International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS)
Volume VI, Issue X, October 2017 | ISSN 2278-2540

A Hybrid Wavelet GP Model for Enhancing Forecasting Accuracy of Time Series Significant Wave Heights

Vinay Anand Dhanvada[1], Dr. Paresh Chandra Deka[2]

[1]*Applied Mechanics and Hydraulics Department, National Institute of Technology Karnataka, Surathkal, Karnataka, India*
[2]*Associate Professor, Applied Mechanics and Hydraulics Department, National Institute of Technology Karnataka, Surathkal, Karnataka, India*

I. INTRODUCTION

Wind waves are very complex in nature. The representation of a wave field is normally done by significant wave height and significant time period which retains much of the insight gained from theoretical studies. Wave prediction is the prediction of wave parameters based on the meteorological and oceanographic data. Wave forecasting is extremely useful in the planning and maintenance of the marine activities. The representation of a wave field by significant height and period has the advantages of retaining much of the insight gained from theoretical studies. Its value has been demonstrated in the solution of many engineering problems. A significant wave height is defined as the average height of the one-third highest waves and it is about equal to the average height of the waves as estimated by an experienced observer. During recent decades, some black-box models have been applied to simulate the wave and the wave heights.

Although the ANNs are useful tools in the time series wave modeling, the obvious disadvantage of the ANNs is that they represent their knowledge in term of a weight matrix that is not accessible to human understanding at present *(Savic et al, 1999)*; in other words, these types of models are so implicit that they cannot be simply used by other investigator.

Therefore, it is still necessary to develop an explicit model for overcoming this problem *(Aytek and Kisi, 2008)*. From this point of view, genetic programming (GP), which is an evolutionary computing method that provides transparent and structured system identification, has been developed *(Savic et al, 1999)*.

Genetic programming has been successfully applied to problems that are complex and nonlinear and where the size, shape, and overall form of the solutions are not explicitly known in advance *(Whigham and Crapper, 2001)*. It also partially alleviates the problem necessary for conceptual model calibration. The state-of-the art applications of the GP in civil engineering have been listed by *Shaw et al, (2004)*. Also, some aspects of GP in hydraulic and sedimentation engineering were mentioned by *Babovic et al. (2001)* and *Aytek and Kisi (2008)* and respectively.

In spite of suitable flexibility of the ANN and GP methods for oceanographic time-series modelling, sometimes there is a limitation when signal fluctuations are highly non-stationary and physical oceanographic processes operate under a large range of scales varying from one day to several decades. In such a situation, ANN and GP approaches may not be able to cope with non-stationary data if pre-processing of input and/or output data is not performed *(Cannas et al. 2006)*. To cope with this problem, the wavelet technique is widely applied to time-series analysis of non-stationary signals *(Nason and Von Sachs, 1999; Labat 2005; Nourani et al. 2009a, 2011)*.

In a recent research, hybrid wavelet neural network model was used to improve forecasting accuracy of time series significant wave height by *Prahlada R. (2011)*. In an even more recent research, a hybrid wavelet-GP approach was used to optimize ANN modeling of Rainfall-runoff Process. The GP component (like ANN) of the model can handle the nonlinearity and non-stationary elements in mean and variance trends, while the wavelet component can deal with seasonal (cyclic) non-stationary elements of the process.

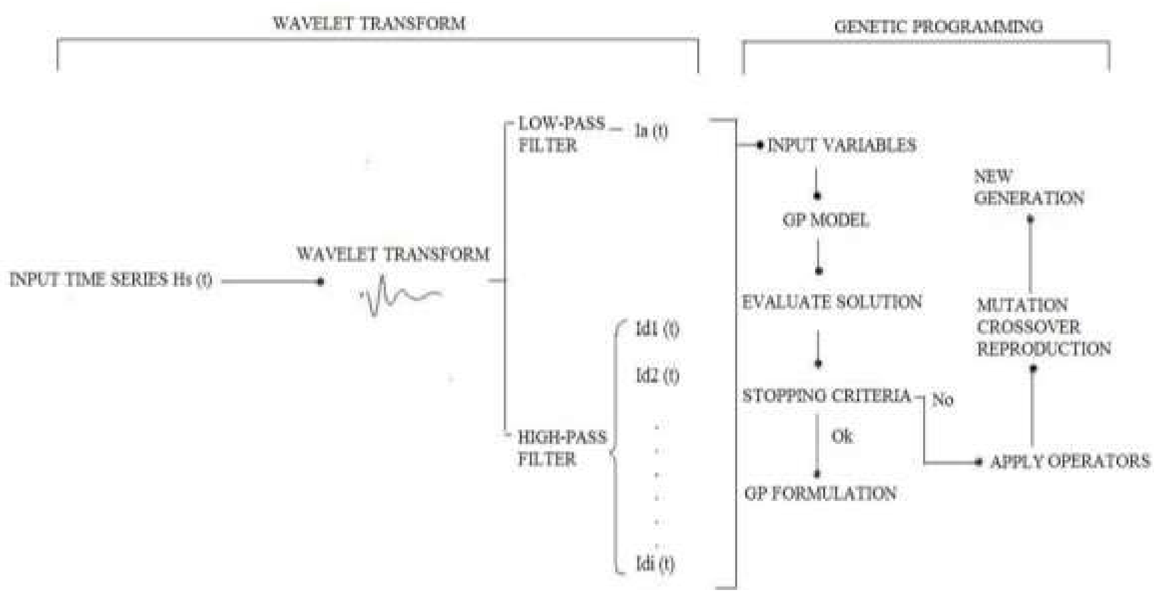
A wavelet transformation is a signal processing tool like Fourier transformation with the ability of analyzing both stationary as well as non stationary data, and to produce both time and frequency information with a higher resolution, which is not available from the traditional transformation. The wavelet transform breaks the signal into its wavelets which are scaled and shifted versions of the original wavelet so called mother or father wavelet. So, it can be seen that the results might get enhanced and more accurate if used in forecasting waves. The present study is carried out with an aim to come up with a best hybrid model for time series predictions for different lead times with the following objectives.

1. To develop a hybrid model which is able to handle both nonlinearity and non-stationary's present in the time series data.

2. To check the performance of hybrid model for different mother wavelets.
3. To investigate the influence of different decomposition levels for 3hr lead time on the model performance.
4. To optimize best possible methodology in hybrid wavelet-GP model for improving the model performance.

II. DATA COLLECTION AND METHODOLOGY

For the present study the data collected from one station was be used. The data used in the current study is processed significant wave height (Hs) of the station SW4 (Latitude 12°56’31” and longitude 74°43’58”) located near west coast of India which was collected from New Mangalore Port Trust (NMPT) during the year 2003 from January 1st to December 31st. The frequency of the data was 3 hourly significant wave heights. The wavelet-genetic programming (WLGP) model combines the strengths of discrete wavelet transform and GP processing to achieve powerful nonlinear approximation ability.



In the WLGP model, the decomposed time series (details and approximation) play various roles in the original time series and the behavior of each is distinct, so the contribution to the original time series varies from each other. The periodicity property of wave heights phenomenon is considered in this structure because the decomposed time series by wavelet analysis produces detailed information about the data trend and periodicity. In the model, in addition to the periodicity effect which is represented by the detailed sub signals, the autoregressive time series property may also be taken into account by considering the original time series as an input to GP model.

III. RESULTS AND DISCUSSION

The parameters for GP such population size, number of generations, mutation frequency and crossover frequency’s are decided based upon previous studies *(Gaur and Deo, 2008).*

Population size – 500

No. of generations – 300

Mutation Frequency – 90%

Crossover Frequency – 50%

% Data Distribution	R^2	BIAS	SI
40-20-40	0.463	1.10	0.281
40-30-30	0.800	0.98	0.201
50-25-25	0.749	0.998	0.143
60-20-20	0.857	0.999	0.121
70-15-15	**0.96**	**1.00005**	**0.038**
80-10-10	0.96	1.00007	0.041

All the plots of the data individually processed through GP are provided in the tabular column above

MOTHER WAVELET	R^2	**BIAS**	**SI**
Haar	0.758	1.002	0.091
Sym_2	0.762	1.006	0.083
$Bior_{1.1}$	0.759	1.008	0.081
$Coif_1$	0.761	1.01	0.090
Db_2	0.762	1.010	0.092
Db_3	**0.923**	**1.0001**	**0.029**
Db_4	0.763	1.009	0.045

All the results for the decomposition level 6 are tabulated above

MOTHER WAVELET	R^2	**BIAS**	**SI**
Haar	0.9226	1.00023	0.034
Sym_2	0.855	1.011	0.056
$Bior_{1.1}$	0.904	1.008	0.037
$Coif_1$	0.854	1.009	0.067
Db_2	0.868	1.010	0.078
Db_3	**0.960**	**1.00005**	**0.024**
Db_4	0.758	1.002	0.076

All the results for the decomposition level 7 are tabulated above

IV. CONCLUSION

The present study analysis and results revealed the following conclusions.

1. The proposed hybrid WLGP model outperformed single GP model for a 3hr lead time prediction.
2. Single GP model had certain issues with selection of distribution of data into training, validation and testing data. It was observed that at 70-15-15 ratio and 80-10-10 ratio, the model performed better when compared to other ratio which had less training data and comparatively more testing data like 40-20-40 or 40-30-30.
3. The statistical variations of the data given from SW4 station were observed to be low to medium. This might be because of the fact that the training data length in relative comparison with the normal data size was less. A 1 hourly data of the same period would have fetched slightly better results.
4. The improvement of results in WLGP model is due to dividing the dataset into multifrequency bands using DWT to make data as a stationary data. GP is good at handling non-stationary data, but it has shown excellence in handling stationary data and hence the proposed model performed very well.
5. It was noticed that as the decomposition level for different wavelets was increased, the performance of the hybrid WLGP model also increased. This enhancement was minute for some wavelets but was noticeable for all the different wavelets. Level 7 of decomposition had less error than level 5 and 6.
6. Selection of proper mother wavelet was also carried out in the present study. Db2 wavelet performed better at decomposition level 5 suggesting that at lower decomposition level it can perform better. Db3 wavelets were best at decomposition level 6 and 7. This may suggest that as the decomposition level increases, if the order of the wavelet increases, the model performs better. But the same was not true in case of Db4 wavelet since it had more error than Db3 wavelet at all the three decomposition levels.

REFERENCES

[1]. Aussem, A., Murtagh, F., (1997). Combing neural network forecasts on wavelet transformed time series. Connect Sci 10(1): 113-121.

[2]. Aytek, A., Kisi, O., (2008). A genetic programming approach to suspended sediment modelling. J. Hydrol. 351, 288-298.

[3]. Babovic, V., Kanizares, R., Jenson, H. R., Klinting, A., (2001). Neural networks as routine for error updating of numerical models. J. Hydrol. Eng. 127(3), 181-193.

[4]. Banzhaf, W., Nordin, P., Keller, R. E., and Francone, F. D. (1998), Genetic Programming: An Introduction: On the Automatic Evolution of Computer Programs and its Applications, Morgan Kaufmann.

[5]. Barricelli, Nils Aall (1954), Esempi numerici di processi di evoluzione, Methodos, pp. 45–68.

[6]. Brameier, M. and Banzhaf, W. (2007), Linear Genetic Programming, Springer, New York.

[7]. Cannas, B., Fanni, A., See. L., Sias, G., (2006). Data preprocessing for river flow forecasting using neural network: wavelet transforms and data partitioning. Phys Chem Earth 31(18): 1164-1171.

[8]. Chen, B. F., Wang, H. D., Chu, C. C., (2007). Wavelet and artificial neural network analysis of tide forecasting and supplement of tides around Taiwan and South China Sea.Ocean Engineering 34, 2161-2175.

[9]. Crosby, Jack, L. (1973), Computer Simulation in Genetics, John Wiley & Sons, London.

[10]. Cramer, Nichael Lynn (1985), "A representation for the Adaptive Generation of Simple Sequential Programs" in Proceedings of an International Conference on Genetic Algorithms and the Applications, Grefenstette, John J. (ed.), Carnegie Mellon University

[11]. Deka, P. C., Mandal, S., Prahlada, R., (2010). Multiresolution Wavelet-ANN model for significant wave height forecasting. Proceedings of National conference on hydraulics and water resources, HYDRO-2010, Dec 16-18th, 230-235.

[12]. Deka, P. C., Prahlada, R., (2012). Discrete Wavelet neural network approach in significant wave height forecasting for multistep lead time. Ocean Engineering 43, 32-42.

[13]. Deo, M. C., Naidu, C. S., (1999). Real time wave forecasting using neural network. Ocean Engineering 35, 191-203.

[14]. Falco, D., Cioppa, A. D., and Tarantino, E., (2005). A genetic programming system for time series prediction and its application to El Nino forecast. Adv. Soft Comput. 32(2005), 151-162.

[15]. Fogel, David B. (2000) Evolutionary Computation: Towards a New Philosophy of Machine Intelligence IEEE Press, New York.

[16]. Fogel, David B. (editor) (1998) Evolutionary Computation: The Fossil Record, IEEE Press, New York

[17]. Gaur, S., Deo, M. C., (2008). Real time wave forecasting using genetic programming. Ocean Engineering 35, 1166-1172.

[18]. Huang. M. C., (2002). Wavelet parameters and functions in wavelet analysis. Science Direct, Ocean engineering 31 (2004). 111-115.

[19]. Jain, P., and Deo, M. C., (2008), Artificial Intelligence Tools to Forecast Ocean Waves in Real Time. The Open Ocean Engineering Journal, 2008, 1, 13-20.

[20]. Kim, T., Valdes, J. B., (2003). Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. ASCE journal of hydrology engineering 8 (6), 319-328.

[21]. Kinsman, B., (1965). Wind waves. Prentice Hall, Englewood Cliffs, NJ Kisi. O., (2009). Neural Networks and wavelet conjunction model for intermittent stream flow forecasting. ASCE journal of hydrology 14, 773-782.

[22]. Kisi O., Shiri J. and Nazemi A. H., (2011). Wavelet-Genetic programming model for predicting short-term and long-term air temperatures.Journal. Civil Eng. Urban. 1(1):25-37.

[23]. Labat, D., (2005). Recent advances in wavelet analysis, Part-I: A review of concepts. J Hydrol 314 (1-4): 275-288.

[24]. Lee, H. S., and Kwon, S. H., (2003). Wave profile measurement by wavelet transforms. Science Direct, Ocean Engineering 30 (2003).2313-2328.

[25]. Legates, D. R., and McCabe, G. J. Jr., (1999). Evaluating the use of goodness-of-fit measures in hydrologic and hydro climatic model validation. Water resource Research. 35(1), 233-241.

[26]. Mallet, S., (1998). A wavelet tour of signal processing. 2nd Ed., Academic Press, SanDiego, CA.

[27]. Massel, S. R., (2001). Wavelet analysis for processing of ocean surface wave records. Ocean Engineering 28, 957-987.

[28]. Nason. G. P., and Von Sachs. R., (1999). Wavelets in time series analysis. Philes. Trans. R. Soc., 357(1760), 2511-2526.

[29]. Nourani, V., Komasi, M., Mano, A., (2009). A multivariate ANN-Wavelet Approach for Rainfall-runoff modeling. Water resource Manage 23:2877-2894. Doi 10.1007/s11269-009-9414-5.

[30]. Nourani, V., Komasi, M., Kisi, O., (2011). Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. Journal of hydrology 402, 41-59.

[31]. Nourani, V., Komasi, M., Mohammad Taghi Alami, (2012). Hybrid Wavelet-Genetic Programming approach to optimize ANN modeling of rainfall-runoff process. Journal of Hydrologic Engineering. (2012).17:724-741.

[32]. Ozger, M., (2010). Significant wave height forecasting using wavelet fuzzy logic approach. Ocean Engineering 37, 1443-1451.

[33]. Savic, A. D., Walters, A. G., Davidson, J. W., (1999). A genetic programming approach to rainfall-runoff modelling. Water Resour. Manage 13, 219-231.

[34]. Shaw. D., Miles. J., and Gray. A., (2004). Genetic programming within civil engineering. Organization of the adaptive computing in design and manufacture Conf. IEEE, New York.

[35]. Sivanandam. S. N., and Deepa. S. N., (2008). Introduction to genetic algorithms. Springer. Berlin, Heidelberg, Germany.

[36]. Shore Protection Manual. Coastal Engineering Research Centre, US Army Corps of Engineers, Washington, DC.

[37]. Suttasupa, Y., Rungraungsilp, S., Pinyopan, S., Wungchusunti, P., and Chongstitvatana, P., (2011). A Comparative Study of Linear Encoding in Genetic Programming, Ninth International Conference on ICT and Knowledge. ISBN 978-1 4577-2162-5/11.

[38]. Wang, W. and Ding. J., (2003). Wavelet network model and its application to the prediction of hydrology. Nature and science, 1(1).

[39]. Whigham, P. A., Crapper, P. F., (2001). Modeling rainfall-runoff using genetic programming. Mathematical and Computer Modeling 33(6-7): 707-721.

[40]. Zhou, H. C., Peng Y., and Liang, G. H., (2008). The research of monthly discharge Predictor-Corrector model based on wavelet decomposition. Water resource management (2008). 22:217-227.