Ocean Systems Engineering, *Vol. 7, No. 4 (2017) 319-328*
DOI: https:// doi.org/10.12989/ose.2017.7.4.319

Potential of regression models in projecting sea level variability due to climate change at Haldia Port, India

Thendiyath Roshni*, Md. Sajid K.[a] and Pijush Samui[b]

Department of Civil Engineering, National Institute of Technology Patna, Ashok Rajpath, Bihar, India-800005

(Received August 18, 2017, Revised October 22, 2017, Accepted November 9, 2017)

Abstract. Higher prediction efficacy is a very challenging task in any field of engineering. Due to global warming, there is a considerable increase in the global sea level. Through this work, an attempt has been made to find the sea level variability due to climate change impact at Haldia Port, India. Different statistical downscaling techniques are available and through this paper authors are intending to compare and illustrate the performances of three regression models. The models: Wavelet Neural Network (WNN), Minimax Probability Machine Regression (MPMR), Feed-Forward Neural Network (FFNN) are used for projecting the sea level variability due to climate change at Haldia Port, India. Model performance indices like PI, RMSE, NSE, MAPE, RSR etc were evaluated to get a clear picture on the model accuracy. All the indices are pointing towards the outperformance of WNN in projecting the sea level variability. The findings suggest a strong recommendation for ensembled models especially wavelet decomposed neural network to improve projecting efficiency in any time series modeling.

Keywords: WNN; MPMR; FFNN; sea-level; statistical downscaling

1. Introduction

Efficacy in the prediction of sea level variability are of prime importance to coastal and ocean engineers, as this will be one of the deciding factors for the off-shore structures, groundwater quality, saline water intrusion etc. As mentioned by the experts, sea level variations enact a great role in the groundwater levels of low lying coastal aquifers (Meyer 1981), coastal rivers (Thain *et al*. 2004) etc. Moreover, our fresh water resources are diminishing as reported by IPCC in the 5th report, and are severely affected by the sea level rise. Sea level rise could trigger the saline intrusion in the coastal aquifers and thereby it affects the groundwater quality, aquatic plants and animals (Chen *et al*. 2016). The climate change can be a dominant concern for the sea level variability and a few literatures (Goharnejad *et al*. 2013, Masciopinto and Liso 2016) remarks about the contribution on this. Worldwide investigations established a global temperature increase of both land and sea by 0.76℃ from 1850-2005 (IPCC, 2007). As noted by Pfeffer *et al*. 2008, the global sea level rise will happen in between 0.18 m to 0.4 m towards the end of 21st

*Corresponding author, Assistant Professor, E-mail: roshni@nitp.ac.in
[a] Master Student, E-mail: mdsajidk1992@gmail.com
[b] Associate Professor, E-mail: pijush@nitp.ac.in

http://www.techno-press.org/?journal=ose&subpage=7 ISSN: 2093-6702 (Print), 2093-677X (Online)

century. Hence this study focuses on sea level rise in regional scale and the region selected is Haldia Port, located in West Bengal, India.

Now the question arises, which all applications can be effectively employed in time series data like sea level. Artificial Intelligence (AI) techniques have been demonstrated successfully in different fields of water resources engineering (Coulibaly *et al*. 2001, Nourani and Mousavi 2016) especially in the downscaling techniques. Hence it would be a valid reason for selecting ANN techniques for statistical downscaling technique in Climate change analysis. It is apparent from the literature review that a combination of AI techniques exploits the strengths of each technique and improves the combined model efficiency (Han *et al*. 2007). Such a recent and widely used model is Wavelet decomposed Neural Network (WNN). Hence this study focuses on comparison of the conventional Feedforward Neural Network and a hybrid Wavelet decomposed Neural Network.

These techniques provide an effective tool in timeseries modeling problems. A novel approach developed by Strohmann and Grudic in 2002 for time series modeling is Minimax Probability Machine Regression (MPMR) algorithm. MPMR is a non-linear regressor and the distinctive feature is that it maximizes the minimum probability of the objective function to be within the limits of the true regression (Strohmann and Grudic 2002). Potential utility of the MPMR model in different applications are cited here. MPM is used for both classification and regression analysis and the applications were restricted to image retrieval, face or pattern recognition etc. A few literature work employs the comparative performance of MPMR regression models with other regression techniques for prediction of seismic ultrasonic attenuation (Kumar *et al*. 2013), forecasting evaporative loss (Deo and Samui 2017) etc. The results presented MPMRs as a convincing alternative to conventional ANN. Study of MPMR further shows promising results for further applications.

Considering the above literature work, a superior performance of MPMR than ANN is visible for the regression analysis in different fields of engineering. Hence in this paper, authors attempted a primal application of MPMR in statistical downscaling technique to project the sea level variability at Haldia Port, West Bengal, India. Authors also compared the performances of MPMR with WNN and FFNN models. The results are compared and analyzed with different statistical metrics to find the potential capability of the developed models in time series modeling.

2. Study area and data acquisition

For the analysis of the sea level variability, Haldia Port of West Bengal, India is selected (Fig. 1). The average elevation of the Haldia Port is 8 m from Mean Sea Level (MSL). Haldia Port plays an important role for the development of the Indian economy and hence many industries are established in this area. It is just near to the mouth of the Hooghly river. The climate in the Haldia region is typical moderate. Generally, in the winter season, temperature varies from 7℃ minimum to 22℃ maximum. But in the Summer season, environment of this area becomes very hot and humid. Observed field data of monthly average mean sea level at Haldia Port is collected from Permanent Service for Mean Sea Level (http://www.psmsl.org/) site for twenty-seven years (1985 to 2012). Input data (1985-2012) for training and testing phase were selected from NCEP/ NCAR Reanalysis Project for the selected area. Input data for the projecting period (2013-2049) are obtained from CMIP5 project (Coupled Model International Project 5): Model MPI-ESM-MR.

The range of data sets used for the present study are shown in Table 1. All the data were normalized between 0 to 1 for model development (Sahoo and Jha 2013).

Table 1 Range of data-set used for model development

	Sea Level (m)	Air Temperature (°K)	Geopotential Height (m)	Relative Humidity (%)	Specific Humidity
Minimum	6.63	261.34	9515.51	34.71	0.005
Maximum	7.73	272.23	9754.65	95.60	0.017
Std. Deviation	0.21	2.72	61.72	19.13	0.004
Average	7.08	267.64	9660.25	68.22	0.011

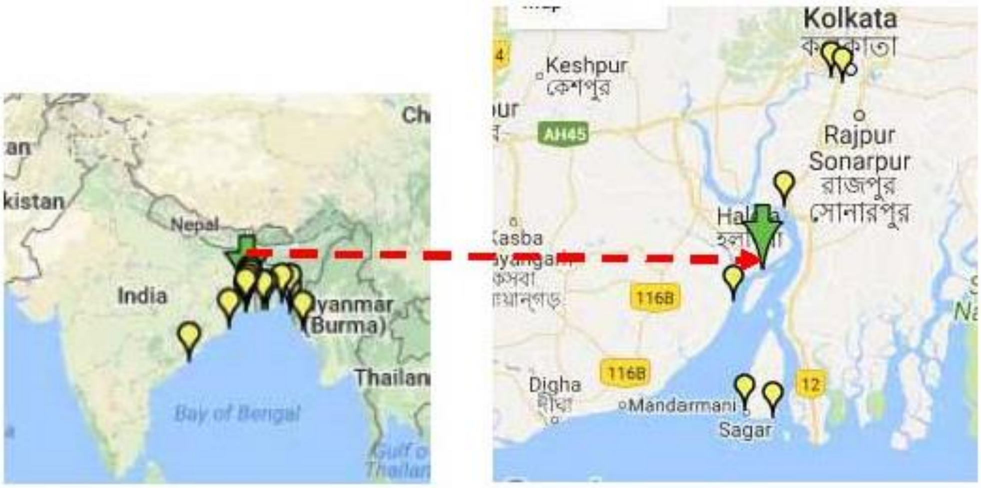


Fig. 1 Location of the study area (www.psmsl.org)

3. Methodology

3.1 Downscaling of data

Downscaling of data is to extract the information from known large scale area to make a prediction to a small scale of our desired area. Statistical downscaling is performed in this study. The two steps involved in statistical downscaling are: to find a relationship between local climate variables and large-scale predictors and secondly to apply such relationships for finding output of global climate model experiments to simulate local climate characteristics in the future. All predictors from NCEP/NCAR reanalysis data may not be potential inputs for the model and hence correlation studies are carried out to find the potential inputs. Correlation studies are done by R-studio software. Different techniques are involved in the statistical downscaling of the data and is reviewed in Section 3.2.

3.2 Overview of techniques used

3.2.1 Feedforward Neural Network (FFNN)

Artificial Neural Network mimics the human brain and it learns the behavior of the system from the data (Hornik *et al*. 1989). It acts as a powerful computational model. Previous literature work found the superior nature of FFNN with Back Propagation algorithm. FFNN training involves two phases: training and testing phase. Correspondingly, whole data is divided in to two 70% and 30% for training and testing. The basic concept involved in FFNN is the adjustment of weight and reduction of errors between the given output and target given by the network. If the errors computed are higher than the estimated value, the network adjusts itself its weight during training phase. This process continues till the error generated between the output and the target is reduced to a desirable limit. The three layers: input layer, hidden layer and output layer are the common structure of the basic NN. The input nodes in the layer are selected based on correlation analysis and is done by R studio software. Root Mean Square Error (RMSE) is used to check the model accuracy for different hidden nodes in the hidden layer. Transfer function used in hidden layer and the output layer is sigmoid function.

3.2.2 Wavelet decomposed Neural Network (WNN)

Wavelet is a class of function that is used to centralize a given function in position and frequency. These functions are used in signal processing and in time series analysis. In recent literatures, wavelet decomposition has used in several fields of hydrology. Wavelet decomposition happens in different steps: splitting of signal by passing in to low-pass filter and high-pass filter and appears two signals, approximate (A) and detail (D). D and A are called as wavelet coefficients. ‘Approximates’ are low frequency and high scale component signal and ‘Details’ are high frequency and low scale component signal. In different levels of decompositions, the wavelet coefficients are further allowed to pass through low pass and high pass filters. The level of decomposition (L) depends on the length of data (N) and is shown in Eq. (1).

$$L = \mathrm{int}\left[\log(N)\right] \tag{1}$$

Once decomposition is over, the decomposed signals are fed as inputs to Feedforward Neural Network. Then the process is repeated as mentioned in section 3.2.1.

3.2.3 Minimax Probability Machine Regression (MPMR)

Minimax Probability Machine analysis is mainly done for classification and regression. In this study regression application of MPM is utilized effectively. MPMR is found as an effective alternative and improvised form of Support Vector Machine (Deo and Samui 2017). Previous work of MPMR reveals its effective application in other fields of engineering like prediction of seismic ultrasonic attenuation (Kumar *et al*. 2013), prediction of fast fading channel (Yang *et al*. 2010) forecasting evaporative loss (Deo and Samui 2017) and a few more.

MPMR follows a regression model for y as

$$y = \left[\sum_{i=1}^{N} \beta_i \, K(x_i, x) + b\right] \pm \varepsilon \tag{2}$$

Where, x and y denotes the inputs and corresponding outputs, $K(x_i, x)$ is kernel function. In this study, Radial Basis Function (RBF) is used as the kernel function. The other variables β, b are outputs obtained by MPMR algorithm and ε shows the limits of error fluctuations. The proposed MPMR algorithm is established in MATLAB 2010.

4. Performance indicators

The goodness-of-fit statistics considered in this study are: Coefficient of Correlation (*R*), maximum determination coefficient value (R^2), Adjusted determination coefficient (*Adj*. R^2), Root Mean Square Error (*RMSE*), Performance Index (*PI*), Nash–Sutcliffe model Efficiency (*NSE*), Root Mean Square Error to observation's standard deviation ratio (*RSR*), Normalized Mean Bias Error (*NMBE*), mean absolute percentage error (*MAPE*) and Variance Account Factor (*VAF*). In addition, a visual comparison of the observed and the developed models, Probability Distribution and Cumulative Distribution plots are analyzed.

$$R = \frac{\sum(h_{oi} \cdot h_{ci})}{\sqrt{\sum h_{oi}^2 \sum h_{ci}^2}} \tag{3}$$

$$R^2 = \frac{\sum_{i=1}^{N}(h_{oi} - \overline{h_{oi}})^2 - \sum_{i=0}^{N}(h_{oi} - h_{ci})^2}{\sum_{i=1}^{N}(h_{oi} - \overline{h_{oi}})^2} \tag{4}$$

$$Adj \cdot R^2 = 1 - \frac{(N-1)}{N-P-1} \cdot \left(1 - R^2\right) \tag{5}$$

$$RMSE = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}(h_{oi} - h_{ci})^2\right)} \tag{6}$$

$$PI = Adj \cdot R^2 + 0.01 \cdot VAF - RMSE \tag{7}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}((h_{oi} - h_{ci})^2}{\sum_{i=1}^{N}((h_{oi} - \overline{h_{oi}})^2} \tag{8}$$

$$RSR = \frac{RMSE}{\frac{1}{N}\sum_{i=1}^{N}((h_{oi} - \overline{h_{oi}})^2} \tag{9}$$

$$NMBE = \frac{\frac{1}{N}\sum_{i=1}^{N}(h_{ci} - h_{oi})^2}{\frac{1}{N}\sum_{i=1}^{N}(h_{oi})} \cdot 100 \tag{10}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|h_{oi} - h_{ci}|}{h_{oi}} \cdot 100 \tag{11}$$

$$VAF = 1 - \frac{VAR(h_{oi} - h_{ci})}{VAR(h_{oi})} \cdot 100 \tag{12}$$

Where, h_{oi} = observed sea level and h_{ci} = calculated sea level, and N = number of observations, P is the input quantity, $\overline{h_{oi}}$ and $\overline{h_{ci}}$ are the mean of the observed values and the calculated values. *RMSE* indicates the discrepancy between the observed and calculated values; the lower the *RMSE* shows higher prediction accuracy. Coefficient of Correlation (R) values near to 1 shows the accurate prediction. *MAPE* is the mean absolute percentage error and it compares the residual error between the calculated or predicted with that of the observed value. Smaller values indicate better performance of the model. *NMBE* shows the potential of the model to predict the deviation from the mean value. Positive values of *NMBE* indicate over-prediction and negative values of *NMBE* show under-prediction of the model. *VAF* represents the ratio of the error variance to the observed data variance and is expressed in percentage. *RSR* is the ratio of *RMSE* to observation's standard deviation ratio. *RSR* adds the benefits of error index statistics. R^2 and *Adj* R^2 evaluate the linear regression between the observed and calculated date, while Nash–Sutcliffe model efficiency evaluates the predictive capability of the developed models. Performance Index (*PI*) examines the accuracy of the developed models. *PI* of a model greater than 1 indicates greater efficiency. In theory, for an ideal model, *RMSE*, and *RSR* should be 0, 100% for *VAF*, 1 for R, R^2, *Adj* R^2 and *NSE* (Chandwani *et al*. 2015, Ceryan 2014). A combination of all these statistical indicators provide an unbiased estimate for prediction ability of the developed regression models.

5. Results and discussions

5.1 Selection of potential predictors

The correlation analysis was carried out by R studio software for all the predictor variables. Of these, the correlation coefficient greater than 0.85 area considered as potential predictors. The potential predictors found by the analysis are Air temperature at 500 mb pressure, Geopotential Height at 500 mb pressure, Specific Humidity at 925 mb pressure and Relative Humidity at 1000 mb pressure and is shown in Fig. 2. The figure reveals a good correlation coefficient with that of sea level data with correlation coefficients more than 0.85.

5.2 Skills of the developed models

To evaluate the model performances of the FFNN, WNN and MPMR models, data were normalized before each simulation. For FFNN model development a conventional three-layered structure is formed with input layer, hidden layer and output layer. The four inputs and one output constitute the number of input nodes and output node. Hidden nodes in the hidden layer are found by trial and procedure and the optimum number was fixed by the criteria of *RMSE*. The hidden layer with 35 nodes satisfied the *RMSE* criteria. The sigmoid function was performed as transfer functions in both layers. In WNN, the inputs are decomposed by passing through high pass filter and low pass filter. The Daubechies-10 mother wavelet (db 10) is selected (Seo *et al*. 2015) and the level of decomposition is found as 2 (Nourani *et al*. 2009) for wavelet decomposition. The decomposed inputs are fed in to the FFNN. Instead of 4 inputs in FFNN, there will be 12 sets of inputs in hybrid WNN. In MPMR application, Radial Basis Function is used as kernel function.

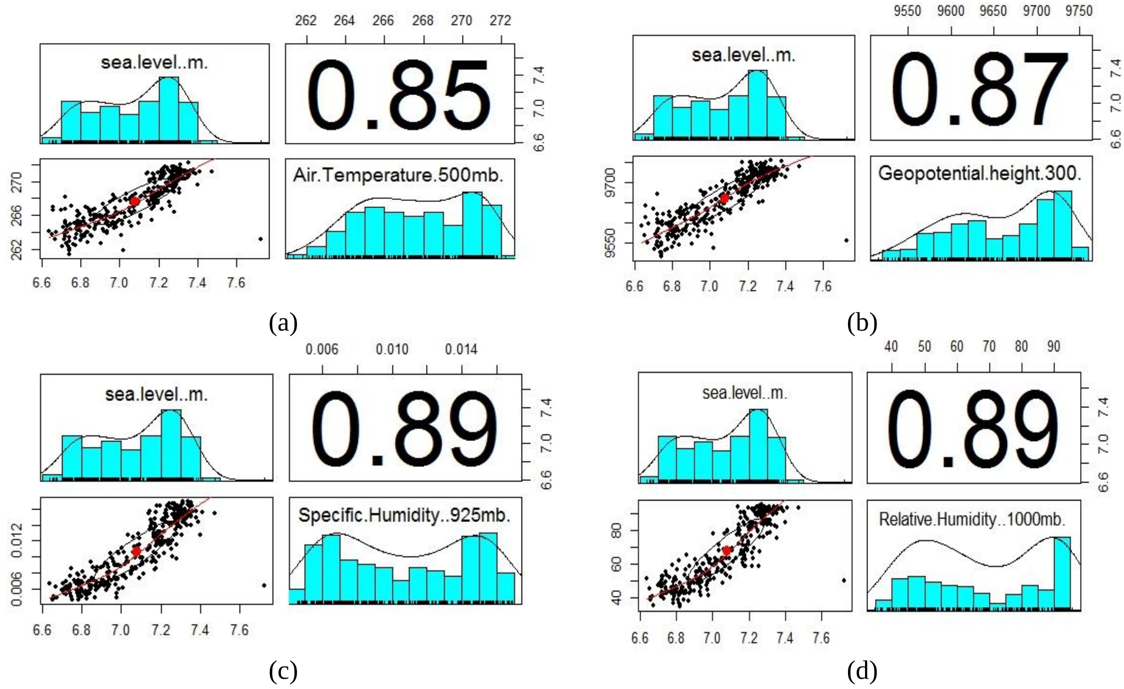


Fig. 2 Correlation Coefficients for sea level against (a) Air Temperature (500 mb), (b) Geopotential Height (300 mb), (c) Specific Humidity (925 mb) and (d) Relative Humidity (1000 mb)

For comparison, only testing period data were considered. A visual comparison, probability distribution function (pdf) and cumulative distribution function (cdf) plots for the models developed are also plotted during the testing period (Figs. 3 and 4). As in the visual comparison (Fig. 3), it is not evident to speak about the better performances. All the three models show a similar sinusoidal pattern as that of the observed sea level during the testing period. The accordant results can also be seen from in Fig. 4(a). It is seen in Fig. 4(b) that Cumulative Distribution Function curves show a flat S-curve, that indicates a lesser over prediction (positive values) and under prediction (negative values) for all the developed models. The results were better analyzed by the performance indicator measures for three developed models (Table 2). The indicators selected were *R*, R^2, *Adj. R*, *RMSE*, *PI*, *NSE*, *RSR*, *NMBE*, *MAPE* and *VAF*. From the model performances for the testing period, the best results were generated for WNN model compared to MPMR and FFNN. Almost similar statistical values were observed for MPMR and FFNN. *PI* values were found in the range of 1.75, 1.68 and 1.61 for WNN, FFNN and MPMR. NMBE shows that WNN was over-predicted and MPMR and FFNN were under predicted during testing period. VAF shows 90.75% for WNN compared to 86.37% and 84.32% for FFNN and MPMR. From these analyses, a slight outperformance of WNN was found. The greater performance of WNN can be ascribed to the decomposition of the input signals, which helps in capturing the non-linear dynamics of the processes.

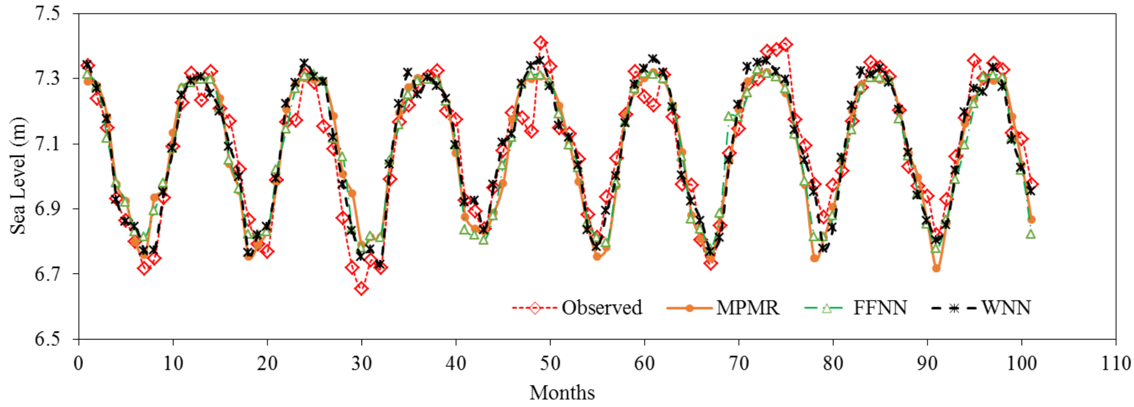
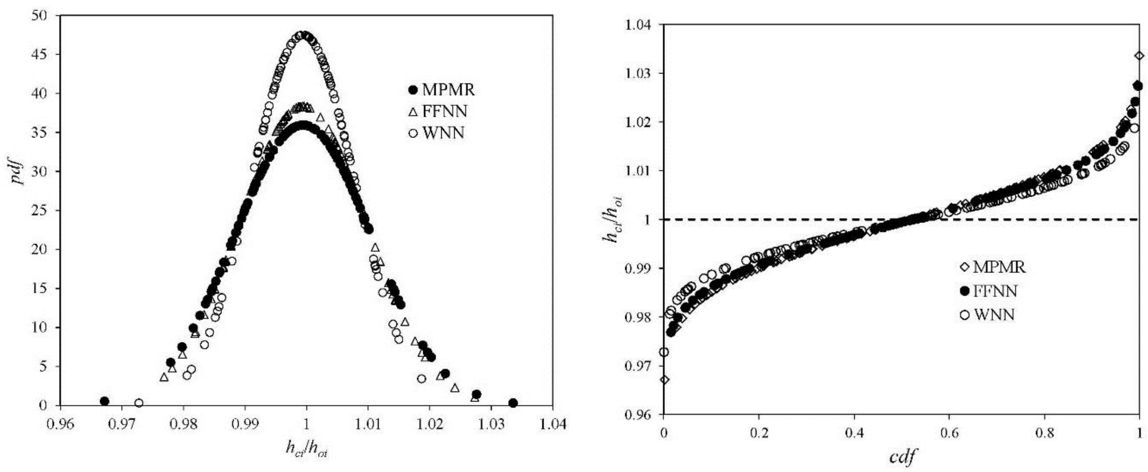


Fig. 3 Visual Comparison of MPMR, WNN and FNN model results with Observed data in testing period (2004-2012)



(a) Probability distribution function (pdf) plot (b) Cumulative distribution function (cdf) plot

Fig. 4 Pdf and Cdf plots of Sea Level Variability for FFNN, WNN and MPMR models for testing period (2004-2012)

Table 2 Performance indicators for MPMR, FFNN and WNN Models

	R	R^2	*Adj* R^2	*RMSE (m)*	*PI*	*NSE*	*RSR*	*NMBE (%)*	*MAPE (%)*	*VAF (%)*
MPMR	0.9217	0.8495	0.8432	0.0776	1.6089	0.8420	0.3975	-0.0631	0.8711	84.3188
FFNN	0.9296	0.8641	0.8584	0.0728	1.6493	0.8518	0.3849	-0.0723	0.8419	86.3753
WNN	0.9543	0.9108	0.9070	0.0595	1.7550	0.9084	0.3027	0.0590	0.6637	90.7455

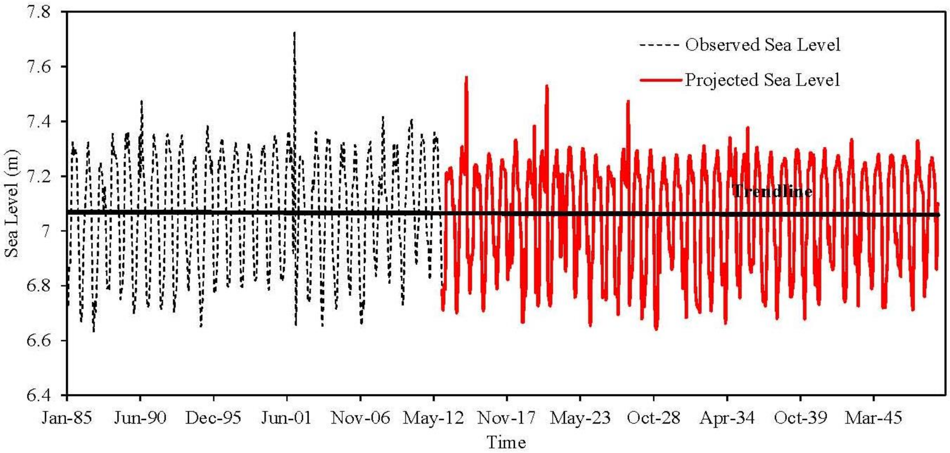


Fig. 5 Time series plot of Observed Sea Level and Future Sea Level (WNN)

With the best model developed (WNN), the sea level variability for the future period is found and plotted in Fig. 5. A trend analysis plot was also made in the plot and found that there is a slight decreasing trend is observed for the entire period selected. On January 2002, a Tsunami of 7.725 m magnitude was noted. A separate analysis of projected sea level did not show any significant trend.

6. Conclusions

Regression models were presented and illustrated in prediction of sea level variability due to climate change at Haldia Port, India. Three models: WNN, FNN and MPMR were developed and evaluated in terms of model accuracy by several statistical indices. All the evaluation metrics showed a unanimous agreement for the outperformance of WNN model. Concurrently, MPMR and FNN models also showed its potential for non-linear mapping of the hydrological processes. The outperformance of the WNN model might be attributed to the multi-level decomposition of the input signals and this decomposition aggravates the effectiveness of the model diagnosis. Hence it is suggested that ensemble model approach might be adopted for enhancing forecasting/ projection accuracies as it takes the strengths of constituent networks.

References

Ceryan, N. (2014), "Application of support vector machines and relevance vector machines in predicting uniaxial compressive strength of volcanic rocks", *J. African Earth Sci.*, **100**, 634-644.

Chandwani, V., Agarwal, V. and Nagar, R, (2012), "Modeling slump of ready mix concrete using genetic algorithms assisted training of Artificial Neural Networks", *Earth Syst. Appl.*, **42**, 885-893.

Chen, W., Chen, K., Kuang, C., David, Z., Zhu, D.Z., He, L., Mao, X., Liang, H. and Song, H. (2016), "Influence of sea level rise on saline water intrusion in the Yangtze River Estuary, China", *Appl. Ocean Res*., **54**, 12-25.

Coulibaly, P., Anctil, F., Aravena, R. and Bobée, B. (2001), "Artificial neural network modeling of water table depth fluctuations", *Water Resour. Res*., **37**, 885-896.

Deo, R.C. and Samui, P. (2017), "Forecasting evaporative loss by least-square support-vector regression and evaluation with genetic programming, Gaussian process, and minimax probability machine regression: Case study of brisbane", *J. Hydrol. Eng*., **22**(6), 1-15.

Goharnejad, H., Shamsai, A. and Hosseini, S.A. (2013), "Vulnerability assessment of southern coastal areas of Iran to sea level rise: evaluation of climate change impact", *Oceanologia*, **55**(3), 611-637.

Han, D., Kwong, T. and Li, S. (2007), "Uncertainties in real-time flood forecasting with neural networks", *Hydrol. Process.*, **21**(2), 223-228.

Hornik, K., Stinchcombe, M. and White, H. (1989), "Multilayer feedforward networks are universal approximators", *Neural Networks*, **2**(5), 359-366.

IPCC. (2007), "Climate change 2007: The scientific basis", (Eds., Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. and Miller, H.L.), Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom/New York City, United States.

Kumar, M., Mittal, M. and Samui, P. (2013), "Performance assessment of genetic programming and minimax probability machine regression for prediction of seismic ultrasonic attenuation", *Earthq. Sci*., **26**(2), 147-150.

Masciopinto, C. and Liso, I.S. (2016), "Assessment of the impact of sea-level rise due to climate change on coastal groundwater discharge", *Sci. Total Environ.*, 672-680.

Nourani, V. and Mousavi, S. (2016), "Spatiotemporal groundwater level modeling using hybrid artificial intelligence-meshless method", *J . Hydrol*., **536**, 10-25.

Nourani, V., Komasi, M. and Mano, A. (2009)," A multivariate ANN-wavelet approach for rainfall–runoff modeling", *Water Resour. Manag*., **23**, 2877-2894.

Sahoo, S. and Jha, M.K. (2013), "Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment", *Hydrogeol. J*., **21**, 1865-1887.

Seo, Y., Kim, S., Kisi, O. and Singh, V.P. (2015), "Daily water level forecasting using wavelet decomposition and artificial intelligence techniques", *J. Hydrol*., **520**, 224-243.

Strohmann, T. and Grudic, G.Z. (2002), "A formulation for minimax probability machine regression", Proc., Advances in Neural Information Processing System, MIT Press, Cambridge, MA, 769-776.

Yang, L., Wang, L., Sun, Y. and Zhang, R. (2010), "Simultaneous feature selection and classification via Minimax Probability Machine". *Int. J. Comput. Intell. Syst*., **3**(6), 754-760.

MK