

---

Dissertations, Theses, and Masters Projects

Theses, Dissertations, & Master Projects

---

2003

## ANN wave prediction model for winter storms and hurricanes

Jun-Young. Kim

*College of William and Mary - Virginia Institute of Marine Science*

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Artificial Intelligence and Robotics Commons](#), and the [Oceanography Commons](#)

---

### Recommended Citation

Kim, Jun-Young., "ANN wave prediction model for winter storms and hurricanes" (2003). *Dissertations, Theses, and Masters Projects*. William & Mary. Paper 1539616716.

<https://dx.doi.org/doi:10.25773/v5-t50q-nz52>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact [scholarworks@wm.edu](mailto:scholarworks@wm.edu).

ANN WAVE PREDICTION MODEL  
FOR WINTER STORMS AND HURRICANES

---

A Dissertation

Presented to

The Faculty of the School of Marine Science

The College of William and Mary in Virginia

In Partial Fulfillment

Of the Requirements for the Degree of Doctor of Philosophy

---

By

Jun-Young Kim

2003

APPROVAL SHEET

This dissertation is submitted in partial fulfillment of

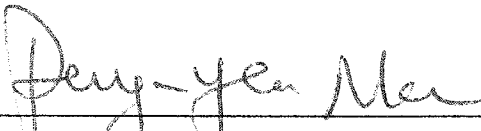
The requirements for the degree of

Doctor of Philosophy



Jun-Young Kim

Approved, December 2003



Jerome P. Y. Maa, Ph.D.  
Committee Chairman/Advisor



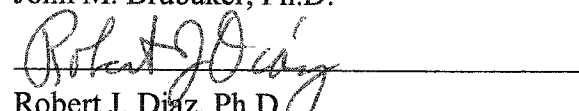
L. Donelson Wright, Ph.D.



Michael C. Newman, Ph.D.



John M. Brubaker, Ph.D.



Robert J. Diaz, Ph.D.



Bernard B. Hsieh, Ph.D.  
U.S. Army Eng Res. and Development Center  
Vicksburg, MS

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	vii
LIST OF TABLES.....	viii
LIST OF FIGURES.....	x
ABSTRACT.....	xxi
CHAPTER I. INTRODUCTION.....	1
1.1. Rationale.....	1
1.2. Problem of Current Numerical Models.....	6
1.3. What is ANN. Its Advantages and Limitations.....	8
1.4. Objectives.....	10
1.5. Outline of Presentation.....	11
CHAPTER II. REVIEW OF ANNs.....	13
2.1. Introduction.....	13
2.2. Historical Background.....	13
2.3. Basic Concept.....	14
2.4. Statistical Relationship.....	18
2.5. Classification of ANNs.....	19



(Continued)

	Page
2.5.1. Classification according to Learning Algorithm.....	19
2.5.2. Classification according to Time Delay.....	20
2.6. Learning Rule.....	26
2.6.1. Early Learning Rules.....	26
2.6.2. Gradient Descent Rule.....	27
2.6.2.1. Two-Layered Network (Input and Output Layers).....	28
2.6.2.2. Three-Layered Network (Input, Hidden, and Output Layers).....	32
2.6.2.3. Problem of Gradient Descent Rule.....	35
2.6.3. GDG.....	35
2.6.4. SCG.....	36
2.7. Neural System Operation.....	38
2.7.1. Pre-Processing.....	38
2.7.2. Selection of Initial Weights.....	39
2.7.3. Post-Processing.....	39
2.7.4. Comparison of Observed and Predicted Values.....	40
2.8. Selection of Optimum Numbers of Hidden Neurons and Iterations.....	44
2.8.1. Early Stopping.....	46
2.8.2. Bias and Variance.....	49

(Continued)

	Page
2.8.3. Trial and Error.....	49
<b>CHAPTER III. REVIEW OF WIND-WAVE MODELS.....</b>	<b>51</b>
3.1. Introduction.....	51
3.2. Monochromatic Waves.....	51
3.3. History of Wind-Wave Prediction Models.....	53
3.3.1. SMB.....	54
3.3.2. Spectrum Model.....	56
3.4. Currently Available Wind-Wave Models.....	58
3.4.1. WAM.....	58
3.4.2. SWAN.....	60
3.4.3. HISWA.....	61
3.4.4. STWAVE.....	62
3.4.5. WAVEWATCH.....	63
3.4.6. GLERL.....	65
3.5. Early ANN Wind-Wave Prediction Models.....	67
<b>CHAPTER IV. TWO TEST CASES FOR CHECKING ANN MODEL CAPABILITY.....</b>	<b>70</b>
4.1. Introduction.....	70
4.2. Linear Case.....	71
4.2.1. Proposed ANN Structure.....	73
4.2.2. Results and Discussion.....	74
4.3. Non-Linear Case (compare with SMB method).....	78

(Continued)

	Page
4.3.1. Method.....	80
4.3.1.1. Data Description.....	80
4.3.1.2. Proposed ANN Structure.....	83
4.3.2. Results and Discussion.....	86
4.4. Conclusions.....	96
CHAPTER V. ANN MODEL FOR WINTER-STORM WAVES.....	98
5.1. Introduction.....	98
5.2. Study Area.....	101
5.3. Data Description.....	104
5.3.1. Wind.....	104
5.3.2. Wave.....	108
5.4. Proposed ANN Structure.....	108
5.5. Time Delay.....	109
5.6. Selection of Learning Algorithm.....	111
5.7. Results of Significant Wave Height Prediction and Discussion .....	114
5.7.1. Three-Layered BPN and ERN.....	114
5.7.2. Three-Layered TDNN.....	126
5.7.3. Effects of Training Data Set Size.....	135
5.8. Results of Zero-Crossing Wave Period Prediction and Discussion.....	161
5.8.1. Three-Layered TDNN.....	161
5.8.2. Four-Layered TDNN.....	169

(Continued)

	Pages
5.9. Conclusions.....	176
<b>CHAPTER VI. ANN MODEL FOR HURRICANE WAVES.....</b>	<b>178</b>
6.1. Introduction.....	178
6.2. Data Description.....	184
6.2.1. Wind and Wave.....	184
6.2.2. Hurricane Track Pattern.....	186
6.3. Input.....	190
6.4. Proposed ANN Structure.....	192
6.5. Results of Significant Wave Height and Discussion.....	193
6.5.1. Effects of RMW as Input.....	193
6.5.2. Effects of Number of Training Data Point.....	200
6.5.3. Effects of Duration of Time Delay.....	208
6.6. Results of Peak Wave Period Prediction and Discussion.....	223
6.7. Conclusions.....	230
<b>CHAPTER VII. DISCUSSION AND CONCLUSION.....</b>	<b>231</b>
7.1. Convergence Rate.....	231
7.2. How Many Data Are Needed for Training?.....	238
7.3. Reason for an Excellent Simulation of the SMB Model.....	239
7.4. Possible Improvements on Wave Period Prediction.....	239
7.5. Possible Extension of ANN Prediction Results to Areas with No	

(Continued)

	Page
Data.....	240
7.6. Possible Future Improvements on ANN Wave Modeling.....	241
7.7. Application of ANN Wind-Wave Prediction Model to Other Places.....	242
7.8. Conclusions.....	243
APPENDICES.....	246
I. Source Code for the SMB Method .....	246
II. Source Code for an ANN Wind-Wave Prediction Model.....	252
LITERATURE CITED.....	253

## ACKNOWLEDGEMENTS

The support and guidance provided by my major advisor professor, Dr. Jerome P.-Y. Maa, throughout the course of this research are gratefully acknowledged. Definition of the topic and selection of appropriate methodologies were greatly facilitated by his willingness to assist and instruct. I also wish to thank my other Advisory Committee members, especially Dr. Bernard B. Hsieh and Michael C. Newman, for constructive reviews of this manuscript.

Successful completion of this project would not been possible without the forbearance, endurance, and unwarranted good humor of several of my fellow students. Notably, Robert Daniel and Jason Sea donated long hours assisting with sincere concerns. Finally, the VIMS Information, Technology and Networking Services are to be commended for their collective interest and assistance in the construction of my data set.

## LIST OF TABLES

Table		Page
4-1	Wind and wave data sets for non-linear test.....	82
4-2	Training and validation data sets.....	84
4-3	Mean Square Error (MSE) and correlation coefficient between reference and predicted wave heights at different number of hidden neurons.....	87
4-4	Mean Square Error (MSE) and correlation coefficient between reference and predicted wave periods at different number of hidden neurons for the prediction of wave height.....	89
4-5	Mean Square Error (MSE), correlation coefficient for the prediction of wave period.....	90
5-1	Information on the selected five wave stations.....	103
5-2	Summary of three selected data sets.....	106
5-3a	Bias, Variance, and Mean Square Error (MSE) for the Back-Propagation Network with different numbers of hidden neurons to predict wave height. The winter storm season of 1999 with 218 points was used for training, and the number of iterations was fixed as 20.....	117
5-3b	Bias, Variance, and Mean Square Error (MSE) for the Elman Recurrent Network with different numbers of hidden neurons to predict wave height. The 1999 winter storms with 218 points was used for training, and the number of iterations was fixed as 20.....	118
5-4	Bias, Variance, and Mean Square Error (MSE) for the Time Delay Neural Network with different numbers of hidden neurons to predict wave height. The 1999 winter storms with 218 data points was used for training, and the number of iterations was fixed as 40.....	132

(Continued)

Table		Page
5-5	Comparison of Mean Square Error (MSE) and computing time for the Back-Propagation Network (BPN), Elman Recurrent Network (ERN), and Time Delay Neural Network (TDNN).....	139
5-6	Bias, Variance and Mean Square Error (MSE) for the Time Delay Neural Network with different numbers of hidden neurons to predict wave height. The 1999 and 2001 winter storms with 403 data points were used for training, and the number of iterations was fixed as 25.....	142
5-7	Comparison of the Mean Square Error (MSE) and computing time for the Time Delay Neural Network when 218 data points for 1999 winter storms and 403 points for 1999 and 2001 winter storms were used.....	150
5-8	Correlation coefficient between the observed and predicted wave height at five wave stations for the Back-Propagation Network (BPN), Elman Recurrent Network (ERN), and Time Delay Neural Network (TDNN).....	151
5-9	Effects of global maximum and minimum wind on the performance of the Time Delay Neural Network with eight hidden neurons and 403 data points for training.....	155
5-10	Correlation coefficient between the observed and predicted zero-crossing wave period at the five wave stations for the four-layered Time Delay Neural Network (TDNN) with 403 training data points.....	168
6-1	Historical cyclonic events over the northwest Atlantic Ocean from 1988 to 2001.....	185
6-2	Maximum correlation coefficients between the observed and predicted wave heights, iterations, computing time, and Mean Square Error (MSE) when zero, six, 12, 18, 24-hour time-delays were used.....	213
6-3	Comparison of the correlation coefficient between the observed and predicted significant wave heights and peak wave periods at the five wave stations.....	215



## LIST OF FIGURES

Figure		Page
1-1	Comparison of observed (solid line) and NWW3 forecasted (dashed line) wave heights at station 42002 (25.17°N and 94.42°W) located in the Gulf of Mexico (after Robert and Christopher, 2002).....	5
1-2	A Typical structure of the Back-Propagation Neural Network with a symbol $I_m H_n O_p$ . Solid lines represent weights and dashed lines indicate biases.....	9
2-1a	The basic unit of biological neurons, which are interconnected each other: dendrites for accepting input signals, soma for processing the input, axon for directing the processed result to the output, and synapse for representing the final output to other neurons.....	16
2-1b	Structure of an Artificial Neural Network, simulating biological neurons. Each input, hidden, and output layer consists of different numbers of artificial neurons. Input layer is for accepting input, X, hidden layer is for processing input signals, and output layer is for representing network result.....	16
2-2	Structure of the Elman Recurrent Network (ERN) with the input, hidden, output layers, and additional context unit. For the prediction ( $O_1 \sim O_p$ ) at the next time level, the ERN will use an external input ( $I_1 \sim I_m$ ) as well as an internal input ( $n_1 \sim n_n$ ) from the context unit.....	22
2-3	An illustration of the Time Delay Neural Network (TDNN). Each input condition, $I_m$ , $I_{m-1}$ , and $I_{m-2}$ has four external durations of time delay at four different time levels (e.g., $t = m$ , $m-1$ , $m-2$ , and $m-3$ for $I_m$ , $t = m-1$ , $m-2$ , $m-3$ , and $m-4$ for $I_{m-1}$ , and $t = m-2$ , $m-3$ , $m-4$ , and $m-5$ for $I_{m-2}$ ). For the prediction of outputs at time level $t = m-1$ ( $O_{m-1}$ ), the TDNN uses $I_{m-1}$ . In order to predict the outputs at time level $t = m$ ( $O_m$ ), if the TDNN uses an external input $I_m$ and an internal input $n_{m-1}$ , this is	

(Continued)

Figure		Page
	called short-term memory. For the same purpose, if the TDNN uses an external input $I_m$ and internal inputs $n_{m-1}$ and $n_{m-2}$ , it is called long-term memory.....	24
2-4a	An illustration of a constant slope linear transfer function where the slope = -1. If bias is not used ( <i>e.g.</i> , $b = 0$ ), the linear function passes through the origin (0, 0) shown by the dashed line. If bias = 1, the linear function moved towards the solid line with the same slope.....	30
2-4b	An illustration of a sigmoid transfer function with a range between 0 and 1. A low initial bias makes a sharp sigmoid function (solid line), while a high initial bias makes the smooth sigmoid function very gently (dashed line).....	30
2-5	An illustration of the learning pathway of a three-layered Artificial Neural Network. $W$ = weight matrix between input and hidden layers, $U$ = weight matrix between hidden and output layers, $F_j$ = transfer function between input and hidden layers, $F_k$ = transfer function between hidden and output layers, $b_1$ and $b_2$ = bias between input and hidden layers, and between hidden and output layers, $e$ = difference between $d$ and $Y$ , $H$ = output at hidden layer, $Y$ = output of $F_k(UH + b_2)$ between hidden and output layers, and $k$ = number of iteration. If the mean square error is larger than a predefined value, which is ideally zero, the ANN will repeat to update weights (notice dotted line) until it meets the criterion.....	42
2-6	An illustration of error paths for a training and validation data set. The output of the first half for the validation data set is within the output range of the first half for the training data set, while the rest 50 % output for the validation data set is out of range for the training data set. Pass 1 and 2 indicate the error path for the training and validation data set, respectively. For the first half output, pass1 = pass2 because they have the same output. However, for the second half output, pass1 and pass2 may have different errors.....	47

(Continued)

Figure		Page
4-1	Wind speeds and corresponding wave heights, which were generated by a linear equation with a constant slope.....	72
4-2	Wave height predictions without pre-processing for the BPN with structures of $I_1H_2O_1$ and $I_1H_{45}O_1$ . (a) For 25 iterations and (b) for 100 iterations, respectively.....	75
4-3	An experiment to demonstrate the need for global maximum and minimum wind speeds for pre-processing. The predicted maximum wave height should be 10 m for the maximum wind speed of 60 m/s.....	76
4-4	Wave height predictions using the global maximum wind speed of 50 m/s (a) and 70 m/s (b). The Back-Propagation Network structures of $I_1H_2O_1$ and $I_1H_{45}O_1$ were used at 25 iterations.....	77
4-5	An assumed location map with five wave stations. Wind fetch is ranked from high to low in order of station 1, 2, 3, 4, and 5. Station 5 has the least wind effects because it is located behind a headland.....	79
4-6	Simulated eight wind (a) and wave events (e.g., wave height $H_s$ and period T) (b). Wind speed was changed from 5 m/s to 27 m/s, and corresponding waves were designated from left to right as No. 1 to No. 8.....	81
4-7	Wave heights and periods at the five imaginary wave stations. Wave heights and periods from the SMB method were multiplied by factor 1 at station 1 (a), 0.8 at station 2 (b), 0.6 at station 3 (c), 0.5 at station 4 (d), 0.4 at station 5 (e), respectively, to simulate wind effects according to different wind fetch.....	85
4-8	Comparison of true and predicted wave heights using the Time Delay Neural Network with a structure of $I_{18}H_{15}O_5$ at 700 iterations when four training events were used for training.....	91

(Continued)

Figure		Page
4-9	Comparison of true and predicted wave periods using the Time Delay Neural Network with a structure of $I_{18}H_{15}O_5$ at 600 iterations when four training events were used for training.....	92
4-10	Correlation coefficients between reference (true) and predictions when four training events were used. (a) For wave height and (b) for wave period. The arrow indicates that predicted wave periods are obviously less than true period compared with wave heights.....	93
4-11	Correlation coefficients of weight matrix among three data sets. (a) For wave height and (b) for wave period.....	95
4-12	Relationship between wave heights and periods. (a) shows simulated waves produced by the SMB method. Wave heights simply increase with wave periods. (b) shows observed wave heights and periods at five stations (44007, 44009, 44025, 44013, and 41009) in February 1998, 1999, and 2001. In general, observed wave heights increase as wave periods increase. However, many large wave heights occurred at small wave periods. This indicates that the relationship between wave height and period is more complicated in real wind-wave generation.....	97
5-1	Storm-wind field over the northwest Atlantic Ocean. (a) On February 24, 00:00 GWT, 1998 and (b) on February 26, 00:00 GWT, 1998. Circles indicate the five wave stations along the east coast of the U.S.....	99
5-2	Observed significant wave height at the five wave stations (44007, 44009, 44025, 44013, and 41009) along the East Coast of the U.S. from February 22 to 28, 1998.....	100
5-3	Location map of 40 wind stations (squares) and the five wave station (circles). Wind stations are regrouped as three types: (1) 15 single-point stations, which represent the wind information in each grid; (2) 19 middle-size ( $2.5^{\circ}E \times 2^{\circ}N$ ) stations, which represent the average of nine single point stations; (3) Six large-size ( $5^{\circ}E \times 4^{\circ}N$ ) stations, which represent the average of 25 single-point stations.....	102

(Continued)

Figure		Page
5-4	Comparison of Mean Square Error between the Gradient Descent with a Variable Learning Rate and Momentum (GDX) and the Scaled Conjugate Gradient (SCG) for the Back-Propagation Network (BPN) and Elman Recurrent Network (ERN) with a structures of $I_{80}H_{10-130}O_5$ . The number of training data was 218 points.....	113
5-5	Effects of the number of hidden neurons and iterations on Mean Square Error (a) for the Back-Propagation Network and (b) for the Elman Recurrent Network. The winter storm season of 1999 with 218 data points was used.....	115
5-6	Effects of the number of hidden neurons on the Mean Square Error. The number of training data and iterations were 218 and 20, respectively.....	119
5-7	Error gradient curves for finding the optimum number of iterations (a) for the Back-Propagation Network and (b) for the Elman Recurrent Network. Six hidden neurons and 1999 winter storm events with 218 data points were used.....	120
5-8	Correlation coefficient between the observed and predicted significant wave heights. The numbers in parentheses indicates the number of iterations.....	122
5-9	Comparison of the observed and predicted significant wave heights Using the Back-Propagation Network and Elman Recurrent Network with a structure of $I_{720}H_6O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed were 20 m/s and -20 m/s. The number of iterations was 20 for the BPN and 30 for the ERN, respectively.....	123
5-10	Correlation coefficient between the observed and predicted significant wave heights. (a) For the Back-Propagation Network with 20 iterations and (b) for the Elman Recurrent Network with 30 iterations.....	125

(Continued)

Figure		Page
5-11	Effects of the numbers of iterations and hidden neurons on Mean Square Error for the Time Delay Neural Network. The 1999 winter storm season with 218 data points was used.....	128
5-12	Error gradient curves for finding the optimum number of iterations for the TDNN. (a) For four hidden neurons, (b) five hidden neurons, (c) six hidden neurons, (d) seven hidden neurons, (e) eight hidden neurons, and (f) nine hidden neurons. Only the data from 1999 winter-storm season with 218 data points were used.....	129
5-13	Effects of the number of hidden neurons on the Mean Square Error (MSE) for the Time Delay Neural Network when 1999 winter storms with 218 data points were used for training.....	133
5-14	Comparison of the correlation coefficient between the observed and predicted significant wave heights for the Time delay Neural Network with different numbers of hidden neurons and iterations. The number on each line indicates the number of hidden neurons.....	134
5-15	Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of $I_{720}H_8O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed was $\pm 20$ m/s. The number of iterations was 40.....	136
5-16	Correlation coefficient between the observed and predicted significant wave heights for Time Delay Neural Network at the five wave stations when 1999-winter storm season with 218 data points was used for training.....	138
5-17	Effects of the numbers of hidden neurons and iterations on the Mean Square Error (MSE) for the Time Delay Neural Network. The total number of training data points was 403.....	141
5-18	Effects of the number of hidden neurons on the Mean Square Error (MSE) for the Time Delay Neural Network.....	143

(Continued)

Figure		Page
5-19	Comparison of the correlation coefficient between the observed and predicted significant wave heights for the Time delay Neural Network with different numbers of hidden neurons and iterations. The number on each line indicates the number of hidden neurons.....	144
5-20	Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of $I_{720}H_8O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed was 20 m/s and -20 m/s. The number of training data points and iterations were 403 and 25, respectively.....	146
5-21	Correlation coefficient between the observed and predicted significant wave heights for the Time Delay Neural Network, which was trained with 403 data points.....	149
5-22	Illustration of the relationship between global maximum and minimum wind, and output range of sigmoid transfer function, e.g., $F = 1/(1+\exp^{-2n})$ , where $n = \text{sum of weight values}$ . When $\pm 20$ m/s is used as a global maximum and minimum wind, an ANN may set a large weight value for a given wind speed than $\pm 80$ m/s. In this case, the transfer function output will not change much (solid line) after $n$ is larger than 2.5. In contrast, because $\pm 80$ m/s of a global wind may have smaller weight values, the output of the transfer function changes non-linearly (dotted line).....	153
5-23	Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of $I_{720}H_8O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed was $\pm 80$ m/s. The number of training data points and iterations were 403 and 30, respectively.....	156
5-24	Correlation coefficient between the observed and predicted significant wave heights for the Time Delay Neural Network with a global maximum and minimum wind of $\pm 80$ m/s.....	158

(Continued)

Figure		Page
5-25	Wind field over the northwest Atlantic Ocean on February 22, 12:00, GWT, 1998. Circles indicate the five wave stations along the East Coast of the U.S.....	160
5-26	Effects of the numbers of hidden neurons and iterations on the Mean Square Error (MSE) for the Time Delay Neural Network. The number of training data points was 403.....	162
5-27	Comparison of the correlation coefficient between observed and predicted zero-crossing wave periods for the Time Delay Neural Network at different number of hidden neurons and iterations. The number on each line indicates the number of hidden neurons.....	163
5-28	Comparison of the observed and predicted zero-crossing wave periods using the Time Delay Neural Network with a structure of $I_{720}H_8O_5$ . (a) For Station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed was $\pm 20$ m/s. The number of training data and iterations was 403 and 50, respectively.....	165
5-29	Correlation coefficient between the observed and predicted zero-crossing wave periods for the Three-Layered Time Delay Neural Network. The number of training data points was 403.....	167
5-30	A typical four-layered Time Delay Neural Network with a symbol $I_m H_{n1} H_{n2} O_p$ . Solid lines indicate weights and dashed lines indicate biases.....	170
5-31	Comparison of the correlation coefficient between the observed and predicted zero-crossing wave periods using four-layered Time Delay Neural Network. The first and second numbers indicate the number of hidden neurons and iterations, respectively.....	171
5-32	Comparison of the observed and predicted zero-crossing wave periods using the Time Delay Neural Network with a structure of	



(Continued)

Figure	Page
$I_{720}H_6H_4O_5$ . (a) For Station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed was $\pm 20$ m/s. The number of training data and iterations was 403 and 30, respectively.....	172
5-33 Correlation coefficient between the observed and predicted zero-crossing wave periods for the four-layered Time Delay Neural Network. The number of training data and iterations was 403 and 30, respectively.....	175
6-1 Hurricane wind fields over the northwest Atlantic Ocean. (a) On September 14, 00:00 GWT, 1999 and (b) on September 16, 00:00 GWT, 1999. Circles indicate the five wave stations along the U.S. east coast.....	182
6-2 Observed significant wave height at the five stations (44007, 44009, 44025, 44013, and 41009) along the East Coast of the U.S. from September 11 to 19, 1999 during hurricane Floyd.....	183
6-3 'Pattern 1' hurricane moves along the East Coast of the U.S. from Florida to Maine (a). 'Pattern 2' hurricane crosses the continent in the middle area between 30°N and 40°N (b).....	187
6-4 (a) 'Pattern 3' hurricane comes from the Bahamas and turns left near station 41009. (b) 'Pattern 4' includes many different types of hurricanes generated far away from the coast in the Atlantic Ocean, move toward one particular coastal area, then either go back offshore or cross the continent.....	188
6-5 (a) 'Pattern 5' hurricane forms at southern and eastern far offshore and moves parallel with the East Coast of the U.S. (b) 'Pattern 6' has a similar location of hurricane track to 'Pattern 5', but it moves eastward open ocean.....	189
6-6 Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of $I_{65}H_3O_5$ with Radius of Maximum Wind (RMW) as input at 70 iterations or	

(Continued)

Figure	Page
$I_{60}H_{10}O_5$ without RMW at 60 iterations. (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009.....	197
6-7 Correlation coefficient between the observed and predicted significant wave heights. Pluses indicate the results with the Radius of Maximum Wind speed (RMW) used, and circles indicate the results without RMW.....	199
6-8 Relationship between the maximum wind speeds and central pressures for the 16 hurricanes from 1995 to 2001.....	201
6-9 Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of $I_{60}H_9O_5$ for 12 hurricanes (287 data points at 100 iterations) and $I_{60}H_{10}O_5$ for 15 hurricanes (306 points at 60 iterations). (a) For Station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A 24-hour time-delay was used.....	203
6-10 Correlation coefficient between the observed and predicted wave heights. Triangles indicate the results when 15 hurricanes were used as training data, and circles indicate the results with 12 hurricanes.....	205
6-11 Five hurricane tracks over the northwest Atlantic Ocean from 1991 to 1993. Number one and two are hurricanes Ana and Bob in 1991, number three and four are hurricanes Danniell and Earl in 1992, and number five is hurricane Emily in 1995.....	207
6-12 Comparison of the observed and predicted hurricane significant wave heights using the Time Delay Neural Network with a structure of $I_{12}H_{10}O_5$ for a zero-hour time-delay, $I_{36}H_8O_5$ for a 12-hour time-delay, and $I_{60}H_9O_5$ for a 24-hour time-delay. (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. The radius of maximum wind speed was not used as input.....	210

(Continued)

Figure		Page
6-13	Correlation coefficient between the observed and predicted wave heights. Triangles indicate the results for a 12-hour time-delay, and circles indicate the results for a 24-hour time-delay.....	212
6-14	Hurricane tracks for hurricanes Bertha from July 9 to 14, 1996 and the Floyd from September 12 to 18, 1999.....	216
6-15	Distances for the five wave stations to hurricane center. (a) For hurricane Bertha in 1996 and (b) for Floyd in 1999.....	217
6-16	Time series of the maximum wind speed and central pressure. (a) For hurricane Bertha in 1996 and (b) for Floyd in 1999.....	218
6-17	Plots of the observed significant wave heights at the five wave stations caused by (a) hurricane Bertha in 1996 and (b) Floyd in 1999.....	220
6-18	Tracks of (a) Hurricane Bertha in 1996 and (b) Floyd in 1999. Dates indicate the time of maximum wave height at stations 41009, 44009, and 44025.....	221
6-19	Comparison of the observed and predicted peak wave periods using the Time Delay Neural Network with a structure of $I_{60}H_4O_5$ with four hidden neurons and 110 iterations. (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009.....	225
6-20	Correlation coefficient between the observed and predicted hurricane peak wave periods.....	227
6-21	Time series of the observed peak wave periods at five wave stations. (a) For hurricane Bertha in 1996 and (b) for Floyd in 1999.....	229

## ABSTRACT

Currently available wind-wave prediction models require a prohibitive amount of computing time for simulating non-linear wave-wave interactions. Moreover, some parts of wind-wave generation processes are not fully understood yet. For this reason accurate predictions are not always guaranteed. In contrast, Artificial Neural Network (ANN) techniques are designed to recognize the patterns between input and output so that they can save considerable computing time so that real-time wind-wave forecast can be available to the navy and commercial ships. For this reason, this study tries to use ANN techniques to predict waves for winter storms and hurricanes with much less computing time at the five National Oceanic and Atmospheric Administration (NOAA) wave stations along the East Coast of the U.S. from Florida to Maine (station 44007, 44013, 44025, 44009, and 41009).

In order to identify prediction error sources of an ANN model, the 100 % known wind-wave events simulated from the SMB model were used. The ANN predicted even untrained wind-wave events accurately, and this implied that it could be used for winter-storm and hurricane wave predictions. For the prediction of winter-storm waves, 1999 and 2001 winter-storm events with 403 data points and 1998 winter-storm events with 78 points were prepared for training and validation data sets, respectively. In general, because winter-storms are relatively evenly distributed over a large area and move slowly, wind information (u and v wind components) over a large domain was considered as ANN inputs. When using a 24-hour time-delay to simulate the time required for waves to be fully developed seas, the ANN predicted wave heights ( $r = 0.88$ ) accurately, but the prediction accuracy of zero-crossing wave periods was much less ( $r = 0.61$ ). For the prediction of hurricane waves, 15 hurricanes from 1995 to 2001 and Hurricane Bertha in 1998 were prepared for training and validation data sets, respectively. Because hurricanes affect a relatively small domain, move quickly, and change dramatically with time, the location of hurricane centers, the maximum wind speed, central pressure of hurricane centers, longitudinal and latitudinal distance between wave stations and hurricane centers were used as inputs. The ANN predicted wave height accurately when a 24-hour time-delay was used ( $r = 0.82$ ), but the prediction accuracy of peak-wave periods was much less ( $r = 0.50$ ). This is because the physical processes of wave periods are more complicated than those of wave heights.

This study shows a possibility of an ANN technique as the winter-storm and hurricane-wave prediction model. If more winter-storm and hurricane data can be available, and the prediction of hurricane tracks is possible, we can forecast real-time wind-waves more accurately with less computing time.

## CHAPTER I

### INTRODUCTION

#### 1.1. Rationale

Water waves are the most important force in erosion, transport, and deposition of sediments on coastal areas. They not only change natural environments by eliminating beaches and barriers or reshaping the maps of shorelines, but also impact human life by undermining waterfront houses and public facilities, eventually making them uninhabitable or unusable.

Waves are subdivided by generation forces: tidal waves, tsunami, and gravity waves are generated by gravitational forces, earthquakes, and wind forces, respectively. Among gravity waves, wave periods of 5 to 15 seconds are most important for coastal communities because substantial wave energies are associated with these particular periods.

Change of atmospheric pressure generates different wind speeds near the sea surface, and big waves are accompanied by strong winds. Summer hurricanes and winter storms are the two strong forces that generate large surface wind waves, which are usually gravity waves. For this reason, it is very important to know the expected waves caused by those strong wind systems.

Hurricane is a term used for a specific type of cyclone that has a low-pressure system and counterclockwise wind circulation with wind speeds exceeding 64 kts in the northern hemisphere: the North Atlantic Ocean, the Northeast Pacific Ocean east of the International Date Line, or the South Pacific Ocean east of 160°E (Neumann, 1993).

Winter storm is a general name to define a strong wind system during cold weather. The wind fields of winter storms appear over a large domain. Wind speeds are relatively even distributed in the corresponding areas. The wind speeds and directions of winter storms vary slowly during each event. For this reason, winter storms give enough time for waves to be developed.

In contrast, hurricanes generate from tropical areas and move to extra-tropical areas. Hurricanes usually move rapidly along a track, and the area affected by hurricanes is restricted to a relatively small domain (*e.g.*, on the order of 100km) compared with that for winter storm systems. The intensity of hurricanes depends on the initial intensity, the thermodynamic state of the atmosphere through which it moves, and the heat exchange with the upper layer of the ocean under the core of hurricanes (Emanuel, 1999). In general, hurricane wind speed is much larger when compared with the winter storm wind field, and the wind speed and direction of hurricanes change rapidly with time and location. In general, there is no wind at the hurricane center. Wind speed increases sharply from the hurricane's center and reaches a maximum at a distance called the radius of maximum wind, beyond which it gradually declines.

The strong wind energy from winter storms and hurricanes is transmitted to the sea, generating large waves, which may or may not impinge coastal areas. However, wind-energy transfer to the sea is not the same between hurricanes and winter storms. The wind energies of winter storms are transferred over large areas. Hurricane waves, however, are restricted to a small region because hurricanes have small wind fields. In other words, the wind energies of hurricanes are transferred to waves by strong winds over a relatively small domain.

Moreover, the corresponding wave energy is changing drastically with time because hurricane-wind energy changes rapidly with time and space. The wind velocity for a particular area is always changing. Thus, hurricanes rarely give enough time for water waves to be fully developed. These are the reasons why hurricane wave prediction is much more difficult than winter-storm wave prediction.

Useful warning systems for hurricanes include predicting hurricane intensity, track, and corresponding waves. Today, with the advance of high-speed computers, it is possible to predict wind speeds, locations, and rainfall of hurricanes while hurricanes are in progress. Mathematical models have been developed to estimate wind speeds (Aberson, 2001; Batts *et al.*, 1980; Georgiou *et al.*, 1983; Georgiou, 1985; McAdie and Lawrence, 2000; Vickery and Twisdale, 1995a and b; Vickery *et al.*, 2000).

Many research divisions in the National Oceanic and Atmospheric Administration (NOAA) analyzed the information about storms and hurricanes for future event predictions. For instance, the Hurricane Prediction Center (HPC) made

86 real-time analyses of hurricane wind fields during the 1995 hurricane season using the technique developed during the reconstruction of the hurricane Andrew's wind field (Powell and Houston, 1996).

The National Center for Environmental Prediction (NCEP) provides worldwide wind forecasts. The Hurricane Prediction Center (HPC) has forecasted hurricane tracks three days in advance since 1964. The National Weather Service (NWS) has provided the forecast from the HPC to the public. Since March 2003, the HPC has been able to forecast hurricanes five days in advance.

The Ocean Modeling Branch (OMB) in the NOAA provides forecasts of wave height and period up to 126 hours in advance, using the NOAA WaveWatchIII (NWW3) model. The Tropical Analysis and Forecasting Branch (TAFB) has provided marine forecasts and warnings for the tropical and subtropical oceans of the Atlantic and eastern Pacific using the NWW3 since 2003.

The Tropical Analysis and Forecasting Branch (TAFB) of the HPC compared observed wave heights, obtained from the National Data Buoy Center (NDBC), and predicted wave heights from advanced NWW3 model at the NDBC station 42002 and others in the Gulf of Mexico on February 1, 2002 during a winter storm. Figure 1-1 compares the observed and predicted wave heights (Robert and Christopher). They predicted only five data points that have a 12-hour period, and the correlation coefficient between the observed wave height and predicted wave height is about 0.78.



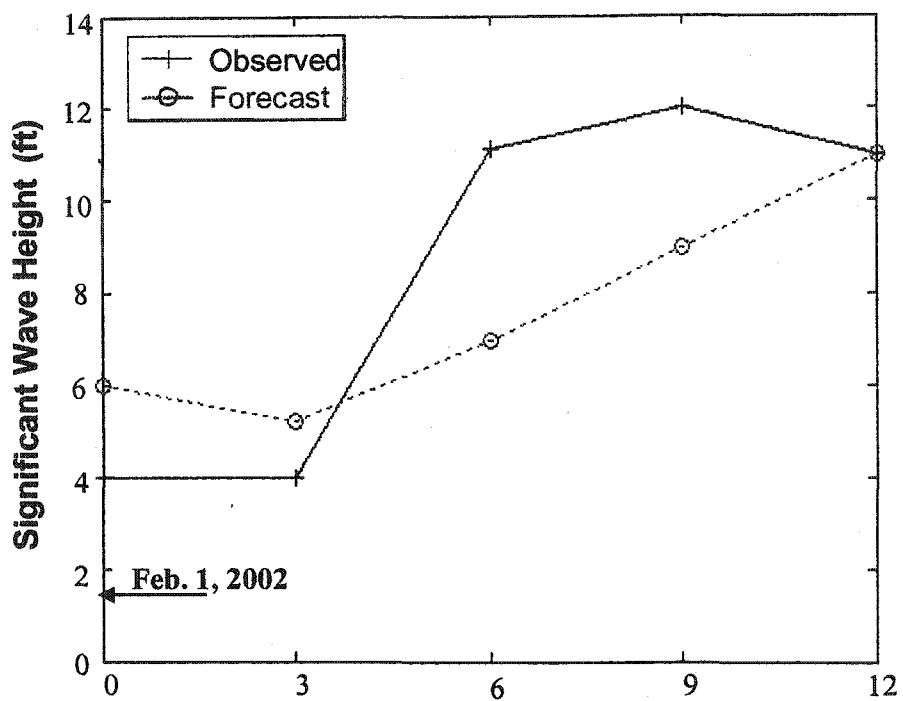


Fig. 1-1. Comparison of observed (solid line) and NWW3 forecasted (dashed line) wave heights at station 42002 (25.17°N and 94.42°W) located in the Gulf of Mexico (after Robert and Christopher, 2002).

Currently available numerical wave models, such as the WAve Model (WAM) and the Simulating WAve Nearshore (SWAN), require prohibitive computing time, and some physical processes are still poorly understood. A more detailed explanation of numerical wind-wave prediction models will be given in Chapter 3.4. In contrast to the improvement in forecasting the intensity and track of hurricanes, there has been comparatively little advance in predictions of storm and hurricane waves. Improvement in computing time and accuracy of the prediction of hurricane waves can save both economic loss and human lives.

Thus, the objective of this study is to develop a faster wave prediction model using an Artificial Neural Network (ANN) technique. Problems of numerical wave models and advantages of an ANN model are explained in the next section.

## 1.2. Problem of Current Numerical Models

Wind-wave generation is a complicated non-linear process that is not yet fully understood. The prediction of wave generations requires knowledge of all processes as well as all inputs. At present, wave prediction models (*e.g.*, WAM and SWAN) solve physical processes numerically. These models include wave-wave interaction terms that are important to represent complicated non-linear characteristics. However, the uncertainty of these non-linear interactions as well as processes of non-linear interaction terms significantly degrade the model accuracy and efficiency. This

means that progress in wave prediction models strongly depends on research of wave-wave interactions and computer advances.

Today, the wind-wave generation models have advanced to third and fourth generation models. However, even the updated wind-wave prediction models are not yet perfect. With respect to physical processes, wind energy inputs and energy dissipations are the least understood processes in wave evolution and are still difficult to adequately simulate. This means that different possible combinations of the two terms can yield the same model results. Hence, even with field verifications, correct wave prediction results do not necessarily mean that wind inputs and energy dissipations have been correctly modeled (Burgers and Makin, 1993).

Low computing speed is another critical weakness of the third generation models. Even though numerical integration of all wave-wave interactions improves prediction accuracy, this integration requires tremendous computer resources. Hasselmann and Hasselmann (1985) provided the Discrete Interaction Approximation (DIA) method, which considers only simplified quadruplet wave interactions to reduce the prohibitive computing time. However, the DIA model still requires considerable computing resources. An improvement on computing speed is only possible when computer power increases by 2000 times (Komatsu and Masuda, 1996). More efficient computing schemes are needed in the future.

### 1.3. What Is ANN? It's Advantages and Limitations

An Artificial Neural Network (ANN) is a system that, in imitation of human brains, learns from past experiences. Biological neurons consist of three functional parts: dendrites for receiving inputs from other sources, soma for processing the inputs in some way, and axons for producing outputs (see Fig. 1-2). An ANN model was designed to simulate these biological functions, so it also has three artificial layers: input, hidden and output layers. In each layer, the number of neurons is selected according to the characteristics of a process. The neurons between the input layer and hidden layer, and between the hidden layer and output layer are interconnected. More information about the basic ANN concept is given in chapter 2.3.

An ANN technique recognizes general patterns and relationships between inputs and corresponding outputs. Before prediction, however, an ANN model must learn through known data sets. This process is called training. The ANN continues to update its weights and biases using summation, multiplication, and transfer functions until it obtains a pre-defined least square error, which is the difference between observed values and model outputs. The core processes of an ANN model self-optimize by trying to find optimum weight values for the errors. After training, the ANN uses the set of best-fitted weights and biases between the neurons for the prediction of future events. (For more information on an ANN model, see chapter 2).

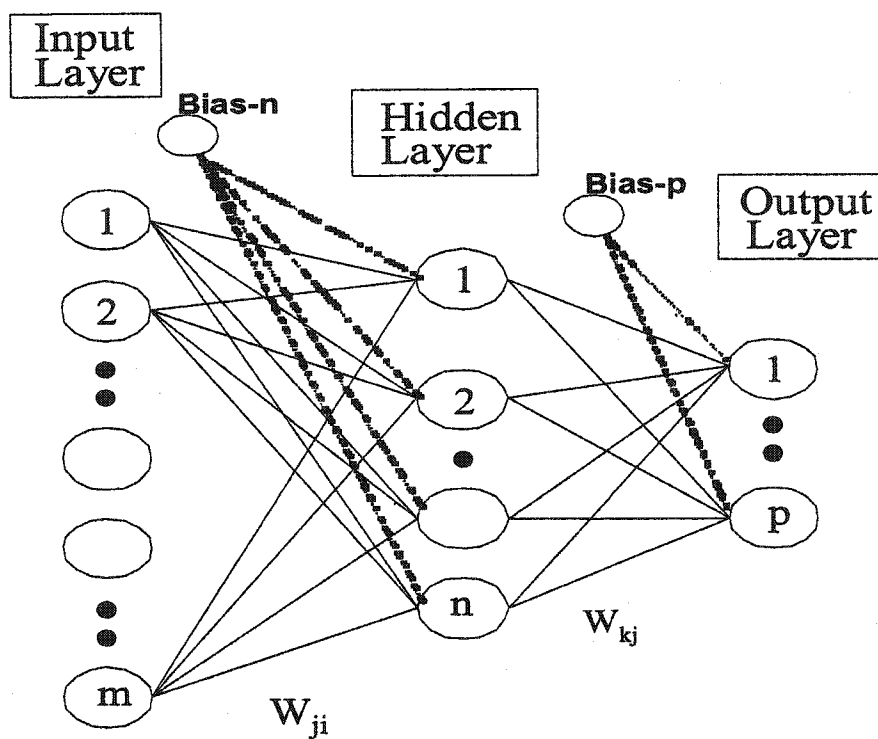


Fig. 1-2. A Typical structure of the Back-Propagation Neural Network with symbol  $I_m H_n O_p$ . Solid lines represent weights and dashed lines indicate biases.

An ANN technique is very useful for wind-wave prediction because of the following reasons. First, an ANN model requires much less computing time compared with traditional numerical models because it does not compute physical processes but finds the differences in patterns between observed inputs and outputs. Reduced computing time is the greatest advantage in using an ANN technique.

Second, an ANN model can avoid error that has resulted from simulating inexactly-known non-linear wave-wave interactions because it does not simulate those processes at all.

Third, an ANN with optimum structure can increase accuracy of wind-wave predictions in the well-trained state (*e.g.*, a sufficient set of training data on wind speed and corresponding wave height). In addition, the training process can be updated whenever new data become available.

There is, however, a critical limitation in using an ANN model. The ANN model cannot be trained without long term or sufficient data on wind and waves. Another limitation is that measurement data of winds and corresponding waves cannot be available everywhere. Therefore, where insufficient or no data are available, an ANN technique is hardly useful to predict wind-waves.

#### 1.4. Objectives

Tsai and Lee (1999), Tsai *et al.* (2002), and Deo *et al.* (2001) used the Back-Propagation Network (BPN) to predict tide level or wind-waves. Although their new

trial of an ANN technique was revolutionary, their prediction results were not satisfactory. For this reason, there are no ANN models available to accurately predict wind-waves yet. More explanation about early ANN studies is given in section 3.5.

The objective of this study is to build ANN wave prediction models for winter-storms and hurricanes with a small computing time at the five stations (44007, 44013, 44025, 44009, and 41009) from Florida to Maine along the east coast of the U.S. Before using ANN techniques, we should first know what is required for ANN modeling and what is its capability for accurate wind-wave prediction.

For this reason, the following five questions were posed and the answers will be given in separate chapters: (1) which is the more efficient ANN scheme and learning algorithm; (2) how to simulate wind-wave generation for ANN modeling and what is(are) the major parameter(s); (3) how to select the optimum number of hidden neurons and iteration; (4) how to identify the sources of prediction errors; and (5) how many wind-wave patterns are needed for training?

### 1.5. Outline of Presentation

This dissertation consists of seven chapters and is presented in the following order. A brief review of the Artificial Neural Network (ANN) is given in chapter II. Review of wind-wave prediction models is given in chapter III. Two test cases for ANN model capability are given in chapter IV, using a linear and the SMB-simulated nonlinear wind-waves. An ANN for winter-storm waves is given in chapter V. An

ANN for hurricane waves is given in chapter VI. Discussion and conclusions are given in chapter VII.



## CHAPTER II

### REVIEW OF ANN

#### 2.1. Introduction

The fundamental background and general features of an ANN model are explained in this chapter. An historical background of an ANN is given first in section 2.2. The basic layout and structural concept of an ANN model are given in section 2.3, including the function of each component of an ANN model. Statistical relationships in an ANN model are explained in section 2.4. Classification of ANN models is given in section 2.5. The various types of ANN learning rules are detailed in section 2.6. A general outline of neural system operation is given in section 2.7, and the optimum number of hidden neurons and iterations is selected in section 2.8.

#### 2.2. Historical Background

Although the study of artificial neurons began in the late 19<sup>th</sup> century, models off the basic theories of human neurons were not possible until computer hardware and software advanced in the 1950s. Farelly and Clark (1954) were the first to simulate the Hebbian theory. They used a digital computer at the Massachusetts

Institute of Technology to construct nodes that represented biological neurons randomly connected with each other. Rosenblatt (1958) opened a new phase of neural network research by presenting an unsupervised learning network, which is called an auto-learning network. This was a pattern regularity detector that learns occurring patterns consistently and regularly. Then, the ANN technique was first applied to solve real world problems (Widrow and Hoff, 1960). Hopfield (1982) clearly showed how such networks could work and what they could do.

Today, ANNs are applied to many areas including defense, industry, commerce, civil engineering, medicine and science for prediction, pattern recognition, classification, signal processing, and data filtering. The promise of ANN seems bright, and development is absolutely dependent on hardware advancement in the future (Kartam *et al.*, 1997; Daniel, 1998).

### 2.3. Basic Concept

An ANN is a system designed to imitate the brain to learn from past experiences, just as children learn to recognize dogs from examples of dogs. After learning, children develop capabilities of generalization beyond the taught data. In the same way, the ANN can recognize patterns from past experience (Haykin, 1994).

The basic unit of the brain is a specific type of cell, the neuron. Up to 20,000 are interconnected and provide the ability to think, remember, and apply previous

experiences to our every action. All these biological neurons have four sub-units: dendrites, soma, axons, and synapses (Fig. 2-1a).

Dendrites receive inputs from external sources, the soma processes the inputs, and the axons perform a non-linear operation and turn the processed inputs into outputs. Signal transmission between neurons occurs at the synapse.

The basic unit of an ANN is the input, hidden, and output layers, which represent corresponding biological neural groups. Each layer may have an unspecified number of artificial neurons. The artificial neurons have much simpler functions than found in biological neurons (see Fig. 2-1b). The number of input parameters determines the number of input neurons, which is represented by a mathematical symbol  $x_i$ . The number of output parameters determines the number of output neurons. However, there is no rule to determine the number of hidden neurons.

Weights are established between the input and hidden layers. The number of interconnections of the neurons between the two layers determines the number of weights. Each weight has a proper numerical value to adequately represent the importance of each input parameter. In other words, an ANN model identifies the optimum weight factor for each input and produces the best output for a given effect of input.

An ANN multiplies the inputs and weight values, and these products are summed arithmetically. The result of the summation is transformed to the output through a transfer function, which has a range from -1 to +1. The transfer function

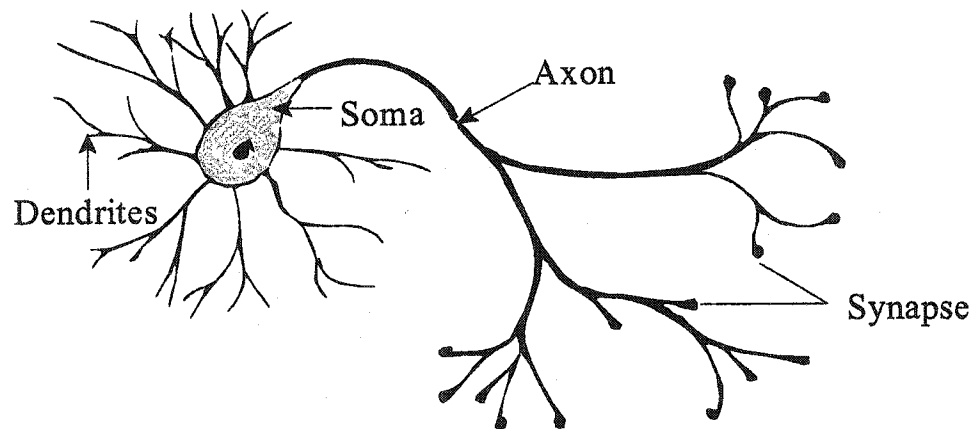


Fig. 2-1 a. The basic unit of biological neurons, which are interconnected each other: Dendrites for accepting input signal, Soma for processing the input, Axon for directing the processed result to the output, and Synapse for representing the final output to other neurons.

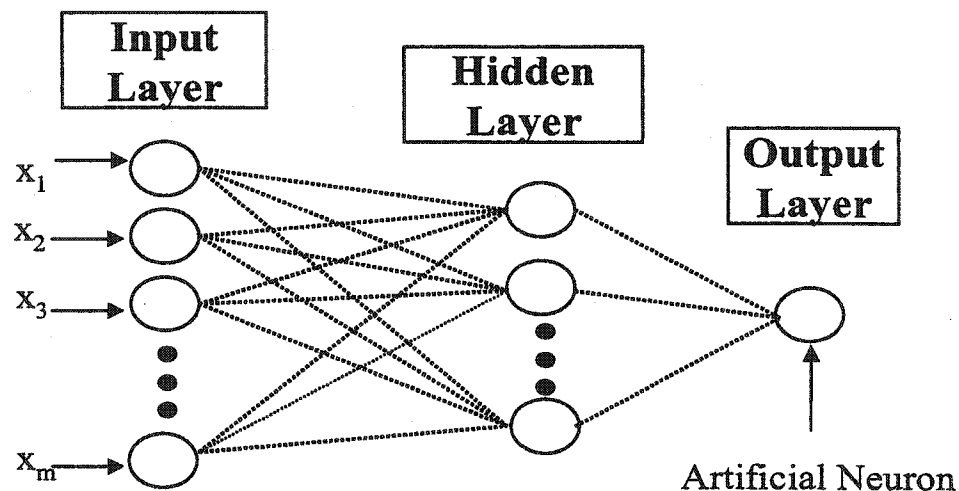


Fig. 2-1 b. A Structure of Artificial Neural Network, which are Simulating Biological Neurons. Each input, hidden, and output layer consists of different numbers of artificial neuron. Input layer is to accept input,  $X$ , hidden layer is to process the input signal, and output layer is to represent the network result.

can process both linear and non-linear problems. Further explanation of transfer function is given in section 2.6.4.1.

The accuracy of the model output is evaluated by the difference between observed values and model outputs. The performance of an ANN model is judged by the 'Least Square Error (LSE)'. A large LSE means that the distributed weight values are improper. Thus, the error signal is fed back to the input and hidden layers. An ANN changes its weight value according to the feedback signal, a process called 'weight update'. The feedback of the LSE and the weight update processes will continue until the LSE is less than a defined value. The repetitive processes of the ANN are called 'iteration'.

The great advantage of an ANN model is that it can self-optimize by finding optimum weight values to reduce the LSE. Hence, an ANN model is easy to use where the relationship between a given input and corresponding output is not clearly known (*e.g.*, non-stationary and time varying environments).

Through these processes, an ANN can learn the patterns from presented representative data. In fact, an ANN model is trained to recognize the hidden relationship between the fed inputs and outputs using their weight values. Afterwards, it can identify the patterns from the learned weight values. Hence, an ANN can be called a physical cellular system that can acquire, store, and utilize experiential knowledge (Zurada, 1992; Daniel, 1998).

## 2.4. Statistical Relationship

There is considerable overlap between the fields of Artificial Neural Networks (ANNs) and statistics. By including feedback topology, most ANNs can be trained to identify relationships effectively from noisy and complicated data when the data set is large. This process is similar to statistical methods, *e.g.*, statistical regression models and the analysis of variance among groups (Francis, 2001; Ripley, 1994).

Data mining aims to find trends, patterns, or regularities in data. It almost always involves a search architecture requiring evaluation of hypotheses at different stages of the search, evaluation of the search output, and appropriate use of the results. Artificial Neural Networks are one of best technical tools for discovering hidden regularities or groupings in data.

Based on a particular theory, one can use a statistical method to find the relationship between variables. Based on the selected theory, measures of uncertainty (*e.g.*, standard deviations) can be generated from probability distributions for given samples. Therefore, statistical applications are often central to data analysis and model fit.

Artificial Neural Networks involve exactly the same kind of model fit. However, ANNs ignore the consequences and importance of particular theories. For instance, many types of ANNs' learning algorithms, such as the Backprop, the Quickprop, and the Levenberg-Marquardt, are modified techniques that use the usual statistical formulae of arithmetic mean such as  $\frac{1}{2n} \sum_{i=1}^n X_i$ , where  $X_i$  = each data point

and  $n$  = the total number of data points. In short, ANNs are used for data mining through learning from a given noisy data set, while a statistic approach must be based on a theoretical formulation such as factor analysis, clustering, and principal component analysis (Hand, 1999; Glymour *et al.*, 1996).

Statistical methods can be used for simulating non-linear systems, *e.g.*, polynomial regression, Fourier series regression, and multivariate adaptive regression splines (Friedman, 1991). However, when using an ANN technique for a non-linear system, efficient handling of many input parameters that are related to each other is an advantage.

## 2.5. Classification of the ANN

There is no standard classification for Artificial Neural Networks (ANNs), but it is possible to roughly classify them according to the learning algorithms and the kinds of data employed.

### 2.5.1. Classification According to Learning Algorithm

According to different learning algorithms, ANNs can be categorized by supervised and unsupervised learning. In supervised learning, both inputs and outputs are provided, and the outputs, so-called desired values, take the role of a teacher for system learning. The ANN processes the inputs and compares its results (*i.e.*, model outputs) with the desired values. Errors are then fed back through the

system, making the system adjust the weights that represent the importance of a particular input (Jordan and Rumelhart, 1992). This back-propagated learning algorithm is the so-called feedback system, because the communication continues until the ANN reaches a pre-determined Least Square Error to predict the right answer (Taylor, 1993; Haykin, 1994).

In unsupervised learning, the representative outputs are not fed back during training. For instance, Kohonen (1995) used an unsupervised ANN for his Self-Organizing Map. This unsupervised ANN is also called a feed forward ANN, because the connections among layers are not iterated. The neurons at the first layer send their outputs to the neurons at the second layer, but they do not receive any input from the second layer neuron.

#### 2.5.2. Classification according to Time Delay

According to whether an ANN uses time delay, either externally or internally, it can be classified as an ANN with or without time delay. If a user externally feeds data from more than one previous time level to the ANN input layer for predicting outputs at the current time level, it is a so-called external time-delay ANN (*e.g.*, Time Delay Neural Network). By contrast, if an ANN uses internal outputs at the previous time level to help predict output at the present time level, it is called an internal time-delay ANN. Comparison of each ANN type is given in the next section.



### 2.5.2.1. BPN

The Back-Propagation Network (BPN) is a popular ANN because it is simple to understand, and it works well for general application. The BPN is a general name for a supervised learning algorithm and can be expressed using the symbol  $I_m H_n O_p$  to represent the input, hidden, and output layers (see Fig. 1-2). In each layer, the number of neurons is represented by subscript  $m$ ,  $n$ , and  $p$ . In general, the numbers of neurons in the input and output layers correspond to the numbers of input parameters and output requirements, respectively. For instance, if there is only one input, then the BPN will use one input neuron at time level  $t = 1$  to predict the output at time level  $t = 1$ . Hence, the BPN does not use external and internal delays.

### 2.5.2.2. ERN

The Elman Recurrent Network (ERN) still consists of the three layers of a Back-Propagation Network (BPN), but uses the additional internal input of the previous time level  $t = 0$  generated from the hidden layer, to predict the output at time level  $t = 1$ . This process results in an increase in the memory of an ANN (Elman, 1990). The place in which the internal input is fed is called the Context Layer (Fig. 2-2).

The ERN has been widely used in applications such as sequence recognition, phonetic representations, and temporal sequence generation. Because it uses information only at time level  $t = 0$ , it is a short-term memory ANN. An easy way to

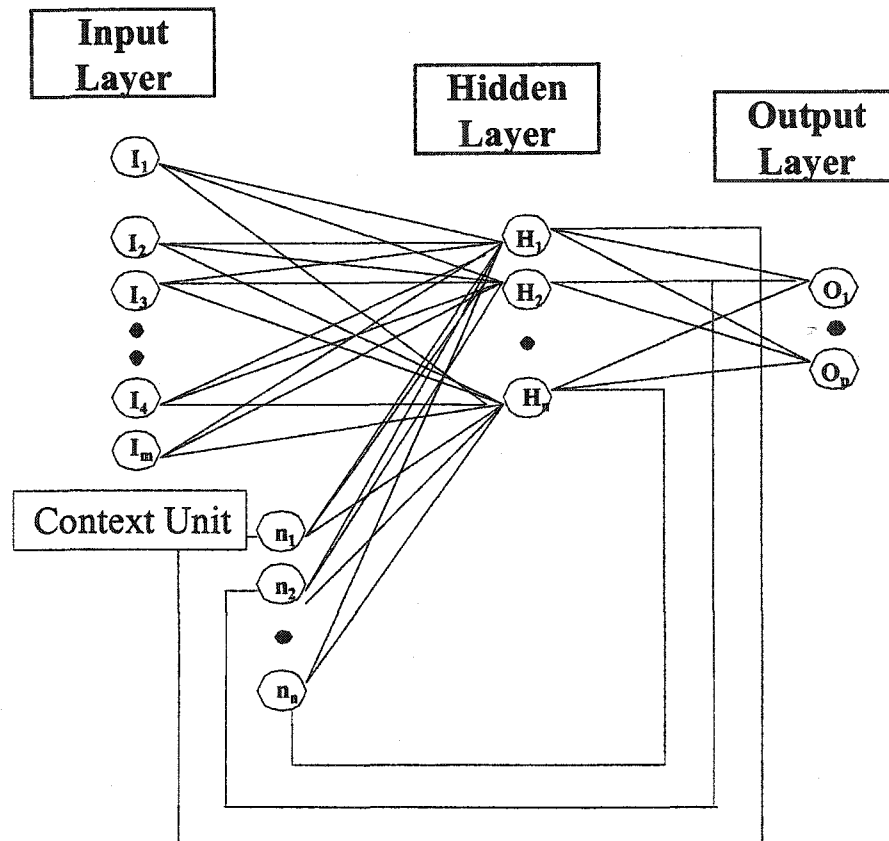


Fig. 2-2. Structure of the Elman Recurrent Network (ERN) with the input, hidden, output layers, and additional context unit. For the prediction ( $O_1 \sim O_p$ ) at the next time level, the ERN will use an external input ( $I_1 \sim I_m$ ) as well as an internal input ( $n_1 \sim n_n$ ) from the context unit.

increase the short-term memory of the ERN is to increase the number of hidden layer neurons (Watrous *et al.*, 1990; Hanes *et al.*, 1994; Elman, 1990).

### 2.5.2.3. TDNN

The Time Delay Neural Network (TDNN) has a time delay at the input and/or in the hidden layers. Depending on the duration of the time delay, there are three different TDNNs: a non-memory TDNN, a short-term memory TDNN, and a long-term memory TDNN.

#### (1) Non-Memory TDNN

The non-memory TDNN uses only an external time-delay. For instance, suppose that there are four external delays. For predicting an output at time level  $t = m$ , the TDNN will use four consecutive inputs that occurred at time level  $t = m, m-1, m-2,$  and  $m-3$  (Fig. 2-3).

#### (2) Short-Term Memory TDNN

The short-term memory TDNN uses both an external and internal time-delay, the latter generated by the hidden layer.

For predicting outputs at time level  $t = m$  (*i.e.*,  $O_m$ ), it uses not only four durations of external time-delay as input at time level  $t_i$ ,  $i = m, m-1, m-2,$  and  $m-3$ , but also the internal result at time level  $t = m-1$  (*i.e.*,  $n_{m-1}$ ), which is provided from the

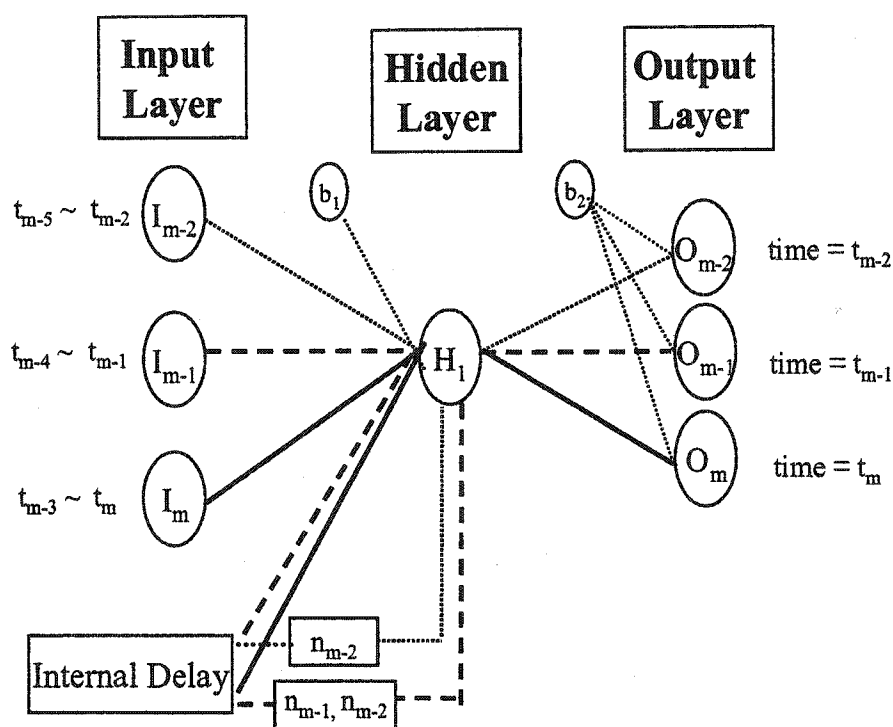


Fig. 2-3. An illustration of the Time Delay Neural Network (TDNN). Each input condition,  $I_m$ ,  $I_{m-1}$ , and  $I_{m-2}$  has four external durations of time-delay at four different time levels (e.g.,  $t = m - 1, m - 2, m - 3$ , and  $m - 4$  for  $I_{m-1}$ , and  $t = m - 2, m - 3, m - 4$ , and  $m - 5$  for  $I_{m-2}$ ). For the prediction of outputs at time level  $t = m - 1$  ( $O_{m-1}$ ), the TDNN uses input condition  $I_{m-1}$ . In order to predict the outputs at time level  $t = m$  ( $O_m$ ), if the TDNN uses an external input  $I_m$  and an internal input  $n_{m-1}$ , this is called short-term memory. For the same purpose, if the TDNN uses an external input  $I_m$  and internal inputs  $n_{m-1}$  and  $n_{m-2}$ , it is called long-term memory.

hidden layer. For predicting new outputs at time level  $t = m+1$  (*i.e.*,  $O_{m+2}$ ), the TDNN model uses both the data from four time steps at time level  $t_i$ ,  $i = m+1, m, m-1, m-2$ , and the previous internal results at time level  $t = m+1$  (*i.e.*,  $n_m$ ). Hence, only one internal result produced at previous time level from the hidden layer, is used for current predictions (see Fig. 2-3).

### (3) Long-Term Memory TDNN

The long-term memory TDNN model uses the external time-delays and internal time-delays that include all previous time levels to predict an output at the current time level (see Fig. 2-3). For instance, suppose that the duration of time delay is four (*e.g.*, time level  $t_i$ ,  $i = m-1, m-2, m-3$ , and  $m-4$ ). For prediction of outputs at time level  $t = m-1$  (*i.e.*,  $O_{m-1}$ ), the TDNN model uses the four durations of external time delay as inputs at time level  $t_i$ ,  $i = m-1, m-2, m-3$ , and  $m-4$  and the previous internal result at time level  $t = m-2$  ( $n_{m-2}$ ). For the prediction of outputs at time level  $t = m$  (*i.e.*,  $O_m$ ), the TDNN model uses four durations of external time delay as inputs at time level  $t_i$ ,  $i = m, m-1, m-2$ , and  $m-3$  and the internal results at time level  $t = m-1$  (*i.e.*,  $n_{m-1}$ ) as well as at time level  $t = m-2$  ( $n_{m-2}$ ). In the same way, to predict outputs at time level  $t = m+10$  (*i.e.*,  $O_{m+10}$ ), the TDNN model will use all previous internal results (*e.g.*,  $n_{m-2}, n_{m-1}, \dots, n_{m+9}$ ) as well as the four consecutive external inputs at time level  $t = m+7, m+8, m+9$  and  $m+10$ .

A long-term memory TDNN seems to be better than the others. Depending on the physical process, however, a long-term memory TDNN may not always be better than others (Clouse *et al.*, 1997).

## 2.6. Learning Rule

Learning rules vary according to the different mathematical algorithms used to update the connection weights. A few basic rules will be explained briefly in this section.

### 2.6.1. Early Learning Rules

Hebb (1949) introduced a principle that became very influential in ANN learning. The characteristic of Hebb's rule is that no desired signals are required, and thus it is one of unsupervised learning algorithms. Only input signals need to be fed to an ANN, and a learning rate is fixed a priori. However, this rule cannot be used for a wave prediction model because the present wave prediction model had to be trained with a set of known data using a supervised learning algorithm.

The delta rule is a variation of Hebb's rule, and it was one of the most commonly used. This rule is based on the idea of continuously modifying the weights of the input connections to reduce the difference between the desired value and the actual network output value, so-called 'delta'. This rule changes the connection weight so as to minimize the Mean Squared Error (MSE). The error is fed

back to the previous layers in the neural network, and this process continues until the MSE reached a defined minimum error. This network is called a back-propagation network because it uses a supervised learning algorithm (Widrow and Hoff, 1960). However, the delta rule is outdated and not usually used at present.

The Hopfield network simulates how human memory works, *e.g.*, a person can be remembered by the type of his hair, eyes, the shape of his nose, his height, the sound of his voice, *etc.* (Hopfield, 1982). The memory of those characters for that person is stored all together as one pattern. Hence, a Hopfield network is defined as an ANN with memory that stores patterns. However, this network is not designed for prediction purposes, so it will not be described in detail. For more information on this rule, see Principe *et al.* (2000).

### 2.6.2. Gradient Descent Rule

The Gradient Descent (GD) is one of the most commonly used supervised learning algorithms for predictive ANN models. For this reason, it will be explained in detail here.

For finding the Least Square Error (LSE), the GD rule is designed to change weight values,  $\Delta w$ , using the following equation; see Eq. 2-1.

$$\Delta W = \eta \frac{\partial E}{\partial W} X \quad (2-1)$$

where E is the error based on a given weight matrix, W, the input column matrix, X, and a learning rate coefficient,  $\eta$  ( $0 < \eta < 1$ ).

The basic mathematical derivations of GD for a two-layered and a three-layered ANN model are available on the Internet ([http://maths.uwa.edu.au/~rkealley/ann\\_all/node1.html](http://maths.uwa.edu.au/~rkealley/ann_all/node1.html)). The following is a brief documentation.

#### 2.6.2.1. Two-Layered Network (Input and Output Layers)

Mathematically, the inputs and the corresponding weights are represented as an input vector,  $X = x_1, x_2, \dots, x_m$ , and weight vector,  $W = w_{11}, w_{12}, \dots, w_{mn}$ . Those inputs and weights are multiplied and summed using the following equation:

$$S = WX + b \quad (2-2)$$

where  $b$  is a bias discussed later.

The result of the summation is transformed to an output through a transfer function between the input and hidden layers. There are two different transfer functions: linear transfer function (*e.g.*, a constant slope or step equations) or non-linear transfer function (*e.g.*, a sigmoid or a Gaussian equation). The non-linear transfer function allows an ANN model to represent complicated non-linear features (Jordan, 1995; Rumelhart and McClelland, 1986; Hornik *et al.*, 1989; Cybenko, 1989; McCullag and Nelder, 1989).

Among those transfer functions, a constant slope function and a sigmoid function are more commonly used. Let  $F$  determine the transfer function, and in the case of a simple constant slope function,  $F$  can be expressed as  $F = aS + c$ , where  $a$  and  $c$  are two constants. On the other hand, a sigmoid transfer function for non-linear case can be expressed as



$$F = \frac{1}{1 + \text{EXP}(-s)} \quad (2-3)$$

The transfer level is based on its net input (*i.e.*,  $S = WX + b$ ). The model result,  $Y$ , for the two-layered ANN structure is the output of a selected transfer function between the input and hidden layers. Thus,  $Y = F(WX + b)$  has a value between -1 and 1.

The above mentioned bias ( $b$ ) is also important to understand. As the complexity of learning task increases, it becomes more difficult for an ANN model to generalize well. Also, it is too expensive to prepare a sufficient number of training data points to ensure good generalization. So, a bias that was learned from the training data set is used for a validation data set. Thus, the problems resulting from an insufficient number of training data points may be reduced if the bias is applied.

The effects of a bias for a linear and non-linear transfer function are explained next. Fig. 2-4a shows an example of a constant slope transfer function with the slope of -1. If the linear function does not use any bias (*e.g.*,  $b = 0$ ), the output of the linear function would be the dashed line. However, if the linear function uses a bias of one (*e.g.*,  $b = 1$ ), the output of the linear function will move toward the solid line and the slope remains the same at -1. Fig. 2-4b shows a sigmoid transfer function that has a bias. A high bias makes the sigmoid function vary gradually, and a low bias makes the sigmoid function vary quickly. In general, an ANN model needs more iterations to find an optimum solution if it does not use bias (Zwart and Vries, 2001, unpublished).

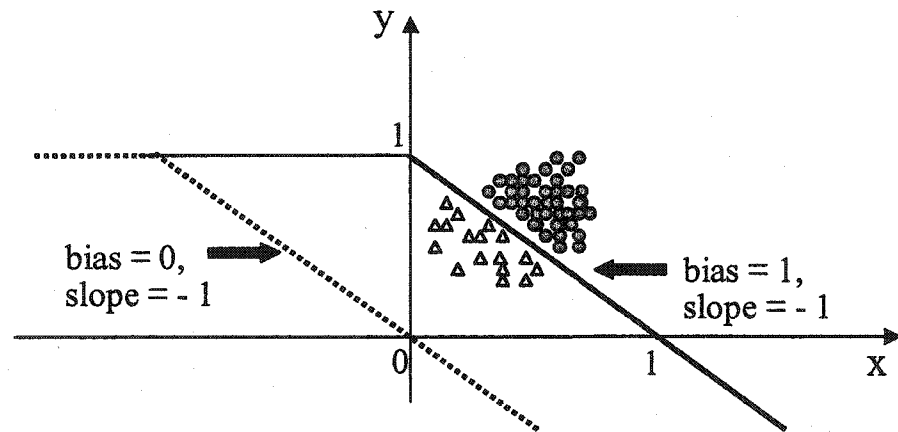


Fig. 2-4a. An illustration of a constant slope linear transfer function where the slope = -1. If bias is not used (e.g.,  $b = 0$ ), the linear function passes through the origin  $(0, 0)$  shown by the dashed line. If bias = 1, the linear function moved towards the solid line with the same slope.

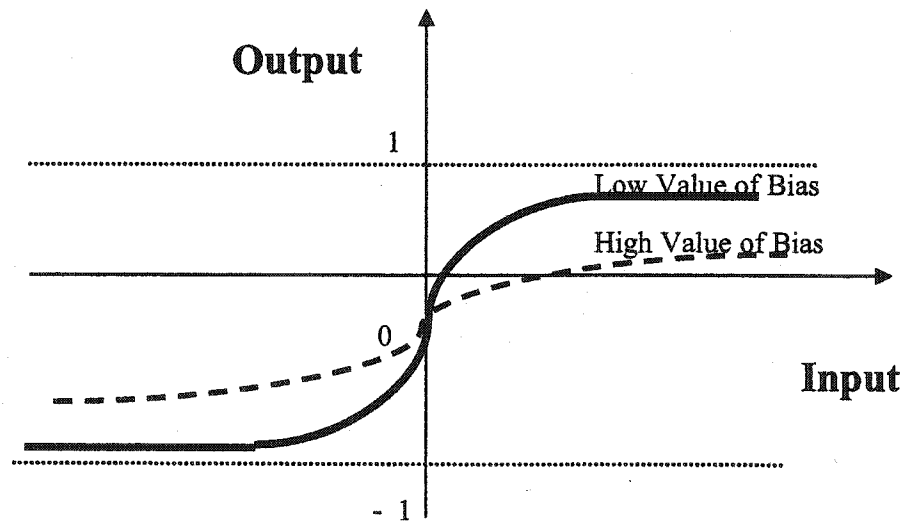


Fig. 2-4b. An illustration of a sigmoid transfer function with a range between -1 and 1. A low initial bias makes a sharp sigmoid function (solid line), while a high initial bias makes the smooth sigmoid function vary gently (dashed line).

The error,  $E$ , is actually an error square, defined as  $E = \sum(d_i - Y_i)^2/2$ , here  $E$  is the difference between desired value,  $d$ , and model output,  $Y$ . If  $E$  is larger than a pre-defined value, the ANN will change its weight according to the GD learning algorithm. The gradient of error in terms of weight,  $W$ , can be written as

$$\begin{aligned} \frac{\partial E}{\partial W} &= e \left( \frac{\partial(d - Y)}{\partial W} \right) \\ &= e \left( \frac{\partial(d - F(s))}{\partial W} \right) \\ &= e \left( \frac{\partial d}{\partial W} - \frac{\partial F}{\partial s} \frac{\partial s}{\partial W} \right) \\ &= e \left[ \frac{\partial d}{\partial W} - F' \left( W \frac{\partial X}{\partial W} + X \frac{\partial W}{\partial W} \right) \right] \end{aligned} \quad (2-4)$$

Because  $d$ ,  $X$ , and  $b$  are independent of  $W$ , *i.e.*,  $\partial X/\partial W = \partial d/\partial W = \partial b/\partial W = 0$ . Thus,  $\partial E/\partial W = -eXF'(s)$ . In the case of a linear transfer function,  $\partial E/\partial W = -eaX$ , where  $a$  is the slope of the linear transfer function.

The change of the weight at the  $(k+1)^{\text{th}}$  iteration is determined by the previous weight and gradient of error in terms of  $W$ ; see Eq. 2-5.

$$W_{k+1} = W_k + \eta \frac{\partial E}{\partial W} \quad (2-5)$$

where  $W_k$  = weight at  $k^{\text{th}}$  iteration

$W_{k+1}$  = weight at  $(k+1)^{\text{th}}$  iteration

$\eta$  = learning rate ( $0 < \eta < 1$ )

The magnitude of the error gradient will change, depending upon the given learning rate.

### 2.6.2.2. Three-Layered Network (Input, Hidden, and Output Layers)

In a three-layered network, the learning algorithm is very similar to that previously explained for a two-layered network. The only difference is that there are two learning processes involved. The first learning is between the input and hidden layers, and the second learning is between the hidden and output layers.

A three-layered network has two transfer functions:  $F_j$  between the input and hidden layers and  $F_k$  between the hidden and output layers. The output between the input and hidden layers is expressed as  $H$ . Thus,  $H = F_j(S_1)$ , where  $S_1$  is the summation,  $WX + b_1$ , between the input and hidden layers. Another output between the hidden and output layers is expressed as  $Y$ . The weight matrix between the hidden and the output layers is expressed as  $U$ , and  $F_k$  is the transfer function. Thus,  $Y = F_k(S_2)$ , where  $S_2$  is the summation,  $UH + b_2$ , between the hidden and output layers.

After model results are produced,  $E$  is calculated. If  $E$  is less than a pre-defined LSE,  $W$  and  $U$  will be used as the final weight matrices. However, if  $E$  is larger than the LSE, an ANN model will change  $U$  and  $W$  according to the GD algorithm, or another selected algorithm.

If using the GD, the gradient of error in terms of U between the hidden and output layers can be expressed as

$$\frac{\partial E}{\partial U} = -eH F_k'(S_2) \quad (2-6)$$

The change of the weight at  $(k+1)^{\text{th}}$  iteration between the hidden and output layers is determined by the previous weight and the gradient of error in terms of U:

$$U_{k+1} = U_k + \eta \left( \frac{\partial E}{\partial U} \right)_k \quad (2-7)$$

where  $U_k$  = weight between the hidden and the output layers at  $k^{\text{th}}$  iteration

$U_{k+1}$  = weight between the hidden and the output layers at  $(k+1)^{\text{th}}$  iteration

$\eta$  = learning rate ( $0 < \eta < 1$ )

The error gradient between the input and hidden layers is related to two transfer functions,  $F_j$  and  $F_k$ , and can be expressed in terms of weight matrices U and W as follows:

$$\frac{\partial E}{\partial W} = \frac{1}{2} \frac{\partial (d-Y)^2}{\partial W}$$

$$= -e \frac{\partial Y}{\partial W}$$

Because  $Y = F_k(S_2)$

$$\frac{\partial E}{\partial W} = -e \frac{\partial F_k}{\partial W} = -e \frac{\partial F_k}{\partial S_2} \frac{\partial S_2}{\partial W}$$

$$\begin{aligned}
&= -eF_k' \frac{\partial(HU + b_2)}{\partial W} = -eF_k' \left( U \frac{\partial H}{\partial W} + H \frac{\partial U}{\partial W} \right) \\
&= -eF_k' U \frac{\partial F_j(S_1)}{\partial W} \text{ because } \partial U / \partial W = 0 \\
&= -eF_k' U \frac{\partial F_j}{\partial S_1} \frac{\partial S_1}{\partial W} \\
&= -eF_k' U F_j' \left( \frac{\partial XW}{\partial W} \right) \\
&= -eUXF_j' F_k'
\end{aligned}$$

$$\text{Thus, } \frac{\partial E}{\partial W} = -eUXF_j'(S_1)F_k'(S_2) \quad (2-8)$$

The change of weight at  $(k+1)^{\text{th}}$  iteration between the input and hidden layers is determined by previous weight and the gradient of error in terms of  $W$  because  $U$  has been already determined between the hidden and output layers; see Eq. 2-5.

Based on the complexity to the problem, more than one hidden layer can be used. If two hidden layers are used, the four-layered ANN structure is expressed as  $I_m H_{n1} H_{n2} O_p$  and includes one more process of the summation, multiplication, and transfer function between two hidden layers. For this reason, the performance of the ANN technique will increase as the number of hidden layers increases.

### 2.6.2.3. The Problem of Gradient Descent

For the GD learning algorithm, there is no guidance to find the optimum learning rate for a given problem. In general, 'trial and error' is the only method and is started with an arbitrarily selected low learning rate. This is because a high learning rate does not guarantee convergence to the LSE. This fixed low learning rate, however, may take a long time to meet the LSE, and there is no proof if the selected learning rate is the optimum (Duda *et al.*, 1997; Rumelhart and McClelland, 1986).

The convergence problem due to a fixed learning rate can be improved if the learning rate changes during the ANN's processes. The Gradient Descent with Variable Learning Rate and Momentum (GDX) and Scaled Conjugate Gradient (SCG) are two of these improved learning algorithms that use a variable learning rate. The characteristics of these two learning algorithms will be explained in the next section.

### 2.6.3. GDX

The GDX algorithm uses a momentum coefficient,  $\alpha$ , and a variable learning rate,  $\eta$  to improve the learning efficiency of an ANN model. The GDX algorithm retains previous weights and effectively smoothes the variations of weights during training. Thus, the effective learning rate is rapid and speeds up the training of the ANN model.

The momentum coefficient is designed to slow the rapid changes of weight. A new weight is updated as a function of previous weights and the error gradient,  $\partial E/\partial W$ ; see Eq. 2-9.

$$\Delta W_{k+1} = \alpha \Delta W_k + \eta \alpha \left( \frac{\Delta E}{\Delta W} \right)_k \quad (2-9)$$

where,  $\alpha$  = less than 1 for stability

The new weight is discarded, if the new error exceeds the old error by more than a predefined ratio (*e.g.*, 1.05), and the learning rate decreases. Otherwise, the new weight is retained. If the new error is less than the old error, the learning rate increases. In general, a default momentum and a learning rate of 0.9 and 0.01, respectively, are used (Rabelo, 1990; Moreira and Fiesler, 1995).

#### 2.6.4. SCG

The Scaled Conjugate Gradient (SCG) is an advanced learning algorithm for calculating a new weight (*e.g.*,  $W_2 = W_1 + \alpha P$ , where  $P$  is a unit vector and  $\alpha$  is the amplitude or length) every iteration using the information from the second-order approximation to find the least square error (Møller, 1993). The unique feature of the SCG is that the new search direction is always conjugate to the previous direction.

For instance, let a given initial weight matrix be  $W_0$ , and suppose an ANN with SCG uses  $W_0$  at the starting point. The first direction,  $P_1$ , is determined by the steepest gradient descent direction,  $r_1$ , which is  $-E_1'(E_1' = \text{the maximum of } (\partial E/\partial W)_1)$ .



In other words, another weight,  $W_1$ , is determined at a point in which  $\partial E/\partial W$  is the maximum.

The step size,  $\alpha_1$  can be calculated using the ratio between  $E_1'$  and  $E_1''P_1$ , here  $E_1'' = (\partial^2 E/\partial W^2)_1$ . Because in calculating  $E''$  every iteration is computationally expensive,  $E_1''P_1$  was approximated as:

$$E_1'' P_1 \approx \frac{\left( E_{(W_1 + \sigma_1 P_1)}' - E_{(W_1)}' \right)}{\sigma_1} + \lambda_1 P_1 \quad (2-10)$$

where  $\sigma_1$  and  $\lambda_1$  are two constants and selected as  $5 \times 10^{-5}$  and  $5 \times 10^{-7}$ , respectively.

A new weight matrix,  $W_2$  is determined by adding the weight ( $W_1$ ), and the product of search direction ( $P_1$ ) and step size ( $\alpha_1$ ):  $W_2 = W_1 + \alpha_1 P_1$ . After knowing the  $W_2$ , the ANN model can calculate a new model output ( $Y_2$ ) and a new error gradient ( $E_2'$ ), and the process repeats.

The new steepest gradient descent direction  $r_2$  is  $-E_2'$ . If  $E_2' \neq 0$ , the ANN will update its new weight and go to the next iteration, otherwise, the weight,  $W_2$ , is the desired value.

For further iteration, a new conjugate search direction  $P_2$  is calculated by adding  $r_2$  and  $\beta_1$ , a modulator factor, defined as  $\beta_1 = (r_2^2 - r_2 r_1)/(P_1^T r_1)$ .

$$P_2 = r_2 + \beta_1 \quad (2-11)$$

In the same way,  $P_2$  is used to calculate  $\alpha_2$  and the next new weight ( $W_3$ ).

## 2.7. Neural System Operation

In this section, the details of establishing an ANN model are given.

### 2.7.1. Pre-Processing

Before training an ANN model, the inputs and outputs are normalized within a range of -1 and 1. This is called pre-processing. Pre-processing is performed using minimum and maximum of the input,  $P_{\min}$  and  $P_{\max}$ ; see Eq. 2-12.

$$P_n = \frac{2(P - P_{\min})}{P_{\max} - P_{\min}} - 1 \quad (2-12)$$

where  $P_n$  = normalized value

$P$  = input and corresponding output

For the pre-processing, choosing maximum and minimum wind speed must be done carefully. For instance, suppose 10 m/s is the maximum wind speed for the training data set and 20 m/s the maximum wind speed for the validation data set. If a maximum wind speed for both training and validation data sets, *e.g.*, 20 m/s, global maximum, is used for the pre-processing, the range of normalized wind speed is 0.5 for the training data set and one for the validation data set. By contrast, if 10 m/s and 20 m/s wind speed are used for the training and validation sets, respectively, so-called local maximum, the range of normalized wind speed is the same for both training and validation data sets.

Further explanation and discussion of global wind speed for pre-processing is given in Chapter 4.2.

### 2.7.2. Selection of Initial Weights

Each element of the input has its own weight for constructing the output. The weight should be different, reflecting the importance of each element. Initial weights can be particularly specified or arbitrarily selected.

An ANN model can also have additional bias neurons between the input and hidden layers, and the hidden and output layers. A constant bias is usually sufficient and is used in ANN modeling.

### 2.7.3. Post-Processing

The results of an ANN model should be explained in the same units as if pre-processing had not been. If there was pre-processing, however, the output results vary between  $-1$  and  $+1$ . These values have to be changed in order to have physical meaning. This process is called post-processing, and the post-processed value is the ANN model output needed. This post-processing is calculated using the following equation:

$$P_p = 0.5(P_n + 1) \times (P_{\max} - P_{\min}) + P_{\min} \quad (2-13)$$

where  $P_p$  = post-processed model output

$P_n$  = ANN model output with pre-process

$P_{\min}$  = minimum output of training data set

$P_{\max}$  = maximum output of training data set

Values of  $P_{\min}$  and  $P_{\max}$  should be carefully selected in post-processing and pre-processing.

For predictions, a new input data set has to be provided to an ANN model. The wind speed will also be pre-processed by using the same minimum and maximum values, which were used in the training data set. The output of the ANN for the prediction is also normalized corresponding to the minimum and maximum values used in the training data set.

#### 2.7.4. Comparison between Desired Value and Model Output

For comparing the performance of an ANN model, four indexes are currently being used: (1) the Mean Square Error (MSE), (2) Root Mean Square Error (RMSE), (3) correlation coefficient ( $r$ ) between two variables, and (4) square correlation coefficient ( $r^2$ ) between the two variables. The MSE indicates that how far model outputs are from their true values. That is to say, the MSE is defined by the mean sum of square deviations between the observed and predicted values (Principe *et al.*, 2000):

$$MSE = \frac{1}{2N} \sum_{i=1}^n (y_i - \bar{y}_i)^2 \quad (2-14)$$

where  $N$  = number of training data points.

The RMSE is the root mean square error between the observed and predicted values, and it is defined by

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2 / 2N} \quad (2-15)$$

where  $y_i$  = observed value, and  $\bar{y}_i$  = predicted value.

If the RMSE or MSE is larger than a pre-defined error, which is ideally zero, the ANN model will adjust the weight between input and hidden layer, and between hidden and output layer until this is smaller than the pre-defined value. A simplified illustration of general ANN procedures for weight update is shown in Fig. 2-5. The weight value that is set at the final iteration during the training process will be used for prediction.

The correlation coefficient ( $r$ ) between the observed and predicted values identifies the strength of the linear relationship between the observed and predicted values rather than estimating prediction errors. For more information about the index of the correlation coefficient between the observed and predicted value, see section 2.8.3.

The square correlation coefficient ( $r^2$ ) between the observed and predicted values gives the proportion of the total variability in the dependent variable  $y$  (predicted value) that can be accounted for by the independent variable  $x$  (observed value). For instance, if  $r = 0.9996$ ,  $r^2 = 0.9996^2 = 0.9992$ . That is to say, 99.92% of the variability in  $y$  is accounted for by  $x$ .

However, there is no guidance for use of a particular index. For instance, Tsai *et al.*, (2002) used the RMSE for determining the optimum learning rate and momentum, and correlation coefficient between the observed and predicted values. On the other hand, Deo *et al.*, (2001) used the MSE and correlation coefficient as agreement indices for the same purpose.

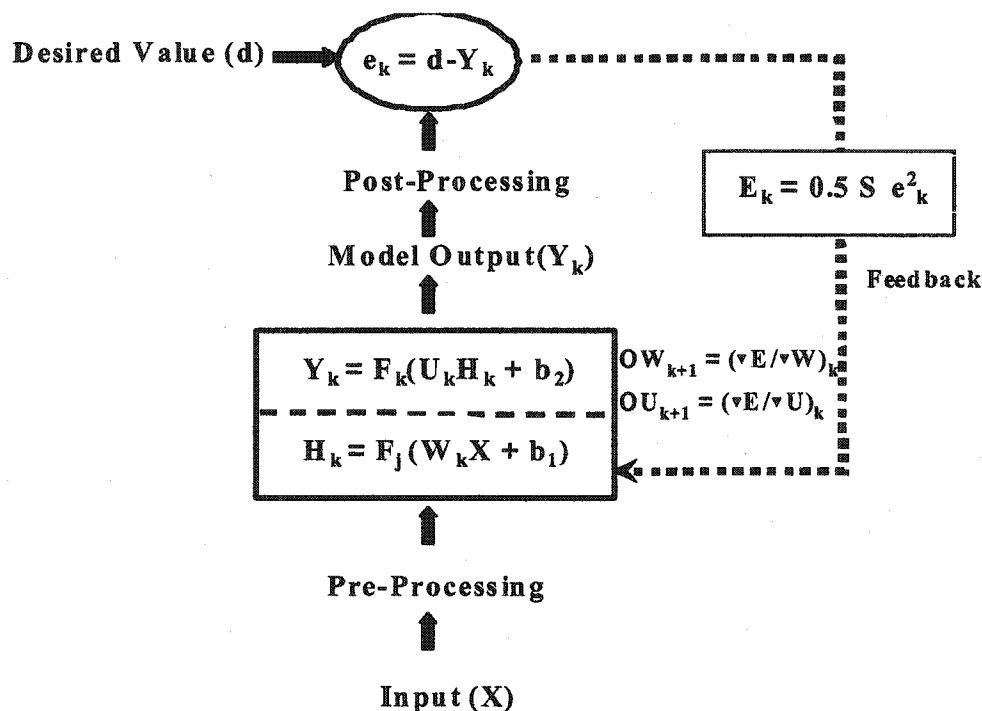


Fig. 2-5. A schematic illustration of the learning pathway of a three-layered Artificial Neural Network.  $W$  = weight matrix between input and hidden layers,  $U$  = weight matrix between hidden and output layers,  $F_j$  = transfer function between input and hidden layers,  $F_k$  = transfer function between hidden and output layers,  $b_1$  and  $b_2$  = bias between input and hidden layers, and between hidden and output layers,  $e$  = difference between  $d$  and  $Y$ ,  $H$  = output at hidden layer,  $Y$  = output of  $F_k(UH + b_2)$  between hidden and output layers, and  $k$  = number of iteration. If the mean square error is larger than a predefined value, which is ideally zero, the ANN will repeat to update weights  $U$  and  $W$  update (notice dotted line) until it meets the criterion.

When using the RMSE or MSE as an agreement index between the observed and predicted value, we can compare the performance of an ANN at a given condition regarding the number of hidden neurons, learning rate, momentum, or training data points. However, these indices do not show exactly what number is the optimum for hidden neurons and iterations in a validation data set. For instance, in general, the RMSE or MSE decreases with an increase in the number of hidden neurons and iterations. However, when using a large number of hidden neurons or iterations, the ANN may be over-structured or over-trained. In this case, the prediction uncertainty for unknown events may increase.

The optimum number of hidden neurons or iterations can be relatively easy to find when using the correlation coefficient between the observed and predicted values because an ANN has a maximum correlation coefficient at the optimum condition. After this optimum, the correlation coefficient may not increase any more while the RMSE or MSE still decreases. However, there still remains a problem that the best correlation coefficient does not guarantee an accurate prediction for other events yet unknown.

In the beginning, we used the MSE in order to find the optimum number of hidden neurons and iterations. But the optimum condition could not be found because the MSE decreased continually as the number of hidden neurons and iterations increased. For this reason, the correlation coefficient between the observed and predicted values was used in this study. In this case, although uncertainties for

other events still remain, the correlation coefficient index will be helpful in finding the optimum condition.

## 2.8. Selection of Optimum Numbers of Hidden Neurons and Iterations

For the best accuracy and computing pace, an ANN model should be used with an optimum number of hidden neurons and iterations. For this reason, the importance of these two variables and their selection is explained in this section.

For a three-layered ANN model with a structure of  $I_m H_n O_p$ , the parameter is determined as follows: parameters =  $(m \times n) + (n \times p)$ . Therefore, selecting the optimum number of hidden neurons will directly determine the number of parameters for the ANN model. Baum and Haussler (1989) suggested that an optimum number of parameters for an ANN model should be less than 10% of the total training data points. For instance, because 218 data points were used for training in this study, the number of parameters should be less than 22. However, an ANN with one hidden neuron has 720 parameters, which is larger than 22 parameters by about 30 times. For this reason, this rule is not sufficient for this study.

Fletcher and Goss (1993) proposed a specific rule to determine the optimum number of hidden neurons with a range between  $(2m + 1)$  and  $(2m^{0.5} + p)$ . For instance, in this study, this rule implies that these should be 58 to 1,441 hidden neurons. However, for multiple input parameters like those in this study, if an ANN uses 58 hidden neurons, the total parameters will be 41,760 which greatly exceeds



218 data points for training by 200 times. Many studies reported that this rule did not give good prediction results (Kuligowski and Barros, 1998; Swingler, 1996; Kim and Barros, 2001). Again, this rule was not used here.

The following five approaches were developed to find an optimum number of hidden neurons and iterations: (1) pruning, (2) regularization, (3) early stopping, (4) trial and error, (5) bias and variance. The pruning method chooses a large number of hidden neurons arbitrarily and then, removes the hidden neurons one by one to find the optimum number of hidden neurons and iterations by using the error sensitivity to the removal of hidden neuron (Hassibi and Stork, 1993; Solla *et al.*, 1990; and Mozer and Smolensky, 1988). However, the selection of reference error sensitivity is unclear and subjective. For this reason, this method is also not used in this study. More information about this selection can be found in Reed (1993).

Regularization was not designed to directly control the number of hidden neurons but to modify the error of an ANN model by decaying a given weight (Tikhonov and Arsenin, 1977). Thus, it also called weight decay method. For instance, the equation,  $xA = y$ , is ill conditioned if a small change of  $y$  due to noises produce an enormous change in the solution for  $x$ . In this case, the total error is to be reduced using a regularization parameter,  $\lambda$ , which limits the growth of weight,  $w$ . It is so-called 'weight decay'. Thus,  $E_p = E + \frac{1}{2} \lambda \sum w^2$ , here  $E_p$  = penalized error and  $E$  = original error. A small value of the regularization parameter (*e.g.*,  $\lambda = 0.00008$ ) could be used (Krogh and Hertz, 1995). However, it is not clear what number of  $\lambda$  is

the optimum value, so this method was not used in this study. Krogh and Hertz (1995) have more details on the weight decay method.

Because the early stopping, 'trial and error', and 'bias and variance' methods were used in this study, details are given in next section.

### 2.8.1. Early Stopping

The early stopping method was developed to find the optimum number of iterations at a fixed number of hidden neurons (Amari *et al.*, 1997; Finnof and Zimmermann, 1993; Sarel, 1994 and 1995; Nelson and Illingworth, 1991). In order to find the optimum number of iterations, a given data set should be divided into a training and validation set. Then, a large number of hidden neurons and a small (as small as possible) iteration have to be arbitrarily selected. The error, the difference between observed and model output of both training and validation data set should be calculated for every trial. The best number of iterations is that for which the error of the validation data set begins to diverge from that of the training data set.

The assumption of this method is that the training and validation data sets are not identical each other. For instance, let the first-half output range of a validation data set be within the output range of a training data set, while letting the second-half output be outside of the training data set. In this case, the error paths of the first patterns for the training and validation data set may pass along the same line (*e.g.*,  $Pass_1 = Pass_2$ ); see Fig. 2-6. However, the error paths for the second-half output for the training and validation data set may be different ( $Pass_1 \neq Pass_2$ ). Sarel (1994)

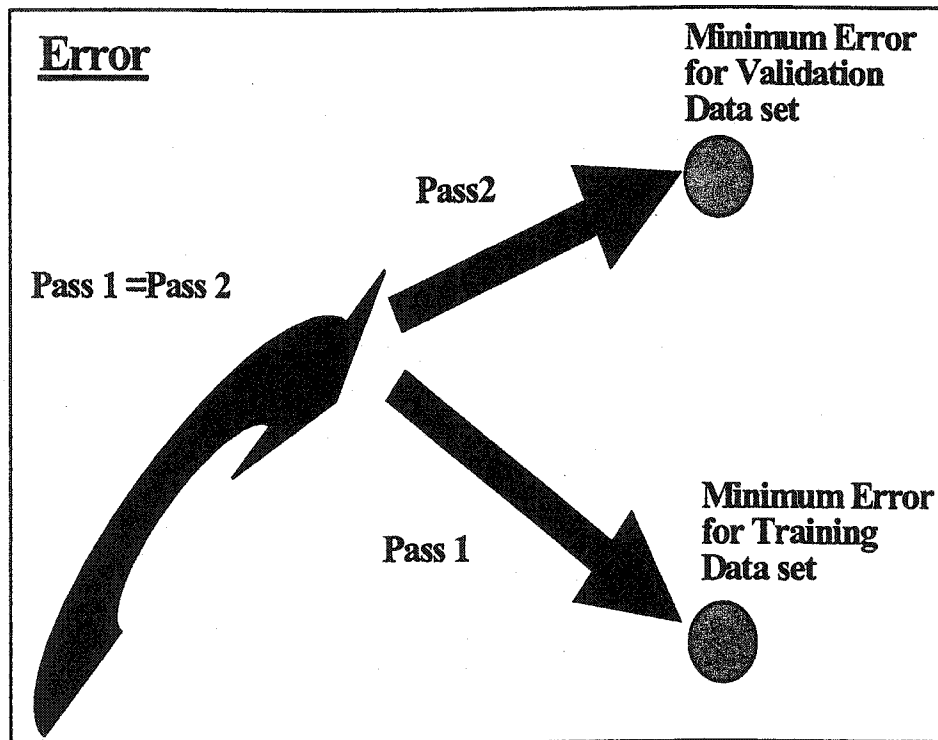


Fig. 2-6. An illustration of error paths for a training and validation data set. The output of the first half for the validation data set is within the output range of the first half for the training data set, while the rest 50 % output for the validation data set is out of range for the training data set. Pass 1 and 2 indicate the error path for the training and validation data set, respectively. For the first half output, pass1 = pass2 because they have the same output. However, for the second half output, pass1 and pass2 may have different errors.

indicated that, at this turning point, the error path for the validation data set is diverging from that for the training data set. If the number of iterations stops early before the ANN model exactly learns all the patterns in the second-half training data set, the ANN may avoid large errors for prediction of unknown patterns.

The early stopping method is faster compared to other previously explained methods, and can be applied successfully to networks in which the number of parameters far exceeds the number of data points (Nelson and Illingworth, 1991).

The early stopping method is applicable only for a fixed small learning rate. The Gradient Descent with variable Learning Rate and Momentum (GDX) and the Scaled Conjugate Gradient (SCG), which were used in this study, do not use a fixed small learning rate but use a variable learning rate that changes during training. Moreover, there is no evidence that the variable learning rate is small enough. For this reason, the early stopping method is not available for GDX and SCG directly.

However, both GDX and SCG use the gradient of error with respect to weight (*e.g.*,  $\partial E/\partial W$ ) to update the weights. In other words, because the change of error in terms of weight between two consecutive iterations is known, the error gradient for a training and validation data set can be used instead of error itself. In the same way, if the input of the second-half for the validation data set is out of the range of input for the second-half for the training data set, the change of error gradient for the validation data set will be different from that for the training data set. Thus,  $\partial E/\partial W$  for the validation and training data is different at this particular number of iterations. Thus,

instead of error, the error gradient was used to select an optimum number of iterations in this study.

### 2.8.2. Bias and Variance

For an ANN model, the number of hidden neurons determines the number of model parameters. From a statistical point of view, if the number of hidden neurons changes, the bias and variance may also change. For this reason, the examination of the bias and variance is important for selecting an optimum number of hidden neurons.

Mathematically, the bias,  $E_b$ , is the difference between the mean of observed values,  $y_m$ , and the mean of model outputs,  $\bar{y}_m$ . Thus,  $E_b = (y_m - \bar{y}_m)^2$ . Also, the variance,  $E_v$ , indicates that how far is each model output,  $\bar{y}_i$ , from the mean of the model outputs,  $\bar{y}_m$ . Thus,  $E_v = 1/n \sum_{i=1}^n (\bar{y}_m - \bar{y}_i)^2$ , here  $n$  = number of data points (Geman *et al.*, 1992).

### 2.8.3. Trial and Error

The trial and error method is one of most widely used for finding the optimum number of hidden neurons and iterations simultaneously (Tsai and Lee, 1999; Tsai *et al.*, 2002; Deo *et al.* 2001; Kim and Barros, 2001). This method tests the efficiency of an ANN model at different combination of hidden neurons and iterations. Here,

the model efficiency of the ANN was determined by the correlation coefficient between observed and predicted values.

The correlation coefficient between the observed and predicted values indicates the strength of the linear relationship between the two variables ranging between -1 and 1. If  $r \neq 0$ , the two variables have a linear relationship, but if  $r = 0$ , there are no linear relationships between the two variables. If one wants to know the percentage of the variability in one variable to the other,  $r^2$  can be used.

The correlation coefficient,  $r$ , can be calculated using the following equation (Davies and Goldsmith, 1972):

$$r = \frac{\sum_{i=1}^n (y_i - y_m)(\bar{y}_i - \bar{y}_m)}{\sqrt{\sum_{i=1}^n (y_i - y_m)^2 \sum_{i=1}^n (\bar{y}_i - \bar{y}_m)^2}} \quad (2-16)$$

where  $n$  = number of data points,

$y_i$  = observed value

$y_m$  = mean of observed values

$\bar{y}_i$  = model output

$\bar{y}_m$  = mean of model outputs

The best correlation coefficient between observed values and model outputs represents the best efficiency of the ANN model, and corresponding numbers of hidden neurons and iterations are the optimum.

## CHAPTER III

### REVIEW OF WIND-WAVE MODELS

#### 3.1. Introduction

Water waves can roughly be classified as monochromatic and random waves, also called regular waves and irregular waves. The study of regular waves can be traced back to the 17<sup>th</sup> century. In this category, only one wave frequency was used to describe wave characteristics, *i.e.*, wave length, wave phase velocity, *etc.*

In this chapter, historical wind-wave prediction models are briefly explored. Monochromatic waves are discussed in section 3.2. The history of early wind-wave prediction models is given in section 3.3. Currently available numerical models are explained in detail in section 3.4. Early studies of ANN wind-wave prediction models are reviewed in section 3.5.

#### 3.2. Monochromatic Waves

Airy (1845) developed a classical linear wave theory that can be applied over an entire wave frequency domain. Stokes (1880) developed a finite amplitude theory, which uses second and higher order approximation to better describe the wave

characteristics for large waves. Wiegel (1960), Kinsman (1965), and Ippen (1966a) presented the development and application of results of those wave theories.

Wave development depends upon wind speed, direction, fetch, and duration. Fetch is the total length of water body over which the wind can act. Wind duration is the total time that a given wind blows over the water body. If wind speed, fetch, and duration increases, the generated wave energy will also increase, resulting in a general decrease of wave numbers (Smith, 1973).

Waves with periods less than 0.1 seconds are called capillary waves, which occur at all stages of wave generation. The continuous random collision and reforming of these capillary waves over time produce well-developed seas that move in the general direction of the wind.

Waves with periods between 1 and 30 seconds are called gravity waves. Among the gravity waves, wave periods from 5 to 15 seconds are the most common and thus, important for coastal communities because a large amount of wave energies are associated with these particular wave periods.

Waves with periods more than 5 minutes are called long period waves. Tidal waves, which are generated by gravitational attraction forces, have 12-hour wave periods. Storm surges, which are generated by a low atmospheric pressure of hurricanes and strong winds, have long wave periods that are more than several hours. Another example of long period waves is Tsunami waves that are generated by earthquakes, submarine landslides, and volcanic explosions. Tsunami waves



usually have periods over 12 hours. For more information about wave classification according to wave periods, see Kinsman (1965).

### 3.3. History of Wind-Wave Prediction Models

For random waves, there are many wave components. Each component wave has its own period and wave height (*i.e.*, energy). To find a representative period and wave height, significant wave height ( $H_s$ ) and wave period ( $T_s$ ) are commonly used to describe the strength of a random wave field.

Munk (1944) defined the significant wave height, also expressed as  $H_{1/3}$ , as the average of the one-third highest waves in an entire wave record. Other terminologies, tenth-wave height,  $H_{10}$ , maximum wave height,  $H_{\max}$ , and root-mean-squared wave height,  $H_{\text{rms}}$ , are also available and have their use for different objectives. The  $H_{10}$  represents the average of the highest 10 percent of all waves.  $H_{\max}$  represents the maximum wave height in the entire wave record. The  $H_{\text{rms}}$  represents the wave height that has root-mean-squared values of the waves.

With respect to wave period, zero-crossing period ( $T_z$ ) and peak wave period ( $T_p$ ) are the two most commonly used wave period to represent a random wave field. The  $T_z$  represents the average period of all wave periods measured when water level crosses the zero-level. The  $T_p$  represents the component which has the largest wave energy. In general,  $T_p$  is about 1.4 times larger than  $T_z$  (Hogben and Dacunha, 1985).

Numerical and empirical studies have contributed to our understanding of wind-wave generation processes (Phillips, 1957; Pierson and Moskowitz, 1964; Kinsman, 1965; Hasselmann *et al.*, 1973; Shore Protection Manual, 1977; World Meteorological Organization, 1998) and continue to expand our understanding. After national efforts led by the U.S. to establish wave models during World War II (WWII), development of wave models accelerated. In the next sections, two historical wave prediction models are briefly presented.

### 3.3.1. SMB

The Sverdrup-Munk-Bretshneider (SMB) model was developed to facilitate military operations during World War II. At that time, the processes of wind-waves and wave-wave interactions were not fully understood. Sverdrup and Munk (1947) found inter-relationships between wind states and waves through empirical observations: waves are changed according to wind speed ( $U$ ), fetch ( $F$ ), and duration ( $t$ ). The SMB used those wind conditions to estimate wave height and period. Later, Bretschneider modified this model in a series (1951, 1952, and 1959).

In general, significant wave height ( $H_s$ ) and wave period ( $T_p$ ) increase when  $F$  and  $t$  increase at a given  $U$  for a constant direction. However,  $H_s$  and  $T_p$  will not increase anymore when wave development comes into equilibrium with  $U$  for a fixed  $F$  or  $t$ . That is to say, a sea is not fully developed for a given  $U$  until a required  $F$  and  $t$  are reached.

The  $H_s$ ,  $T_p$ , and minimum duration for fully developed sea,  $t$ , can be determined by  $U$  and  $F$  using the following equations:

$$\frac{gH_s}{U^2} = 0.283 \tanh\left[0.0125 \left(\frac{gF}{U^2}\right)^{0.42}\right] \quad (3-1)$$

$$\frac{gT_p}{2\pi U} = 1.20 \tanh\left[0.077 \left(\frac{gF}{U^2}\right)^{0.25}\right] \quad (3-2)$$

$$\frac{gt}{U} = K \exp\left\{\left[A \left(\ln\left(\frac{gF}{U^2}\right)\right)^2 - B \ln\left(\frac{gF}{U^2}\right) + C\right]^{\frac{1}{2}} + D \ln\left(\frac{gF}{U^2}\right)\right\} \quad (3-3)$$

where  $g$  = gravitational acceleration

$\ln$  = natural log

$K = 6.5882$

$A = 0.0161$

$B = 0.3692$

$C = 2.2024$

$D = 0.8798$

Equations 3-1, -2, and -3 are shown for the significant wave height, period, and duration, respectively. The use of Eqs. 3-1 to 3 for prediction has been described in details in the Shore Protection Manual (1977), so will not be repeated here. A computer program written in Matlab was also developed. This program is listed in

Appendix I and used for checking the performance of an ANN model compared with the SMB method.

### 3.3.2. Spectrum Models

The second-generation (spectrum) model was produced in the 1960s and early 1970s. In this model, wave growths are assumed to stop when it reaches a universal saturation level. The intensity of a wave-field can be described by integration of wave energy and frequency. In the beginning, wave energy increases with increase of wave frequency, however the energy does not continue to increase after a certain point. This phenomenon is called a fully developed sea. The number of powers of frequency determines an energy peak. However, these models usually overestimate the influence of wind inputs and underestimate the strength of non-linear effects.

Pierson, Neumann, and James (1955) developed a wave spectrum for estimating significant wave height,  $H_s$ , and zero-crossing wave period,  $T_z$ . A random wave may have many wave components, and each component may have a different energy. Thus, a random wave field can be described by a wave energy spectrum.

From a given wave spectrum,  $E(f)$ ,  $H_s$  and  $T_z$  can be calculated, but wave spectrum moments ( $m_n$ ) should be calculated first. The n-order wave moment is determined by integrating wave frequency,  $f$ , and corresponding wave energy,  $E(f)$ .

Thus,  $m_n = \int_0^{\infty} f^n E(f) df$ . If  $n = 0, 1,$  and  $2,$  above results will be zero, first, and second-order moments, which are expressed as  $m_0, m_1,$  and  $m_2,$  respectively. The  $H_s$

and  $T_z$  are determined by zero-order moment and the ratio between zero-order and second-order moments, respectively. Thus,  $H_s = 4\sqrt{m_0}$  and  $T_z = \sqrt{m_0 / m_2}$ .

Philips (1957) was the first person to propose a wave spectrum model as  $E(f) = \alpha g^2 f^{-5}$ . Bretschneider (1959) improved the spectrum as  $E(f) = A f^{-5} e^{(-Bf^{-4})}$ . Later, Pierson and Moskowitz (1964) suggested another improvement based on the wave spectra measured in the North Atlantic Ocean:

$$E(f) = 8.1 \times 10^{-3} g^2 (2\pi)^{-4} f^{-5} e^{(-0.74(\frac{g}{Uf})^4)} \quad (3-4)$$

where  $U$  = mean wind speed measured at 19.5m above sea surface

Results from the JOint North Sea Wave Project (JONSWAP) suggested that the wind-wave spectrum in a growing phase had a much sharper peak than the Pierson-Moskowitz spectrum and provided a peak enhancement factor as:

$$E(f) = \alpha g^2 (2\pi)^{-4} f^{-5} e^{\left\{ \frac{5}{4} \left( \frac{f_m}{f} \right)^4 + \ln \gamma \cdot e^{\left[ \frac{-(f-f_m)^2}{2\sigma^2} \right]} \right\}} \quad (3-5)$$

where  $f_m$  = peak frequency

$\alpha$  = Phillip's constant or equilibrium range constant

$\gamma$  = peak enhancement factor

$\sigma = 0.07$  for  $f \leq f_m$

$0.09$  for  $f > f_m$

For more information on JONSWAP wave spectrum, see Hasselmann *et al.* (1973).

### 3.4. Currently Available Wind-Wave Models

The third generation of wave prediction models developed in the late 1980s overcame those non-linear problems by using wave-wave interaction terms (SWAMP, 1985). Wave energy at a specific point on the sea surface was described as a function of frequency, wave direction, and position, and the energies were balanced by energy sources and sinks.

Currently available numerical models for wave hindcast or prediction include WAM, SWAN, HISWA, STWAVE, and GLERL. The basic concept of currently used wave prediction models will be explained in detail in the next section.

#### 3.4.1. WAM

The WAve Model (WAM) is the third generation of wave prediction models that solve the wave transport equation explicitly without any presumptions on the shape of the wave spectrum in deep water (WAMDI, 1998). The propagation of wave energy, which is balanced by wave energy on the sea surface, can be described as a function of longitudinal ( $\varphi$ ), and latitudinal ( $\lambda$ ), and wave direction ( $\theta$ ) using the following equation:

$$\frac{\partial E}{\partial t} + \frac{\partial C_{\varphi} E}{C_{\varphi}} + \frac{\partial C_{\lambda} E}{C_{\lambda}} + \frac{\partial C_{\theta} E}{C_{\theta}} = S \quad (3-6)$$

where  $C_{\varphi} = C_{g\varphi} + u$

$$C_{\lambda} = C_{g\lambda} + v$$

$$C_{\theta} = C_{g\theta} + U, \text{ here } U = u + v$$

$C_{g\phi}$  and  $C_{g\lambda}$  = wave group velocity in latitudinal and longitudinal directions

$u$  and  $v$  = depth and time averaged current velocity in  $x$  and  $y$  directions

$S$  = net wave energy source and sink

The four terms on the left of equation 3-6 indicate the change of wave energy with time and the change of wave energy in moving wave group velocity and current velocity in terms of three variables:  $\phi$ ,  $\lambda$ , and  $\theta$ . The right term in equation 3-6 indicates net energy source and sink from wave energy input, energy dissipation due to white-capping, and non-linear wave interaction.

Although accuracy of the model design was improved by including a non-linear interaction term, one big problem of the WAM is that solving the non-linear wave interaction term takes too much computational time. Hasselmann and Hasselmann (1985) have developed the Discrete Interaction Approximation method, which considers only simplified quadruplet wave interactions, to reduce prohibitive computing time. But it still needs considerable computational resources. According to Komatsu and Masuda (1996), a 2000-fold improvement on computational power is needed and an increase is unlikely to occur soon.

With respect to prediction accuracy, the WAM usually underestimates wind-waves by 10 % and swells by 30% (Wen *et al.*, 1999). Nonetheless, the WAM, or its derivatives are currently running for prediction of wind waves.

### 3.4.2. SWAN

The Simulating WAVes Nearshore (SWAN) is a wind-wave prediction model, which has been developed to compute random, short-crest wave conditions in small scale coastal regions and inland waters (Booij *et al.*, 1999). The irregular waves are described by the two-dimensional density spectrum of wave actions: wave frequency and position, which are conserved in the presence of ambient currents (Whitham, 1974). Hasselmann *et al.* (1973) mathematically described the propagation of wave action density that is balanced by net wave energy, in terms of x and y directions, wave direction ( $\theta$ ), and relative wave frequency (f), as follows:

$$\frac{\partial N}{\partial t} + \frac{\partial C_x N}{\partial x} + \frac{\partial C_y N}{\partial y} + \frac{\partial C_\theta N}{\partial \theta} + \frac{\partial C_f N}{\partial f} = \frac{S}{f} \quad (3-7)$$

where N = action density (E/f)

$$C_x = C_{gx} + u$$

$$C_y = C_{gy} + v$$

$$C_\theta = C_{g\theta} + U, \text{ here } U = u + v$$

$$C_f = C_{gf} + U$$

$$C_{gx} \text{ and } C_{gy} = \text{wave group velocity in x and y directions}$$

$$u \text{ and } v = \text{current velocity in x and y directions}$$

$$S = \text{net energy source and sink}$$

The left terms in equation 3-7 indicate the change of action density with time, location, wave moving direction, and energy transfer among each component in terms of four variables: x and y directions,  $\theta$ , and f.



When compared with the WAM, the SWAN has one more variable,  $f$ , because of the possible depth effect. For this reason, SWAN is expected to require even more computational time than the WAM. As another shortcoming, the SWAN cannot solve the wave diffraction process.

### 3.4.3. HISWA

The HIndcasting Shallow water WAVes (HISWA) was developed by Booij *et al.* (1985) for wave predictions at shallow depth areas, which are strongly influenced by coastal morphology (*e.g.*, islands, bars, shoals, and channels). The HISWA uses a parameterized process on the frequency domain to reduce the huge computation time. The parameterization is formulated in the zero and first order of spectrum moments ( $m_0$  and  $m_1$ ) in each spectral direction. Using this momentum equation, the evolution of  $m_1$  of the action density spectrum is induced.

$$\frac{m_1}{t} + \frac{C_x^+ m_1}{x} + \frac{C_y^+ m_1}{y} + \frac{C_\theta^+ m_1}{C} = C_f^* m_0 + S_1 \quad (3-8)$$

where  $C_x^+$  = propagation speed of  $m_1$  in x direction

$C_y^+$  = propagation speed of  $m_1$  in y direction

$C_\theta^+$  = propagation speed of  $m_1$  in wave direction

$C_f^* m_0$  = effect of time variations in currents and depth on the mean frequency

$S_1$  = net generation and dissipation of  $m_1$

The left hand terms in equation 3-8 indicate the change of  $m_1$  with time, and the flux change of  $m_1$  in terms of three variables:  $x$  and  $y$ , and  $\theta$ . The right hand terms in

equation 3-8 indicate the effects of time variation in current and depth on the mean frequency, and net generation and dissipation of  $m_1$ .

HISWA takes less computational time compared with that for SWAN because it does not use a variable for wave frequency. HISWA, however, is not effective when actual wave directions do not fall within a certain direction boundary, and the situation is not stationary. In other words, the wave spectrum is discrete only in limited directions and the shape of the frequency spectrum has to be prescribed due to parametric frequency. For this reason, HISWA is restricted to the prescribed shape of spectrum (Holthuijsen *et al.*, 1997; Resio, 1987 and 1988; Resio and Perrie, 1989).

#### 3.4.4. STWAVE

The STeady state irregular WAVE (STWAVE) which simulates wave energy transformation is easy to apply and flexible for near shore wind-generated wave growth and propagation (Resio and Perrie, 1989). This model is based on the assumption of steady-state condition, which means that wave energy is always considered to be in equilibrium with time. For this reason, the STWAVE has no local time derivative term, and it does not provide information on wave evolution. The important features of the STWAVE are that the wave spectrum in shallow depths has a depth independent equilibrium range, and the growth of spectral peak frequency is limited.

The governing equation of this model solved the spectrum energy in the moving group velocity of the spectral peak:

$$\frac{C_{gx} E(f, \theta)}{x} + \frac{C_{gy} E(f, \theta)}{y} = S \quad (3-9)$$

where  $E(f, \theta)$  = the energy spectrum in frequency and direction

$C_{gx}$  and  $C_{gy}$  = group velocity in x and y directions

$S$  = net energy source and sink

The left hand terms of equation 3-9 indicate the change of wave spectrum energy in moving group velocity in terms of x and y directions. The right of the equation indicates net energy source and sink.

The computational speed is fast because the STWAVE has assumed that waves are in a steady state, and they move with the spectral center. But the accuracy of wind-wave prediction has not been clearly reported so far.

#### 3.4.5. WAVEWATCH

The development of the WAVEWATCH has three different phases: WAVEWATCH I, II, and III. The WAVEWATCH I was first developed by Tolman (1989 and 1991) at Delft University of Technology. Tolman (1992) improved the governing equations, the model structures, numerical methods, and physical parameterizations to create the WAVEWATCH II.

The Ocean Modeling Branch at the National Center for Environmental Prediction (NCEP) has developed the WAVEWATCH III as a new global wave forecast system (Tolman, 1997). The basic concept of WAVEWATCH III is like the WAM, except it uses wave number as an additional variable. The major governing

equation of WAVEWATCH III can be described mathematically by the following equation:

$$\frac{\partial N}{\partial t} + \frac{\partial C_x N}{\partial x} + \frac{\partial C_y N}{\partial y} + \frac{\partial k N}{\partial k} + \frac{\partial \theta N}{\partial \theta} = \frac{S}{\sigma} \quad (3-10)$$

where  $C_x = C_{gx} + u$

$$C_y = C_{gy} + v$$

$C_{gx}$  and  $C_{gy}$  = group velocity in x and y directions

u and v = depth and time averaged current velocity in x and y directions

$$k = -\frac{\partial \sigma}{\partial d} \frac{\partial d}{\partial s} - k \frac{\partial u}{\partial s}, \text{ here } d = \text{mean depth}$$

$$\theta = -\frac{1}{k} \left( \frac{\partial \sigma}{\partial d} \frac{\partial d}{\partial m} - k \frac{\partial u}{\partial m} \right)$$

here m = a coordinate perpendicular direction of net S

N = wave action density

S = net energy source and sink

k = wave number ( $2\pi / L$ )

$\theta$  = wave direction

m = perpendicular coordinate to net energy source

$\sigma$  = intrinsic frequency ( $\sigma^2 = gk \tanh(kh)$ , here h = mean water depth)

The left hand terms in equation 3-10 indicate the change of wave spectral action density with time, and the change of action density flux in terms of four domains: x and y directions, wave number, and direction. The right side of the equation indicates the change of energy source and sink in terms of wave frequency.

One implicit assumption of WAVEWATCH III is that depth and current as well as wave field vary with time and space much more than corresponding scales of a single wave.

The WAVEWATCH III is expected to be more accurate than the WAM because it is designed to solve for one additional parameter, wave number. This model is expected to need significantly more computationally expensive due to its higher-order accurate numerical scheme (Tolman, 1997).

#### 3.4.6. GLERL

The Great Lakes Environmental Research Laboratory (GLERL) wave prediction model was developed in the National Ocean and Atmospheric Administration for forecasting wave height in the Great Lakes (Schwab *et al.*, 1984).

The GLERL uses local momentum for deep-water wave prediction, which assumes that the potential energy is equal to the kinetic energy in the wave fields. The momentum force,  $F$ , is determined by the mass,  $m$ , and acceleration rate (*i.e.*,  $F = ma$ ). The momentum component in  $x$  and  $y$  directions are expressed as:

$$M_x = g \int_0^{\infty} \int_0^{2\pi} \frac{F(f, \theta)}{C(f)} \cos \theta d\theta df \quad (3-11)$$

$$M_y = g \int_0^{\infty} \int_0^{2\pi} \frac{F(f, \theta)}{C(f)} \sin \theta d\theta df \quad (3-12)$$

where  $M_x$  and  $M_y$  = momentum in  $x$  and  $y$  directions

$C$  = phase velocity

The governing equation of those momentums is not based on the energy transport equation but the local momentum balance equation. In other words, wind force,  $\tau^w$ , is balanced by the local and accelerated change of momentum with respect to time,  $t$ , and  $x$  and  $y$  directions:

$$\frac{dM_x}{dt} + \frac{\partial T_{xx}}{\partial x} + \frac{\partial T_{xy}}{\partial y} = \frac{\tau^w}{\rho_w} \quad (3-13)$$

$$\frac{dM_y}{dt} + \frac{\partial T_{yx}}{\partial x} + \frac{\partial T_{yy}}{\partial y} = \frac{\tau^w}{\rho_w} \quad (3-14)$$

where  $\rho^w$  = air density

$T_{xx}$ ,  $T_{xy}$ ,  $T_{yx}$ , and  $T_{yy}$  = wave radiation stress tensor

According to linear theory, group velocity is equal to half of the phase velocity, thus radiation stress tensors are:

$$T_{xx} = \frac{g}{2} \int_0^{\infty} \int_0^{2\pi} F(f, \theta) \cos^2 \theta d\theta df \quad (3-15)$$

$$T_{xy} = T_{yx} = \frac{g}{2} \int_0^{\infty} \int_0^{2\pi} F(f, \theta) \sin \theta \cos \theta d\theta df \quad (3-16)$$

$$T_{yy} = \frac{g}{2} \int_0^{\infty} \int_0^{2\pi} F(f, \theta) \sin^2 \theta d\theta df \quad (3-17)$$

The wave spectrum of the GLERL was assumed to agree with the Joint North Sea Wave Project spectrum, the so-called JONSWAP spectrum, which has the three

parameters: peak frequency ( $f_m$ ) Phillips equilibrium range parameter ( $\alpha$ ), and constant ( $\sigma$ ) (see Eq. 3-5).

The problem with GLERL is that the magnitude of the total momentum from the air to water used to generate waves is not clearly known, and there is no direct measurement. For this reason, the weakness of the GLERL model is that the momentum fraction has to be adjusted to fit the observed wave heights to this model prediction. (For more information on the GLERL model, see Schwab *et al.*, 1984.

### 3.5. Early ANN Wind-Wave Prediction Models

Tsai and Lee (1999) used the Back-Propagation Network (BPN) to estimate tidal level at the Taichung harbor and Mitour coast in Taiwan. Two observed consecutive tide levels and the difference in tide level between observed and predicted tide levels at time  $t-1$  and  $t-2$  to predict tidal elevation at time  $t$  were used for ANN inputs.

As a training data set, they regrouped one month of tidal levels in January, 1995 into five different data sets according to data span: one, three, five, eight, 15, and 30-day. As a validation set, five different data sets observed in 1996 were used: one, two, three, five, and 12-month.

The prediction results were satisfactory with the correlation coefficients larger than 0.9 between observed and predicted wave heights when using only one-day tide levels for training.

However, in order to predict the tide level at time =  $t$ , the observed tide level at time  $t-1$  and  $t-2$  from predicted results are still needed for preparing ANN inputs. That is to say, the predicted tide-level at a previous time step must be used as an ANN input for current-time prediction. For this reason, it is hard to say the BPN was used to predict tide levels for unknown events.

Tsai *et al.* (2002) used the BPN model to estimate wave height at Taichung Harbor in Taiwan. Significant, tenth, maximum, and mean wave heights at two nearby wave stations were used as ANN inputs.

They collected two training data sets with one-month spans in September 1994 and 1995, and another two validation data sets with three-month spans from December 1994 to February 1995 and from December 1995 to February 1996, respectively.

The prediction results were satisfactory, and the correlation coefficient between observed and predicted wave heights for two validation sets was larger than 0.9. However, for the prediction of future waves, waves at two nearby wave stations must be known in advance. For this reason, their study is not reliable for practical wind-wave prediction.

Deo *et al.* (2001) used a three-layered BPN to predict significant wave height and zero-crossing wave period at a near-shore wave station with 16m-depth, several kilometers away from Karmar and Mumbai in India.

Wind and wave data with 900 points measured every three hours were prepared from March 1988 to July 1988 and from December in 1988 and May 1989. Among



those data, 720 points were used for training and another 180 points were used as validation data. They used wind speed, fetch, and duration as ANN input. However, the prediction results were not accurate.

In the second trial, they separated previously collected data by different wind patterns, *e.g.*, monsoon and fair weather season, and the BPN was used to predict only wave heights for monsoons. At this time, the wind speed and one previous time step were used as inputs. However, the prediction results were still bad.

They then prepared another data set with 168 points measured at a relatively deep-water (75m) wave station offshore, from October 1992 to January 1993. For training and validation data sets, 134 and 34 points were used, respectively. For reducing uncertainties resulting from sudden shift in winds, at this time, weekly-averaged wind information at about four wind stations were considered. The prediction results improved so that the correlation coefficient between observed and predicted wave height became 0.77. However, wave period prediction was still poor.

There are two drawbacks in the Deo *et al.* (2001) study. (1) It is well known that wind direction as well as wind speed is important in generating wind-waves. However, they did not use wind direction as input. (2) Winds blowing away from the wave recorder can generate swells, which can be observed at the wave recorder after several hours. Thus, wind information on the wind fields upwind side as well as the spot of interest is important for wind-wave prediction.

## CHAPTER IV

### TWO TEST CASES FOR CHECKING ANN MODEL CAPABILITY

#### 4.1. Introduction

There are two basic questions regarding the use of Artificial Neural Network (ANN) techniques for wind-wave predictions: (1) How complicated are wind-wave systems, and (2) what is the ultimate capability of ANN techniques? Theoretically, the number of training observations and the optimum ANN structure for wave predictions should change according to the level of complexity in wind-wave systems. However, there is no reference to identify ANN capabilities for wind-wave predictions. Thus, without the understanding of ANN capabilities, it is hard to determine to what extent prediction errors are caused by the complexities of wind-wave systems for a selected ANN structure or by limitation of the ANN technique itself.

For this reason, we carried out two experiments for (1) a simple linear case and (2) a complicated non-linear case to observe how an ANN model responds differently between the two contrasting cases. Because the linear case has a simple relationship between inputs and outputs and has no other error sources, we can find how to properly use an ANN model for accurate prediction in terms of structures and procedures in terms of structures and procedures. For the non-linear case, because we

can set the model conditions and the number of training data sets, the error sources must be studied.

In this chapter, the ANN capabilities of predicting linearly and non-linearly simulated wind-waves were checked. The test of the linear case for ANN modeling is given in section 4.2. The test of non-linear case for ANN modeling and conclusions are given in sections 4.3 and 4.4, respectively.

#### 4.2. Linear Case

For this study, a pure linear function between wind speed and corresponding wave height was assumed. For a training data set, wind and waves were generated by a simple linear equation,  $H = U/6$  where  $H$  = wave height and  $U$  = wind speed. For the first 10 hours, wind speed increases from 0 m/s to 30 m/s. Afterwards,  $U$  linearly decreases to zero at time = 20 hours. For the validation data set,  $H$  and  $U$  are also generated by the same equation and time span, but wind speed increases to 60 m/s (Fig. 4-1).

In order to determine the effects of pre-processing, the Back-Propagation Network (BPN) was tested with or without pre-processing. If the pre-processing was used, the input and output in the training and validation data sets should be normalized by a selected pair of global maximums and minimums. For this study, two wind speeds of 50 m/s and 70 m/s were arbitrarily selected as the global maximum. When 50 m/s was used as a global maximum wind speed, the maximum

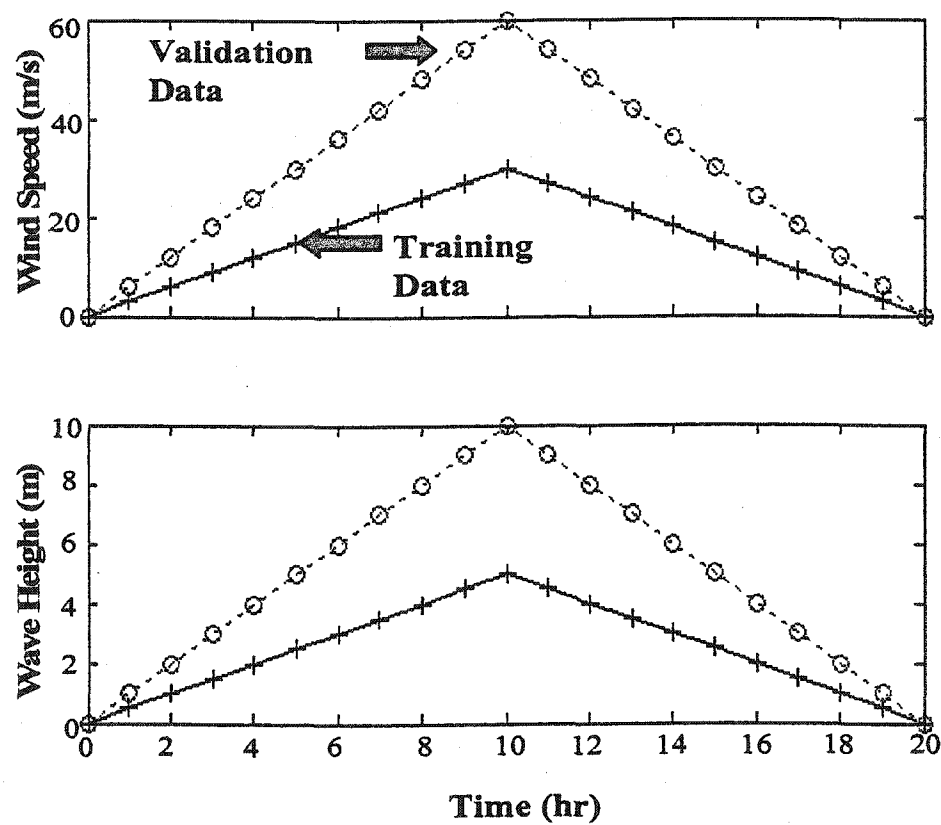


Fig. 4-1. Wind speeds and corresponding wave heights, which were generated by a linear equation with a constant slope to prepare training and validation data.

wind speed of 60 m/s for the validation data set was normalized as 1.17. But when using 70 m/s, it was normalized as 0.86. For global minimum wind speeds, 0 m/s was used for both the training and validation data sets.

Second, in order to understand the effects of maximum and minimum wind speeds on the pre-processing, local and global maximum and minimum wind speeds were tested, respectively. When local maximum and minimum wind speeds were used, both training and validation data sets were normalized between  $-1$  and  $+1$ . For instance, with a maximum input of 30 m/s for wind in the training set and a maximum input of 60 m/s for wind in the validation data set, both were changed to  $+1$ . On the contrary, if a global maximum wind speed (60 m/s) was used, the maximum input for the training set was changed to  $+0.5$ . But the maximum input for the validation set was changed to  $+1$ .

Third, in order to determine the effects of number of hidden neurons, comparative numbers of two and 45 hidden neurons were arbitrarily selected. Because the number of hidden neurons determines the number of parameters for ANN models, we can also observe how ANN prediction results can change with the number of model parameters.

#### 4.2.1. Proposed ANN Structure

In this experiment, the BPN and the scaled conjugate gradient learning algorithm were used. The number of iterations was arbitrarily selected as 25. Only one wind speed was used at one station as input. Thus, the number of input and

output neurons is one, respectively ( $m = 1$  and  $p = 1$ ). With regard to the number of hidden neurons, two and 45 ( $n = 2$  and 45) were used. Thus, the BPN structures of  $I_1H_2O_1$  and  $I_1H_{45}O_1$  were proposed for this study.

#### 4.2.2. Results and Discussion

When pre-processing was not used, the BPN structures of  $I_1H_2O_1$  and  $I_1H_{45}O_1$  did not predict the 10 m-maximum wave height perfectly but predicted only about 9 m at 25 iterations, respectively (Fig. 4-2a). Notice that the predicted wave height became 10 m when the number of iterations increased to 100 (Fig. 4-2b).

When a local maximum and minimum wind speed was used for the pre-processing, two BPN structures predicted only 5 m as the maximum wave height (Fig. 4-3). By contrast, when the BPN used a global maximum wind speed of 50 m/s or 70 m/s, both structures of  $I_1H_2O_1$  and  $I_1H_{45}O_1$  predicted the 10 m-maximum wave height exactly (Fig. 4-4). It was clear that the prediction results improved, and the BPN needed only 25 iterations when global maximum and minimum wind speeds were used.

Conclusions of the studies on the linear case for ANN models were the following. (1) Pre-processing should be used for ANN models, and (2) a global maximum and minimum wind speed should be used although the magnitude of these two selected values may not be critical for the linear case.

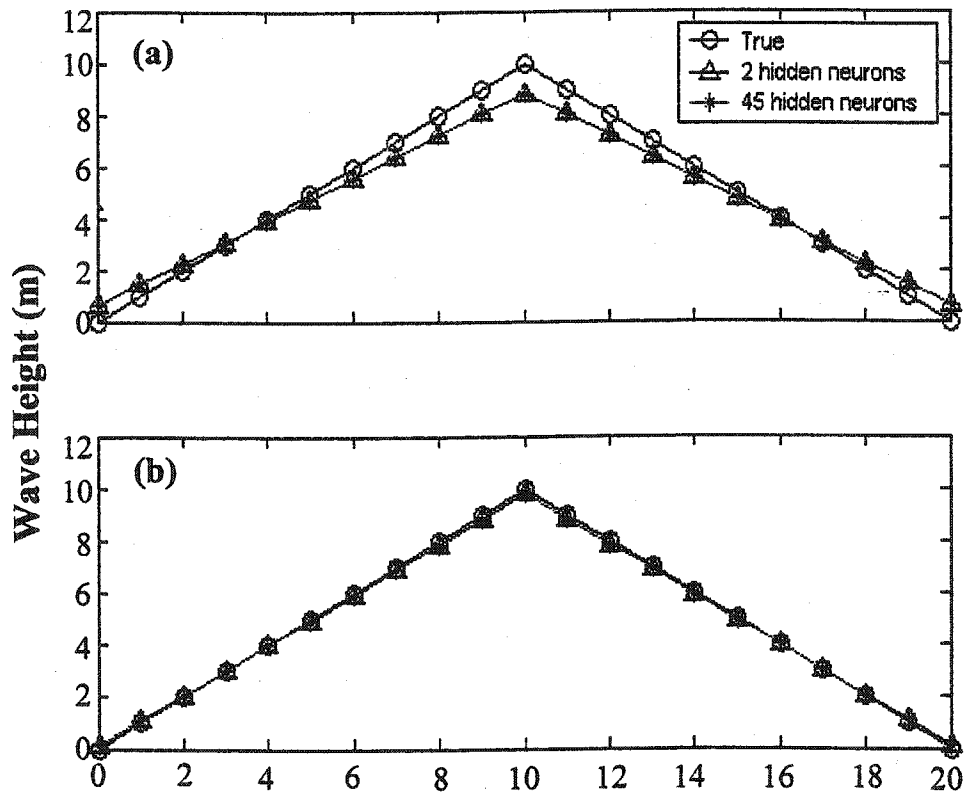


Fig. 4-2. Wave height predictions without pre-processing for the BPN with structures of  $I_1H_2O_1$  and  $I_1H_{45}O_1$ . (a) For 25 iterations and (b) for 100 iterations, respectively.

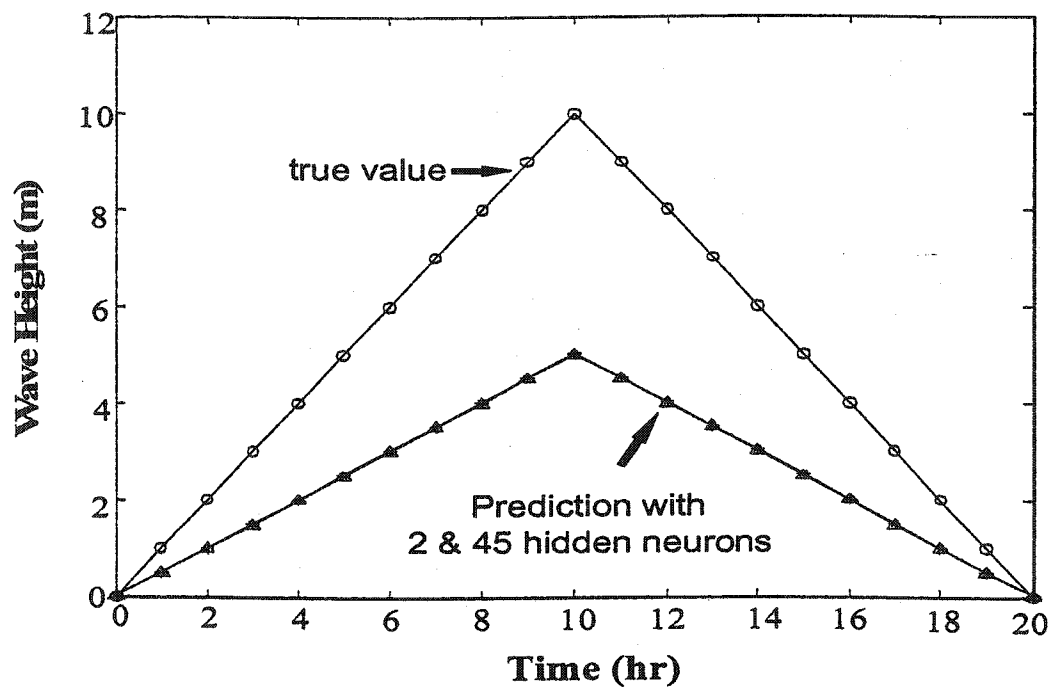


Fig. 4-3. An experiment to demonstrate the need for a global maximum and minimum wind speed for pre-processing. The predicted maximum wave height should be 10 m for the maximum wind speed of 60 m/s.



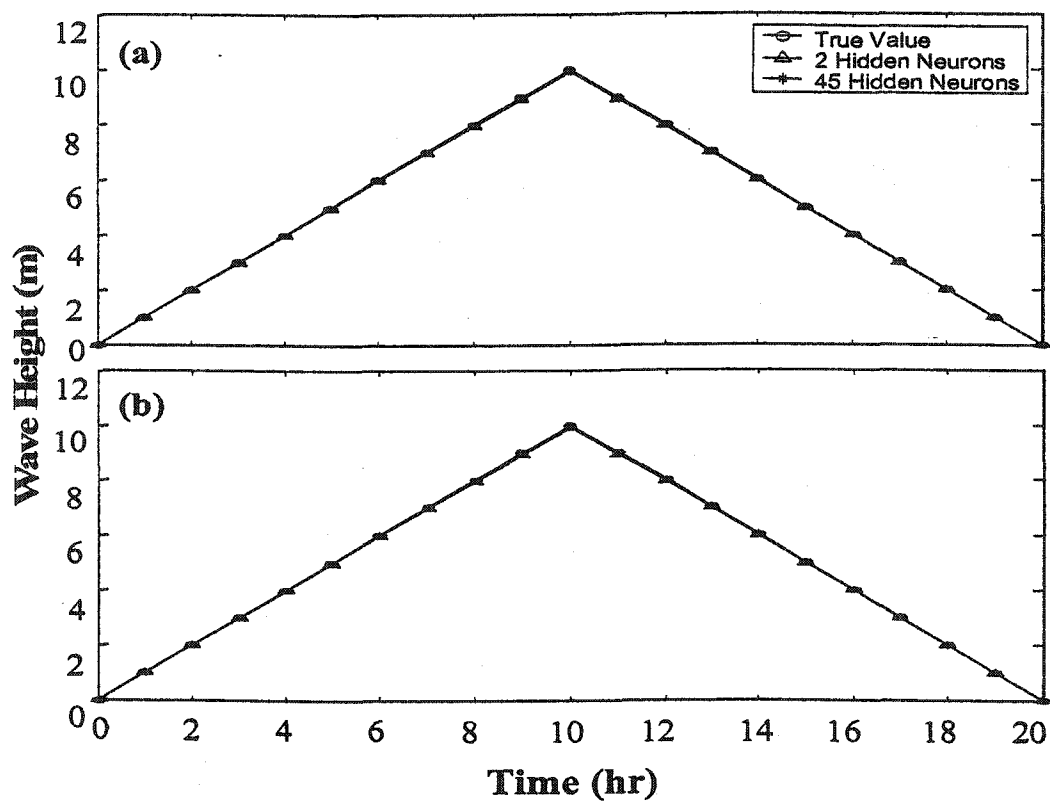


Fig. 4-4. Wave height predictions using the global maximum wind speed of 50 m/s (a) and 70 m/s (b). The Back-Propagation Network structures of  $I_1H_2O_1$  and  $I_1H_{45}O_1$  were used at 25 iterations.

### 4.3. Non-Linear Case (compared with SMB method)

This test was designed to check the feasibility of using an ANN for predicting non-linear responses. It is assumed that there are five stations for wave heights and one wind station to represent wind within the wind field (Fig. 4-5). Wave heights at these five stations are arbitrarily proportional to that given in station 1.

The objective of this test was to identify error sources in a simple non-linear wind-wave system. If prediction errors are due to improper ANN model structures and an insufficient number of training data, then prediction accuracy can improve when the optimum condition for the ANN model is found, and more training data are provided. However, if the reason is related to non-linearity and/or other unknown effects, it would be difficult to improve prediction accuracy, but at least it will indicate the suitability of ANN modeling.

For this reason, this test was designed to address: (1) the sufficient number of training data points, (2) the optimum number of hidden neurons and iterations, and (3) the optimum structure of an ANN model for simple wind-wave predictions.

For evaluating the efficiency of an ANN model, the correlation coefficient ( $r$ ) of 0.9 between reference and predicted values was used. That is to say, if the correlation coefficient was more than 0.9, it was considered that a correct model structure was found. In contrast, if  $r$  was less than 0.9, the ANN was assumed to have an insufficient number of training data and/or improper ANN structures. If large numbers of training data did not improve prediction accuracies, the ANN was tested by increasing the number of iterations and hidden neurons.

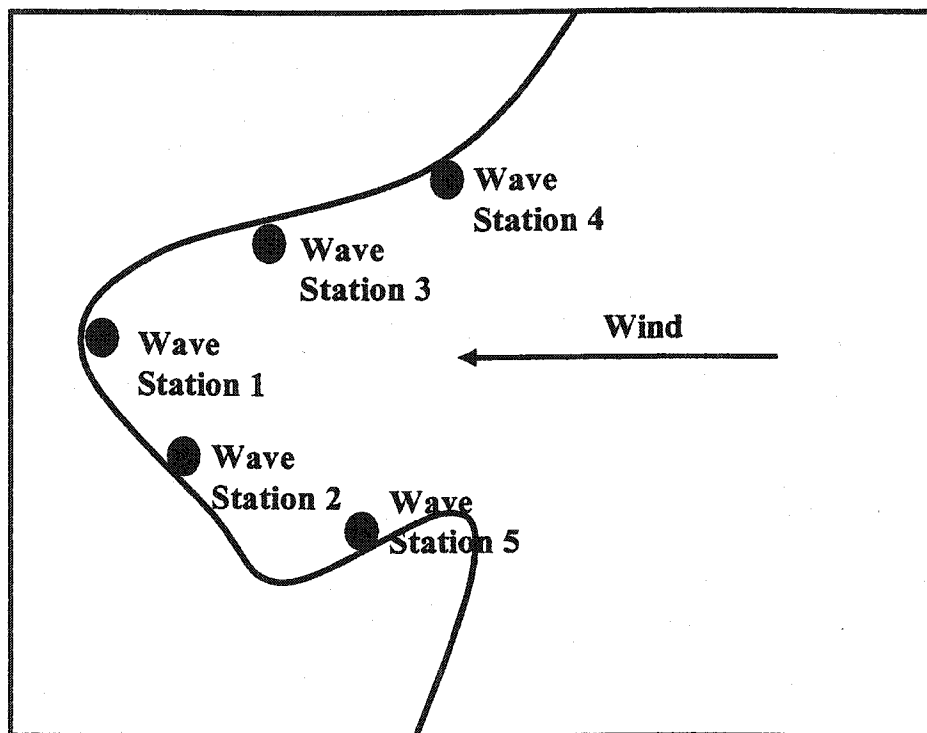


Fig. 4-5. An assumed location map with five imaginary wave stations. Wind fetch is ranked from high to low for station 1, 2, 3, 4 and 5. Station 5 has the least wind effects because it is located behind a headland.

### 4.3.1. Method

#### 4.3.1.1. Data Description

Non-linear wind-wave data were produced by using the SMB model developed for estimating significant wave heights and zero-crossing wave periods using wind speed, fetch, and duration. More details about the SMB were given in chapter 3.

For applying the SMB method, it was assumed that a representative fetch for this study area was 500 km. Wind speed was assumed to arbitrarily change every hour between 5 m/s and 27 m/s from zero to 799 hours. Thus, the total number of wind data points was 800, which consisted of eight events named from the left to right as No.1 to No.8 (see Fig. 4-6a). The corresponding significant wave height and zero-crossing wave period produced by the SMB are shown in Fig. 4-6b. The ranges of the wave height and wave period were 0.18 m-9.83 m and 1.67 seconds-12.52 seconds, respectively. Details of the simulated winds and waves used are given in Table 4-1.

Event one, three, and six were arbitrarily chosen as validation data to compare with prediction results for the ANN model. The sequence of these three events was randomly changed into event number 6-1-3 to make certain that the ANN did not memorize the same event sequence of 1-3-6. Thus, the total number of validation data points was 245.

For training data, five different data sets were prepared. For instance, the sequence of wind-wave events was randomly changed to 2-6-4-8-5-3-7-1. In order to

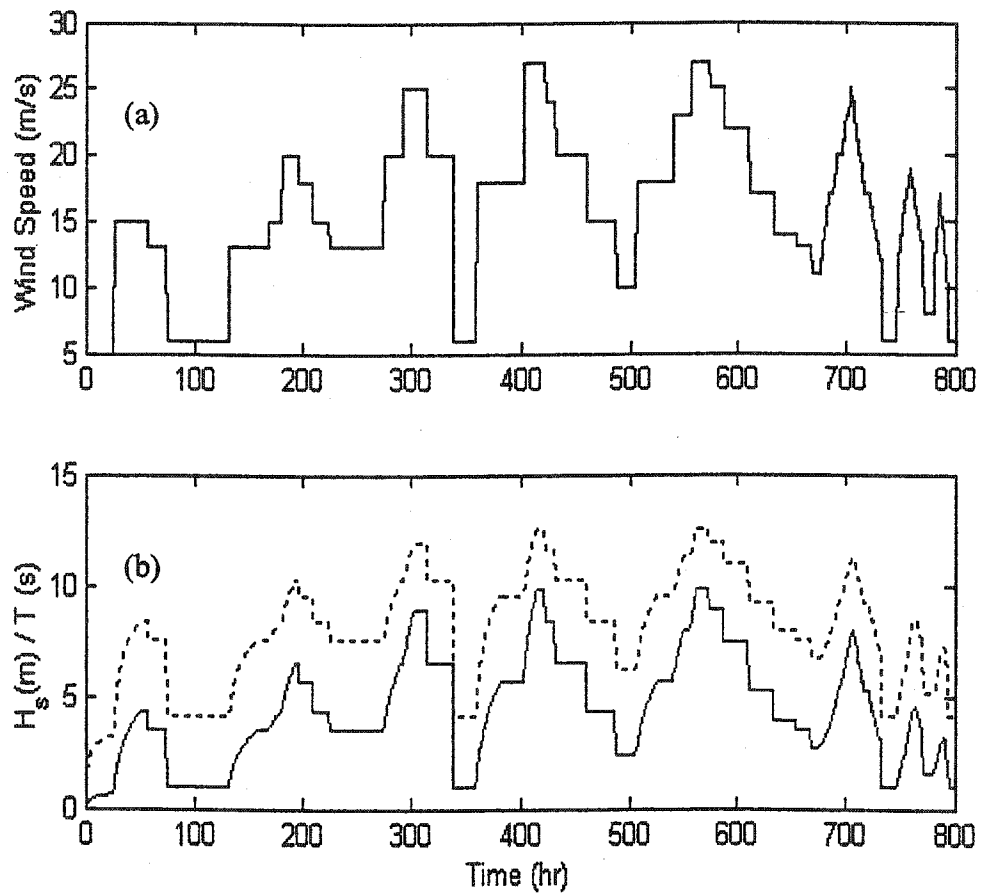


Fig. 4-6. Simulated eight wind (a) and wave events (e.g., wave height  $H_s$  and period  $T$ ) (b). Wind speed was changed from 5 m/s to 27 m/s, and corresponding waves were designated from left to right as No. 1 to No. 8.

Table 4-1

Wind and wave data sets for non-linear test

Type	Events	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	Total
Wind Speed (m/s)	Min.	5	6	6	6	10	6	6	6	5
	Max.	15	20	25	27	27	25	19	17	27
Wave Height (m)	Min	0.18	0.98	0.98	0.98	2.38	0.98	0.98	0.98	0.18
	Max	4.42	6.62	8.90	9.83	9.83	8.02	4.57	3.25	9.83
Wave Periods (m/s)	Min.	1.67	4.16	4.16	4.16	6.28	4.16	4.16	4.16	1.67
	Max	8.46	10.30	11.92	12.52	12.52	11.31	8.58	7.25	12.52
Data Points		80	170	100	140	180	65	40	25	800

determine how much training data are needed for accurate prediction, the number of training data was reduced as follows: (1) the eight events with the sequence of 2-6-4-8-5-3-7-1 were all used, which indicated that the training data had 100 % coverage of all validation data; (2) seven events with the sequence of 2-4-8-5-3-7-1 were used as the training data, omitting event 6 and providing 92 % coverage; (3) six events with the sequence of 2-4-8-5-3-7 were used, omitting events 1 and 6, and providing 82 % coverage; (4) five events with the sequence of the 2-4-8-5-7 were used, providing coverage of 69 %, and omitting events 1, 3 and 6; (5) only four events with the sequence of 2-8-5-7 were used, covering only 52 % of the validation data. Details of all training data sets are given in Table 4-2.

#### 4.3.1.2. Proposed ANN Structure

The Time Delay Neural Network (TDNN) and the scaled conjugate gradient learning algorithm were used in this test.

Wave heights and periods produced by the SMB model were assumed to be different at five stations (stations 1 to 5). The simulated wave heights and periods were modified by multiplying a factor of 1, 0.8, 0.6, 0.5, and 0.4 for station 1 to 5, respectively (Fig. 4-7) so as to add some complexity to the data set.

Wind speed and fetch were used as inputs to the ANN model. Thus, the number of input neurons was set as two ( $m = 2$ ). For predicting wave heights or periods at

Table 4-2

## Training and validation data sets

Training Data Set	Data Points (%)	Validation Data Set	Known Events (%)
8 events (2-6-4-8-5-3-7-1)	100	3 events (6-1-3; all trained)	100
7 events (2-4-8-5-3-7-1)	92	3 events (6-1-3; Event 1 and 3 trained)	66.6
6 events (2-4-8-5-3-7)	82	3 events (6-1-3; Event 3 trained)	33.3
5 events (2-4-8-5-7)	69	3 events (6-1-3; none trained)	0
4 events (2-8-5-7)	52	3 events (6-1-3; none trained)	0



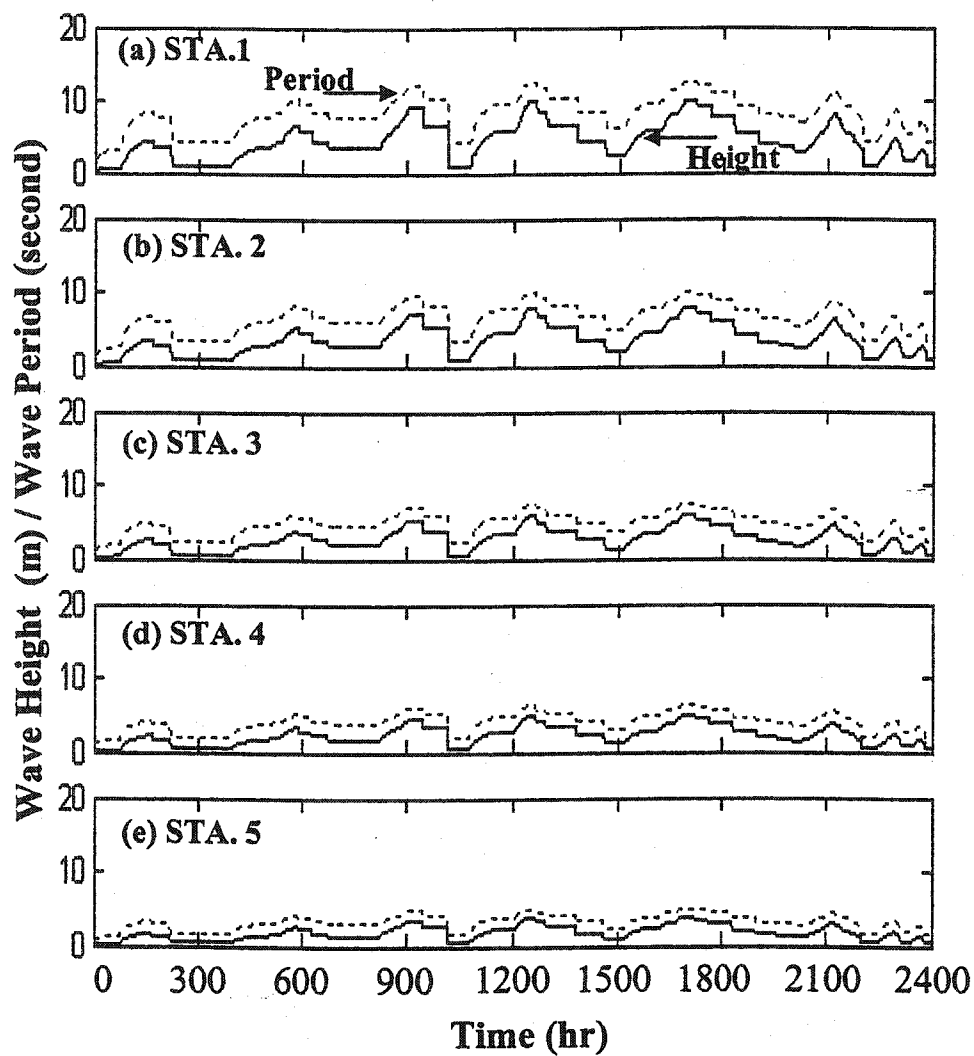


Fig. 4-7. Wave heights and periods at the five imaginary stations (See Fig. 4-5). Wave heights and periods produced by the SMB method were multiplied by factor 1 at station 1 (a), 0.8 at station 2 (b), 0.6 at station 3 (c), 0.5 at station 4 (d), 0.4 at station 5 (e), respectively, to simulate wind effects according to different wind fetch.

five stations, the TDNN needs five output neurons ( $p = 5$ ). The duration of time-delays was set as one-day because it takes time to generate a fully developed wave field, and the 500 km fetch is a long distance for waves to travel. In order to consider a one-day duration of time-delays, eight previous time steps as well as a current-time wind speed and fetch are used because the time interval for wave heights and periods was three hours. Thus,  $J = 9$ . However, there is no rule to determine the number of hidden neurons, and so, the number of hidden neurons must be determined by checking the TDNN performances. For this reason, the TDNN structure of  $I_{18}H_nO_5$  was proposed for this case study.

#### 4.3.2. Results and Discussion

For finding the optimum number of hidden neurons, 6, 10, 12, 15, 20, 30, and 50 hidden neurons were used, while the number of iterations was fixed as 500. The correlation coefficients between reference and predicted wave heights, corresponding to different numbers of hidden neurons, were compared. Eight events were used for training data, and number 6, 1, and 3 events were used as a validation data set.

Regardless of the number of hidden neurons, the correlation coefficients between reference and predicted wave heights were over 0.93 when 100% data coverage cases were used. This implies that the number of hidden neurons did not strongly affect the ANN prediction capability, at least when all the data are available.

The results are shown in Table 4-3.

Table 4-3

Mean Square Error (MSE) and correlation coefficient (r) between reference and predicted wave heights at different number of hidden neurons

Hidden Neurons	6	10	12	15	20	30	50
Iterations	500	500	500	500	500	500	500
MSE	0.02701	0.02757	0.02527	0.02344	0.02733	0.01225	0.01048
r	0.94	0.94	0.93	0.94	0.93	0.96	0.96

For this reason, 15 was arbitrarily chosen as the number of hidden neurons and used to test all different training data sets. The prediction results of wave heights and periods are given in Table 4-4 and 5. Figures 4-8 and 4-9 plot the true and predicted wave heights, and true and predicted wave periods, respectively, at the five wave stations. The correlation coefficients for wave heights and periods are given in Fig. 4-10. In general, the prediction results are satisfactory except that the TDNN did not recognize the rapid change of wave heights and periods between each event. For instance, reference values of wave heights and periods sharply decreased between each event and almost became zero at the end, while the predicted wave heights and periods gently decreased.

With regard to the number of iterations, prediction accuracy improved as the number of iterations increased for both wave heights and periods at a given number of hidden neurons. The correlation coefficient ( $r$ ) between the reference and predicted wave heights and periods for five training data sets were more than 0.95 when the number of iterations was over 500. When using 200 iterations, the similar correlation coefficients were still 0.90 or more (over 80 % variability of the predicted values can be accounted for by references values). This indicates that the TDNN model was not strongly affected by the number of hidden neurons when the number of iterations was large.

For understanding of the difference in patterns among training data sets, the correlation coefficient of the weight matrix between training and validation data sets was compared. If the training data set has the same features as the validation data set,

Table 4-4

Mean Square Error (MSE), correlation coefficient when different number of hidden neurons and iterations were used for the Prediction of wave height.

Events	Data Points	Hidden Neurons	Iteration	MSE	Correlation Coefficient
8 events	100 %	15	100	0.06829	0.86
			200	0.04421	0.91
			300	0.02647	0.93
			500	0.02344	0.94
			600	0.01710	0.96
7 events	92 %	15	100	0.06853	0.85
			200	0.04320	0.90
			300	0.02024	0.92
			500	0.01956	0.96
6 events	82 %	15	100	0.07297	0.86
			200	0.03919	0.92
			300	0.03161	0.93
			500	0.02031	0.95
			600	0.01774	0.95
			700	0.01532	0.96
5 events	69 %	15	100	0.06373	0.86
			200	0.03094	0.92
			300	0.02516	0.92
			500	0.01812	0.93
			600	0.0618	0.94
			700	0.01462	0.95
4 events	52 %	15	100	0.05743	0.90
			200	0.03322	0.92
			300	0.02710	0.92
			500	0.01826	0.93
			600	0.01461	0.94
			700	0.01149	0.96

Table 4-5

The Mean Square Error (MSE), correlation coefficient when different number of hidden neurons and iterations were used for the prediction of wave period.

Events	Data Points	Hidden Neurons	Iteration	MSE	Correlation Coefficient
8 events	100 %	15	100	0.07357	0.87
			200	0.04954	0.91
			300	0.03429	0.93
			500	0.02267	0.95
7 events	92 %	15	100	0.07349	0.86
			200	0.05206	0.93
			300	0.03891	0.94
			500	0.02292	0.95
6 events	82 %	15	100	0.07691	0.93
			200	0.04334	0.93
			300	0.03507	0.94
			500	0.02479	0.95
5 events	69 %	15	100	0.05648	0.92
			200	0.04178	0.93
			300	0.03032	0.94
			500	0.02406	0.95
4 events	52 %	15	100	0.09957	0.89
			200	0.06786	0.93
			300	0.0555	0.94
			500	0.0433	0.94
			600	0.03945	0.95

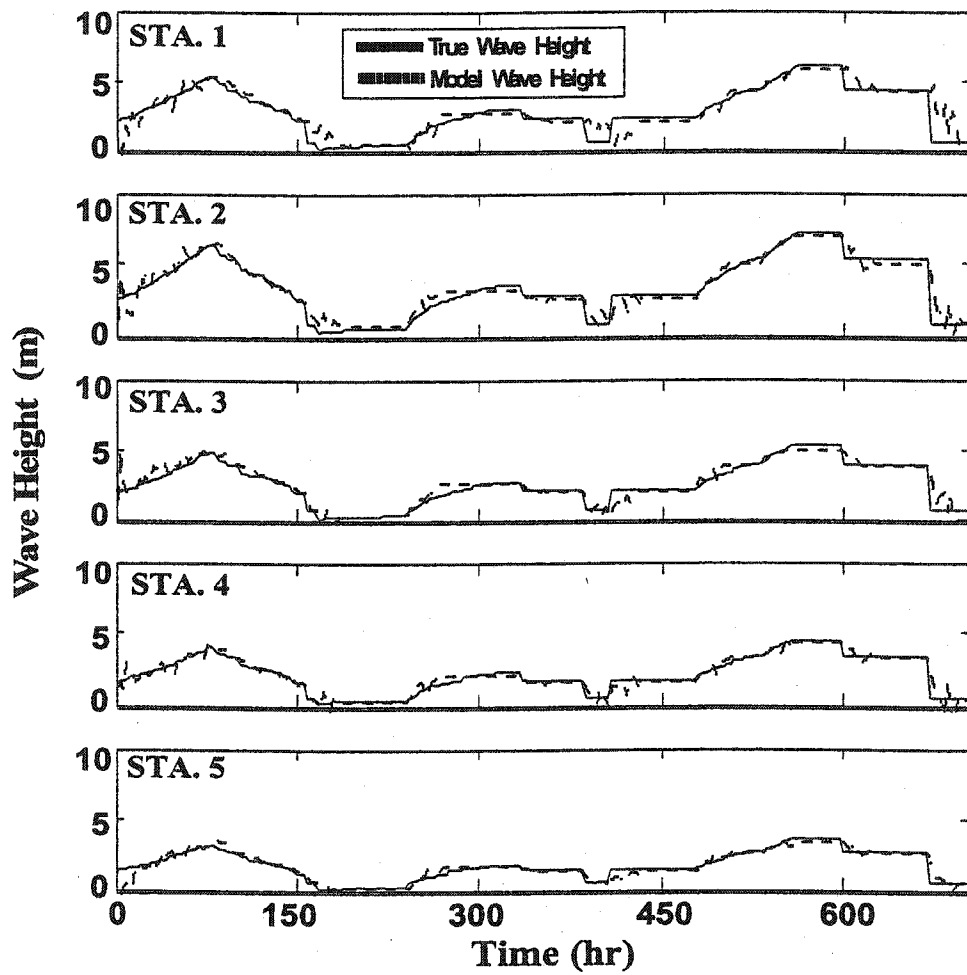


Fig. 4-8. Comparison of true and predicted wave heights using the Time Delay Neural Network with a structure of  $I_{18}H_{15}O_5$  at 700 iterations when four training events were used.

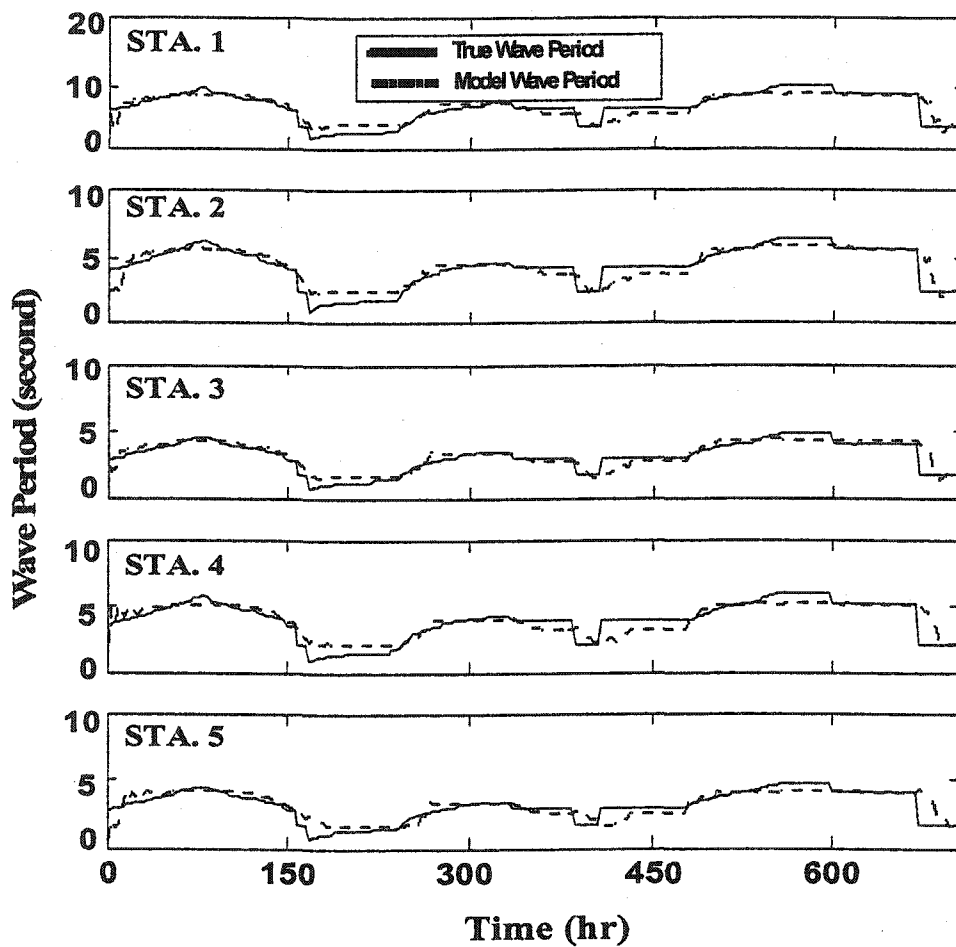


Fig. 4-9. Comparison of true and predicted wave periods using the Time Delay Neural Network with a structure of  $I_{18}H_{15}O_5$  at 600 iterations when four training events were used.



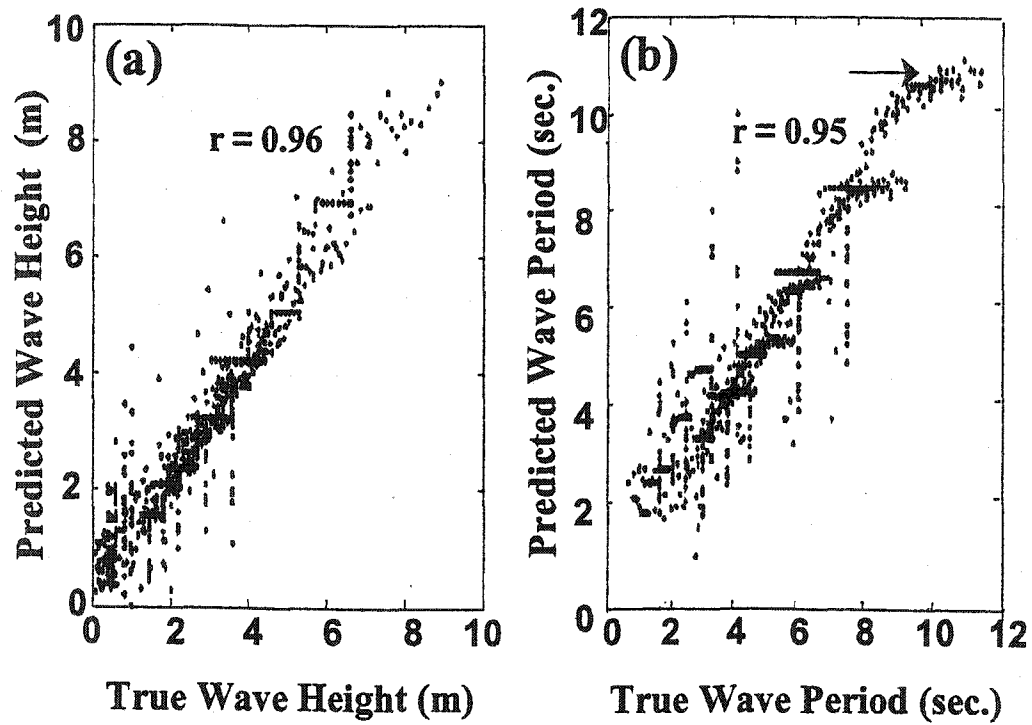


Fig. 4-10. Correlation coefficient between reference (true) and predictions when four training events were used. (a) For wave height and (b) for wave period. Notice that predicted maximum wave periods are obviously less than true maximum periods compared with wave heights.

weight matrices for these two data sets should be identical. If not, the weight matrix produced with a large number of training data should be closer to validation data than that produced with a smaller training data set.

Weight matrices of two training data sets of 552 and 416 points were compared with those of another training data set of 800 points. The number of weights between the input and hidden neurons was determined by the number of hidden neurons ( $n = 15$ ) and inputs ( $m = 2J = 18$ ). Thus, the total number of elements in the weight matrix was 270. The number of iterations tested started from zero and increased to 100 with an increment of 10 iterations for wave heights and from zero to 200 with an increment of 20 for wave periods.

Figure 4-11 shows the change of the correlation coefficient of weight matrices for wave height and period with increasing number of iterations. The correlation coefficient of the weight matrix among three data sets was almost identical for both wave heights and periods. However, when examining these figures, the correlation coefficient of the weight matrix between 800 and 552 data points was larger than that between 800 and 416 data points for both wave heights and periods as the number of iterations increased.

For these reasons, it can be concluded that: (1) the TDNN model was not strongly affected by the number of hidden neurons when the number of iterations was sufficiently large; (2) the prediction accuracy improved as the number of iterations

increased for both wave height and period at a given number of hidden neurons; (3) the prediction accuracy improved with an increase in the number of training events.

Despite a small change in prediction accuracies, the prediction results of the TDNN were satisfactory in this non-linear case. Such good predictions may result for the following reasons: (1) there is only one pattern for the ANN to recognize, *i.e.*, equations 3-1 and 3-2, (2) for the application of the SMB model, it was assumed that only one-wind field affects all five-wave stations, (3) only a single fetch of 500 km was used to produce the reference data. In other words, the simulated wave heights increase simply with periods because it used only one 500 km fetch. In contrast, for observed waves, the wave heights do not always increase (Fig. 4-12). This implies that wind-wave systems are more complicated in nature, and may include several wind fields at the same time.

#### 4.4. Conclusions

The TDNN predicted simulated linear and non-linear wind-waves accurately when it was appropriately structured and trained, although it slightly underestimated a large wave period. This indicates that an ANN technique can be used to predict wind-waves in a real environment, however the prediction results are expected to be somewhat inaccurate because of the complexity of wind-waves.

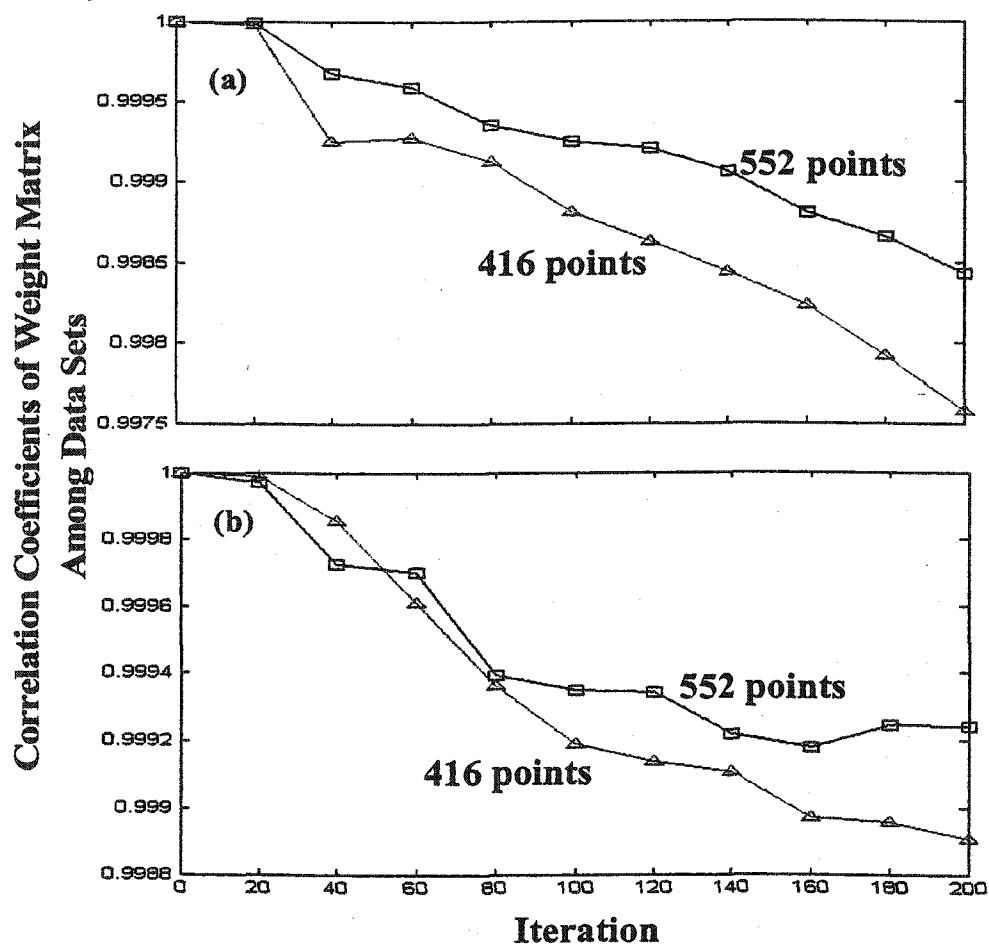


Fig. 4-11. Correlation coefficients of weight matrix among three data sets. (a) For wave height and (b) for wave period.

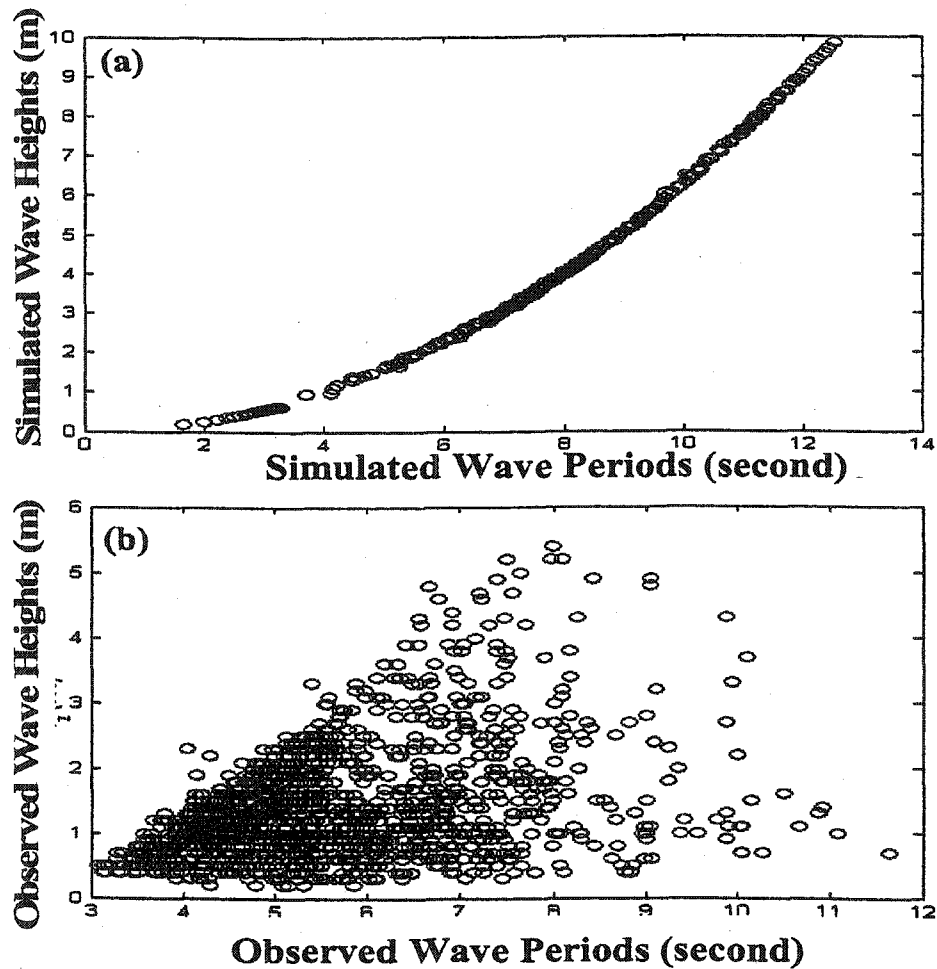


Fig. 4-12. Relationship between wave heights and periods. (a) shows simulated waves produced by the SMB method. Wave heights simply increase with wave periods. (b) shows observed waves at five stations (44007, 44009, 44025, 44013, and 41009) in February, 1998, 1999, and 2001. In general, observed wave heights increase as wave periods increase. However, many large wave heights occurred at small wave periods. This indicates that the relationship between wave height and period is more complicated in real wind-wave generation.

## CHAPTER V

### ANN MODEL FOR WINTER-STORM WAVES

#### 5.1. Introduction

Winter storm is a general name for a strong wind system during cold weather. Wind fields of winter storms appear over a large domain and usually move slowly. Wind speeds are relatively evenly distributed throughout the domain. For these reasons, storm waves may have enough time to generate a fully developed sea.

Because of the large wind field for winter storms, the affected area is also large, and the storm waves can be observed along the entire East Coast of the U.S (Fig. 5-1). Winter-storms with low air pressures usually come from the north and confront winds blowing eastward, forming an extensive and slow moving front. Wave heights increase nearly simultaneously along the East Coast of the U.S. as the storms come from the north, and the temporal-variations of wave heights are similar along the East Coast except at station 41009 (see wave heights observed at the five stations Fig. 5-2).

In this chapter, predictions of storm-wave height and period are given. The study area is given in section 5.2; data description is in section 5.3; a proposed ANN structure is provided in section 5.4; the development of a time-delay mechanism is given in section 5.5; selection of an efficient learning algorithm is in section 5.6;

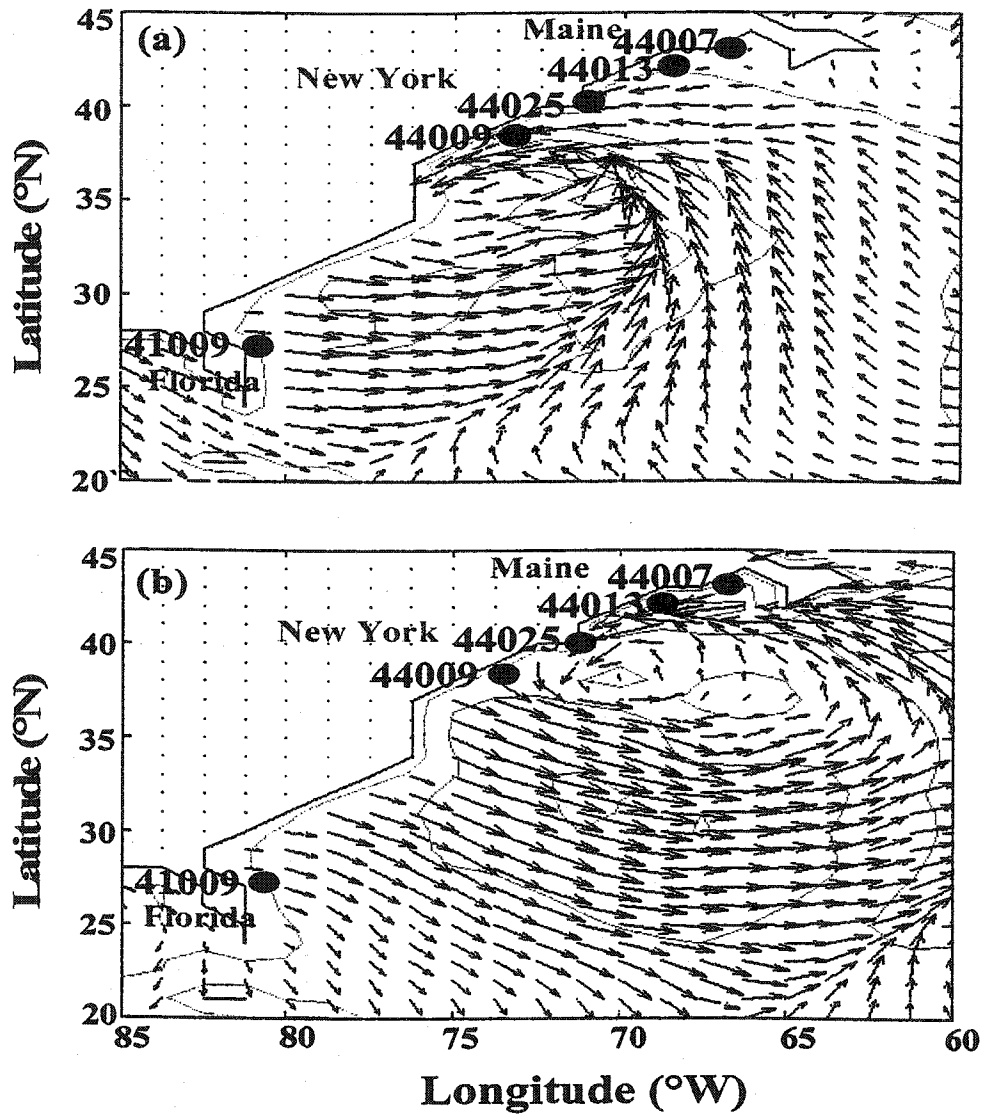


Fig. 5-1. Storm-wind field over the northwest Atlantic Ocean. (a) On February 24, 00:00 GWT, 1998 and (b) on February 25, 00:00 GWT, 1998. Circles indicate the five wave stations along the East Coast of the U.S.

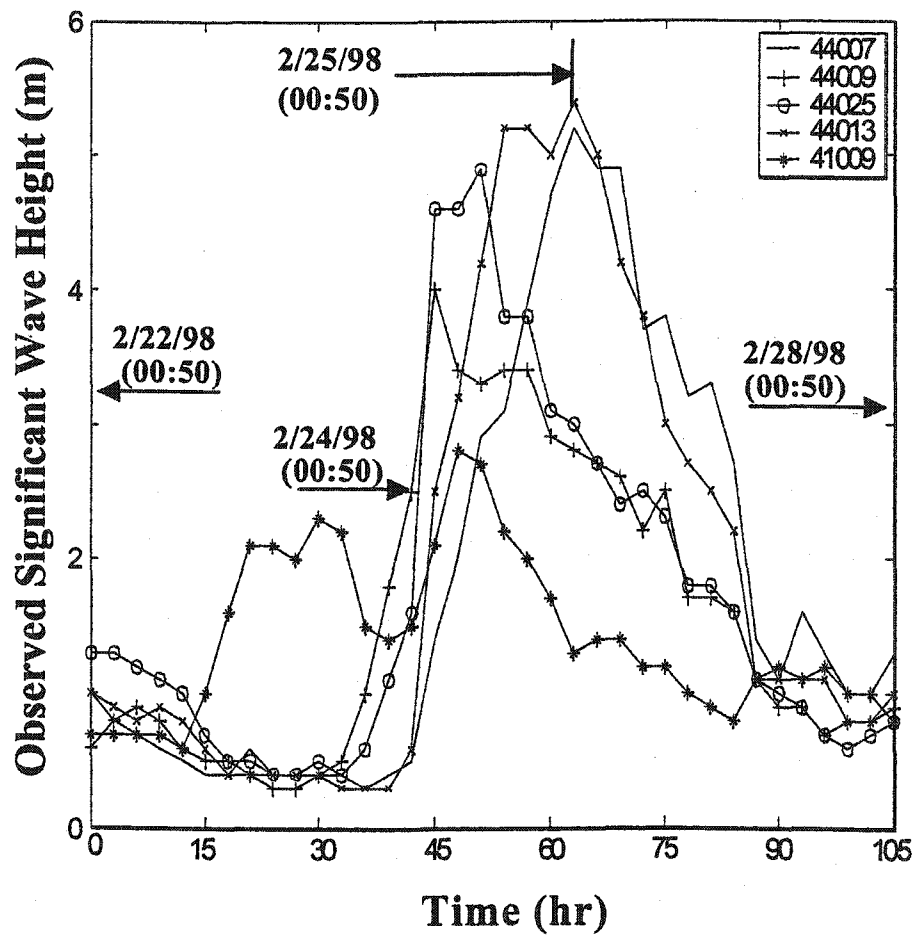


Fig. 5-2. Observed significant wave height at the five wave stations (44007, 44009, 44025, 44013, and 41009) along the East Coast of the U.S. from February 22 to 28, 1998.



results of significant wave height prediction and discussion with three-layered Back-Propagation Network (BPN), Elman Recurrent Network (ERN), and Time Delay Neural Network (TDNN) are given in section 5.7; results of zero-crossing wave period prediction with a three or four-layered TDNN are in section 5.8; and conclusions are in section 5.9.

## 5.2. Study Area

The East Coast of the U.S. was chosen as the study area to develop an artificial neural network model for winter-storm wave prediction. The National Data Buoy Center (NDBC) from the National Oceanic and Atmospheric Administration (NOAA) has more than 10 wave stations along the East Coast of the U.S. Among those stations, seven are located offshore (44007, 44013, 44025, 44009, 44014, 41004, and 41009). However, only the following five offshore stations (44007, 44009, 44025, 44013, and 41009; see Fig. 5-3), which extend from Florida to Maine, were chosen because they have complete wave records during the period of our study. More information on the five offshore wave stations is given in Table 5-1.

As shown in Fig. 5-1, the area dominated by winter-storms may cover a few hundred kilometers in the on-offshore direction as well as the shore-parallel direction. For this reason, wind information over the Northwest Atlantic Ocean should be

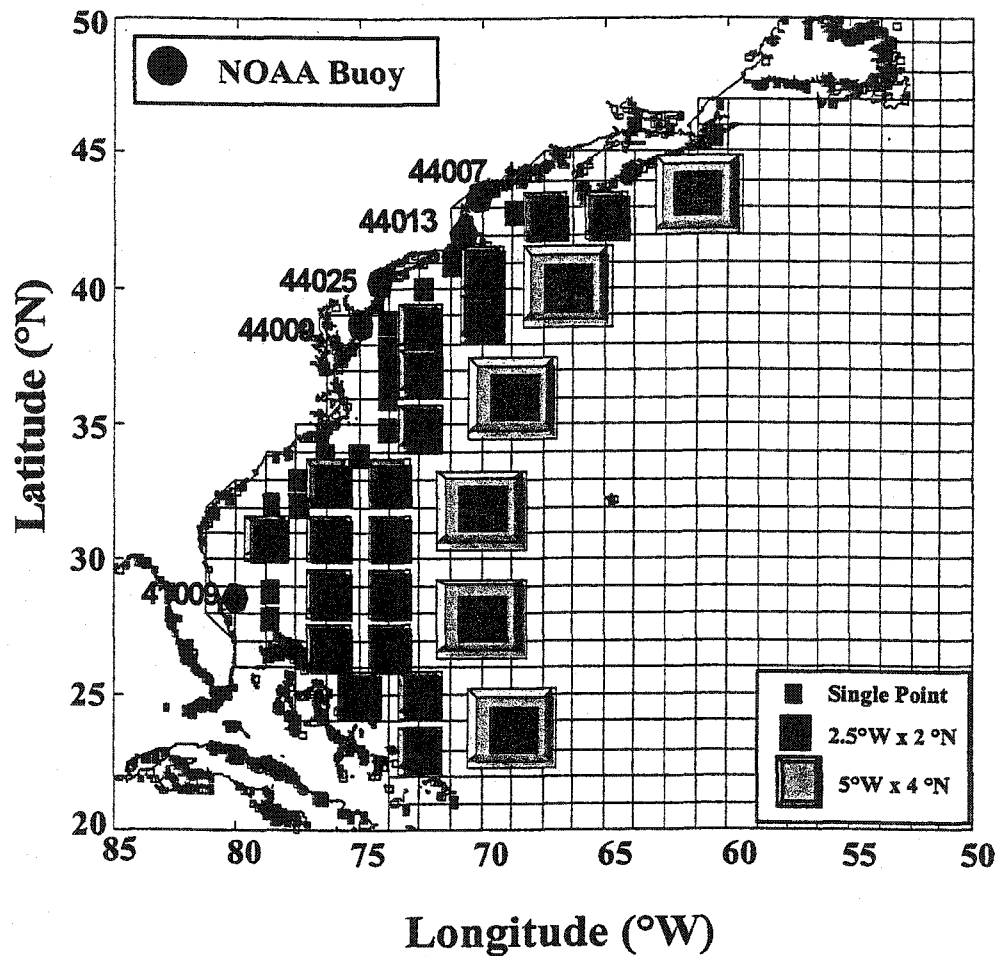


Fig. 5-3. Location map of 40 wind stations (squares) and five wave station (circles). Wind stations are regrouped as three types: (1) 15 single-point stations, which represent the wind information in each grid; (2) 19 middle-size (2.5°E x 2°N) stations, which represent the average of nine single point stations; (3) Six large-size (5°E x 4°N) stations, which represent the average of 25 single-point stations.

Table 5-1

Information on the selected five wave stations

Station	Location		Water Depth (m)	Area
44007	43°-31'-53" N	70°-08'-40" W	18.9	Portland
44013	42°-21'-14" N	70°-41'-29" W	55.0	Boston
44025	40°-15'-01" N	73°-10'-00" W	40.0	Long Island
44009	38°-27'-49" N	74°-42'-07" W	28.0	Delaware Bay
41009	28°-30'-01" N	80°-10'-03" W	42.0	Canaveral

considered for ANN model inputs. This constitutes a spatial domain from approximately 23°N to 45°N and 58.75°W to 80°W.

### 5.3. Data Description

#### 5.3.1. Wind

The National Climate for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) provided 'reanalyzed' atmospheric data for public use. They use data from land, surface, ships, aircrafts, satellites, other data sources that have good quality, and an assimilation system to calculate the best estimated global atmospheric data. Different types of output are being created and archived every year since 1997 to meet different user's needs.

All the reanalyzed meteorological data are stored in 'grib' format, which is a highly compressed binary file. Data are archived by month, and the global domain stretches from 0°E to 360°E and from 78°S to 78°N. The resolution is 1.25° in east/west directions and 1° in north/south directions, providing a total number of points on longitudinal and latitudinal directions of 288 and 157, respectively. Thus, the grand total number is 45,216 points for each 'reanalysis.'

The reanalyzed wind data have an interval of three hours. Meteorological data from 1998 to 2001 were downloaded and the software 'wgrib' (which was developed by Ebisuzaki at the NCEP and is available on the internet at <ftp://wesly.wwb.noaa.gov/pu/wgrib>) was used to get the wind velocity components u and v. Other available

information was not used. The wind data in 2000 was not used here because station 44025 did not have wave records during February 2000.

A few winter-storms, which occurred from February 14, 1998 to March 1, 1998, were chosen as one of the data sets. Not all the wind data between February 14 and March 1 were used. Only those wind data that represent a winter storm, from beginning to the end, were used. Thus, 78 data points were collected (Table 5-2). The second winter storm season, which occurred from February 1, 1999 to March 1, 1999, was selected as the second data set with 218 data points. A third winter storm season, from February 1 to 28, 2001, was selected as the third data set with 185 data points. Among the three data sets, storm events in 1999 were used as a training set because they have more data points than those in either 1998 or 2000. Later, the data from 2001 were added for training to improve the prediction.

It is well known that waves generated far away can affect the observation at each wave station along the coast. In order to consider long-distance waves, wind stations were regrouped into three types according to the distance from the east coast of the U.S.: (1) 15 nearby wind stations, which used wind information on each grid directly; (2) 19 middle-range wind stations, which represented the average of nine grids (the size of each wind station =  $2.5^{\circ}\text{W} \times 2^{\circ}\text{N}$ ); and (3) Six long distant wind stations, which represented the average of 25 grids (the size of each wind station =  $5^{\circ}\text{W} \times 4^{\circ}\text{N}$ ). Thus, the total number of wind stations was 40; see Fig. 5-3.

For wind information, either u and v wind components or wind speed and direction can be used. It was not clear at the beginning of this study which choice

Table 5-2

Summary of the three selected data sets

Type	Data Set 1	Data Set 2	Data Set 3
Span	2/14/98 ~ 3/1/98	2/1/99 ~ 3/1/99	2/1/01 ~ 2/28/01
Number of Data Point	78	218	185

was a better one for inputs. For this reason, the effects of each choice were tested to find the best choice of wind information for ANN wind-wave prediction.

Theoretically, there were two problems in using wind speed and direction were used as ANN inputs. First, the unit and magnitude of absolute numerical values in wind direction were much larger than wind speed. For instance, wind speed may change from 0 m/s to 30 m/s, and wind direction may change for  $0^\circ$  to  $360^\circ$ . The maximum difference between the wind speeds and directions are 30 and 360, respectively. In this case, an ANN may view the change of 360 in wind direction as more important. Thus, it will put more weight on the change of 360 in wind direction than on the change of 30 in wind speed. Another critical problem was the difference between numerical and physical meanings of wind direction. For instance, there is no difference of wind directions between  $0^\circ$  and  $360^\circ$  from a physical point of view. In ANN, however, it recognizes a difference of 360 numerically. In this case, the ANN model will put more weights on wind direction changes, resulting in an incorrect prediction on weight height.

When using wind speed and direction as inputs for two ANN models (BPN and ERN), the above effects could be observed in the prediction results. For instance, the predicted wave heights had negative values and rapidly changed between consecutive wave heights at all wave stations. For this reason, u and v wind components only were used for ANN inputs in the rest of the studies.

### 5.3.2. Wave

The National Data Buoy Center (NDBC) has provided wave records since 1982 for station 44007, 1984 for station 44013 and 44009, 1975 for station 44025, and 1988 for station 41009. All the above five stations are along the east coast of the U.S. This wave information is now available on the internet (<http://www.ndbc.noaa.gov/rnd.shtml>). Significant wave height and zero-crossing wave period were extracted from the NDBC data set for the same periods shown in Table 5-2. These data sets are required for training and validation of an ANN model. More information about the five wave stations is given in Table 5-1.

### 5.4. Proposed ANN Structure

The purpose of this study was to predict storm-generated wave height and period at the five stations (*e.g.*, 44007, 44013, 44025, 44009, and 41009) along the East Coast of the U.S. Because the generation of wave height and period was physically different (see section 3.3.2), the development of wave height prediction was separated from the wave period prediction. Thus, the number of output requirements for each case is five ( $p = 5$ ).

The number of ANN inputs was determined by the number of wind stations (*i.e.*, 40 stations), and *u* and *v* wind components. Thus, the total number of inputs was 80 ( $m = 80$ ) if there was no time-delay.



However, there was no rule to determine the optimum number of hidden neurons,  $n$ . The number of hidden neurons has to be determined by checking the ANN performance. For this reason, an ANN structure of  $I_{80}H_nO_5$  was proposed in this study where time-delay was not considered.

### 5.5. Time Delay

In Fig. 5-3, the longest wind field selected in this study is about 10 degrees (roughly 1,100 km) from west to east and 18 degrees from south to north (roughly 2,000 km). Assuming all the waves observed at the selected five wave stations were generated within this area, the next step is to consider how to select the 'time-delay' for ANN modeling. It is well known that any wind field requires time to develop large waves. This is a well-known factor called 'duration.' For instance, given a strong wind but short duration, the wave field will not be fully developed, and large waves cannot be produced. This is an indication that a time-delay mechanism is needed for an ANN to better predict the wind-wave relationship.

Previous studies (*e.g.*, the SMB method) indicate that the higher the wind speed, the shorter the duration for a wind-wave system to fully develop. For instance, according to the SMB model, a 50 m/s wind over a fetch of 1,000 km will require 22 hours to fully develop to a wave height of 27.7 m and period of 20.9 seconds. When the wind speed is reduced to 20 m/s over the same 1,000 km fetch, it takes 38 hours to become a fully developed sea with wave height of 8.1 m and period of 11.5 seconds.

This example indicates that it is impossible to have a single time-delay to mimic the different duration encountered in winter storms.

There is another factor that may affect the ANN wind-wave predictions. For a large domain with many different wind fields, waves could be generated from each wind field and move into another wind-wave system, and all of them will reach the observation station with different attenuations. How to simulate this process is another challenge. Assuming that one single time-delay can handle the real wind-wave generation in a large domain may be an over-simplifying assumption. However, it is worth trying, and the following attempts are based on this assumption.

For testing the need of a time delay, two algorithms (BPN and ERN) with the scaled conjugate gradient learning algorithm were used with pre-processing. For this study, a 24-hour duration was chosen to consider the required time-delay for wind-wave generation and compared to a zero-hour duration. Two, 10, and 45 hidden neurons were used ( $n = 2, 10, \text{ and } 45$ ) to observe the difference between a small and large number of hidden neurons. The number of inputs was 80 ( $m = 80$ ) because  $u$  and  $v$  wind components at 40-wind stations were used.

For considering a 24-hour time-delay, wind information from eight previous time steps as well as a current time was used because wind and wave data were available every three hours. Hence, the number of inputs changes to  $m \cdot J$  with  $m = 80$  and  $J = 9$ .

Hence, the BPN and ERN structures are  $I_{80}H_2O_5$ ,  $I_{80}H_{10}O_5$ , and  $I_{80}H_{45}O_5$  for no time-delay ( $J = 1$ ), and  $I_{720}H_2O_5$ ,  $I_{720}H_{10}O_5$ , and  $I_{720}H_{45}O_5$  for 24-hour time-delay ( $J =$

9). The prediction results were better when a 24-hour time delay and 45 hidden neurons were used, although the predicted wave heights still can occasionally be negative. The above results indicate that a 24-hour time-delay should be used in further studies.

### 5.6. Selection of Learning Algorithm

In order to select a better learning algorithm, two supervised learning algorithms were tested: (1) the Gradient Descent with a Variable Learning Rate and Momentum (GDX) and (2) a Scaled Conjugate Gradient (SCG) which was developed to overcome the drawbacks of the gradient descent rule. For more information on the GDX and SCG, see sections 2.6.5 and 2.6.6, respectively. The performances of these two learning algorithms were compared in terms of the Mean Square Error (MSE) for two rather simple ANN schemes: BPN and ERN without time delay. It was assumed that the better-performing learning algorithm would also and consistently perform better for other ANN schemes, *e.g.*, TDNN. The number of inputs and output requirements was 80 and 5 ( $m = 80$  and  $p = 5$ ). For training data, 218 data points from 1999 were used.

The number of hidden neurons was arbitrarily selected as 60 ( $n = 60$ ), and the number of iterations was set as 100, 500 and 1000, respectively, to compare the computing time. Because this test was tried in the very beginning of this study, the importance of pre-processing was not recognized, and the BPN and ERN did not use

pre-processing. Thus, BPN and ERN structures of  $I_{80}H_{60}O_5$  were established. No difference in computing time was observed between the BPN and ERN.

For comparing the MSE between observed and predicted wave heights, the number of hidden neurons was increased from 10 to 130 with an increment of 10 hidden neurons between each trial for both the GDX and SCG learning algorithms. The number of iterations was fixed at two arbitrarily selected values: 500 and 1000, respectively. Thus, the BPN and ERN structures of  $I_{80}H_nO_5$ , where  $n = 10 \rightarrow 130$ , were established.

In general, the MSE of the GDX and SCG decreased as the number of iterations increased for the combination of BPN with GDX, BPN with SCG, and ERN with GDX. At the beginning of this experiment, the BPN and ERN with the GDX were tested first, and the results of the MSE were only slightly different between the two ANN schemes. For this reason, it was assumed that the MSE for the ERN with SCG would be similar to that for the BPN with SCG, and only the BPN with the SCG was tested later. The final result was that the MSE for BPN with SCG was less than that of BPN and ERN with GDX (Fig. 5-4). For this reason, the SCG learning algorithm was selected for use used in further study.

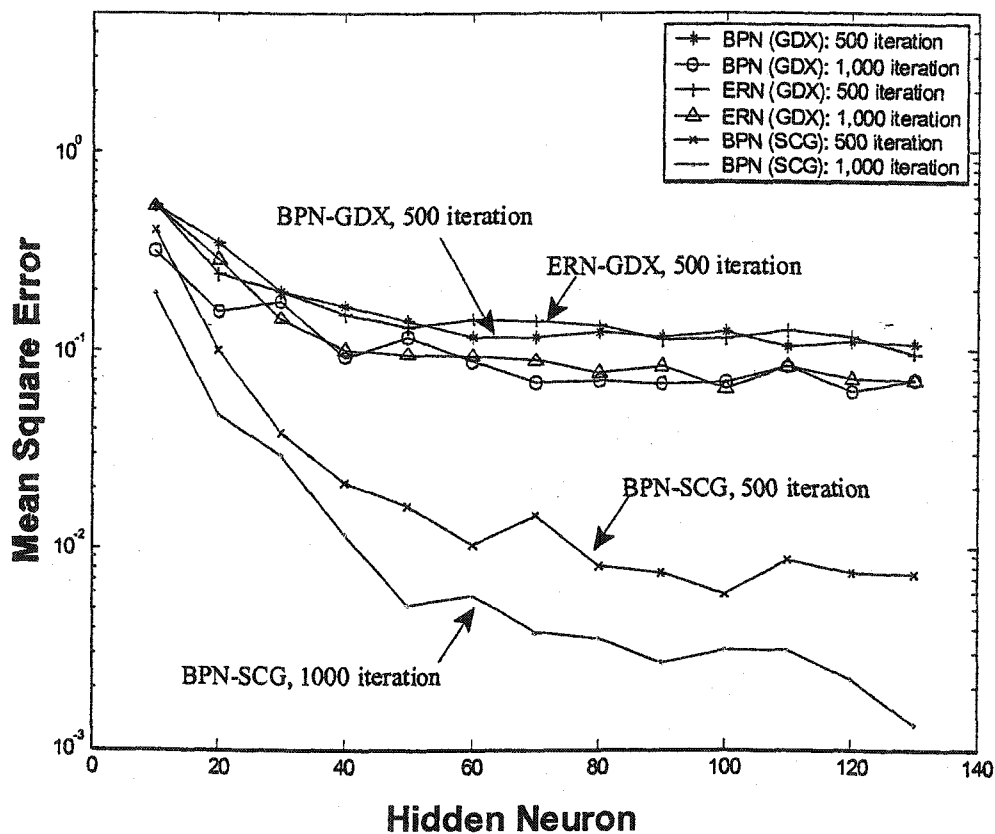


Fig. 5-4. Comparison of Mean Square Error between the Gradient Descent with a Variable Learning Rate and Momentum (GDX) and the Scaled Conjugate Gradient (SCG) for the Back-Propagation Network (BPN) and Elman Recurrent Network (ERN) with a structure of  $I_{80}H_{10-130}O_5$ . The number of training data was 218 points.

## 5.7. Results of Significant Wave Height Prediction and Discussion

### 5.7.1. Three-Layered BPN and ERN

Although the BPN algorithm has no time-delay capability, and the ERN algorithm has a short time memory (one time-delay), time-delays were used in an artificial manner. That is to say, by using wind data from every three hours for a 24-hour period ( $J = 9$ ) as inputs, instead of from only one time ( $J = 1$ ), it was assumed that time-delay information was incorporated into the ANN. This is a rather naive approach, but might produce some insight into the importance of a time-delay. Thus, the simple BPN and ERN algorithms were used first.

The observed maximum and minimum values of  $u$  and  $v$  wind components were 17.8 m/s and -19.8 m/s for the training data set and 17.8 m/s and -19.1 m/s for the validation data set, respectively. For this reason, global maximum and minimum were selected as 20 m/s and -20 m/s for pre-processing.

Figure 5-5 shows the learning curves for the BPN and ERN using the training data set from 1999 winter storms with 218 data points. The number of hidden neurons increased from one to 10 at an increment of one, and the number of iterations increased from 10 to 100 at an increment of one iteration between each trial.

The MSE did not further improve after 20 iterations when the BPN model used one or two hidden neurons. Similar results were observed when the ERN model used one, two, or three hidden neurons. This suggests that a small number (less than three) of hidden neurons are not sufficient for wave height prediction.

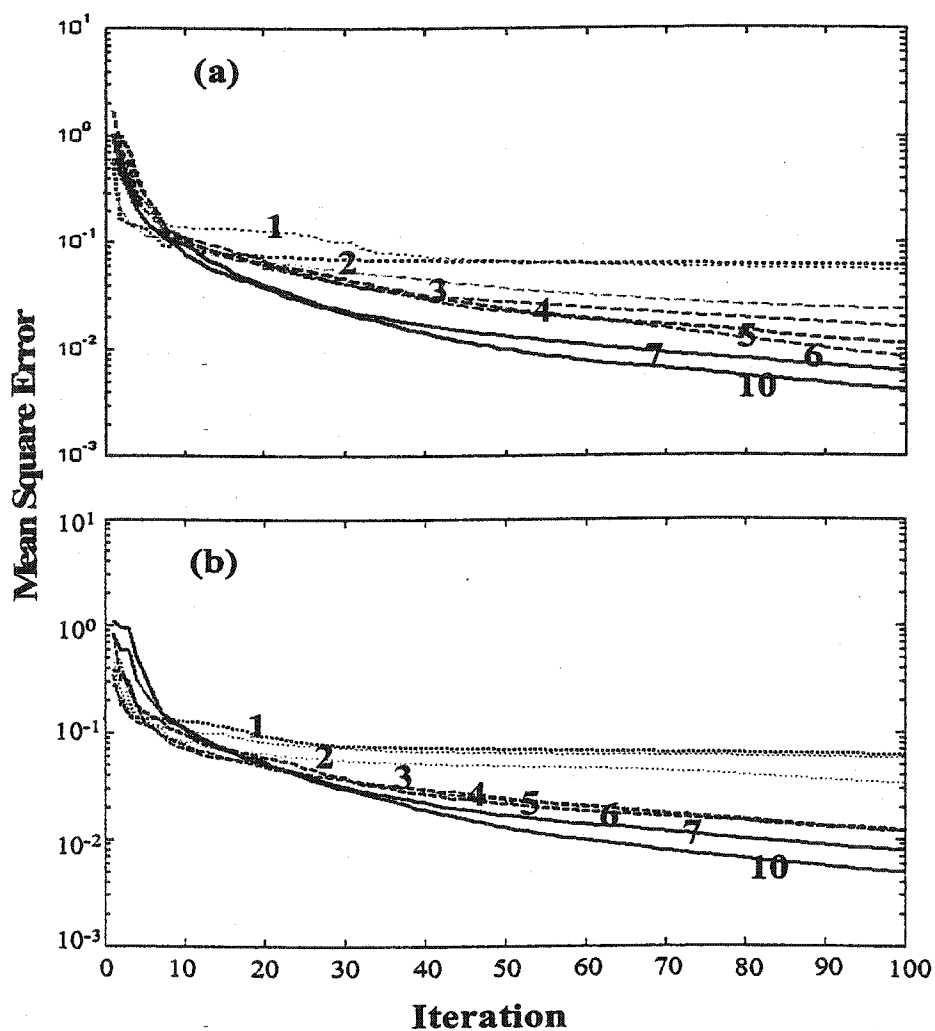


Fig. 5-5. Effects of the number of hidden neurons and iterations on Mean Square Error (a) for the Back-Propagation Network and (b) for the Elman Recurrent Network. The winter storm events of 1999 with 218 data points were used.

The MSE improved gradually with increasing iteration when more hidden neurons were used. With seven to 10 hidden neurons, the MSE of the BPN and ERN models significantly reduced to  $10^{-2}$ .

For checking the change of the bias, variance, and MSE, the number of hidden neurons was increased from one to 10 with an increment of one between each trial. The number of iterations was fixed at 20. The results of bias and variance are given in Table 5-3. The variance and bias were much smaller than the MSE for both BPN and ERN models. The MSE decreased with increasing number of hidden neurons. However, it did not decrease when the number of hidden neurons was larger than seven for both the BPN and ERN (Fig. 5-6). For this reason, the number of hidden neurons was fixed at six.

In order to find the optimum of training iterations, so that the ANN model will not be 'over-trained', two different data sets were used and their error gradients compared (Fig. 5-7). In this particular case, the first data set was the wind-wave data set from the 1999 winter storm season, and the second data set was the wind-wave data from the 1998 winter storm season.

The error gradients for the BPN and ERN were compared by changing the number of iterations and hidden neurons. Figure 5-7 shows divergent points on the gradient curves around 20 iterations for the BPN and 16 iterations for the ERN. But the error gradient is still too large for the ERN with 16 iterations ( $> 10^{-7}$ ), and the next closest point between two lines is at iteration 30. For this reason, 20 and 30 iterations were chosen as the optimum for the BPN and ERN, respectively.



Table 5-3a

Bias, Variance, and Mean Square Error (MSE) for the Back-Propagation Network with different number of hidden neurons to predict wave height. The 1999 winter storms with 218 points was used for training, and the number of iterations was fixed as 20.

Number of Hidden Neurons	$y_m$	$\hat{y}_m$	Bias	Variance	MSE
1	1.3269	1.3152	1.35E-04	1.50E-27	0.075
2	1.3269	1.3150	1.42E-04	1.72E-26	0.130
3	1.3269	1.3133	1.84E-04	1.92E-28	0.067
4	1.3269	1.3241	7.35E-06	2.70E-27	0.059
5	1.3269	1.3528	6.75E-04	5.65E-27	0.064
6	1.3269	1.3559	8.43E-04	2.23E-26	0.049
7	1.3269	1.3328	3.49E-05	5.27E-28	0.034
8	1.3269	1.3669	1.60E-03	8.04E-28	0.044
9	1.3269	1.3573	9.24E-04	1.16E-27	0.034
10	1.3269	1.3561	8.57E-04	4.29 E-27	0.038

$y_m$  = mean of observed wave heights

$\hat{y}_m$  = mean of predicted wave heights

Table 5-3b

Bias, Variance, and Mean Square Error (MSE) for the Elman Recurrent Network with different number of hidden neurons to predict wave height. The 1999 winter storms with 218 points was used for training, and the number of iterations was fixed as 20.

Number of Hidden Neurons	$y_m$	$\hat{y}_m$	Bias	Variance	MSE
1	1.3269	1.3575	9.38E-04	1.29 E-26	0.087
2	1.3269	1.2766	1.50E-03	2.24E-26	0.099
3	1.3269	1.2999	7.28E-04	2.45 E-27	0.058
4	1.3269	1.2925	1.12E-03	4.05 E-27	0.062
5	1.3269	1.2932	1.10E-03	7.35 E-27	0.059
6	1.3269	1.3218	2.53E-05	7.20 E-28	0.049
7	1.3269	1.3169	9.97E-04	4.02 E-27	0.052
8	1.3269	1.3328	3.50E-05	7.69 E-29	0.035
9	1.3269	1.3265	1.37E-07	9.97 E-30	0.037
10	1.3269	1.3246	5.03E-04	2.50 E-27	0.036

$y_m$  = mean of observed wave heights

$\hat{y}_m$  = mean of predicted wave heights

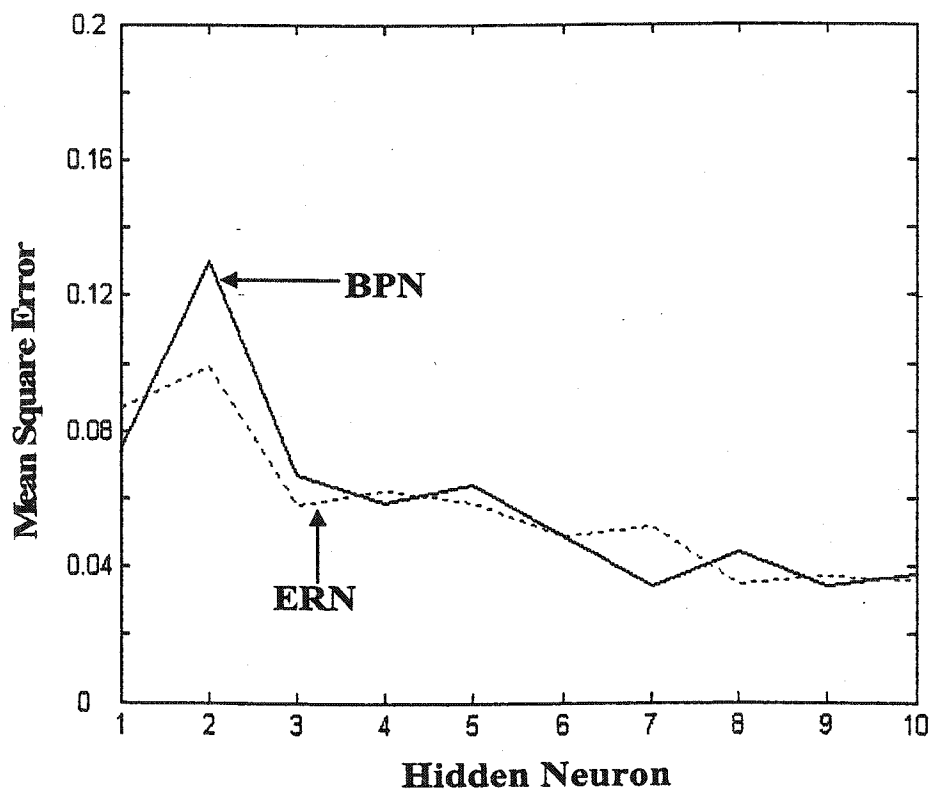


Fig. 5-6. Effects of the number of hidden neurons on the Mean Square Error. The number of training data and iterations were 218 and 20, respectively.

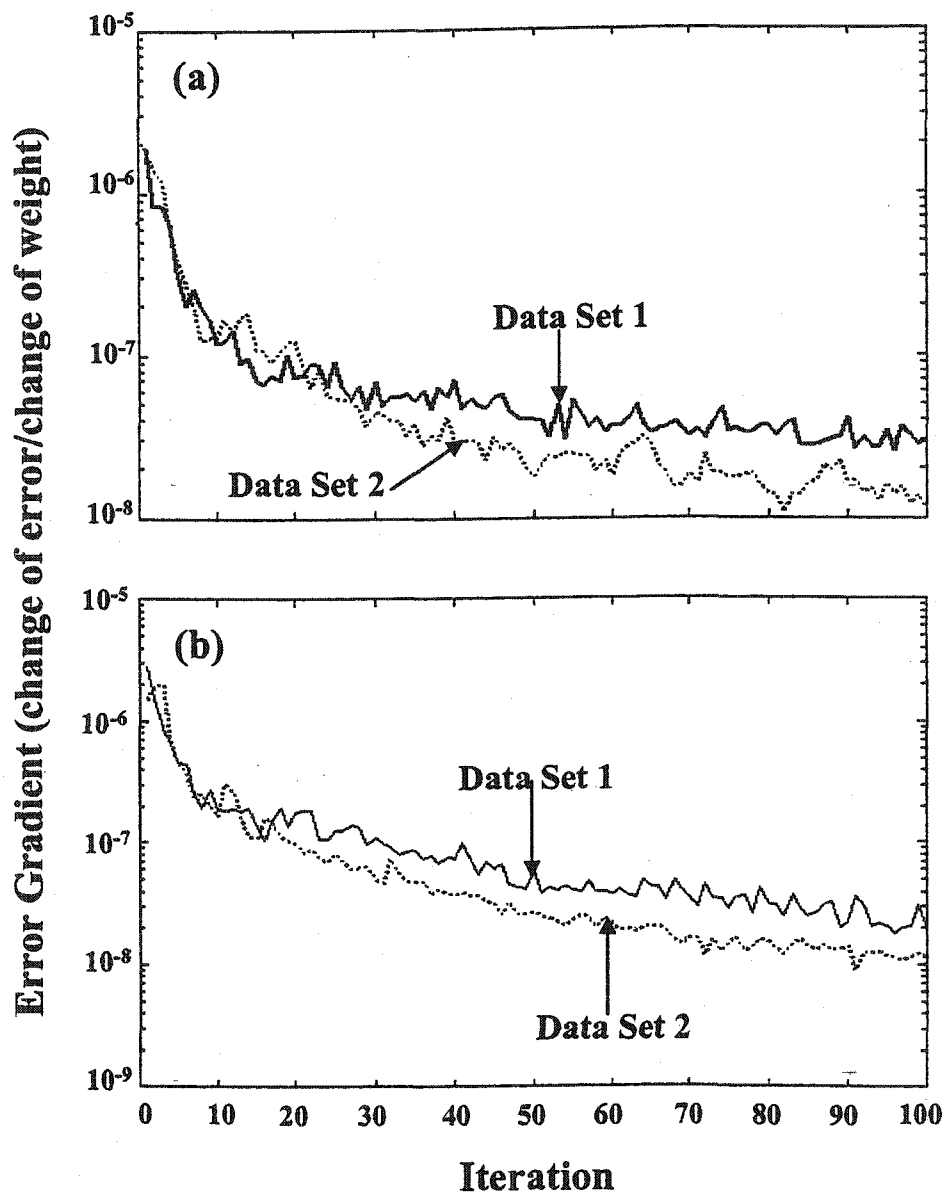


Fig. 5-7. Error gradient curves for finding the optimum number of iterations (a) for the Back-Propagation Network and (b) for the Elman Recurrent Network. Six hidden neurons and 1999 winter storm events with 218 data points were used.

Figure 5-8 shows the correlation coefficient ( $r$ ) between the observed and predicted wave heights for both models when five, six, seven, and 10 hidden neurons were used. Exercises with smaller numbers of hidden neurons were not done because they had larger MSE. The number of iterations ranged from 10 to 50 with an increment of five iterations between each trial. The best correlation coefficient was observed as 0.70 and 0.77 ( $r^2 = 0.49$  and  $0.59$ ) at 20 and 30 iterations for the BPN and ERN, respectively.

Figure 5-9 compares the observed and predicted wave heights for the BPN and ERN at the five wave stations. When compared with prediction results with wind speed and direction as ANN inputs, the number of occurrences for negative wave height was much reduced. However, at station 44013, the predicted wave height still has negative values at 09:50, 15:50, and 18:50, on February 16, 1998 for the BPN, and at 21:50, on February 16, 1998 for the ERN. Some common characteristics of the BPN and ERN might explain why two ANN models have negative heights. So, a better ANN algorithm (*i.e.*, the TDNN) was identified and tested with the same conditions as the BPN and ERN in the next section.

Figure 5-10 shows the correlation coefficient ( $r$ ) between the observed and predicted wave heights. The low correlation coefficient of 0.77 between the observed and predicted wave height compared with the prediction results of the SMB simulated wind-waves may be due to an insufficient number of training data points and the complicated wind-wave systems. The training data set size used was only 218 points from one winter storm season. In general, if the training data set increases, the

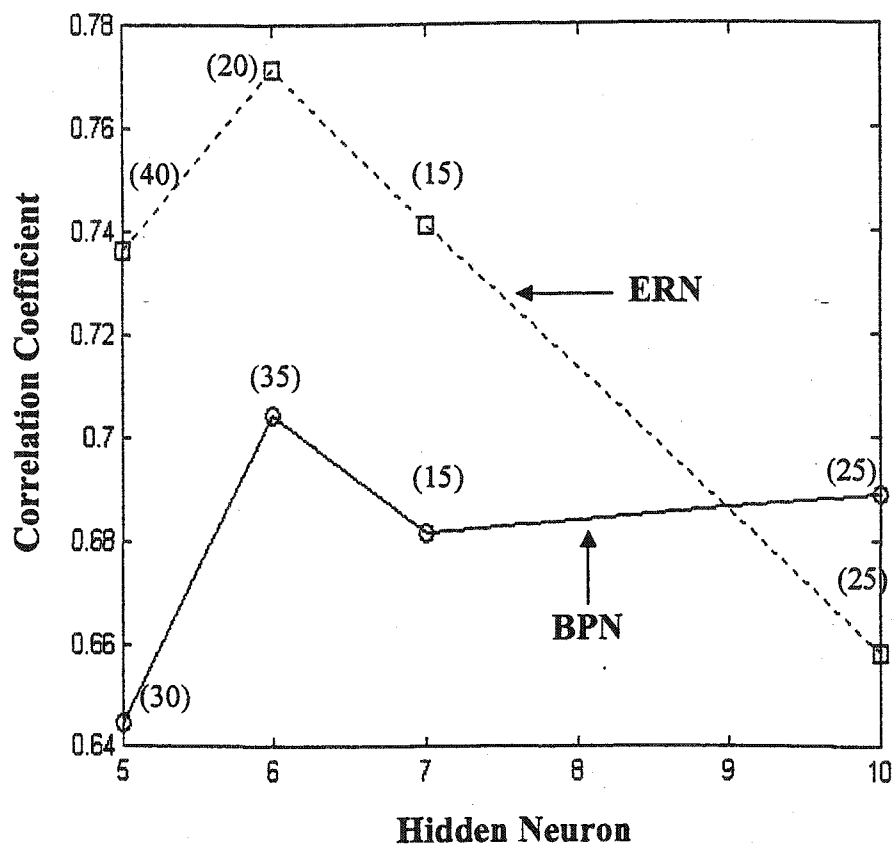


Fig. 5-8. Correlation coefficient between the observed and predicted significant wave heights. The numbers in parentheses indicate the number of iterations.

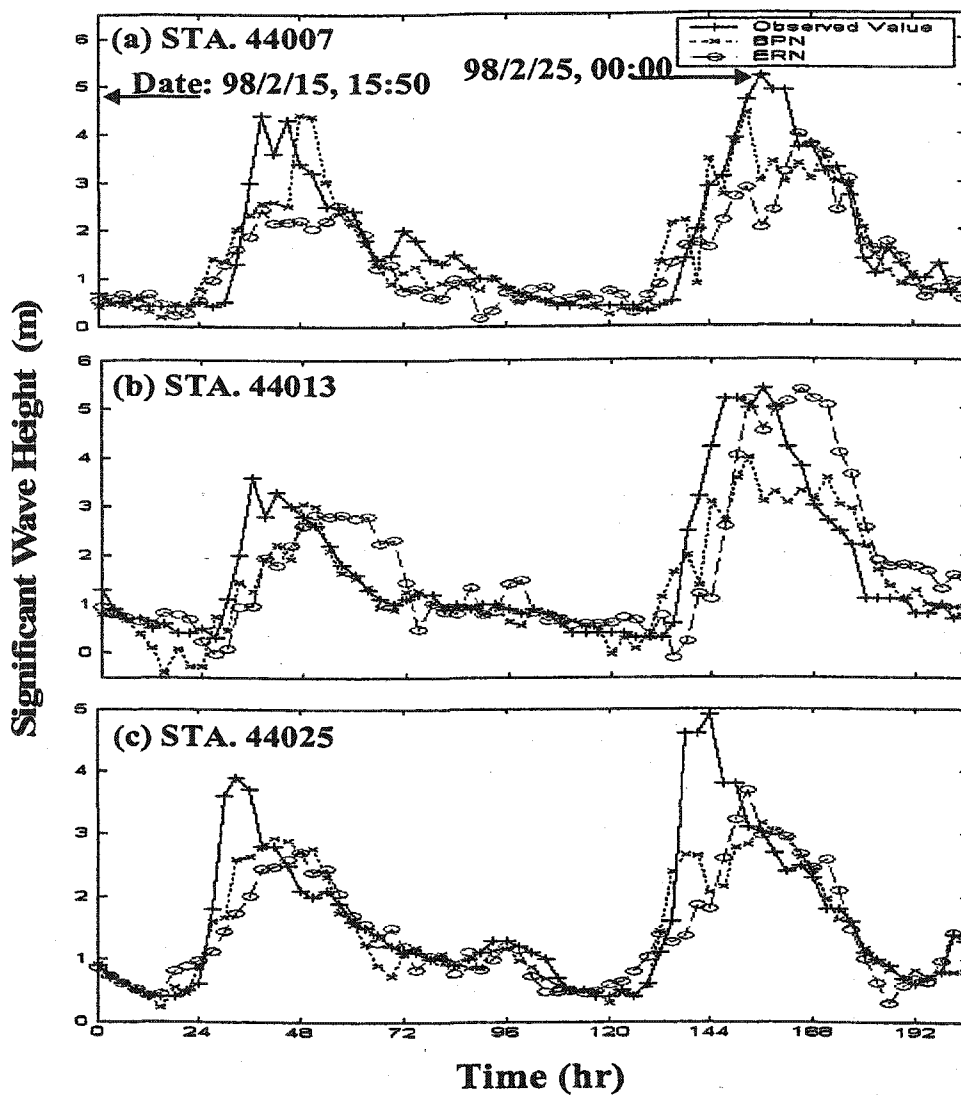


Fig. 5-9. Comparison of the observed and predicted significant wave heights using the Back-Propagation Network and Elman Recurrent Network with a structure of  $I_{720}H_6O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed were 20 m/s and -20 m/s. The number of iterations was 20 for the BPN and 30 for the ERN, respectively.

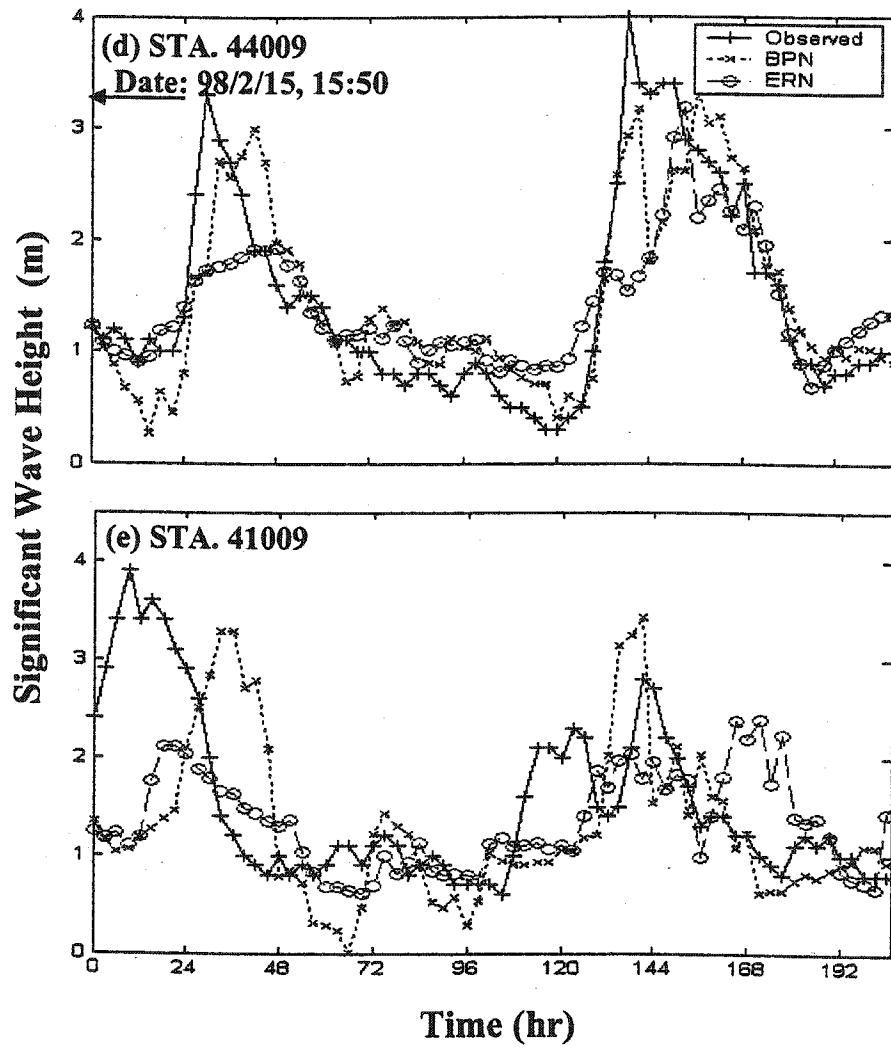


Fig. 5-9. (continued)



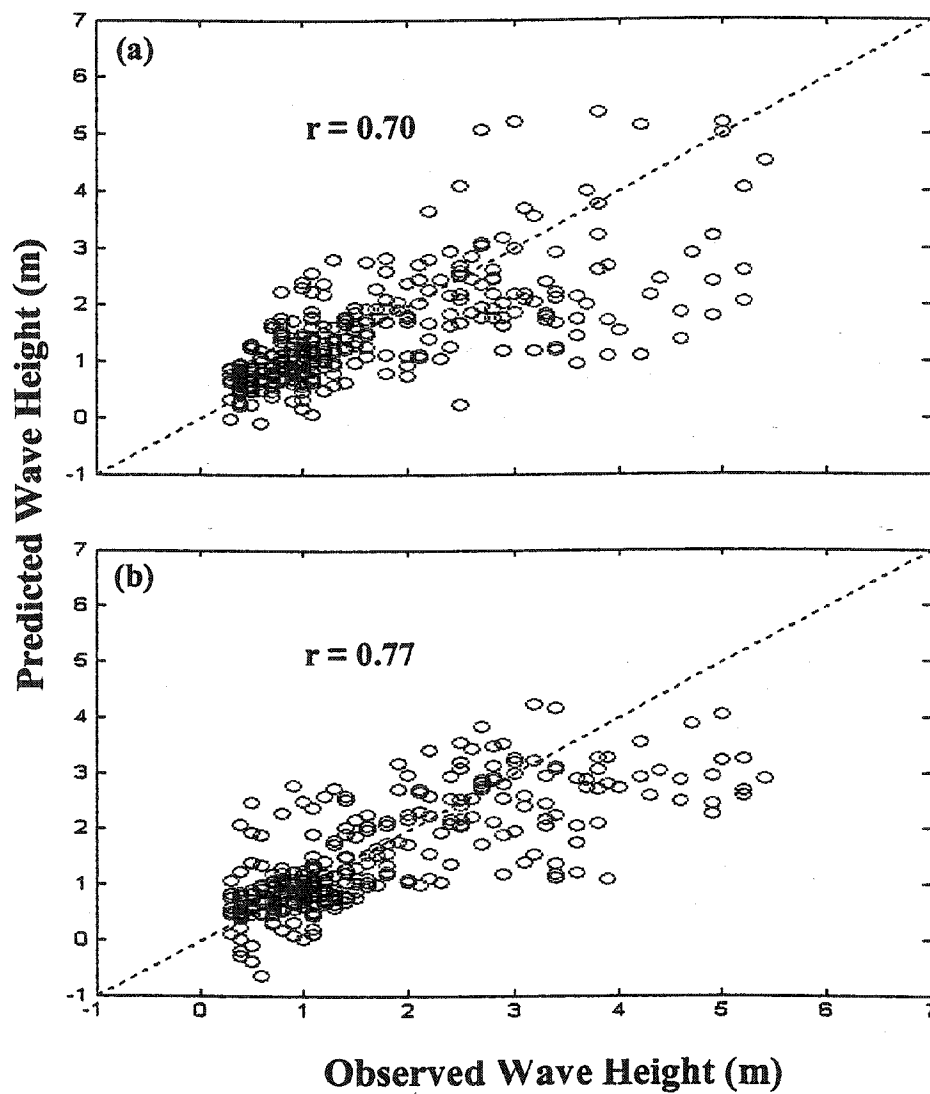


Fig. 5-10. Correlation coefficient between the observed and predicted significant wave heights. (a) For the Back-Propagation Network with 20 iterations and (b) for the Elman Recurrent Network with 30 iterations.

coverage of training patterns to validation data will increase, and thus, the prediction accuracy will improve. For this reason, more training data points should be tested to address this problem in further studies.

The prediction results at station 41009 (see Fig. 5-9e) were much less accurate than those of other stations. It is not clear yet what is (are) the reason(s) of this low accuracy. What is clearly evident is that station 41009 is very distant and different from the other four stations, and the correlation coefficients of wave height between 44007 and 44013, 44025, 44009, and 41009 are 0.72, 0.64, 0.42, and -0.01, respectively. This fact indicates that temporal variation of wave height at station 41009 is very weakly related to that at northern stations, if at all.

As shown in Fig. 5-1, the wind system near station 41009 was clearly different from that in the northern area. This explains why the correlation coefficient is near zero between station 41009 and other stations. Apparently, for station 41009, there are too many input data (*i.e.*, wind at northern area), which are irrelevant to the wave observed at station 41009. Or, additional wind stations on the south side should be included.

### 5.7.2. Three-Layered TDNN

For considering possible long-term memory, the Time Delay Neural Network (TDNN) algorithm with the scaled conjugate gradient learning scheme was used. The input conditions were the same as the previous for the BPN and ERN models.

Figure 5-11 shows the learning curves when the training data set used only the single winter season of 1999 with 218 points. The number of hidden neurons was increased from one to 10 with an increment of one, and the number of iterations was increased from 10 to 100 with an increment of one iteration between each trial.

In general, the MSE improved as the number of hidden neurons and iterations increased. The MSE did not improve when a small number of hidden neurons were used (*e.g.*, one and two). When a large number of hidden neurons, from seven to 10, was used, the MSE improved significantly through all iterations.

Similar to that for the BPN and ERN, the error gradient for two data sets (78 data points from the 1998 winter storm season and 218 points from 1999 winter storm season) was compared with different numbers of hidden neurons and iterations. The number of hidden neurons was increased by one from four to nine because small numbers of hidden neurons are not a good choice based on previous experience using the BPN and ERN. The number of iterations was increased by one iteration from 10 to 100 between each trial.

Figure 5-12 shows the results of estimating the optimum number of iterations. Use of four, five, six, or eight hidden neurons had no remarkable divergent points through all iterations. Results with seven hidden neurons indicate a small number of iterations (*e.g.*, 10) with large error gradients ( $> 10^{-7}$ ), and thus cannot be used. Results from the nine hidden neurons had a divergent point at 40 iterations. This indicates that nine hidden neurons and 40 iterations might be the optimum.

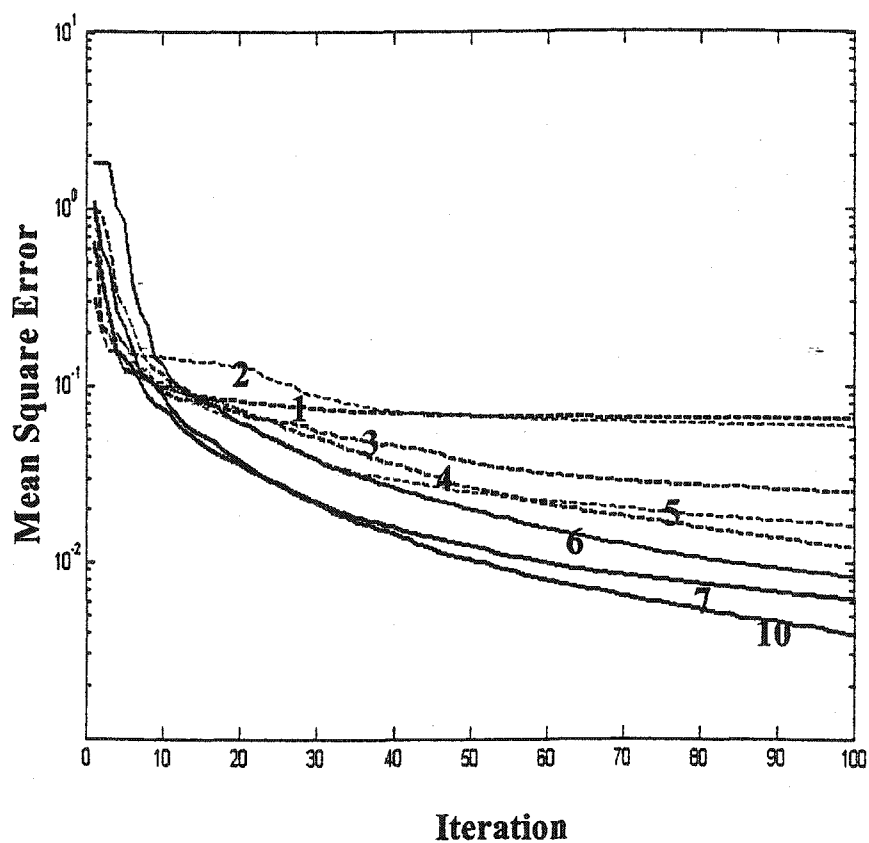


Fig. 5-11. Effects of the numbers of iterations and hidden neurons on Mean Square Error for the Time Delay Neural Network. The 1999 winter-storm events with 218 data points were used.

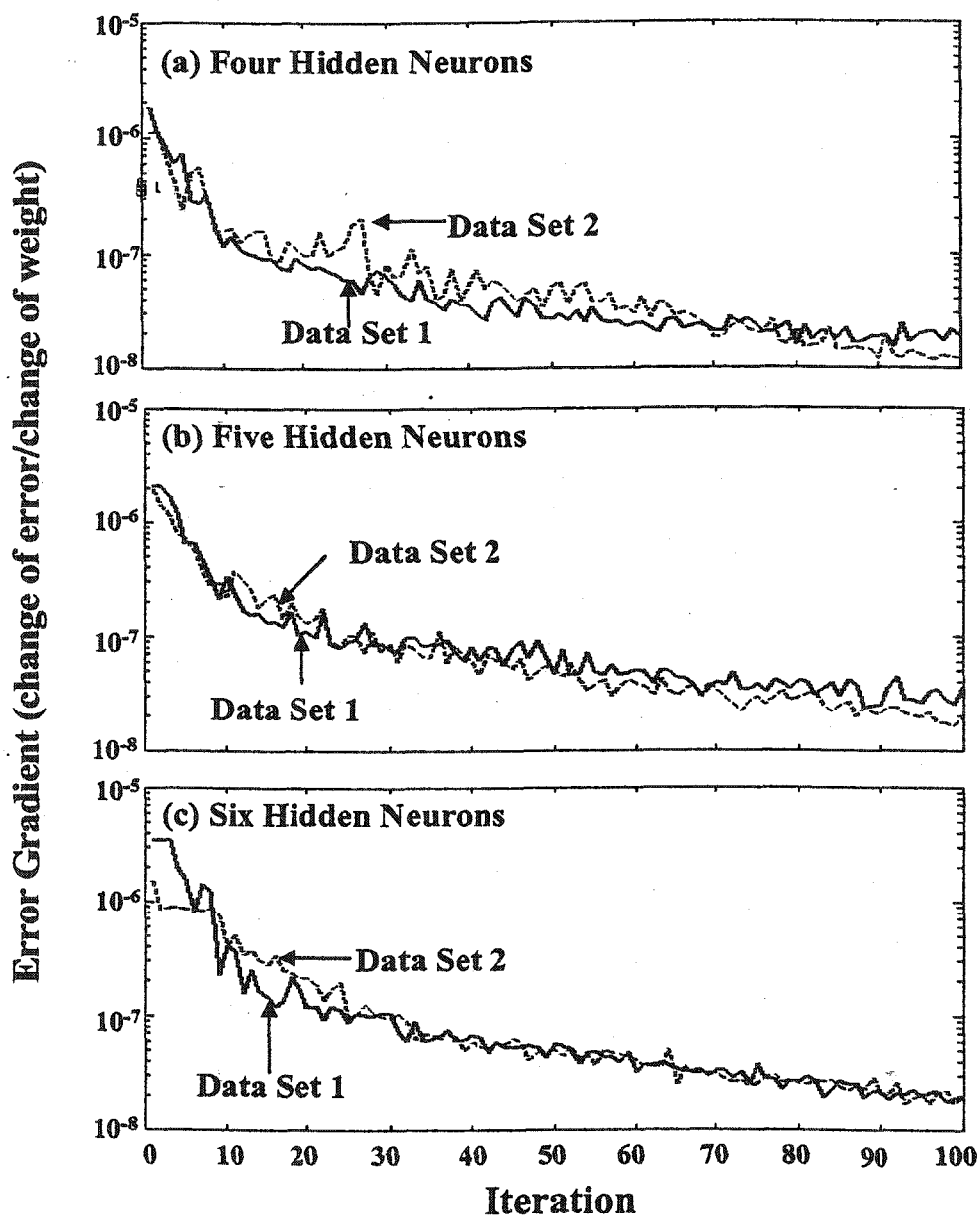


Fig. 5-12. Error gradient curves for finding the optimum number of iterations for the TDNN. (a) For four hidden neurons, (b) five hidden neurons, (c) six hidden neurons, (d) seven hidden neurons, (e) eight hidden neurons, and (f) nine hidden neurons. Only the data from 1999 winter-storm season with 218 data points were used.

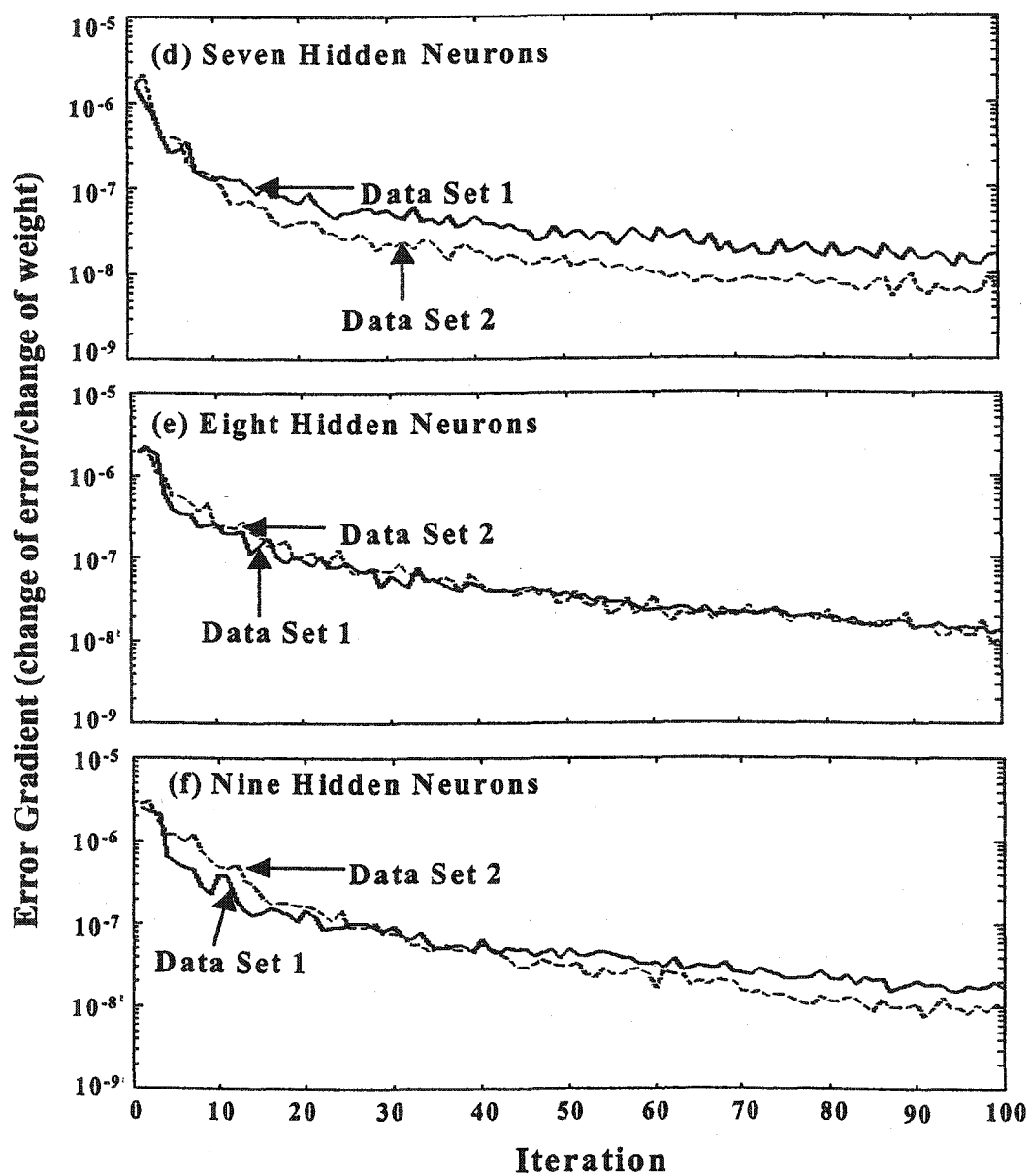


Fig. 5-12. (continued)

For the assessment of bias, variance, and MSE, the data set of 1999 winter storm season with 218 points was used. The number of hidden neurons was increased from one to 10 with an increment of one hidden neuron between each trial, whereas the number of iterations was fixed at 40. The results are shown in Table 5-4. The changes in the bias and variance were much less than those for the MSE. In general, the MSE decreased as the number of hidden neurons increased (Fig. 5-13). However, that decrease ended when the number of hidden neurons was more than seven.

The above studies indicate that the number of hidden neurons should be between six and nine, and the number of iterations should be around 40. Nevertheless, none of the above results guarantee that the optimum number of hidden neurons and iterations for the TDNN have been found.

The last effort used a trial and error method with different numbers of hidden neurons and iterations. The correlation coefficient between observed and predicted wave height was compared. The number of hidden neurons was changed from three to 10 in increments of one, and the number of iterations was increased from 10 to 50 in increments of five between each trial until the correlation coefficient decreased again (Fig. 5-14). The correlation coefficient for eight hidden neurons is clearly better than that for other hidden neurons after the number of iterations was larger than 25. The best correlation coefficient between observed and predicted wave heights was 0.82 ( $r^2 = 0.67$ ) when eight hidden neurons and 40 iterations were used. This is close to the results shown in Figs. 5-12 and 13.

Table 5-4

Bias, Variance, and Mean Square Error (MSE) for the Time Delay Neural Network with different number of hidden neurons. The winter storm season for 1999 with 218 data points was used for training, and the number of iterations was fixed as 40.

Number of Hidden Neurons	$y_m$	$\hat{y}_m$	Bias	Variance	MSE
1	1.3269	1.3259	5.0E-07	1.2E-27	0.0702
2	1.3269	1.347	4.1E-04	1.2E-26	0.0723
3	1.3269	1.3208	3.3E-05	9.8E-27	0.0464
4	1.3269	1.3167	9.8E-05	1.0E-26	0.0296
5	1.3269	1.3177	7.9E-05	2.3E-27	0.0363
6	1.3269	1.3208	3.4E-05	3.1E-27	0.0266
7	1.3269	1.3379	1.3E-04	1.3E-27	0.0159
8	1.3269	1.3411	2.1E-04	6.6E-27	0.0177
9	1.3269	1.3219	2.2E-05	1.8E-27	0.0177
10	1.3269	1.3263	1.1E-07	8.8E-27	0.0147

$y_m$  = mean of observed wave heights

$\hat{y}_m$  = mean of predicted wave heights



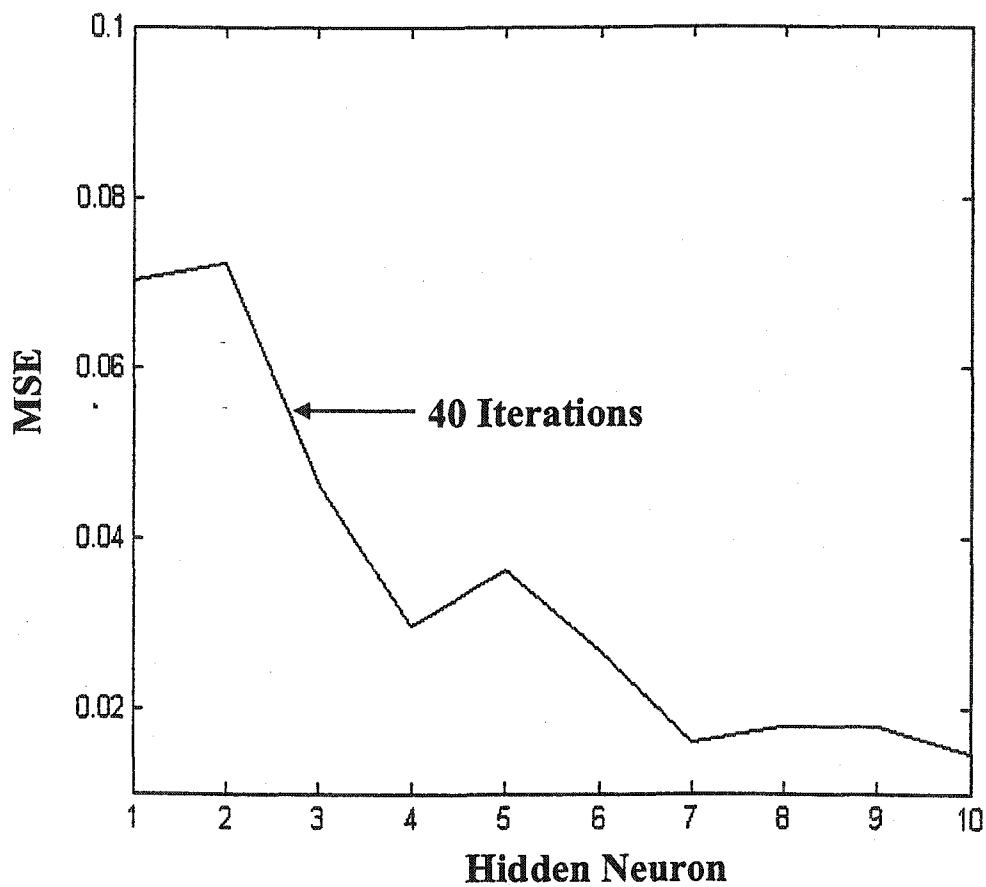


Fig. 5-13. Effects of the number of hidden neurons on the Mean Square Error (MSE) for the Time Delay Neural Network when 1999 winter storms with 218 data points were used for training.

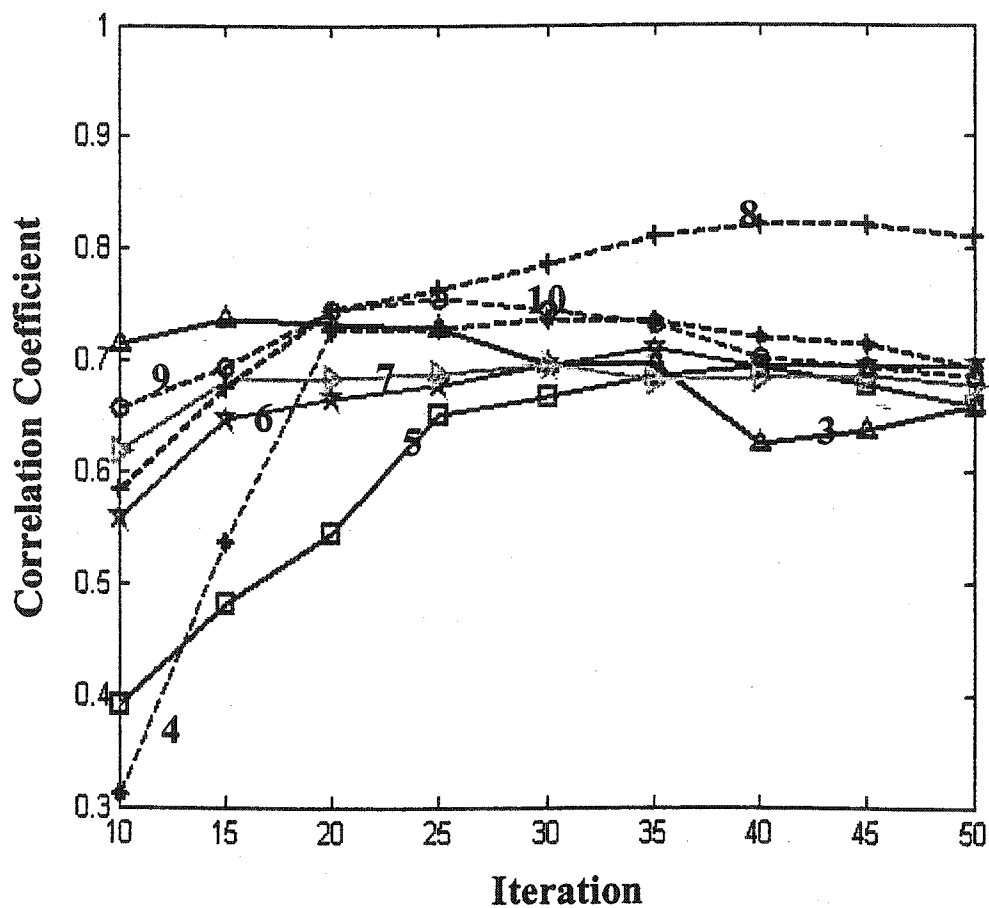


Fig. 5-14. Comparison of the correlation coefficient between the observed and predicted significant wave heights for the Time Delay Neural Network with different numbers of hidden neurons and iterations. The number on each line indicates the number of hidden neurons.

Figure 5-15 plots the observed and predicted wave heights together at five wave stations along the east coast of the U.S. Figure 5-16 shows the correlation coefficient between the observed and predicted wave heights. When compared with the results for the BPN and ERN, the TDNN has no negative wave heights at any wave station. This indicates that the TDNN algorithm is better than the BPN and ERN. However, because the prediction results at station 41009 were still poor, more training data were added in the next trials.

Structures of the BPN, ERN, and TDNN models are summarized in Table 5-5. The MSE of the TDNN model is 0.018, which was 2.7 times less than that of the BPN model (0.049) and 1.5 times less than the ERN model (0.026). The unit computing time (total computing time/number of iterations) of the TDNN model was 1.34 seconds, which was smaller than those of the BPN and ERN. The reason for the slow speed of the BPN and ERN may be the use of artificially provided time-delays. Because error gradient curves from the previous section failed to show the optimum number of hidden neurons and iteration clearly, they were omitted from further studies.

### 5.7.3. Effect of Training Data Set Size

In this section, the effects of the number of training data points were observed. The third winter storm season in 2001 with 185 data points were added to the 218 points in 1999. Thus, the total number of training data is 403 points, an increase of 85 %.

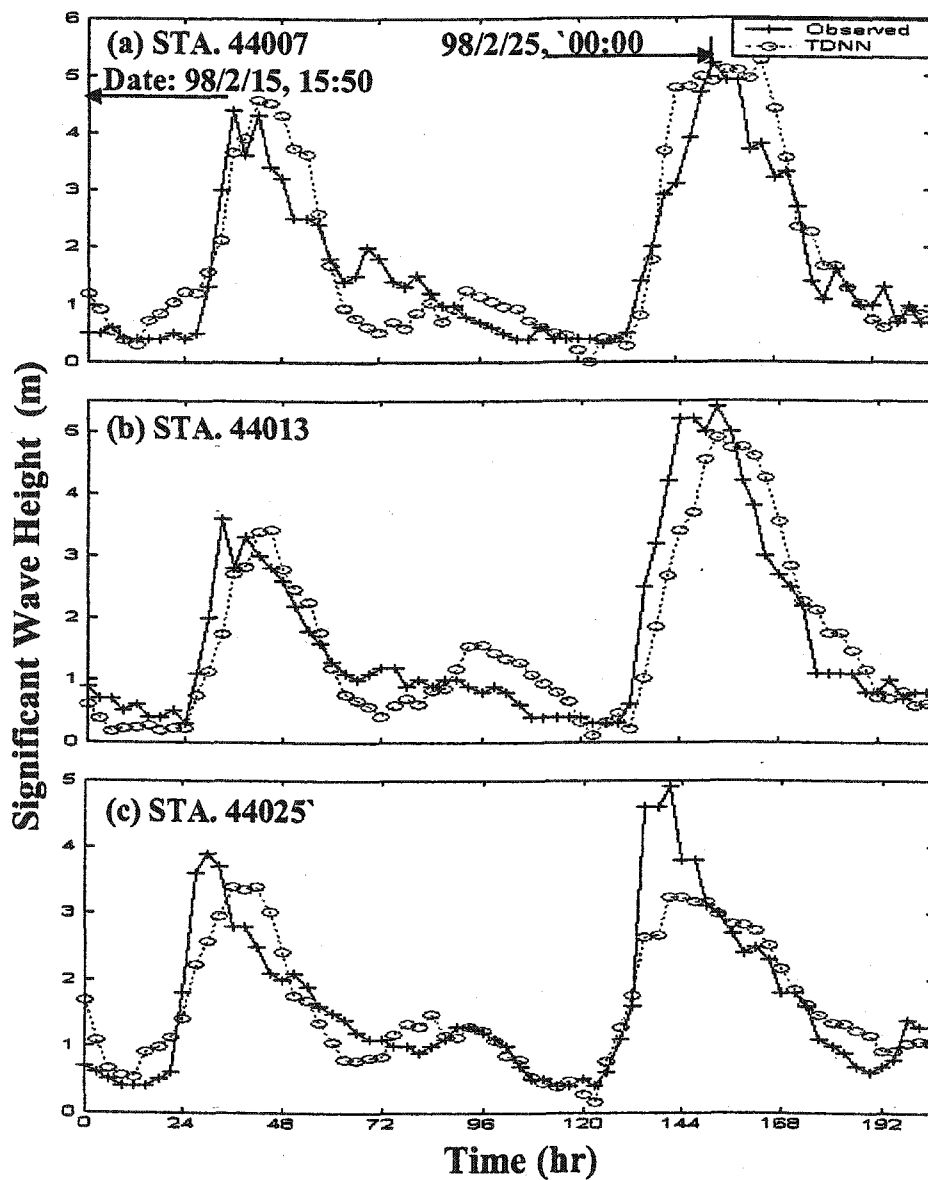


Fig. 5-15. Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of  $I_{720}H_8O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed was  $\pm 20$  m/s. The number of iterations was 40.

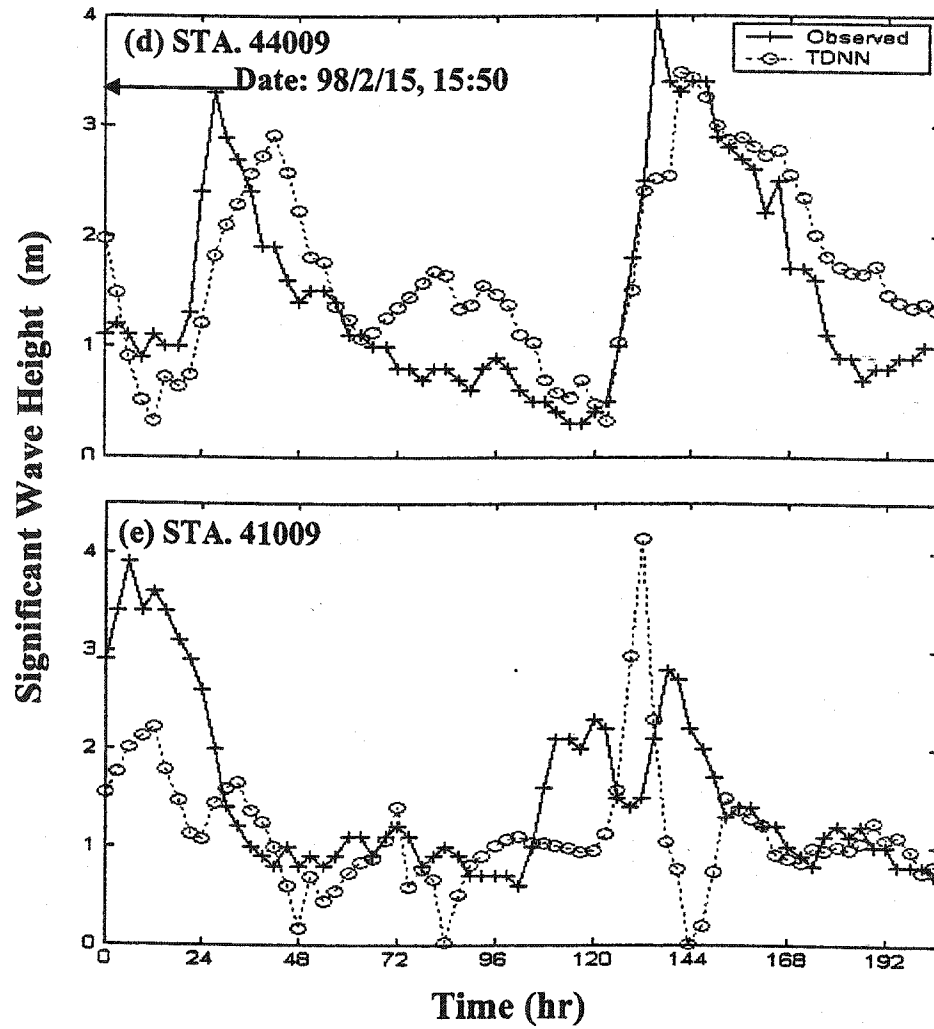


Fig. 5-15. (continued)

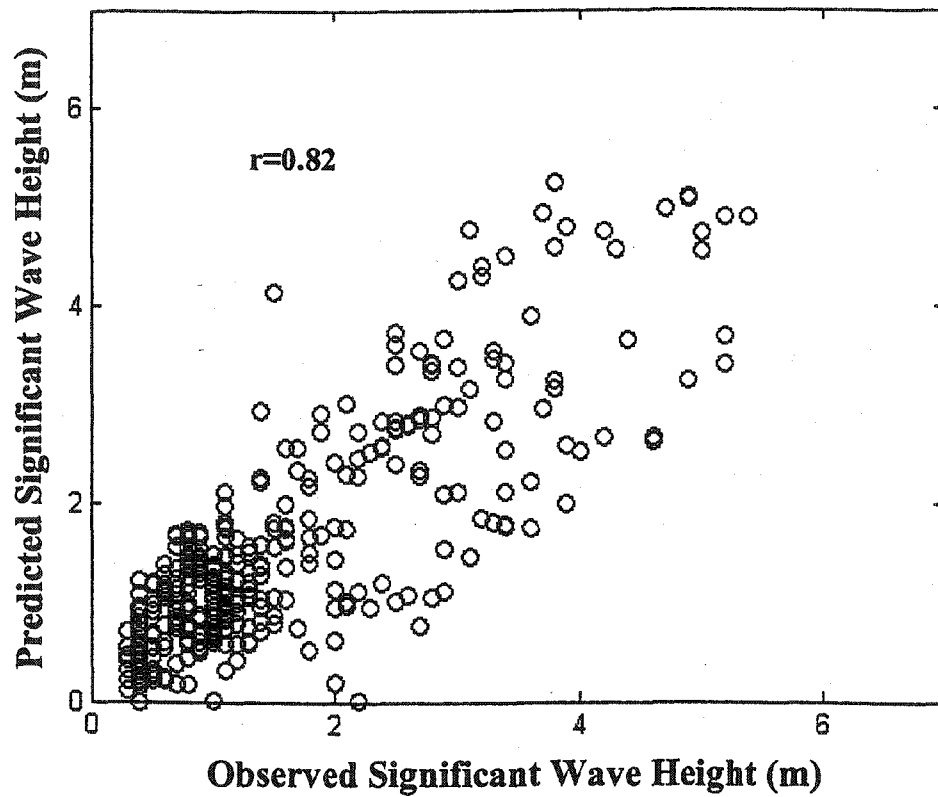


Fig. 5-16. Correlation coefficient between the observed and predicted significant wave heights for the Time Delay Neural Network at the five wave stations when 1999-winter storm season with 218 data points was used for training.

Table 5-5

Comparison of Mean Square Error (MSE) and computing time for the Back-Propagation Network (BPN), Elman Recurrent Network (ERN), and Time Delay Neural Network (TDNN).

Network Type	Structure	Iterations	MSE	Computing Time (Sec.)	Time/Iteration
BPN	720-6-5	20	0.049	33.5	1.67
ERN	720-6-5	30	0.026	164.8	3.66
TDNN	720-8-5	40	0.018	53.7	1.34

For the validation data set, the data for the 1998 winter season was used as before.

The maximum and minimum of  $u$  and  $v$  wind components for the 2001 winter storm season was 20.9 m/s and  $-19.3$  m/s. For this reason, a global maximum and minimum of 20 m/s and  $-20$  m/s was used as before.

Figure 5-17 shows the learning curves with the new data set. The number of hidden neurons was increased from one to 10, and the number of iterations was increased from 10 to 100 between each trial. The MSE did not improve for a small number of hidden neurons (*e.g.*, one and two). The MSE behaved similar to those for the BPN, ERN (see Fig. 5-5), and TDNN (see Fig. 5-11), except that the curves for seven to 10 hidden neurons were almost identical.

In order to check the change on the bias, variance, and MSE for the training set with 403 data points, the number of hidden neurons was increased from one to 10 between each trial, and the number of iterations was fixed at 25. The results are shown in Table 5-6. The bias and variance are much less than the MSE for all hidden neurons. In general, the MSE decreased when the number of hidden neurons increased (Fig. 5-18). However, the MSE did not decrease when using more than seven hidden neurons.

Figure 5-19 shows the correlation coefficient between the observed and predicted wave heights, prepared as in Fig. 5-14 with only 218 data points. Correlation coefficients for all hidden neurons with 403 data points generally improved. Especially, seven to 10 hidden neurons have a high correlation coefficient of 0.80 after 15 iterations.



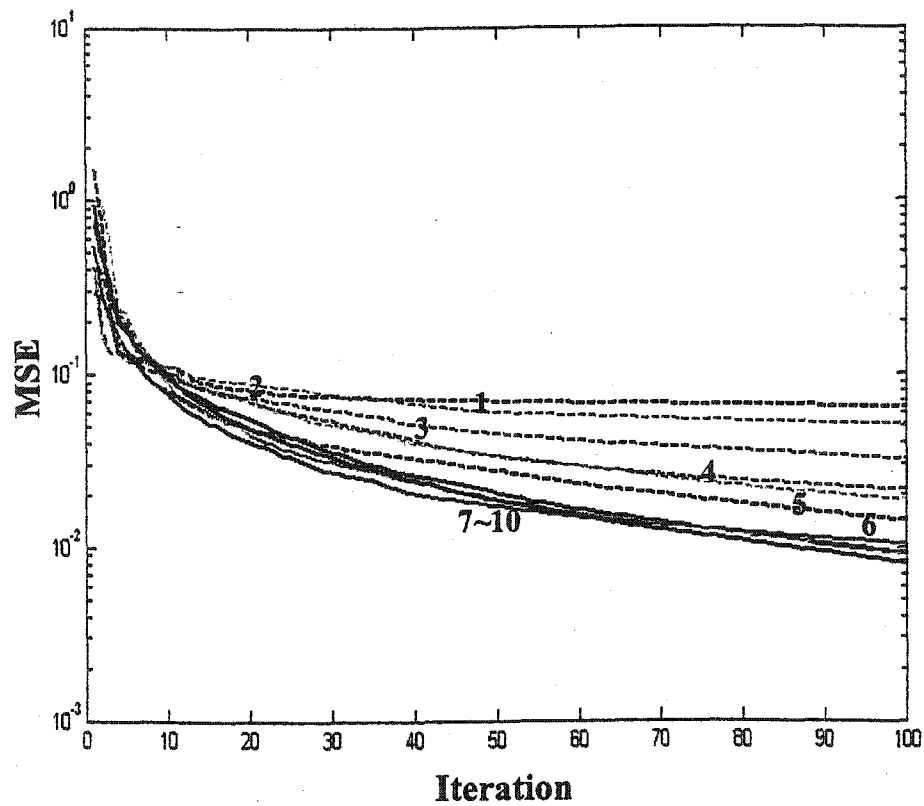


Fig. 5-17. Effects of the numbers of hidden neurons and iterations on the Mean Square Error (MSE) for the Time Delay Neural Network. The total number of training data points was 403.

Table 5-6

Bias, Variance and Mean Square Error (MSE) for the Time Delay Neural Network with different numbers of hidden neurons to predict wave height. The 1999 and 2001 winter storms with 403 data points were used for training, and the number of iterations was fixed as 25.

Number of Hidden Neurons	$y_m$	$\hat{y}_m$	Bias	Variance	MSE
1	1.2157	1.2092	4.20E-05	3.80E-27	0.07630
2	1.2157	1.2208	2.65E-05	3.76E-26	0.08280
3	1.2157	1.2329	2.98E-04	9.25E-27	0.06770
4	1.2157	1.2252	9.19E-05	1.23E-26	0.06130
5	1.2157	1.2422	7.02E-04	3.71E-27	0.05960
6	1.2157	1.2371	4.61E-04	4.60E-26	0.04380
7	1.2157	1.2237	6.45E-05	3.96E-26	0.03300
8	1.2157	1.192	5.60E-04	2.08E-26	0.04380
9	1.2157	1.2136	4.42E-06	7.87E-27	0.03640
10	1.2157	1.2103	2.89E-05	1.97E-27	0.04030

$y_m$  = mean of observed wave heights

$\hat{y}_m$  = mean of predicted wave heights

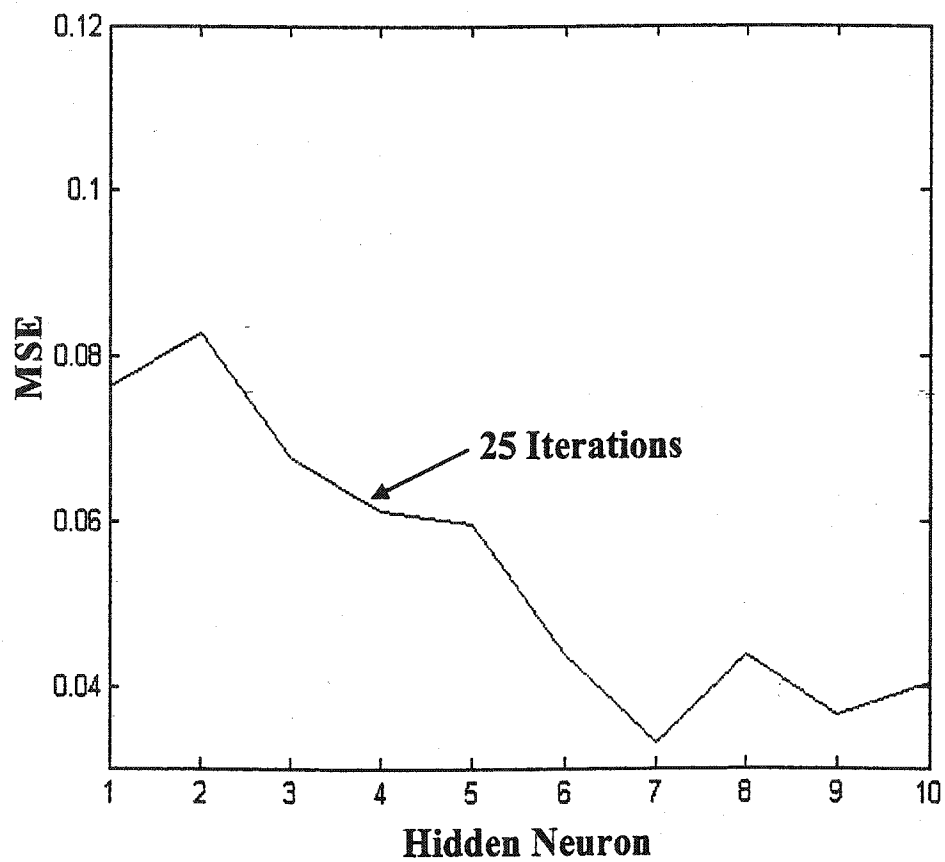


Fig. 5-18. Effects of the number of hidden neurons on the Mean Square Error (MSE) for the Time Delay Neural Network.

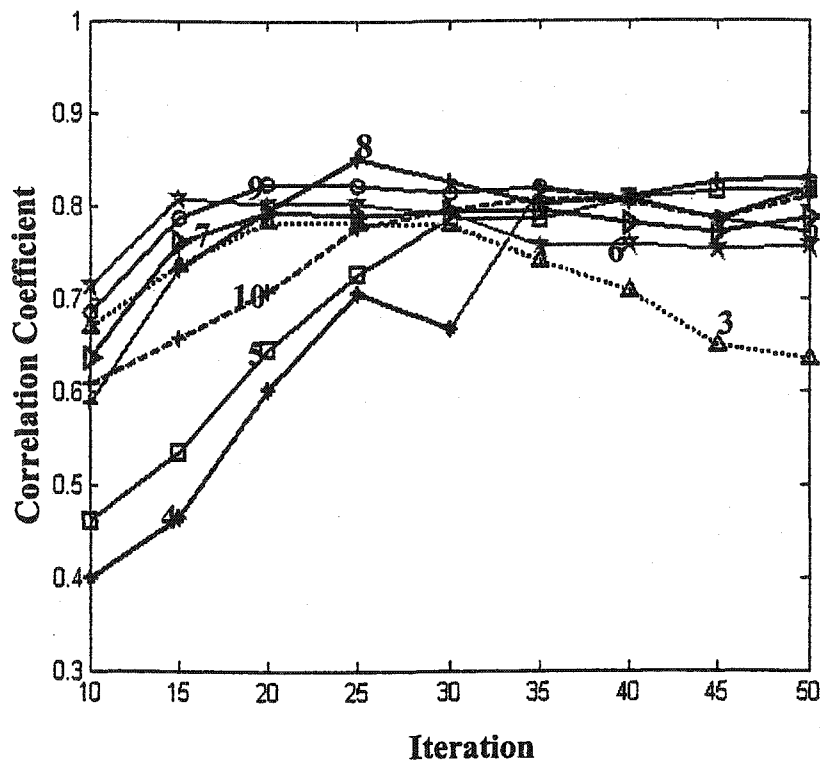


Fig. 5-19. Comparison of the correlation coefficient between the observed and predicted significant wave heights for the Time delay Neural Network with different numbers of hidden neurons and iterations. The number on each line indicates the number of hidden neurons.

The best correlation coefficient was 0.85 ( $r^2 = 0.72$ ) when eight hidden neurons and 25 iterations were used. That is to say, only 70 % of variation can be explained by the model.

Figure 5-20 plots the observed and predicted wave heights at the five stations along the east coast of the U.S. when the TDNN model used  $I_{720}H_8O_5$  at 25 iterations. Like the BPN and ERN, the prediction accuracy at station 41009 was still not good despite the improved correlation coefficient 0.60 vs. 0.34. In addition, the predicted largest wave heights at station 44013 and 44025 were less than those of observed wave heights by 1 m to 1.5 m. At station 44007, one of the predicted wave heights was negative on February 16, 06:50, 1998. The reason for this negative wave height is unclear.

After examining the relationship between the number of iterations and training data points, one question was posed: "Is the negative wave height prediction caused by the decrease of iterations from 40 to 25 when the number of training data was increased from 218 to 403?" Therefore, iteration numbers greater than 25 were tested. When the number of iterations was 30, the predicted wave heights were all positive at the five stations with a slight reduction in the correlation coefficient ( $r$ ) to 0.83. It is clear that the problem of negative wave height is strongly related to the number of iterations. As shown, when the number of iterations is small, the error can be large, and this large error may cause negative wave height when the wave heights itself is small.

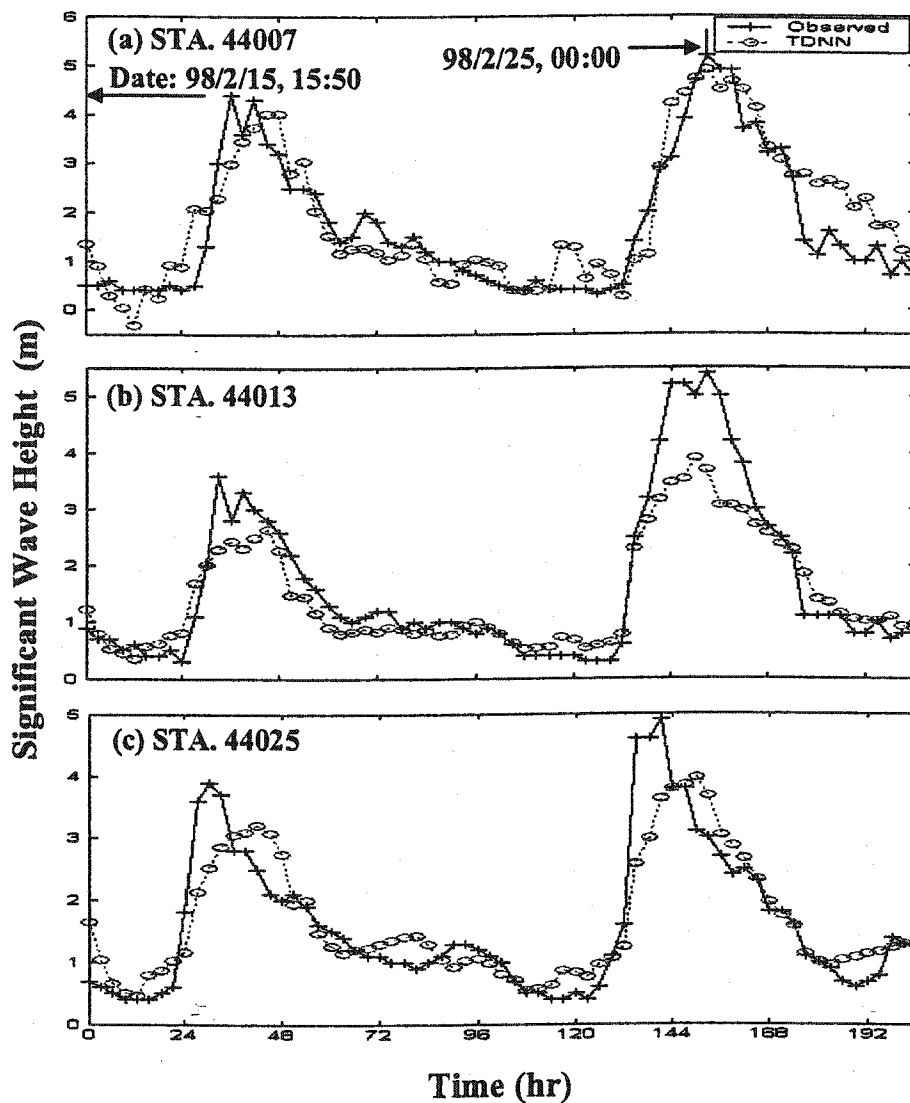


Fig. 5-20. Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of  $I_{720}H_8O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed was  $\pm 20$  m/s. The number of training data points and iterations were 403 and 25, respectively.

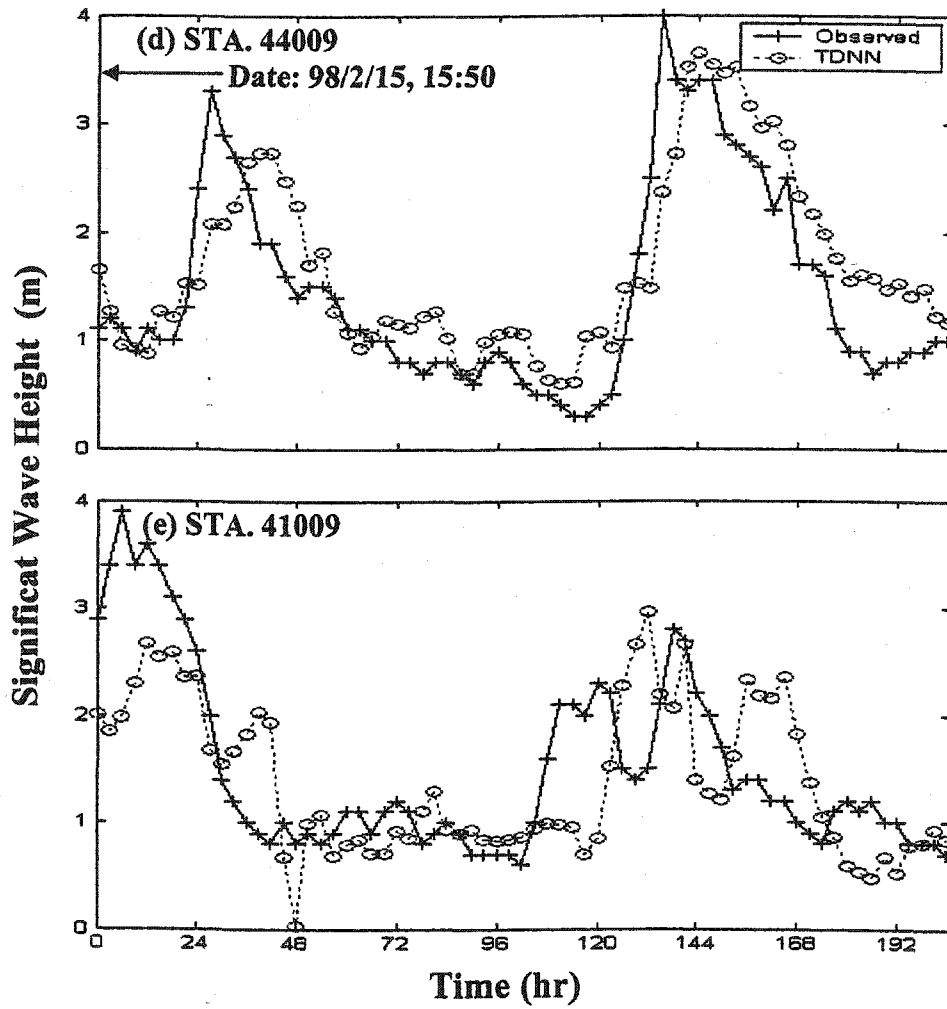


Fig. 5-20. (continued)

The next question is: "Which number of iterations (25 or 30) is better for storm wave height prediction?". Considering that negative wave height has no meaning at all and occurred only when the wave height was small, it is of little importance and the number of iterations for the best correlation coefficient may be used as the optimum (*i.e.*, 25). Figure 5-21 shows the correlation coefficient ( $r$ ) between the observed and predicted wave heights at 25 iterations. The unit-computing time (total computing time/number of iterations) increases from 1.34 seconds to 2.56 seconds. The summary of the prediction results for training data set with 218 and 403 is given in Table 5-7.

The correlation coefficient between the observed and predicted wave heights for the BPN, ERN, and TDNN at each wave station was compared (Table 5-8). Clearly, the TDNN algorithm with two winter storm seasons for training is much preferable. The maximum correlation coefficient of 0.96 ( $r^2 = 0.92$ ) was observed at station 44013 and the minimum of 0.60 ( $r^2 = 0.36$ ) at station 41009.

When the 2001 winter storm season with 185 data points was added for training, two characteristic things were observed. The prediction accuracy increased at four wave stations (44013, 44025, 44009, and 41009) with only a slight reduction (from 0.92 to 0.90) at a single station (44007). The correlation coefficient at station 41009 increased from 0.38 to 0.60. Although the larger number of training patterns has better predictive accuracy, there is an obvious need of more winter-storm patterns for training if prediction accuracy at that station is to be improved.



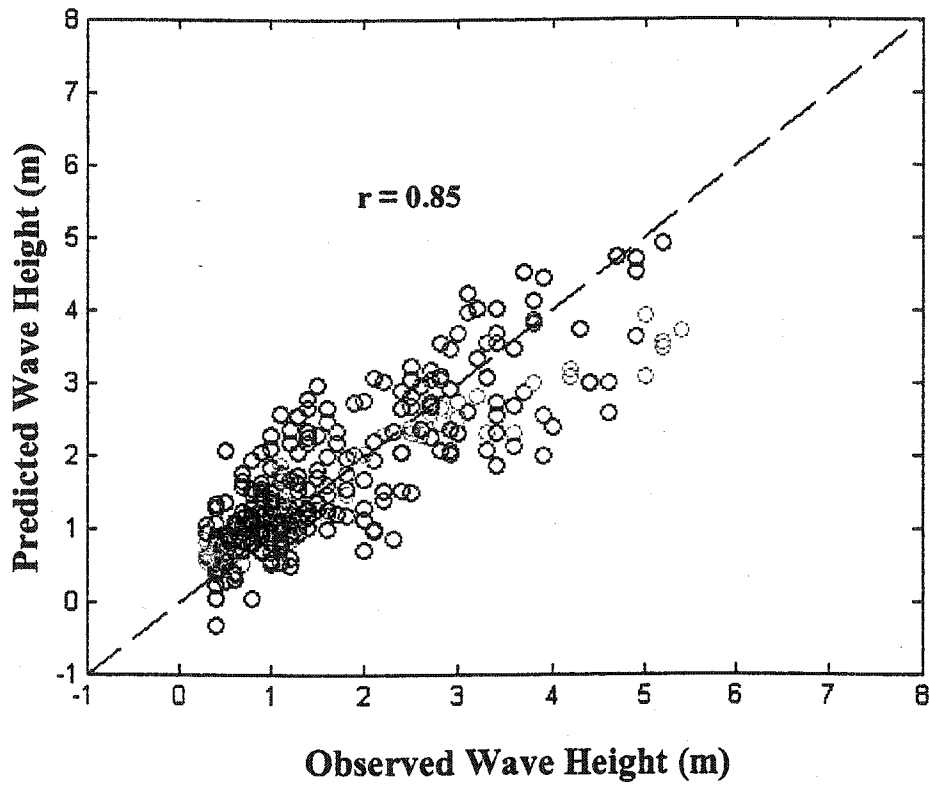


Fig. 5-21. Correlation coefficient between the observed and predicted significant wave heights for the Time Delay Neural Network, which was trained with 403 data points.

Table 5-7

Comparison of the Mean Square Error (MSE) and computing time for the Time Delay Neural Network when 218 data points for 1999 winter storms and 403 points for 1999 and 2001 winter storms were used for training.

Type	Structure	Iteration	MSE	Computing Time (Sec)	Time/Iteration
TDNN (218 points)	720-8-5	40	0.0176	53.7	1.34
TDNN (403 points)	720-8-5	25	0.0438	63.9	2.56

Table 5-8

Correlation Coefficient between the observed and predicted wave height at five Wave stations for the Back-Propagation Network (BPN), Elman Recurrent Network (ERN), and Time Delay Neural Network (TDNN).

Type	44007	44013	44025	44009	41009
BPN (218)	0.84	0.73	0.72	0.80	0.35
ERN (218)	0.81	0.82	0.84	0.88	0.31
TDNN (218)	0.92	0.89	0.86	0.79	0.38
TDNN (403)	0.90	0.96	0.88	0.84	0.60

-(Number) indicates the number of training data points

One possible reason for negative values at station 44007 and underestimated wave heights at stations 44013 and 44025 is the sensitivity of ANN performance to the range of normalized wind. For instance, the TDNN has negative wave height when small wave heights (less than 0.5 m) occur (see Fig. 5-20a). In pre-processing, we used  $\pm 20$  m/s as a global maximum and minimum wind speed, resulting in normalized wind between  $-1$  and  $1$ . In the case of a large range of wave heights, relatively small changes in wind ( $0.5$  m/s) will not be easy to identify. When using a larger global maximum and minimum wind speed, the input range of the wind decreased with reduced output of transfer function (Fig. 5-22).

For instance, as previously explained in Chapter 2, an ANN weights each input between  $-1$  and  $1$  according to its importance and adds up these weight values, which are used as input in a non-linear transfer function. As shown in Fig. 5-22, the output of the transfer function will be between  $-1$  and  $1$  if the summed weight value is between  $-5$  and  $5$ .

Now, let's think about the normalized wind input again. When using a large global wind (e.g.,  $\pm 80$  m/s), the normalized wind inputs will be less than those when  $\pm 20$  m/s global wind are used. If the normalized value is relatively small, the ANN may designate lower weight values and thus, the product of input and weight values becomes smaller. This implies that the maximum input of the activation function may be less than five, better allowing the ANN to recognize changes in wind speed.

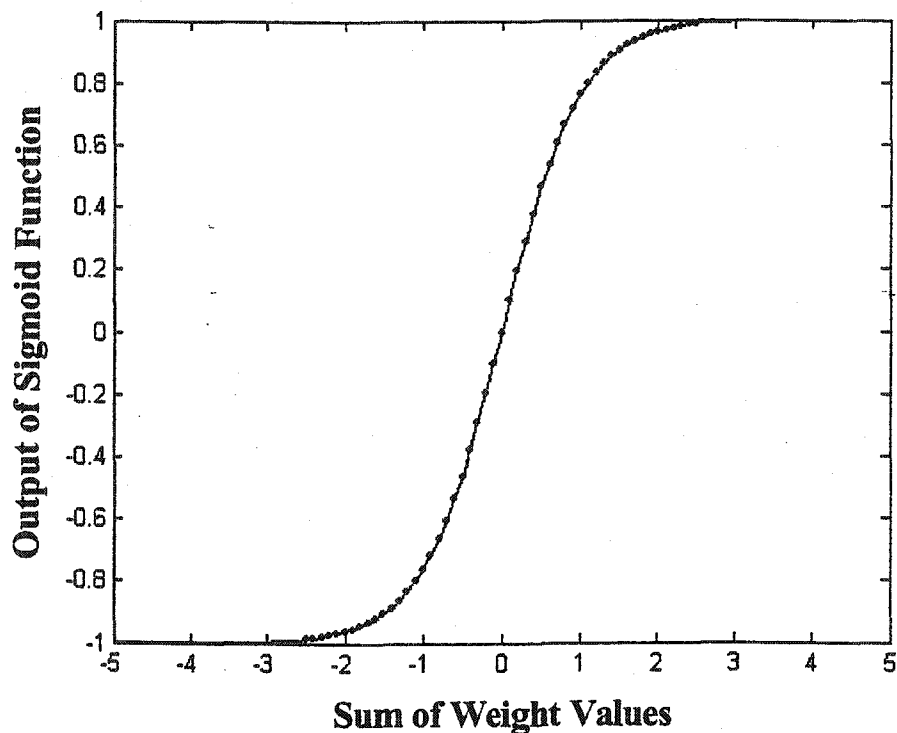


Fig. 5-22. Illustration of the relationship between global maximum and minimum wind, and output range of sigmoid transfer function, e.g.,  $F = 1/(1+\exp^{-2n})$ , where  $n = \text{sum of weight values}$ . When  $\pm 20$  m/s is used as a global maximum and minimum wind, an ANN may set a large weight value for a given wind speed than  $\pm 80$  m/s. In this case, the transfer function output will not change much (solid line) after  $n$  is larger than 2.5. In contrast, because  $\pm 80$  m/s of a global wind may have smaller weight values, the output of the transfer function changes non-linearly (dotted line).

For this reason, a global maximum and minimum wind more than  $\pm 20$  m/s was tested to reduce the input wind range after normalization using the same number of hidden neurons and training data set. For instance, the global wind was changed from  $\pm 30$  m/s to  $\pm 100$  m/s with an increment of  $\pm 10$  m/s between each trial. The optimum iteration was selected as one of the iterations around 25 having best results and no negative wave heights (*e.g.*, 20, 25, or 30).

In general, the prediction accuracy gradually improved as a larger global maximum and minimum wind was used. The MSE was least as 0.032 when a global maximum and minimum of  $\pm 20$  m/s was used. However, the correlation coefficient ( $r$ ) between the observed and predicted wave heights was 0.86 ( $r^2 = 0.74$ ) when using  $\pm 80$  m/s as global winds at 25 iterations, and the correlation coefficient improved to 0.88 ( $r^2 = 0.77$ ) when iterations were increased to 30. However, the TDNN did not improve prediction accuracy after  $\pm 80$  m/s. All results for the TDNN with various global maximum and minimum winds are given in Table 5-9.

Figure 5-23 shows the observed and predicted wave heights at the five wave stations when the TDNN used  $\pm 80$  m/s as the global winds. The predicted wave heights at stations 44013 and 44025 were much improved and almost the same as the observed wave height. At station 44007, no negative wave height was predicted.

The correlation coefficient between the observed and predicted wave heights at the five stations is given in Fig. 5-24. In general, the correlation coefficient increased slightly compared with the results using  $\pm 20$  m/s global winds. The correlation coefficient ( $r$ ) increased from 0.90 to 0.92 at station 44007, 0.88 to 0.89 at station

Table 5-9

Effects of global maximum and minimum wind on the performance of the Time Delay Neural Network with eight hidden neurons and 403 data points for training.

Global Wind (m/s) (Minimum, Maximum)	Iterations	MSE	Correlation Coefficient
(-20, 20)	25	0.044	0.85
(-20, 20)	30	0.036	0.83
(-30, 30)	30	0.032	0.83
(-40, 40)	30	0.039	0.82
(-50, 50)	20	0.051	0.85
(-60, 60)	20	0.064	0.85
(-70, 70)	25	0.061	0.86
(-80, 80)	30	0.058	0.88
(-90, 90)	25	0.075	0.81
(-100, 100)	25	0.069	0.84

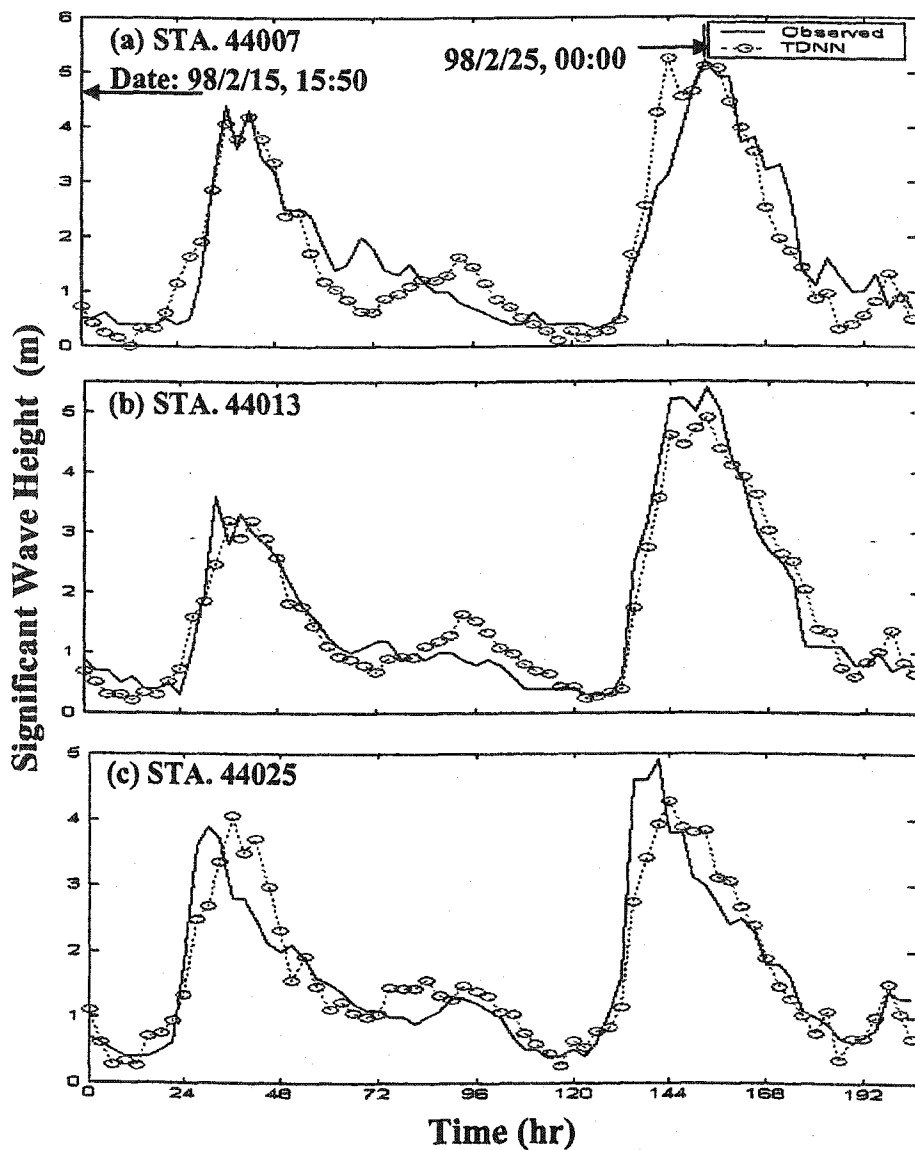


Fig. 5-23. Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of  $I_{720}H_8O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed were 80 m/s and -80 m/s. The number of training data points and iterations were 4 0 3 and 3 0, respectively.



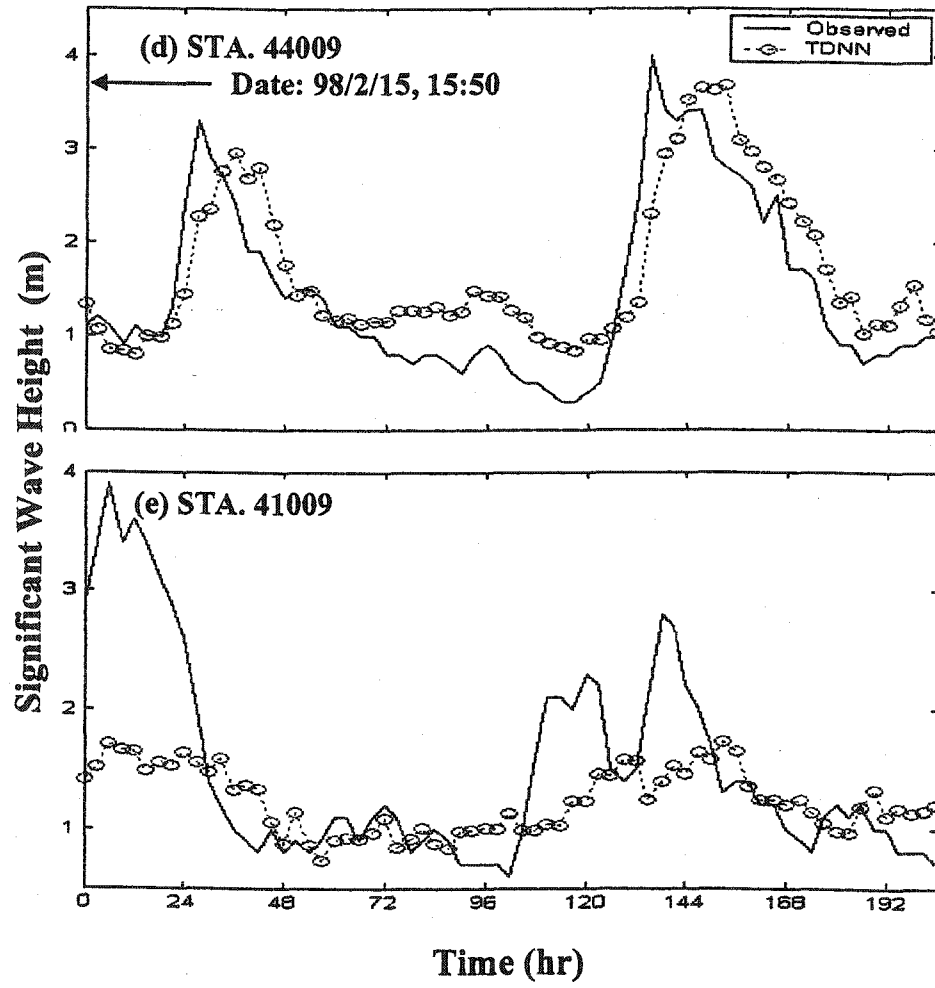


Fig. 5-23. (continued)

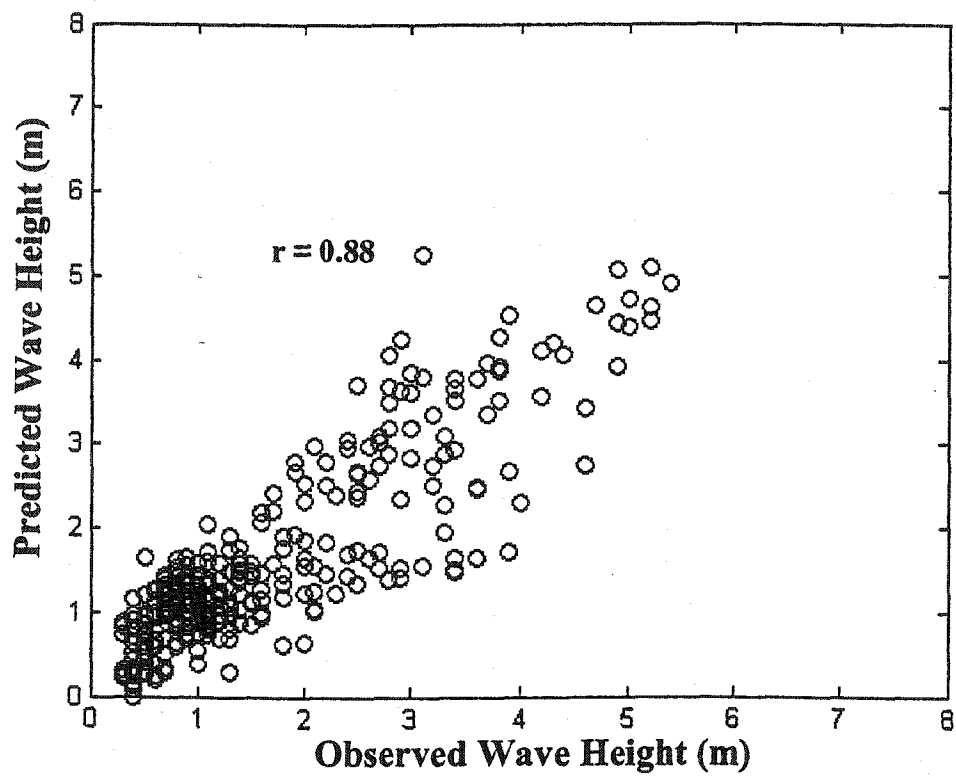


Fig. 5-24. Correlation coefficient between the observed and predicted significant wave heights for the Time Delay Neural Network with a global maximum and minimum wind speed of  $\pm 80$  m/s.

44025, 0.84 to 0.85 at station 44009, and from 0.60 to 0.65 at station 41009. Station 44013 remained at 0.96.

The TDNN did not greatly improve the prediction accuracy at station 41009. This implies that waves at station 41009 may be generated by another wind system developed in the area south of Florida. For instance, Fig. 5-25 shows the wind systems over the Northwest Atlantic Ocean on February 22 at 12:00 GWT, 1998. At this time, the southern wind system blowing westward extended to Florida and affected station 41009, leaving the mid-latitudes unaffected. In other words, the observed waves at station 41009 were not generated by the same winter-storms impacting northern stations. For this reason, for predicting wind-waves more accurately, more wind stations should be considered in the southern part of station 41009.

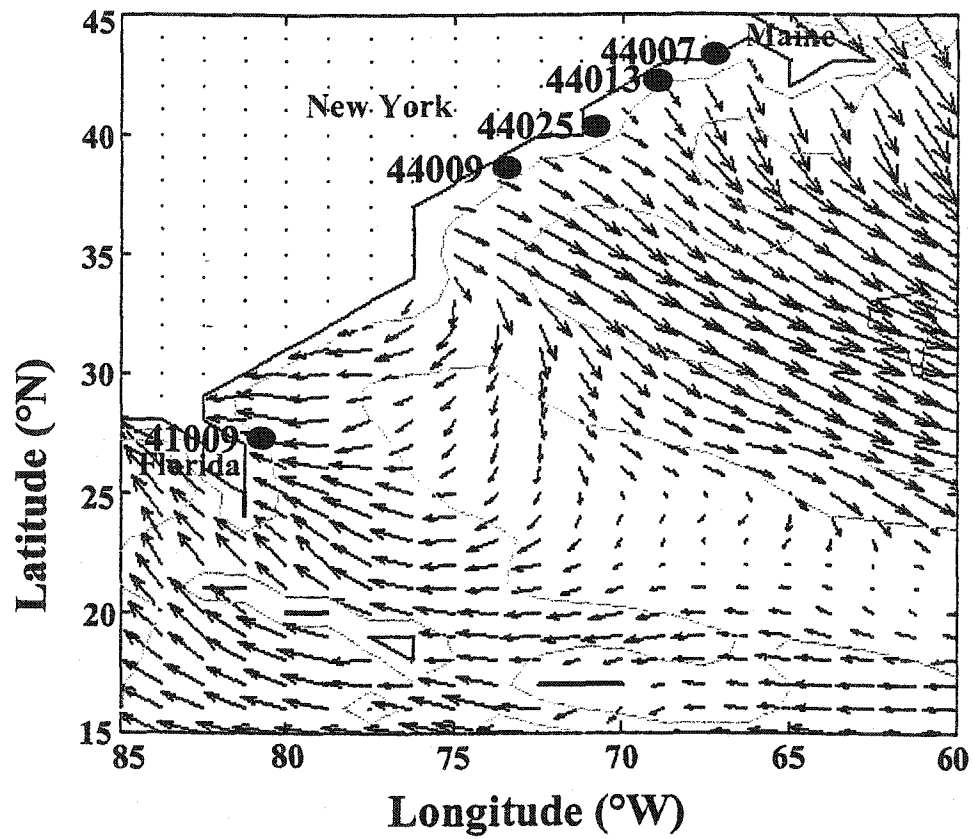


Fig. 5-25. Wind field over the northwest Atlantic Ocean on February 22, 12:00 GWT, 1998. Circles indicate the five wave stations along the East Coast of the U.S.

## 5.8. Results of Zero-Crossing Wave Period Prediction and Discussion

Since the Time Delay Neural Network shows the best performance, only the TDNN algorithm was used to study wave period prediction. The bias, variance, and gradient error curves were not used here for finding the optimum number of hidden neurons and iterations because those parameters did not show the optimum choices clearly from previous experiment. Instead, the optimum numbers were selected from the trial and error method by comparing correlation coefficients between observed and predicted zero-crossing wave periods directly.

### 5.8.1. Three-Layered TDNN

The Time Delay Neural Network (TDNN) model with scaled conjugate gradient learning algorithm was used to predict zero-crossing wave period. The total number of training data points was 403 from the winter storms in 1999 and 2001. The number of inputs was 80 ( $m = 80$ ), and 24 hours was used as the duration of time delays ( $J = 9$ ). Thus, the TDNN structure of  $I_{720}H_nO_5$  was used.

Figure 5-26 shows learning curves for different selections of hidden neurons and iterations with a training data set of 403 points. The MSE gradually improved as the number of iterations increased if the number of hidden neurons was more than three. This is similar pattern to that of wave height prediction.

The correlation coefficient between the observed and predicted wave heights was compared using different numbers of hidden neurons and iterations (Fig. 5-27). The TDNN model had the best correlation coefficient, 0.58 ( $r^2 = 0.34$ ), when it used

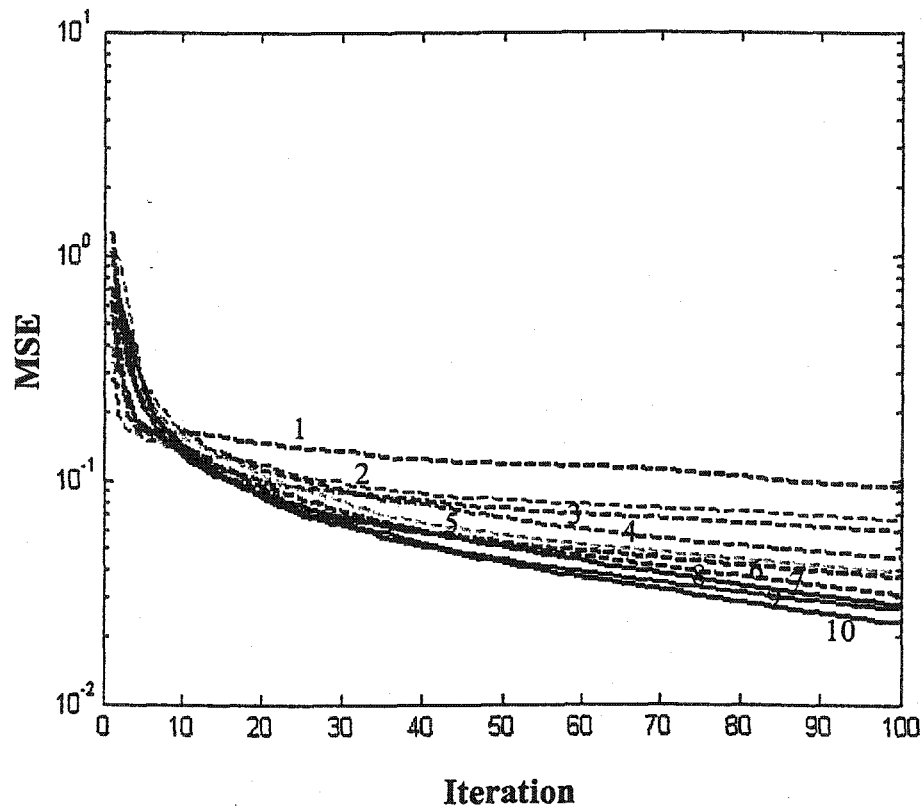


Fig. 5-26. Effects of the numbers of hidden neurons and iterations on the Mean Square Error (MSE) for the Time Delay Neural Network. The number of training data points was 403.

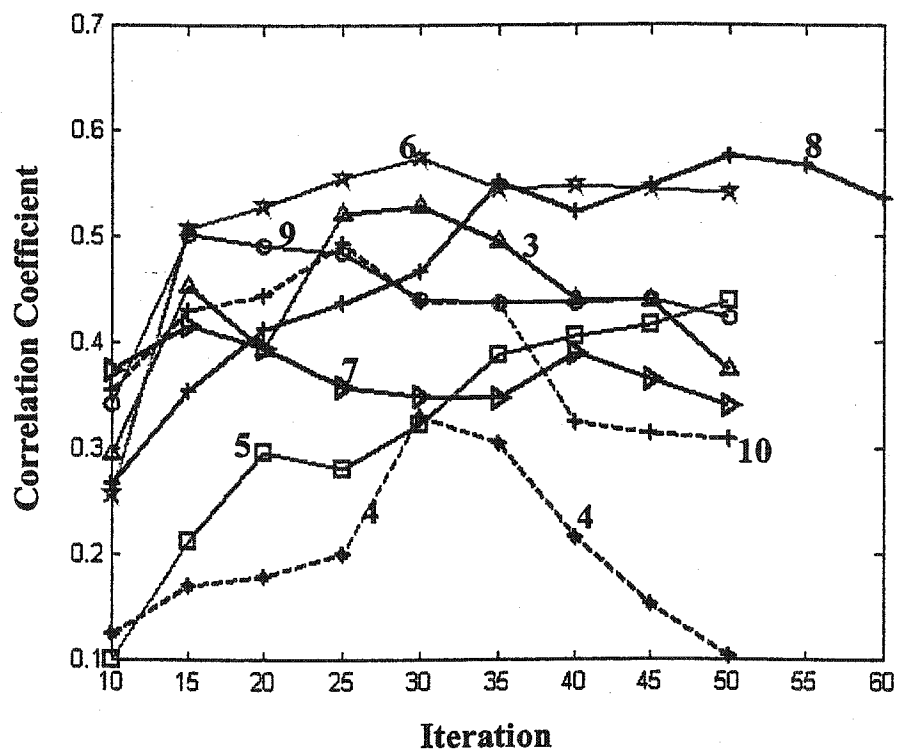


Fig. 5-27. Comparison of the correlation coefficient between the observed and predicted zero-crossing wave periods for the Time Delay Neural Network at different numbers of hidden neurons and iterations. The number on each line indicates the number of hidden neurons.

eight hidden neurons and 50 iterations. Thus, eight hidden neurons and 50 iterations were used for training and prediction.

Figure 5-28 plots the observed and predicted wave periods at five wave stations. In general, the TDNN predicted the general trend of the zero-crossing wave period reasonably well. When compared with wave heights, the temporal variation of the observed period was more complicated than wave height, and thus the prediction of zero-crossing wave period was much less accurate than that of wave height.

Figure 5-29 shows the relatively low correlation coefficient between the observed and predicted wave periods. Station 44025 had the maximum-correlation coefficient of 0.69 ( $r^2 = 0.48$ ), while the minimum correlation coefficient was 0.27 ( $r^2 = 0.07$ ) at station 41009 (Table 5-10).

The best correlation coefficient ( $r$ ) of 0.58 between observed and predicted wave periods was much less than that for wave heights (0.85). This indicates that wave period prediction is more difficult than wave height prediction because wave period is a more complicated physical process.

In order to observe the difference of ANN performances with range of wind input,  $\pm 80$  m/s was used as a global maximum and minimum for pre-processing using the same hidden neurons and iterations. In contrast to wave height, the prediction results of wave period were less accurate, so, a four-layered TDNN model, which uses an additional hidden layer, was tested in the next section.



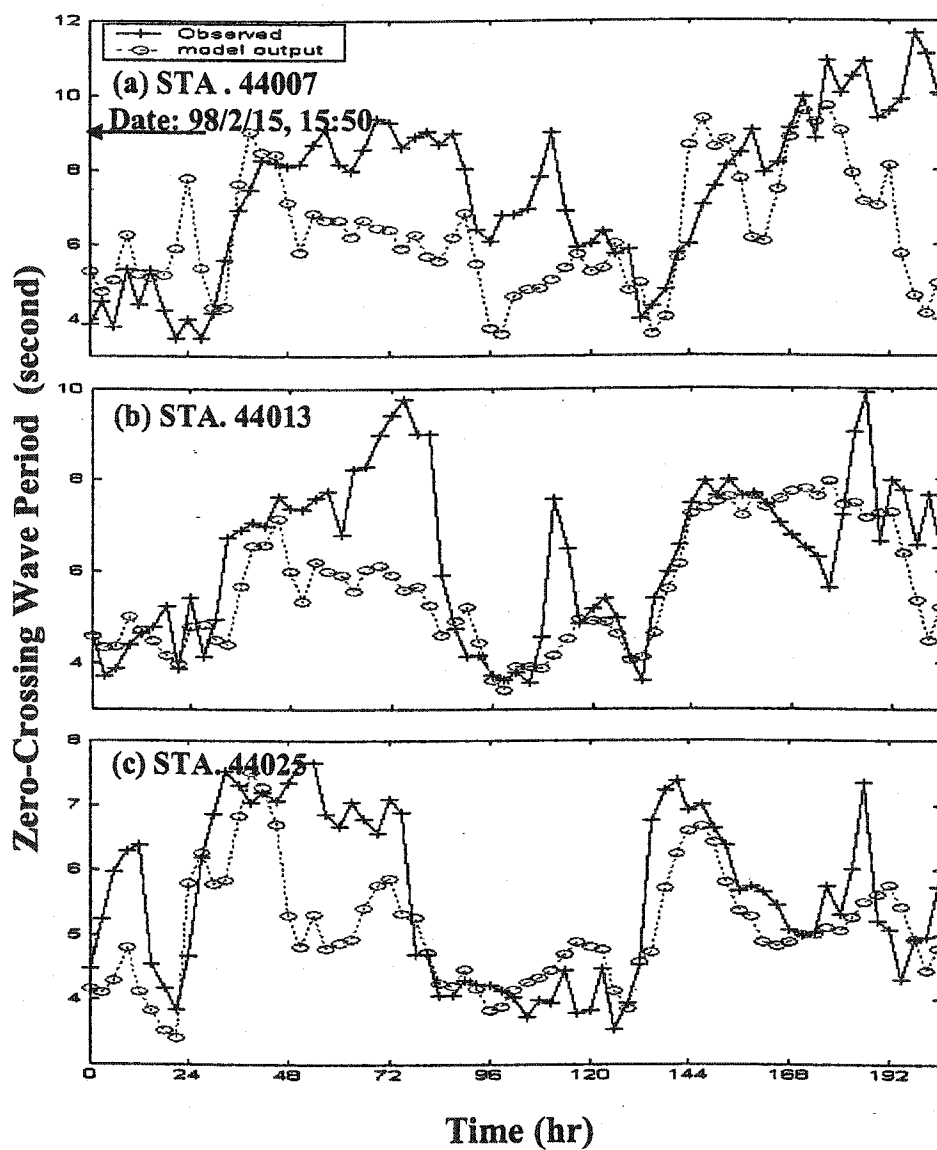


Fig. 5-28. Comparison of the observed and predicted zero-crossing wave periods using the Time Delay Neural Network with a structure of  $I_{720}H_8O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed was  $\pm 20$  m/s. The number of training data and iterations was 403 and 50.

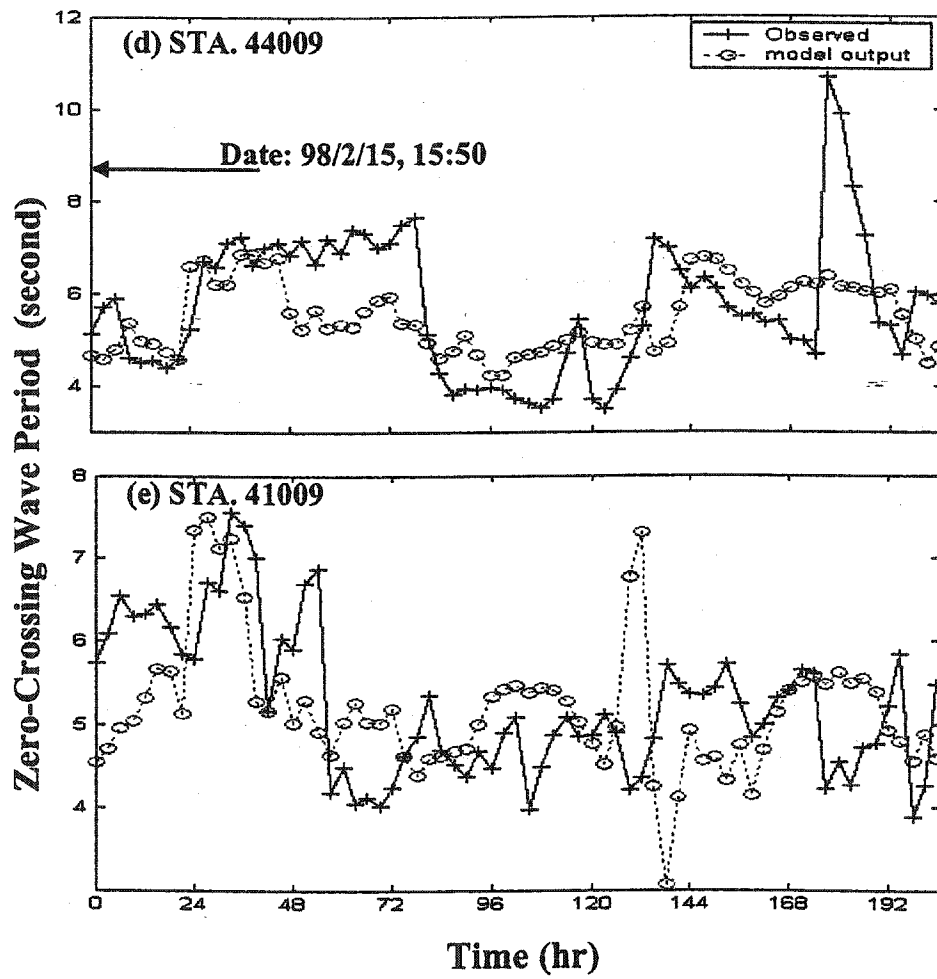


Fig. 5-28. (continued)

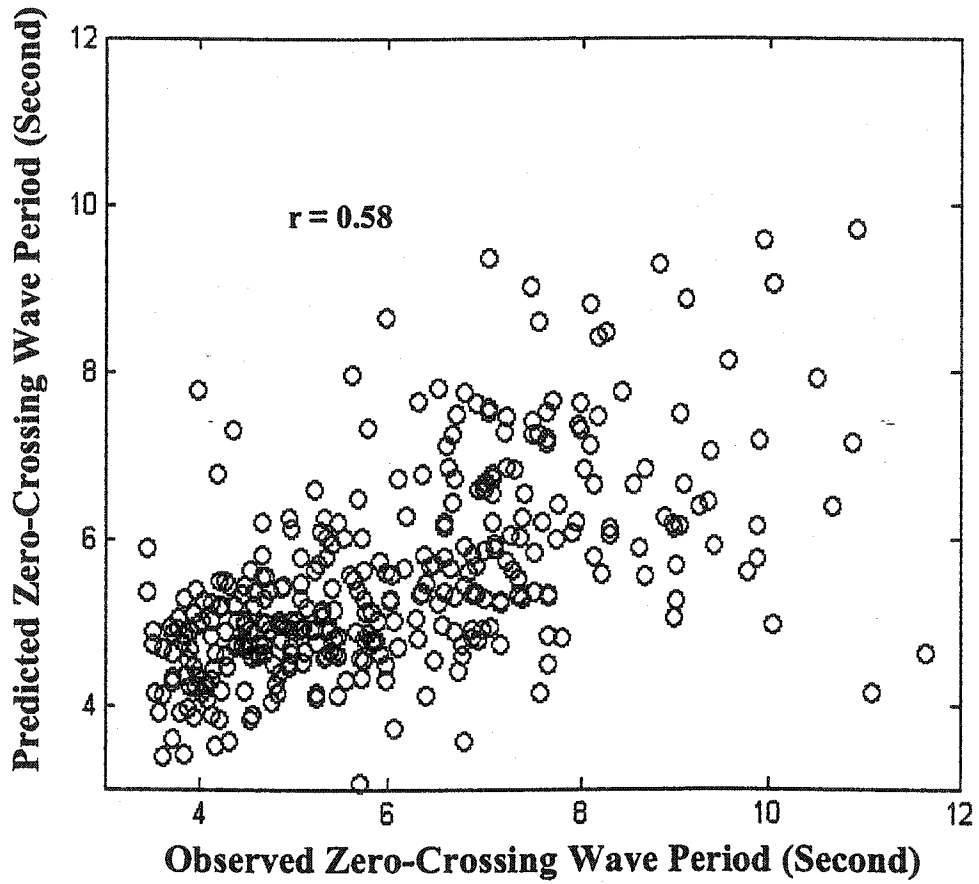


Fig. 5-29. Correlation coefficient between the observed and predicted zero-crossing wave periods for the three-layered Time Delay Neural Network. The number of training data points was 403.

Table 5-10

Correlation coefficient between the observed and predicted zero-crossing wave period  
at the five wave stations for the four-layered Time Delay Neural Network (TDNN)  
with 403 training data points.

Type	44007	44013	44025	44009	41009
Three-Layered TDNN	0.41	0.65	0.69	0.52	0.27
Four-Layered TDNN	0.63	0.64	0.58	0.49	-0.20

### 5.8.2. Four-Layered TDNN

One of advantages of ANN models is that it is easy to increase the number of hidden layers to accommodate the complexity of a system. A four-layered TDNN model was used because the initial prediction result of wave periods with a three-layered TDNN was not satisfactory. Because two hidden layers are used in the four-layered structure, the summation of weight matrix and transfer function are additionally processed between the first and second hidden layer (Fig. 5-30). Since all the inputs and outputs remain the same, the TDNN structure of  $I_{720}H_{n1}H_{n2}O_5$  was established.

Following results from the previous three-layered ANN, eight hidden neurons and 50 iterations were selected. It is not clear how to select the number of second hidden neurons, so, the trial and error method described next was tried.

Figure 5-31 compares the correlation coefficient between the observed and predicted wave periods at the five wave stations. The number of neurons in both hidden layers was increased from three to 10 in increments of one. The number of iterations increased from 10 with an increment of 10 between each trial until the coefficient decreased again. The best correlation coefficient ( $r$ ) of 0.61 ( $r^2 = 0.37$ ) was observed when six hidden neurons at the first layer, four hidden neurons at the second hidden layer, and 30 iterations were used.

Figure 5-32 plots the observed and predicted wave periods at the five wave stations. The prediction results were still not as good as those for wave height.

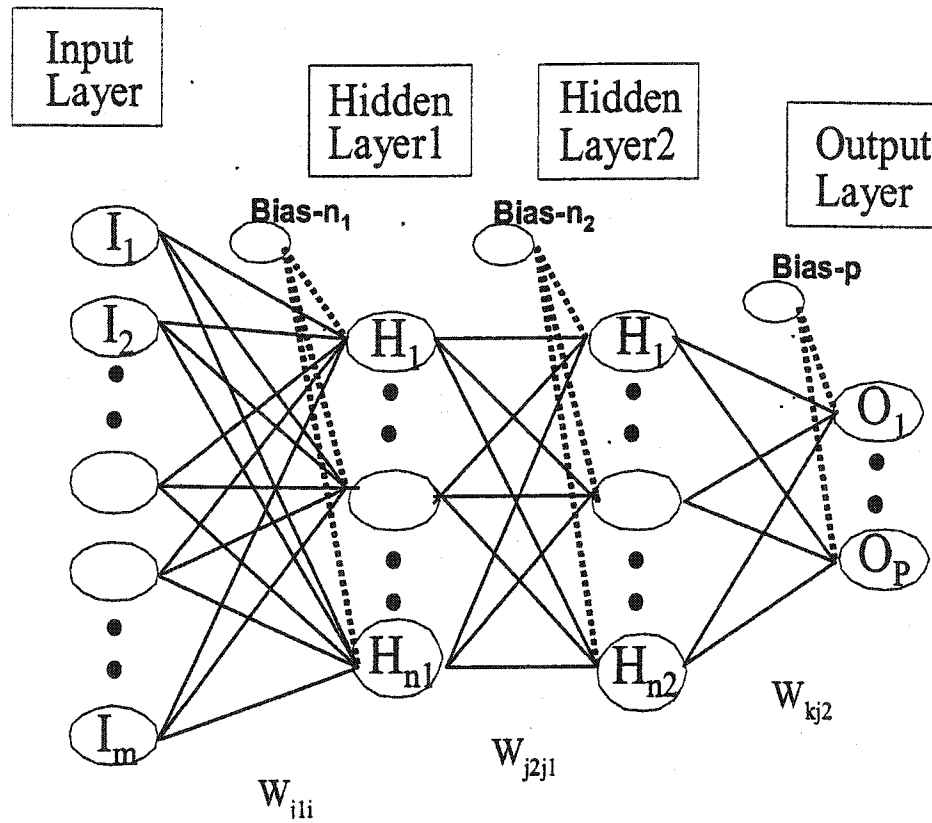


Fig. 5-30. A typical four-layered Time Delay Neural Network with a symbol  $I_m H_{n1} H_{n2} O_p$ . Solid lines indicate weights and dashed lines indicate biases.

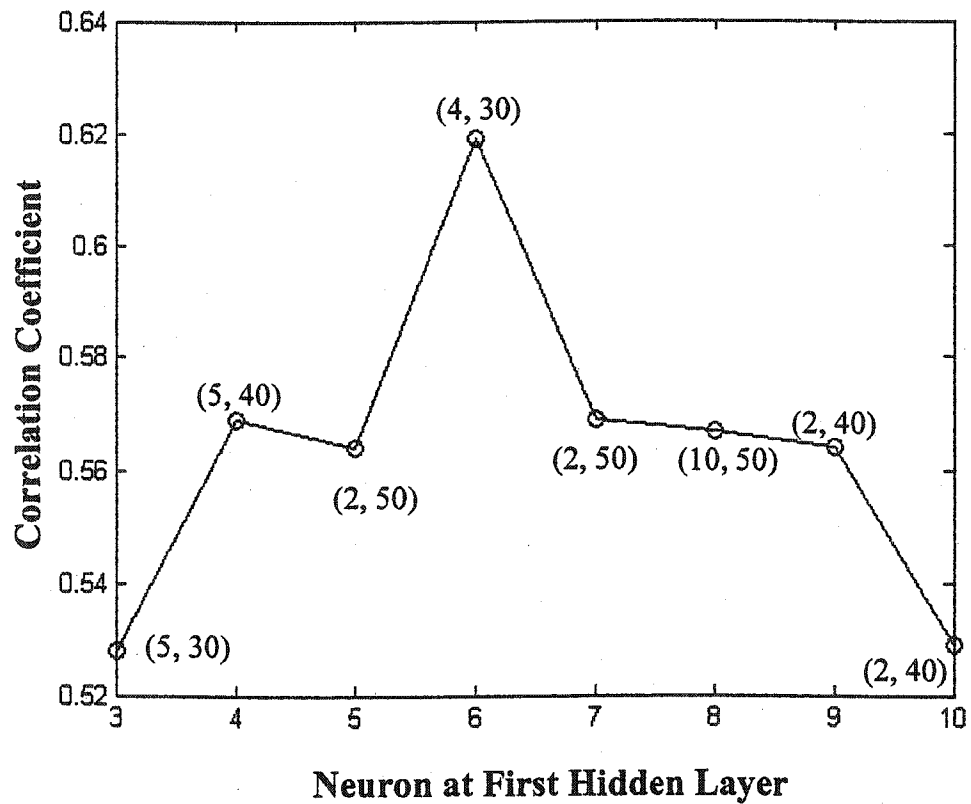


Fig. 5-31. Comparison of the correlation coefficient between the observed and predicted zero-crossing wave periods using the four-layered Time Delay Neural Network. The first and second numbers indicate the number of hidden neurons and iterations, respectively.

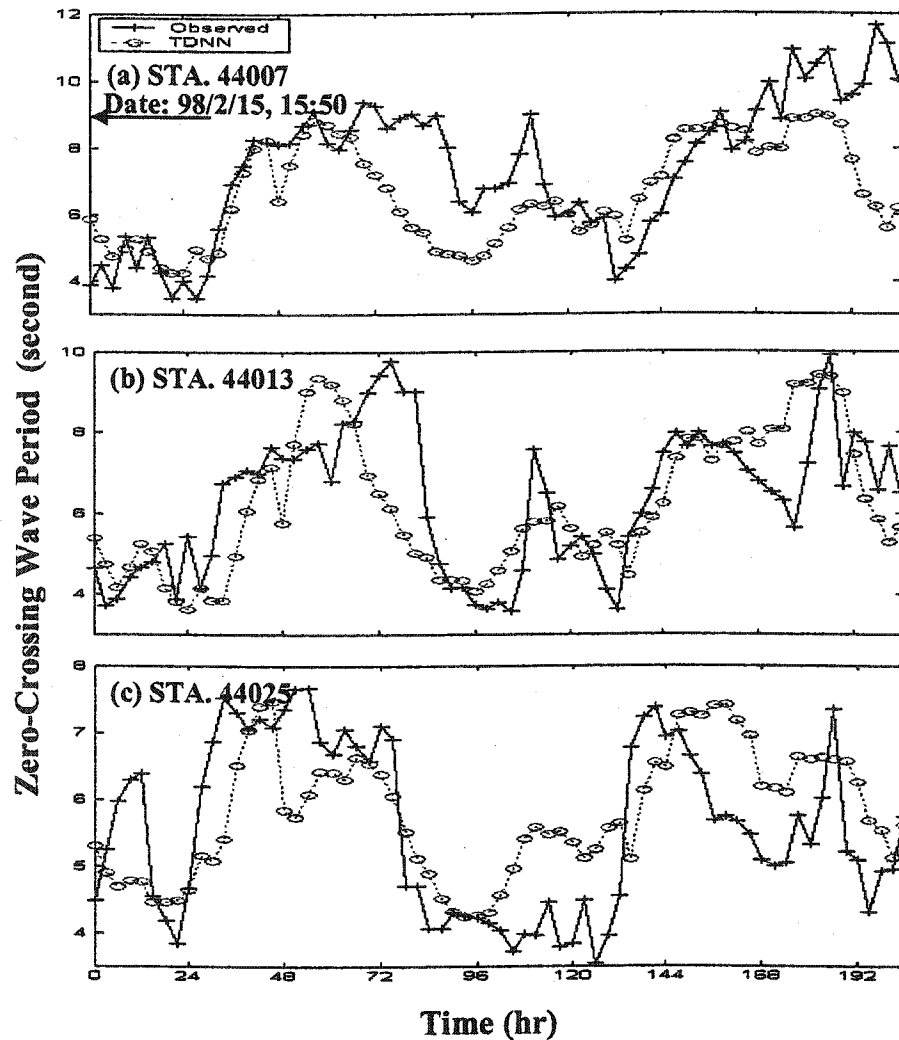


Fig. 5-32. Comparison of the observed and predicted zero-crossing wave periods using the Time Delay Neural Network with a structure of  $I_{720}H_6H_4O_5$ . (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A global maximum and minimum wind speed was  $\pm 20$  m/s. The number of training data and iterations was 403 and 30, respectively.



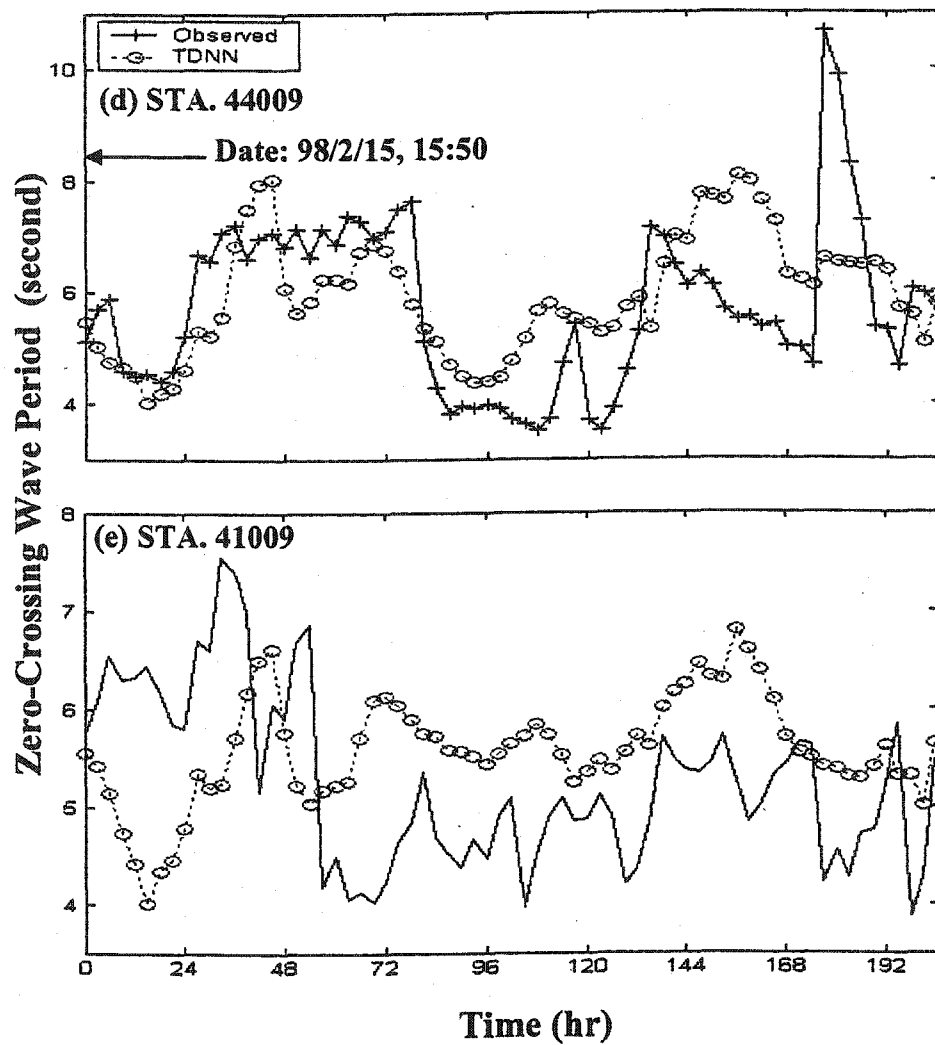


Fig. 5-32. (continued)

Figure 5-33 shows the correlation coefficient between the observed and predicted wave periods. The correlation coefficient improved slightly from 0.58 to 0.61 when two hidden layers were used. For the correlation coefficient at each wave station for the three-layered and four-layered TDNN, again see Table 5-10.

Only one additional hidden layer was tested. Theoretically, the number of hidden layers can increase infinitely according to the complexity of a given wind-wave system. But the improvement obtained with one additional hidden layer was small (about 5%), implying that further increase in the number of hidden layers may not be warranted.

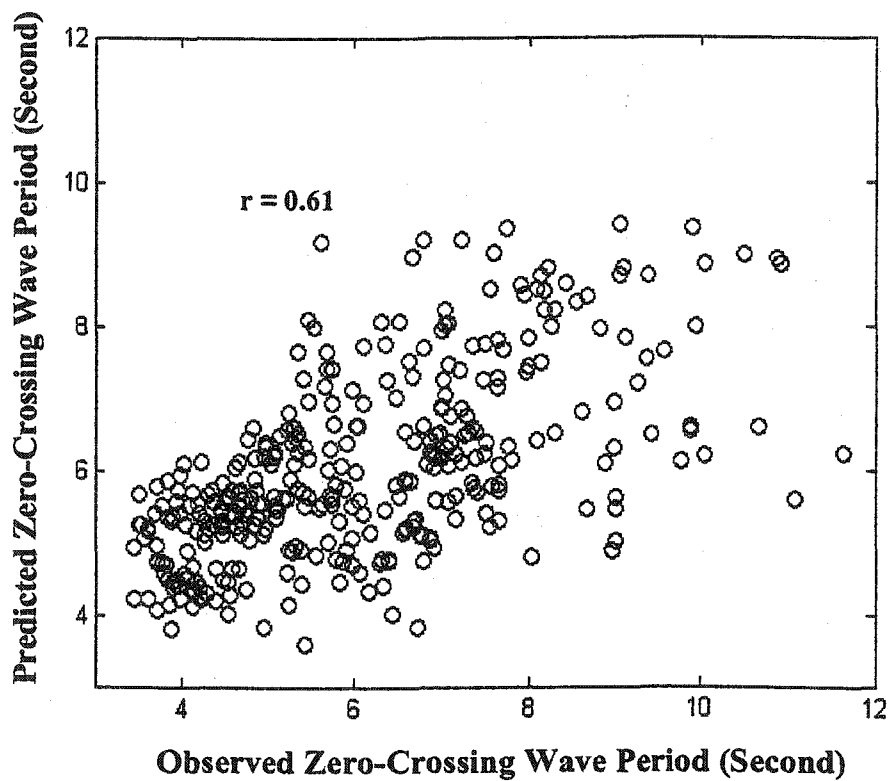


Fig. 5-33. Correlation coefficient between the observed and predicted zero-crossing wave periods for the four-layered Time Delay Neural Network. The number of training data and iterations was 403 and 30, respectively.

### 5.9. Conclusions

The prediction results of ANN wind-wave models are satisfactory for wave heights. Compared with simulated non-linear wind waves, the prediction of real-storm waves is more difficult because there are many possible patterns. The prediction accuracy of wave periods was much lower than that of wave heights for the TDNN model.

The prediction accuracy obtained by using the TDNN algorithm is much better than that for the other two approaches. This is because of the time delay function. The ERN algorithm has a limited short-term memory, thus produces a result that is better than the BPN, but cannot compete with the TDNN. Regarding model efficiency, ANN prediction results improved for both wave heights and periods as the number of training data and hidden neurons was increased.

The range of normalized wind speed is another factor that affects the results of wind-waves prediction. When global maximum and minimum winds were increased to  $\pm 80$  m/s, the TDNN could predict peak wave height more accurately at station 44013 and 44025, and had no negative wave height at station 44007. Moreover, the correlation coefficient between the observed and predicted wave height increased to 0.88 ( $r^2 = 0.77$ ), from the 0.85 ( $r^2 = 0.72$ ) for  $\pm 20$  m/s global wind. However, the prediction results were less accurate for wave periods when  $\pm 80$  m/s was used.

The prediction results of wave height and period for all types of ANN models were much less accurate at station 41009 because of the possibly different wind fields in the southern part of the Northwest Atlantic Ocean. The prediction results of wave

height for the TDNN are expected to improve if the training data set size is sufficient. This implies that the winter storms affecting more northern stations do not produce severe seas at station 41009 at all. The severe waves at station 41009 may be generated from somewhere else.

## CHAPTER VI

### ANN MODEL FOR HURRICANE WAVES

#### 6.1. Introduction

Cyclones have high wind speed with organized cloud systems and are differently named, depending upon the region in which they occur (Neumann, 1993). A hurricane is the name in the North Atlantic Ocean, the Northeast Pacific Ocean east of the International Date Line (IDL), or the South Pacific Ocean east of 160°E. Cyclones in the Northwest Pacific Ocean west of the IDL or a severe tropical cyclone in the Southwest Pacific Ocean west of 160°E and Southeast Indian Ocean east of 90°E are called typhoons. It is called a severe cyclonic storm in the North Indian Ocean and a tropical cyclone in the Southwest Indian.

In general, tropical cyclones are categorized according to the maximum wind speed. A tropical depression has an organized cloud system with thunderstorms and a maximum sustained wind of 33 kts or less. A tropical storm is an organized cloud system with strong thunderstorms, and a maximum sustained wind speed from 34 kts to 63 kts. A hurricane is an intense tropical weather system with strong thunderstorms, and a maximum sustained wind speed greater than 64 kts.

Atlantic hurricanes are further subdivided into the 5-category Saffir- Simpson Scale, which is used to give an estimate of the potential property damage and flooding along the U.S. east coast. Category 1 has wind speeds of 64 – 82 kts, which may cause minimum damage. Category 2 has wind speeds of 83 – 95 kts, which may cause moderate damage. Category 3 has wind speeds of 96 – 113 kts, which may produce extensive damage. Category 4 has wind speeds of 114 – 135 kts, which may generate extreme damage. Category 5 has wind speeds of over 135 kts, which may cause catastrophic damage.

Unlike winter storms, hurricanes move quickly with time, but their speed is not constant. The Hurricane Prediction Center (HPC) has provided the plots of hurricane tracks over the Northwest Atlantic Ocean since 1958. Their results show that some hurricanes accelerate as they move from tropical areas to subtropical and mid-latitude areas. Other hurricanes move slowly in tropical areas, faster in sub-tropical areas, and slow again in the mid-latitudes. The irregularity of the moving speed of hurricanes changes the duration of hurricanes at each particular site. The different duration at a particular site changes the waves because wave period and height are directly determined by the duration of hurricanes.

It is important to understand the characteristics of hurricanes before using an ANN model. If hurricanes have the same characteristics as those of winter storms, then the ANN technique developed in the previous chapter can be used directly. If not, the different characteristics of the hurricanes and the way to use those features of hurricanes must be understood and established.

Summer hurricanes and winter storms are similar in that both of them have high wind speeds. The characteristics of the wind systems and dynamics, however, are extremely different. The winter storm wind fields appear over a large domain. The wind speeds are relatively evenly distributed in the corresponding areas, and the wind speed and direction of the storm are nearly constant during each event. Also, the front of winter storm wind systems usually moves slowly. For these reasons, winter storm waves become fully developed seas. (For more information on winter storms, see chapter 5.1.)

In contrast, hurricanes usually move rapidly, and the area affected by hurricanes is restricted to a small domain when compared with winter storms. In the northern hemisphere, the maximum wind speed always lies in the northeast quadrant of hurricanes because wind speed to the right of hurricanes is greater than that to the left. Wind speed and direction at a specific location change drastically according to the time and location of hurricane centers. The effects of the hurricane motion on the wind field decrease as the distance from the area of maximum wind speed increases (Shore Protection Manual, 1977).

For this reason, the effects of wind energy on waves differ between hurricanes and storms. The wind energies of storms are transferred over large areas, so the distributions of wave heights can be observed in a wider area at a given time. Hurricane waves, however, are restricted to a small region because hurricanes have small but strong wind fields.



Moreover, because of the relatively fast movement of hurricanes, wave heights vary greatly depending upon the corresponding wind energy. The area in which wind speed and direction are reasonably constant is always very small, and so generated waves are rarely fully developed seas. Thus, prediction of hurricane waves is much more difficult than prediction of storm waves.

For instance, Fig. 6-1 shows the wind field of hurricane Floyd, which occurred from September 12 to 18, 1999, over the Northwest Atlantic Ocean. Only a restricted small domain was affected by strong hurricane winds. Figure 6-2 shows the observed significant wave heights at the five stations along the east coast of the U.S. during hurricane Floyd. The change of wave height at different wave stations corresponded with the hurricane track. For instance, as hurricane Floyd moved northward along the east coast, wave height rapidly increased at the nearest station (41009) to 10 m on September 15, 1999, while other stations showed low wave heights. Wave height gradually increased in the sequence of station 44009, 44025, 44013, and 44007. Hence, the only place at which the hurricane wind significantly affected the wave heights was over the area at the hurricane front line.

In this chapter, the results of hurricane waves prediction are given. Data description is given in section 6.2, input in section 6.3, proposed ANN structure in section 6.4, results of significant wave height prediction and discussion in section 6.5, results of peak wave period prediction and discussion in section 6.6, and conclusions in section 6.7.

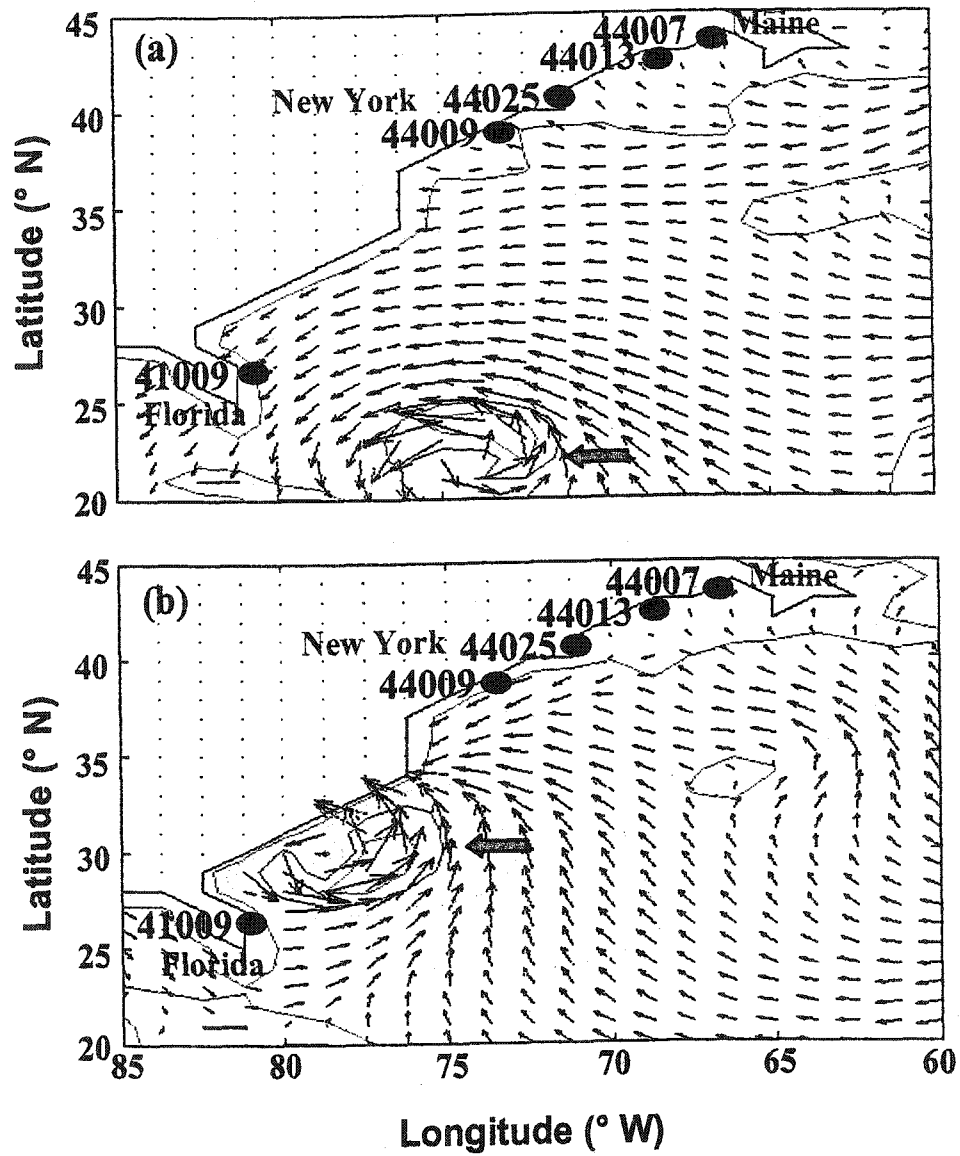


Fig. 6-1. Hurricane wind fields over the northwest Atlantic Ocean. (a) On September 14, 00:00 GWT, 1999 and (b) on September 16, 00:00 GWT, 1999. Circles indicate the five wave stations along the east coast of the U.S.

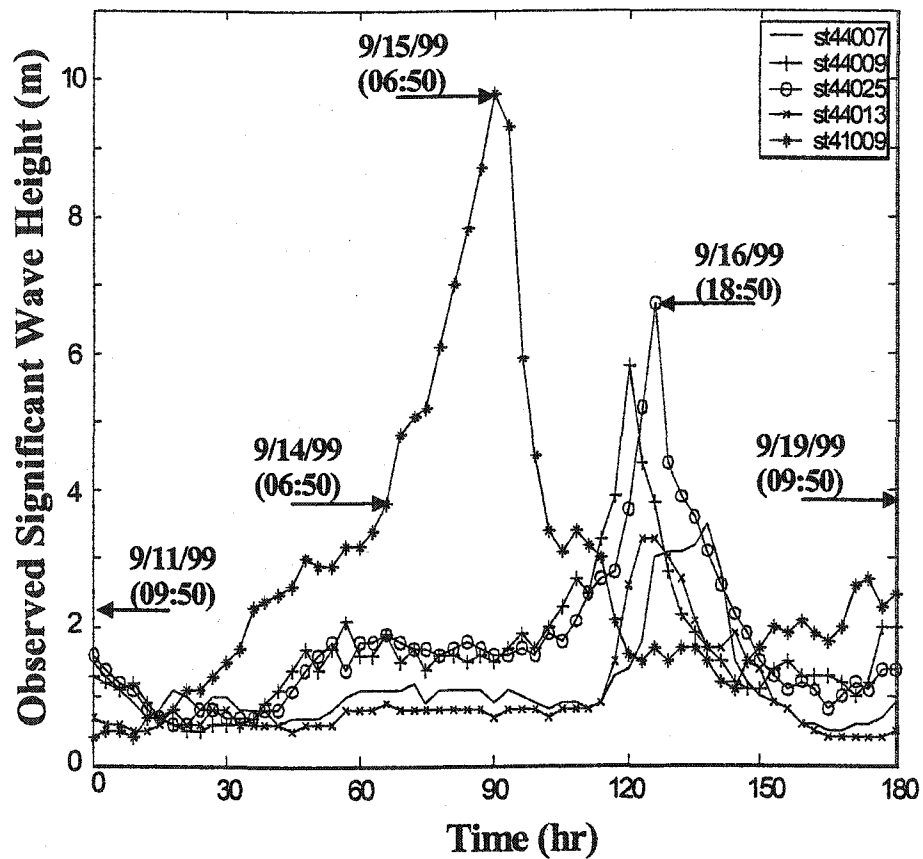


Fig 6-2. Observed significant wave height at the five wave stations (44007, 44009, 44025, 44013, and 41009) along the East Coast of the U.S. from September 11 to 19, 1999 during hurricane Floyd.

## 6.2. Data Description

### 6.2.1. Wind and Wave

The National Hurricane Center (NHC) has provided information on the maximum wind speed, central pressure, and latitude and longitude of hurricane centers since 1958. The Risk Prediction Initiative (RPI) has provided the Radius of Maximum Wind (RMW) as well as the above information since 1989.

The NHC and RPI presented the data in time intervals of six hours (*e.g.*, 00, 06, 12 and 18). The National Data Buoy Center (NDBC) has provided wave records since 1975 for station 44025, 1982 for station 44007, 1984 for station 44013 and 44009, 1988 for station 41009.

This study assembled available hurricane information from the Northwest Atlantic Ocean. The maximum wind speed, central pressure, location of hurricane centers, and the RMW were used as input conditions. The reasons why those inputs are essential for hurricane predictions are explained in the next section.

Both the HPC and RPI have presented those input data since 1989. The NDBC, however, had no completed hurricane wave records in 1989 and 1990. For this reason, only 23 hurricanes from 1991 to 2001 were used in this study (Table 6-1).

Table 6-1

Historical cyclonic events over the northwest Atlantic Ocean from 1988 to 2001

Year	No.	Name	Date	Wind Speed (kts)	Saffir-Simpson Scale	RMW	Wave Record
1988	1	Alberto	8/6-8/7	35	-	No	Yes
	2	Chris	8/26-8/29	80	1	No	Yes
1989	1	Hugo	9/19-9-22	140	5	Yes	No
1990	1	Bertha	7/24-8/1	70	1	Yes	No
	2	Lili	10/11-10/15	65	1	Yes	No
1991	1	Ana	7/2-7/4	45	-	Yes	Yes
	2	Bob	8/16-8/19	100	3	Yes	Yes
1992	1	Danniel	9/22-9/26	45	-	Yes	Yes
	2	Earl	9/27-10/3	55	1	Yes	Yes
1993	1	Emily	8/27-9/2	100	3	Yes	Yes
1995	1	Erine	7/30-8/2	80	1	Yes	Yes
	2	Marilyne	9/17-9/20	100	3	Yes	Yes
1996	1	Bertha	7/9-7/14	100	3	Yes	Yes
	2	Edourd	8/29-9/3	125	4	Yes	Yes
	3	Fran	9/2-9/9	105	3	Yes	Yes
	4	Hortense	9/11-9/14	120	4	Yes	Yes
1997	1	Ana	6/30-7/5	60	1	Yes	Yes
1998	1	Bonnie	8/21-8/28	100	3	Yes	Yes
	2	Danniel	8/28-9/2	90	2	Yes	Yes
1999	1	Dennis	8/24-9/8	90	2	Yes	Yes
	2	Floyd	9/12-9/18	130	4	Yes	Yes
2000	1	Florence	9/10-9/16	70	1	No	Yes
2001	1	Humberto	9/22-9/24	90	2	No	Yes

### 6.2.2. Hurricane Track Pattern

The effects of hurricane waves on the U.S. east coast differ according to the tracks followed by the hurricanes. For this reason, the selected 23 hurricane events were analyzed in six patterns of direction and location of the hurricane. Pattern 1 comes from the Bahamas and passes along the East Coast of the U.S. from Florida to Maine. Hence, the effects of this hurricane can be felt along the coast from the south to the north (Fig. 6-3a). Pattern 2 has a similar track to Pattern 1, but it crosses the continent between 30°N and 40°N and moves further North or Northeast. Thus, the effects of these hurricanes are restricted to the southern coastal area (Fig. 6-3b). Pattern 3 hurricanes come from Bahamas. It, however, bends to the Gulf of Mexico near Florida station 41009. Thus, the effects of 'Pattern 3' hurricanes would be restricted on the Florida region (Fig. 6-4a).

Pattern 4 includes many different types of hurricanes, which are generated far away from the East Coast in the Atlantic Ocean, move toward one particular coastal area, and then go back offshore or make landfall. The effects of these hurricanes concentrate on one particular region that depends on the hurricane track (Fig. 6-4b). Pattern 5 hurricanes form at southern and eastern locations far offshore and move parallel with the east coast of the U.S. Thus, they have little or no direct effect on the east coast of the U.S. (Fig. 6-5a). Pattern 6 has similar locations of hurricane formations as Pattern 5, but they move eastward toward open ocean. Thus, no effects of this pattern of hurricane are expected on the east coast of the U.S. (Fig. 6-5b).

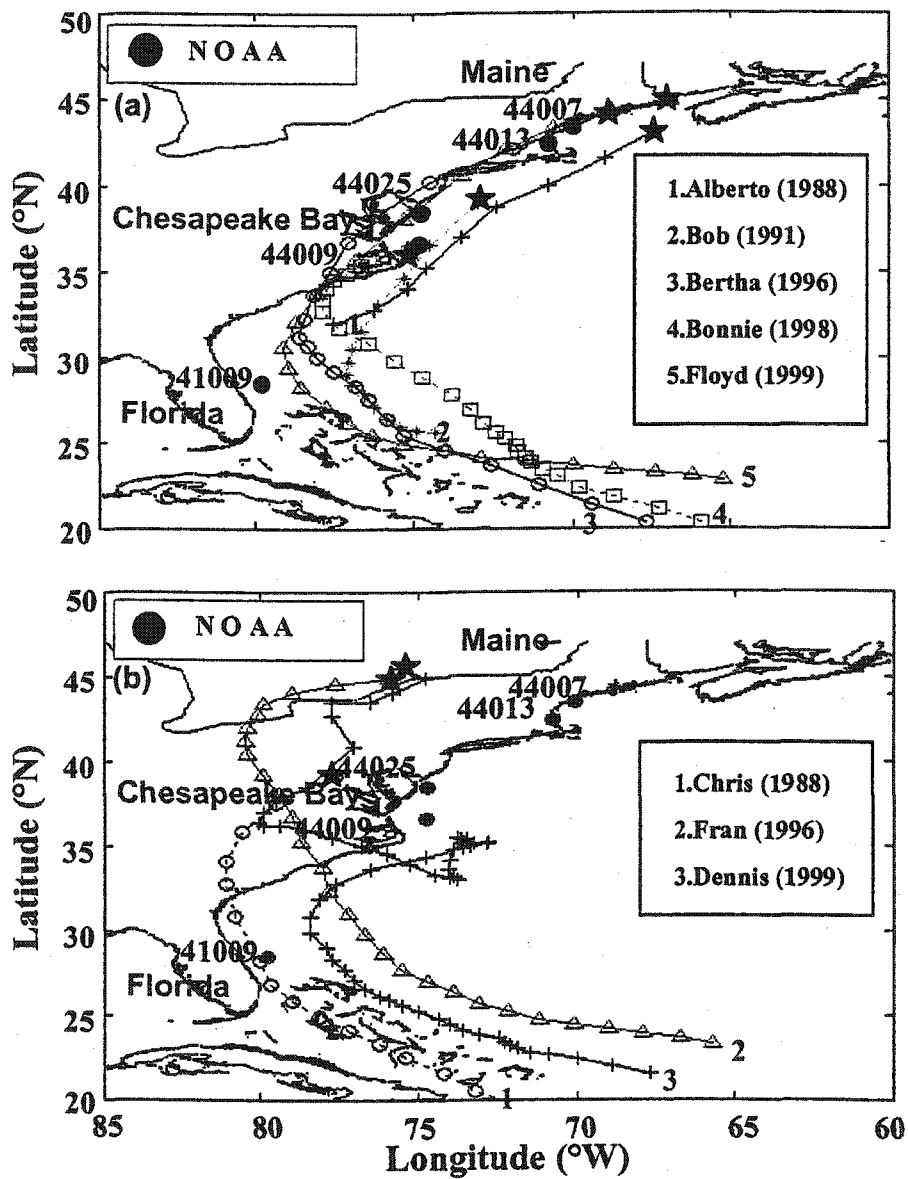


Fig. 6-3. 'Pattern 1' hurricane moves along the East Coast of the U.S. from Florida to Maine (a). 'Pattern 2' hurricane crosses the continent in the middle area between 30°N and 40°N (b).

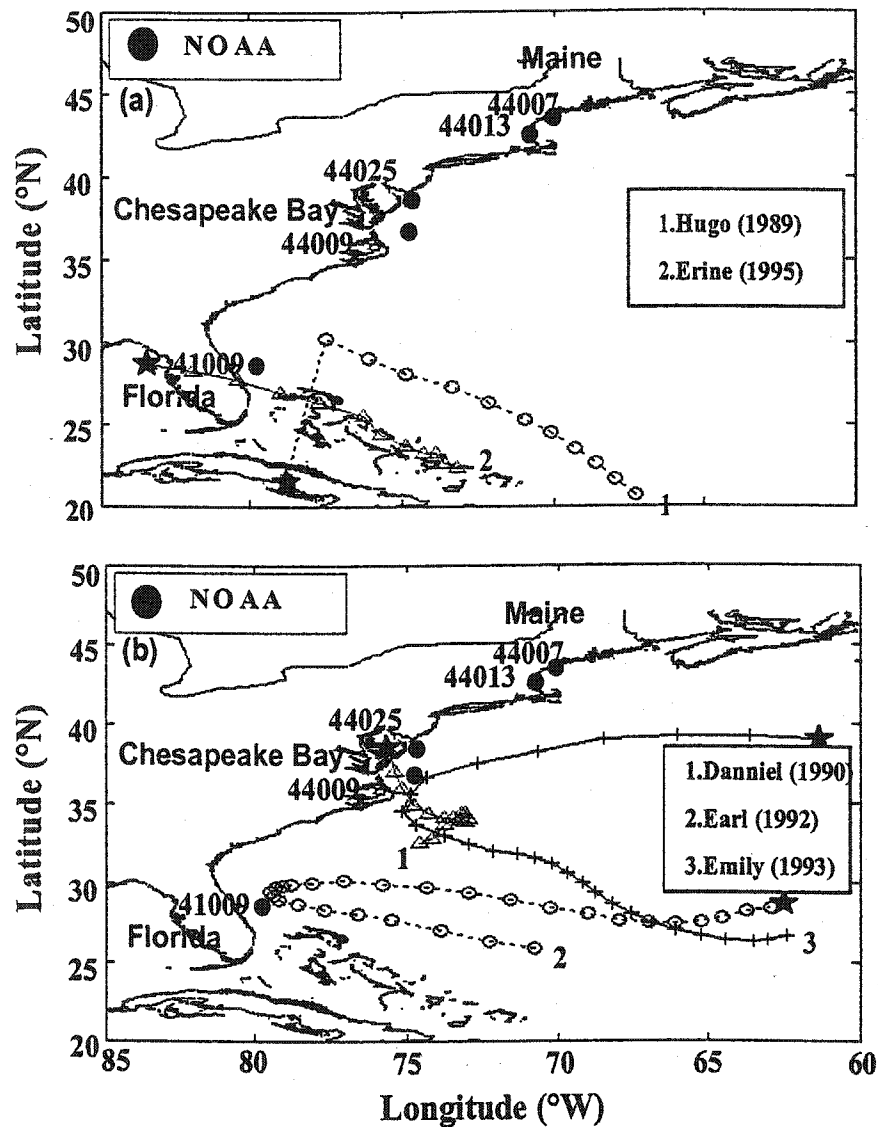


Fig. 6-4. (a) 'Pattern 3' hurricanes come from the Bahamas and turn left near Station 41009. (b) 'Pattern 4' includes many different types of hurricanes generated far away from the coast in the Atlantic Ocean, move toward one particular coastal area, then either go back offshore or cross the continent.



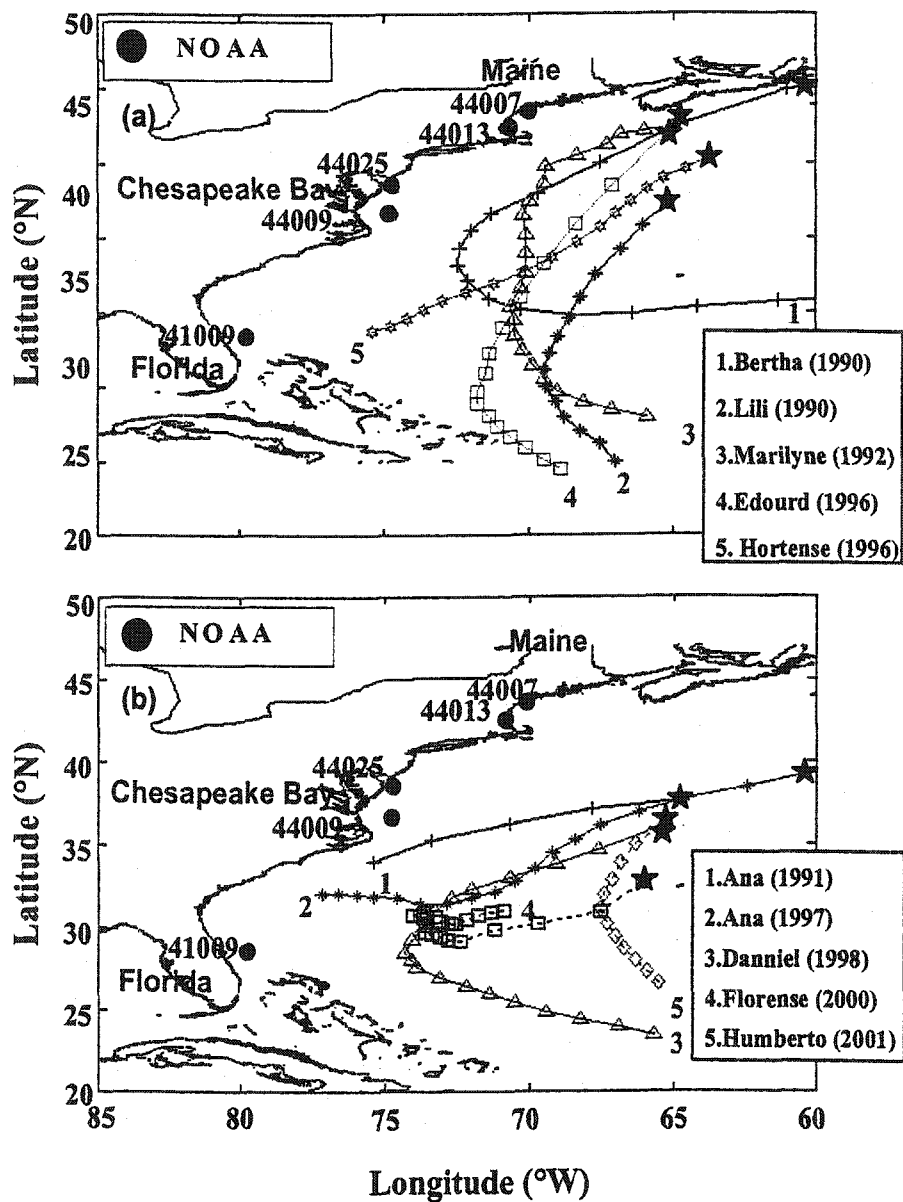


Fig. 6-5. 'Pattern 5' hurricane forms at southern and eastern far offshore and moves parallel with the East Coast of the U.S. (a). 'Pattern 6' has a similar location of hurricane track to 'Pattern 5', but it moves eastward open ocean (b).

### 6.3. Input

In order to find optimum input conditions for an ANN model, currently used hurricane models were examined (e.g., the Bretschneider wave model, and the Sea, Lake and Overland Surges from Hurricanes (SLOSH) model).

The following Bretschneider wave model estimates wave height and period at the position of maximum wind speed of a slowly moving hurricane (Shore Protection Manual, 1977):

$$H_s = 16.5e^{\frac{RMW\Delta P}{100}} \left[ 1 + \frac{0.208\alpha V_F}{\sqrt{U_R}} \right] \quad (6-1)$$

$$T_s = 8.6e^{\frac{RMW\Delta P}{200}} \left[ 1 + \frac{0.104\alpha V_F}{\sqrt{U_R}} \right] \quad (6-2)$$

where  $H_s$  = significant wave height (in feet)

$T_s$  = corresponding significant wave period (in seconds)

RMW = radius of maximum wind (in nautical miles)

$\Delta P = P_n - P_o$ , where  $P_n$  is the normal pressure of 29.92 inches of mercury, and

$P_o$  is the central pressure of the hurricane

$V_F$  = forward speed of the hurricane in knots

$U_R$  = maximum sustained wind speed in knots, specified at 10 m above mean sea surface at radius R where

$U_R = 0.865 U_{\max}$  (for a stationary hurricane)

$U_R = 0.865 U_{\max} + 0.5V_F$  (for a moving hurricane)

$U_{\max}$  = maximum gradient wind speed at 10 m above the water surface

$$U_{\max} = 0.868 [73(P_n - P_o)^{1/2} - R(0.575f)]$$

$f$  = coriolis force ( $= 2\omega \sin\phi$ )

$\alpha$  = a coefficient depending on the forward speed of the hurricane and the increase in effective fetch length, because a hurricane is moving. It is suggested that  $\alpha = 1.0$  for a slowly moving hurricane

The Hurricane Prediction Center (HPC) currently uses the SLOSH model to estimate storm surge heights and winds from historical, hypothetical, or predicted hurricanes. The SLOSH model uses the maximum wind speed, difference between central and normal pressures (29.92 inches of mercury), latitude of the hurricane centers, and RMW as inputs (Jelensnianski, 1984; Jarvinen and Lawrence, 1985).

The above two models have several important parameters in common: the maximum wind speed, central pressure, location of hurricane centers, and RMW. In other words, those parameters are the keys to understanding the hurricane characteristics of rapid temporal-spatial changes. For this reason, those four parameters were selected as major inputs for an ANN model in this study.

#### 6.4. Proposed ANN Structure

The Time Delay Neural Network (TDNN) with the scaled conjugate gradient learning algorithm was used to predict hurricane significant wave height and peak period at the five wave stations (*i.e.*, station 44007, 44013, 44025, 44009, and 41009) along the east coast of the U.S. Thus, the number of output neurons was five ( $p = 5$ ).

The maximum wind speed, central pressure, and position of the hurricane centers were used as ANN inputs. The latitude and longitude of hurricane centers were converted into x- and y-directional distances (km) from the five wave stations, *i.e.*, 10 data points for each position, increasing the total input neurons to 12 ( $m = 12$ ). If the RMW were additionally used as an input, the number of input neurons would be 13 ( $m = 13$ ). A trial and error method was used for finding the optimum number of hidden neurons ( $n$ ) by checking the ANN performance.

In contrast to winter storms, hurricanes are strongly non-stationary because of dramatic change in the maximum wind speed and central pressure with time and location. For this reason, instead of fixing one time-delay for hurricane waves, zero, six, 12, 18, 24-hour time delays were tested to find the best for ANN hurricane-waves prediction modeling. Hence, the TDNN structures of  $I_{12J}H_nO_5$  and  $I_{13J}H_nO_5$ , where  $J = 1, 2, 3, 4,$  and  $5$ , were established.

## 6.5. Results of Significant Wave Height Prediction and Discussion

As previously mentioned, prediction results differ according to different input parameters, number of training data, and time-delay. For this reason, the effects of input parameter (*e.g.*, the Radius of Maximum Wind), training data points, and time-delay were observed in this section.

The RMW is one of the major input parameters used to estimate hurricane wave height and period, as previously discussed. However, no equations can estimate the RMW exactly. For instance, Jelesnianski *et al.* (1992) and Vickery *et al.* (2000) provided some equations that use the central pressure and latitude of hurricanes as input to estimate the RMW, but the correlation coefficient between observed and predicted RMW was unacceptably low (less than 0.3). This indicates that those models not only are inaccurate for predicting the RMW, but also that the RMW is not closely related to the central pressure and latitude of hurricanes. However, the RMW is still essential to generate predictions of hurricane waves in the ocean. This problem may raise the question, "Can the RMW improve the prediction accuracy of an ANN model?" For this reason, the effects of the RMW were tested in next section.

### 6.5.1. Effects of RMW as Input

The Radius of Maximum Wind (RMW) is the distance from hurricane centers in which wind speed is usually zero to the place in which a maximum wind speed occurs. Jelesnianski *et al.* (1992) provided an equation to estimate the RMW along the east coast of the U.S. The equation was based on the fact that the RMW is

directly changed according to the position of the hurricane center and central pressure. They proposed two equations for two different cases: one applies when hurricane centers are at 30°N and the other applies when hurricane centers are south or north of 30°N.

Vickery *et al.* (2000) combined two different formulas into one comprehensive equation, the so-called global model. Hence, users can use only one equation to calculate the RMW, regardless of the location of hurricane centers. They provided three different global models to estimate the RMW along the east coast of the U.S.:

$$\ln(\text{RWM}) = 2.636 - 0.00005086\Delta P^2 + 0.0394899\psi, r^2 = 0.2765 \quad (6-3)$$

$$\ln(\text{RWM}) = 2.097 - 0.0187793\Delta P + 0.00018677\Delta P^2 + 0.0381328\psi, r^2 = 0.2994 \quad (6-4)$$

$$\ln(\text{RWM}) = 2.173 + 0.0056748\Delta P + 0.0416289\psi, r^2 = 0.2544 \quad (6-5)$$

where  $\ln$  = natural log

$P$  = central pressure

$\psi$  = latitude of the hurricane center.

According to the above equations, the RMW is determined by the central pressure and latitude of hurricanes. However, notice that the correlation coefficients,  $r$ , of the three equations are relatively low, indicating that the RMW is poorly correlated those inputs.

Wind-generated waves are mainly affected by the intensity of maximum wind speed and central pressure of hurricanes. If the RMW is not directly related to wind-generated waves, then there is no reason to include it as an input for an ANN hurricane wave prediction model.

Another question still remains. "Is the role of the RMW for an ANN model and the SLOSH wave model the same?" For the SLOSH model, the RMW is used to compute numerical processes. If other inputs are uncertain, the initial RMW can be adjusted to fit the model outputs. Therefore, it is relatively unimportant whether the RMW is correct or not for use in the SLOSH model.

On the contrary, an ANN model uses the RMW as an independent input. In this study, the central pressure, maximum wind speed, position of hurricane centers, and the RMW were used as inputs for the ANN model to find the importance of RMW compared with other inputs.

In case that the RMW is not a major factor that affects wind-waves, to use the RMW as input may not be necessary. Moreover, the ANN model might produce irrelevant weight values if the RMW is used unnecessarily. The result might have higher error or lower correlation coefficient value between observed and predicted waves.

For this reason, the objective of this experiment is to identify whether the RMW is, in fact, an important input for an ANN wind-wave prediction model. Sixteen hurricane events with 327 data points were collected from 1991 to 1999. Among those 327 data points, 21 points for hurricane Bertha in 1996 were used for a

validation data set to compare with the prediction. The other 306 points were used as a training data set. Hence, prediction data points comprise 6.9 % of the training data points.

A 24-hour time delay ( $J = 5$ ) was used for a Time Delay Neural Network (TDNN) model. The maximum wind speed, central pressure, and longitudinal and latitudinal distances between hurricane centers and the five wave stations were used. Thus, the TDNN structure of  $I_{60}H_nO_5$  was established. For another input condition, the RMW was added to the basic 12 inputs with the same duration of time delay. Thus, the total number of inputs was increased to 13 yielding another TDNN structure of  $I_{65}H_nO_5$ .

For finding the best correlation coefficient between the observed and predicted hurricane wave heights, the number of hidden neurons was changed from three to 10 with an increment of one hidden neuron, and the number of iterations was changed from 10 to 100 with increments of 10 iterations between each trial. The optimum number of hidden neurons was nine or 10, whether the TDNN used the RMW, or not.

Figure 6-6 shows the observed and predicted significant wave heights together at five wave stations for the TDNN structure of  $I_{65}H_9O_5$  at 70 iterations when the RMW was used as an input or  $I_{60}H_{10}O_5$  at 60 iterations without the RMW. The time series of predicted wave heights generally agreed with that of observed wave heights.

Figure 6-7 shows the correlation coefficient ( $r$ ) between the observed and predicted wave heights. The best  $r$  was 0.75 ( $r^2 = 0.56$ ), and the Mean Square Error (MSE) was 0.0516 when the RMW was used as an input. The best correlation



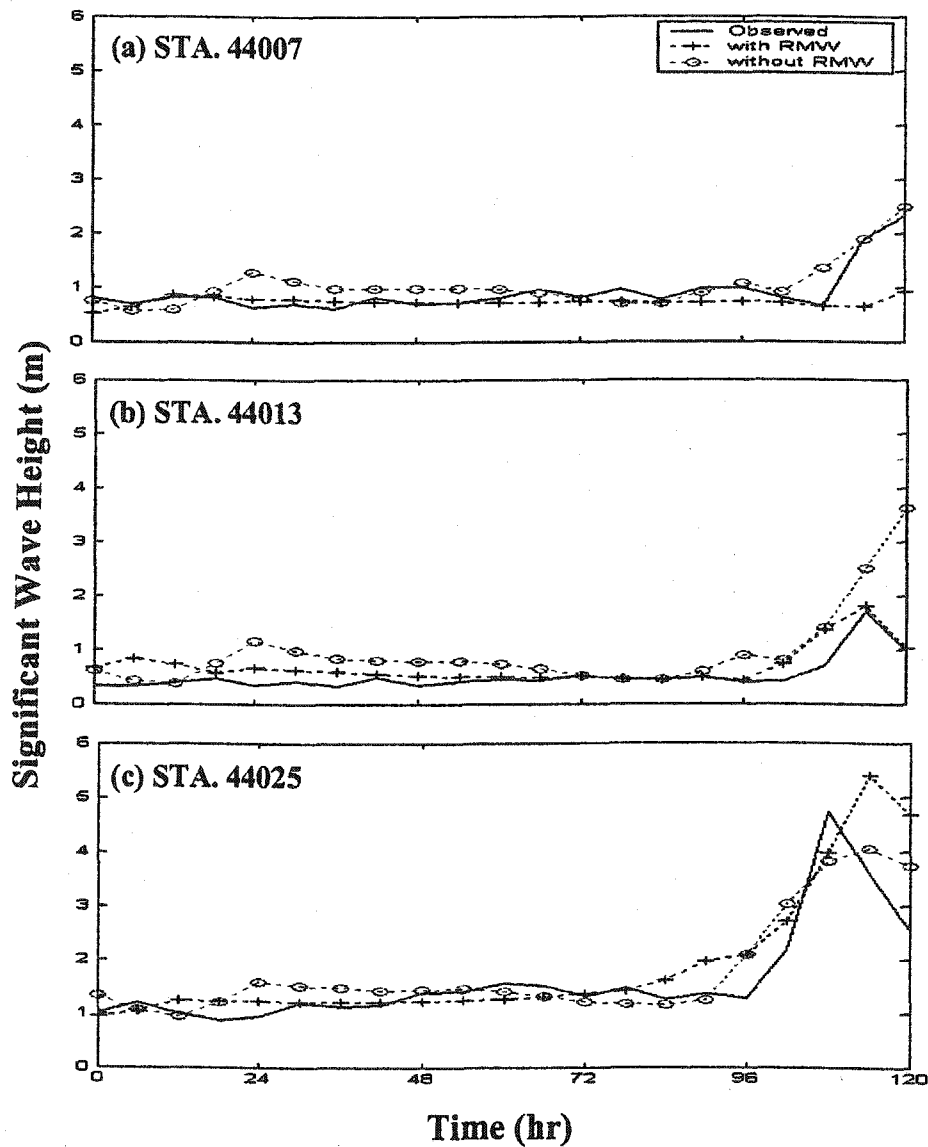


Fig. 6-6. Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of  $I_{65}H_9O_5$  with Radius of Maximum Wind (RMW) as Input at 70 iterations or  $I_{60}H_{10}O_5$  without RMW at 60 iterations. (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A 24-hour time-delay was used.

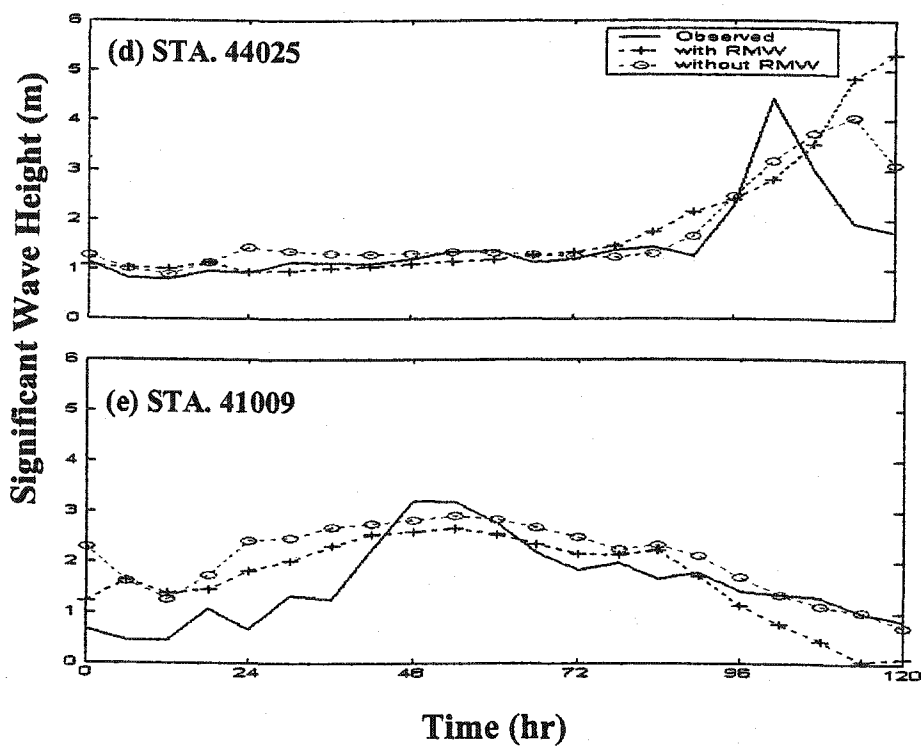


Fig. 6-6. (continued)

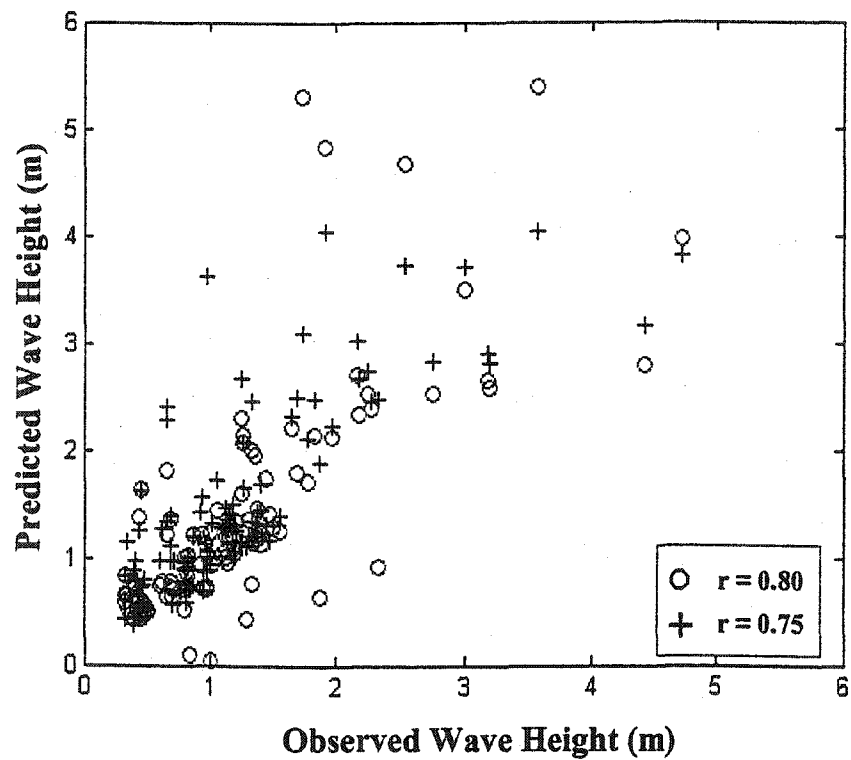


Fig. 6-7. Correlation coefficient between the observed and predicted significant wave heights. Pluses indicate the results with the Radius of Maximum Wind speed (RMW) used as an input, and circles indicate the results without RMW.

coefficient increased up to 0.80 ( $r^2 = 0.64$ ), and the MSE decreased to 0.0503 when the TDNN did not use the RMW.

Although the RMW is one of the important input parameters for the mathematical and numerical models, it does not improve the prediction capability of an ANN model for hurricane wave heights. The reason why the RMW does not help to improve prediction is that, as previously explained, the RMW has no strong relationship between the position of hurricane centers and central pressure (Vickery *et al.* 2000). For this reason, the RMW is not an essential input condition for an ANN hurricane wave prediction model. On the other hand, the maximum wind speed clearly increases as the central pressure of hurricanes decreases (Fig. 6-8).

#### 6.5.2. Effects of Number of Training Data Point

Theoretically, an ANN model can predict more accurately if it was trained with more data. For a validation data set, 21 data points from hurricane Bertha in 1999 were used. For training data, two different training data sets were prepared: (1) 15 hurricanes from 1991 to 1999 with 306 data points; (2) 12 hurricanes from 1995 to 2001 with 287 data points. Therefore, the ratio of the number of training data to validation data is 6.9 % and 7.3 % respectively.

The maximum wind speed, central pressure, and longitudinal and latitudinal distances of hurricane centers from the five wave stations were used as input factors (excluding the RMW).

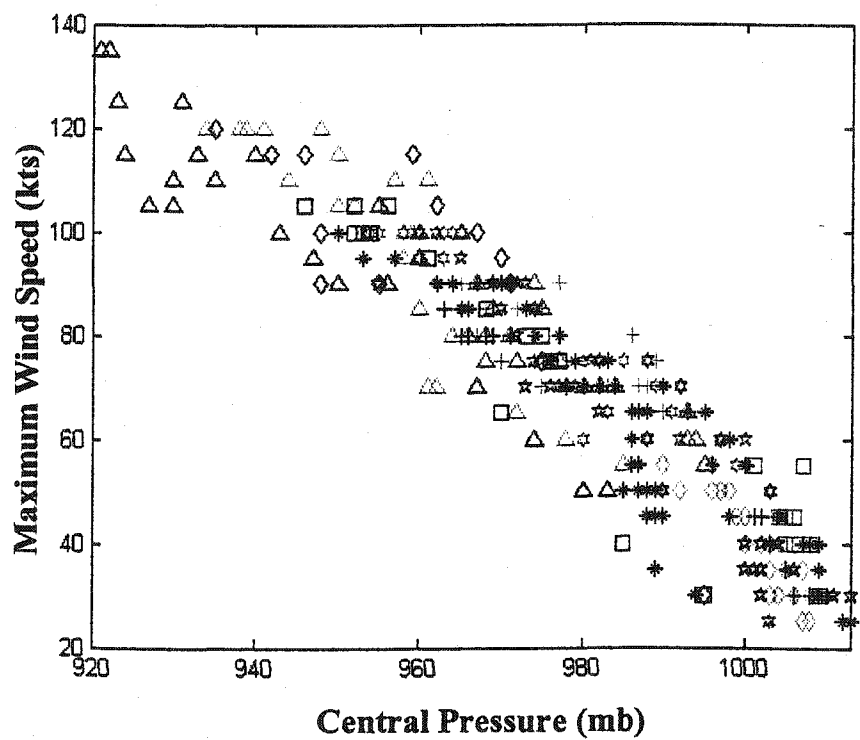


Fig. 6-8. Relationship between the maximum wind speeds and central pressures for the 16 hurricanes from 1995 to 2001.

A 24-hour time delay ( $J = 5$ ) was used for input conditions. Thus, the TDNN structure of  $I_{60}H_nO_5$  was used.

The number of hidden neurons was changed from three to 10 with an increment of one hidden neuron between each trial to find the best correlation coefficient between observed and predicted hurricane wave heights. The number of iterations was changed from 10 to 100 in increments of 10. The optimum number of hidden neurons was observed as nine or 10 when 12 or 15 hurricanes were used for training, respectively.

Figure 6-9 compares the observed and predicted wave heights at the five stations. The TDNN structures of  $I_{60}H_9O_5$  at 100 iterations or  $I_{60}H_{10}O_5$  at 60 iterations were used when 12-hurricanes with 287 data points or 15-hurricanes with 306 points were used for training, respectively. In general, the temporal variation of predicted wave height is similar to that of observed wave height.

Figure 6-10 shows the correlation coefficient between the observed and predicted wave heights. The best correlation coefficient was 0.82 at nine hidden neurons and 100 iterations, with a Mean Square Error (MSE) of 0.0394 when 12 hurricane events were used as training data. The correlation coefficient, however, was 0.80 and the MSE increased to 0.0451 when 15 hurricanes with 306 points were used with 10 hidden neurons and 60 iterations.

The number of data points of the 12-hurricane data set is 7 % less than that from the 15-hurricanes data set. Theoretically, an ANN should predict wave heights better as the number of training data increases. The test results here, however, do not

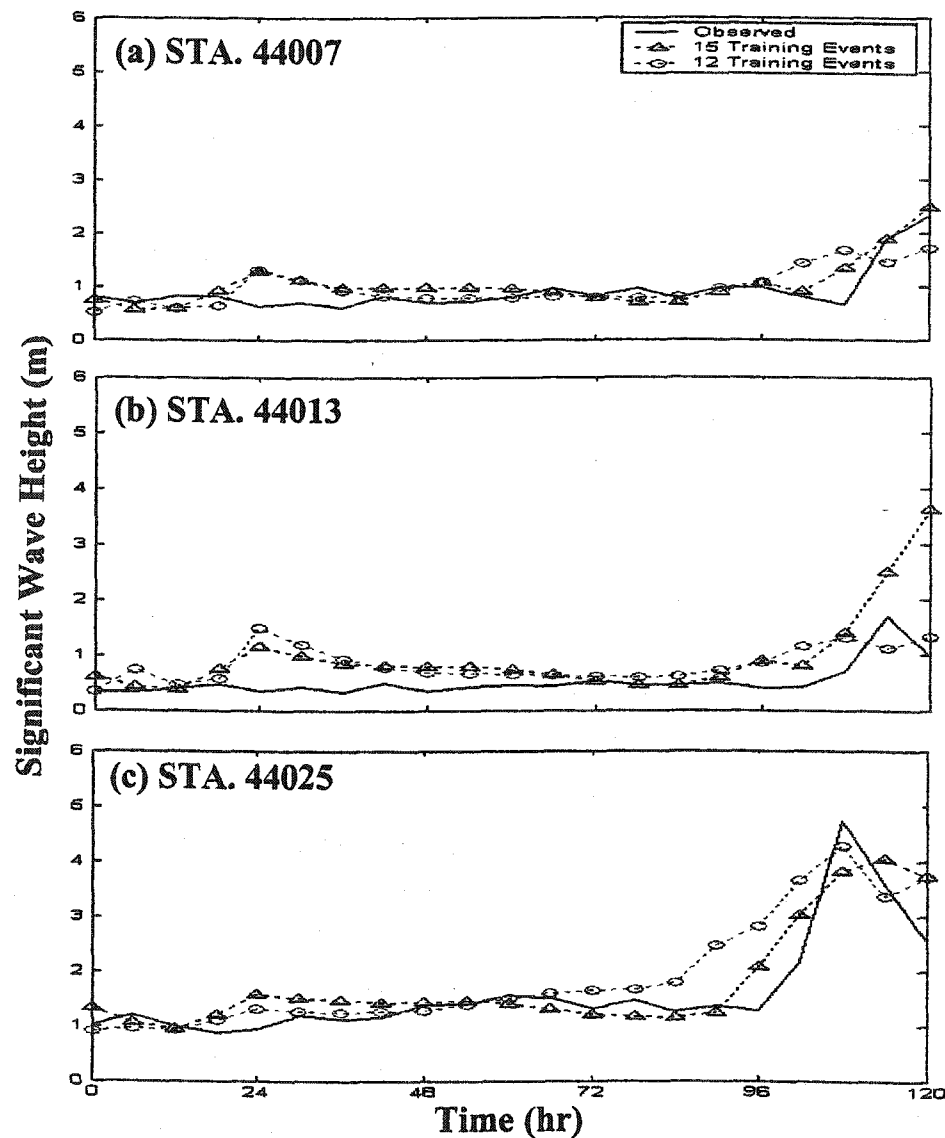


Fig. 6-9. Comparison of the observed and predicted significant wave heights using the Time Delay Neural Network with a structure of  $I_{60}H_9O_5$  for 12 hurricanes (287 data points at 100 iterations) and  $I_{60}H_{10}O_5$  for 15 hurricanes (306 points at 60 iterations). (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. A 24-hour time-delay was used.

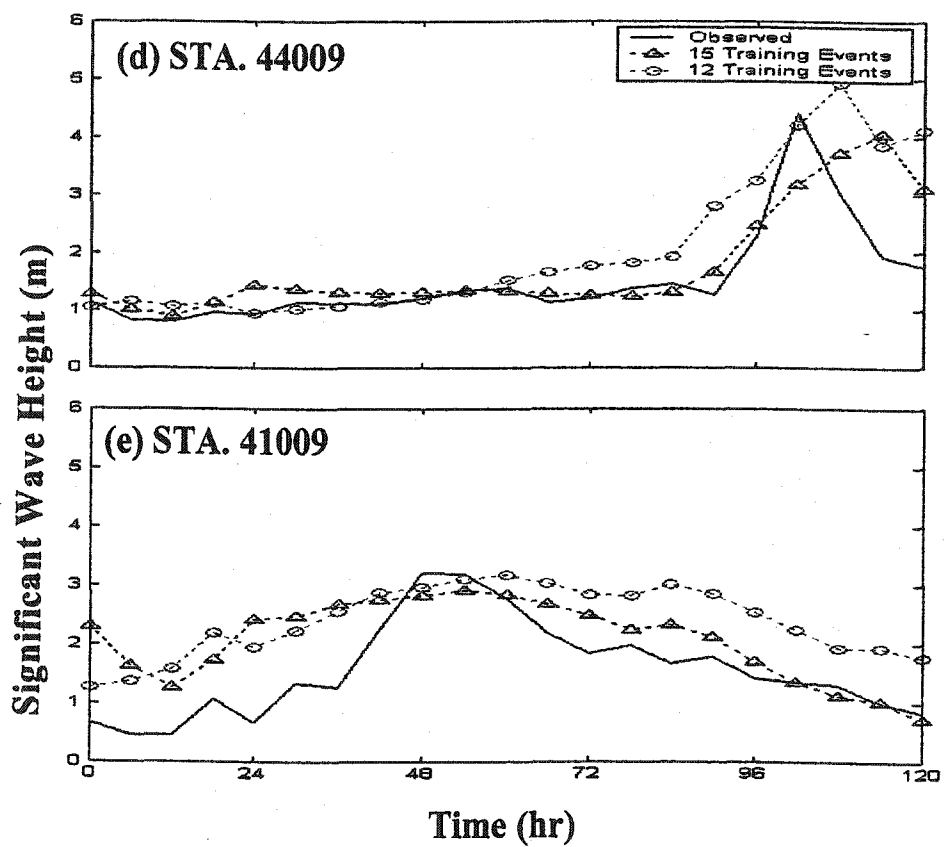


Fig. 6-9. (continued)



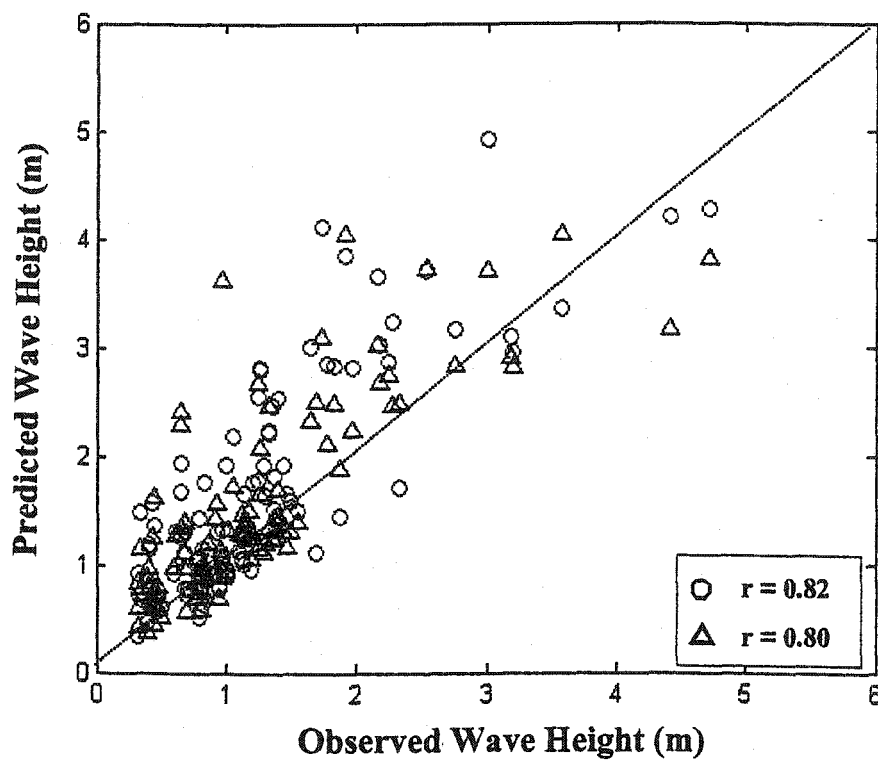


Fig. 6-10. Correlation coefficient between the observed and predicted wave heights. Triangles indicate the results when 15 hurricanes were used as training data, and circles indicate the results with 12 hurricanes.

agree with this basic rule of an ANN model.

This discrepancy can be explained by distinguishing between a simple increase in the number of training data points and possible changes in the similarity of hurricane tracks used in training and validation sets. A quantitative increase of training data does not always guarantee that the increased data include hurricane tracks to those in the validation data set.

In other words, even though the number of training data increases, the qualitative pattern similarity of training data may not increase proportionally or even might decrease. Thus, similarity of hurricane track pattern between training and validation data sets should be considered as an important condition when using a larger data.

Fifteen hurricanes include five hurricanes (Ana and Bob in 1991, Danniell and Earl in 1992, and Emily in 1993) different than those in the 12 hurricanes data set. However, the extra five hurricanes have complete different tracks from Hurricane Bertha in 1999 (Fig. 6-11). Thus, the data domain increased, but not the data similarity. For this reason, the prediction accuracy did not improve in spite of more training data. On the other hand, if all the patterns were included in the training data set, more training data would produce better prediction.

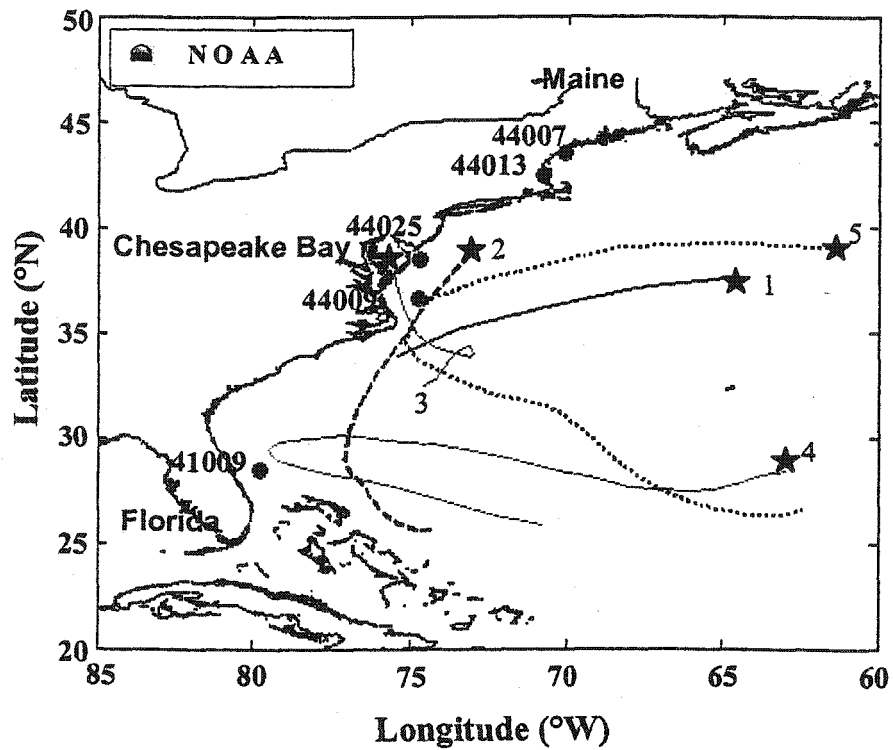


Fig. 6-11. Five hurricane tracks over the northwest Atlantic Ocean from 1991 to 1993. Number one and two are hurricanes Ana and Bob in 1991, number three and four are hurricanes Danniell and Earl in 1992, and number five is hurricane Emily in 1995.

### 6.5.3. Effects of Duration of Time Delay

Group velocity of deep-water waves is determined by wave period (Ippen, 1966b). For hurricanes, because the maximum wind speed, central pressure, and location of hurricane centers change drastically with time, and there is no evidence to assume that the duration of hurricanes is constant, hurricane-wave period might also change more frequently. This changes the group velocity of hurricane waves as well as duration to have a fully developed sea, which can affect wind-waves observed at the five stations. For this reason, it is not possible to set a fixed time delay of hurricane wind waves for an ANN modeling.

The optimum time-delay has not been determined yet in this region. Moreover, identifying the optimum duration is difficult due to the non-stationary characteristic of hurricanes. For this reason, it was decided to use a trial and error method using different time-delays: zero, six, 12, 18, and 24 hours. The results of those time-delays were compared at the same conditions to find the optimum of this study area.

Thirteen hurricanes with 304 data points from 1995 to 2001 were prepared. Among 304 data points, 21 points for Hurricane Bertha in 1999 were used as the validation data set. The other 283 data points were used for training. Thus, the total number of validation data is only 4.2 % of the number of the training data.

The maximum wind speed, central pressure, longitudinal and latitudinal distances of hurricane centers from the five wave stations were used as inputs; the Radius of Maximum Wind (RMW) was excluded. Thus, the Time Delay Neural

Network (TDNN) structure of  $I_{12J}H_nO_5$  was employed, where  $J = 1, 2, 3, 4$  and  $5$  corresponding to zero, six, 12, 18, and 24-hour time delays, respectively.

In order to find the maximum correlation coefficient between the observed and predicted hurricane wave heights, the number of hidden neurons was increased from three to 10 by increments of one, and the number of iterations was changed from 10 to 100 with increments of 10 iterations between each trial. The optimum hidden neurons were 10, seven, eight, and nine for zero and 18, six, 12, and 24-hour time delays, respectively.

Figure 6-12 plots the observed and predicted wave heights at five stations when only zero, 12, and 24-hour time-delays were used because the results of six and 18-hour time-delay were similar to the other predictions. The results indicate that a zero-hour time-delay may have underestimated wave height at station 44025 and 44009 remarkably. The other two time-delays have similar and satisfactory results.

Figure 6-13 shows the correlation coefficient ( $r$ ) between the observed and predicted wave heights when a zero-hour time-delay with 10 hidden neurons and 70 iterations, a 12-hour time-delay with eight hidden neurons and 80 iterations, and a 24-hour time-delay with nine hidden neurons and 100 iterations used. The correlation coefficient 0.82 ( $r^2 = 0.67$ ) was better when a 24-hour time-delay was used, while the TDNN with a 12-hour time-delay shows a lower correlation coefficient, 0.80 ( $r^2 = 0.64$ ). All the results of zero to 24-hour time delay can be seen in Table 6-2.

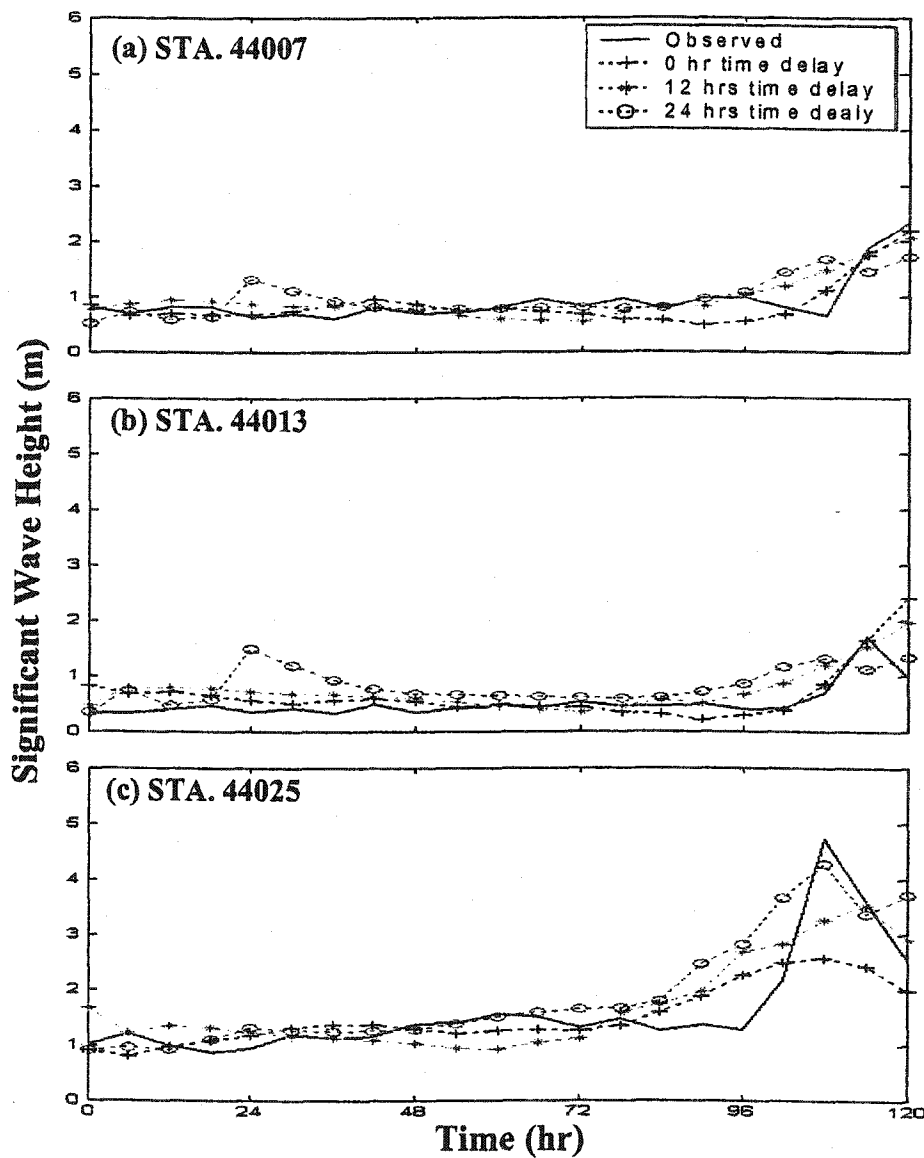


Fig. 6-12. Comparison of the observed and predicted hurricane significant wave heights using the TDNN with a structure of  $I_{12}H_{10}O_5$  for a zero-hour time-delay,  $I_{36}H_8O_5$  for a 12-hour time delay, and  $I_{60}H_9O_5$  for a 24-hour time delay. (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009. The radius of maximum wind speed was not used as input.

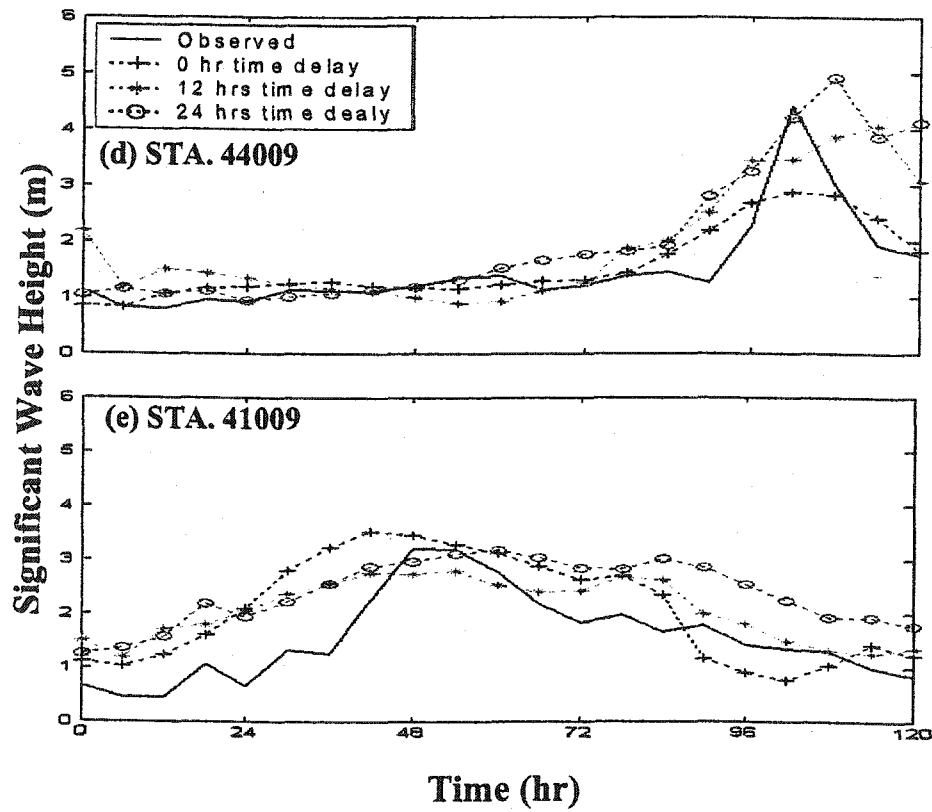


Fig. 6-12. (continued)

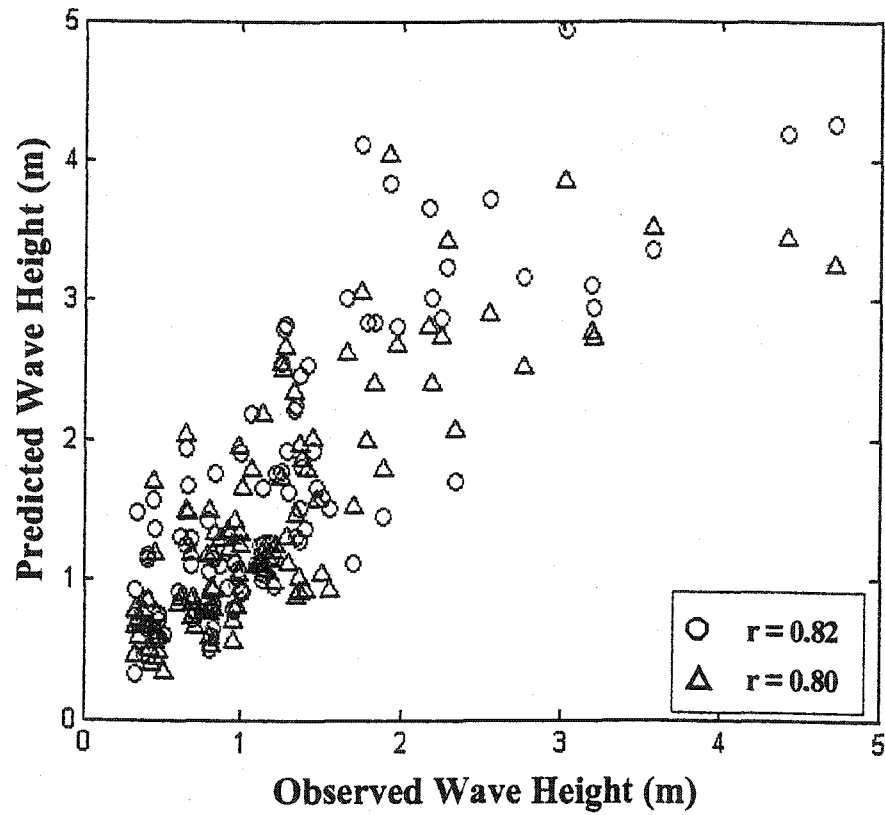


Fig. 6-13. Correlation coefficient between the observed and predicted wave heights. Triangles indicate the results for a 12-hour time-delay, and circles indicate the results for a 24-hour time-delay.



Table 6-2

Maximum correlation coefficients between the observed and predicted significant wave heights, iterations, computing time, and Mean Square Error (MSE) when zero, six, 12, 18, and 24-hour time-delays were used

Time Delay	0-Hour	6-Hour	12-Hour	18-Hour	24-Hour
Iteration	70	70	80	60	100
Time	15.5	15	17	16	21.3
MSE	0.0478	0.0531	0.0483	0.0479	0.0394
Correlation Coefficient	0.78	0.72	0.80	0.80	0.82

The correlation coefficient between the observed and predicted significant wave heights for each station is given in Table 6-3. The maximum correlation coefficient ( $r$ ) was 0.87 ( $r^2 = 0.76$ ) at station 41009, while the lowest was 0.37 ( $r^2 = 0.14$ ) at station 44013. In general, the correlation coefficient was better (greater than 0.80) at southern stations (44025, 44009, and 41009) that have large wave heights than at northern stations (44007 and 44013) with smaller wave heights.

An ANN model should accurately predict if the validation data set has hurricane tracks similar to training data sets, so the similarity between training and validation data sets was examined. Hurricane Bertha occurred from July 9, 06:00 to July 14, 06:00, 1996. The duration is five days and the total data points are 21 with an interval of six hours. It moved from Florida to Maine along the east coast of the U.S. The maximum wind speed varied from 55 kts to 100 kts and central pressures changed from 960 mb to 995 mb.

Hurricane Floyd, which was one of training data sets occurred from September 12 to 17, 1999. The duration was also five days, providing a total of 21 data points. Hurricane Floyd had a track similar to Hurricane Bertha from Florida to Maine (Fig. 6-14). The variation in  $x$  and  $y$  distances between hurricane center and five wave stations was almost the same for Bertha and Floyd (Fig. 6-15).

The maximum wind speed of Floyd varied from 50 kts to 125 kts, and the central pressures decreased from 983 mb to 921 mb before increasing again to 983 mb. Figure 6-16 shows the temporal variation of the maximum wind speed and

Table 6-3

Comparison of the correlation coefficient between the observed and predicted significant wave heights and peak wave periods at the five wave stations

Type	44007	44013	44025	44009	41009
Wave Height	0.51	0.37	0.84	0.81	0.87
Wave Period	0.44	0.45	0.51	0.55	0.37

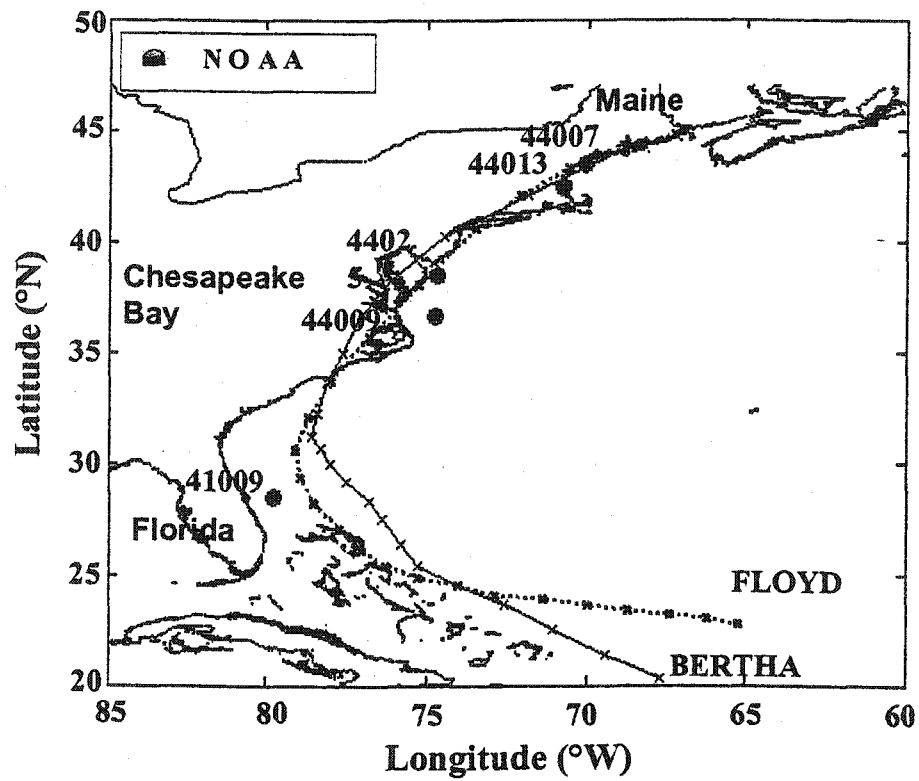


Fig. 6-14. Hurricane tracks for hurricanes Bertha from July 9 to 14, 1996 and Floyd from September 12 to 18, 1999.

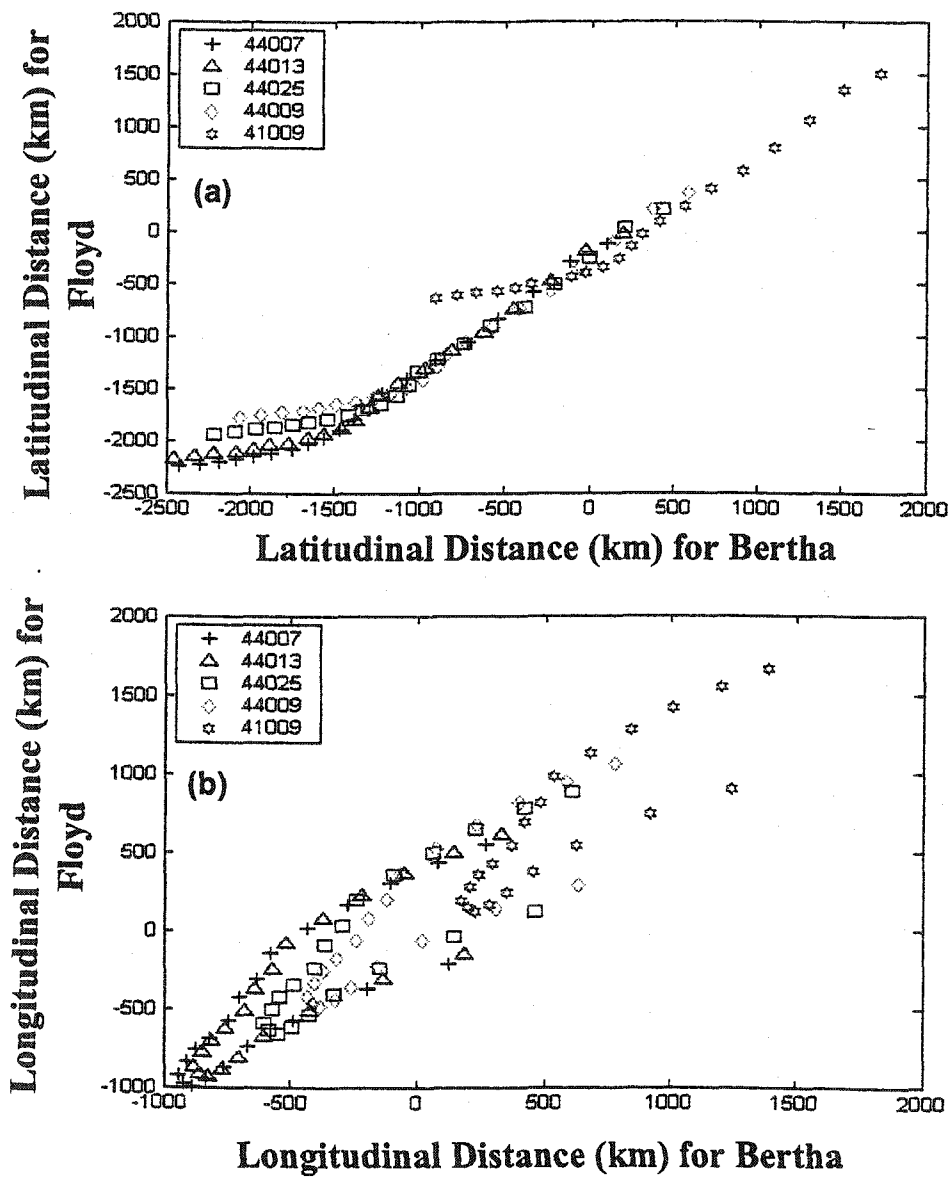


Fig. 6-15. Distances for the five wave stations to hurricane center. (a) For hurricane Bertha in 1996 and (b) for Floyd in 1999.

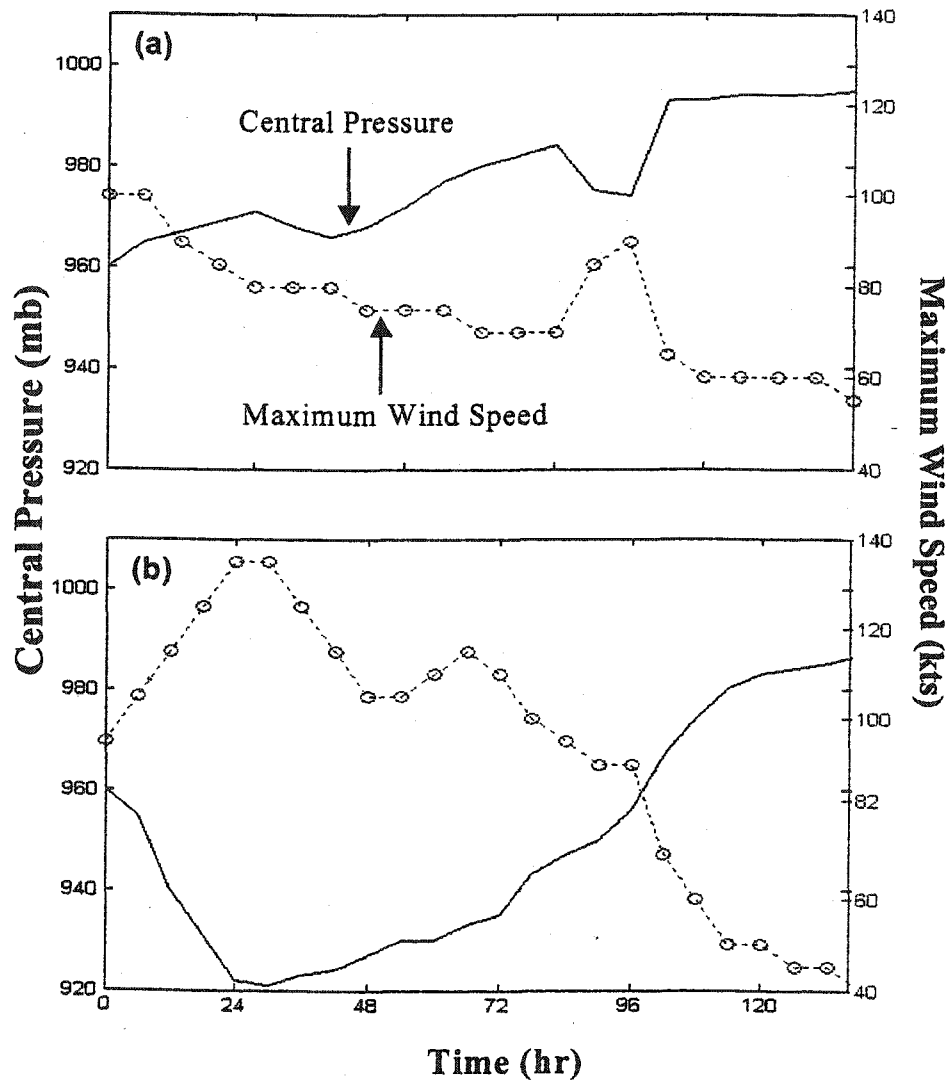


Fig. 6-16. Time series of the maximum wind speed and central pressure. (a) For hurricane Bertha in 1996 and (b) for Floyd in 1999.

central pressures of Floyd and Bertha. Hurricane Floyd has a much higher maximum wind speed and lower central pressure than hurricane Bertha. It can be noticed that the maximum wind speed decreases as the central pressure increases.

The wave records obtained from the National Data Buoy Center indicate that the maximum wave heights for hurricanes Floyd and Bertha were 8.41 m at station 41009 and 4.72 m at station 44025, respectively. Hurricane Floyd had a higher maximum wave height than Hurricane Bertha, however the temporal variation of wave height is similar. The wave height at station 41009 began to rise first, followed in order by stations 44009, 44025, 44013, and 44007 (Fig. 6-17).

Figure 6-18 marked hurricanes tracks to compare the time of the maximum wave height recorded at a wave station and the corresponding location of hurricane centers. Notice that intervals between two consecutive locations of hurricane centers become large at higher latitudes. This indicates that the moving speed of hurricane accelerates as the hurricanes moves toward the north.

Dates and hours in Fig. 6-18 indicate the time when the maximum wave heights were observed at station 41009, 44009, and 44025. The maximum wave height at station 41009 occurred when the location of the hurricane center of Floyd and Bertha moved toward the station. However, at station 44009 and 44025, the maximum wave height occurred after the hurricanes had passed those stations. That is to say, when a hurricane is located in low latitudes, it moves slowly and allows more time for the strong wind forces to generate large waves. In contrast, when a hurricane located in

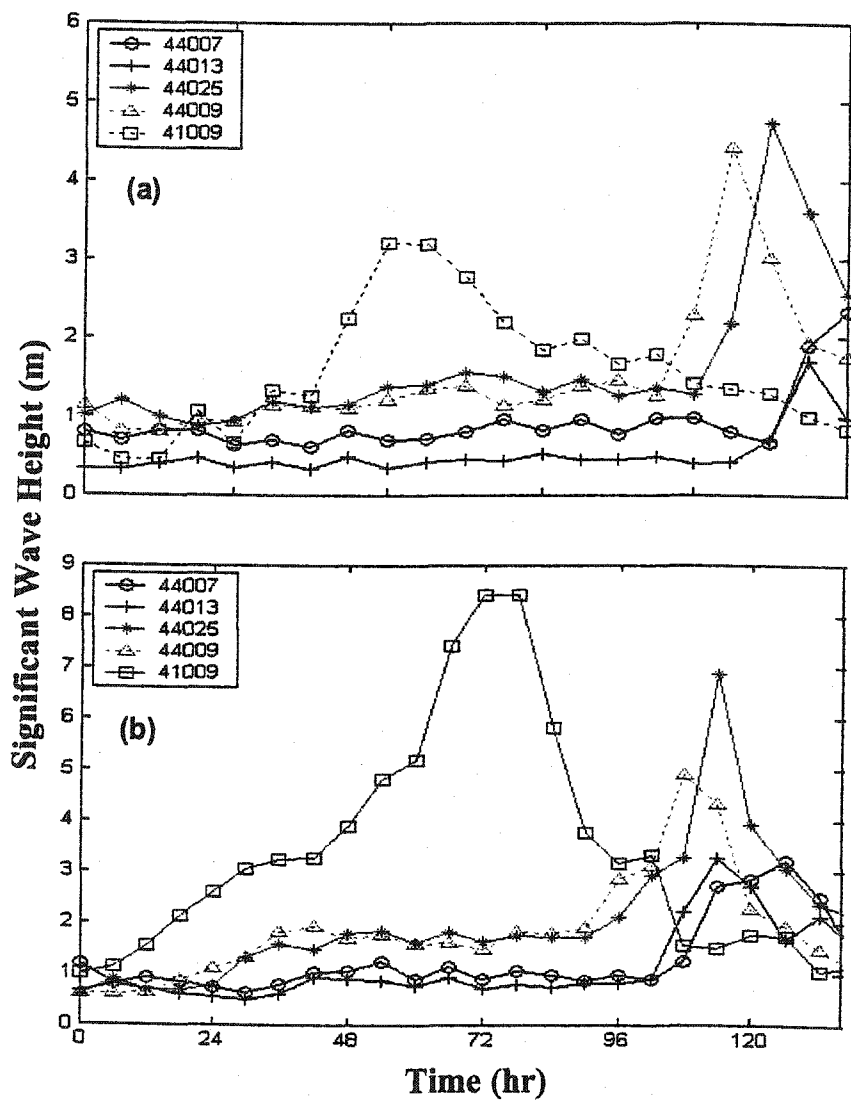


Fig. 6-17. Plots of the observed significant wave heights at the five wave stations caused by (a) hurricane Bertha in 1996 and (b) Floyd in 1999.



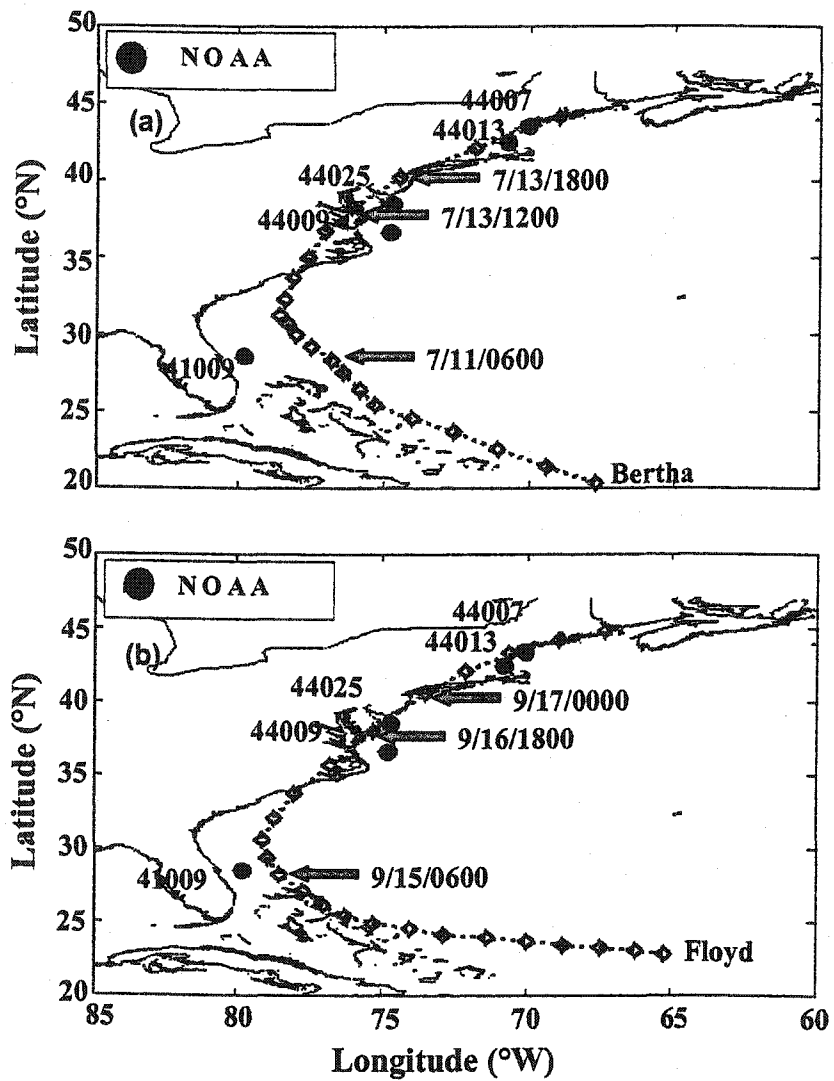


Fig. 6-18. Tracks of (a) hurricane Bertha in 1996 and (b) Floyd in 1999. Dates indicate the time of the maximum wave height at stations 41009, 44009, and 44025.

high latitude areas, it moves more rapidly and there is a shorter time for the weaker wind forces to generate large waves. It was the previously generated large waves that moved to the wave stations to produce maximum wave heights. Because of the slow wave propagation speed, the maximum wave occurred later, after passage of the hurricane center.

Thus, general features of Hurricane Bertha in 1996 and Floyd in 1999 can be summarized as follows: (1) the maximum wind speeds were mainly determined by the differences in central pressures; (2) the stronger the maximum wind speed, the longer the distance a hurricane affects; (3) the stronger the maximum wind speed, the higher the wave height at a given condition; (4) the further north the hurricane moves, the faster its progress.

The TDNN model predicted the significant wave heights accurately except for station 44013 because hurricane Floyd and Bertha were in the same pattern (Pattern 1). When compared with winter-storm waves, the prediction accuracy was relatively small for the hurricane waves. For instance, the maximum correlation coefficient between the observed and predicted winter-storm waves was 0.96 ( $r^2 = 0.92$ ) at station 44013, while the maximum coefficient for hurricane waves was 0.87 ( $r^2 = 0.76$ ) at station 41009.

One of possible reasons is the slight difference in tracks between training (Hurricane Floyd in 1999) and validation data (Hurricane Bertha in 1996). For instance, hurricane Floyd passed closer to Florida than the Bertha, but between 30°N and 40°N, hurricane Floyd moved along the East Coast shoreline, while hurricane

Bertha moved much farther inshore (see Fig. 6-14). Hence, Floyd could affect the five wave stations more strongly than Bertha.

If Hurricane Bertha and Floyd had the same pattern, an ANN could predict Hurricane Floyd-generated wave height well, as it did for SMB-simulated wind-waves that had only one pattern. However, because the pattern of hurricanes for training was different than that for the validation data set, hurricane-waves prediction accuracy was slightly less than storm-waves prediction accuracy.

#### 6.6. Results of Peak Wave Period Prediction and Discussion

Both peak wave period and zero-crossing wave period are important for practical applications. Earlier, an Artificial Neural Network (ANN) model was used to predict the zero-crossing period for winter storms (Chapter 5.8). In spite of a low correlation coefficient between observed and predicted wave periods, the ANN model predicted similar temporal variations compared with that of the observed. This indicates that the ANN model may be used to predict the zero-crossing wave period in this region.

However, there have been no studies on the capabilities of the ANN model to predict peak wave periods in the western Atlantic. For this reason, we present here the prediction results of peak wave periods using the Time Delay Neural Network (TDNN).

Information on peak periods can be obtained from the National Data Buoy Center at the five wave stations. Thirteen hurricanes from 1995 to 2001 with 283 data points were assembled for the training data set, and Hurricane Bertha in 1999 with 21 data points was used for the validation data set. Thus, the ratio of the number of validation data to training data is only 4.2 %.

The maximum wind speed, central pressure, and longitudinal and latitudinal distances of hurricane centers from the five wave stations were used as input factors. A 24-hour time delay ( $J = 5$ ) was used for input conditions, again excluding the Radius of Maximum Wind speed (RMW). Hence, the TDNN structure of  $I_{60}H_nO_5$  was employed.

In order to determine the correlation coefficient between observed and predicted hurricane wave periods, the number of hidden neurons was changed from three to 10 in increments of one. The number of iterations was increased from 10 in increments of 10 iterations between each trial until the correlation coefficient again declined. Because a 24-hour time-delay showed the best prediction results from the previous hurricane wave height prediction, a one-day time delay was used ( $J = 5$ ).

Figure 6-19 plots the observed and predicted peak wave periods at five stations (44007, 44013, 44025, 44009, and 41009). Except in the initial 24 hours, predicted peak wave period was almost a constant, about 10 seconds. Figure 6-20 shows the correlation coefficient ( $r$ ) between the observed and predicted peak wave periods. The best correlation coefficient was 0.50 ( $r^2 = 0.25$ ), and the Mean Square Error

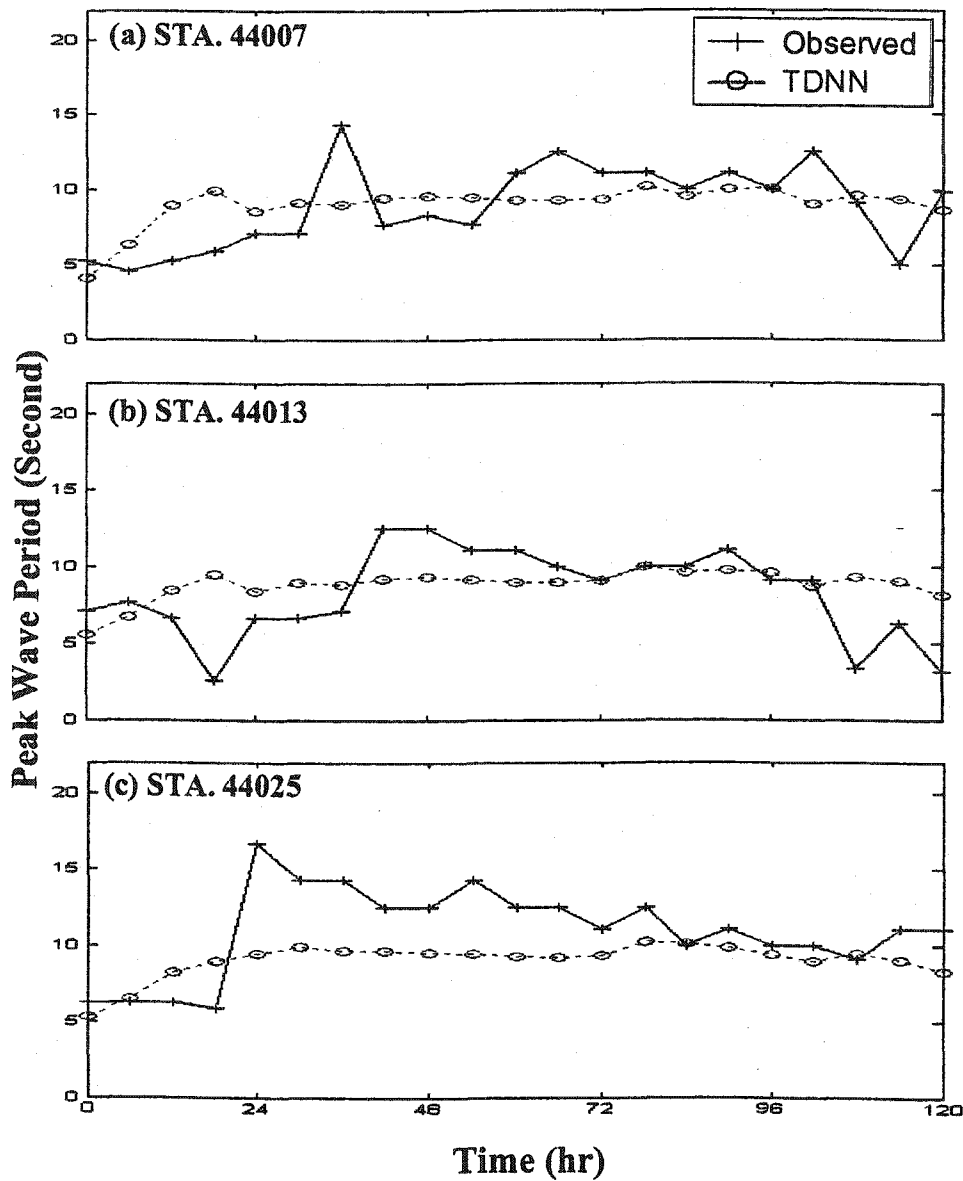


Fig. 6-19. Comparison of the observed and predicted peak wave periods using the Time Delay Neural Network with a structure of  $I_{60}H_4O_5$  at 110 iterations. (a) For station 44007, (b) station 44013, (c) station 44025, (d) station 44009, and (e) station 41009.

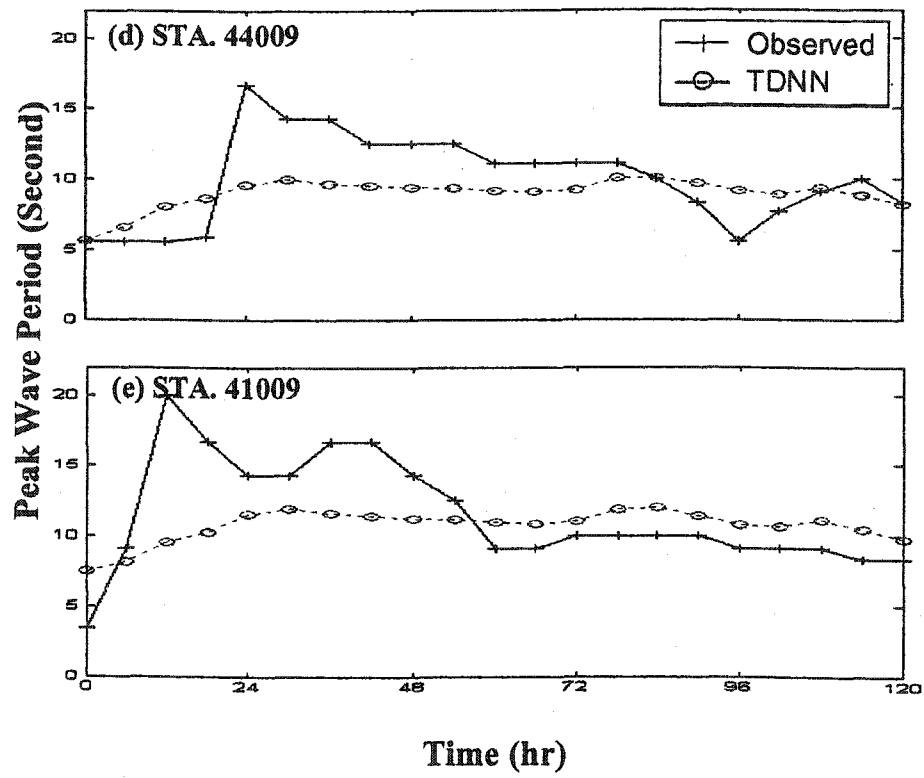


Fig. 6-19. (continued)

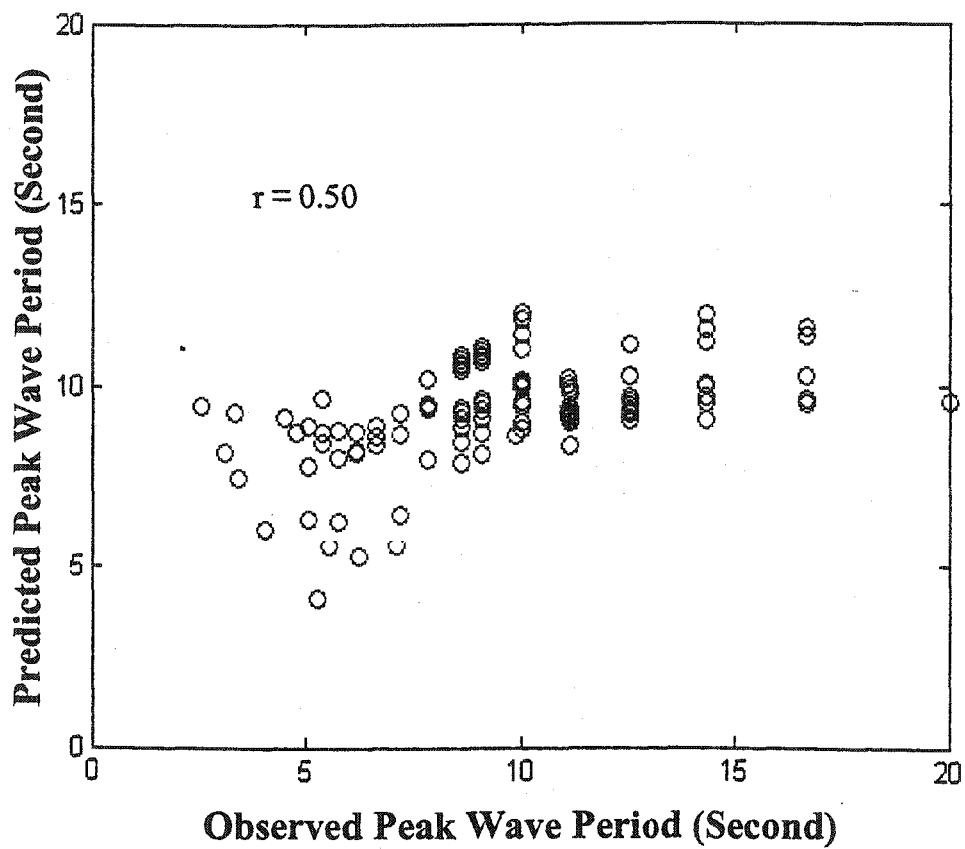


Fig. 6-20. Correlation coefficient between the observed and predicted hurricane peak wave periods.

(MSE) was 0.1184 when four hidden neurons and 110 iterations were used. For more information about the correlation coefficient at five stations, see Table 6-3.

The best correlation coefficient of 0.50 is slightly less than that of 0.58 for winter-storm zero-crossing wave period prediction. The TDNN model did not accurately predict wave periods for hurricane Bertha in 1996. But it is difficult to decide which result is better because the TDNN model was used to predict peak wave period in this study of hurricanes, instead of the zero-crossing wave period. The prediction of zero-crossing wave periods for hurricanes should be addressed in further studies.

Hurricane Bertha (the validation hurricane) produced a maximum wave period of about 20 seconds at station 41009 and 15 seconds at the other stations at the beginning of the event. But the predicted wave period was almost constant with time as about 12 seconds at station 41009 and 10 seconds at the other four stations. The major reason can be revealed from Fig. 6-21. This figure presents the variations of wave period with time for Hurricane Bertha and Floyd (the training data). The peak wave period of Hurricane Floyd did not change as much, and varied only between 10 to 15 seconds at the five wave stations. From this fact, it is apparent that the TDNN model cannot predict the wave periods for hurricanes Bertha because the patterns of wave periods are different between Hurricane Bertha and Floyd.



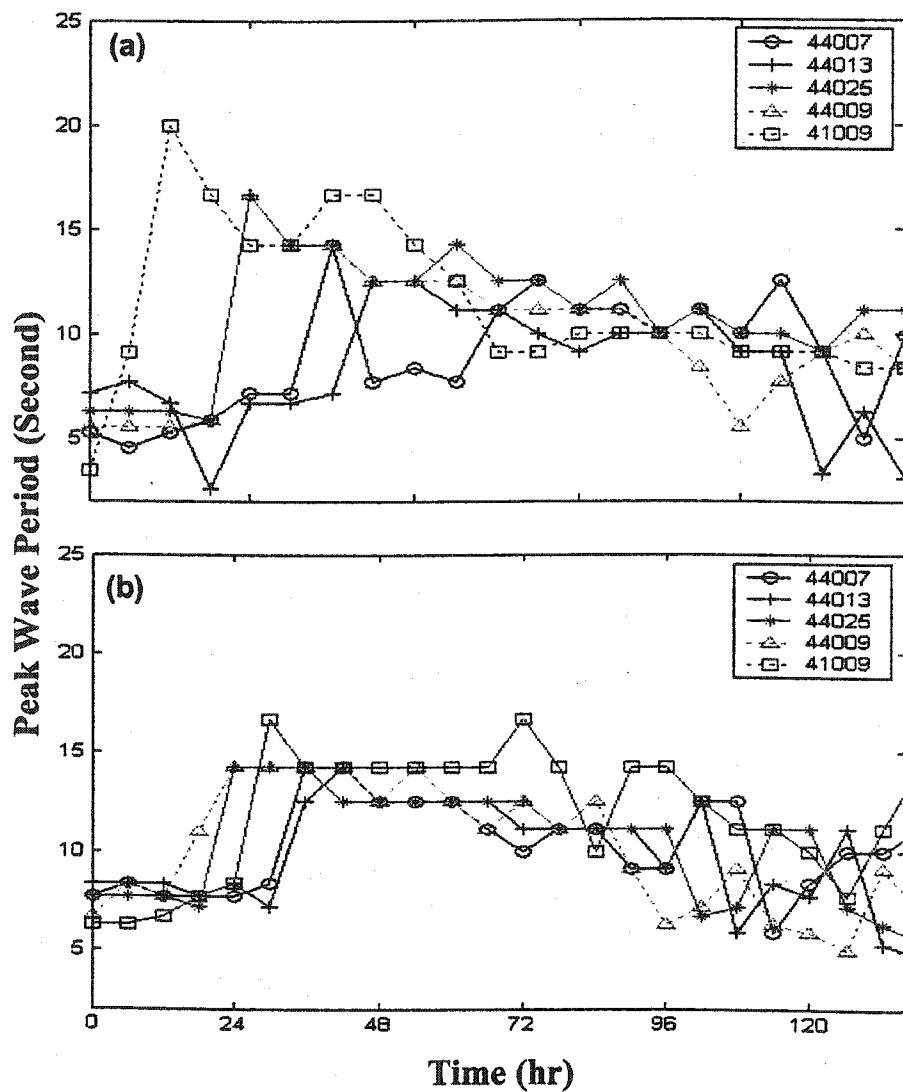


Fig. 6-21. Time series of the observed peak wave periods at five wave stations (a) for hurricane Bertha in 1996 and (b) for Floyd in 1999.

### 6.7. Conclusions

The TDNN wave model produced reasonably good predictions of hurricane-generated wave height when the maximum wind speed, central pressure, and location of hurricane centers were used as input. The required computing time was less than a few minutes for both hurricane wave height and period predictions. The best time-delay was 24-hours. But the prediction results for hurricane-generated peak wave periods were still not satisfactory. Increasing the number of training data points did not greatly improve prediction accuracies because of difference in patterns between training and validation data sets.

The prediction accuracy of hurricane waves at each station was much less than that of winter-storm waves, and this disparity in forecast accuracy due to the major differences between hurricanes and storms. Slight differences in hurricane patterns between training (hurricane Floyd) and validation data (hurricane Bertha) may reduce accuracy in hurricane wave prediction.

This study shows the feasibility of using ANN techniques for hurricane-wave prediction model. If more hurricane data becomes available, and prediction of hurricane tracks is possible, we can forecast real-time hurricane waves more accurately with less computing time.

## CHAPTER VII

### DISCUSSION AND CONCLUSIONS

The following subjects are discussed in this chapter: (1) convergence rate including most affecting factors for accurate prediction (*e.g.*, simulation of a time-delay, similarity between training and validation data, and pre-processing); (2) the quantity of data needed for training; (3) the reason for an excellent simulation of the SMB model; (4) possible improvement of wave period prediction; (5) possible extension of ANN prediction results to areas with no data; (6) possible future improvements on wave modeling, and (7) application of ANN wind-wave prediction model to other places.

#### 7.1. Convergence Rate

In this study, we have seen that ANN model convergence rate and prediction accuracy changed according to the size of training data set, the similarity in patterns between training and validation data, and the complexity of a physical process. The causes and effects of convergent rates are summarized and discussed in the next section.

### Which Learning Algorithm Is Better?

In this study, the two most updated learning algorithms: the Scaled Gradient Descent (SCG) and the Gradient Descent with a Variable Learning Rate and Momentum (GDX) show little difference in computing time. However, in terms of the mean square error, the SCG was better than the GDX. This implies that the SCG approach has a better algorithm (*i.e.*, using the maximum of error gradient,  $\partial E/\partial W = 0$ , to find the optimum solution) to update the weights for finding the least square error.

### What Is the Best Way for Finding the Optimum Number of Hidden Neurons and Iterations?

Learning curves, error gradient curves, bias, and variance were available for the Back-Propagation Network (BPN) and Elman Recurrent Network (ERN) for finding the optimum number of hidden neurons and iterations. For the Time Delay Neural Network (TDNN), there was no clear index to find the optimum number of hidden neurons and iterations except the Mean Square Error (MSE) and the correlation coefficient ( $r$ ) between observed and predicted values. Thus, these two parameters (MSE and  $r$ ) were used. For more information about these indices, see Chapter 2.7.4 and 2.8.3.

Using MSE and  $r$  between the observed and predicted values, the optimum number of hidden neurons and iterations can be determined for a given wind-wave events (data set), however, the optimum does not necessary always good for other

events. There is one possible way to reduce uncertainty in the check of the optimum number of hidden neurons and iterations by swapping the training and validation data if this data set is sufficiently large. Theoretically, the optimum number of hidden neurons and iterations should be different if the training and validation data have a different pattern. In this case, we might try to use the average of the two cases and check if a better prediction result can be achieved. So far, this is just a possible approach. Further study to verify this possibility is pending for more studies.

#### Which ANN Algorithm Is the Best?

The prediction accuracy was best with the TDNN with long-term memory, progressively poorer for the Elman Recurrent Network (ERN) with short-term memory and the Back-Propagation Network (BPN) with no-memory. This implies an external time-delay is necessary for using ANN to do wind-wave predictions. However, because the prediction accuracy among the three models were not compared using the same conditions, a slight change of this conclusion is possible. In terms of computing time, the difference in computing speed was negligible small among the three ANN wind-wave prediction algorithms.

#### What Is the Difference in ANN Inputs for Winter-Storm and Hurricane Wave Prediction?

For winter storm-wave predictions, wind components (u and v) specified at the selected 40-wind stations must be used to include the effects of a large wind field on

wave generation. For hurricane-wave prediction, the maximum wind speed, pressure of hurricane centers, x- and y-directional distances from the hurricane centers to these five wave stations were ANN inputs for simulating the characteristics of hurricanes: have strong wind within a relatively small domain (*i.e.*, about 100 km), move quickly, change dramatically with time and location and thus, affect only the nearest small area.

#### Simulation of Time-Delay

The importance of considering time-delay for an ANN wind-wave prediction model was already explained with examples based on the Sverdrup-Munk-Bretschneider (SMB) method given in Chapter 5.5. Depending on wind speed, the duration time required for the waves to be fully developed seas differs.

For winter-storm wave prediction, we used one representative time-delay (24 hours) because wind fields are relatively evenly distributed and stationary. That is to say, the duration for a corresponding wind speed is possible. When a 24-hour time delay was used, the prediction of storm-wave height was fairly accurate ( $r = 0.88$ ), but the wave period prediction was much less ( $r = 0.61$ ).

However, hurricanes are different. For instance, the wind field for a hurricane is small, the wind speed within the domain is high, and a hurricane usually moves very fast. Because hurricanes are more dynamic, the duration for a fully developed sea also changes drastically with time. For this reason, setting a fixed time-delay for an ANN hurricane-wave prediction model was not possible. In this study, five

different time-delays (*i.e.*, zero, six, 12, 18, and 24-hour) were tested. When a 24-hour time-delay was used, the TDNN showed the best prediction accuracy ( $r = 0.82$ ) for hurricane wave height, while the prediction accuracy for wave periods was not satisfactory ( $r = 0.50$ ). This might be a coincidence. The correlation coefficient of wave height was 0.80 when a 12-hour time-delay was used. This indicates that there is no much difference in prediction accuracy between a 12 and 24-hour time-delays.

#### Pre-Processing Is Needed?

When pre-processing was used, the model efficiency of the BPN and ERN has a significant improvement, *i.e.*, the computing time is less and prediction accuracy is better. Because the TDNN has a different learning algorithm, such as an external time-delay, the pre-processing efficiency was not compared with the other two models. In general, the wind input was pre-processed using a global maximum and minimum of  $\pm 20$  m/s, which normalized the input wind velocity between  $-1$  and  $1$ .

However, the ANN performances were sensitive to the global maximum and minimum wind speeds. When using  $\pm 80$  m/s as the global winds, the correlation coefficient between the observed and predicted wave heights increased from 0.85 to 0.88. The predicted peak-wave height also improved at station 44013 and 44025, and station 44007 no longer showed negative wave heights.

When further examining the ANN modeling, it is noticed that the non-linear transfer function might play a role in carrying this result. For the selected non-linear transfer function  $(1/1+\exp(-WX))$ , where  $X$  is the normalized wind speed and  $W$  is the

weight matrix), the change of output is flattened even when the input has a significant change. For example, for two input  $WX = 2$  and  $4$  (a change of one fold), the output will be  $0.88$  and  $0.99$  (a change of only  $14\%$ ), which is not reflecting the big change in input.

When  $\pm 20$  m/s was used, the change of wind speed from  $15$  m/s to  $20$  m/s only causes a small change in the output. In contrast, if  $\pm 80$  m/s of global winds was used, the normalized wind input will be much less, but the slope of response function is much high, so the ANN may still be able to produce a noticeable change in the output. Hence, the ANN could produce a higher wave height corresponding to a strong wind.

#### Why Station 41009 Has Much Less Prediction Accuracy

When comparing the correlation coefficient between the observed and predicted wave height, station 41009 had much less prediction accuracy than the other four wave stations located at northern areas. As previously explained in chapter 5.7.1 and 5.7.3, insufficient training data points, complicated wind-wave patterns, and another wind field system that was not considered in this study are the possible reasons.

At the beginning of this study, winter-storm waves at the selected five stations were assumed to be affected by the strong winds from the northwest part of the Atlantic Ocean. For this reason, not many wind stations on the southeast side of station 41009 were considered for wind stations. However, from this study, it is clear that observed waves at station 41009 are not generated by winter storm wind from



northwest of the Atlantic Ocean blowing toward the west, but they were generated from wind fields between  $0^{\circ}\text{N}$  and  $20^{\circ}\text{N}$ . For this reason, more wind stations off southern Florida should be used for a better prediction at station 41009.

One may argue that the effects of the Gulf Stream, which flows from the south to the north, may affect the wind-waves generation pattern, so the ANN prediction is not good. However, those effects were already represented as the observed waves. In other words, because an ANN model was trained with wind-waves data affected by the Gulf Stream, the ANN should include those effects already, so, this effect might not be the reason of a low accuracy.

#### Why the Prediction Accuracy Is Lower for Hurricane Waves

The prediction accuracy for hurricane waves (with a correlation coefficient  $r = 0.82$  and  $0.58$  for wave height and wave period) was less than that for winter-storm waves ( $r = 0.88$  and  $0.61$ , respectively). The difference may be caused by the reason of lack of sufficient training data for hurricanes. Considering the dramatical difference among hurricanes, it is difficult to have two closely similar hurricanes. For example, we only used 15 hurricanes during 13 years, but none of these hurricanes were similar enough to provide a clear pattern for ANN to remember and use. This problem may be overcome when there are more hurricanes, or when a well-established numerical model based on physical processes can be used to provide more data for training.

## 7.2. How Many Data Are Needed for Training?

In nature, many different types of winter-storm wave generation patterns can be expected, and it is clear that the prediction accuracy of winter-storm wave height increases as the number of training data is increased (*e.g.*, the correlation coefficient,  $r$ , improved from 0.82 to 0.88 when training data increases from 218 to 403 points).

Nonetheless, one must draw a line on how many data for training is sufficient. The only possible way to determine the number of sufficient training data points is to select a threshold correlation coefficient between the observed and predicted waves. Tentatively, if the correlation coefficient exceeds 0.90, the number of training data points should be satisfactory. Based on this statement, a little more data set for training is needed for both winter storm and hurricane wave predictions.

For winter-storm waves, because the difference in wind-wave patterns is relatively small among each event, one more year's data should be enough.

In contrast, for hurricane waves, the difference in pattern between each event is very different. Therefore, the prediction accuracy of hurricane wave height did not increase with an increase in the number of training data (*e.g.*, from 12 hurricanes to 15 hurricanes) because the pattern of those three additional data sets were different from those of the validation data set. For this reason, to find sufficient patterns is more important than to find more events. A similar pattern between training and validation data sets is the key to increase the accuracy of predictions. If only measurements are allowed to use, one may expect that data from many more years are needed to meet the  $r = 0.9$  criterion.

### 7.3. Reason for an Excellent Simulation of the SMB Model

When using non-linear wind-wave data sets simulated from the Sverdrup-Munk-Bretschneider (SMB) method, the TDNN model predicted wave height and period accurately, even though it used only four of eight data sets as training data, which did not cover the validation data set at all. In this test, we assumed only one wind-field system, which has one wind direction but different wind speed and fetch to the imagined five wave stations. That is to say, only one wind-wave pattern was considered.

For this reason, the TDNN easily understood the wind-wave pattern and was able to recognize untrained events. That is to say, although the validation events were not used for training, the ANN predicted wave height and period accurately because it already knew the non-linear relationship.

### 7.4. Possible Improvements on Wave Period Prediction

The prediction accuracy of wave period was much poorer than wave height for both winter storms and hurricanes because of complicated physical processes in wave period generation. There may be two possible ways to improve ANN performances for wave period prediction: (1) by using a large number of hidden layers and (2) by implementing an ANN technique that can handle different time-delays for the multiple wind-wave patterns in nature.

For the first case, we have already tried the TDNN with two hidden layers. When a four-layered TDNN was used, the correlation coefficient between the observed and predicted wave periods were  $r = 0.61$ , which was a little better than 0.58 for a three-layered TDNN. However, because the improvement was so small, a test with more hidden layers was not attempted.

As previously mentioned in Chapter 5.5, considering time-delay is important because of the time needed for waves to be fully developed seas. Currently, to simulate multiple time-delays for all possible wind speeds, however, is not possible. If this problem can be resolved in the future, wave period prediction should be able to be improved.

#### 7.5. Possible Extension of ANN Prediction Results to Areas with No Data

The major condition for being able to use an ANN wind-wave prediction model is that sufficient data are available for training. If that condition is not met, an ANN technique is not helpful. For this reason, I tried to introduce two methods for areas with no or insufficient data: (1) ANN models can be used in combination with other type of numerical wind-wave prediction models, a so-called 'hybrid model'. For example, the SWAN or WAM can produce wave heights and periods using wind speed and direction, then after collecting the wind and model generated wave data set, an ANN model can be trained and used to predict future events; (2) linear interpolation: assuming the difference in spatial domain between the nearest wave

stations can be obtained by linear interpolation, then ANN model is still useful. For instance, a wave station in Rhode Island is located between stations 44013 and 44025. Let's suppose a 4 m-wave height for station 44025 and a 10 m-wave height for station 44013, and Rhode Island is located in the center of the two stations. If the linear interpolation of 7 m wave height at the station in Rhode Island is applicable, then the ANN is still a fast technique to estimate wave condition at the Rhode Island station.

#### 7.6. Possible Future Improvements on ANN Wave Modeling

In the future, the ANN will be more valuable for real-time wind-waves prediction because of more available data. The prediction accuracy will continue to improve as the efforts to measure wind waves continue.

An ANN model that can handle several natural wind-wave systems together is needed. Currently, fixed wind stations are provided by users, and not changeable during the training and prediction. It would be better if an ANN model can be used to identify wind stations and group them as an effective wind fetch during the training and validation. If this development happens, the ANN technique will be much more powerful and valuable to predict winter-storm and hurricane waves.

As previously mentioned in Chapter 5.5, a hybrid model that combines an ANN and other physical processes based numerical models shall be used together to supplement each other.

An ANN model can be used to improve the efficiency of currently used numerical models that are based on simulation of physical processes. For instance, an ANN technique may substitute for one part of numerical processes as a so-called embedded model. For the numerical wind-wave models such as the SWAN and WAM, simulating the wave-wave interaction is very complicated and requires much computing time. Although they use discrete interaction approximation but still require significant computing time. That is where an ANN approach can be used for the recognition of wave-wave interactions pattern and save much computing time.

#### 7.7. Application of ANN Wind-Wave Prediction Model to Other Places

An ANN wind-wave prediction model can be relatively easily applied to other places. However, because each area may have different physical environments for wind-wave generations (e.g., different fetch), different considerations and even different ANN structures should be considered.

Suppose we need to forecast wind-generated wave heights at Honolulu, Hawaii, then we need to know that winds blowing near New Zealand may generate large waves at Honolulu after a few days because the physical distance between the two places is over 5000 km. Thus, a much long time-delay and a much large areas are needed.

In contrast, suppose we need to predict wind-generated waves at North Chesapeake Bay area in Maryland. The average length and width of the Chesapeake

Bay is about 400 km and 30 km respectively, much less than the domain of the northwest Atlantic Ocean. The northern part of the Chesapeake Bay is closed and only the southern area is connected with the Atlantic Ocean. The effects of winter storms can be negligible at the Baltimore harbor because the winter-storm winds usually come from the north and thus, the wind fetch is very limited. However, hurricanes can affect many cities on the coastline along the Chesapeake Bay because strong winds might come from the south and generate large waves because of 400 km fetch. The duration of a time-delay must be less than that given in this study.

For harbors in the Hampton, we also must consider the effects of the Chesapeake Bay mouth. The open Bay mouth can allow the transfer of waves generated in the southern part of the northwest Atlantic Ocean into the Chesapeake Bay. For this reason, this effect must also be considered for this particular application.

### 7.8. Conclusions

From this study, we reached the following conclusions about ANN wave prediction model for winter storms and hurricanes:

- (1) The Scaled Conjugate Gradient learning algorithm is better than the Gradient Descent with a Variable Learning Rate and Momentum.
- (2) The performance of Time Delay Neural Network is better than the Back-Propagation Network and Elman Recurrent Network model.

(3) In order to select the optimum number of hidden neurons and iterations, the trial and error method is better than the Mean Square Error (MSE) or Root Mean Square Error (RMSE).

(4) U and v wind components are the necessary wind inputs for ANN prediction model. Wind speed and direction cannot be used because of the ambiguity.

(5) The maximum wind speed, longitudinal and latitudinal distance between hurricane centers and wave stations, central pressure of hurricanes can be used as ANN inputs for hurricane-generated wave prediction because of the relatively small wind field and fast moving characteristics.

(6) Pre-processing is necessary to improve model efficiency and performance. Especially, the sensitivity of ANN prediction accuracies to the global wind speed has to be checked.

(7) An increase in training data set size improves the prediction accuracy for winter-storm waves. A reference correlation coefficient (*i.e.*,  $r_{\min} = 0.9$ ) between observed and predicted waves can be used for determining the optimum number of training data points.

(8) Station 41009 is not affected by winter-storms but by a different wind field. For accurate prediction, the wind information at areas further south of Florida must be considered.



(9) The prediction accuracy of wave heights is better than that of wave periods for both winter storms and hurricanes. For more accurate wave period prediction, time-delays for different wind-wave patterns should be included in ANN modeling.

(10) Hurricane waves prediction is more difficult than winter storm waves prediction because of the lack of hurricane wave data.

(11) The prediction accuracy depends on the similarity in hurricanes between training and validation data.

(12) The ANN prediction result can be extended to areas with no data if a proper interpolation technique can be used.

(13) A hybrid model can be used to help for overcoming some of the drawbacks of ANN modeling. A hybrid modeling using the strength of both modeling technique should be the next effort.

## APPENDIX

## I. Source Code for the SMB Method

## 1. SMB Method for Wave Predictions at Deep Water, Open Ocean

```

%
% 1 nautical mile = 6080 ft = 1853.658 m
% 1 knots = 1 nautical mile/hr = 0.5149 m/s
%
infile = 'c:\waveprediction\inp';
[fid, message] = fopen(infile);
if fid == -1
message
end
%
% Read data
%
title = fgetl(fid);
disp(title);disp(fgetl(fid));
disp(fgetl(fid));
a = fscanf(fid, '%d %d %d', [3,inf]);
fclose(fid);
fetch_in = a(3,:); % in meter
u_in = a(2,:); % in m/s
time_given_in = 3600*a(1,:); % in seconds
ncase = length(u_in);
outfile = 'e:\smbmodel\smb_test.out';
[fid, message] = fopen(outfile,'w');
if fid == -1
message
end
%
fprintf(fid,'%s\n', title);
fprintf(fid,' Time Wind Speed Fetch Dur_min;_act Hs_f T_f Hs T\n');
fprintf(fid,' hr knots (m/s) NM (km) hr hr m s m s\n');
fetch = fetch_in(1);
u = u_in(1);
duration = time_given_in(1);

```

```

time_elp = time_given_in(1)/3600;
cont = 0;
for ic = 1:ncase-1
disp(['elapsed time=', num2str(time_elp)]);
f_nm = fetch/1853.658;
f_km = fetch/1000;
u_kt = u/0.5149;
dur_a = duration/3600.0;
min_dur = minimum_duration(fetch, u);
min_a = min_dur/3600.0;
[Hs, Period] = HandT(fetch, u);
%
% Find the equivalent wind fetch if the given duration is less than the required
% for a fully developed sea
%
if duration < min_dur
fetch_try=fetch - 100; % reduce the fetch by 100 m
time_try = min_dur;
while time_try > duration
time_try=minimum_duration(fetch_try, u);
fetch_try=fetch_try - 100;
end
%
% Duration limited wind wave
%
[Hs_1, Period_1]= HandT(fetch_try, u);
else
Hs_1=Hs;
Period_1=Period;
end
%
% Save the results
%
format='%5.1f %4.1f%4.1f %6.1f%6.1f %5.1f%5.1f%6.2f%6.2f %6.2f'
fprintf(fid, format, time_elp, u_kt, u, f_nm, f_km, min_a, dur_a, Hs, Period, Hs_1,
Period_1);
%
% Check if it is needed to calculate wave energy and shaft to a different wind speed
% for the same wind velocity, move to the next step
%
if u_in(ic+1) == u_in(ic)
fetch=fetch_in(ic+1);
u=u_in(ic+1);
duration=duration + time_given_in(ic+1) - time_given_in(ic);

```

```

time_elp=time_given_in(ic+1)/3600.0;
else
%
% For a different wind speed, move to the middle of the two time step, upgrade
% wave height and period first, then find the energy and then find the corresponding
% wind speed, fetch, and duration
%
fetch = fetch_in(ic);
u = u_in(ic);
duration = duration + 0.5*(time_given_in(ic+1) - time_given_in(ic) );
time_elp = time_elp + 0.5*(time_given_in(ic+1) - time_given_in(ic) )/3600.0;
min_dur = minimum_duration(fetch, u);
[Hs, Period] = HandT(fetch, u);
%
% Find the equivalent wind fetch if the given duration is less than the required
% for a fully developed sea
%
if duration < min_dur
fetch_try = fetch - 100; % reduce the fetch by 100 m
time_try = min_dur;
while time_try > duration
time_try = minimum_duration(fetch_try, u);
fetch_try = fetch_try - 100;
end
%
% Duration limited wind wave
%
[Hs_1, Period_1] = HandT(fetch_try, u);
dur_new = time_try/3600.0;
else
Hs_1 = Hs;
Period_1 = Period;
dur_new = duration/3600.0;
end
f_nm = fetch/1853.658;
f_km = fetch/1000;
u_kt = u/0.5149;
%
% fprintf(fid, format, time_elp, u_kt, u, f_nm, f_km, min_a, dur_new, Hs,Period,
% Hs_1, Period_1);
energy = Hs_1*Hs_1*Period_1*Period_1;
%
% Find out the corresponding wind fetch and duration for the new given wind speed
%

```

```

u = u_in(ic+1);
fetch = 0;
energ_new = 0;
No_of_itr = 0;
while energ_new < energ
%
f energ_new > energ;
break
end
if No_of_itr > 180000;
break
end
%
fetch = fetch+100;
[Hn, Period_n] = HandT(fetch, u);
energ_new = Hn*Hn*Period_n*Period_n;
end
duration = minimum_duration(fetch, u);
dur_new = duration/3600.0;
f_nm = fetch/1853.658;
f_km = fetch/1000;
u_kt = u/0.5149;
%
% Print the equilibrium wave condition at the middle of two consecutive elapse time
% fprintf(fid, format, time_elp, u_kt, u, f_nm, f_km, dur_new, dur_new, Hs,Period,
% Hn,Period_n);
% find the equilibrium fetch at the end of the second wind velocity
%
time_elp = time_elp + 0.5*(time_given_in(ic+1) - time_given_in(ic))/3600.0;
fetch_try = fetch + 100; % increasee the fetch by 100 m
time_try = duration;
duration = duration + 0.5*(time_given_in(ic+1) - time_given_in(ic));
while time_try < duration
time_try = minimum_duration(fetch_try, u);
fetch_try = fetch_try + 100;
end
%
if fetch_try > fetch_in(ic+1)
fetch_try = fetch_in(ic+1);
else
fetch_try = fetch_try;
end
end
%

```

```

fetch=fetch_try;
end
end
    fclose(fid);

```

## 2. Minimum Duration

```

function [time_req] = minimum_duration(fetch, speed);
%
% This function use the SMB wave prediction method to calculate
% the minimum duration (in second) required to reach a fully developed
% sea for a given wind speed (in m/s) and fetch (in meters)
%
% 1 nautical mile = 6080 ft = 1853.658 m
% 1 knots = 1 nautical mile/hr = 0.5149 m/s
%
K = 6.5882;
A = 0.0161;
B = 0.3692;
C = 2.2024;
D = 0.8798;
grav = 9.8;
p = grav*fetch/(speed*speed);
term1 = A*log(p)*log(p);
term2 = B*log(p);
term3 = D*log(p);
term4 = sqrt(term1 - term2 + C);
term5 = K*exp(term4 + term3);
time_req = term5*speed/grav;

```

## 3. Wave Height and Period

```
function [Hs, Period] = HandT(fetch, speed)
%
% Calculate wave height H (m) and period T (s) for a fully developed wind wave
% using SMB method. The input parameter fetch is in meter and wind speed is in
% m/s
%
% 1 nautical mile = 6080 ft = 1853.658 m
% 1 knots = 1 nautical mile/hr = 0.5149 m/s
%
grav = 9.8;
p = grav*fetch/(speed*speed);
term6 = 0.0125*p^0.42;
Hs = 0.283*speed*speed/grav*tanh(term6);
term7 = 0.077*p^0.25;
Period = 2.40*pi*speed/grav*tanh(term7);
```

## II. Source Code for an ANN Wind-Wave Prediction Model

```
%  
% Set input and output  
%  
input = [wind];  
output = [wave];  
%  
% Set global maximum and minimum wind  
%  
max = 20; % Global maximum  
min = -20; % Global minimum  
%  
% Pre-Processing  
%  
pn = 2*(inp-min)/(max-min)-1;  
  
%  
% Set Time Delay Neural Network with the Scaled Conjugate Gradient learning  
% algorithm, eight hidden neurons (S1), five output neurons (S2), non-linear and  
% linear transfer function at first and second processing % layers  
%  
net = newfftd(pn, [number of time delay] , [S1 S2], {'tansig', 'purelin'}, 'trainscg');  
%  
net = init(net); % Start training an ANN  
net.trainParam.epochs = 30; % Number of iterations = 30  
y = sim(net, pn); % Produce ANN model results  
%  
% Post-Processing  
%  
pp = postmnmx(y);
```



## LITERATURE CITED

- Aberson, S.D., 2001. The Ensemble of Tropical Cyclone Track Forecasting Models in the North Atlantic Basin (1976-1998). *Bull. Amer. Meteor. Soc.*, Vol. 82, pp. 1895-1904.
- Airy, G.B., 1845. On Tides and Waves, Encyclopaedia Metropolitana.
- Amari, S., Murata, N., Muller K.R., Finke, M., and Yang, H.H., 1997. Asymptotic Statistical Theory of Overtraining and Cross-Validation. *IEEE Transactions on Neural Networks*, Vol. 8, No.5, pp. 985-996.
- Baum, E.B. and Haussler, D., 1989. What Size Net Gives Valid Generalization? *Neural Computation*, Vol. 1(1), pp. 151-160.
- Batts, M.E., Cordes, M.R., Russell, L.R., Shaver, J.R., and Simiu, E., 1980. Hurricane Wind Speeds in the United States. Rep. No. BSS-124, Nat. Bureau of Standards, U.S. Department of Commerce, Washington, D.C.
- Booij, N., Rid R.C., and Holthuijsen, L.H., 1999. A third-Generation Wave Model for Coastal Regions: Model description and Validation. *J. Geophys. Res.*, 104(4), pp. 7649-7666.
- Booij, N., Holthuijsen, L.H., and Herbers, T.H.C., 1985. The Shallow Water Wave Hindcast Model HISW, Physical and Numerical Background. Delft University, Faculty of Civil Engineering, Department of Hydraulic Engineering.
- Bretschneider, C.L., 1951. The Generation and Decay of Wind Waves in Deep Water. Tech. Rep. No. 155-46, Ins. Of Engineering Research, Univ. of California, Berkeley, August.
- Bretschneider, C.L., 1952. Revised Wave Forecasting Relationships. *Proc. of the 2<sup>nd</sup> Conference on Coastal Engineering*, ASCE, Council on Wave Research.
- Bretschneider, C.L., 1959. Wave Variability and Wave Spectra for Wind-Generated Gravity Waves. Tech. Memo. No. 118, Beach Erosion Board, U.S. Army Corps of Engineers, August, pp. 192.

- Burgers, G. and Makin, V.K., 1993. Boundary-Layer Model Results for Wind-Wave Growth. *J. Phys. Oceanogr.*, pp. 23-372.
- Clouse, D.S., Giles, C.L., Horne, B.G., and Cottrell, G.W., 1997. Time Delay Neural Networks: Representation and Induction of Finite State Machines. *IEEE Transactions on Neural Networks*, Vol. 8, No. 5, pp. 1065-1070.
- Cybenko, G., 1989. Approximators by Suspensions of a Sigmoid Function. CSRD Report No. 856, Center for Supercomputing in Civil Engineering, Vol. 8., No. 2, pp. 234-251.
- Daniel, K., 1998. Artificial Neural Networks: Univ. of Saint Louis, individual project within MISB-420-0. News of U.S. Department of Commerce, 2003. Washington D.C., 20230. NOAA 30-024.
- Davies, O.L. and Goldsmith, P.L., 1972. Statistical Methods in Research and Production, pp. 235. Tweeddale Court, Edinburgh EH 1YL.
- Deo, M.C., Jha, A., Chaphekar, A.S., and Ravikant, K., 2001. Neural Networks for Wave Forecasting. *Ocean Engineering*, Vol. 28, pp. 889-898.
- Duda, R.O., Hart, P.E., and Stork, D.G., 1997. Pattern Classification(2<sup>nd</sup> ed.). pp. 35.
- Elman, J.L., 1990. Finding Structure in Time. *Cognitive Science*, Vol. 14, pp. 179-211.
- Emanuel, K.A., 1999. Thermodynamic Control of Hurricane Intensity. *Nature*, Vol. 401, pp. 665-669.
- Farely, B. and Clark, W.A., 1954. Simulation of Self-Organizing Systems by Digital Computer. *IRE Transactions on Information Theory*, Vol. 4, pp. 76-84.
- Finnof, W.H.F. and Zimmermann, H., 1993. Improving Model Selection by Nonconvergent Methods. *Neural Networks*, Vol. 6, pp. 771-783.
- Fletcher, D.S. and Goss, E., 1993. Forecasting with Neural Networks: An Application Using Bankruptcy Data. *Information Management*, Vol. 24, pp. 159-167.
- Francis, L., 2001. Neural Networks Demystified. Francis Analytic and Actuarial Data Mininf, Inc., pp 245-319.
- Friedman, J.H., 1991. Multivariate Adaptive Regression Spline. *Annals of Statistics*, Vol. 19, pp. 1-141.

- Geman, S., Bienenstock, E., and Doursat, R., 1992. Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, Vol. 4, pp. 1-58.
- Georgiou, P.N., 1985. Design Wind Speeds in Tropical Cyclone-Prone Regions. Ph.D Thesis, Fac. of Engrg. Sci., University of Western Ontario, London, Ont., Canada.
- Georgiou, P.N., Davenport, A.G., and Vickery, B.J., 1983. Design Wind Speeds in Regions Dominated by the Tropical Cyclones. *J. of Wind Engineering and Industrial Aerodynamics*, Amsterdam, The Netherlands, Vol. 13(1), pp. 139-152.
- Glymour, C., Midigan, D., Pregibon, D., and Smyth, P., 1996. Statistical Inference and Data Mining. *Communications of the ACM*, Vol., 39, No.11, pp. 35-41.
- Hand, D.J., 1999. Statistics and Data Mining: Intersecting Disciplines. *ACM SIGKDD*, Vol. 1, Issue1, pp. 16-19.
- Hanes, M.D., Ahalt, S.C., and Krishnamurthy A.K., 1994. Acoustic to Phonetic Mapping Using Recurrent Neural Networks. *IEEE Transactions on Neural Networks*, Vol. 5(4), pp. 659-662.
- Hasselmann, K., Barnett T.P., Bouwes E., Carlson H., Cartwright K., Enke J.A, Ewing, H., Gienapp, D.E., Hasselmann D.E., Kruseman P., Meerburg A., Olbers D.J., Richter K., Sell W., and Waden H., 1973. Measurement of Wind-Wave Growth and Swell Decay during the Joint North Sea Wave Project, *Dtsch. Hydrogr. Z. Suppl.*, 12, A8.
- Hasselmann, S. and Hasselmann, K., 1985. Computations and Parameterizations of the Nonlinear Energy Transfer in a Gravity\_Wave Spectrum. Part 1: A new method for efficient computations of the exact nonlinear transfer integral. *J. Phys. Oceanogr.*, pp. 15-1369.
- Hassibi, B. and Stork, D.G., 1993. Second Order Derivatives for Network Pruning: Optimal Brain surgeon. *Advances in Neural Information Processing Systems*, Vol. 5, pp. 164-171.
- Haykin, S., 1994. *Neural Networks: A Comprehensive Foundation*, NY: Macmillan, p. 2.
- Hebb, D.O., 1949. *The Organization of Behavior*, Wiley, New York.
- Hogben, N. and Dacunha, N.M.C., May 1985. Wave Climate Synthesis: Some Recent Advances. *Proceedings of the Offshore Technology Conference*, Paper No. OTC 4938. OTC, Houston.

- Holland, G.J., 1997. The Maximum Potential Intensity of Tropical Cyclones. *J. of Atmos. Sci.*, 54, pp.2519-2541.
- Holthuijsen, L.H., Booij, N., and Padilla-Hernandez, R., 1997. A Curvil-Linear, Third Generation Coastal Wave Model. *J. of Coastal Dynamics*, Plymouth, pp. 128-136.
- Hopfield, J.J., 1982. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, Vol. 79, pp. 2554-2558.
- Hornik, K., Stinchcombe, M., and White, H., 1989. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 2, pp. 359-366.
- Ippen, A.T., 1966a. Waves and Tides in Coastal Processes. *J. of the Boston Society of Civil Engineering*, Vol. 53, No. 2, pp. 158-181.
- Ippen, A.T., 1966b. Estuary and Coastline Hydrodynamics, McGraw-Hill, New York.
- Jarvinen, B.R. and Lawrence, M.B., 1985. An Evaluation of the SLOSH Storm Surge Model. *Bull. Amer. Meteor. Soc.*, Vol. 66, pp. 1408-1411.
- Jelesnianski, C.P., 1984. SLOSH (Sea, Lake, and Overland Surges from Hurricanes), NOAA Technical Memorandum. Silver Spring, Md.: NOAA.
- Jelesnianski, C.P., Chen, J., and Shaff, W.A., 1992. SLOSH: Sea, Lake and Overland Surges from Hurricanes, NOAA Tech. Rep. NWS 48.
- Jordan, M.I., 1995. Why the Logistic Function? A Tutorial Discussion on Probabilities and Neural Networks. MIT Computational Cognitive Science Report 9503.
- Jordan, M.I. and Rumelhart, D.E., 1992. Forward Models: Supervised Learning with a Distal Teacher. *Cognitive Science*, Vol. 16, pp. 307-354.
- Jordan, M.J., 1986. Attractor Dynamics and Parallism in a Connectionist Sequential Machine. *Proceedings of 8<sup>th</sup> Annual Conference of the Cognitive Science Society*, pp. 531-546.
- Kartam, N., Flood L., and Garrett Jr. J.H., 1997. Artificial Neural Networks for Civil Engineers: Fundamentals and Applications. ASCE, NY.

- Kim, G. and Barros, A.P., 2001. Quantitative Flood Forecasting Using Multisensor Data and Neural Networks. *J. of Hydrology*, Vol. 246, pp. 45-62.
- Kinsman, B., 1965. Wind Waves, Their Generation and Propagation on the Ocean Surface. NJ: Prentice Hall.
- Kohonen, T., 1995. The Self-Organizing Map. *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1464-1480.
- Komatsu, K. and Masuda, A., 1996. A new scheme of nonlinear energy transfer among wind waves: RIAM method-Algorithm and performance, *J. of Oceanogr.*, pp. 52-509.
- Krogh, A. and Hertz, J.A., 1995. A Simple Weight Decay Can Improve Generalization. In Advances in Neural Information Processing Systems 4. Moody, J. E., Hanson, S. J., and Lippmann, W. P., editions. Morgan Kaufmann Publishers, San Mateo CA, pp. 950-957.
- Kuligowski, R.J., and Barros, A.P., 1998. Experiments In Short-Term Precipitation Forecasting Using Artificial Neural Networks, *Monthly Weather Rev.*, Vol. 126, pp. 470-482.
- Lang, K.J., Waibel, A.H., and Hinton, G.E., 1990. A Time-Delay Neural-Network Architecture for Isolated Word Recognition. *Neural Networks*, Vol. 3 (1), pp. 23-44.
- Mai, S., Ohle N., and Zimmermann C., 1999. Applicability of Wave Models in Shallow Coastal Waters. *Proceedings of the 5<sup>th</sup> Int. COPEDEC*, Cape Town, South Africa.
- McAdie, C.J. and Lawrence, M.B., 2000. Improvements to Tropical Cyclone Track Forecasting in the Atlantic Basin, 1970-1998. *Bull. Amer. Meteor. Soc.*, Vol. 81, pp. 989-999.
- McCullag, H.P. and Nelder, J.A., 1989. Generalized Linear Models, 2<sup>nd</sup> ed., London: Chapman & Hall.
- Møller, M.F., 1993. A scaled Conjugate Gradient Algorithm for Fast Supervised Learning. *Neural Networks*, Vol. 6, pp. 525-533.
- Moreira, M. and Fiesler, E., 1995. Neural Networks with Adaptive Learning Rate and Momentum Terms. IDIAP Technical Report. (<http://citeseer.nj.nec.com/moreira95neural.html>).

Mozer, M.C. and Smolensky, P., 1988. Skeletonization: A technique for trimming the fat from a Network via Relevance Assessment. *Advances in Neural Information Processing System*. Touretzky, D. S. edition, pp. 107-115.

Munk, W.H., 1944. Proposed Uniform Procedure for Observing Waves and Interpreting Instrument Records, Wave Project, Scripps Institute of Oceanography Lajolla, California.

Nelson, M.C. and Illingworth, W.T., 1991. A Practical Guide to Neural Nets, Reading. MA: Addison-Wesley, pp. 165.

Neumann, C.J., 1993. Global Overview-Chapter 1: Global Guide to Tropical Cyclone Forecasting. WMO/TC-No. 560, Report No. TCP-31, World Meteorological Organization; Geneva, Switzerland.

Philips, O.M., 1957. On the Generation of Waves by Turbulent Wind. *J. Fluid Mechanics*, Vol. 2, pp. 417-445.

Pierson, W. J., Neumann, G., and James, R., 1955. Practical Methods for Observing and Forecasting Ocean Waves by Means of Wave Spectra and Statistics. U.S. Navy Hydrographic Office, Washington D.C., pp. 284.

Pierson, W.J. and Moskowitz, 1964. A Proposed Spectral Form for Fully Developed Seas Based on the Similarity Theory of S.A. Kitaigorodskii. *J. Geophys. Res.*, pp. 5181-5191.

Powell, M.D. and Houston, S.H., 1996. Hurricane Andrew's Landfall in South Florida. Part II: Surface Wind Fields and Potential Real Time Applications. *Weather Forecasting*, Vol. 11, pp. 329-349.

Principe, J.S., Euliano, N.R., and Lefebvre, W.C., 2000. Neural and Adaptive Systems-Fundamentals Through Simulations. NY: John Wiley & Sons, pp.11-12.

Rabelo, L. C., 1990. A Hybrid Artificial Neural Network and KBES Approach to Flexible Manufacturing System Scheduling. UMI Dissertation Services, Ann Arbor, MI.

Reed, R., 1993. Pruning Algorithms-A Survey. *IEEE Transactions on Neural Networks*, 4, pp. 740-747.

Resio D.T., 1987. Shallow-Water Waves, Part I: Theory. *J. Waterways Port, Coast and Ocean Engineering*, Vol. 113, pp. 264-281.

- Resio D.T., 1988. Shallow-Water Waves, Part II: Data Comparison. *J. Waterways Port, Coast and Ocean Engineering*, 113, pp. 50-65.
- Resio, D. and Perrie, W., 1989. Implications of an F-4 Equilibrium Range for Wind-Generated Waves. *J. of Phys. Oceanogr.*, Vol. 19 (2), pp. 193.
- Ripley, B.D., 1994. Neural Network and Related Methods for Classification. *J. of the Royal Statistical Society*, Series B, Vol. 56, pp. 409-456.
- Robert, M. and Christopher, C.R., 2002. NOAA Wave Watch III (NWW3) Model Biases in Gulf of Mexico and Western Atlantic: An Operational Case Study and Impact on Tropical Prediction Center Marine Forecasts. (<http://www.erh.noaa.gov/er/mhx/marsem02/Wavewatch2.pdf>)
- Rosenblatt, F., 1958. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, Vol. 65, pp. 386-408.
- Rumelhart, D.E. and McClelland, J.L. (Eds.), 1986. *Parallel Distributed Processing*. Cambridge, Mass: MIT Press.
- Rumelhart, D.E. and Zipser, 1985. Feature Discovery by Competitive Learning. *Cognitive Science*, Vol. 9, pp. 75-112.
- Sarel, W.S., 1994. Neural Networks and Statistical Models. In SAS Institute Inc., *Proceedings of the 19<sup>th</sup> Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc., pp. 1538-1550.
- Sarel, W.S., 1995. Stopped Training and Other Remedies for Over-fitting, *Proceedings of the 27<sup>th</sup> Symposium on the Interface of Computing Science and Statistics*, pp. 352-360.
- Schwab, D.J., Bennett, J.R., Liu, P.C., and Donelan, M.A., 1984. Application of a Simple Numerical Wave Prediction Model to Lake Erie. *J. of Geophysical Research*, Vol. 89, no. C3, pp. 3586-3592.
- Shore Protection Manual, 1977. Coastal Engineering Research Centre, US Army Corps of Engineers, Washington DC.
- Smith, F.G.W., 1973. *The Seas in Motion*. NY: Thomas Y. Crowell Co., pp. 248.
- Smith, M., 1996. *Neural Networks for Statistical Models*. Boston: International Thomson Computer Press, ISBN 1-850-3284-0.

Solla, S., Cun, Y, L, and Denker, J., 1990. Optimal Brain Damage. *In Advances in Neural Information Processing Systems*, Vol. 2, pp. 598-605.

Stokes, G.C., 1880. On the Theory of Oscillatory Waves, *Mathematical and Physical Papers*, Vol. 1. Cambridge: Cambridge University Press.

Sverdrup, H.U. and Munk, W.H., 1947. Wind Sea And Swell: Theory of Relations for Forecasting. U. S. Navy Hydrographic Office, Washington, D. C., pp. 44.

SWAMP, 1985. An Inter-Comparison Study of Wind Wave Prediction Models, Part 1: Principal Results and Conclusions in Ocean Wave Modeling. NY: Plenum Press, pp. 44.

Swingler, k., 1996. Applying Neural Networks: A Practical Guide. CA: Academic, California, pp. 303.

Taylor, J.G., 1993. The Promise of Neural Networks. London: Springer-Verlag.

Tikhonov, A., and Arsenin, V., 1977. Solution of Ill-Posed Problems. John Willy Sons, New York.

Tolman, H.L., 1989. The Numerical Model WAVEWATCH: A Third Generation Model for The Hindcasting of Wind Waves on Tides in Shelf Sea. *Communications on Hydraulic and Geotechnical Engineering*, Delft University of Techn., ISSN 0169-6548, Rep. No. 89-2, pp. 72.

Tolman H.L., 1991. A Third Generation Model for Wind Waves on Slowly Varying, Unsteady and Inhomogeneous Depths and Currents. *J. Phys. Oceanogr.*, Vol. 21, pp. 782-197.

Tolman H.L., 1992. Effects of Numerics on the Physics in a Third Generation Wind-Wave Model. *J. Phys. Oceanogr.*, Vol. 22, pp. 1095-1111.

Tolman H.L., 1997. User Manual and System Documentation of WAVEWATCH-III Version 1.15. NOAA/NWS/NCEP/OMB Technical Note, Vol. 151, pp. 97.

Tsai, C.P. and Lee, T.L., 1999. Back-Propagation Neural Network in Tidal-Level Forecasting. *J. of Waterway, Port, Coastal and Ocean Engineering*, Vol. 125, pp. 195-202.

Tsai, C.P., Lin, C., and Shen, J.N., 2002. Neural Network for Wave Forecasting among Multi-Stations. *J. Ocean Engineering*, Vol. 29, 1683-1695.



- Vickery, P.J., Skerlj, P.F., and Twisdale, L.A., 2000. Simulation of Hurricane Risk in the U.S. Using Empirical Track Model. *J. of Structural Engineering*, pp. 1222-1237.
- Vickery, P.J., and Twisdale, L.R., 1995a. Wind-Field and Filling Models for Hurricane Wind-Speed Prediction. *J. Struct. Engineering*, Vol. 12(11), pp. 1700-1709.
- Vickery, P.J., and Twisdale, L.R., 1995b. Prediction of Hurricane Wind Speeds in the United States. *J. Struct. Engineering*, Vol. 12(11), pp. 1691-1699.
- WAMDI, 1998. (13 authors, including Cardone, V.J. and Greenwood, J.A.). The WAM Model a Third Generation Ocean Wave Prediction Model. *J. of Phys. Oceanog.*, Vol. 18, pp. 1775-1810.
- Watrous, R. L., Ladendorf, B., and Kuhn, G., 1990. Complete Gradient Optimization of a Recurrent Network Applied to /b/, /d/, /g/ Discrimination. *J. of the Acoustical Society of America*, Vol. 78(1), pp.375-380.
- Wen S.C., Qian C.C., Ye, A.L., 1999. Wave Modeling Based on an Adopted Wind-Wave Directional Spectrum. *J. Ocean Univ. Qingdao*, pp. 29-345.
- Whitham, G.B., 1974. Linear And Nonlinear Waves. Wiley, New York.
- Widrow, B. and Hoff M.E., 1960. Adaptive Switch Circuit. NY: IRE WESCON Convention Record.
- Wiegel, R.L., 1960. A Presentation of Cnoidal Wave Theory for Practical Application. *J. of Fluid Mechanics*, Vol. 7, pp. 273-286.
- World Meteorological Organization, 1988. Guide to Wave Analysis and Forecasting. Secretariat of the World Meteorological Organization, Geneva, No. 702.
- Zurada, J.M., 1992. Introduction to Artificial Neural Systems. PWS, Boston, p15.
- Zwart, J.P. and Vries, J.D., 2001. The Effects of Using Biases in Perceptron Networks on Classification (unpublished).