STORM WAVE PREDICTION USING NN AND GP

1.2

S. B. Charhate and M. C. Deo*

Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India

1 INTRODUCTION

The rise in water level along with accompanying waves during cyclones causes severe damage to coastal installations and further poses serious threat to shipping and other routine ocean activities. The prediction of resulting values of these variables during a storm period is therefore essential to protect lives and properties and also to issue appropriate warnings to coastal and ocean operators. The increase in sea levels and also the rise in wave heights during a storm period are highly uncertain in space as well as in time and thus calls for application of soft computing tools such as neural networks and genetic programming.

Since recent past the technique of neural network (NN) has also been employed by some researchers as complementary to the numerical schemes (Lee, 2006; Tseng et al. 2007; Lee 2008). Neural networks have now become an established computing tool for many applications in ocean engineering (Jain and Deo, 2006). Neural networks do not assume any mathematical model a priori and hence are more flexible in data mining. Lack of requirement of process knowledge and that of any exogenous information, data error tolerance and easy adaptability to new observations are some of the additional attributes that this scheme possesses.

In recent past another soft computing tool, namely genetic programming (GP) has become available to practicing engineers and it would be of interest to see how it performs against that of NN. The present study is oriented along this direction.

* *Corresponding author: M.C. Deo*
Tel: +91(022)25767301; fax:+91(022)25767302
.*E-mail address*: mcdeo@civil.iitb.ac.in

Like genetic algorithm (GA) the concept of genetic programming also follows the principle of 'survival of the fittest' borrowed from the process of evolution occurring in nature. But unlike GA its solution is a computer program or an equation as against a set of numbers in the GA and hence it is convenient to use the same as a regression tool. A good explanation of various concepts related to GP can be found in Koza (1992).

In developing countries it becomes difficult to collect data of a large number of parameters required to accurately forecast ocean parameters resulting from cyclones and their movements. Hence in this study an attempt is made to use only those parameters than can be collected through a single wave rider buoy deployed at the station of forecast.

Another specialty of this study is that it attempts to provide forecasts up to a longer lead time of 12 hr and beyond during a cyclone.

2. THE DATABASE USED

In the present study the storm wave activity over and above the storm surge generated due to hurricanes that occurred in the Gulf of Mexico during the years: 2003 to 2005 is considered. The major hurricanes namely, Ivan, Dennis, Katrina, Rita are included in the study for model training and testing. Hurricane Ivan was category-4 hurricane that prevailed from Sept. 2 to 24, 2004, while Hurricane Dennis was smaller in intensity (category-3) as well as in time duration (July 5-13, 2005). Hurricane Katrina was most severe (category-5), although it lasted only for five days (August 27-31, 2005). Hurricane Rita happened during Sept. 19 to 24, 2005 and was of category-5 in intensity.

The records available at two wave rider buoy locations in the Gulf of Mexico, namely, 42001 and 42039 (Fig. 1) are considered in this work. The measurements made during the above mentioned hurricane periods were downloaded from the website: http://www.ndbc.noaa.gov/.

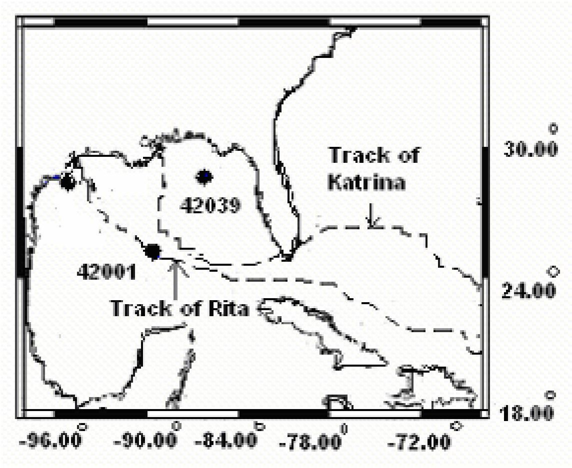

Fig.1 Tracks of 'Rita' and 'Katrina' and data buoy locations in the Gulf of Mexico (www.ndbc.noaa.gov)

Storm waves are influenced by major causative factors such as wind speed along with its direction, fetch and duration. In addition the characteristics and track of the storm and its forward speed also influence the wave distribution in a given region. Our objective was to explore the feasibility of using only the wave buoy measurements and hence in order to evaluate storm waves using such observations we have used four variables as input, namely, wind speed, wind direction; wind gust and barometric pressure observed during isolated storm events. It may be noted that the parameter of wind speed here indicates its average value over a short recording interval while that of the wind gust means its maximum value over such an interval. During cyclones the barometric pressure plays a crucial role and hence the same is included as an additional input. The effect of all other causal parameters including time duration and fetch has been accounted for by using preceding observations of the time series as an additional input. By considering such a segment of preceding measurements every time, a time series modeling component is added to the analysis. While modeling any time series it is assumed that causative factors get reflected in the very occurrence of the time series values and therefore if we model the sequence of such observations, the same should be regarded as sufficient to forecast the future occurrence.

3. GP AND NN

The GP may be viewed as an extension of GA. In GP a random population of individuals (equations or computer programs) is created, the fitness of individuals is evaluated and then the 'parents' are selected out of these individuals. The parents are then made to yield 'offspring's' by following the process of reproduction, mutation and cross-over. The creation of offspring's continues in an iterative manner till a specified number of offspring's in a generation are produced and further till another specified number of generations are created. The best offspring (an equation or a computer program) resulting in this process is the solution of the problem. Appendix I gives the major steps in implementing GP.

Applications of GP in the field of ocean engineering are conspicuous by their absence. But most recently Charhate et al. (2007) have applied GP to predict coastal currents in tide dominated areas in the Gulf of Khambhat, India and found it to be more useful than traditional harmonic analysis and NN. The application of GP to storm surge prediction is difficult to find among normally available published works.

Some investigators in recent past working with other random variables than the storm surge have compared performance of GP with traditional statistical methods as well as NN. Drecourt (1999) reported that GP handles peak river flows better while NN takes care of the input noise efficiently. Muttil and Liong (2004) found performance of GP marginally better than NN. On similar lines Wighnam and Crapper (2001) noticed that when the rainfall-runoff correlation was strong then GP is better than lumped model; otherwise it does not have much relative advantage. Hong and Rao (2003) found that GP results are comparable with NN and both produced more accurate values than the statistical methods. GP can automatically select input and tell which are more important (and thus produce parsimonious results as per Hong and Rosen (2002) unlike the NN and further the GP results are understandable. Babovic et al. (2004) show the accuracy as well as exhaustiveness of the GP equations compared with traditional regressions, but opine that a combination of data driven and theory guided models are better in performance. Keijzer et al. (2005) also support this adding that gaps in the knowledge can be filled up by this and that the same advantage exists even if synthetic data are used rather than actual observed. Fonlupt (2001) employed GP to evaluate ocean component (phytoplankton and others) from sunlight luminance and reflectance and found GP better than polynomials fits and rivaling NN.

The NN used in this work is of feed forward type, which has been most commonly used in the past studies and where the information flows only in the forward direction, i. e. from the input layer to the output layer and through a hidden layer – all layers consisting of a set of neurons or computational elements. The basic concepts and details of the working of a NN can be seen in text books like Wu (1994) and Wasserman (1988). Jain and Deo (2006) provide a review of applications of NN in ocean engineering. A variety of learning algorithms have been tried in this work to impart training to the network and these ranged from ordinary error back propagation to advanced search based techniques. The Levernberg-Marquardt algorithm (Demuth et al., 1998) turned out to be the best method of training in the enc and hence the same has been adopted for testing.

4. ANALYSIS AND RESULTS

The statistical measures of correlation coefficient, R, root mean square error, RMSE, and mean absolute error, MAE, have been used in this study to compare the GP and NN estimations with actual observations.

As stated earlier measurements at two stations: 42001 and 42039 have been used in this study.

At station 42001 the total number of hurricane events recorded was 5, 7 and 11 respectively for the hurricane in the years: 2003, 2004 and 2005. The hurricanes that occurred in the seasons of 2003 and 2005 were used to build the GP and NN models while those of the hurricane season of the year 2004 were used to test these models. The model training and testing has been done on storm-to-storm basis.

The number of hurricane events recorded at the other station no. 42039 in the years of 2003, 2004 and 2005 were 8, 9 and 8 respectively. The hourly values for the hurricane season of 2003 and 2004 were used to build the GP and NN models with similar data division for both methods, while the hurricane season of 2005 was used to test these models.

Alternative GP and NN models were built as per the following functional relations:

$$Hs_{(t)} = f(Ws, Wd, Wg, \Pr, Hs_{(t-1)}) \quad (1)$$

$$Hs_{(t)} = f(Ws, Wd, Wg, \Pr, Hs_{(t-1)}, Hs_{(t-2)}) \quad (2)$$

$$Hs_{(t)} = f(Ws, Wd, Wg, \Pr, Hs_{(t-1)}, Hs_{(t-2)}, Hs_{(t-3)}) \quad (3)$$

$$Hs_{(t)} = f(Ws, Wd, Wg, \Pr, Hs_{(t-1)}, Hs_{(t-2)}, Hs_{(t-3)}, Hs_{(t-4)}) \quad (4)$$

Where $Hs_{(t)}$, $Hs_{(t-1)}$ = significant wave heights at time *t, t-i*; Ws = wind speed, Wd = wind direction, Wg = wind gust and $\Pr$ = pressure at time t.

A comparison of the outcome of this exercise with the actual measurements showed that the GP and NN models calibrated and tested on the basis of the model formation shown in eq. (4) produced the best evaluations at both the locations. Table 1 (column 3) shows an overall testing performance of the GP and NN models. Figures 2 and 3 show a qualitative performance of the GP models at location 42001 in terms of time history and scatter plots, while Fig. 4 and 5 show the scatter and time series plot at site 42039. These Figures along with Table 1 thus reveal that the use of an additional input as surrogate to unobserved causative factors in terms of preceding time series values indeed paid rich dividends, especially when overall error criteria of RMSE and MAE were considered. Table 1 also indicates that the GP exhibits better performance than the NN.

It is thus apparent that most accurate predictions are possible if one uses a GP model that involves a combination of causal as well as temporal mapping. In such a case it is possible to obtain the output with 0.15 m RMSE and 0.06 m MAE at station 42039 and those with 0.07m RMSE and 0.03 MAE at location 42001.

Table 1 Testing performance of NN, GP models

1	2	3		
Station ID	Method	Model performance with $Hs_{(t)} = f(Ws, Wd, Wg, \Pr, Hs_{(t-1)}, Hs_{(t-2)}, Hs_{(t-3)}, Hs_{(t-4)})$		
		R	RMSE (m)	MAE (m)
42001	NN GP	**0.98** **0.99**	0.21 0.07	0.10 0.03
42039	NN GP	**0.97** **0.99**	0.31 0.15	0.20 0.06

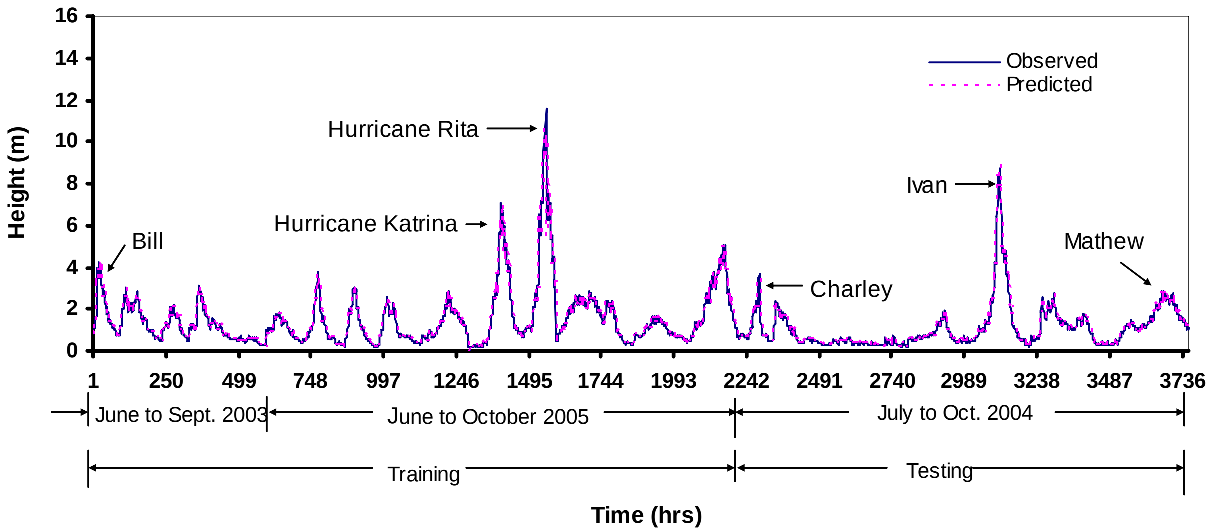

Fig. 2 Time series plot at St. 42001: Training and testing results of GP model

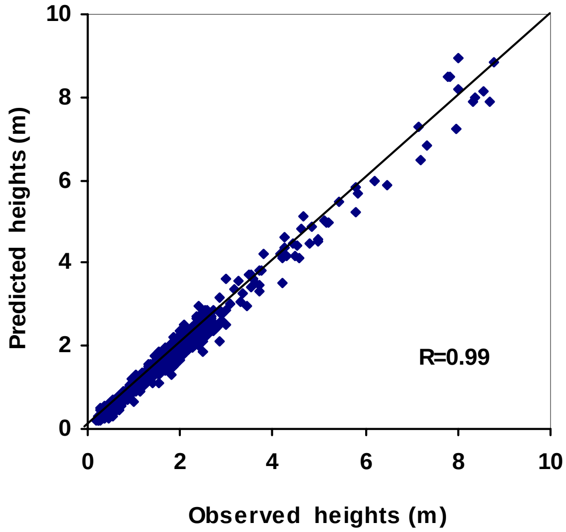

Fig. 3 Scatter plot at St. 42001

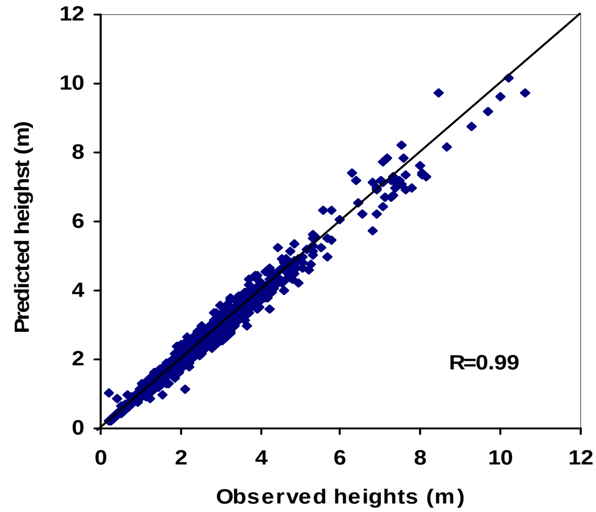

Fig.4 Scatter plot at St. 42039

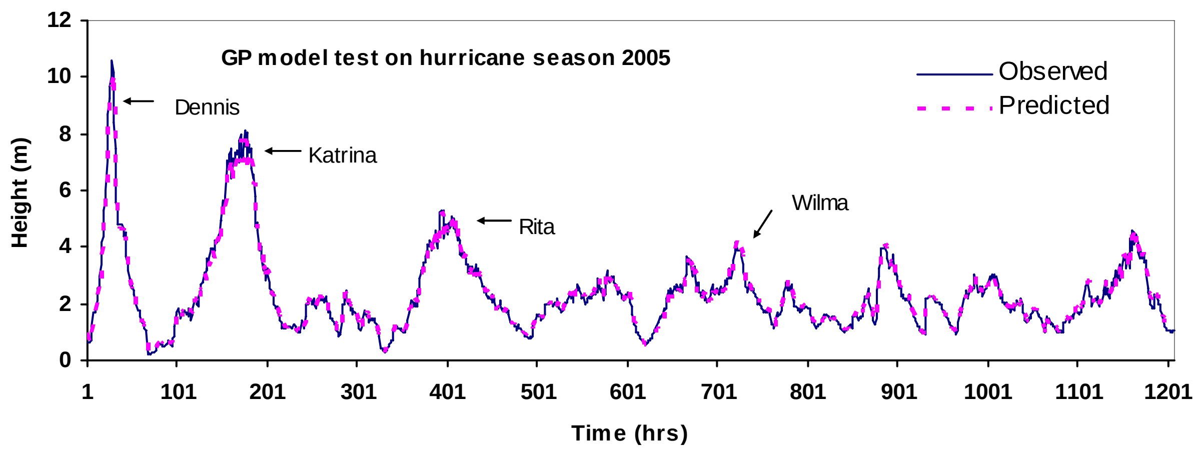

Fig. 5 Time series plot at St. 42039: testing results of GP model

5. FORECASTING FOR FUTURE TIME STEPS

The earlier sections dealt with wave estimations at the same time step. From practical applications it would be useful if these values are predicted over multiple time steps in the future. In the present study we have built GP and NN models for prediction over a lead time up to 24 hr at stations 42001 and 42039. The sample size and the division of dataset for training and testing of these models are similar to the previous case of estimation at the same time step. The GP, NN models are calibrated and tested as per the following equation:

$$Hs_{(t+i)} = f(Ws, Wd, Wg, \Pr, Hs_{(t)}, Hs_{(t-1)}, Hs_{(t-2)}, Hs_{(t-3)}) \quad (5)$$

Where the subscript t+i denotes the value at time t+i and t is the current time.

By trial it was found sufficient to go up to the third time step (t-3) in the past. Figures 6 to 9 show as examples the testing performance of GP models in terms of scatter and time series plots for 6 hr and 12 hr ahead forecasts for station 42001 while Figures 10 to 13 indicate the same for station 42039. Tables 2 and 3 show an overall testing performance of the GP and NN models at stations 42001 and 42039 respectively. The results based on GP are more attractive than those arrived at from the use of NN.

It thus appears that the prediction over a period of subsequent 12 hr can be made with RMSE = 0.48 m and MAE = 0.28 m at location 42001 and with RMSE = 0.78 m and MAE = 0.49 m at site: 42039.

Table 2 Prediction at station 42001

Time interval in hrs	Method	(R)	RMSE m	MAE m
3	GP NN	0.98 0.97	0.20 0.26	0.11 0.17
6	GP NN	0.95 0.94	0.27 0.38	0.16 0.23
9	GP NN	0.93 0.90	0.31 0.46	0.20 0.27
12	GP NN	0.92 0.84	0.48 0.62	0.28 0.39
18	GP NN	0.78 0.72	0.57 0.79	0.36 0.48
24	GP NN	0.69 0.63	0.65 0.88	0.47 0.56

Table 3 Prediction at station 42039

Time interval in hrs	Method	R	RMSE (m)	MAE (m)
3	GP NN	0.97 0.96	0.32 0.44	0.19 0.28
6	GP NN	0.91 0.90	0.59 0.68	0.27 0.43
9	GP NN	0.90 0.88	0.65 0.73	0.35 0.52
12	GP NN	0.82 0.76	0.78 0.95	0.49 0.64
18	GP NN	0.69 0.62	0.85 1.19	0.55 0.77
24	GP NN	0.58 0.49	0.97 1.28	0.69 0.87

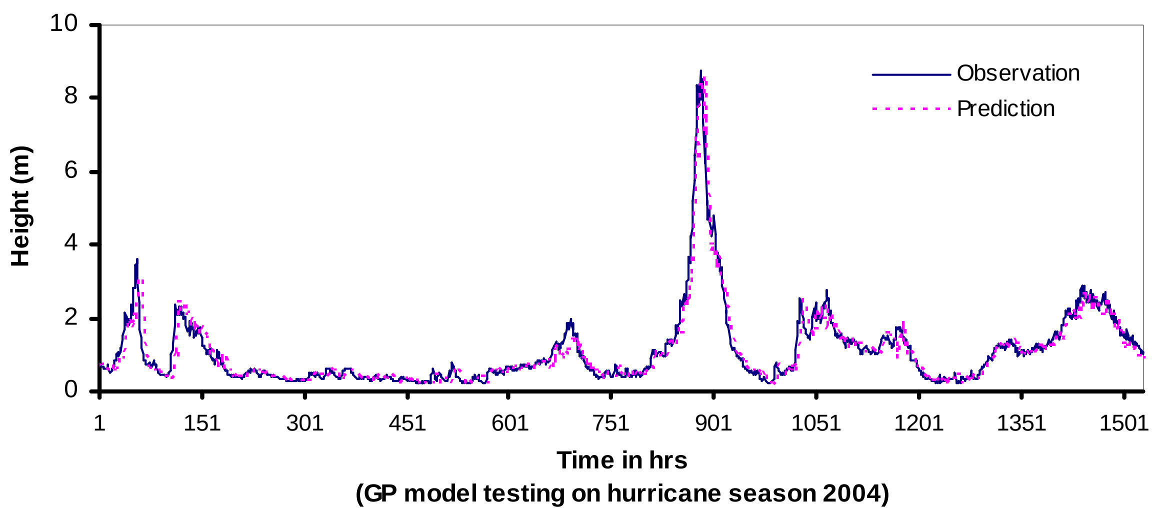

Fig. 6 Time series plot for 6 hr ahead forecast at St. 42001

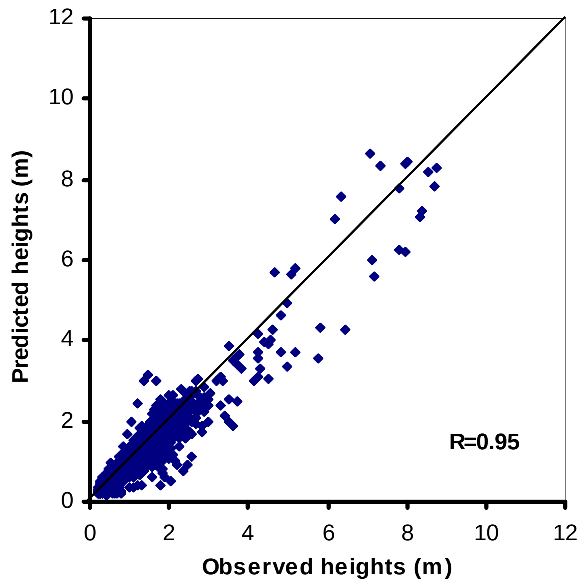

Fig. 7 Scatter plot for 6 hr ahead forecast at St. 42001

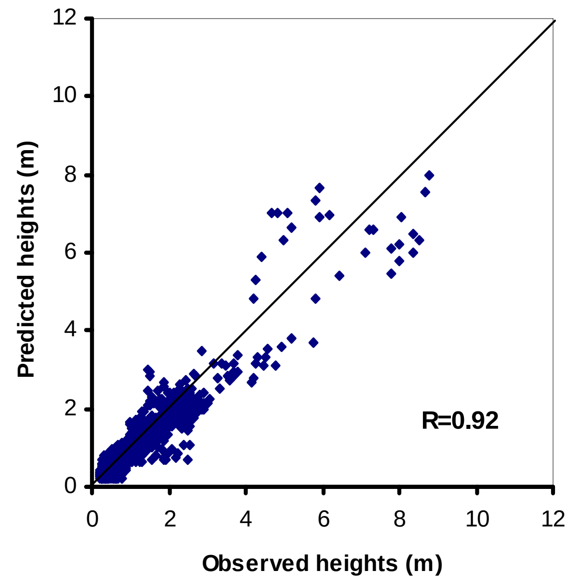

Fig. 8 Scatter plot for 12 hr ahead forecast at St. 42001

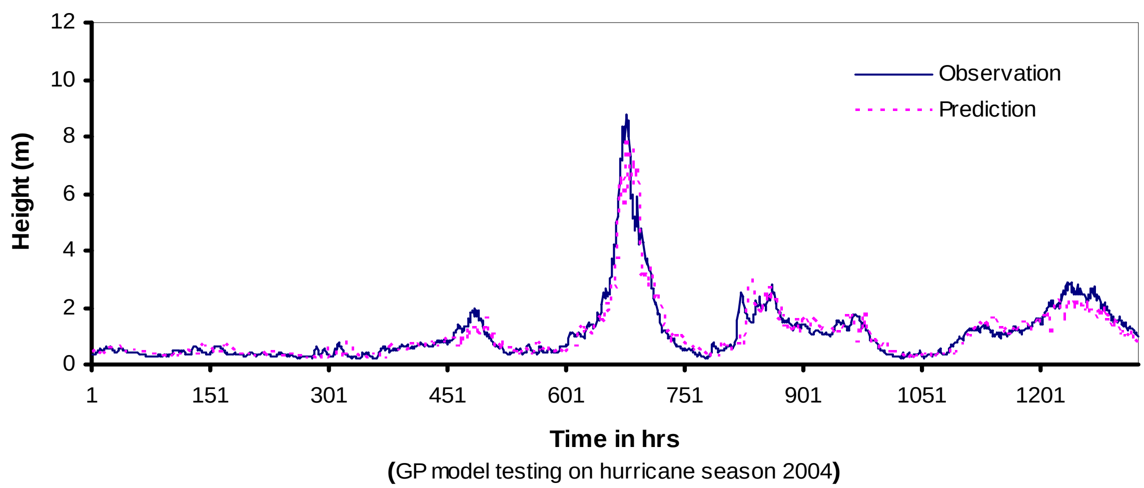

Fig. 9 Time series plot for 12 hr ahead forecast at St. 42001

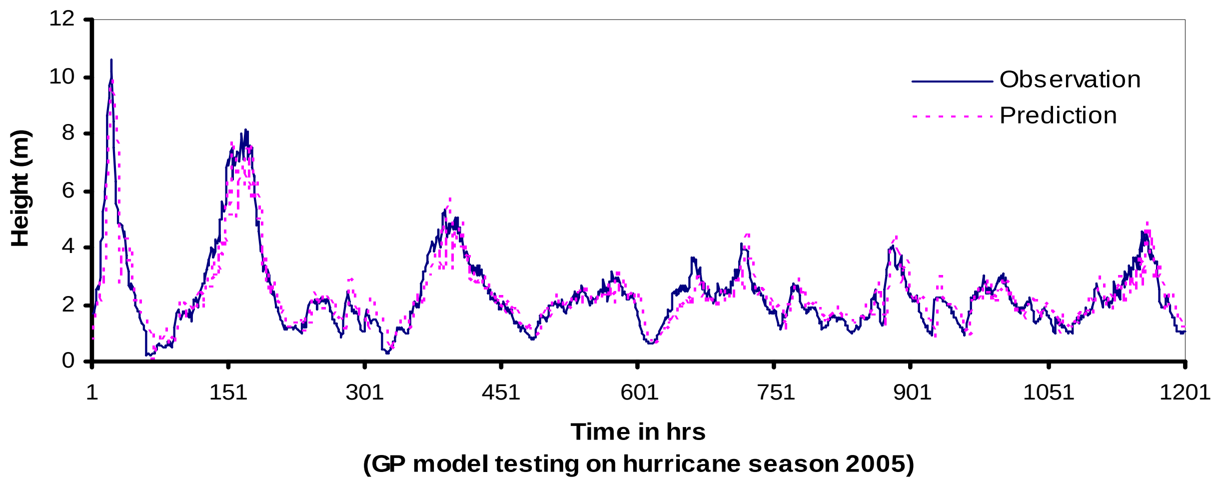

Fig. 10 Time series plot for 6 hr ahead forecast at St. 42039

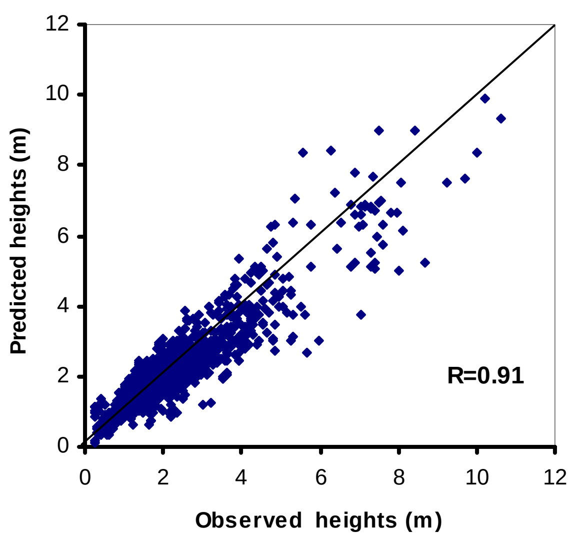

Fig. 11 Scatter plot for 6 hr ahead forecast at St. 42039

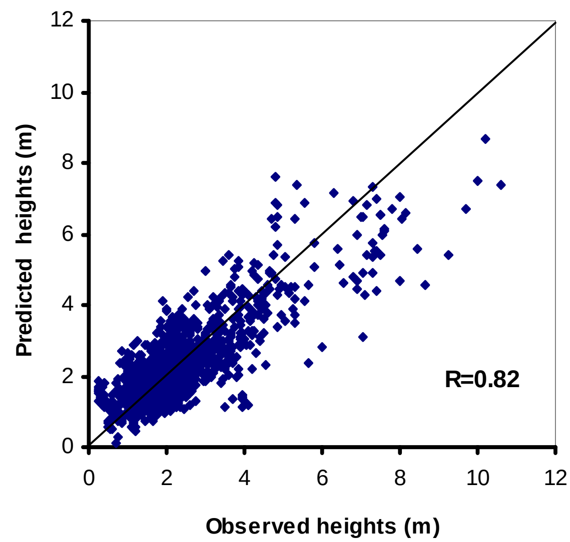

Fig. 12 Scatter plot for 12 hr ahead forecast at St. 42039

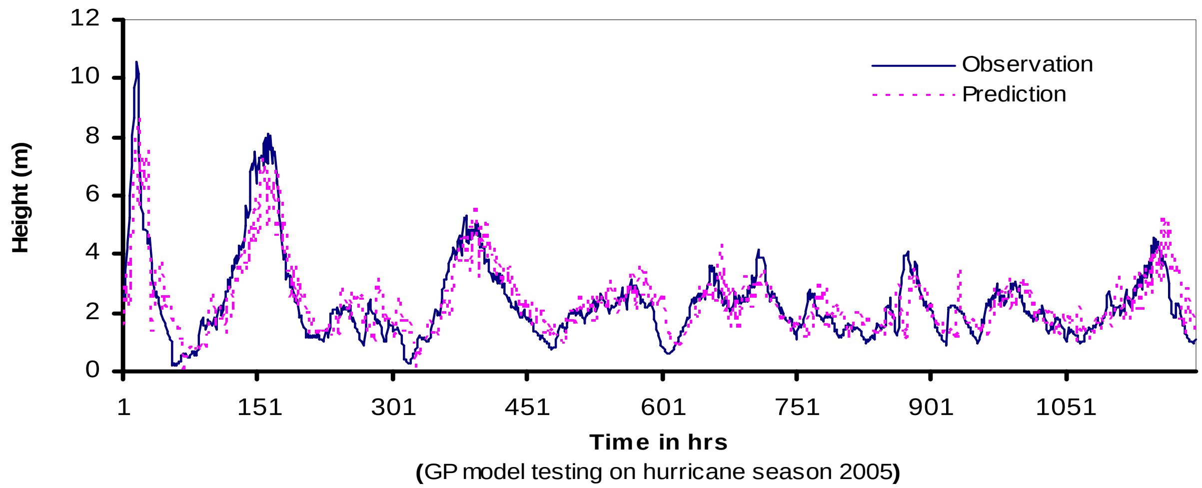

Fig. 13 Time series plot for 12 hr ahead forecast at St. 42039

6. CONCLUSIONS

An attempt has been made in this study to evaluate and predict storm waves at a given ocean location, where wave rider buoy measurements are only available. Genetic programming and neural network based models were developed for this purpose. It was found that a hybrid method based on causal as well as temporal values collected by the wave buoy satisfactorily predicted waves during the hurricanes. Although results of the GP model rivaled those of the NN scheme, the accuracy achieved through GP was marginally but consistently better than that of the NN. The GP also handled the prediction of higher values better than the NN method. The estimation at the same time step by GP resulted in producing such values with RMSE of only 0.07 m and MAE of only 0.03 m at the location 42001. Similarly GP was able to predict 12-hr ahead values at this site with RMSE of 0.48 m and MAE of 0.28 m. The better predictions at large lead times by GP indicated that it can understand weaker dependency structures satisfactorily.

9 LIMITATIONS

The study involved the use of the NN architecture of feed forward type. The networks were trained using most efficient training schemes. Although maximum efforts was made to come up with the best NN model, adoption of alternative network architectures such as RBF, GRNN and ANFIS could have improved performance of the NN.

Appendix 1: IMPLEMENTATION OF GP

The implementation of the GP was done as per the following four steps:

A. An initial population of individuals (equations or programs) of a certain size is created by randomly picking up the same from a set of terminals (consisting of input variables and constants) and functions (involving operators like, multiplication, addition, subtraction, division, square root, log, etc.). As an example consider a program $[(-q + (\pi)^{1/2})/ 3\ p]$ in the form of a tree structure as in Fig. 14. A population of random trees representing the programs is initially constructed and genetic operations are performed on these trees to generate individuals with the help of two distinct sets; the terminal set T and the function set F. For Fig. 14:

$$\{ -\ +\ \surd\ / \} \subseteq F \quad \text{and } \{ \pi, 3, p, q\} \subseteq T$$

In order to generate a new tree one has to pick randomly from $T \cup F$, until all branches end up in terminals.

B. The fitness of each individual in a population is evaluated through a criterion like the root mean square error.

C. The individuals or parents are selected probabilistically through a tournament involving comparing two parents at a time and thereafter short listing the winner for further competition.

D. New offspring's (individuals) are generated from these parents by following procedures a, b and c below:

a. Cross-over: Two individuals are chosen as per the fitness. The cross-over is performed. In the cross-over two random nodes are selected from inside such program (parents) and thereafter the resultant sub-trees are swapped, generating two new programs as in Fig 15. The resulting individuals are inserted into the new population. Individuals are increased by 2.

b. Mutation: One individual is selected as per the fitness. The mutation is performed. In the mutation a sub-tree is replaced by another one randomly (Fig.16). The mutant is inserted into the new population. Individuals are increased by 1.

c. Reproduction: The best program is copied as it is as per the fitness criterion and included in the new population. Individuals are increased by 1.

E. If the number of individuals (offspring's) equals a maximum (selected) number, the number of generations is increased by 1 and we go to step F; otherwise the individuals are increased by repeating steps B-E.

F. If the number of generations is equal to a certain maximum value, the program is terminated; otherwise steps B-E are repeated.

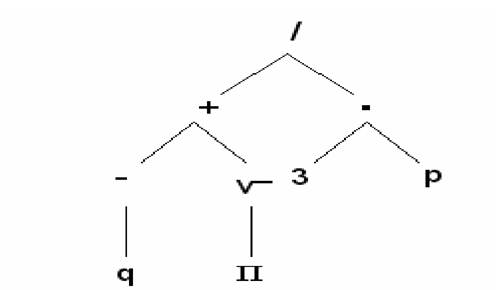

Fig. 14 Program [(-q + $(\pi)^{1/2}$)/ 3 p] in the form of a tree structure

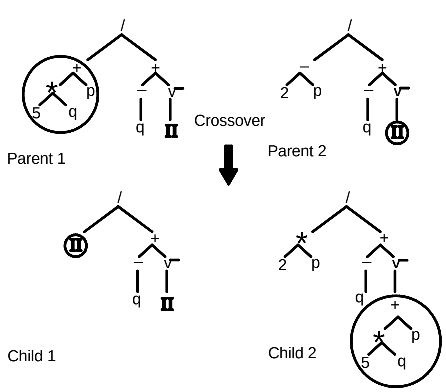

Fig. 15 The Cross-over

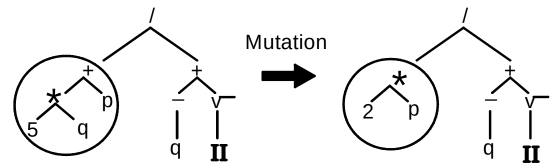

Fig. 16 The Mutation

References

Babovic, V., Keijzer, M., Aguilera, D.R., and Harrington, J., 2004: An Evolutionary Approach to Knowledge Induction: Genetic Programming In Hydraulic Engineering, ASCE Conference Proceeding.

Charhate, S. B., Deo, M. C., and Sanil Kumar V., 2007: Soft and hard computing approaches for real time prediction of currents in a tide dominated area, Journal of Engineering for the Maritime Environment, Proceedings of the Institution of Mechanical Engineers, London, Part M, (221), M4, 147-166

Demuth H, Beale M, and Hagen M.,1998: Neural network toolbox user's guide. 48 Natick (MA, USA): The Mathworks Inc.

Drecourt, J.P., 1999: Application Of Neural Networks and Genetic Programming to Rainfall Runoff modeling. Danish Hydraulic Institute (Hydro-Informatics Techonologies HIT, June, D2K-0699-1.

Fonlupt, C., 2001: Solving the ocean color problem using genetic programming. Journal of Applied Soft Computing, 1, 63-72.

Hong, Y.S. and Rao, Bhamidimarri, 2003: Evolutionary self-organising modelling of a municipal wastewater treatment plant. Journal of Water Research, 37(6), 1199-1212.

Hong, Y.S., and Rosen, M.R.,2002: Identification of an urban fractured-rock aquifer dynamics using an evolutionary self-organizing modelling. Journal of Hydrology, 259(1-4), 89-104.

Jain, P., and Deo, M.C., 2006: Neural networks in ocean engineering. International Journal of Ships and Offshore Structures, Woodhead Publications, Cambridge, U.K., 1(1), 25-35.

Keijzer, M., Baptist, M., Babovic, V., Collet, P., and Uthurburu., 2005: Genetic And Evolutionary Computation Conference Proceedings of the 2005 conference on Genetic and evolutionary computation. Washington DC, USA: 1999–2006.

Koza, J. R., 1992: Genetic Programming on the Programming of Computers by Means of Natural Selection. Bradford Book, MIT Press.

Lee, T.L., 2006: Neural network prediction of a storm surge. Elsevier, Ocean Engineering, 33, 483-494.

Lee,T.L., 2008: Back-propagation neural network for the prediction of the short-term storm surge in Taichung harbor, Taiwan, Elsevier Journal of Engineering Applications of Artificial Intelligence,doi:org/10.1016/j.engappai.2007.03.00.2

Muttil, N., and Liong, S.Y., 2004: A Superior Exploration-Exploitation Balance in Shuffled Complex Evolution. Journal of Hydraulic Engineering, ASCE: 211–220.

Tseng, C.M., Jan, C.D., Wang, J.S., and Wang, C.M., 2007: Application of artificial neural networks in typhoon surge forecasting. Elsevier Journal of Ocean Engineering. 34, 1757-1768.

Wasserman, P.D., 1988: Neural Computing-theory and Practice, Van Nostrand Reinhold, N.Y.,

Whigham, P. A. and Crapper, P. F., 2001: Modeling Rainfall-Runoff using Genetic Programming, Mathematical and Computer Modeling Canberra, Australia, 33, 707-721.

Wu, J.K., 1994: Neural Networks and Simulation Methods, Marcel Dekker, New York 1994.