one of the implementations of each of the larger models had a performance better or equivalent to the small [1,1] model.

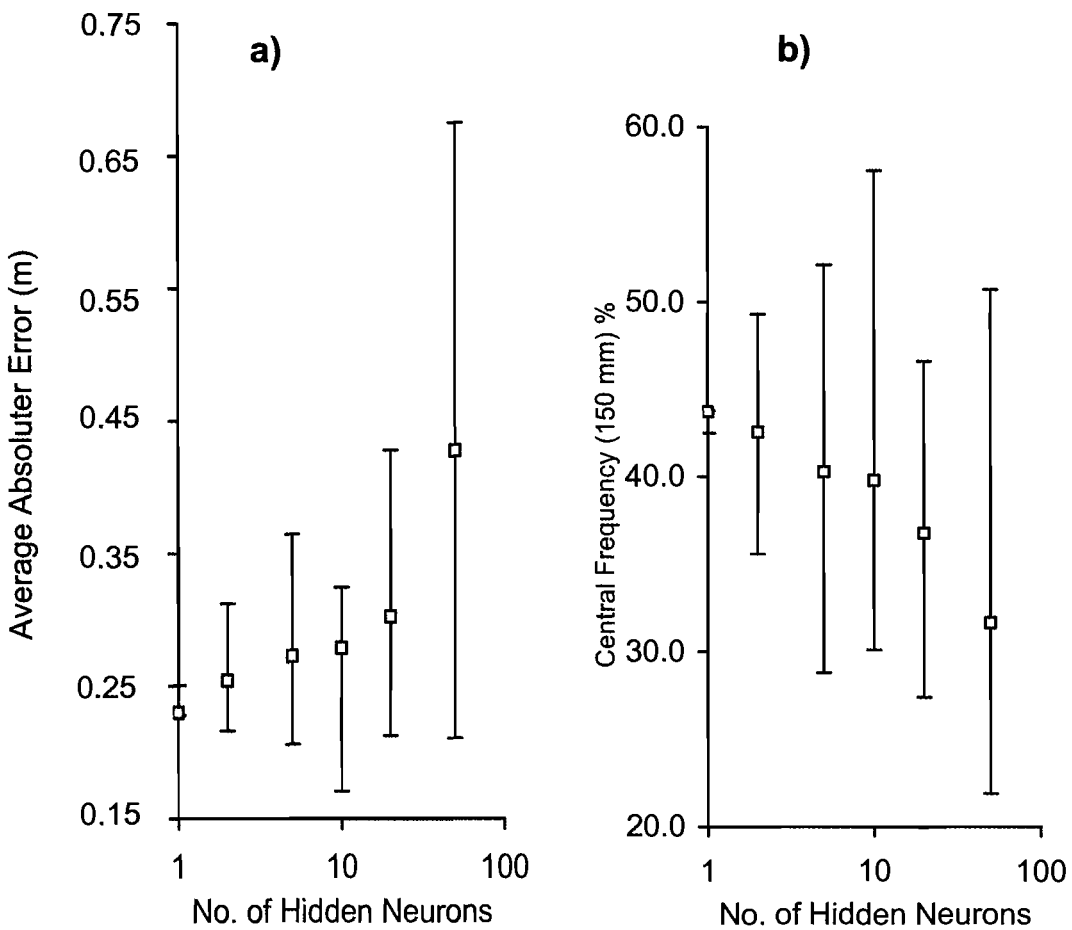


Figure 6-8 ANN performance and variation during the storm event of Feb 19, 1993. Squares mark the average value for each size ANN. Bars show the range of variation of the forecasts.

6.5 Experiment Suite #4: Artificial Neural Network model / CS3 model comparison

Experiment suite #4 compares the performance of the Artificial Neural Network (ANN) model to the CS3 model on a yearly basis and using a 72-hour storm

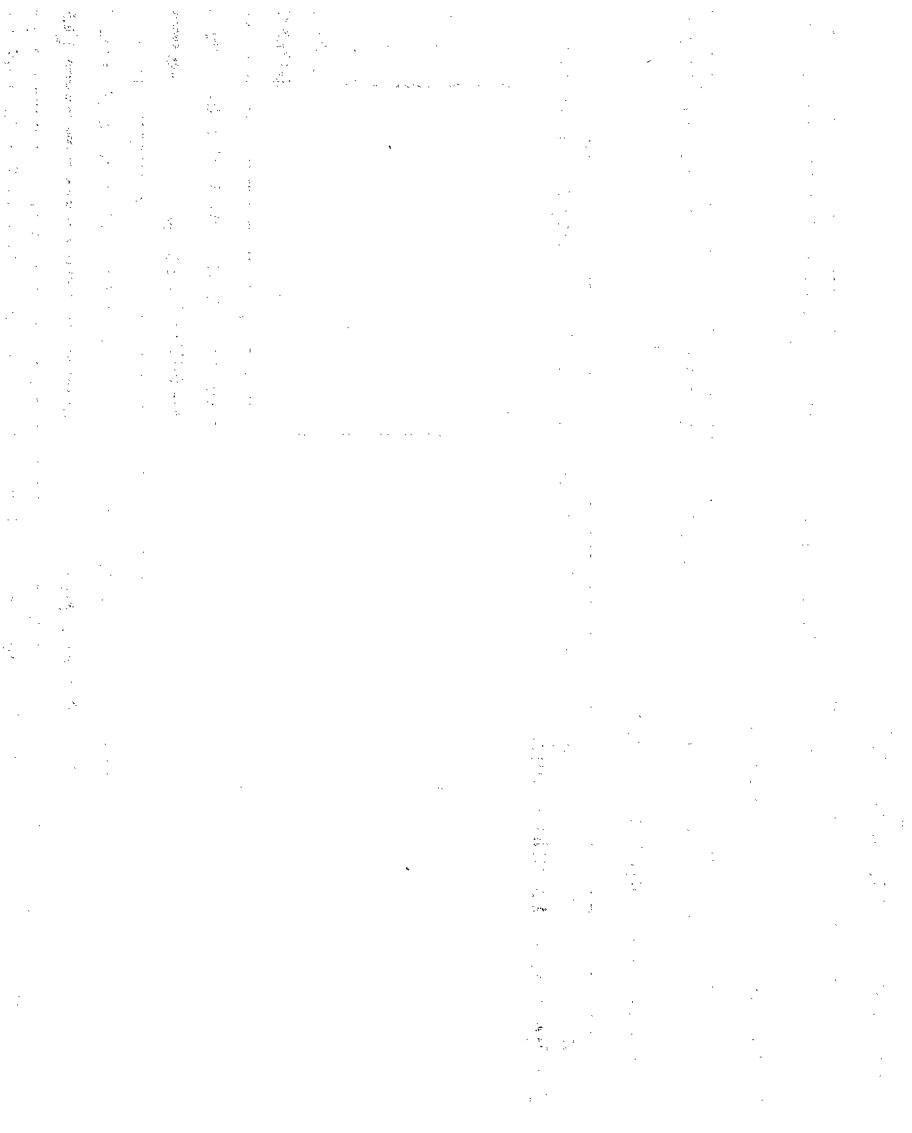
window. The artificial neural network model used for this comparison uses Sheerness as the primary station and Immingham as the secondary station. 24-hours of previous water level residuals were used for each station. The network was trained using year 1997 data, validated using year 2001 data and tested on years 1999 and 2002. The data set years used for this experiment were changed for two reasons. First, the CS3 archived model (full year sets) were limited to years 1999 and 2002; this restricted testing to these two years. Second, since testing was on year 1999, it could not be used as a validation year (2001 was used).

The CS3 model predicts new water levels every 6 hours for time periods designated T+00 to T+36 (Figure 6-9a). The data is archived only for times T+00 to T+05 for each forecast (Figure 6-9b). This archived format only allows comparison of the 1, 3, and 5-hour forecasts at 6-hour intervals as shown in Figure 6-9c, Figure 6-9d and Figure 6-9e. When the CS3 model is used for surge forecasts at a location with a tide gauge, the model bias can be removed by calculating the model error at time T+00 and subtracting it from future forecasts. Three statistical methods were used for comparing the accuracy of the forecasts in this experiment, root mean square error (RMSE), average absolute error (AAE), and central frequency (CF).

6.5.1 Experiment 4.1: Model comparison using yearly performance

For the years 1999 and 2002 the yearly performances of both models were compared using the root mean square error, central frequency, and average absolute error. The results are shown in Table 6-3. The ANN model out-performs the CS3 model in all forecast intervals for both testing years. The one hour forecast for the ANN model was 0.05 m compared to 0.12 m for the CS3 model with the bias removed. For 5-hour forecasts, the ANN model performs only slightly better, with a RMSE of 0.15 m compared to 0.18 m for the CS3. It is interesting to note that the CS3 model performed less well with its bias removed for the 3 and 5 hour forecasts than without removing it. Removal of the bias in the CS3 model only helps in short -term forecasts. When comparing the Central Frequency (15cm), The ANN model out-performs the CS3 model in short 1-hour forecasts with a CF (15cm) of 98.2% for the ANN model compared to only 64.2% for the CS3 model with bias removed. For 5-hour forecast intervals, the ANN model performed with a CF of 83.6% compared to 64.3% for the CS3 model alone.

When comparing the average absolute error (AAE), the ANN model out-performs the CS3 model in short 1-hour forecasts with an AAE of 0.03 m for the ANN model compared to 0.10 m for the CS3 model with bias removed. For 5-hour forecast intervals the ANN model performed at 0.10 m compared to 0.14 m for the CS3 model alone. Similar results were found when comparing results from the year 2002 data set.



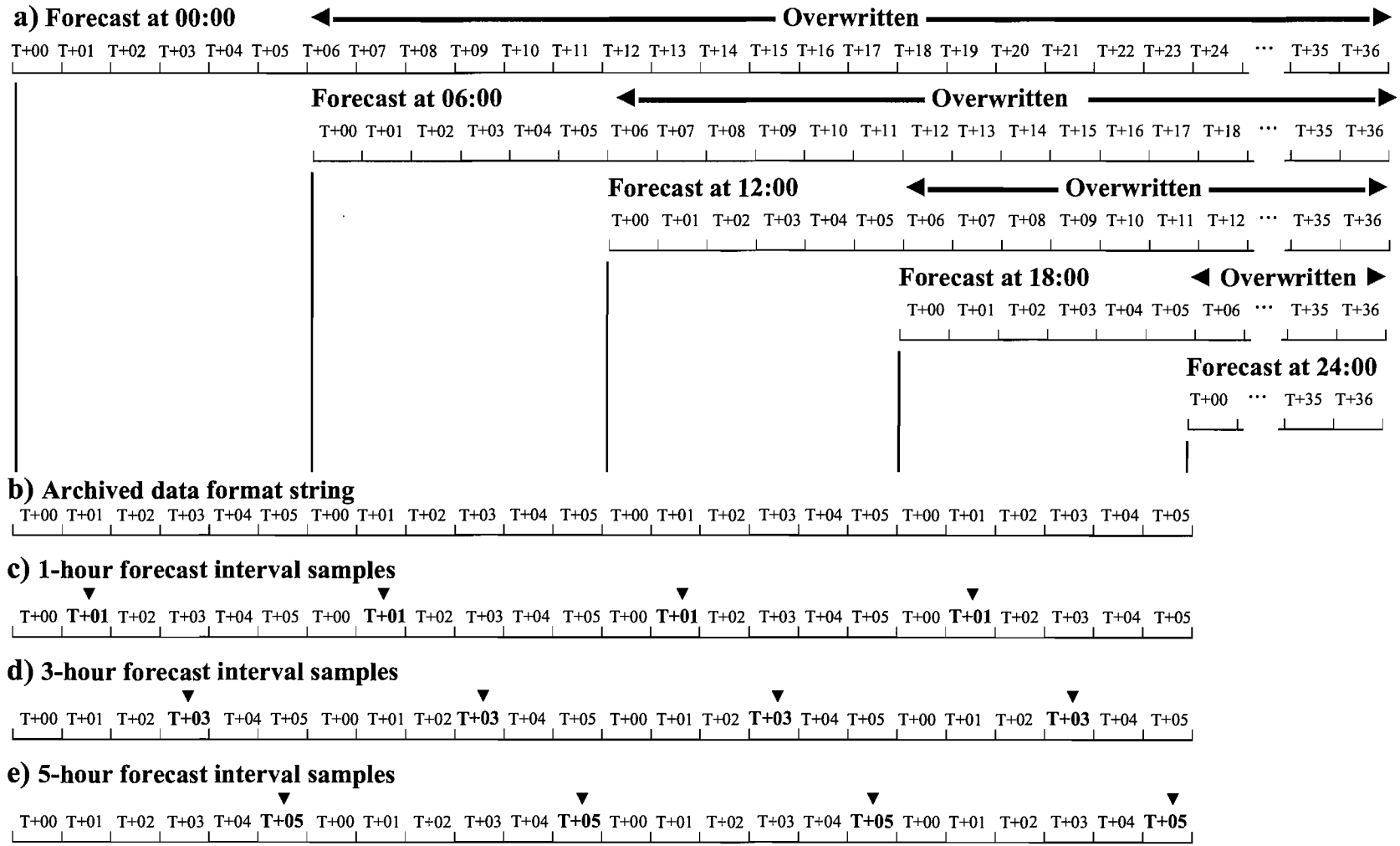


Figure 6-9 CS3 model archived data format. a) Shows original data and how each additional forecast overwrites the T+06 – T+36 forecast intervals from the previous forecast. b) Shows the final archived data format. c) Shows location of 1-hour forecast samples. d) Shows location of 3-hour forecast samples. e) Shows location of 5-hour forecast samples.

Table 6-3 Yearly comparison, Artificial Neural Network (ANN) vs. Storm Tide Forecasting Service (STFS). Models with the best performances are shaded in grey.

Year	Forecast Interval	Root Mean Square Error(m)			Central Frequency (±15cm)%			Average Absolute Error (m)		
		ANN	CS3	CS3 (Bias Removed)	ANN	CS3	CS3 (Bias Removed)	ANN	CS3	CS3 (Bias Removed)
(1)	(2)	(3)	(4)	(5)	(5)	(6)	(5)	(7)	(8)	(5)
1999	1	0.05	0.18	0.12	98.22	59.45	64.23	0.03	0.15	0.10
	3	0.15	0.18	0.22	84.59	61.10	59.36	0.09	0.14	0.17
	5	0.15	0.18	0.22	83.63	64.32	61.48	0.10	0.14	0.18
2002	1	0.09	0.24	0.12	97.88	52.81	65.25	0.04	0.18	0.09
	3	0.13	0.24	0.23	82.19	54.32	55.36	0.09	0.18	0.17
	5	0.15	0.23	0.24	82.60	53.49	51.88	0.10	0.18	0.19

6.5.2 Experiment 4.2: Model comparison using storm performance

Another, more important comparison of the respective model performances can be made when analyzing individual storm events. Because most of the time water levels are near normal elevations, most surge models can be expected to perform well when a high percentage of the time no storms are present. Yearly statistics do not show how well a model performs at the time we are most interested, i.e. during a storm event. Because of this, the performance of the models is analyzed during storm events. For the purpose of this research, a storm event is identified when surge elevations exceed 1.0 m. These storm events were evaluated using the archived CS3 model results and a simple [1,1] ANN model using 24-hours of previous water level residuals from the primary and secondary stations. The results are shown in Table 6-4 for a 1-hour forecast comparison, Table 6-5 for a 3-hour forecast comparison, and in Table 6-6 for a 5-hour forecast comparison. Note that 1-hour forecasts are of little value for public safety and shipping concerns, and are presented here for purposes of model comparison only.

1-hour forecast comparison

Performances of the two models were compared using the root mean square error (RMSE) and the average absolute error (AAE) of each model during a 72-hour storm window. For the year 1999, seven storms with surges greater than 1.0 m were found in the CS3 model's 1-hour forecast data-set (Figure 6-9c), and three storms were found for the year 2002. The results are shown in Table 6-4. The model's performance for individual storms and their averages are shown for each year. The models with the lowest error levels are shaded in grey. For the 7 storm events in 1999, the ANN model had the better performance with a RMSE of 0.08 m compared to 0.13 m for the CS3 model with bias removed. When using AAE as

a reference the ANN model also had the better performance with a value of 0.07 m compared to 0.11 m for the CS3 model with bias removed. For the 3 largest storm events in 2002, the CS3 model with bias removed had the better average performance with a RMSE of 0.18 m compared to 0.21 m for the ANN model. When using AAE as a reference the ANN model had the better performance with a value of 0.12 m compared to 0.14 m for the CS3 model with bias removed.

Table 6-4 Comparison of 1-hour forecasts using ANN and CS3 models. Results are from values sampled during a 72–Hour Storm window. Models with the best average performance are shaded grey. Peak surge elevation and times are indicated in columns 9 and 10.

Year	Storm #	ANN RMSE (m)	CS3 RMSE (m)	CS3 RMSE (Bias Removed)	ANN AAE (m)	CS3 AAE (m)	CS3 AAE (Bias Removed)	Elev (m)	Time (Hrs)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1999	1	0.18	0.18	0.10	0.11	0.15	0.07	1.57	853
	2	0.07	0.24	0.16	0.05	0.21	0.13	1.20	913
	3	0.08	0.18	0.10	0.07	0.12	0.08	1.06	1135
	4	0.05	0.20	0.16	0.04	0.15	0.14	1.14	1273
	5	0.08	0.29	0.14	0.06	0.24	0.11	1.67	7435
	6	0.08	0.35	0.13	0.06	0.28	0.10	1.01	7933
	7	0.12	0.22	0.13	0.09	0.14	0.10	1.38	8419
Average		0.08	0.24	0.13	0.07	0.18	0.11		
2002	1	0.10	0.30	0.22	0.08	0.24	0.18	1.61	1225
	2	0.09	0.36	0.17	0.07	0.28	0.12	1.18	1267
	3	0.43	0.30	0.15	0.20	0.24	0.12	1.85	7201
Average		0.21	0.32	0.18	0.12	0.25	0.14		

3-hour forecast comparison

For the year 1999, six storms with surges greater than 1.0m were found in the CS3 model's 3-hour forecast data-set (Figure 6-9d), and three storms were found for the year 2002. The results are shown in Table 6-5. The model's performance for individual storms and their averages are shown for each year. The models with the lowest error levels are shaded in grey. For the six storm events for 1999, the CS3 model with bias removed had the best performance with a RMSE of 0.21 m compared to 0.23 m for the ANN model. When using average absolute error (AAE) as a statistic, the ANN and the CS3 model with bias performed equally, with both models performing with a value of 0.17 m. For the 3 largest storm events for 2002, the CS3 model with bias had the better performance with a RMSE of 0.32 m compared to 0.35 m for the ANN model. When using AAE as a reference the CS3 model with bias had the better performance with a value of 0.26 m compared to 0.30 m for the ANN model.

Table 6-5 Comparison of 3-hour forecasts using ANN and CS3 models. Results are from values sampled during a 72–Hour Storm window. Models with the best performance are shaded grey. Peak surge elevation and times are indicated in columns 9 and 10.

Year	Storm #	ANN RMSE (m)	CS3 RMSE (m)	CS3 RMSE (Bias Rem-oved)	ANN AAE (m)	CS3 AAE (m)	CS3 AAE (Bias Rem-oved)	Elev (m)	Time (Hrs)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1999	1	0.32	0.17	0.21	0.24	0.15	0.18	1.14	849
	2	0.16	0.19	0.18	0.10	0.15	0.15	1.12	1113
	3	0.22	0.20	0.17	0.15	0.16	0.15	1.60	1137
	4	0.21	0.23	0.32	0.18	0.18	0.26	1.23	7437
	5	0.16	0.23	0.26	0.13	0.18	0.21	1.32	8019
	6	0.31	0.27	0.33	0.22	0.22	0.28	1.16	8085
Average		0.23	0.22	0.21	0.17	0.17	0.18		
2002	1	0.35	0.32	0.38	0.32	0.27	0.30	1.36	1227
	2	0.38	0.33	0.43	0.35	0.26	0.39	1.02	1263
	3	0.32	0.31	0.35	0.23	0.24	0.31	1.46	7197
Average		0.35	0.32	0.39	0.30	0.26	0.33		

5-hour forecast comparison

For the Year 1999, seven storms with surge values greater than 1.0m were found in the CS3's 5-hour forecast data-set (Figure 6-9e), and three storms were found for the year 2002. The results are shown in Table 6-6. The model's performance for individual storms and their averages are shown for each year. The models with the lowest error levels are shaded in grey. For the six largest storm events for 1999, the CS3 model with bias had the best performance with a RMSE of 0.19 m compared to 0.31 m for the ANN model. When using AAE as a reference the CS3 model with bias had the best performance with a value of 0.15 m compared to 0.22 m for the ANN model. For the 3 largest storm events for 2002, the CS3 model with bias had the best performance with a RMSE of 0.23 m compared to 0.48 m for the ANN model. When using AAE as a reference the CS3 model with bias had the best performance with a value of 0.20 m compared to 0.28 m for the ANN model.

Table 6-6 Comparison of 5-hour forecasts using ANN and CS3 models. Results are from values are sampled during a 72–Hour Storm window. Models with the best performance are shaded grey. Peak surge elevation and times are indicated in columns 9 and 10.

Year	Storm #	ANN RMSE (m)	CS3 RMSE (m)	CS3 RMSE (Bias Rem-oved) (m)	ANN AAE (m)	CS3 AAE (m)	CS3 RMSE (Bias Rem-oved) (m)	Elev (m)	Time (Hrs)
(1)	(2)	(3)	(4)	(5)	(5)	6)	(5)	(7)	(8)
1999	1	0.56	0.15	0.22	0.29	0.11	0.16	1.09	371
	2	0.36	0.13	0.17	0.27	0.10	0.14	1.80	839
	3	0.36	0.15	0.19	0.28	0.12	0.16	1.82	851
	4	0.21	0.12	0.20	0.18	0.09	0.15	1.01	1139
	5	0.18	0.26	0.26	0.13	0.20	0.21	1.24	7433
	6	0.29	0.32	0.27	0.24	0.25	0.21	1.09	7931
	7	0.21	0.20	0.27	0.17	0.17	0.22	1.06	8423
Average		0.31	0.19	0.23	0.22	0.15	0.18		
2002	1	0.42	0.19	0.28	0.29	0.18	0.23	1.08	1223
	2	0.37	0.30	0.45	0.24	0.25	0.39	1.50	1265
	3	0.65	0.19	0.34	0.31	0.17	0.30	1.74	7199
Average		0.48	0.23	0.35	0.28	0.20	0.31		

In conclusion, for the forecast intervals studied, the ANN performed better than the CS3 model with or without a bias correction when comparing yearly statistics. When measuring storm performances, the statistical analysis show that the ANN model performs better in the short 1-hour forecasts, results were mixed with similar model performances in the 3-hour forecasts but the CS3 model without bias removed performs better in the longer 5-hour forecasts. In general the CS3 model performance was consistent, with the size of the forecast errors independent of the forecast interval, while the ANN performance always decreased when the forecast interval was increased. Forecast performances for the CS3 model were significantly improved by removal of the bias for short 1-hour forecasts, but were actually reduced for most of the 3-hour forecasts, and all of the 5-hour forecasts.

6.6 Experiment Suite #5: Ensemble Forecasting

Experiment suite #5 is designed to explore how using ensemble forecasting can increase forecasting accuracy. This is done by running the following four experiments:

- Experiment 5.1 will determine how the accuracy and variability of a model changes as the number of repetitions are increased in an ensemble.
- Experiment 5.2 will determine how the accuracy and variability of a model changes as the size of an ANN structure is increased.

- Experiment 5.3 will determine the effect of structure size on peak water and surge elevations
- Experiment 5.4 will determine the effect of over-training on forecast variance.

All Ensemble forecasts are reported to the 0.001 m accuracy in this case to better illustrate the case to case variability. Each forecast can include up to 50 individual repetitions of a single ANN model.

6.6.1 Experiment 5.1: Ensemble forecasts, changing the number of repetitions used per ensemble

Experiment 5.1 is performed to see how changing the number of repetitions used for an individual ensemble forecast effects the variance. All models in this experiment used the same [5,1] structure. The storm event selected for this experiment was November 5, 1999. This storm was selected because it was a typical moderate size storm with a maximum water level residual (or surge height) of 1.24 meters. Sheerness was used as the primary station, and Immingham was used as the secondary station, 24 hours of previous data was used from each station. The data sets that are used for this experiment are: test year 1999, training year is 1997, and validation year is 2001.

For reference, 20 individual ANN models were run and their average absolute error and variance is calculated, and the results are shown in Table 6-7, column #2. These 20 individual or single run ANN models are used as a reference to measure the performance of the ensemble models. The ensemble models were run using 5, 10, 20, and 50-repetitions of the original ANN model. Each of the ensemble models was run 20 times. This resulted in a total of 1720 runs of the original ANN model. The average absolute error (AAE) of each run is shown in Table 6-7, columns 3 – 6. The average of the AAE for all 20 runs is shown for each model in Table 6-7. Values are for a 4-hour forecast, and sampled during a 72-hour storm window. The results show that the average AAE for 20 single runs used for reference is 0.207 m. The average AAE for the 5-repetition ensemble models is 0.207 m, and decreases to 0.202 m for a 50-repetition ensemble forecast. The results also show that the variance of the ensemble models decreases as the number of repetitions used for each ensemble is increased.

The variance of each model is calculated and also expressed as a percent of the reference model. The reference model (essentially an ensemble model with only one repetition) had a variance of 5.7×10^{-5} m^2. The 5 repetition ensemble

model has a variance of $2.7x10^{-6}$ m^2 or 47.8% of the variance of the reference model. The 50 repetition ensemble model has a variance of $3.0x10^{-6}$ m^2 or 5.1% of the variance of the reference model. This shows a significant decrease in the variance of the ensemble model when compared to the simple or single ANN model. The reference models average absolute error ranges from 0.195 m to 0.222 m, compared to the 50-repetition ensemble models average absolute error ranges from 0.198 m to 0.205 m.

Table 6-7 Average absolute error (m) for 4-hour Ensemble forecasts changing the number of repetitions used per model. 72–hour storm window, November 5, 1999, Primary station: Sheerness, Secondary station: Immingham, Test Year 1999, Training Year 1997, Validation Year 2001. Note: All forecasts used a [5,1] structure.

Ensemble Run Number	◄Few	Number of Repetitions Per Ensemble (Repetitions)			Many►
	1	5	10	20	50
(1)	(2)	(3)	(4)	(5)	(6)
1	0.211	0.205	0.209	0.207	0.205
2	0.209	0.213	0.206	0.205	0.202
3	0.212	0.201	0.207	0.206	0.202
4	0.197	0.212	0.204	0.199	0.203
5	0.213	0.207	0.209	0.202	0.203
6	0.205	0.208	0.206	0.203	0.203
7	0.210	0.206	0.207	0.201	0.201
8	0.195	0.203	0.204	0.205	0.198
9	0.211	0.204	0.203	0.207	0.204
10	0.209	0.204	0.212	0.205	0.203
11	0.212	0.218	0.199	0.206	0.203
12	0.197	0.211	0.200	0.199	0.201
13	0.213	0.199	0.206	0.202	0.205
14	0.205	0.202	0.199	0.203	0.202
15	0.210	0.208	0.202	0.201	0.202
16	0.195	0.209	0.205	0.205	0.202
17	0.203	0.206	0.205	0.202	0.203
18	0.222	0.208	0.198	0.204	0.201
19	0.195	0.217	0.206	0.202	0.202
20	0.212	0.200	0.206	0.206	0.199
Average	0.207	0.207	0.205	0.204	0.202
Variance (m^2)	0.000057	0.000027	0.000013	0.000006	0.000003
% of Reference (Single repetition model)	100.00	47.82	23.27	10.60	5.09

Real time use of ensemble forecasting requires saving of the weights and biases during each training iteration. These can be saved as an ASCII text file or in Excel format. These values can be re-applied later using any programming language and real-time sea-level data to predict surge elevations in a sub-second process.

6.6.2 Experiment 5.2: Ensemble forecasts, varying the ANN structure size

Experiment 5.2 was performed to see how the structure size is related to the variance of a forecast. This experiment used the same storm as experiment 5.1. Sheerness was used as the primary station, and Immingham was used as the secondary station. 24 hours of previous data was used from each station. The test year was 1999, the training year was 1997, and the validation year was 2001.

The average of the AAE for all 20 runs is shown for each model in Table 6-8. The reference forecast (column #2) is the intermediate results from a single 20 repetition [1,1] ensemble forecast. Values are for a 4-hour forecast, and sampled during a 72–Hour Storm window. The average AAE for 20 single [1,1] ANN models is 0.206 m. The average AAE for the [1,1] ensemble models is 0.204 m. The error decreases as the models become more complex, with a minimum average AAE of 0.201 m when using the [10,1] ensemble model. As the models become more complex, the average AAE increases, with the most complex model having an average AAE of 0.204 m.

The variance of the ensemble models increases as the ANN structure becomes more complex. The 20 single ANN models used for reference had a variance of 2.83×10^{-4} m^2. The [1,1] ensemble model had the lowest variance of 4.00×10^{-6} m^2 or 1.29% of the variance of the reference model. The variance increased as the models grew more complex with the [30,1] ensemble model having a variance of 3.40×10^{-5} m^2 or 11.8% of the reference model.

The results show that the use of the ensemble approach decreases considerably model variance for all ANN structures. The smallest variance is obtained for the smallest ANN structures [1,1] at 1.4% of the reference variance and increases with ANN structure size up to 11.8% for the [30,1] structure. The best performance as measured by the absolute average error is reached for the [10,1] structure at 0.201 m with the error increasing to 0.204 m for the largest and the smallest ANN structures. The best modelling strategy will be to select a rather small ANN [1,1] to [10,1], with the final selection using the more robust model with the lowest variance.

Table 6-8 Average absolute error (m) for 4-hour Ensemble forecasts changing the number of neurons used in the hidden layer. 72–hour storm window, November 5, 1999, Primary station: Sheerness, Secondary station: Immingham, Test Year 1999, Training Year 1997, Validation Year 2001. Note: All ensemble forecasts used 20 repetitions each.

Ensemble Run Number (1)	Reference Forecast (Single run) [1,1] (2)	Structure Size ◄Simple				Complex►
		Ensemble [1,1] (3)	Ensemble [5,1] (4)	Ensemble [10,1] (5)	Ensemble [15,1] (6)	Ensemble [30,1] (7)
1	0.245	0.206	0.207	0.204	0.203	0.205
2	0.198	0.202	0.205	0.206	0.199	0.199
3	0.228	0.202	0.206	0.204	0.202	0.199
4	0.207	0.204	0.199	0.198	0.205	0.200
5	0.204	0.201	0.202	0.199	0.190	0.196
6	0.192	0.205	0.203	0.201	0.199	0.197
7	0.194	0.203	0.201	0.197	0.205	0.210
8	0.198	0.205	0.205	0.202	0.198	0.207
9	0.198	0.203	0.207	0.200	0.201	0.203
10	0.245	0.206	0.205	0.199	0.203	0.208
11	0.198	0.202	0.206	0.203	0.201	0.215
12	0.228	0.203	0.199	0.203	0.204	0.210
13	0.207	0.204	0.202	0.205	0.200	0.198
14	0.205	0.201	0.203	0.196	0.210	0.212
15	0.192	0.206	0.201	0.198	0.202	0.198
16	0.195	0.203	0.205	0.205	0.201	0.205
17	0.199	0.206	0.202	0.201	0.202	0.206
18	0.198	0.203	0.204	0.198	0.204	0.204
19	0.193	0.203	0.202	0.200	0.202	0.214
20	0.193	0.208	0.206	0.207	0.201	0.203
Average	0.206	0.204	0.203	0.201	0.202	0.204
Variance (m^2)	0.000283	0.000004	0.000006	0.000010	0.000014	0.000034
% of Reference Forecast	100.00	1.29	2.14	3.65	5.11	11.84

6.6.3 Experiment 5.3: Effect of Ensemble ANN structure on accuracy of maximum surge and water levels predictions

Experiment 5.3 evaluates the error of the model at the instant of maximum surge height and maximum total water level during a storm. For people living along the coast, the peak or maximum water level reached during a storm is a very important concern. Also of interest to ocean scientists and engineers is the maximum storm surge elevation, which is very critical if it occurs near or during a high tide. These events are identified and compared separately in this experiment.

Experiment 5.3.1 is evaluates the model performance during the instant of maximum surge height during a storm. This experiment uses ensemble forecasts of the ANN models and varies the number of hidden neurons from 1 to 30. This experiment uses the same storm examined in experiments 5.1 and 5.2 and was selected because it was a typical November storm with a maximum surge height

of 1.24 m. Sheerness was used as the primary station, and Immingham was used as the secondary station. 24 hours of previous data was used from each station. The test year was 1999, the training year was 1997, and the validation year was 2001.

For this experiment, 20 [1,1] ANN models are run and the error of the model at the time of maximum surge is calculated. The results are shown in Table 6-9. The Reference Forecast (column #2) is the intermediate results from a single 20 repetition [1,1] ensemble forecast. These 20 individual or single run ANN models are used as a reference to measure the performance of the ensemble models. Next, five different size ensemble models are tested using [1,1], [5,1], [10,1], [15,1] and [30,1] sized ANN structures. For the five different ensemble models each used 20 repetitions for each run, and each ensemble model was run 20 times. Each ensemble model calculated a surge elevation at the time of maximum surge. The results for these models are shown in columns 3–7 in Table 6-9. The average value for each model type is shown at the bottom of Table 6-9. The average error decreases with the size of the models as the models become more complex, with a minimum error of 0.677 m when using the [30,1] ensemble model.

In general, the variance of the forecasted maximum surge elevation increases with the number of hidden neurons as the ANN structure becomes more complex. The 20 single ANN models used for reference had an average variance of 1.48×10^{-1} m^2. The [1,1] ensemble model had the lowest average variance of 2.51×10^{-4} m^2 or 1.69% of the variance of the reference model. Although variance grew as the model complexity was increased, there is an anomaly with the [30,1] structure, where the variance decreased.

Table 6-9 4-hour forecast Ensemble model error (m) during time of maximum surge, 13:00 UTC November 5, 1999. Primary station: Sheerness, Secondary station: Immingham, Test Year 1999, Training Year 1997, Validation Year 2001. Note: All ensemble forecasts use 20 repetitions each.

Ensemble Run Number	Reference Forecast Single run	Structure Size ◄Simple				Complex►
	[1,1]	[1,1]	[5,1]	[10,1]	[15,1]	[30,1]
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1.044	0.829	0.778	0.721	0.752	0.694
2	0.753	0.799	0.760	0.752	0.713	0.625
3	1.007	0.806	0.744	0.748	0.719	0.646
4	0.896	0.812	0.722	0.750	0.697	0.597
5	0.787	0.791	0.746	0.699	0.534	0.556
6	0.712	0.835	0.752	0.711	0.707	0.658
7	0.723	0.806	0.773	0.685	0.733	0.739
8	0.766	0.804	0.769	0.722	0.690	0.694
9	0.787	0.788	0.778	0.687	0.700	0.675
10	1.045	0.829	0.760	0.721	0.667	0.778
11	0.753	0.799	0.744	0.716	0.690	0.700
12	1.007	0.806	0.722	0.765	0.719	0.747
13	0.895	0.812	0.746	0.746	0.693	0.655
14	0.787	0.790	0.752	0.673	0.747	0.711
15	0.712	0.835	0.773	0.808	0.665	0.677
16	0.723	0.806	0.770	0.750	0.704	0.675
17	0.766	0.803	0.703	0.751	0.689	0.680
18	0.73	0.806	0.706	0.697	0.858	0.680
19	0.678	0.842	0.753	0.701	0.740	0.655
20	0.71	0.799	0.744	0.699	0.534	0.700
|Average| (m)	0.814	0.810	0.750	0.725	0.698	0.677
Variance(m^2)	0.014843	0.000251	0.000498	0.001100	0.004800	0.002516
% of Reference	100.00	1.69	3.35	7.41	32.34	16.95

It is apparent from these results that ANNs have difficulties predicting maximum surge elevations during large storm events. This poor performance during maximum surge is thought to occur because of the acceleration of the tide due to increased water depth due to the storm surge. Since the tide behaves as a coastally trapped long wave, its speed of propagation is determined by the water depth. The increase in water depth caused by the surge accelerates the tide signal, causing it to arrive sooner than anticipated which is interpreted as surge. Prediction of maximum surge is not as important as maximum total water elevation which is required by engineers and flood managers and is discussed next.

Experiment 5.3.2 determines the effect of ANN structure size on model performance during the instance of maximum total water elevation during a storm. The maximum total water elevation event occurs when the tidal and surge components combined reach the maximum elevation during a storm. The storm

event selected for this experiment was again November 5, 1999. Sheerness was used as the primary station, and Immingham was used as the secondary station. 24 hours of previous data was used from each station. The test year was 1999, the training year was 1997, and the validation year was 2001. All ensemble models use 20 repetitions.

For this experiment, 20 single run [1,1] ANN models were run and the error of the model at the time of maximum water level was calculated. The results are shown in column 2 of Table 6-10. The Reference Forecast (column #2) is the intermediate results from a single 20 repetition [1,1] forecasts. These 20 individual or single run ANN models are used as a reference to measure the performance of the ensemble models. Next, five different size ensemble models were tested using [1,1], [5,1], [10,1], [15,1] and [30,1] sized ANN structures. Each of the five different ensemble models used 20 repetitions for each run, and each ensemble model was run 20 times. Each ensemble model calculated a water level value at the time of maximum water level and its error calculated. The results for these models are shown in columns 3–7 in Table 6-10. The average error for each model type is shown at the bottom of Table 6-10. The size of the average error increases as the models become more complex, with a minimum average error of 0.008 m when using the [1,1] model, increasing to a maximum error of 0.102 m when using the [30,1] ensemble model.

Table 6-10 4-hour forecast Ensemble model error (m) during time of peak water elevation, 21:00 UTC November 5, 1999, Primary station: Sheerness, Secondary station: Immingham, Test Year 1999, Training Year 1997, Validation Year 2001. Note: All ensemble forecasts use 20 repetitions each.

Ensemble Run Number	Reference Forecast Single run	Structure Size ◄Simple				Complex►
	[1,1]	[1,1]	[5,1]	[10,1]	[15,1]	[30,1]
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.048	0.007	0.033	0.016	-0.025	-0.160
2	-0.015	0.006	-0.004	0.009	-0.027	-0.128
3	0.030	-0.002	-0.011	-0.022	-0.051	-0.120
4	0.005	-0.006	0.005	-0.015	0.027	-0.081
5	0.004	0.003	-0.005	-0.025	-0.017	-0.068
6	-0.025	0.009	0.033	0.034	-0.024	-0.086
7	-0.018	-0.012	0.024	-0.033	-0.022	-0.078
8	-0.014	0.016	0.002	-0.001	-0.056	-0.013
9	-0.013	-0.013	0.033	-0.025	-0.053	-0.189
10	0.048	0.007	-0.004	-0.033	-0.007	-0.141
11	-0.015	-0.006	-0.011	0.000	-0.114	-0.018
12	0.030	-0.002	0.005	-0.035	0.006	-0.082
13	0.005	-0.006	-0.005	-0.017	-0.016	-0.049
14	0.004	0.003	0.033	-0.061	-0.001	-0.108
15	-0.025	0.009	0.024	-0.004	-0.035	-0.134
16	-0.018	-0.012	0.002	-0.054	-0.046	-0.093
17	-0.013	0.016	0.012	0.009	0.003	-0.160
18	-0.025	-0.008	0.020	-0.018	-0.027	-0.123
19	-0.041	-0.006	0.004	-0.019	-0.084	-0.116
20	-0.019	-0.002	0.005	0.034	-0.053	-0.093
|Average| (m)	0.017	0.008	0.014	0.025	0.037	0.102
Variance(m^2)	0.000622	0.000079	0.000241	0.000649	0.001048	0.002068
% of Reference	100.00	12.72	38.71	104.45	168.57	332.57

The variance of 20 [1,1] ensemble models was significantly less then 20 individual [1,1] ANN models and was the smallest of all the ensemble forecasts. The 20 single ANN models used for reference had a variance of $6.22 \times 10^{-4} m^2$, while the [1,1] ensemble model had a variance of 7.90×10^{-5} m^2 or 12.7% of the variance of the reference model. The variance grew rapidly as the complexity of the models increased to 2.068×10^{-3} m^2 or 332% of the variance of the reference model.

It is apparent from the results shown in Table 6-10 that ensemble forecasting methods predict peak water elevations very well. The simple [1,1] ensemble model performed best with a average error of 0.008 m. As the models became more complex, the average error increased from 0.014 m for the [5,1] model to 0.102 m for the [30,1].

6.6.4 Experiment 5.4: Ensemble forecasts, effect of overtraining on forecast variance.

To determine the effect of over-training when using ensemble averaging, two sets of networks were set up. The first used the same training method and set-up used in Experiment 5.3. This set-up uses the validation technique for early stopping used in experiment suite #3 to prevent over training. Networks trained with this method rarely exceed 15 epochs before stopping. The second ensemble ANN is set up to stop training only after 3000 epochs. The purpose of the over-training is to reduce the bias portion of the error; the variance portion of the error is reduced later by the ensemble averaging. The results of the experiment are shown in Table 6-11. It is immediately obvious that the error levels are reduced by using the overtraining technique. The model using early stopping method performed the worst with an AAE of 0.203 m compared to the model using no early stopping method and 20 repetitions performance of 0.181 m.

This experiment was only used to test the performance of the method, and not to find an optimal solution at a particular station. Many more models would need to be tested to determine which ensemble modelling method/set-up actually has the best performance. It is important to note that the method of using over-training to reduce the variance while averaging ensembles was found after all the experiments for this project were completed. It is obvious now that this technique is quite useful in reducing modelling errors when using ensemble averaging and will be incorporated in future experiments.

Table 6-11 Effect of over-training in ensemble forecasting. 4-hour forecast, 72–hour storm window, November 5, 1999, Primary station: Sheerness, Secondary station: Immingham, Test Year 1999, Training Year 1997, Validation Year 2001 (when used).

Method of early stopping (1)	Number of repetitions (2)	AAE (m) (3)	RMSE (m) (4)	CF (%) (5)	Error during maximum surge (m) (6)	Error during maximum water Elevation (m) (7)
Validation	20	0.203	0.282	47.9	0.721	0.015
None	20	0.181	0.259	57.5	0.578	0.004
None	40	0.181	0.263	57.5	0.610	0.048
None	100	0.190	0.277	57.5	0.631	0.072

6.7 Experiment Suite #6: Engineering Application

This experiment utilizes the results from the previous experiments to apply an ANN model at a new location. Silvertown was selected for the location because of its close proximity to the Thames Barrier (the Silvertown tide gauge is located only

500m downstream from the Thames Barrier). Models at this station have to cope with the changing stage height elevation of the estuary due to fresh water flow, hydrodynamic effects of the Thames Barrier, and the constricting effects of the Thames as it changes from wide estuary to a narrow estuary. The mean spring tidal range changes from 5.27 m at Sheerness, to 6.50 m at Silvertown, an increase of 1.23 m. The mean neap tidal range changes from 3.28 m at Sheerness, to 4.40 m at Silvertown, an increase of 1.12 m.

This experiment will examine the accuracy of an ANN forecast at this new location without taking into account the barrier closures, and river stage height. This experiment was performed at this location to see how a simple ANN using water level data only would cope with a very complex hydrodynamic location. Because data availability for Silvertown was limited to the years 2000 – 2004, this prohibited using the same storm events which had been focused on during the previous experiments.

For this experiment, the validation, training, and testing data sets are selected arbitrarily as Experiment #1.1 showed no model dependence on such selection. The validation data set uses year 2000 data, the training data set uses year 2001 data and the testing data set uses year 2002 data. 24 hours of previous data is used for the primary and secondary stations based on results from Experiment #1.2. The optimum secondary station location varies depending on the forecast interval, based on results from Experiment #2.3. A complex [10,1] 20-repetition ANN ensemble model is used, based on results from Experiment #3.1 to reduce forecast variance, and for better prediction of peak storm surge elevation. The models are tested using both the early-stopping and over-training methods.

A cross-correlation analysis was run to estimate the additional tidal and surge lag periods for changing water levels travelling from Sheerness to Silvertown. These lag periods were compared to the results from experiment 2.3. The 56 km of additional distance up the Thames Estuary from Sheerness to Silvertown increased the tidal lag period ~2 hours and the surge lag period ~1 hour. The results for the Silvertown cross-correlation tests are shown in Table 6-12.

Table 6-12 Cross-correlation analysis test for Silvertown station (Years 2000 – 2002).

Station Name (1)	Tidal Lag Period(hrs) (2)	Surge Lag Period(hrs) (3)
Sheerness	2	1
Immingham	11	7
North Shields	12	10
Wick	17	16

3, 6, 12, and 12 hour forecasts were calculated using information obtained from experiment 2.3 to select the optimal secondary station locations. It was determined that the additional 1-hour required for the surge to travel up the Thames Estuary did not alter the optimal location for the secondary stations when restricting the selection to Immingham, North Shields, or Wick.

The location was modelled during two different storm events. The first selected is February 20, 2002, an event chosen because the Thames Barrier was open. The second event selected is April 27, 2002, an event chosen because the Thames Barrier was closed and presented only to see how the model would perform. These storm events are shown in Table 6-13, where maximum water level and maximum surge-level are also indicated. During the February 20, 2002 (2 days after neap tide minimum) storm event, a 1.184 m increase in maximum water level occurred during the surge's 56 km travel up the Thames Estuary between Sheerness and Silvertown (close to normal for a neap tide), and a 1.781 m water level increase occurred during the April 27, 2002 (1 day before the spring tide maximum) storm event. The water level increases can be further broken down into a normal tidal increase, and an increase in the surge elevation. For the barrier open case the surge increased 0.52 m while for the closed barrier case the surge increased 0.01 m.

Table 6-13 Comparison of storm surge statistics for Sheerness/Silvertown.

Storm #	Date	Barrier open/closed	Sheerness		Silvertown		Δ Max water level (difference between stations)	Δ Max surge-Level (difference between stations)
			Max water level (m)	Max Surge-level (m)	Max water level (m)	Max Surge-level (m)		
1	February 20, 2002	Open	5.203	1.623	6.387	2.141	+1.18	+0.52
2	April 27, 2002	Closed	6.238	0.855	8.019	0.867	+1.78	+0.01

The model performances for an event when the Thames Barrier was open are shown in Table 6-14. 24 previous water level residuals are used from each station

for each forecast. During this event the maximum water elevation was relatively low (6.39 m), but the maximum surge level was 2.14 m. During testing, it was soon evident that the model using validation as an early stopping method to prevent overtraining performed much better than the model using no early stopping method. For the 3-hour forecasts the validation method had an average absolute error of 0.36 m compared to the model using no early stopping performance of 0.40 m. For the 24-hour forecast interval, the validation method had an average absolute error of 0.57 m compared to 1.43 m for the model using no early stopping.

It was found that the over-training technique for reducing bias in ensemble forecasts used in Experiment 5.4 performed very poorly at the Silvertown location during large surge events. The model performed reasonably well for the short term 3 – 6 hour forecasts, but the overtraining caused extreme and erratic model behaviour during the large storm events, especially during the 12 and 24 hour forecasts. During the longer term forecasts, the model would perform normally during the period up to and including the surge event peak, but then would exhibit large erratic behaviour for several hours afterward before returning to normal.

Table 6-14 Barrier Open - Silvertown/Immingham using a 20 repetition ensemble [10,1] ANN. IMM = Immingham, NSH = North Shields, WIC = Wick. Storm #1 date: February 20, 2002 (Hrs: 1188-1260)

Early Stopping Method Used	Forecast Interval	Station Name	AAE (m)	RMSE (m)	CF(15cm) %	Error during maximum surge level (m)	Error during maximum water level(m)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Validation	3	IMM	0.356	0.433	24.66	0.986	0.054
	6	IMM	0.410	0.509	21.92	1.323	0.104
	12	NSH	0.529	0.684	13.70	1.980	0.467
	24	WIC	0.586	0.772	12.33	1.928	0.429
None	3	IMM	0.395	0.543	26.03	0.982	0.072
3000	6	IMM	0.615	1.185	28.77	0.949	0.141
Epochs	12	NSH	0.805	1.332	17.80	1.873	0.205
	24	WIC	1.432	3.424	12.33	1.923	0.391

For comparison purposes, the model results for Sheerness are presented in Table 6-15 before the surge's 56 km travel up the River Thames. 24-hours of previous water level residuals are used from each station for each forecast. Analysis was performed using a 72-hour storm window. For this experiment, data from year 2002 was used for testing, year 2001 for training, and year 2000 was used for verification. The model at Silvertown performed similarly to that at Sheerness. The most notable difference between the two locations was the

performance of the model when predicting the maximum surge levels. For a 12-hour forecast, the model error during maximum surge increased from 1.38 m at Sheerness to 1.98 m at Silvertown. Since the barrier was open during this test, the difference between the locations can probably be attributed to acceleration of the tide due to the increased water depth from the surge.

Table 6-15 Comparison – Sheerness/Immingham, using a 20 repetition ensemble [10,1] ANN. IMM = Immingham, NSH = North Shields, WIC = Wick. Storm #1 date: February 20, 2002 (Hrs 1188-1260)

Early Stopping Method Used	Forecast Interval	Station Name	AAE (m)	RMSE (m)	CF(15cm)%	Error during maximum surge level (m)	Error during maximum water level(m)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Validation	3	IMM	0.286	0.349	30.1	0.754	0.128
	6	IMM	0.434	0.515	19.2	1.159	0.260
	12	NSH	0.530	0.636	11.0	1.380	0.230
	24	WIC	0.499	0.622	19.0	1.640	0.431

The results of the model performance for when the Thames Barrier was closed are shown in Table 6-16. 24-hours of previous water level residuals are used from each station for each forecast. Analysis was performed using a 72-hour storm window. For this experiment, data from year 2002 was used for testing, year 2001 for training, and year 2000 was used for verification. Although the maximum water levels were very high during this test (8.019 m), the maximum surge level was only 0.867 m. During this event, the 3-hour forecasts from both models had an average absolute error of 0.210 m. For the longer 24-hour forecasts the model using validation clearly performed better with an average absolute error of 0.290 m compared to 0.328 m for the model using over-training.

Table 6-16 Barrier Closed - Silvertown/Immingham using a 20 repetition ensemble [10,1] ANN. IMM = Immingham, NSH = North Shields, WIC = Wick. Storm #2 date: April 27, 2002 (Hrs: 2785-2857)

Early Stopping Method Used	Forecast Interval	Secondary Station Name	AAE (m)	RMSE (m)	CF(15cm)%	Error during maximum surge level (m)	Error During m maximum water level (m)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Validation	3	IMM	0.210	0.298	53.4	0.556	0.556
	6	IMM	0.219	0.287	43.8	0.566	0.566
	12	NSH	0.244	0.316	42.5	0.384	0.384
	24	WIC	0.290	0.385	38.4	0.862	0.862
Overtraining	3	IMM	0.210	0.304	52.43	0.530	0.530
	6	IMM	0.233	0.327	50.68	0.573	0.573
	12	NSH	0.295	0.418	36.99	0.393	0.393
	24	WIC	0.328	0.420	30.14	0.832	0.832

It is interesting to note that the ANN model performed better during storm #2 with a closed barrier than it did during storm #1 when the barrier was open. This is due to the fact that the size of the surges associated with the two storms was more a factor than the hydrodynamic effects caused by the closed barrier. Also it is interesting to note that the maximum surge level and maximum water elevation occurred at the same time. This is because when the barrier is closed, the location becomes a terminus or ending station for all water moving up the Thames, causing the surge to remain until the tide withdraws.

In conclusion, the results from the previous five experiments have allowed the construction of an acceptable working model at this new location much more quickly than if the model was developed from scratch. However, the new model is far from optimal. Much more analysis of the model's performance using different storms is needed with the barrier in open and closed positions. A proper ANN model would need much more historical data than was available at the time of this project.

7 Discussion

Researchers working with the Texas Ocean Observation Network have often commented on the potential benefit that would be gained from additional offshore tide gauge stations located in the path of approaching surges from storms and hurricanes. In the North Sea, the tide gauges are aligned such that the storm surge propagation follows the gauges in a path along the coastline. Due to the unique configuration of these gauges, the researcher selected this area to model storm surge movement. Given the non-linear relationships between forcings and surges, using ANN's seemed a promising methodology to predict surge movement and growth.

One of the main goals of the research was to show that artificial neural networks can be used to accurately predict storm surge elevations in the North Sea. This research has successfully demonstrated that a simple ANN model can accurately predict storm surge elevations at the Sheerness tide station. The research has further demonstrated that including secondary station information as additional input to the model significantly improves ANN forecast accuracy. The ANN model was to use minimal data resources, and be computationally efficient. The most accurate model currently available (CS3) requires input from very large and computationally expensive meteorological models. This research uses only historical sea level data and does not require input from any other data source. The current ANN model is able to compete with the more computationally expensive numerical models for 3-hour or shorter forecast intervals. Several different model configurations were developed for this research each systematically increasing in complexity. The first model was a single station ANN model used primarily to test possible differences between training data sets. Model complexity was then increased to include a secondary station, and larger ANN structures. Finally ensemble models were developed to reduce variance, and overtraining was used to minimize bias and overall forecast errors. The results will now be discussed in the context of each experiment.

Experiment Suite #1 tested the artificial neural network models performance while varying the amount and selection of data used for training.

Experiment 1.1 tested the performance of the ANN while varying the size and selection of the training data set. It was concluded that although increasing

the size of the training data set does improve ANN performance, it is insignificant in proportion to the substantial increase in training time. Hence one year data sets can be used to train an ANN model without a significant reduction in performance.

Experiment 1.2 found that the performance of the ANN does not change significantly with the selection of a specific or individual training year when analyzed on a yearly basis. Note that the influence of the training year on performance during individual storms was not tested, and is reserved for future work.

Experiment Suite #2 tested the performance of an ANN using different configurations for the inputs. These inputs consisted of different numbers of historical or previous water level residuals measured either at a single tide gauge location, or at two tide gauge stations, utilizing the secondary tide gauge station to provide additional information to the model.

Experiment 2.1 used a single or primary tide gauge station only. The number of previous water level residuals used for each forecast was varied from 1 to 48 hours. The results indicated that although including more information from previous water level residuals increases performance, little benefit is found when including more than 24-hours. For a 3-hour forecast the AAE decreases from 0.18 m to 0.12 m when increasing the number of previous water level residuals used from 1 to 24 hours. Increasing the number of previous water level residuals used to 48 hours decreases the AAE to 0.11 m but has the negative effect of increasing training times by a factor of two. The conclusion is that very little useful information is contained in this "older data" and it contributes little to short-term forecasts. Because of this, later models included only 24-hours of previous water level residual data, significantly reducing training times, without a significant sacrifice in performance.

Experiment 2.2 tested the performance of an ANN using two tide gauge stations. The results indicated that this significantly increased the ANN performance. The addition of a second tide gauge station gives the ANN important information about the approaching surge. This information allows the ANN to correlate peak surge and water levels at the secondary station to future surge and water levels at the primary station. This additional information from the secondary station reduced the average absolute error of the 3-hour forecast found in experiment 2.1 from 0.12 m to 0.08 m.

Experiment 2.3 found that for a secondary station to be effective in predicting a storm surge event, it must provide information of the storm surge existence at the time of the forecast. The optimal secondary station location is one where its distance to the primary station is similar to the distance a storm would travel during the forecast interval. For the 3 and 6-hour forecasts, the optimum secondary site location was Immingham with a minimum AAE of 0.08 m and 0.09 m respectively. Immingham is located 337 km north of sheerness resulting in a typical surge travel time of 6 hours; the other stations, North Shields and Wick were located further north resulting in longer surge travel time of 6 to 14 hours. For the 12-hour forecast North Shields was the best secondary station location with a minimum AAE of 0.13 m. For a 24-hour forecast, Wick performed best as the secondary station location with a minimum AAE of 0.16 m. If the selected station is located too close to the primary station for the forecast interval, it will not include any new significant information from the approaching surge, and the information it does supply will be very similar to that of the primary station, adding little knowledge to the ANN to increase performance. If the selected station is located too far away from the primary station for the forecast interval, the information it contributes applies best only to longer forecasts, and has little correlation on what happens at the primary station in the short term.

Experiment Suite #3 investigated how model performance is affected by changes in ANN topology. The topology of the ANN was changed by varying the number of hidden neurons used in the model. It was hypothesized that increasing the size of the ANN structure would allow it to predict a more complicated time series. Note that increasing the size of an ANN structure can substantially increase training times.

Experiment 3.1 found that more complex models have the potential to better capture large surges. Unfortunately as the number of hidden neurons is increased, the variability of the ANN model also increases.

Experiment 3.2 tested the variability of individual ANN model runs. It was found that variability increases as the size of the ANN structure increases. Although the average performance decreased as the structure size is increased, the best overall performance was always obtained by an individual instance of a large ANN structure. The potential for large structure ANNs to better model surge

events is further explored through the use of ensemble models in Experiment suite #5.

Experiment Suite #4 compared the performance of a simple ANN model with the CS3 model used by the Storm Tide Forecasting Service (STFS). The archived STFS data format limited comparisons to 1 – 5 hour forecast intervals. Since the CS3 model forecasts were observed at location with a tide gauge and included a T+00 (or NOW) observation, a bias could be calculated and removed to improve the CS3 model performance.

Experiment 4.1 showed that when analyzing the "yearly performance" of the models the ANN always performed better than the CS3 model. Although the removal of the bias significantly improved the CS3 model performance on 1-hour forecasts, this was not enough to beat the performance of the ANN model. It is interesting to note that removal of the bias actually degraded yearly performances of the longer 5-hour forecasts, and that the bias correction at T+00 is only helpful for short term predictions.

Experiment 4.2 analyzed the ANN/CS3 performances using a 72-hour storm window. For 1-hour forecasts the results were mixed, with the ANN model performing better than the CS3 model for 1999 but not in 2002 where the CS3 model with the bias removed performed best. For the 3-hour forecasts all the models (ANN, CS3, and CS3 with bias removed) had virtually identical performances, with differences of 0.02 m or less. For the 5-hour forecasts the CS3 model performed best. Perhaps one of the reasons that the CS3 performed better was that meteorological forcing may become more important with the increasing size of the surge event (which is included in the CS3 model's forcing parameters). It was also noted that the CS3 model was very consistent, with error values very similar, independent of the forecast interval, while at the same time the ANN error values always increased with an increase in the forecast interval. Future work that includes additional information such as tides and/or real time meteorological data should lead to improvements in ANN model performance for longer forecast intervals.

Experiment Suite #5 was designed to assess the potential of ensemble forecasting methods used to reduce forecasting errors. The model assessment in this area was concentrated on the storm events only. The first part of the

experiment was concerned with changing the number of repetitions used for each ensemble. The experiment also tested the effect of changing the number of hidden neurons on performance. Ensemble forecasting techniques were used to improve the performance of ANN models during maximum water level and surge-level events.

Experiment 5.1 tested the model performance for several ANN structure sizes while also changing the number of repetitions for each ensemble. The results showed that ensemble ANN models have significantly less variance than single ANN models, with the variance decreasing as the number of repetitions per ensemble is increased.

Experiment 5.2 tested the accuracy of an ANN ensemble as the size of its structure or number of hidden neurons is increased. The results showed that increasing the size of an Ensemble structure had no significant effect on accuracy. Variance was found to be reduced to 1.3% of the original single ANN model [1,1] ensemble, and then increased in proportion to the number of hidden neurons used.

Experiment 5.3 tested the effect of ANN structure size on prediction accuracy for both peak surges and peak water levels. The results showed that the more complex models performed best when predicting peak surge levels while the simpler models performed best when predicting peak water levels. The results also showed that ANN models perform poorly in predicting peak surge-level elevations, but very well in predicting peak water level elevations. It is thought that one of the reasons the ANN performed so well in predicting peak water level elevations is that for the relatively short 4-hour forecasts, the model has seen both the peak surge at the secondary and primary stations. The knowledge that the peak surge has already passed both stations and the knowledge of the peak surge elevation increase between secondary and primary station likely provides the ANN with the additional information necessary to make very accurate peak water level predictions.

Experiment 5.4 tested the use of overtraining to reduce the bias portion of the error of individual ANN models, combined with the use of ensemble techniques to minimize the variance of the error. This combination of techniques produced the lowest individual forecast errors when forecasting water or surge peaks during this research. In previous ANN models early stopping was used to prevent over-training of the models. While over-training is generally viewed as a condition that is

to be avoided when using a single ANN model, (because it increases variance), it has been found to be beneficial when used with ensemble averaging (where the variance is removed). During the period of maximum surge, the average absolute error for a 20-repetition ANN ensemble was decreased from 0.721 m for a model using validation compared to 0.578 m using overtraining. During the period of maximum water elevation, the average absolute error was decreased from 0.015 m for a single 20-repetition ensemble model using validation to 0.004 m for a 20-repetition ensemble model using overtraining. This level of improvement was unexpected, and shows great potential and need for further testing.

Experiment Suite #6 was designed to assess the performance of an ANN model at a new location using only the techniques used to those developed in experiments 1 – 5. Silvertown was selected as the new location for this experiment because of its location 500 m downstream from the Thames Barrier. By virtue of this location, accurate surge level forecasts would be challenging at Silvertown. When the barrier is closed during low tide, incoming waters that normally flow past the barrier have nowhere to go, causing higher tide levels than with the barrier open. Conversely when the barrier is closed during a high tide, water levels down-stream from the barrier are lower than they would have been if the barrier remained open. This research trains the models with the assumption that the barrier is closed only for a very small percentage of the time during the year.

Cross-correlation analysis at Silvertown found that lag periods for surges and tides were only increased by 1-2 hours compared to the lags for the Sheerness tide station. The results show that the optimal secondary station locations used for the various forecast intervals at Sheerness were indeed directly applicable to the new location.

The model was tested during open and closed barrier conditions. The results showed that when using the methods refined in experiments 1-5, accurate surge elevation predictions at Silvertown are difficult, whether the barrier is open or closed. The results show that the ANN model under-predicts maximum surge levels when the Thames Barrier is closed, with errors ranging from 0.530 m for a 3-hour forecast, to 0.832 m for a 24-hour forecast. This is expected because incoming storm surges would normally pass through the barriers, but now have nowhere to go, thus causing higher surge levels than expected. However, results

show that even when the Thames Barrier is open, the model also significantly under predicts the maximum surge elevation and surge component at the time of maximum water level. This under-prediction was also found to a lesser extent for the other cases treated in this research. The under-prediction is attributed to the acceleration of the tide caused by the increased water depth resulting from the surge that occurs during its 56 km journey up the Thames Estuary from Sheerness. This additional water depth accelerates the tidal signal, leading to its early arrival and early peak of the surge. This feature is not currently captured by the ANN model, possibly because tides and meteorological information are not included in training.

Although the ensemble method that utilized over-training to reduce bias was very promising in experiment 5.4, no significant performance improvements were observed at the Silvertown location. The large storm events that were selected to test the model caused it to behave very erratically after experiencing large surge events. This behaviour lasts for several hours after the peak surge event and adversely affects the model performance as compared to validation trained models. It is thought that a selective procedure that removes significantly different ensemble member prediction(s) from the computation of the forecast could be used to control this behaviour in the future.

Future work

Most of the ANN models used in this research "under predicted" the maximum storm surge elevations. The primary cause for this prediction error is thought to be the acceleration of the tide caused by the increase in water depth in the region of the storm. It has been shown that the amount of surge generated by a specific storm event is modulated by the current state or elevation of the tide (Horsburgh and Wilson 2007). Artificial neural networks will not be able to predict a variable if the forcings that are driving this variable are not included in the training data set. By including the tidal elevation as an additional input, the artificial neural network model may learn to better predict maximum surge heights.

Other future work could include the use of additional or tertiary tide stations. It is hypothesized that training an artificial neural network using three or more stations may reduce forecast errors by detecting a change in the rate of surge growth during its propagation. This change in growth rate could be caused by

strengthening of a storm system, and be better modelled using additional information from a third tide gauge station.

The inclusion of real time wind and barometric data in the training and operation of the ANN models could also significantly increase the performance of the models by providing the latest storm information without a substantial decrease in computational efficiency. Atmospheric predictions could also be added, but with a negative effect on the overall computational efficiency of the method.

Future work should compare ANN / CS3 performance for longer forecast intervals. Currently CS3 model data is archived to only T+5 hours. Future comparisons should also include peak surge and water elevations. Finally future work should test the significance of training year selection on ANN storm performance. It is thought that training with a dataset that includes several significant storms could increase ANN performance in forecasting such events.

8 Conclusions

A model based on artificial neural networks was developed to predict storm surge magnitudes and arrival times at selected locations in the North Sea. This research was first to use artificial neural networks to predict surge elevations in the UK. The research explored the significance of the selection of a training year and the size of the training data set has on model performance when predicting surge elevations. No significant difference in performance was found when changing the training year when testing on a yearly basis. Also no significant benefit was found when increasing the size of the training data set to include more than a single year.

This research pioneered the use of multiple secondary station locations to predict surge elevation for different forecast intervals. It was found that an artificial neural network's performance for prediction of storm surge was greatly increased by the use of a secondary station that has already measured the effects of the approaching storm. The optimal location of the secondary station is found where the distance from the primary station is the same as the distance the storm will travel during the forecast interval.

This project was the first to use Ensemble forecasting for prediction of storm surge elevations. Several types of ensemble model configurations were tested. First, the number of repetitions used for each ensemble was varied. It was found that the ensemble model variance was directly proportional to the inverse of the number of repetitions used. Next, the number of hidden neurons was changed, and it was found that performance accuracy was increased, but at the expense of an increase in variance. Finally, the bias of the ANN forecast error was reduced by using an over-training technique. The overtraining led to an increase in variance which was subsequently removed by ensemble averaging. The results showed a significant increase in accuracy, particularly when predicting maximum water levels during storm events.

The evaluation method used in this research uses a more critical method for measuring model performance than what is typically used in the literature. This method assesses the performance of the models using a 72-hour window centred on individual storm events rather than over a full year. This method was used because it was felt that previous assessments did not focus enough on the time when model performance is most critical, during storms.

The current artificial neural network model developed for this research is able to achieve similar level of accuracy as the more computationally expensive numerical model (CS3) at forecast intervals up to 3-hours. The accuracies were evaluated based on yearly datasets, and using 72-hour storm windows. These accuracies varied by year and for each storm system tested. Yearly forecasts at Sheerness had an average absolute error which varied from 0.08 m for a 3-hour forecast to 0.17 m for a 24-hour forecast. The 72-hour storm window forecasts had an average absolute error which varied from 0.08 m for a 1-hour forecast, to 0.31 m for a 5-hour forecast. The ensemble model accuracies were measured at the instance of peak surge and peak water elevation. The best 4-hour forecast of peak surge elevation was 0.677 m with a [30,1] model using 20 repetitions during a 1.67 m surge on November 5th, 1999. This poor performance was likely related to acceleration of the tide due to increased water depth caused by storm surge, and was not learned by the ANN as the tide level itself was not included in the training. One of the most significant results of this research was the ANN's ability to accurately predict peak water elevations. A simple [1,1] ensemble model using 20 repetitions had an average 4-hour forecast error of 0.008 m. When over-training is included to reduce the model bias, the error is further reduced to 0.004 m.

References

As-Salek, J.A.,and Yasuda, T. (2001), "Tide–Surge Interaction in the Meghna Estuary: Most Severe Conditions", *Journal of Physical Oceanography,* 31(10), 3059–3072

Barbounis, T. G., and Theocharis, J. B. (2006). "Locally recurrent neural networks for long-term wind speed and power prediction." *Neurocomputing*, 69(4-6), 466-496

Bargagli, A., Carillo, A., Pisacane, G., Ruti, P. M., Struglia, M. V., and Tartaglione, N. (2002). "An Integrated Forecast System over the Mediterranean Basin: Extreme Surge Prediction in the Northern Adriatic Sea." *Monthly Weather Review*, 130(5), 1317-1332.

Bazartseren, B., Hildebrandt, G., and Holz, K. P. (2003). "Short-term water level prediction using neural networks and neuro-fuzzy approach." *Neurocomputing*, 55(3-4), 439-450.

Blain, C. A. (1997). "Modeling methodologies for the prediction of hurricane storm surge." *Recent Advances in Marine Science Technology '96*, 177-189.

Blumberg, A. F., and G. L. Mellor. (1987). "A description of a three-dimensional coastal ocean circulation model" *Coastal Estuar. Sci.*, Hamburg, 1-16.

BODC. (2004). "U.K. National Tide Guage Network." British Oceanographic Data Centre, http://www.bodc.ac.uk/.

Box, G., Jenkins, G. (1970). *Time series analysis: Forecasting and control*, Holden-Day, San Francisco.

Brion, G. M., and Lingireddy, S. (1999). "A neural network approach to identifying non-point sources of microbial contamination." *Water Research*, 33(14), 3099-3106.

Chakraborty, K., Mehrota, K., Mohan, C. K., and Ranka, S. (1992). "Forecasting the Behavior of Multivariate Time Series Using Neural Networks." *Neural Networks*, 5, 961-970.

Cox, D. T., Tissot, P., and Michaud, P. (2002). "Water Level Observations and Short-Term Predictions Including Meteorological Events for Entrance of Galveston Bay, Texas." *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 128(1), 21-29.

Darwin, G. H. (1883). "Reports of the committee for harmonic analysis of tides." British Association for the Advancement of Science, London.

Deo, M. C., Jha, A., Chaphekar, A. S., and Ravikant, K. (2001). "Neural networks for wave forecasting." *Ocean Engineering*, 28(7), 889-898.

Deo, M. C., and Sridhar Naidu, C. (1999). "Real time wave forecasting using neural networks." *Ocean Engineering*, 26(3), 191-203.

Dixon, M. J. and Tawn, J. A. (1994). Estimates of Extreme Sea Conditions: Extreme Sea-Levels atthe UK A-Class Sites: Site-By-Site Analyses, *Proudman Oceanographic Laboratory*, Internal Document No. 65.

Doodson, A. T. (1921). "The harmonic development of the tide-generating potential." *Proceedings of the Royal Society of London* 100, 305-329.

Flather, R. A., Proctor, R. (1983). "Prediction of the North Sea Storm Surges Using Numerical Models: Recent Developments in the U.K." North Sea Dynamics, W. L. J. Sunderman, ed., Spring-Verlag, Berlin, 299-317.

Flather, R., Smith, J, Richards, J., Bell, C., Blackman, D. (1998), "Direct estimates of extreme surge elevations from a 40 year numerical model simulation and from observations.", *Global Atmos Ocean Sys,* 6: 165-176

Flather, R. A., Williams, J. A. (2000). "Climate change effects on storm surges: methodologies and results.", Climate scenarios for water-related and coastal impact (ed. J. Beersma, M. Agnew,D. Viner & M. Hulme). ECLAT-2 Workshop, Report No. 3, pp. 66–78., The Netherlands: KNMI.

Flather, R. A., Williams, J. A. (2004). "Future Development of Operational Storm Surge and Sea Level Prediction.", *Proudman Oceanographic Laboratory, Internal Document* No. 165., 70 pp

Gica, E., Teng, M.H. (2002). "Numerical simulation of storm surge generated by Hurricane Iwa in Hawaii." *ASCE Engineering Mechanics Conference June 2-5, 2002*, Columbia University, New York, NY.

Gorban, A., Rossiev, A. (2002) "Recovering data gaps through neural network methods.", *International Journal of Geomagnetism and Aeronomy., 7(2), 191-197*

Haigh, I. (2004). "An improved understanding of surge propagation through the Solent.", Volumes 1 and 2. ABPmer Report R1074. Environment Agency, West Sussex

Harris, R. A. (1897). "'Manual of Tides', Appendices to Reports of the United States Coast and Geodetic Survey.

Hecht-Nielsen, R. (1989). *Neurocomputing*, Addison-Wesley Longman Publishing Co., Inc.

Hagan, M., Demuth, H., and Beale, M. (1996). *Neural Network Design*, PWS Publishing Company, Boston MA, USA.

Horsburgh K.J., and Wilson C. (2007), "Tide-surge interaction and its role in the distribution of surge residuals in the North Sea.", *Journal of Geophysical Research,* Accepted for publication.

Huang, W., Murray, C., Kraus, N., and Rosati, J. (2003). "Development of a regional neural network for coastal water level predictions." *Ocean Engineering*, 30(17), 2275-2295.

IPCC (2007): Climate Change 2007: The Physical Scientific Basis. Summary for policymakers. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. http://www.ipcc.ch/SPM2feb07.pdf

Ishiguro, S. (1976), "Highest Surge in the North Sea.", *Institute of Oceanographic Sciences*, Report No 36, 31pp.

Jelesnianski, C., Chen, J., Shaffer, W., Gilad, A. (1984). "SLOSH - A Hurricane Storm Surge Forecast Model." *OCEANS*, 16, 314-317.

Kim, G., and Barros, A. P. (2001). "Quantitative flood forecasting using multisensor data and neural networks." *Journal of Hydrology*, 246(1-4), 45-62.

Lee, T. L. (2004). "Back-propagation neural network for long-term tidal predictions." *Ocean Engineering*, 31(2), 225-238.

Leuliette, E. W, Nerem, R.S., and Mitchum, G.T. (2004), "Calibration of TOPEX/Poseidon and Jason altimeter data to construct a continuous record of mean sea level change.", *Marine Geodesy*, 27(1-2), 79-94.

Luettich, R. A., Westerink, J. J., and Scheffner, N. W. (1992). "ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves Coasts and Estuaries, Report 1: Theory and Methodology of ADCIRC-2DDI and ADCIRC-3DL." U.S. Army Engineers Waterways Experiment Station, Vicksburg, MS, USA.

Luettich, R., Hudgins, J., and Goodall, O. (1996). "Initial results from a combined tide and storm surge forecast model of the U.S. East Coast, Gulf of Mexico and Caribbean Sea." *15th Conference on Weather Analysis and Forecasting, American Meteorological Society*, Norfolk, VA, USA, 547-550.

Liu, h., Xie, L., Pietrafesa, L., Peng, M. (2006). "Effect of ocean surface waves on storm surge and coastal flooding." *American Meteorological Society's 27th Conference on Hurricanes and Tropical Meteorology* Monterey, CA., USA.

Makarynskyy, O. (2004). "Improving wave predictions with artificial neural networks." *Ocean Engineering*, 31(5-6), 709-724.

Makarynskyy, O., Makarynska, D., Kuhn, M., and Featherstone, W. E. (2004). "Predicting sea level variations with artificial neural networks at Hillarys Boat Harbour, Western Australia." *Estuarine, Coastal and Shelf Science*, 61(2), 351-360.

MathWorks, Inc. (2002). "MATLAB Neural Network Toolbox." Natick, MA, USA.

McCulloch, W. S., Pitts, W. (1943). "A Logical Calculus of Ideas Immanent in Nervous Activity." *Bull. Mathematical Bio-phyisics*, 5, 115-133.

Meteorological Office, U. K. (2003). "Flood of memories." http://www.metoffice.com/bookshelf/outlook/200301/53flood.html.

Meteorological Office, U. K. (2006). "The Unified Model System." http://www.metoffice.com/research/nwp/numerical/unified_model/unified_model.html.

Moatar, F., Fessant, F., and Poirel, A. (1999). "pH modeling by neural networks. Application of control and validation data series in the Middle Loire River." *Ecological Modeling*, 120(2-3), 141-156.

More, A., and Deo, M. C. (2003). "Forecasting wind with neural networks." *Marine Structures*, 16(1), 35-49.

Mostella, A., Duff, J. S., and Michaud, P. R. (2002). "Harman and Harmpred: Web-based Software to Generate Tidal Constituents and Tidal Forecasts for the Texas Coast." *Proc. of 19th AMS Conf. on Weather Analysis and Forecasting/15th AMS Conf. on Numerical Weather Prediction,* San Antonio, Texas, USA.

Naftaly, U., Intrator, N., and Horn, D. (1997). "Optimal ensemble averaging of neural networks." *Network: Comput. Neural Syst.*, 8, 283-296.

Nerem, R. S. (1999). "Measuring very low frequency sea level variations using satellite altimeter data." *Global and Planetary Change 20(* 20), 157-171.

NOAA. (1999). "NOS Procedures for Developing and Implementing Operational Nowcast and Forecast Systems for PORTS." National Oceanic and Atmospheric Administration (NOAA), Center for Operational Oceanographic Products and Services, National Ocean Service (NOS), U.S. Department of Commerce.

NTSLF. (2005). "UK Tide Guage Network." United Kingdom National Tidal & Sea Level Facility, http://www.pol.ac.uk/ntslf/tgi/.

NTSLF. (2006). "Chart datum and ordnance datum." United Kingdom National Tidal & Sea Level Facility, http://www.pol.ac.uk/ntslf/tides/datum.html.

Prandle, D., and Wolf J. (1978). "Interaction of surge and tide in the North Sea and River Thames.", *Geophysical Journal of the Royal Astronomical Society,* London, A233, 407-418.

Pratt, I. (1995), "The storm surge of 21 February, 1993.", *Weather* 50(2), 42-48.

Pugh, D.T. (1987). "Tides, surges, and mean sea level: a handbook for engineers and scientists.", Wiley, Chichester, 472pp.

Rajasekaran, S., Thiruvenkatasamy, K., and Lee, T.-L. (2005). "Tidal level forecasting using functional and sequential learning neural networks." *Applied Mathematical Modeling*, In Press, Corrected Proof.

Recknagel, F., French, M., Harkonen, P., and Yabunaka, K., (1997). "Artificial neural network approach for modeling and prediction of algal blooms." *Ecological Modeling*, 96(1-3), 11-28.

RMS. (2003). "1953 U.K. Floods, 50 - year Retrospective." Risk Management Solutions, Inc, http://www.rms.com/Publications/1953_Floods_Retrospective.pdf.

Rossiter, J. R. (1954), "The North Sea storm surge of 31 January and 1 February 1953.", *Philosophical Transactions of the Royal Society A*, 246, 371–400.

Rossiter, J.R. (1961), "Interaction between tide and surge in the Thames.", *Geophysical Journal of the Royal Astronomical Society*, 6, 29-53

Rumelhart, D. E., Hintont, G. E., and Williams, R. J. (1986). "Learning representations by back-propagating errors." *Nature*, 323(6088), 533-536.

Rumelhart, D. E., Durbin, R., Golden, R., and Chauvin, Y. (1995). *Backpropagation: Theory, Architectures, and Applications*, Hillsdale.

Royal Society A (2005),"The Big Flood: North Sea Storm Surge", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, London,* McRobie, A., Spencer, T., Gerritsen, H. eds., 63, 1261-1491.

Saunders, R.D., and Clarke, D.C. (1998). "Statistical Hydraulic Analysis Robin David Saunders (SHARDS), Masters Thesis, School of Civil Engineering and the Environment, University of Southampton.

Schureman, P. (1958). *Manual of Harmonic Analysis and Prediction of Tides*, U.S. Department of Commerce, Coast and Geodetic Survey, Special Publication No. 98, U.S. Government Printing Office, Washington.

See, L., and Abrahart, R. J. (2001). "Multi-model data fusion for hydrological forecasting." *Computers & Geosciences*, 27(8), 987-994.

Sfetsos, A. (2002). "A novel approach for the forecasting of mean hourly wind speed time series." *Renewable Energy*, 27(2), 163-174.

Shennan, I., and Horton, B. (2002). "Holocene Land- and Sea-level changes in Great Britain." *Journal of Quaternary Science*, 17(5-6), 511-526.

Shi, F., Sun, W., and Wei, G. (1998). "A WDM method on a generalized curvilinear grid for calculation of storm surge flooding." *Applied Ocean Research*, 19, 275-282.

Solomatine, D. P., Rojas, C. J., Velickov, S., and Wüst, J. C. (2000). "Chaos theory in predicting surge water levels in the North Sea." *4-th International Conference on Hydroinformatics*, Iowa, USA

STFS. (2004). "Storm Tide Forecasting Service - Annual Report to the Department of the Environment, Food and Rural Affairs for the 2003/4 operational season."

Sztobryn, M. (2003). "Forecast of storm surge by means of artificial neural network." *Journal of Sea Research*, 49(4), 317-322.

Tang, Y.M., Holloway, P.,and Grimshaw, R. (1997), "A Numerical Study of the Storm Surge Generated by Tropical Cyclone Jane.", *Journal of Physical Oceanography,* 27(6), 963-976

Thirumalaiah, K., Deo, M.C. (1998). "River Stage Forecasting Using Artificial Neural Networks." *Journal of Hydrologic Engineering*, 3(1), 26-32.

Thompson, W. (1867). "Reports of the committee for harmonic analysis of tides." British Association for the Advancement of Science, London.

Tissot, P. E., Cox, D. T., and Michaud, P. (2001). "Neural network forecasting of storm surges along the Gulf of Mexico." *Proceedings of the Fourth International Symposium Waves 2001, Sep 2-6 2001*, San Francisco, CA, 1535-1544.

Tissot, P. E., Cox, D. T., Sadovski, A., Michaud, P. R., and Duff, S. (2004). "Performance and Comparison of Water Level Forecasting Models for the

Texas Ports and Waterways." *PORTS 2004 Conference*, Houston, TX, USA.

Tissot, P.E., Duff, S., Michaud, P.R., Rizzo, J., Jeffress, G., (2005), "DNR-TCOON: An integrated observational and operational forecast system for the Gulf of Mexico.", 85th American Meteorological Society Annual meeting, AMS Forum: Living in the Coastal Zone., San Diego, CA, USA, January 2005.

Tissot, P. E., Michaud, P. R., and Cox, D. T. (2003). "Optimization and Performance of a Neural Network Model Forecasting Water Levels for the Corpus Christi, Texas, Estuary." *3rd Conference on the Applications of Artificial Intelligence to Environmental Science, Long Beach, CA, USA, February 2003.*

Tsai, C., and Lee, T., (1999). "Back-propagation neural network in tidal-level forecasting." *Journal of Waterway, Port, Coastal and Ocean Engineering*, 125(4), 195-202.

Ultsch, A., and Röske, F. (2002). "Self-organizing feature maps predicting sea levels." *Information Sciences*, 144(1-4), 91-125.

Verlaan, M., Zijderveld, A., de Vries, H., Kroos, J., (2005), "Operational storm surge forecasting in the Netherlands: developments in the last decade", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 363(1831), 1441 - 1453

Walden, A. T., Prescott, P., and Webber, N. B. (1982). "An alternative approach to the joint probability method for extreame high sea level computations." *Coastal Engineering*, 6, 71-82.

Wordie, J.M., Waverley, L., Brunt, D., Williams, W.W. (1953), "The East Coast Floods: Discussion", *The Geographical Journal*, 3(119), 295-298

Xie, L., Pietrafesa, L., Peng, M. (2003). "Incorporation of a Mass-Conserving Inundation Scheme into a Three Dimensional Storm Surge Model." *Journal of Coastal Research*, 20(4), 1209-1223.

Appendix 1

1.1 Tide-surge analysis at Sheerness

Analysis of tide and surge was performed at the Sheerness station using MATLAB surge analysis tools developed by Haigh (2005). This analysis allows a better understanding of the tide and surge interaction and how to develop new tools used to predict them.

1.1.1 Tidal Data

The tidal data collected at the Sheerness gauge was obtained from the Proudman Oceanographic Laboratory (POL) web site (http://www.pol.ac.uk/ntslf/data.html). The years used are indicated in
Table A1 -1. The data also included the surge residual component of the levels derived from harmonic analysis performed by POL.

Table A1-1 General tidal data information

Chart datum / Ordinance Datum Conversion	Data Length (years)	Start Date - End Date
2.90	38	1952, 1958, 1965-1975, 1980-2005

1.1.2 Analysis of the Surge Component

A frequency distribution of surge levels for Sheerness was calculated and can be seen in Figure A1-1. To a first approximation the distribution of the surge residuals shows a classic bell-shaped appearance of the normal or Gaussian distribution. This is expected for a large sum of random variables. However, there are important differences between this and the theoretical normal distribution, (Pugh, 1987). The plot of the distribution shows extended positive and negative tails which include major surge events. The distribution also indicates the tendency for the more frequent occurrence of larger positive surge residuals than negative residuals, shown by the positive skew in the distribution. This observation matches previous findings for other British Ports, (Pugh, 1987).

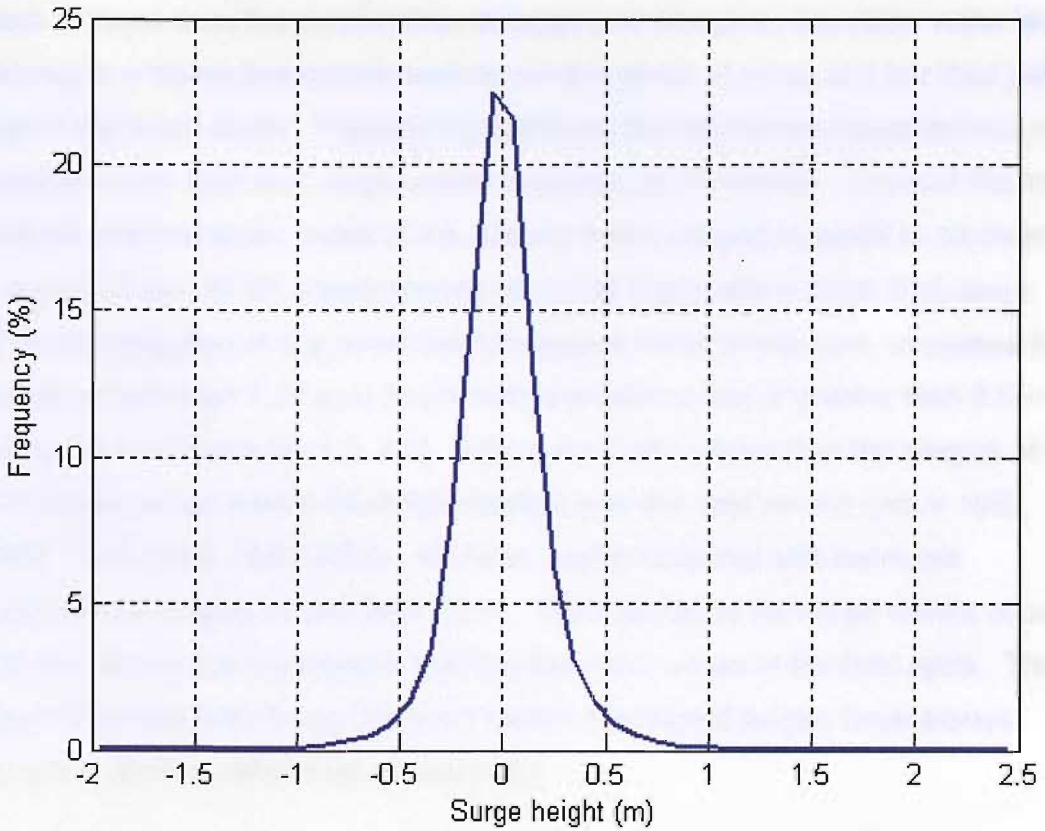


Figure A1-1 Surge frequency plot for sheerness station 1990.

1.2 Extreme Water Levels

Extreme water levels occur when surges and tides combine to give high water levels. The timing of the surge is of particular importance, as a large surge event that occurs during low tide can go unnoticed. Extreme water levels generally occur when moderate to high surges occur over mid to high tidal heights. An improved understanding of the causes of extreme sea levels at a site can be obtained by determining whether a site is either tide or surge dominated (Dixon and Tawn, 1994).

Plotting surge levels against the corresponding predicted tidal level can be used to determine whether or not a site is surge or tide dominated. Such a plot for the Sheerness water-level data can be seen in Figure A1-2. This plot was created using the 100 highest water-level events (Blue/dark circles) in the data record (years 1952, 1958, 1965-1975, 1980-2006) and the 100 highest surge events (Red/light circles). For both type of events, the predicted tidal component is plotted against the surge component at the time of maximum water level. The diagonal lines that cross the plot are contours of equal total water level and also indicate the total measured water-level or harmonic + surge components. Along

each of these lines the combination of surge and tide gives the same water level. Moving to a higher line corresponds to combinations of surge and tide that yield higher still water levels. Figure A1-2 highlights the interesting characteristics of extreme water level and surge events occurring at Sheerness. The plot illustrates that the extreme water levels at this site are being caused by small to moderate surges (0.5 to 1.25 m) superimposed on a tidal signal with a large tidal range. With the exception of one event the 100 largest water levels were all caused by surges of less than 1.25 m in height with a predicted tide of greater than 2.5 m (Note the lone event at ~1.5, 2.0). Figure A1-2 also shows that the heights of the 100 largest surge events (Red/light circles) over the data record (years 1952, 1958, 1965-1975, 1980-2006). All these events occurred with harmonic components heights of less than 2.5 m. The majority of the surge events occurred with elevations that correspond with the low-water stage of the tidal cycle. This figure indicates that during the years tested, the largest surges have always occurred at times other than at high water.

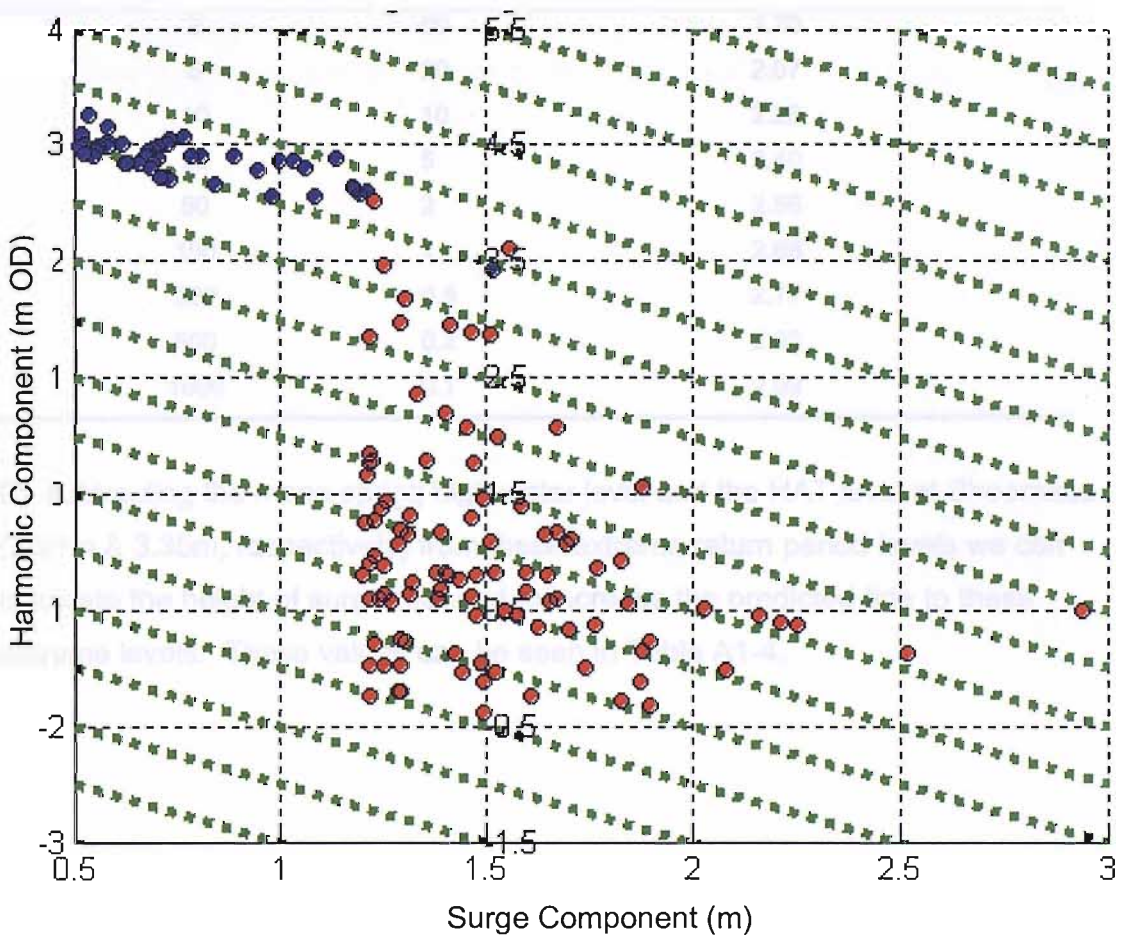


Figure A1-2 100 Largest surge and extreme water-level events.

1.2.1 Characteristics of Extreme Events

Mean spring and neap values, along with LAT (lowest astronomical tide) and HAT (highest astronomical tide) values for Sheerness are presented in Table A1-2. Extreme water-level and surge return period intervals were calculated for the Sheerness tide station for 2, 5, 10, 20, 50, 100, 200, and 500 year events. This was done using SHARDS, a statistical hydraulic analysis program created by Saunders R.D. and Clarke, D.C. (1998) using the Annual Maximum Method. The results can be seen in Table A1-3.

Table A1-2 Tidal levels at Sheerness (Chart Datum)

LAT	MLWS	MLWN	MHWN	MHWS	HAT
-0.07	0.54	1.46	4.74	5.81	6.25
-2.97(OD)	-2.36(OD)	-1.42(OD)	1.84(OD)	2.91(OD)	3.35(OD)

Table A1-3 Return periods for Surge at Sheerness

Return Period (years)	AEP (%)	Surge Component (m)
2	50	1.70
5	20	2.07
10	10	2.23
20	5	2.40
50	2	2.56
100	1	2.68
200	0.5	2.77
500	0.2	2.90
1000	0.1	2.99

By subtracting the mean spring high water level and the HAT level at Sheerness (2.91m & 3.35m, respectively) from these extreme return period levels we can calculate the height of surge required to increase the predicted tide to these extreme levels. These values can be seen in Table A1-4.

Table A1-4 Surge heights required to raise high water of a mean spring and a HAT tide to specific return period levels

Return Period (years)	Of A Mean Spring Tide To Return Period Level (m)	Of A HAT Tide To Return Period Level (m)
2	0.91	0.21
5	1.10	0.40
10	1.22	0.52
20	1.32	0.62
50	1.45	0.75
100	1.54	0.84
200	1.63	0.93
500	1.74	1.04

In order to try and understand some of these features of extreme surges and water level events further, the characteristics of the 20 highest water level and surge events were examined in more detail over the data length (years 1952, 1958, 1965-1975, and 1980-2005). The dates of the 20 highest water level events can be seen in Table A1-5 and surge events in Table A1-6. The table shows the maximum water levels experienced with these events and the corresponding predicted tide and surge levels that combined to achieve these observed water levels.

Table A1-5 Highest recorded water level events at Sheerness for years 1952, 1958, 1965-1975, 1980-2005

Event	Date and Time of Extreme Water Level	Measured Extreme Water Level (m ODN)	Predicted Harmonic Water Level (m ODN)	Corresponding Surge Level (m)	Return Interval
1	10/12/1965 13:00	4.023	2.884	1.139	100
2	14/12/1973 03:00	3.903	2.872	1.031	20
3	14/12/1993 00:15	3.866	2.806	1.060	20
4	29/10/1996 13:30	3.865	2.868	0.997	20
5	15/11/1993 00:45	3.830	3.065	0.765	10
6	24/12/1988 13:00	3.818	2.641	1.177	10
7	16/12/2005 13:00	3.817	2.605	1.212	10
8	25/11/1973 00:00	3.803	2.620	1.183	10
9	22/02/2004 14:30	3.785	2.896	0.889	10
10	20/02/1996 01:00	3.784	3.050	0.734	10
11	28/02/1998 13:45	3.783	3.244	0.539	10
12	25/01/1993 14:00	3.772	2.578	1.194	5
13	08/10/1998 14:00	3.753	3.339	0.414	5
14	28/01/1994 12:45	3.735	2.790	0.945	5
15	02/02/1983 03:00	3.732	2.786	0.946	5
16	07/10/1990 14:00	3.726	3.144	0.582	2
17	08/02/2001 12:30	3.724	2.999	0.725	2
18	11/01/1993 14:00	3.718	2.912	0.806	2
19	18/09/2001 01:00	3.690	3.210	0.480	2
20	12/02/2001 15:45	3.689	2.987	0.702	2

Of the 20 largest water-level events, the 5 smallest events exceeded the 1 in 2-year return period; 4 events exceeded a 1 in 5 year return period; 7 events exceeded a 1 in 10 return interval; 3 events exceeded a 1 in 20 year return period; and the largest event exceeded a 1 in 100 return interval. 9 events occurred as a result of a combination of a surge with a tide with a predicted high water height larger than that of the MHWS value (2.91m), and the remaining 11 events were within 0.30m of predicted MHWS value. The largest surge level at the time of high water was 1.139m, indicating that extreme water level events at Sheerness are indeed being caused by moderate surges with larger astronomic tidal levels. The 20 largest surge events are given in Table A1-6 which shows the maximum surge height that was experienced during the event and the corresponding measured

and predicted water levels at this time. The largest surge height experienced at Sheerness over the data
interval was 2.94m, which corresponds to a 1 in 500 year return interval event.

Table A1-6 Highest recorded surge events at Sheerness for years 1952, 1958, 1965-1975, 1980-2005

Event	Date and Time of Extreme Water Level	Measured Extreme Water Level (m ODN)	Predicted Harmonic Water Level (m ODN)	Corresponding Surge Level (m)	Return Interval
1	21/02/1993 08:45	1.943	-0.998	2.941	500
2	14/02/1989 11:00	1.153	-1.368	2.521	20
3	29/09/1969 10:00	1.123	-1.131	2.254	10
4	06/03/1968 12:00	1.096	-1.116	2.212	5
5	01/02/1983 23:00	1.108	-1.055	2.163	5
6	12/12/1990 16:00	0.569	-1.510	2.079	5
7	05/02/1999 11:45	1.045	-0.987	2.032	2
8	24/11/1981 19:00	0.616	-1.280	1.896	2
9	26/12/1985 18:00	0.065	-1.831	1.896	2
10	08/02/2004 21:30	0.531	-1.351	1.882	2
11	30/01/2000 03:45	1.933	0.054	1.879	2
12	13/12/1974 07:00	0.253	-1.619	1.872	2
13	21/11/1971 22:00	0.907	-0.940	1.847	2
14	02/04/1973 20:00	1.253	-0.576	1.829	2
15	19/02/1996 08:15	0.046	-1.783	1.829	2
16	06/04/1967 07:00	1.131	-0.640	1.771	2
17	04/01/1984 09:00	0.644	-1.121	1.765	2
18	30/01/2003 18:15	0.238	-1.502	1.740	2
19	10/01/1995 10:00	1.323	-0.384	1.707	2
20	27/10/2002 23:00	0.523	-1.178	1.701	2

It is important to note that only 2 of these extreme surge events led to or contributed to the 20 largest extreme water levels experienced in at Sheerness. The 5th largest surge event resulted in the 15th largest extreme water level event on record, and the 15th largest surge event resulted in the 10th largest water-level. On the later, the surge event reached a maximum height of 1.8m at 08:15 on the 19th of February 1996 around the time of low water. 17 hours later the surge had decreased to 0.73m but when combined with a large astronomic tide of 3.05m resulted in an extreme water level event equivalent to a 1 in 10 year event. With

each of the other 19 largest surge events although the surge level was still moderately large when high water occurred, high water levels were not experienced because the astronomical tide ranges were much lower.

It is also interesting to note that if any of these 20 surge events had occurred in such a way that the time of maximum surge was around high water of a mean spring tide, a 1 in 1000 year or larger extreme return period event would have occurred. If the 20 events had occurred in such a way that the time of maximum surge was around high water of a tide about the same height as Highest Astronomical Tide the resultant events would have all been greater than a 1 in 1000 year extreme return period event.

Tide and surge interaction

Tidal / surge interaction plays a large part in the magnitude of and arrival times of large surge events in the North Sea. The speed of propagation of storm surge and is proportional to the water-depth. The tidal range of the North Sea and its relatively shallow depth causes the speed of propagation of a surge to be affected by tidal phase. Tidal circulation in the North Sea rotates counter-clockwise around the primary amphidromic point, causing tides to propagate from north to south along the East coast of the UK. The circulation is shown for the largest tidal constituent (M2) in Figure A1-3 where Co-tidal lines (radiating outward from amphidrome) connect points experiencing the same phase of the tide. Co-range lines (circling amphidrome) connect points of equal tidal range. Figure A1-3 also shows that it takes approximately 15 hours for a tide to travel from Wick to Sheerness. Analysis of storm data from tidal records show that a typical storm surge would take 12 to 20 hours to traverse this distance, and indicates that tidal and storm surges propagate at approximately the same speed, allowing interaction between the two phenomena.

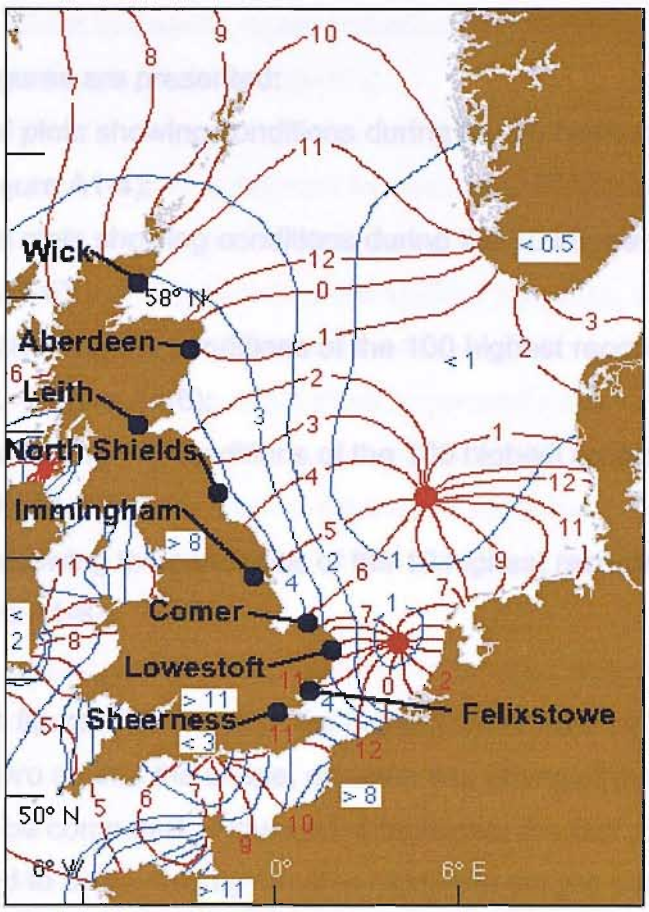


Figure A1-3 Tidal circulation in the North Sea (M2 Constituent)

Tide-surge interaction is caused by frictional resistance, associated with strong currents, and variations in the speed of wave propagation, which modify a surge in the presence of the tide (Walden et al., 1982) Changes in water level, due to the tide, alter the propagation and generation of surge events. Analysis of tide and surge was performed at the Sheerness station using MATLAB surge analysis tools developed by Haigh (2004). This analysis allows a better understanding of the tide and surge interact and how to develop new tools used to predict them.

1.3 Profile Analysis

To better understand a typical storm surge profile at the Sheerness tidal station it is important that not only the height and timing of the surge on the astronomical tide are included accurately but also the shape and duration of the surge profile. The shape and duration of extreme water levels and surge events have been examined by plotting the observed water levels, predicted tide and surge profiles over the duration of the extreme events.

The following figures are presented:

- Individual plots showing conditions during the 10 highest recorded water level events (Figure A1-4);
- Individual plots showing conditions during the 10 largest recorded surge events (Figure A1-5);
- A figure showing the conditions of the 100 highest recorded water level events overlain. (Figure A1-6);
- A figure showing the conditions of the 100 highest recorded surge events overlain. (Figure A1-7);
- A figure showing the conditions of the 10 highest recorded surge events overlain. (Figure A1-8);

In each of these figures the timing of the highest water level or largest surge has been offset to zero so that the shape, duration and timing of the surge relative to tidal phase can be compared. Figure A1-4 highlights the fact that the largest water level events tend to be caused by small to moderate surges occurring around high water of a relative large tide. For the majority of the 100 largest recorded water level events the surge profile is very noisy and lacks any definite shape. This is clear from Figure A1-6, where the 100 largest surge events are overlain. This corresponds with similar work undertaken in the English Channel (Haigh, 2004). At a number of sites within the English Channel the moderate surge levels experienced around the extreme water level event were caused by local wind arfects or a number of weak low pressure systems over the area.

Another interesting feature of these plots is that for a number of events the time of extreme water level occurs a few hours after the peak of the surge. This backs up the point made earlier, that some sort of interaction seems to be occurring similar to that of the southern North Sea, which tends to result in the peak of the surge occurring on the rising tide. The 10 largest surge events recorded are plotted individually in Figure A1-5. In these figures the surge profiles although all different do appear to show a general shape. This is more clearly seen in Figure A1-8 where the 10 events are overlain. Previous research (Haigh, 2004) has shown that these larger surge events tend to occur primarily as a result of large strong low- pressure systems and not local meteorological conditions. Figure A1-5 shows that the peak of the surge in the majority of the 10 largest

events occurs around low water, again indicating that some sort of tide/surge interaction process appears to be occurring.

It is clear from these plots for the extreme water level and surge events that the shape of the surge profile is different for each event. This is to be expected as weather effects are present at all time and space scales, and small changes in weather can lead to large changes in water levels. However, as previously mentioned there does appear to be a general shape that can be identified for large surge events. Before the surge event there is generally a small negative surge. Over the next 15 hours the surge increases up to a maximum level. Over the next 15 hours the surge tends to decrease, then again increase in height slightly at about 10 hours after the peak, and then decrease down to zero. In total the event last for about 30 hours.

For the majority of the 10 largest surge events on record, the peak of the surge tends to occur within ±3 hours either side of low water. The secondary small peak in surge appears to then occur around the next low water. This feature of the surge shape can also be seen in Figure A1-7, for the surge profiles experienced during large water level events, although on a smaller scale. The surge events tend to reach a peak around the time of low water before or after the extreme water level is experienced, again suggesting interaction between the surge and tide.

In order to examine in more detail the typical shape, duration, and timing of the surge profile experienced during the extreme water level and surge events, the 100 largest water-level and surge events were plotted (Figure A1-6 and Figure A1-7). As before, the plots show a large variation in their surge profile, but they emphasize the point that extreme water level events tend to be caused by smaller, weaker surges. Whereas the larger strong surge events tend to occur around the time of low water and do not tend to lead to extreme water levels. These large events do appear to have a general surge shape that has duration of approximately 30 hours.

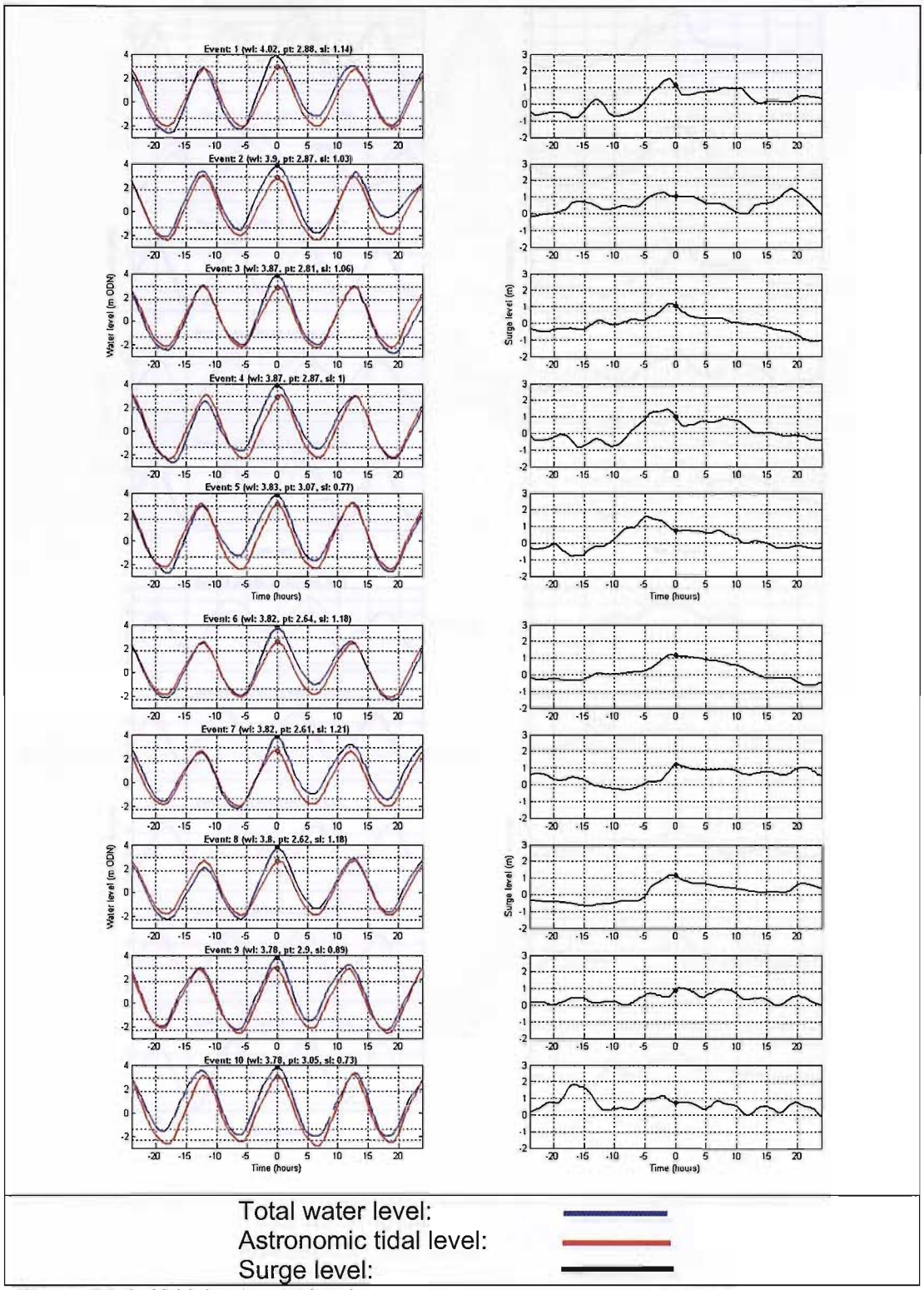


Figure A1-4 10 highest water level events

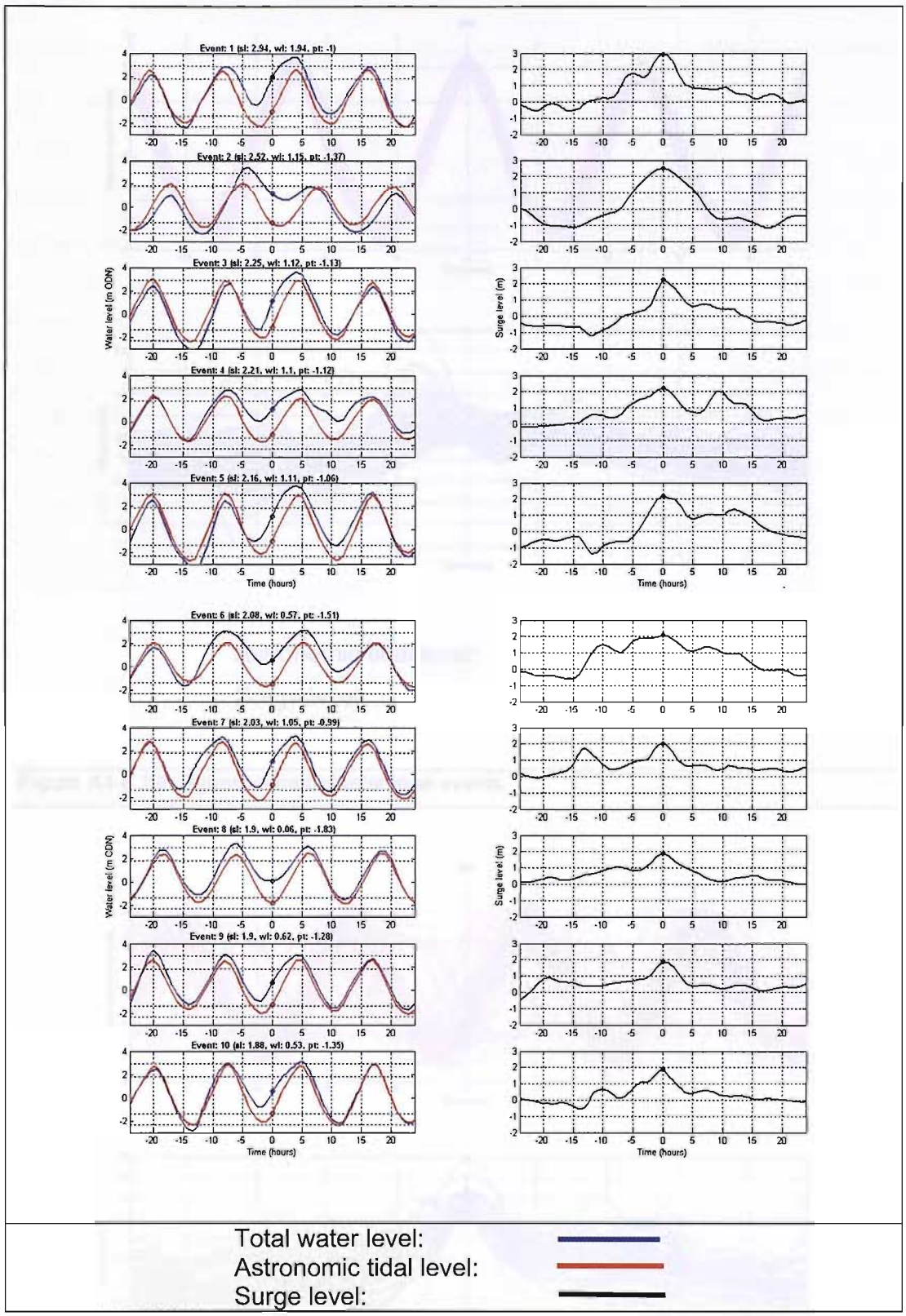


Figure A1-5 10 highest surge events

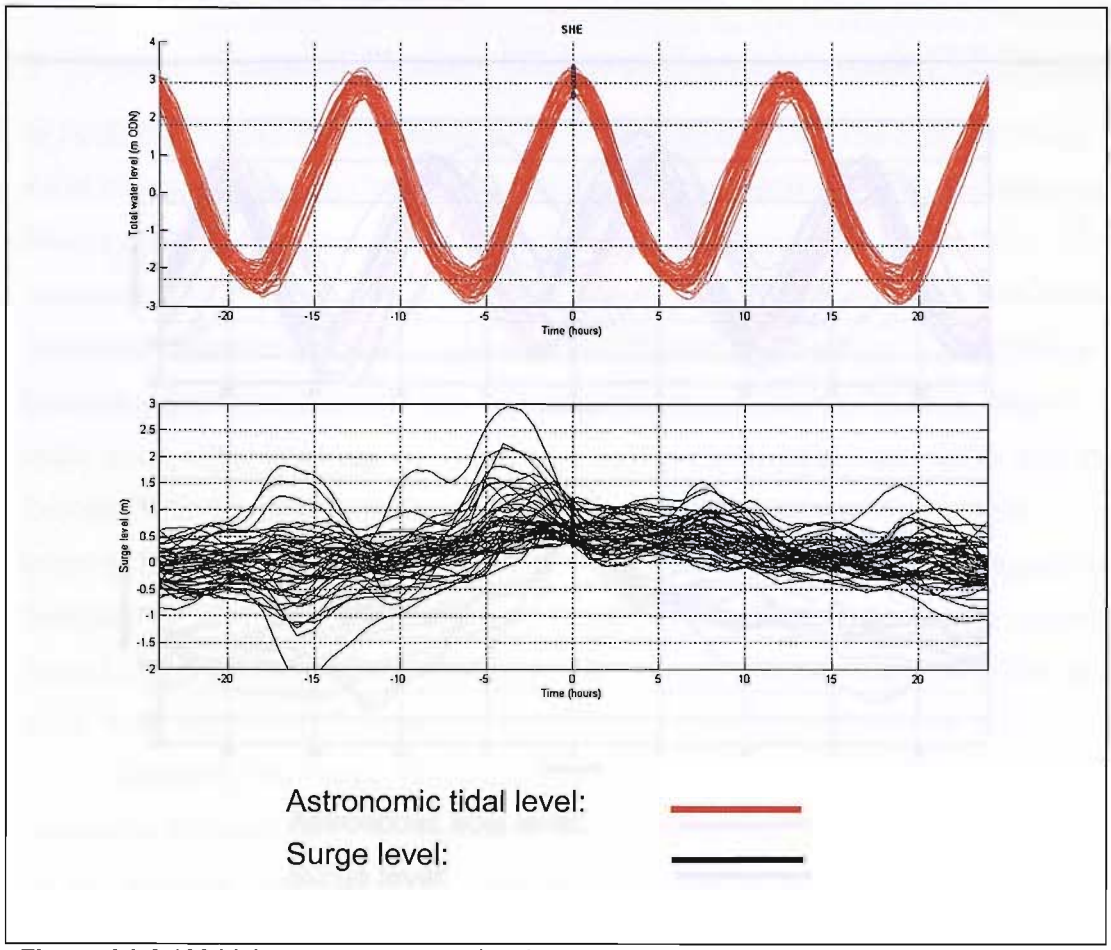


Figure A1-6 100 highest extreme water level events

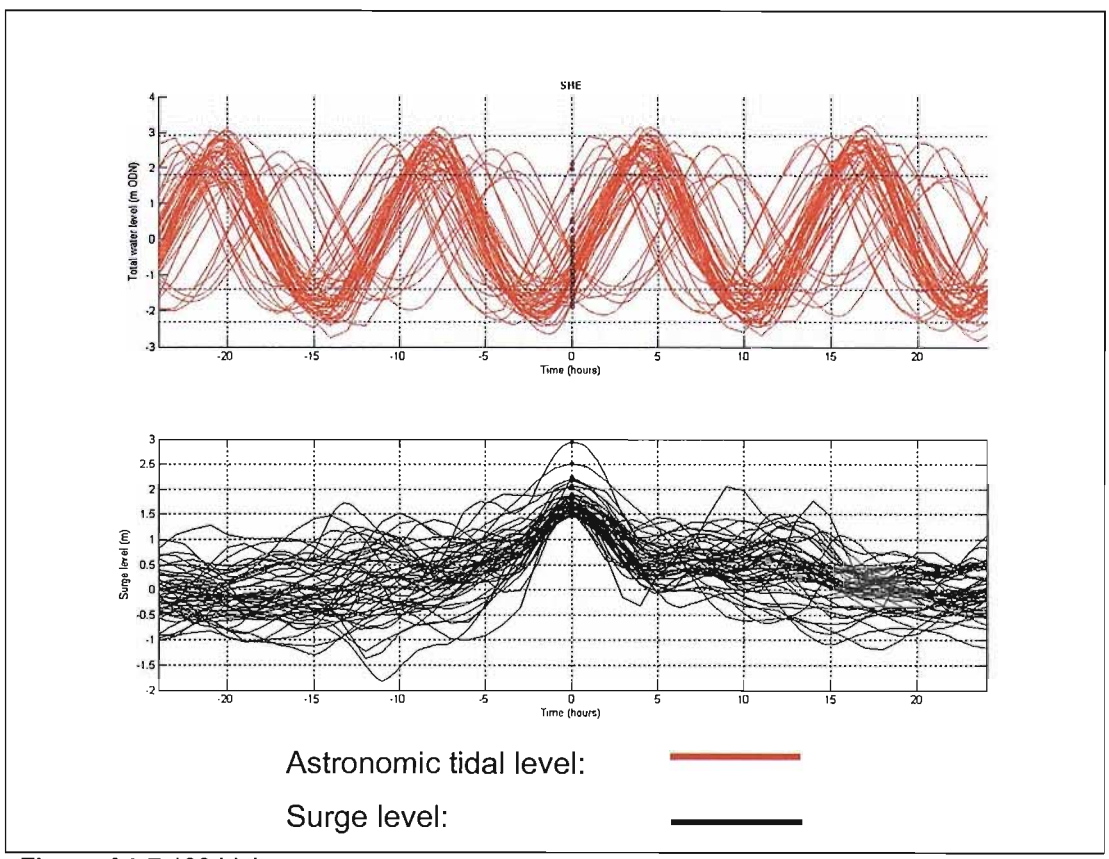


Figure A1-7 100 highest extreme surge events

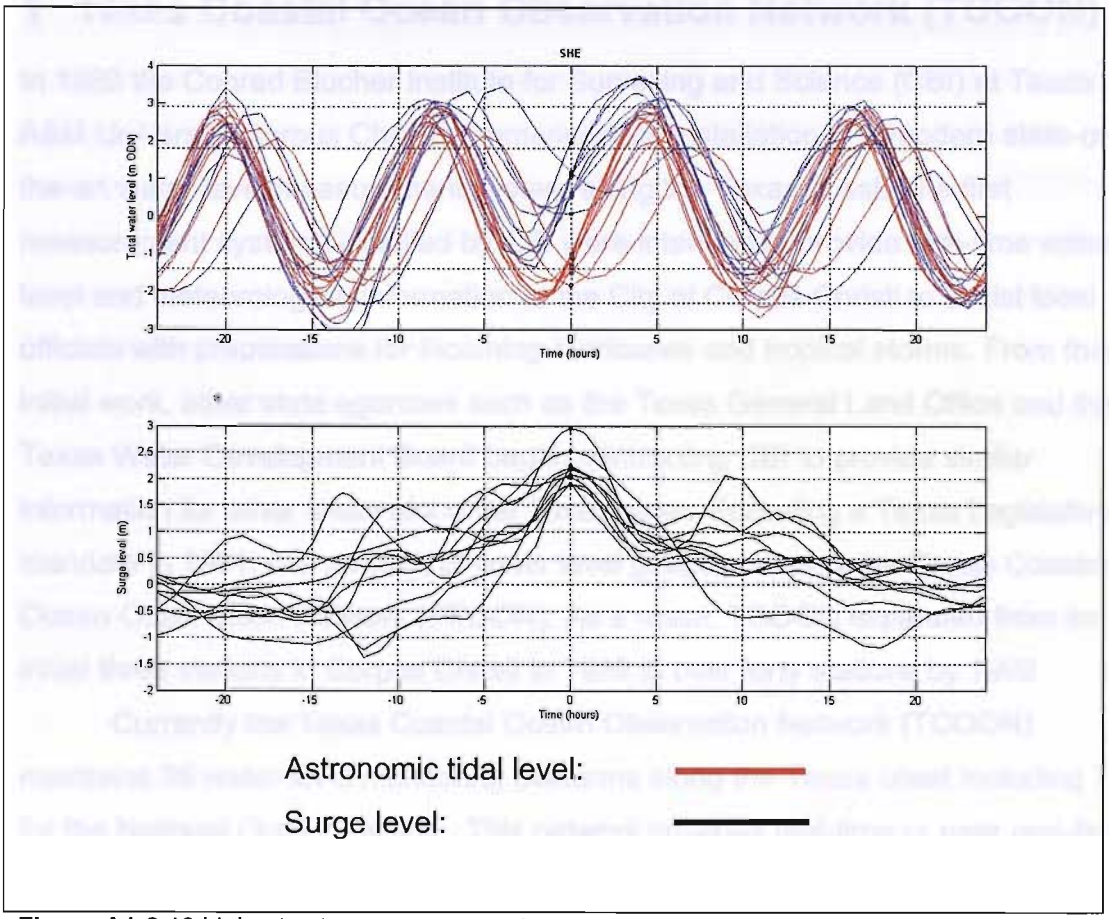


Figure A1-8 10 highest extreme surge events

1 Texas Coastal Ocean Observation Network (TCOON)

In 1989 the Conrad Blucher Institute for Surveying and Science (CBI) at Texas A&M University-Corpus Christi commenced the installation of a modern state-of-the-art water-level measurement system along the Texas coast. The first measurement systems installed by CBI were intended to provide real-time water-level and meteorological information to the City of Corpus Christi to assist local officials with preparations for incoming hurricanes and tropical storms. From this initial work, other state agencies such as the Texas General Land Office and the Texas Water Development Board began contracting CBI to provide similar information for other areas along the Texas coast. Following a Texas Legislative mandate in 1991, this network of water level gauges became the Texas Coastal Ocean Observation Network (TCOON). As a result, TCOON expanded from an initial three stations in Corpus Christi in 1989 to over forty stations by 1992.

Currently the Texas Coastal Ocean Observation Network (TCOON) maintains 36 water-level monitoring platforms along the Texas coast including 7 for the National Ocean Service. This network provides real-time or near real-time coastal measurements such as water levels, wind speeds and wind directions, barometric pressures as well as other variables such as dissolved oxygen, salinity, water currents and wave climates depending on the station. The primary use of the data has been to establish tidal datums, but increasingly the network has provided data for many other uses including the commercial shipping industry, recreational boaters, sailors and windsurfers, the shrimp and fishing industry, marine construction, and decision-makers responsible for marine safety and emergency evacuation in the event of an approaching hurricane.

A distinctive feature of the TCOON network is its unique data management software which provides data in real-time or near real-time to its sponsors and the community through the World Wide Web and through automated phone services. The software and procedures were developed on the principle that all user interaction with the data management system takes place via web-based interfaces. Such interactions include for example site visit and maintenance reports and chain-of-custody records. Sponsors, scientists and other potential users can access all TCOON data in a variety of graphic and text-based formats from http://lighthouse.tamucc.edu/. The TCOON software and hardware were built and combined with a primary goal of reliability. The data are housed in ordinary

PC-base computers rather than sophisticated proprietary systems such that parts or whole systems can be replaced quickly if needed. The software is based on open source technologies such as Linux and Perl such that TCOON is not subject to changes in proprietary systems and has the flexibility to replace software components as new technologies become available or as the needs of TCOON's sponsors evolve. TCOON has started providing access to all its software under the General Public License and aims at co-developing future versions of the web-based software and procedures.

```
%*********************************************************************
%   Copyright Daniel B. Prouty 2007.  Code used for reading directly
%   one year of raw data extracted from the British Oceanographic
%   Data Centre (BODC). The time is in UTC and the metric
%   system is used for all measurements.  The code first removes
%   the 12 comment lines at the top of the file then creates
%   and intermediary file where only the data is stored with no
%   comments or titles.  The data in the intermediary file is then
%   loaded into a matrix.  The code first determines what percentage
%   of the data is missing(NaN).  The code then looks for gaps and
%   spikes in the data.  The missing data is interpolated linearly
%   across the gaps and the spikes are smoothed linearly.  In this
%   version, the water level gaps are filled in by interpolating
%   linearly the water level difference.
%*********************************************************************
tic;
clear;
fprintf('----------------- Starting Program --------------------\n');

% ! Remember to set all missing data to '9999' i.e. both NA data and RM
data

Station = 'WIC';
Year = '2001';

Year_a = strcat(Year,'_A_');    % Adds '_A_' to match new file name
format
Path = strcat('C:\AAA\',Station,'\');
FileName = strcat(Year_a,Station);
InputFileName = strcat(FileName);
OutputFileName = strcat(FileName,'NN');
DataAnalysisFileName = strcat(Path,FileName,'Analysis.txt');
LoadFileName = strcat(Path,InputFileName,'.txt');
SaveFileName = strcat(Path,OutputFileName,'.txt');
TempFile = 'TempFile';
FID1 = fopen(LoadFileName,'rt');
Comment='#';
StationNumber = 999;

ONE='END';
% ******************************************************
% Transfers the actual data to a temporary working file
% ******************************************************

FID1 = fopen(LoadFileName,'rt');            % Open Input file
[YearDay, Time, pwl, res, offswl, harmwl]=textread(LoadFileName,'%d+%d %d
%d %d %d','headerlines',1);
YearOne = [YearDay, Time, pwl, res, offswl, harmwl];
Status = fclose(FID1);                      % Close Input file
fprintf('--------------- Data Loaded into Matrix ----------------\n');
% *********************************************************************
% Analyzes each data stream check for missing data and data gaps and
% prints a report
% *********************************************************************

FID4 = fopen(DataAnalysisFileName,'wt');
DataSetLength = size(Time);
%Id = ['YearDate'; 'Time    '; 'pwl     '; 'res     '; 'offswl  '; 'harwl
'];
%Identifiers = cellstr(Id);
for i=1:6
    Identifiers(i,:)='        ';
```

```
end
Identifiers(1,:)='YearDate';
Identifiers(2,:)='Time    ';
Identifiers(3,:)='pwl     ';
Identifiers(4,:)='res     ';
Identifiers(5,:)='offswl  ';
Identifiers(6,:)='harmwl  ';

fprintf(FID4,'Data originating from BODC raw data file: %s
\n\n',InputFileName);
fprintf(FID4,'Number of data series read from the raw data file including
date and hour time series: %i \n',DataSetLength(1));
fprintf(FID4,'Name of data series read from the raw data file:\n');
for i =2:6
    ColumnHeader = Identifiers(i);
    CH = char(ColumnHeader);
    fprintf(FID4,'  %s\n',Identifiers(i));
end
fprintf(FID4,'\nNumber of data points per series: %i
\n',DataSetLength(1));
fprintf(FID4,'Representing %i days of data\n',DataSetLength(1)/24);
fprintf(FID4,'Starting time of the data series: Day %d @
%d\n',YearDay(1),Time(1));
fprintf(FID4,'Ending time of the data series: Day %d @
%d\n',YearDay(size(YearDay)),Time(size(Time)));

% Analyzes the missing data for each data series starting with the
percentage of missing data

for i = 1:6
    Flagged(i)=0;
    LongestFlagged(i)=0;
    ConsecutiveFlagged(i)=0;
    LongestFlaggedStart(i)=0;
    LongestFlaggedEnd(i)=0;
    DataSetStatus(i)=0;
end

for i = 3:6
    for j = 1:DataSetLength
        if  YearOne(j,i) == 9999
            Flagged(i)=Flagged(i)+1;
            ConsecutiveFlagged(i)=ConsecutiveFlagged(i)+1;
            if j == DataSetLength
                if ConsecutiveFlagged(i) > LongestFlagged(i)
                    LongestFlagged(i) = ConsecutiveFlagged(i);
                    LongestFlaggedStart(i) = j-LongestFlagged(i)+1;
                    LongestFlaggedEnd(i) = j;
                end
            else
                if YearOne(j+1,i) ~= 9999
                    if ConsecutiveFlagged(i) > LongestFlagged(i)
                        LongestFlagged(i) = ConsecutiveFlagged(i);
                        LongestFlaggedStart(i) = j-LongestFlagged(i)+1;
                        LongestFlaggedEnd(i) = j;
                    end
                    ConsecutiveFlagged(i) = 0;
                end
            end
        end
    end
end
```

```
for i = 3:6
    fprintf(FID4,'Missing data for time series "%s": %i data points or
%5.2g
percent\n',Identifiers(i,:),Flagged(i),100*(Flagged(i)/DataSetLength(1)))
;
    if (100*(Flagged(i)/DataSetLength(1))) > 10
        DataSetStatus(i)=0;
    else
        DataSetStatus(i)=1;
    end
end

% Identifies the longest missing data set for the time series

fprintf(FID4,'\n');
for i = 3:6
    fprintf(FID4,'Longest gap for time series "%s": %i data points
\n',Identifiers(i,:),LongestFlagged(i));
    fprintf(FID4,'     Starting at data point %i and ending at data point
%i\n',LongestFlaggedStart(i),LongestFlaggedEnd(i));
end

% ********************************************************************
% Interpolates linearly accross the data gaps for the relevant series
% (not for data sets with more than 10% of the data missing).  Also
% builds a time series of the interpolated points.
% ********************************************************************

for j = 1:DataSetLength(1)
    if pwl(j) == 9999
        GapStart=j;
        while pwl(j) == 9999 & j<=DataSetLength(1)
            j=j+1;
        end
        if j<DataSetLength(1)
            GapEnd=j-1;
            slope=(offswl(GapEnd+1)-offswl(GapStart-1))/(GapEnd-
GapStart+2);
            for k=GapStart:GapEnd
                offswl(k)=offswl(GapStart-1)+(k-GapStart+1)*slope;
                pwl(k)=harmwl(k)+offswl(k);
            end
        end
    end
end


% ***********************************************************
% Computes and prints the statistitics for the haronic
% model
% ***********************************************************

HarmSkill = SkillSet(harmwl,pwl,harmwl,150,300);

SkillLabel(1,:)='Average Error:                                    ';
SkillLabel(2,:)='Average Absolute Error:                           ';
SkillLabel(3,:)='Standard Deviation of the Error:                  ';
SkillLabel(4,:)='Root Mean Square of the signal:                   ';
SkillLabel(5,:)='Root Mean Square Error:                           ';
```

```
SkillLabel(6,:)='Normalized RMS Error:                                    _';
SkillLabel(7,:)='Central Frequency (X=15 cm):                             _';
SkillLabel(8,:)='Positive Outlier Frequency (X=30 cm):                    _';
SkillLabel(9,:)='Negative Outlier Frequency (X=30 cm):                    _';
SkillLabel(10,:)='Maximum Duration of Positive Outlier (X=30 cm):  _';
SkillLabel(11,:)='Maximum Duration of Negative Outlier (X=30 cm):  _';
SkillLabel(12,:)='Worst Case Outlier Frequency (X=15 cm):          _';
SkillLabel(13,:)='Worst Case Outlier Frequency (X=30 cm):          _';

fprintf(FID4,'\n\nSkills for Harmonic model for year
%i:\n',str2num(Year));
fprintf(FID4,'****************************************\n\n');
fprintf(FID4,'                    Skill                 Tide Tables\n');
for i=1:13
    fprintf(FID4,'%s   %6.4f\n',SkillLabel(i,:),HarmSkill(i));
end

status = fclose(FID4);

% ***********************************************************
% Saves the interpolated data sets into a NN ready data file
% ***********************************************************

FID5 = fopen(SaveFileName,'wt');
for i=1:DataSetLength
    count = fprintf(FID5,'%d   %d   %d   %d
%d\n',YearDay(i),Time(i),pwl(i),harmwl(i),res(i));
end
status = fclose(FID5);
toc;
plot_all_V5;
fprintf('----------------    Done !!!          -----------------\n');
```

```
%*************************************************************************
**
%*    Copyright 2007 Daniel Prouty
%*    This program performs a correlation analysis using two tide gauge
%*    stations.  The maximum time shift used to correlate the tide
%*    stations is 24 hours (See MaxLagShift).  The program can also be
%*    used to correlate other columns of data in the same files (See
%*    note on Surge1 and Surge2)
%*************************************************************************
**
%clf                    % Clear active figure (comment out to over-plot)
Station1 = 'she';       % Primary Station
Station2 = 'nsh';       % Secondary station
Year = '1992';          % Testing year
MaxLagShift = 24;       % The Maximum time offset to use +/-

fname1 = strcat('C:\AAA\',Station1,'\',Year,'_A_',Station1,'NN.txt');
fname2 = strcat('C:\AAA\',Station2,'\',Year,'_A_',Station2,'NN.txt');

A = load (fname1);      % Load Primary data
B = load (fname2);      % Load Secondary data

Surge1 = A(:,5);    %Note use 3 for Total water-levels and 5 for surge
                    %residuals
Surge2 = B(:,5);    %Note use 3 for Total water-levels and 5 for surge
                    %residuals
[c,lags] = xcorr(Surge1,Surge2,MaxLagShift,'coeff'); % Run Correlation
                                                     % and Normalize to
                                                     % 1.0
%[c,lags] = xcorr(Surge1,Surge2,MaxLagShift);        % Run Correlation
                                                     % only
[Y,I] = max(c);         % Find maximum correlation
clc;                    % clear console
fprintf('The best lag period = %d\n',I-MaxLagShift); % print lag period
fprintf('The Cross-Covariance at peak = %3.3f\n',Y); % print corr value

figure(100);                                         % start new fig
Result = num2str(I-MaxLagShift);     % Convert result to string
plot(lags,c);                        % Plot lag Period
hold on;                             % Keep graph and plot more
plot(I-MaxLagShift-1,Y,'.','MarkerFaceColor','k');   % Plot Marker on
                                                     % max
text(I-MaxLagShift-1,Y,Result);                      % Plot Max text
```

```
%*************************************************************************
%*   Copyright 2007 Daniel Prouty
%*   This program compares CS3/ANN models and requires Matlab export
%*   files from the ANN program.  This program also requires POL Archived
%*   data sets.
%*   This program compares CS3 and CS3 model with bias (at T+00)
%*   correction to The ANN model. This program extracts storms with
%*   surges greater than the value given by PEAK and compares common
%*   values depending on the forecast interval.  This program also
%*   computes model performances for yearly intervals also.
%*
%*************************************************************************

% PEAK Storm surge height for storm selection
PEAK = 1.0;
% f = Forecast interval +1 (since data starts at T+00)
f = 1 + 1;

%------------------
%load 1 hour ANN model forcast
load 1_2002_SHE_IMM.mat

%conver to metres
Data = (new_matrix')/1000;

%Surge = water level - astronomical tides
Mann = Data(:,3) - Data(:,1);
Measured = Data(:,2) - Data(:,1);
%------------------

%------------------
%load in POL Data
Mpol = load('sheerness/JAN_DEC_02.txt');
[a,b] = size(Mpol);

Mpol = Mpol';
Mpol2 = reshape(Mpol, a*b, 1);
le = length(Mpol2);

%Just want to compare 1 hour forecast
%------------------

%------------------
%Just 1 hour forecasts
ts1 = 0:1:(le-1); % TIME STAMP for using every value
ts2 = (f-1):6:le; % TIME STAMP for using every 6th value

%subset
Mann_6 = Mann(f:6:le);          %subset for ANN - Measured
Mpol2_6 = Mpol2(f:6:le);        %subset for POL - Measured
Measured_6 = Measured(f:6:le); %subset for every 6th measured
POL_Zero_Forecast = Mpol2(1:6:le);          %POL value at zero forecast
*************
Measured_at_Zero = Measured(1:6:le);        %Actual measurement at zero
*************
Bias = Measured_at_Zero - POL_Zero_Forecast %Bias
**********************************
%------------------
%plot data

figure(1);
subplot(2,1,1)
```

```
plot(ts2, Mann_6, 'sr');
hold on
plot(ts2, Mpol2_6, 'ob');
plot(ts1,Measured(1:le), '-+k');
grid
xlabel('Hours');
ylabel('Surge Elevation (m)')
title('One-Hour Forecast', 'fontweight', 'bold');
legend('ANN', 'STFS', 'Measured')
%Difference plot
subplot(2,1,2)
plot(ts2, Mann_6-Measured_6, '-r');
hold on
plot(ts2, Mpol2_6-Measured_6, '-b');
grid
xlabel('Hours');
ylabel('Surge Elevation (m)')
title('All Forecast Values', 'fontweight', 'bold');
legend('ANN-Measured', 'STFS-Measured')
set(gcf,'Color',[1,1,1])
%------------------

% RMSE FOR YEAR ------------------------------------------------------
RMS_ANN_1 = sqrt((sum((Mann_6-Measured_6).^2))/(le/6));
RMS_POL_1 = sqrt((sum((Mpol2_6-Measured_6).^2))/(le/6));
AAE_ANN_1 = (sum(abs(Mann_6-Measured_6)))/(le/6);
AAE_POL_1 = (sum(abs(Mpol2_6-Measured_6)))/(le/6);
AAE_POL_BIAS_CORRECTED = (sum(abs(Mpol2_6 - Measured_6 +
Bias)))/(le/6);%***************
RMS_POL_BIAS_CORRECTED = sqrt((sum((Mpol2_6 - Measured_6 +
Bias).^2))/(le/6));%*********
%--------------------------------------------------------------------
%Central frequency computing area ----------------------------------
Atotal = 0;
Aover = 0;
for q = 1:size(Mann_6)
    if (abs(Mann_6(q)-Measured_6(q))<=0.15)
        Atotal = Atotal +1;
    else
    Aover = Aover+1;
    end
end
ANN_CF = 100*(Atotal/(Atotal+Aover));
%-----------------------
Ptotal = 0;
Pover = 0;
for q = 1:size(Mpol2_6)
    if (abs(Mpol2_6(q)-Measured_6(q))<=0.15)
        Ptotal = Ptotal +1;
    else
    Pover = Pover+1;
    end
end
POL_CF = 100*(Ptotal/(Ptotal+Pover));
%--------------------------------------------------------------------
%--------------------------------------------------------------------
% RNSE FOR STORM

%find events greater than peak
i = find(Measured_6>PEAK);

%loop through each event and calculate RMS's
for j = 1:length(i)
```

```
    ts2_SS = ts2(i(j)-6:1:i(j)+6);

    Mann_SS = Mann_6(i(j)-6:1:i(j)+6);

    Mpol2_SS = Mpol2_6(i(j)-6:1:i(j)+6);
    BiasPol_SS = (Mpol2_6(i(j)-6:1:i(j)+6)) + (Bias(i(j)-
6:1:i(j)+6));%*****************
    Measured_SS = Measured_6(i(j)-6:1:i(j)+6);

    AAE_ANN_SS(j) = (sum(abs(Mann_SS-Measured_SS)))/13;
    AAE_POL_SS(j) = (sum(abs(Mpol2_SS-Measured_SS)))/13;
    AAE_BiasPol_SS(j) =(sum(abs(BiasPol_SS-
Measured_SS)))/13;%**************************
    RMS_ANN_SS(j) = sqrt((sum((Mann_SS-Measured_SS).^2))/13);
    RMS_POL_SS(j) = sqrt((sum((Mpol2_SS-Measured_SS).^2))/13);
    RMS_BiasPol_SS(j) =sqrt((sum((BiasPol_SS-
Measured_SS).^2))/13);%**********************
    %--------------PLOT STORM +36 and -36 HOURS-------------
    subplot(2,1,1)
    plot(ts2_SS, Measured_SS, 'dm', 'markerfacecolor', 'k');
    %-----------------------------------------------------
end

subplot(2,1,1)
plot(ts2(i), Measured_6(i), 'ok', 'markerfacecolor', 'k');
%-----------------------------------------------------------------
clc;
fprintf('\n\n\nOne-Hour Forecast - Yearly Stastistics');
fprintf('\nANN RMSE for 1-year = %5.4f',RMS_ANN_1);
fprintf('\nPOL RMSE for 1-year = %5.4f',RMS_POL_1);
fprintf('\nPOLB RMSE for 1-year =
%5.4f',RMS_POL_BIAS_CORRECTED);%**********************
fprintf('\nANN AAE for 1-year = %5.4f',AAE_ANN_1);
fprintf('\nPOL AAE for 1-year = %5.4f',AAE_POL_1);
fprintf('\nPOLB AAE for 1-year =
%5.4f',AAE_POL_BIAS_CORRECTED);%**********************
fprintf('\nANN CF%%(15) for 1-year = %5.2f%%',ANN_CF);
fprintf('\nPOL CF%%(15) for 1-year = %5.2f%%',POL_CF);
fprintf('\n\nStorm Statistics');
fprintf('\nNumber of events found > %2.2f meters = %d', PEAK, length(i));
for j = 1:length(i)
fprintf('\nStorm #%d RMSE for ANN = %5.4f  POL = %5.4f  POLBIAS = %5.4f
Storm Peak of %5.3fm @ %d
Hours',j,RMS_ANN_SS(j),RMS_POL_SS(j),RMS_BiasPol_SS(j),Measured_6(i(j)),i
(j)*6-5);
end
for j = 1:length(i)
fprintf('\nStorm #%d AAE for ANN = %5.4f  POL = %5.4f  POLBIAS = %5.4f
Storm Peak of %5.3fm @ %d
Hours',j,AAE_ANN_SS(j),AAE_POL_SS(j),AAE_BiasPol_SS(j),Measured_6(i(j)),i
(j)*6-5);
end
```

Closure dates					
Year	Month	Day	Year	Month	Day
1983	2	1	2000	12	16
1985	12	26	2001	1	10
1987	3	29	2001	1	11
1988	12	24	2001	2	8
1990	2	20	2001	2	8
1990	2	27	2001	2	9
1990	2	28	2001	2	9
1990	3	2	2001	2	10
1990	9	19	2001	2	11
1990	10	7	2001	2	12
1992	10	25	2001	3	10
1993	1	11	2001	3	11
1993	1	25	2001	3	12
1993	2	21	2001	2	13
1993	10	13	2001	4	10
1993	10	14	2001	9	17
1993	10	15	2002	1	31
1993	10	16	2002	3	1
1993	11	14	2002	4	27
1993	12	13	2002	11	7
1994	1	28	2003	1	1
1995	1	1	2003	1	2
1995	2	1	2003	1	3
1995	2	1	2003	1	3
1995	2	2	2003	1	4
1995	12	23	2003	1	4
1996	2	19	2003	1	5
1996	2	19	2003	1	5
1996	2	20	2003	1	6
1996	10	29	2003	1	6
1998	2	28	2003	1	7
1998	10	8	2003	1	7
1998	11	6	2003	1	8
1999	12	3	2003	1	8
1999	12	25	2003	1	21
1999	12	26	2003	1	22
1999	12	26	2003	1	22
1999	12	27	2003	1	23
1999	12	27	2003	1	23
2000	11	13	2004	2	22
2000	12	11	2004	11	13
2000	12	12	2005	2	13
2000	12	13	2005	14	2
2000	12	14	2005	4	8
2000	12	14	2005	12	16
2000	12	15	2005	12	17
2000	12	15	2006	2	28
2000	12	16			

Development of Artificial Neural Networks for Storm Surge Prediction in the North Sea

Daniel B. Prouty[1], Philippe Tissot[2], and Arif Anwar[3]

Abstract:

A model based on Artificial Neural Networks (ANNs) was developed to predict storm surge magnitudes and arrival times at selected locations in the North Sea. The model predicts storm surges based on past measured water levels using one or more tidal stations along the North Sea. The work presented focuses on the application and performance of the model at the Sheerness tide station near the entrance of the River Thames in the United Kingdom. Model performance is analyzed for single station input as well as with the addition of inputs from other stations. These secondary stations are located at baseline distances of 337 to 945 km from the Sheerness Station. The selection of an optimal secondary station location depends primarily on the forecast interval and correlates with the storm propagation time between the secondary and primary station. The ANN model performance is analyzed on a yearly basis as well as based on a set of individual storms to better understand the performance and the potential of the model during extreme events. The absolute average errors for a 3-hour prediction at Sheerness using Immingham as a secondary station range from 84 to 102 mm for the years 1990 – 2002 and the Central Frequencies of 150 mm range from 79.1% to 89.9%. For the same years the absolute average error for a 12-hour prediction using North Shields as a secondary station range from 110 to 143 mm and the Central Frequencies of 150 mm range from 68.1% to 75.9%. While training times for an ANN can be substantial, model application can be virtually instantaneous; making ANNs well suited for real-time emergency management situations.

CE Database keywords: Water levels; North Sea; Storm propagation; Storm surges; Artificial neural networks; Prediction

[1]University of Southampton, SO17 1BJ, United Kingdom. E-mail: Prouty@soton.ac.uk

[2]Texas A&M University – Corpus Christi, Corpus Christi, Texas 78412, E-mail: ptissot@lighthouse.tamucc.edu

[3]Lecturer, School of Civil Engineering and the Environment, University of Southampton, SO17 1BJ, United Kingdom. E-mail: A.A.Anwar@soton.ac.uk

Background

Storm surge events are meteorologically induced water level changes, and are defined as the difference between measured water level and tidally predicted water levels (Komar 1998, Pugh 2004). These surge events are caused by regional differences in barometric pressure and associated wind shear on the water surface. Every year numerous storms enter the North Sea and effect Western Europe including the United Kingdom (Komar 1998). Rising sea levels, regional subsidence, and increasing storminess are all increasing the risk and impact from storm surge events. The continuing development of accurate predictive models is important for the safety of growing coastal communities and navigation. This paper investigates the use of Artificial Neural Networks (ANNs) as a tool to predict storm surge propagation in the North Sea.

In the United Kingdom, several methods for predicting water levels are being used today. Tidal analysis is computed by harmonic analysis of previous water level records and is therefore primarily based on periodic astronomical forcings (Schureman, 1941). The method works well for large portions of the year but meteorological effects can significantly influence water levels and can introduce substantial errors in tidal predictions. These additional major forcings on water levels are fundamentally different in their variability with tidal influences being periodic, while weather forcings are fast changing, and mostly aperiodic. To include these additional forcings the United Kingdom's Meteorological Office have developed finite element models (NTSLF, 2005). These model predictions provide substantial improvements over harmonic forecasts but also require large amounts of meteorological and oceanographic data, computer time and are updated only 4 times per day.

This paper investigates the potential of an alternate methodology based on Artificial Neural Networks (ANNs) to predict water levels and storm surges at the Sheerness Tide Station located at entrance of the River Thames. Sheerness is the primary tidal station used to help determine when to open or close the Thames Barrage, an important decision with a substantial economic impact. The model development takes advantage of the large set of observations archived by the National Tidal and Sea Level Facility (NTSLF) for stations along the UK coastline and storm propagation characteristics in the North Sea. Measurements from these tide stations provide information on the advancing storm. The model uses the non linear modeling capability of ANNs (Rumelhart et al. 1995) to predict future water levels at the target station. Finally this paper reports on the use of more complex ANNs to improve forecasting during the largest storms and studies the performance and robustness of the models as the size of the ANN hidden layer is increased.

Review of Relevant ANN Applications

An Artificial Neural Network (ANN) is an information processing method that is inspired by the way biological nervous systems, such as the human brain, process information (Rumelhart et al. 1995, Hecht-Nielsen 1989). The key element of this method is the structure of the information processing system. It is composed of a number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN can be configured for a specific application, such as pattern recognition or data classification. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. Artificial neural networks work in a similar way with the ANN parameters, weights and biases, being adjusted as part

of a learning process supervised or unsupervised (Hagan et al. 1996). In the past 15 years the number of ANN applications has expanded to include environmental, financial and engineering problems. The use of neural networks for time series forecasting has been studied and proven successful for a number of cases. Chakraborty et al. (1992) used non-linear modeling of multivariate time series to predict future prices which consistently outperformed statistical models. Their work was shown to be quite useful in solving other problems in the fields of dynamical system modeling, recognition, prediction and control (Tang et al. 1991).

Neural networks have been applied successfully to a number of coastal and riverine cases, such as the forecasting of physical or water quality parameters (Brion and Lingireddy 1999), modeling of coastal algal blooms (Recknagel et al. 1997), and estimation of daily pH levels of rivers (Moatar et al. 1999). Neural networks are also increasingly used for the forecasting of flooding along rivers (Kim and Barros 2001) and have been incorporated into multi-model data fusion techniques for hydrological forecasting (See and Abrahart 2001).

For water level modeling, Tsai and Lee (1999) used neural networks to predict hourly tidal levels over long durations (up to one year) using very short term (one day) hourly tidal record for training. Results are interesting but the lead time for each forecast is only one hour and has limited applicability for forecasting purposes (Kumar et el. 2001). This type of model may be well suited for the initial operation of new or temporary water level monitoring stations. In Australia, Makarynskyy et al. (2004) computed total water levels including both tidal and meteorological components as output from a single ANN. This approach required large ANN structures to interpret the combined astronomical and meteorological signals. The method may be desirable for cases when existing historical data is

too limited to generate accurate tidal coefficients but require complex ANN structures. Such models can be time consuming to train, and are susceptible to poor generalization, especially when training relies on relatively small data sets. Researchers in the Baltic Sea, (Sztobryn 2003) were looking to improve upon existing numerical models for the forecasting of sea levels. While their existing models performed well during average conditions, performance was poor during storm events. Four different ANN models were tested using 24 hours of previous sea level values as well as 6 hour forecasts of wind speed and direction. The ANN models provided significant improvement over previous models during storms. Cox et al. (2002) modeled the tidal and weather induced water level components separately and showed that ANNs provided an effective approach to model the non-linear relationship between weather induced forcings and water levels. This work will follows the same methodology of Cox et al. (2002) and model tidally and weather induced water level changes separately.

In a related application along the south shore of Long Island, New York Huang et al. (2003) developed a regional neural network to predict water levels at a temporary location based on water levels measured at permanent National Oceanic and Atmospheric Administration (NOAA) tidal stations located about 60 km-100 km away from the inlets. The model was developed to re-construct long-term historic water levels using remote temporary sea level measurement stations and provided very good results for both tidal and non-tidal historical water levels. The present work is following a similar approach but focuses on predictive modeling with an emphasis on storm events and deals with larger distances of 337 km to 945 km between project stations.

The structure and training methodology of ANNs for the modeling of water levels is also a significant concern of this work. Researchers (Rajasekaran et al., 2005) compared the accuracy of sequential learning neural networks and functional networks for tide prediction along the coast of Taiwan based on short historical data sets (30 days). Sequential learning neural networks allow the researcher to determine the optimum number of hidden neurons as part of the training process by adding neurons sequentially one at a time and minimizing the error. The addition of neurons continues until a minimum error is reached. For both methodologies (sequential and functional networks) the predicted tidal levels correlated well with the measured tidal levels. Lee (2004) also showed that small ANN structures, using 1 neuron in the hidden layer had the best performance for the forecast of water levels. Similarly Tissot et al. (2003) showed that very simple ANN structures consisting of one hidden layer with one or two hidden neurons, led to significant improvements over harmonic forecasts for intervals up to 48 hours in the Gulf of Mexico. These studies also showed that the choice of transfer functions did not lead to significant changes in performance (Tissot et al. 2003) and showed promising results for the prediction of water levels during tropical storms (Tissot et al., 2004). This project uses small ANN structures as described by Tissot et al. (2004) and Rajasekaran et al. (2005), and further explores the potential of more complex ANN structures to model large storm events. Also, the tidal regimes of the North Sea are very different from locations examined in other works, including the Gulf of Mexico and the Taiwan Strait. As described in subsequent sections the characteristics of storm propagation are also unique to the North Sea. The models described in this work were designed to take advantage of the local dynamics.

Storm Propagation and Storm Surge Forecasting in the UK

Most large storms affecting the North Sea originate in the North Atlantic Ocean and move southward affecting locations along the coastlines of both the UK and Continental Europe. As these low-pressure systems move, the associated storm surge (water level difference between the measured and harmonically predicted water level) moves with them. These surges can cause large scale flooding and numerous casualties, as in the 1953 storm (Met Office, 2003) in the United Kingdom and Europe. After the floods of 1953, the UK government set up a committee, led by Lord Waverly to investigate and report on the event. The Committee recommended that a national flood warning system be established. This resulted in the creation of the Storm Tide Warning Service, today known as the Storm Tide Forecasting Service (STFS). The Tide Gauge Inspectorate at the Proudman Oceanographic Laboratories (POL) United Kingdom is responsible for maintaining the network of gauges, and archiving the data (NTSLF, 2005).

A comparison of harmonically predicted water levels, and measured water levels are presented in Fig. 1 for a 1993 storm event recorded at the Sheerness Station. The surge component or the difference between measured and harmonically predicted water levels is also presented at the bottom of Fig. 1. A storm surge peak value of 3000 mm occurred approximately 3 hours before the high tide, resulting in an overall maximum water level of only 1500 mm above peak harmonic water levels for the tidal cycle.

Fig. 2 Illustrates the progression of a December 1990 storm and its associated storm surge recorded at four tide stations: Wick, North Shields, Immingham and Sheerness. At Wick (the most northerly station) the peak storm surge is 300 mm above predicted harmonic water level. As the storm moves southward, the maximum surge levels at North Shields and Immingham occur

respectively 6 hours and 8 hours later, and 1200 mm and 1700 mm above the harmonic predicted water level. The storm surge reaches the Sheerness station approximately 16 hours later with a peak elevation of 2100 mm above predicted harmonic water level. The growing size of the storm surge can be explained by a combination of the North Sea bathymetry relative to the storm path, and meteorological influences. When moving southward, the relative depth of the North Sea shallows, and its width (East/West) narrows. As a consequence, storm surge height increases to compensate for the decreased volume of the basin (this is also exacerbated by deepening low pressure from a strengthening storm). Storm surge propagation along the coast is illustrated for several other cases in Fig. 3. These figures were created by determining the peak water levels at each station, then subtracting the peak water level height at Wick (giving the relative change in storm surge elevation). As can be seen from Fig. 3a, the time lag in storm surge, as well as the sizes of the storm surges at each location (Fig. 3b), are relatively consistent between storms. The development of the ANN models is in large part premised on taking advantage of the relatively consistent storm characteristics and time delay between the storm surges at the northern stations as compared to the surge at the targeted Sheerness station.

Artificial Neural Network Modeling of Storm Surges

To better understand the behavior and potential of ANN models for storm surge predictions, three suites of experiments were conducted. The first suite of experiments dealt with the selection of a training year. Tests were run to determine the affect on performance of an ANN when changing the size and selection of data sets used for training. The second suite of experiments is concerned with determining the optimal design of a neural network for the

prediction of water levels and the performance of the resulting model as compared to other standard models. The third suite of experiments focuses on ANN performance and in particular how varying the structure of an ANN impact upon the robustness and storm performance of the model. Finally an engineering example is presented, demonstrating ANN application using optimal configurations found in the previous experiments.

For all experiments, the ANN models were developed, trained, and tested within the MATLAB R13 computational environment and the related Neural Network Toolbox (The MathWorks, Inc., 2002). The computer used was a 3.0 GHz Pentium 4 PC running Windows XP. The models were trained using the Levenberg-Marquardt back-propagation algorithm as implemented within the MATLAB Neural Network Toolbox. The training was conducted by selecting at least one full year of water levels and by assembling input vectors, each consisting of a time series of previous storm surge levels, i.e. the difference between water level measurements and harmonically predicted water levels, from one or more tide stations. Weights and biases were randomly assigned at the beginning of each training session and their values were updated during each iteration such that the error between ANN output and target (or observed) is progressively minimized. Training times varied between a few minutes and several hours depending on the size of the ANN. Although training times can be lengthy, it should be emphasized that for real-time applications, generating water level forecasts is a sub-second process. Once the ANN models are trained, they are ideally suited for streamed forecasting (an automatically generated, real-time forecast based on streaming data).

The performances of the models in this work are assessed based on criteria used by NOAA for the development and implementation of operational nowcast and forecast systems (NOAA, 1999). A single forecasting error, or e_i is defined as the difference between the predicted value p_i and the observed value, r_i or $e_i = p_i - r_i$. The models are assessed by averaging the individual errors over the full data sets, often one year of water level measurements and forecasts. The statistical parameters used to evaluate the models performance for this paper are the Average Absolute Error (AAE) between predictions and measurements and the Central Frequency (CF) of 150 mm. The Central Frequency (CF) is the percentage of predictions that are within 150 mm of the measured water levels. The 150 mm selected for the Central Frequency (CF) measure is the requirement typically used by NOAA and is based on NOAA's estimates of pilots' needs for under keel clearance value (NOAA 1999). In the UK the Storm Tide Forecasting Service (STFS) model performance is measured by a similar method, but is called a "skill measure". This is similar to the CF described by NOAA, but uses a less strict value of 200 mm. This paper will use the NOAA AAE and CF (150 mm) for performance analysis.

The overall impact of the storms and the differences in storm impact at the different study stations can be observed in an analysis of the yearly performance of the tide tables presented in Table 1. In Table 1, the AAE of the harmonic prediction is compared with the measured water levels and the CF (150 mm) for the tide tables with respect to the measured water levels. The lower AAE at Wick is due to its proximity to deep water compared to Sheerness (Pugh, 2004). The high AAE at Sheerness provides a large target for the ANN model to improve upon. Table 1 also shows the completeness and availability of water level data for

the project stations. While the data is of excellent quality with less than 3% missing for most years, gaps must be removed or filled prior to ANN modeling as ANNs typically have difficulties handling gaps (Agapkin et al. 2003). The gaps were filled by adding to the harmonic prediction a linearly interpolated surge component of the water levels. The linear interpolation was computed based on the surge component of the water level on each side of the gap.

Tissot et al. (2001) have suggested tidal and weather components should be modeled separately, the former using harmonic models, the later using ANN. This work follows that suggested principle. For this work, the computation of the harmonic component was based on the United States National Ocean Service (NOS) model. The computations followed the National Oceanic and Atmospheric Administration (NOAA) standards as described by Schureman (1941). The harmonic analysis was implemented with Harman/Harmpred (Mostella et al., 2002), a web based harmonic analysis program. The data sets were obtained on-line through the National Tidal and Sea Level Facility (NTSLF), with the original data collected by the National Tide Gauge Network, supported by the Proudman Oceanographic Laboratory (POL).

Experiment Suite 1: Model performance varying training data

The first part of Experiment Suite 1, Experiment 1.1, concentrates on how the training year is selected and on the impact of the training set length on ANN performance. In Experiment 1.2, input to the ANN models was constructed by using data from a single primary station, while varying the number of previous water level measurements used for input from 1 to 48 hours. For Experiment 1.3 additional previous water levels from a secondary station were added. The numbers of previous water level measurements used at both the primary and

secondary station were varied from 1 to 48 hours. The focus of Experiment 1.4 was to explore the changes in model performance while selecting secondary stations at different locations along the UK North Sea coastline.

Experiment 1.1 Selection of training year

A schematic of a typical ANN model used for this study is presented in Fig. 4. The structure used for the first suite of experiments is a two layer ANN, using one output neuron, one hidden neuron. The optimum ANN topology including number of hidden neurons will be discussed as part of the second suite of experiments. A full year was selected as the minimum data set length for the training of the ANNs to include seasonal variations of water levels. Testing was performed to evaluate the importance of the selection of a particular training year including the possible impact of year to year variability in the frequency and magnitude of storms. Testing was also performed to assess potential improvements when including longer training periods. A series of basic [1,1] models making 3-hour predictions were successively trained, using each year of water levels at the Sheerness station from 1990 to 1999 with an input consisting of 48 hours of previous water levels. For each of the training year the model was then tested on all other years. The early stopping method used to prevent over-training in Experiment Suite 1 is to stop training when a preset error level is reached (1 mm). The results are displayed in Table 2. For each training year, the average absolute errors over all testing years were computed. These errors were all within a 9 mm range, varying from 146 mm (1996) to 155 mm (1991). For all test cases the 1997 training year led to the lowest AAE.

Experiment 1.2 Varying training data set lengths

The lengths of the training sets were varied from 1 year (1990), 3 years (1990-1992), 5 years (1990-1994), and 10 years (1990-1999). All ANNs were trained

using 48 hours of previous water levels. Models were then tested on the 2001 data set. Although the AAE decreased from 153 mm (1 year) to 146 mm (10 years), virtually the same average absolute error of 146 mm for year 2001 can also be obtained by training on the year 1997 data set alone. A substantial disadvantage of using a large training set is that computational time increases from 20 minutes (one-year data set) to 20 hours (10 year data set) for the computer used for this study. Based on these results, training was conducted using a 1 year data set (1997), for the remainder of the study.

Experiment Suite 2: Model performance varying input parameters

Experiment 2.1 Single-Station ANN

The ANN was first trained using Sheerness as the only (or primary) station. The forecast time periods used were: 3, 6, 12 and 24 hours. The number of previous water levels included during training varied, using 1, 3, 6, 12, 24, 36, and 48 hours for each forecast period. No secondary station was used in this experiment. Fig. 5 shows changes in ANN performance for a 1x1 ANN trained for various forecasting times. The models were trained on the 1997 data set, and tested on the 1999 data set.

Model improvement is the largest for the 3-hour forecasts. When increasing the number of previous water level measurements in the input from 1 to 24 hours, the AAE over the testing set decreases from 170 mm to 109 mm an improvement of 61 mm. The improvements are modest for 6 and 12-hour forecasts with a performance improvement of 39 mm and 18 mm respectively. For the 24-hour forecasts, only 13 mm of improvement is observed. In all cases, very little improvement is found when including more than 24 hours of previous water level measurements from the primary station, indicating that little additional information is included in this “older data”.

Experiment 2.2 Two-Station ANN

For this experiment, additional data from a second water level measuring station is included in the ANN training set. This secondary station, is located north of the primary station (a direction towards most approaching North Sea storms), and provides the ANN with additional information of a storm's presence before it can be measured at the primary station. The ANNs were trained on the 1997 data set, tested using the 1999 data set.

In this experiment the number of previous measurements used for the primary station (Sheerness) is maintained constant at 24-hours, while the number of previous measurements from the secondary station (Immingham), and is incremented from 1 to 24 hours. Results are presented in Fig. 6. For short forecast intervals (3-6 hours) the AAE decreases significantly when increasing the length of the secondary station input time series. In comparison the best performance of the previous ANNs (from Experiment 2.1) using only one station as input, with the 3-hour forecasts using two stations, the AAE decreases from 115 mm to 84 mm (with 24 hours of previous water level measurements from both stations), and the 6-hour forecasts improving from 141 mm to 96 mm. The 12-hour and 24-hour forecasts show little improvement when including information from Immingham as a secondary station. This result is not surprising given that storm surge propagation time from the Immingham to Sheerness is approximately 5 to 7 hours. For a 12-hour and 24-hour forecast, the storm surge has yet to reach the Immingham station and therefore the data from Immingham is of little help for the forecast of the upcoming Sheerness storm surge.

Experiment 2.3 Two-station ANN, varying secondary location

The performance of the ANN is analyzed by varying the location of the secondary station. Three secondary stations Immingham, North Shields and Wick are

selected. The stations are respectively 337 km, 510 km and 945 km north of Sheerness. The results are shown in Fig. 7 for each forecast interval. For each test, 24 hours of secondary station data was used, and the number of measurements used from the primary station varied from 1 to 48 hours. For a 3-hour or 6-hour forecast at Sheerness, Immingham (337 km north) was the best choice for a secondary station (Fig. 7a and 7b), as it produced the lowest AAE. For a 12-hour forecast at Sheerness North Shields (510 km North) was the best choice for a secondary station (Fig. 7c). For a 24-hour forecast, Wick (945 km North) performed the best, although only marginally better than using no secondary station (Fig 7d).

The effectiveness of a secondary station location is determined by its proximity to the primary station. The selection of the secondary station should be based on the forecast interval and its distance from the primary station. This effective range varies with the forecast interval. For example, in Fig. 7c when using Immingham for a 12-hour forecast, the AAE is 150 mm, this error drops to 130 mm when switching the secondary station to North Shields. For a secondary station to be effective in predicting a storm surge event, it must provide some information about the storms existence at the time of the forecast. With the addition of the storm's information (a recorded increase in surge height) from the secondary station, this secondary data set can then be used by the ANN to help predict the surge height and arrival time at the primary station. The optimal secondary station location is one which its distance from the primary is the same as what a typical storm would travel during the forecast interval.

Experiment Suite 3: ANN topology and performance

In this suite of experiments we investigate how model performance is affected by changes in ANN topology. While the performance is tracked for both yearly averages and during storms, the discussions will focus primarily on the storm performance because of its importance for safety and commerce. Experiment 3.1 varies the ANN structure size, by changing the number of hidden neurons used. Experiment 3.2 tests the robustness (or repeatability) of the ANN model. For all experiments in Suite 3, Sheerness is used as the primary station, and Immingham as the secondary station. The training year used is 1997, testing year 1993, and validation year 1999. 24-hours of previous water level measurements are used for training for each station. During initial runs of Experiment 3, models that used more than 5 hidden neurons experienced overtraining problems, causing large variations in predictions. Because of this, verification vectors were used for all Suite 3 experiments.

Experiment 3.1: ANN performance varying structure size

In this experiment, the variability of ANN models with different number of hidden neurons was determined. 3-hour ANN forecasts were calculated for each of the following size models: 1,2,5,10,20, and 50 hidden neurons. Fig 8 showing ANN's performance varying the number of hidden neurons used. The increase in the number of hidden neurons leads to a better maximum storm surge prediction than a [1,1] ANN structure (1 hidden neuron, 1 output neuron). However the use of 10 or more hidden neurons also leads to increased noise and variability in the predictions, primarily during storm events

Experiment 3.2: ANN performance variability.

The tests performed for experiment 3.1 were repeated 20 times with random starting values for the weights and biases. Fig. 9 shows the results of four ANN test runs using the same initial starting parameters, and illustrates visually, the variability of individual ANN forecasts. The variability arises from Matlab's Levenberg Marquart algorithm, which assigns random values when initializing weights and biases during training. This variability was found to only be significant when using larger ANNs (greater than 5 hidden neurons).

The results of these experiments show that ANN topology can have a considerable impact on storm performance. While very little change is observed for the yearly averages, significant differences are observed during short-term evaluation of storm events. The performance of different size ANNs are further illustrated in Fig. 10. Small ANNs have the best average performances, but overall the best performances are reached for individual instances of large ANNs. However the repeatability and variability of large ANN predictions during storms is a concern. The error bars in the figure 10 illustrate the range of AAE and CF obtained for the 20 runs for each case using 1, 2, 5, 10, 20, and 50 hidden neurons. The average performance decreases (increase in the average AAE and decrease in the average CF(150mm)) when increasing the number of hidden neurons. The performance range of the predictions also increases significantly with the number of hidden neurons. Based on this experiment, selecting smaller ANNs and in particular [1,1] ANNs leads on to better average performance during the selected storm. However the best performance based on both AAE and CF was obtained for one of the [10,1] models and at least one of the implementations

of each of the larger models had a performance better or equivalent to the small [1,1] model.

Engineering Application - Model performance during storm events

The most important use of the ANN model in this project is the forecast of water levels and peak timing during a storm event. This section uses the best model developed in the previous sections to demonstrate an ANN application in storm forecasting. The largest storm event of the Sheerness data set for years 1990-2003 was selected as an example (Feb 21, 1993). Fig. 11 Shows comparison of measured water levels and the 3-hour ANN forecast. During the 72-hour storm window, the maximum residual of the ANN forecast was 1004 mm; which occurred approximately 3-hours before high tide. This was also the time of maximum storm surge which measured 3041 mm. The ANN forecast lagged behind the measured water levels during the flood portion of tidal-cycle and essentially matched the measured water levels during the ebb tide. As the storm surge peak approached, the ANN forecast soon reduced this error to only 72 mm during the water level maximum, and matched the storm's arrival time precisely.

Conclusion

The United Kingdom coastline provides a data-rich location for testing and training ANNs in the North Sea. Numerous long-term water level measuring stations are available on the North Sea coastline. This large data set enables ANNs to be trained and optimized for many different forecast intervals and locations. The UK's North Sea coastline is ideally suited for using ANN applications to predict storm surge heights and arrival times. Large North Sea storms events normally originate in the North Atlantic then turn into the North Sea and propagate southward. The long linear progression of tidal stations along the UK coastline allows data to be

gathered during many parts of a storm's lifetime. Typically, Large North Sea storms move parallel to or at low angles with the coastline. This combination of regional coastal orientation, typical North Sea storm tracks, and high density of tidal measuring stations allows the use of secondary station data to provide a significant improvement in short term forecasts (3-12 hours).

Tests using North Sea data indicate that increasing the size of the training data set from 1 to 10 years increases prediction accuracy, but can significantly increase training times. Tests also indicate that although using more previous water–level measurements when training ANNs increases the accuracy of forecasts, little improvement is found when using more than 24 hours. The addition of secondary stations can significantly increase forecast accuracy, but the selection of a secondary station is very important in optimization of an ANN forecast. Secondary stations must be selected depending on the forecast interval, and based on the range to the primary station. The optimal secondary station location is one whose distance is approximately the same distance from the primary station that a typical storm would traverse during the forecasting interval. Future work needs to be done to test the use of more than two stations.

When restricting analysis to an individual storm event (72-hours), this work shows that on average, small ANNs (1-2 neurons in the hidden layer) perform better than larger ANNs. The use of large ANNs (10 or more neurons in the hidden layer) can provide more accuracy (during an individual training event), but with an increase in variability. This variability can be significant, and on average, large networks do not perform as well as small networks. Ongoing work is testing techniques such as ensemble forecasting, where the predictive output of several

large ANNs are combined, and variability "averaged out" to better predict storm surge peak heights while reducing error overall.

ANNs can provide significant forecast accuracy at new installations where archived data is in the order of 1-year in size. This project has shown that little difference is found in which year is selected for training of an ANN. Although differences of 40 mm were found in performances for yearly averages of ANN's, any year can be used to obtain acceptable results. Remarkably, this includes events beyond the range of the training data set. For example, in this paper the ANN was able to forecast a 3 m storm surge (a 1/100 year storm event) within 0.1m after using a training year data set with a maximum storm surge of only 1.2m. Normally using standard statistical methods this would not be possible without a much larger data set. The ANN is able to (with the help of a secondary station) scale measurements beyond what it is trained on.

References

Agapkin, O. A., Orlov, Y. V., Persiantsev, I. G., and Dolenko, S. A. (2003). "Preprocessing ultrasonic scanning data with the help of Hopfield-style neural network." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 502(2-3), 520-522.

Brion, G. M., and Lingireddy, S. (1999). "A neural network approach to identifying non-point sources of microbial contamination." *Water Research*, 33(14), 3099-3106.

Chakraborty, K., Mehrota, K., Mohan, C. K., and Ranka, S. (1992). "Forecasting the Behavior of Multivariate Time Series Using Neural Networks." *Neural Networks*, 5, 961-970.

Cox, D. T., Tissot, P., and Michaud, P. (2002). "Water Level Observations and Short-Term Predictions Including Meteorological Events for Entrance of Galveston Bay, Texas." *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 128(1), 21-29.

Hagan M. T., H Bemurth, and M Beale, (1996), "Neural Network Design", Pws Publishing Co.

Hecht-Nielsen, R. (1989). *Neurocomputing*, Addison-Wesley Longman Publishing Co., Inc.

Huang, W., Murray, C., Kraus, N., and Rosati, J. (2003). "Development of a regional neural network for coastal water level predictions." *Ocean Engineering*, 30(17), 2275-2295.

Kim, G., and Barros, A. P. (2001). "Quantitative flood forecasting using multisensor data and neural networks." *Journal of Hydrology*, 246(1-4), 45-62.

Komar, P. D. (1998). *Beach processes and sedimentation* Prentice-Hall, Englewood Cliffs, N.J.

Kumar, A., Minocha, V. K., Mandal, S., Medina, J. R., Walton Jr, T. L., Garcia, A. W., Tsai, C.-P., and Lee, T.-L. (2001). "Back-Propagation Neural Network in Tidal-Level Forecasting." *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 127(1), 54-59.

Lee, T.-L. (2004). "Back-propagation neural network for long-term tidal predictions." *Ocean Engineering*, 31(2), 225-238.

Makarynskyy, O., Makarynska, D., Kuhn, M., and Featherstone, W. E. (2004). "Predicting sea level variations with artificial neural networks at Hillarys Boat Harbour, Western Australia." *Estuarine, Coastal and Shelf Science*, 61(2), 351-360.

MathWorks (2002). "Neural Network Toobox". The MathWorks Inc., Natick, Mass.

Met. Office, (2003). Meteorological Office, U.K. "Flood of memories." <http://www.metoffice.com/bookshelf/outlook/200301/53flood.html> (Nov. 8, 2005).

Moatar, F., Fessant, F., and Poirel, A. (1999). "pH modelling by neural networks. Application of control and validation data series in the Middle Loire river." *Ecological Modelling*, 120(2-3), 141-156.

Mostella, A., Duff, J. S., and P.R. Michaud, (2002). "Harman and Harmpred: Web-based Software to Generate Tidal Constituents and Tidal Forecasts for the Texas Coast", Proc. of 19th AMS Conf. on Weather Analysis and

Forecasting/15th AMS Conf. on Numerical Weather Prediction (San Antonio, Texas), August 2002.

NOAA (1999). "NOS Procedures for Developing and Implementing Operational Nowcast and Forecast Systems for PORTS." Center for Operational Oceanographic Products and Services, National Ocean Service (NOS) National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce.

NTSLF (2005). National Tidal & Sea Level Facility (United Kingdom) "U.K. Tide Gauge Network." <http://www.pol.ac.uk/ntslf/tgi/> (Nov. 8, 2005)

Pugh, D. (2004). *Changing Sea Levels: Effects or Tides, Weather and Climate*, Cambridge University Press, Cambridge.

Recknagel, F., French, M., Harkonen, P., and Yabunaka, K.-I. (1997). "Artificial neural network approach for modelling and prediction of algal blooms." *Ecological Modelling*, 96(1-3), 11-28.

Rajasekaran, S., Thiruvenkatasamy, K., and Lee, T.-L. (2005). "Tidal level forecasting using functional and sequential learning neural networks." *Applied Mathematical Modelling*, In Press, Corrected Proof.

Rumelhart, D. E., Durbin, R., Golden, R., and Chauvin, Y. (1995). *Backpropagation: Theory, Architectures, and Applications*, Hillsdale.

Rumelhart, D. E., Hintont, G. E., and Williams, R. J. (1986). "Learning representations by back-propagating errors." *Nature*, 323(6088), 533-536.

Schureman, P. (1941). *Manual of Harmonic Analysis and Prediction of Tides.* Special Publication No. 98. U.S. Coast and Geodetic Survey.

See, L., and Abrahart, R. J. (2001) "Multi-model data fusion for hydrological forecasting" *Computers & Geosciences* 27 (2001) 987-994.

Sztobryn, M. (2003). "Forecast of storm surge by means of artificial neural network." *Journal of Sea Research*, 49(4), 317-322.

Tang, Z., de Almeida, C., and Fishwick, P. A. (1991). "Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology." *Simulation*, 57(5), 303-310.

Tetko, I. V., Livingstone, D. J., and Luik, A. I. (1995). "Neural Network Studies. 1. Comparison of Overfitting and Overtraining." *J. Chem. Inf. Comput. Sci.*, 35(5), 826 - 833.

Tissot, P. E., Cox, D. T., and Michaud, P. (2001) "Neural network forecasting of storm surges along the Gulf of Mexico." *Proceedings of the Fourth International Symposium Waves 2001, Sep 2-6 2001*, San Francisco, CA, 1535-1544.

Tissot, P. E., Michaud, P. R., and Cox, D. T. (2003). "Optimization and Performance of a Neural Network Model Forcasting Water Levels for the Corpus Christi, Texas, Esturary." *3rd Conference on the Applications of Artificial Intelligence to Environmental Science, Long Beach, California, February 2003.*

Tissot, P. E., Cox D. T, Sadovski, A., Michaud, P. and Duff, S. (2004), "Performance and Comparison of Water Level Forecasting Models for the Texas Ports and Waterways" *Proceedings of the PORTS 2004 Conference, Houston, TX*, May 23-26, 2004.

Tsai, C.-P., and Lee, T.-L. (1999). "Back-propagation neural network in tidal-level forecasting." *Journal of Waterway, Port, Coastal and Ocean Engineering*, 125(4), 195-202.

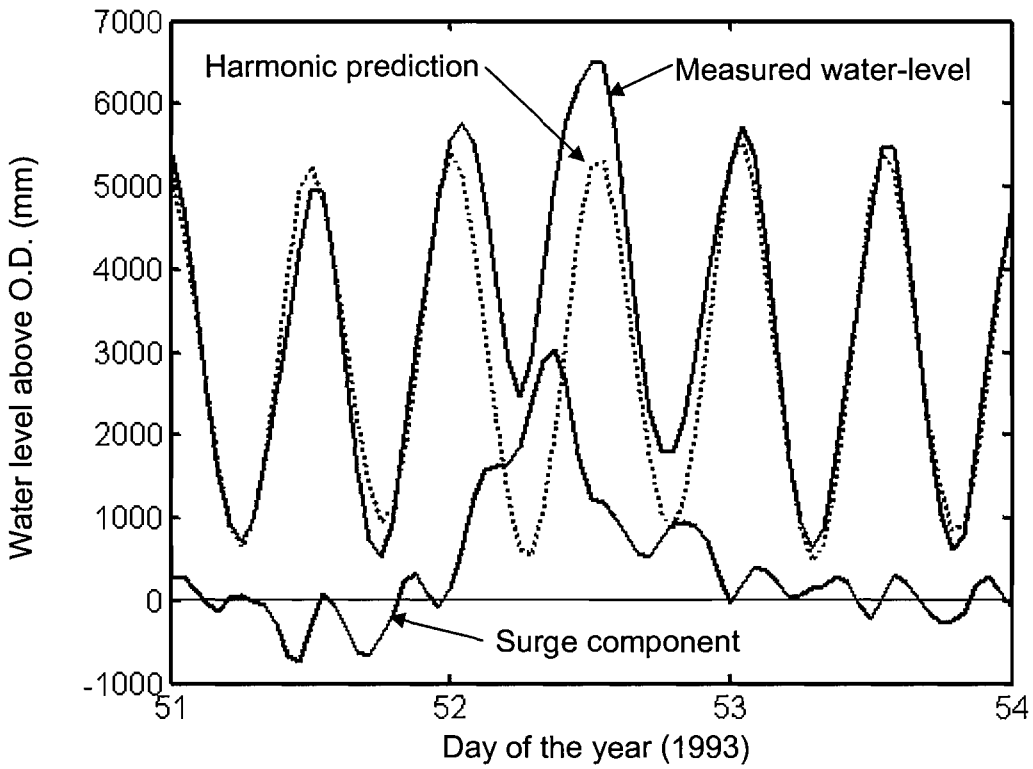


Fig. 1. Illustration of water level changes during a storm passage.

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

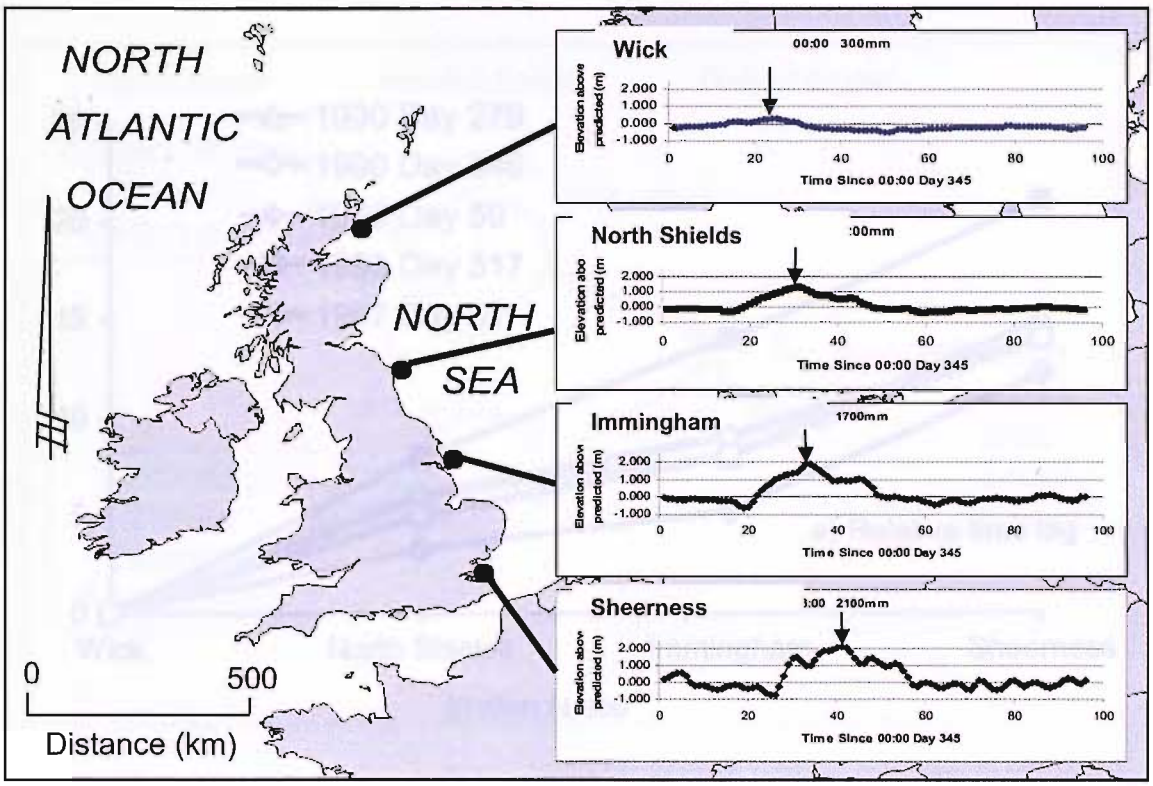


Fig. 2. Propagation of storm on December 11, 1990

TWO COLUMN FIGURE
CAPTIONS ARE LOCATED IN SEPARATE FILE

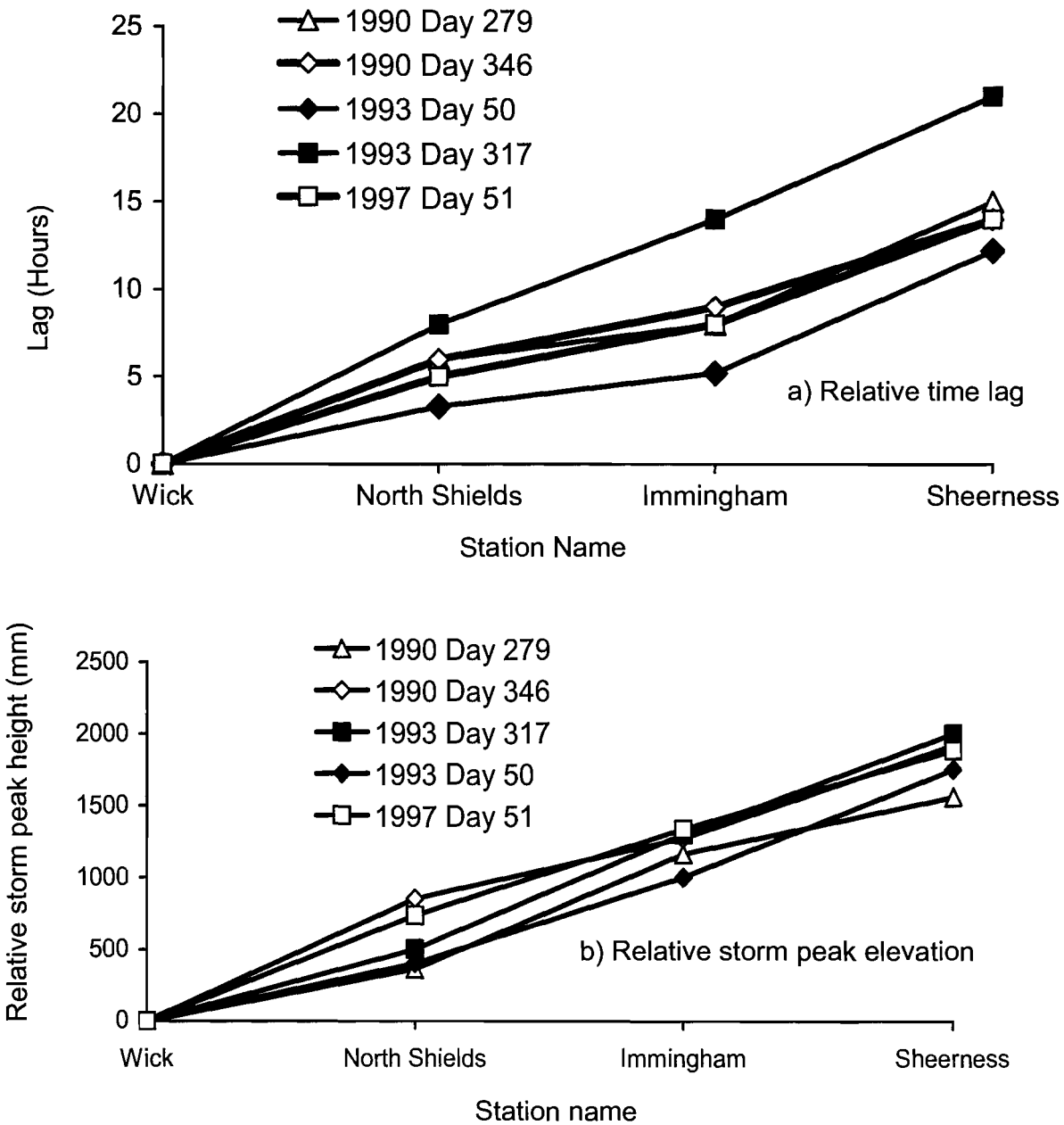


Fig. 3. Dynamics of storm surges between Wick and Sheerness

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

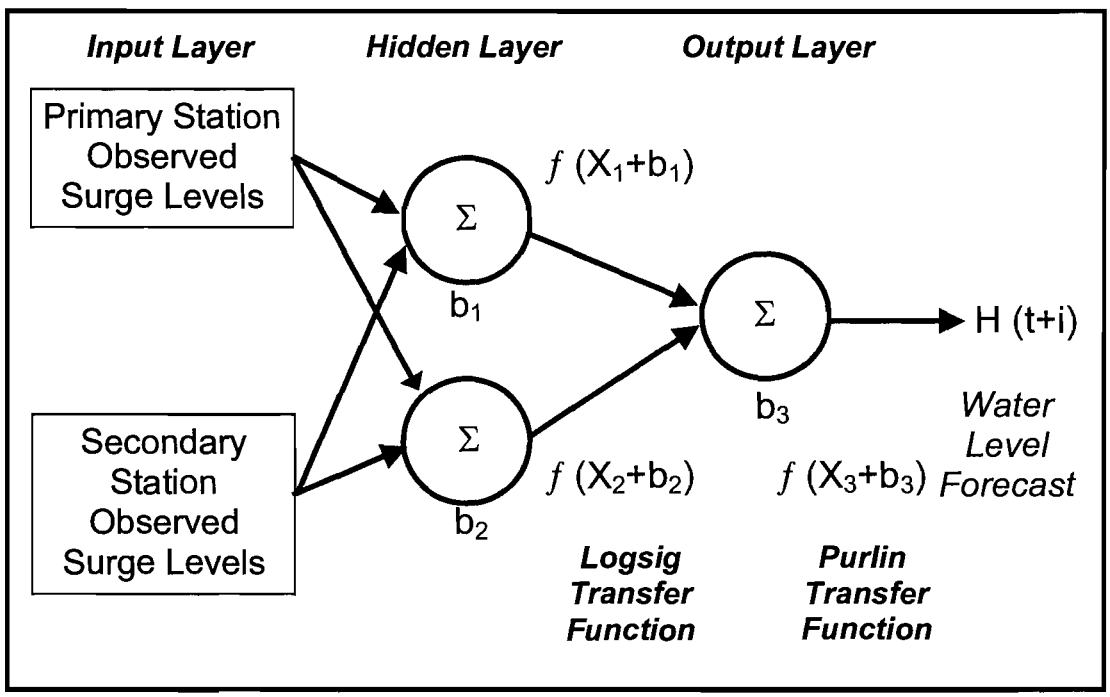


Fig. 4. ANN schematic using two neurons in the hidden layer.

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

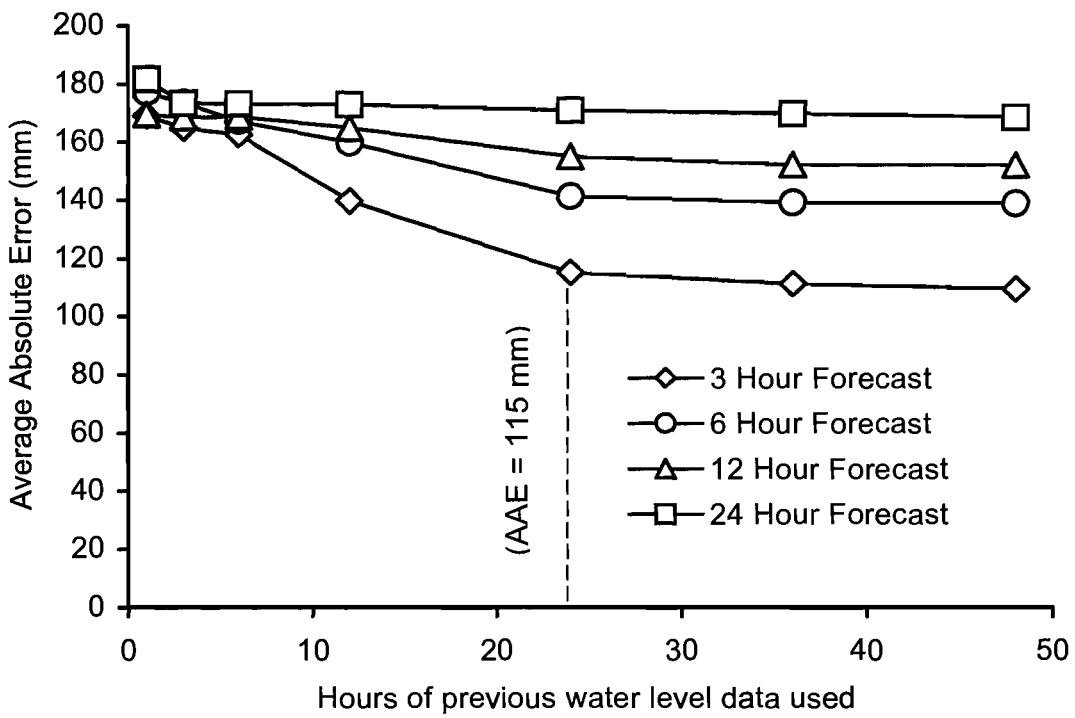


Fig. 5. Performance of a single station ANN model.

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

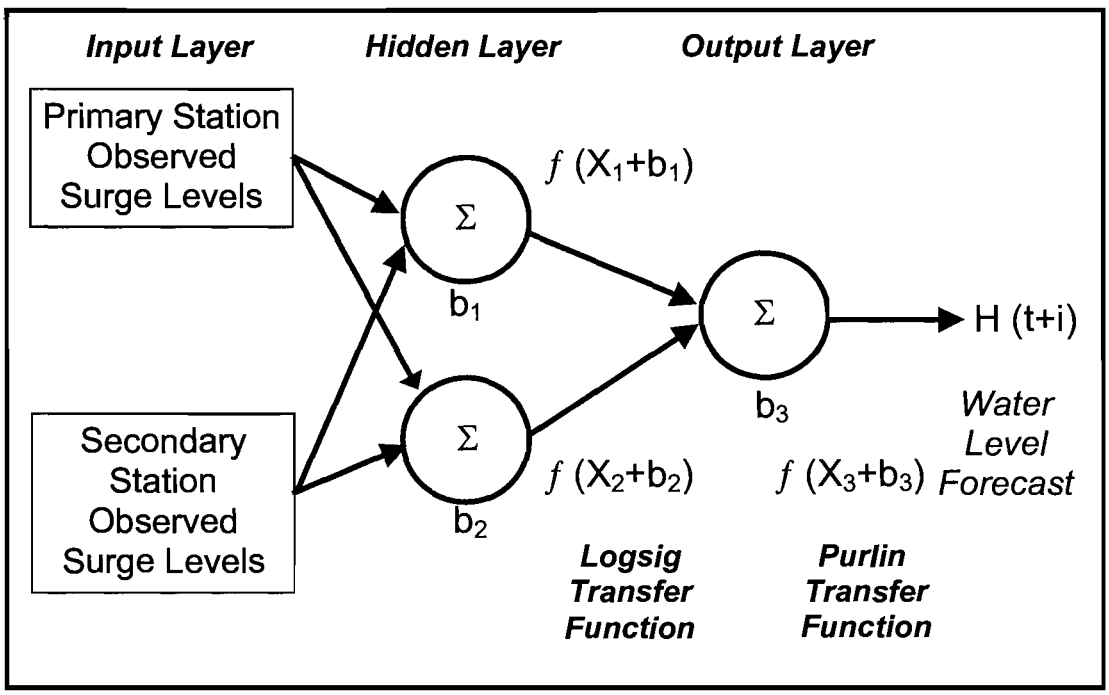


Fig. 4. ANN schematic using two neurons in the hidden layer.

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

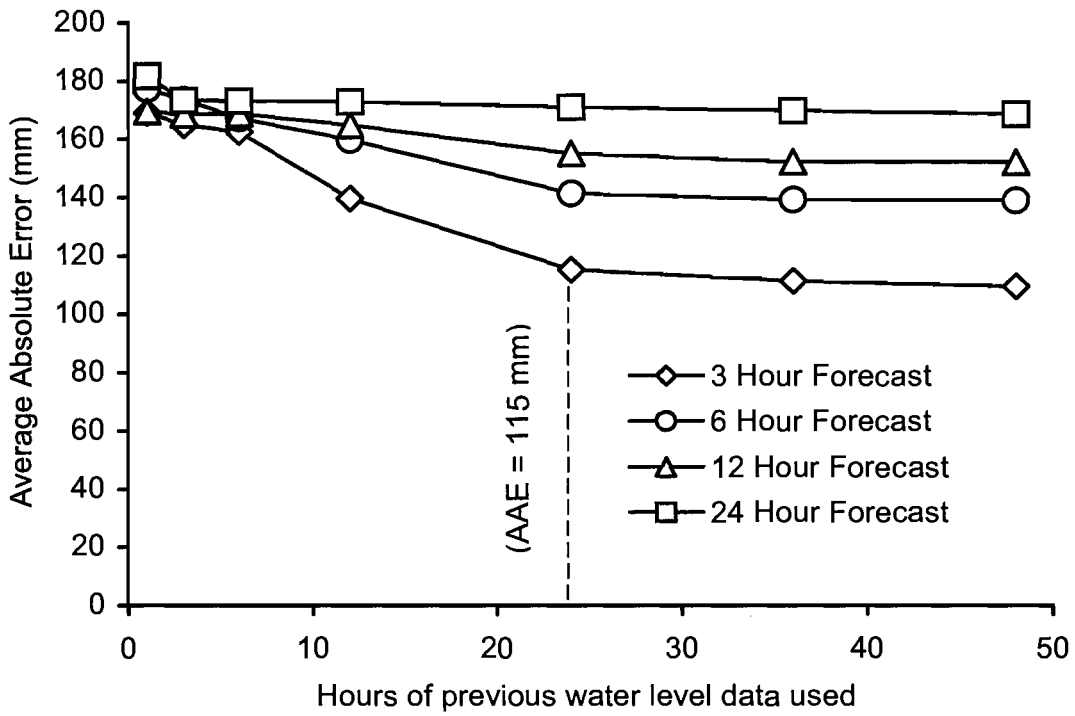


Fig. 5. Performance of a single station ANN model.

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

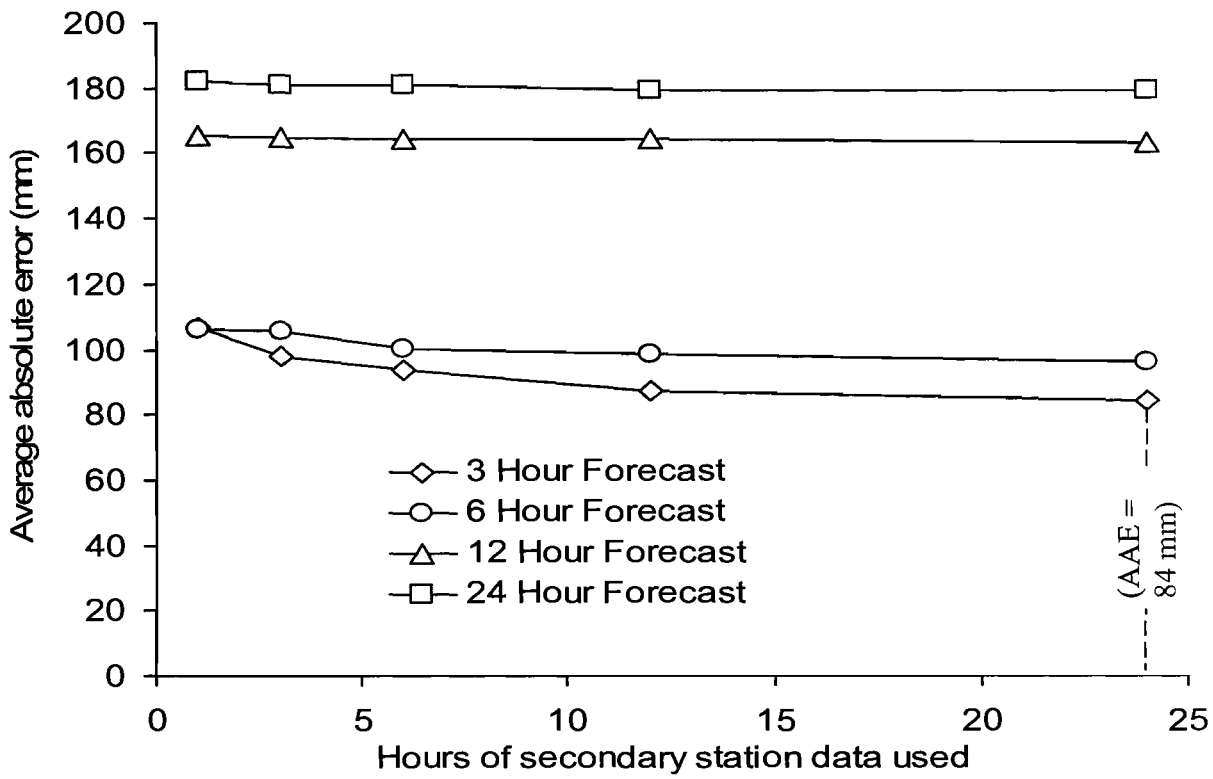


Fig. 6. Performance of a two-station ANN model - varying the secondary station input data.

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

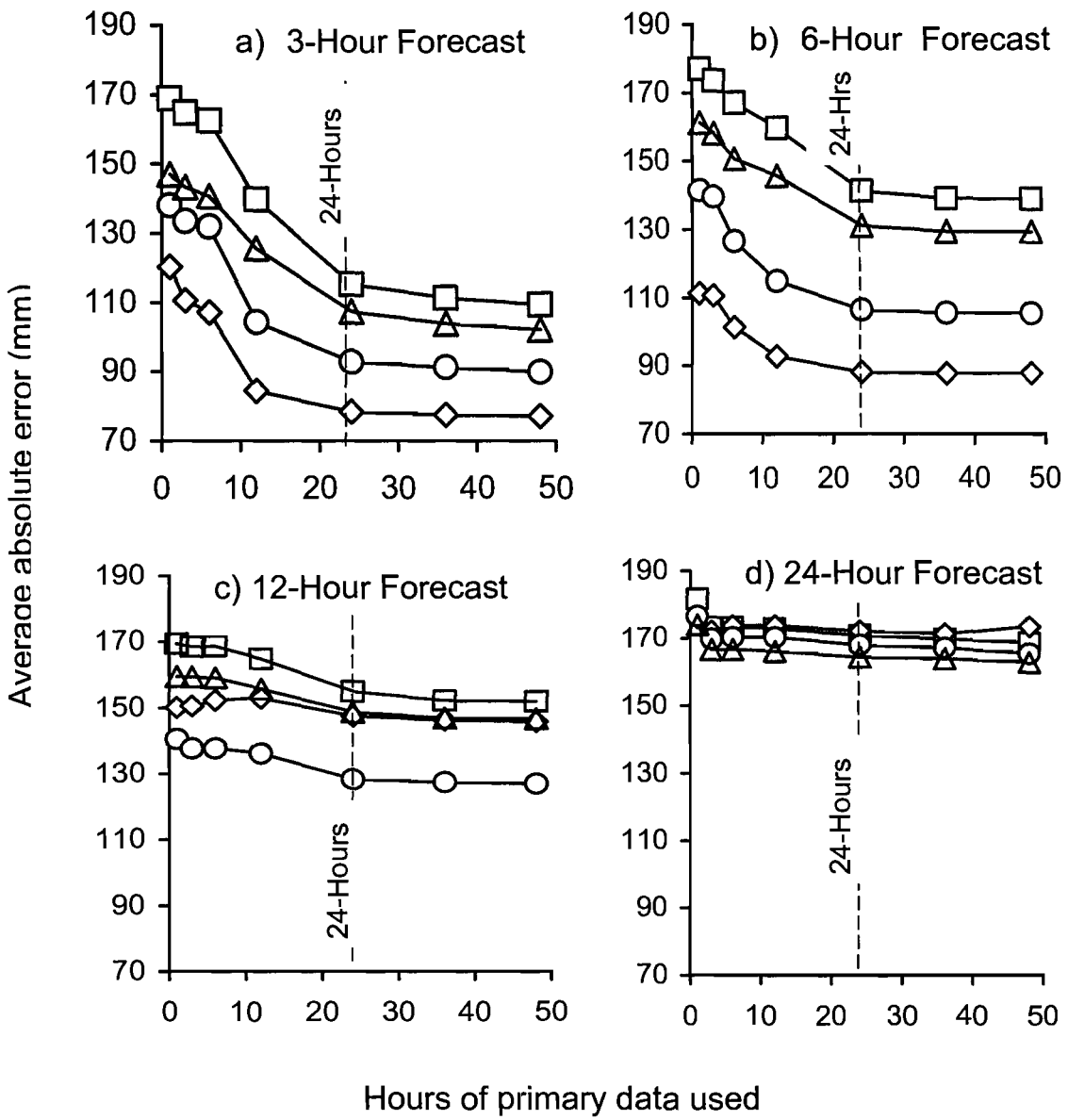


Fig. 7. Performance of a two-station ANN model - varying the secondary location. Square–No secondary station, Diamond–Immingham, Circle–North Shields, Triangle–Wick.

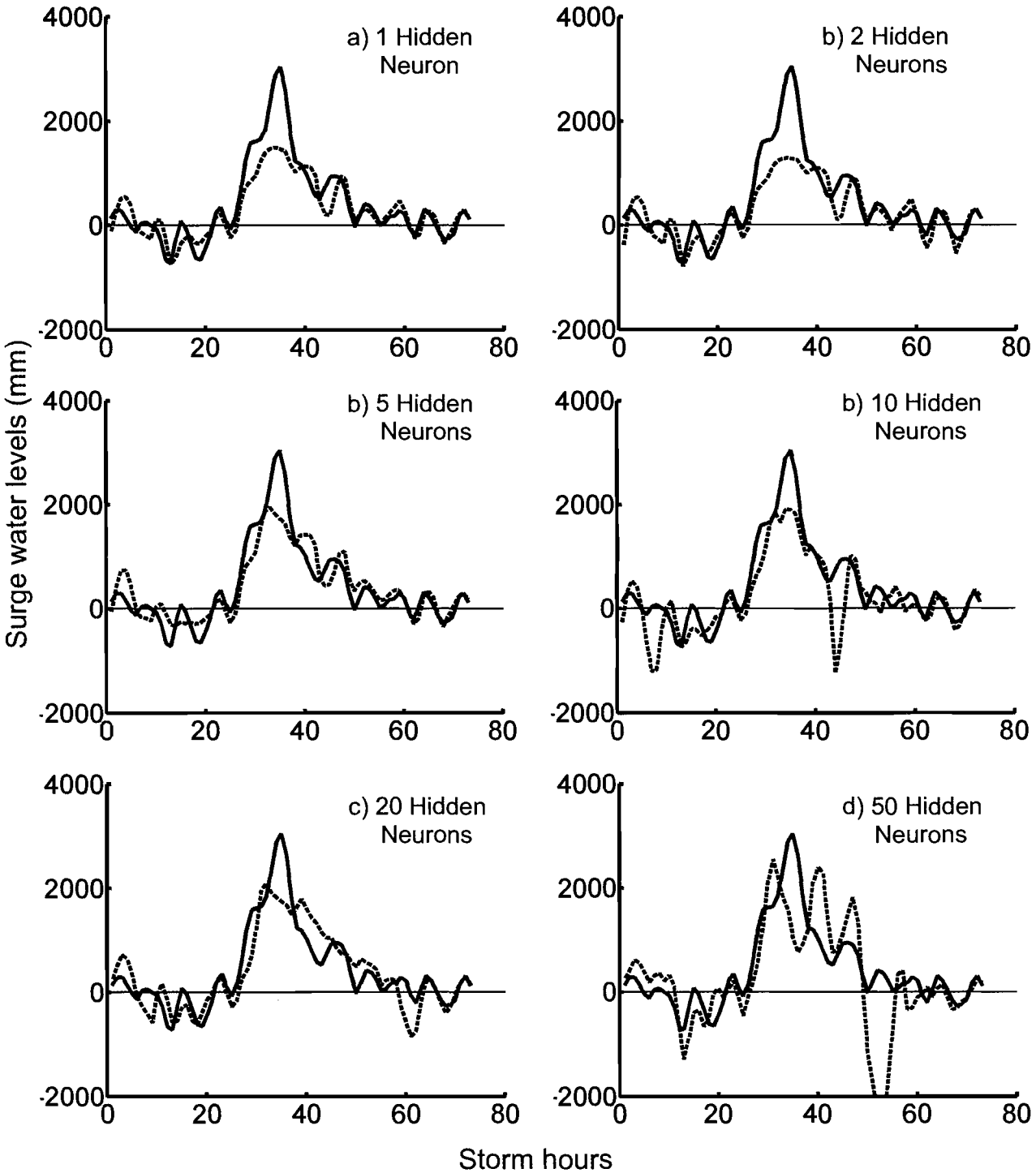


Fig. 8. Performance of ANN - varying the number of hidden neurons. Dashed line - ANN Model, Solid line - Measured surge-level.

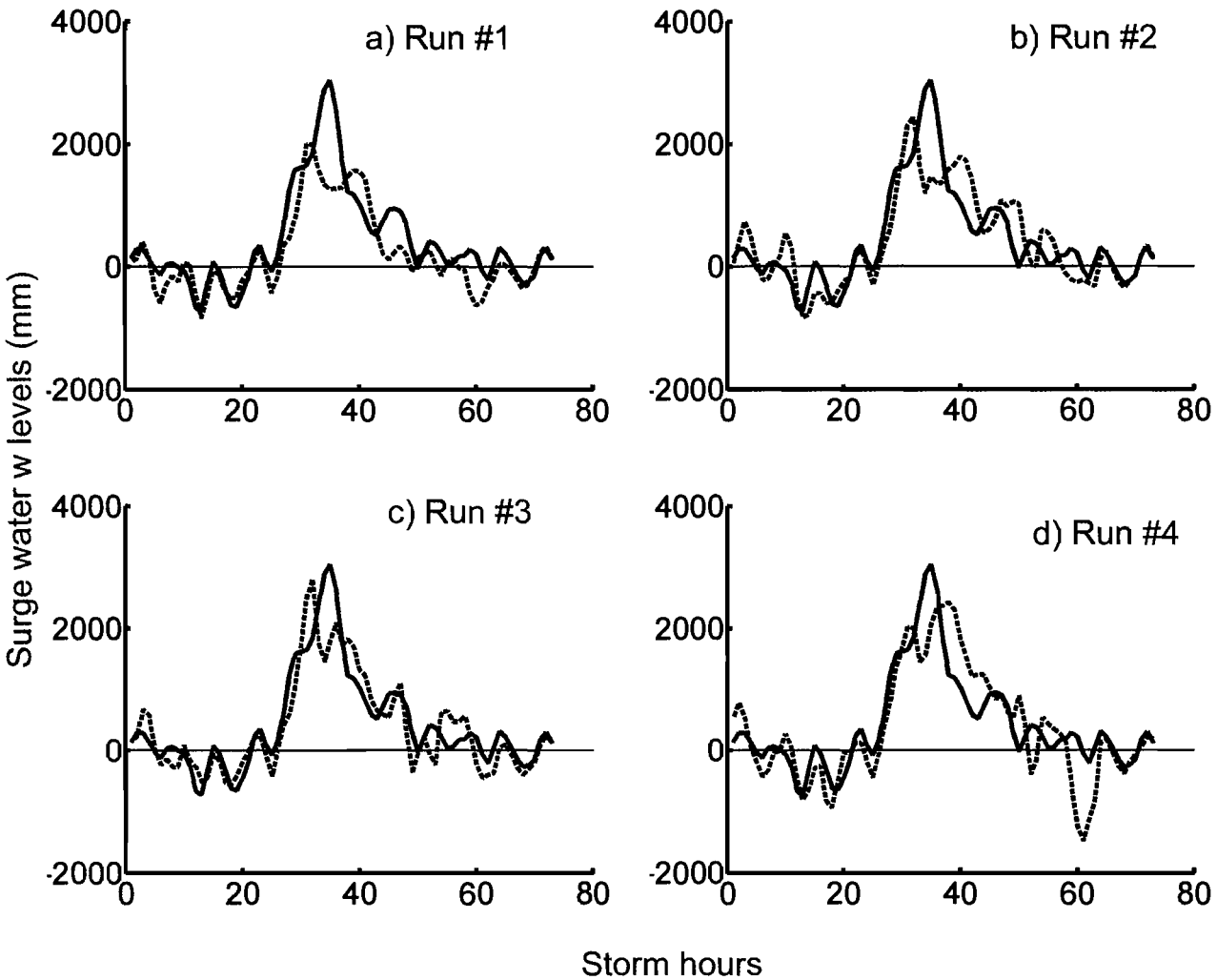


Fig. 9. Performance of ANN during multiple runs using the same ANN structure (20 hidden Neurons).

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

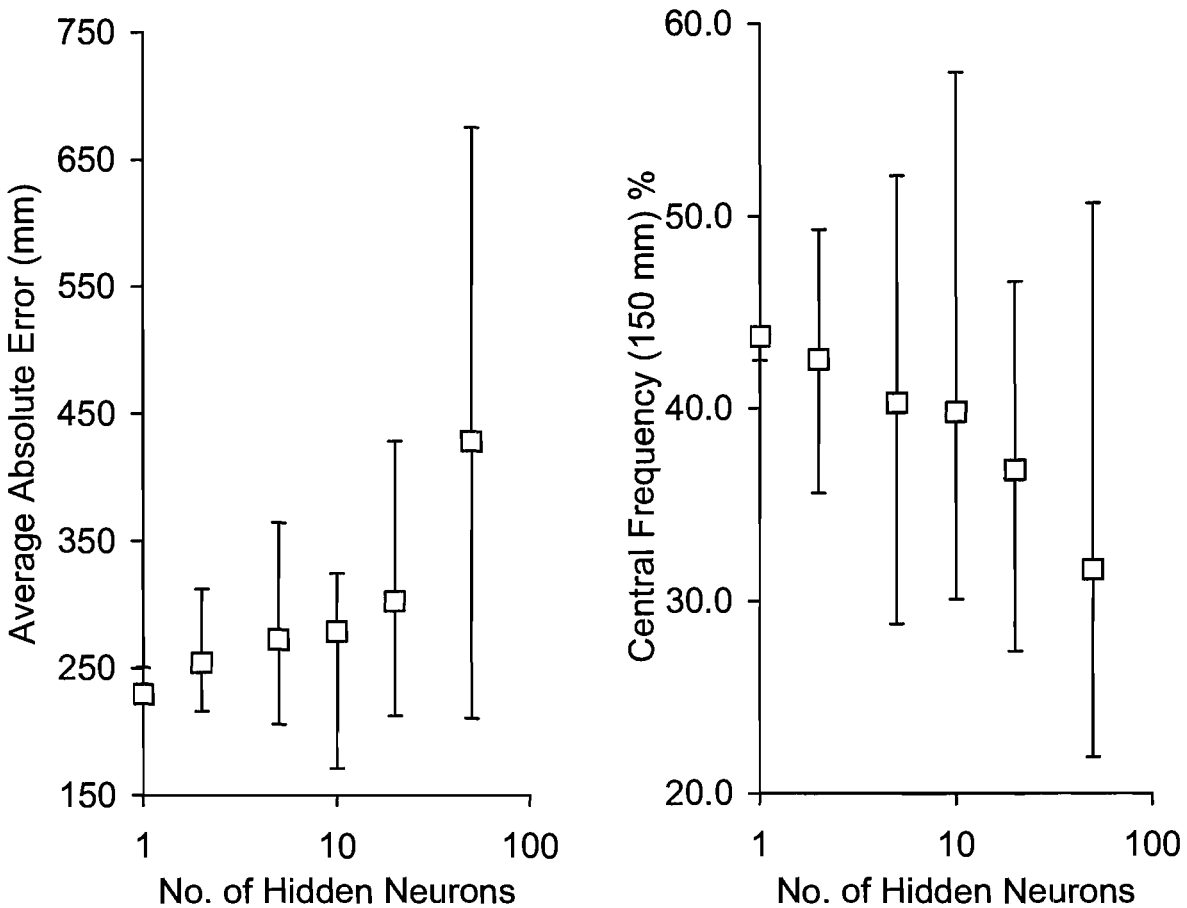


Fig. 10. ANN performance and variation during a storm event - Feb 6, 1990

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

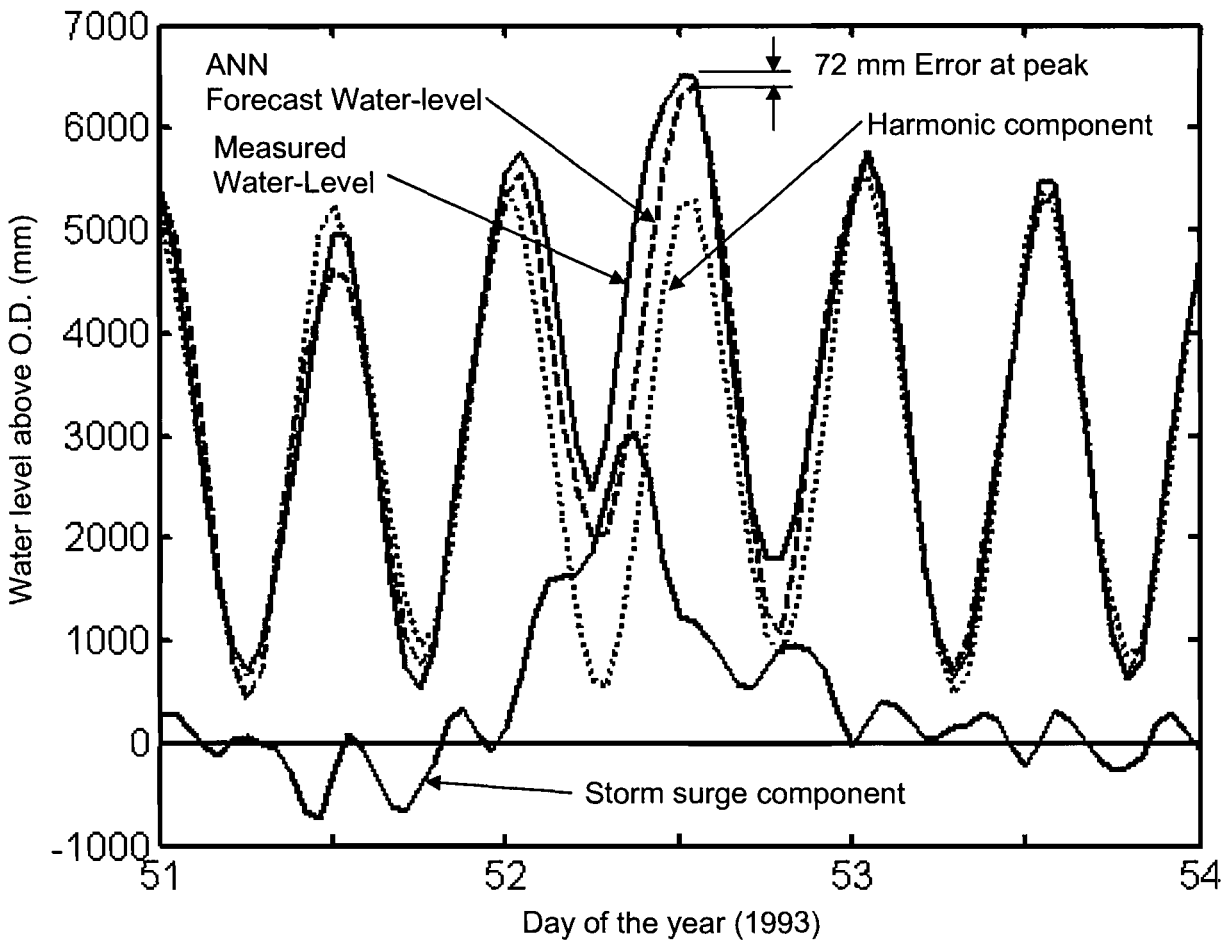


Fig. 11. ANN performance during a storm at the Sheerness Station (Feb 6 – Feb 10, 1993)

TO BE PHOTO-REDUCED TO 50% ORIGINAL SIZE
CAPTIONS ARE LOCATED IN SEPARATE FILE

Table 1. Tide station availability statistics and performance of harmonic predictions for years 1990 – 2000.

		Wick			North Shields			Immingham			Sheerness		
		Central Freq % (150 mm)	AAE (mm)	% Data Avail-able	Central Freq % (150 mm)	AAE (mm)	% Data Avail-able	Central Freq % (150 mm)	AAE (mm)	% Data Avail-able	Central Freq % (150 mm)	AAE (mm)	% Data Avail-able
Year	1990	59.9	154	99	67.8	135	95	59.3	167	100	49.6	205	100
	1991	61.7	145	100	61.0	145	92	57.7	163	100	46.9	202	100
	1992	66.6	132	98	68.8	124	100	59.6	154	100	51.0	189	100
	1993	66.4	130	100	52.8	154	78	59.1	167	98	51.4	198	100
	1994	68.2	123	99	71.4	120	93	59.7	157	100	52.4	188	99
	1995	76.3	100	77	70.5	124	100	59.3	162	98	53.0	194	77
	1996	65.9	127	100	71.9	118	100	56.7	163	87	52.3	186	98
	1997	67.6	127	99	71.5	119	100	62.7	148	99	54.8	174	98
	1998	67.7	124	97	75.3	110	87	58.9	165	94	52.0	190	100
	1999	65.2	133	98	72.5	119	100	64.3	148	97	53.3	188	98
	2000	66.0	136	98	70.2	126	98	61.0	156	94	51.4	196	95
	Average	66.5	130	97	68.5	127	95	59.8	159	97	51.6	192	97

PRINT 100% (2 COLUMNS)

Table 2. Average Absolute Error (mm)

		Training Year									
		1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Test Year	1990		170	165	169	166	167	172	166	172	176
	1991	159		150	149	148	150	148	150	162	170
	1992	156	151		153	150	152	153	150	159	163
	1993	164	160	158		154	156	158	154	163	168
	1994	156	154	152	149		148	149	148	155	160
	1995	157	159	155	153	152		154	152	157	161
	1996	149	142	143	139	137	138		137	145	151
	1997	142	143	139	137	135	136	134		139	143
	1998	157	162	159	153	152	152	150	148		148
	1999	159	168	164	158	156	156	155	152	149	
	Average	153	155	153	148	147	148	147	146	148	151
	Rank	8	10	9	6	3	4	2	1	5	7