Indian Journal of Geo Marine Sciences
Vol. 49 (08), August 2020, pp. 1350-1358

Improving numerical current prediction with Model Tree

S Dauji*[,a] & M C Deo[b]

[a]Bhabha Atomic Research Centre, Anushaktinagar, Mumbai, India & Homi Bhabha National Institute, Anushaktinagar, Mumbai – 400 094, India
[b]Civil Engineering Department, Indian Institute of Technology Bombay, Mumbai – 400 094, India

*[E-mail: dauji_saha@yahoo.com; acad.dauji@gmail.com]

Received 25 March 2019; revised 24 September 2019

A method to improve the real time predictions of ocean currents on the basis of a machine learning technique called model tree is proposed. It consists of forming an error time series obtained as the difference between the numerical prediction and the actual measurement of the current at a given time step, carrying out time series prediction as per the technique of model tree and predicting the error for a future time step. Subtraction of such error from the numerically predicted current produces the improved current magnitude for the next time step. The suggested procedure is applied at two deepwater locations in the Indian Ocean. The numerical current model under investigation is code named: HYCOM, while corresponding current observations are those coming from a measurement program called: RAMA. It was found that such method of error subtraction yielded more accurate predictions than those based only on the numerical modelling. This is judged from analysing certain error statistics as well as by comparison with the random walk time series prediction method. The predictions up to five days in advance are satisfactorily done in this manner.

[**Keywords**: Current observations, Current prediction, Model tree, Numerical ocean model, Ocean currents]

Introduction

The operational or short-term prediction of oceans currents is routinely provided to the user community on the basis of ocean circulation models that solve the governing hydrodynamic equations using a suitable numerical scheme. Such predictions are used for carrying out different activities such as shipping and towing, tracking of oil spill or other type of pollution, monitoring of coastal water quality, warnings for beach activities like sports and swimming and also for fishing expeditions. There are various reasons due to which numerical predictions may not be accurate enough at a specific site of interest. These include uncertainties in the forcing functions and boundary conditions, accuracy of bathymetry and related input, simplifying assumptions made in the governing equations. In order to reduce the effect of such uncertainties the model parameters are required to be tuned beforehand or data assimilation is carried out in which the real time measurements are assimilated into the model using certain algorithms and the model is re-run[1]. However, it is likely that as an outcome the resulting computations put heavy demand on time and resources.

In India, the agency: Indian National Centre for Ocean Information Services (INCOIS)[2], located at Hyderabad, routinely predicts the ocean currents on operational basis. The numerical simulation results of currents are obtained from a typical 3-dimensional global ocean circulation model: HYCOM[3] (Hybrid Coordinate Ocean Model) that provides current predictions at short intervals into the future. There are a few previous works in which these model predictions were post-processed using typically the technique of artificial neural network (ANN)[4,5]. The success of the ANN for this application has prompted the authors to use another well-established machine learning approach in the form of model tree. Model tree solves the problem with a tree-like representation wherein the problem domain gets split into many sub-domains and linear regression of the historical data gives the solution within each sub-domain. When a new query (input) is posed appropriate model is picked up and the output is produced using it. It is thus regarded as piece-wise linear model.

The potential of the tool of model tree (MT) as a data driven tool has not been fully utilised in engineering applications and hence its past applications are limited in hydrology[6-8], particularly in ocean engineering[9]. Garg *et al.*[9] had earlier used model tree to predict ocean currents in tide-dominated areas like gulfs and creeks in time series prediction mode. The authors concluded that the MT was faster, by orders of magnitude, compared to other soft computing tools such as ANN and genetic programming and yet produced comparable forecasts.

The aim of this work is therefore to improve on the numerical current predictions done at daily intervals by INCOIS by post-processing such predictions with MT as was earlier performed with ANN[4,5] by the authors.

Materials and Methods

Model trees

The model tree, also known as decision tree or regression tree, essentially is an extension of classification procedure, wherein the computational process follows a tree structure[10,11]. Starting from a root node or a decision box, the model tree branches out, based on the decision at each node, onto numerous other nodes and leaves. The entire input or parameter domain gets partitioned into sub-domains based on certain criteria and for each of these sub-domains, multiple linear regression models are developed on the basis of historical data. Thus, the model tree approximates the non-linear relationship between the dependent and independent variables, with piecewise linear models. In our case the dependent variable is the error in predicted current (from numerical model) at a subsequent time step while the independent variable/s are a selected segment of (say 5, 6, …) preceding errors in current values. There are various algorithms for development of model tress, which vary in their domain splitting criterion. The criteria employed for demarcation of sub-domains may be minimization of the entropy in sub-domains, or collecting as many samples as possible belonging to the same class or any other. M5 algorithm of Quinlan[12] is one very popular and widely applied algorithm, in which the domain splitting is executed with the help of standard deviation of the class value reaching a particular node[6]. The error at a particular node is represented by the standard deviation at that node. The criterion that maximizes such error reduction at a node is chosen for splitting the domain at the node.

The development of a model tree and the domain splitting process is depicted in Figure 1. This is called the calibration or training process. For easy graphical representation, 2-D input domain with parameters x_1 and x_2 has been chosen. The splitting criterion in M5 algorithm is based on the difference between the standard deviation of the total training set and the weighted standard deviation of any i^{th} sub-domain. Typical mathematical representation would be:

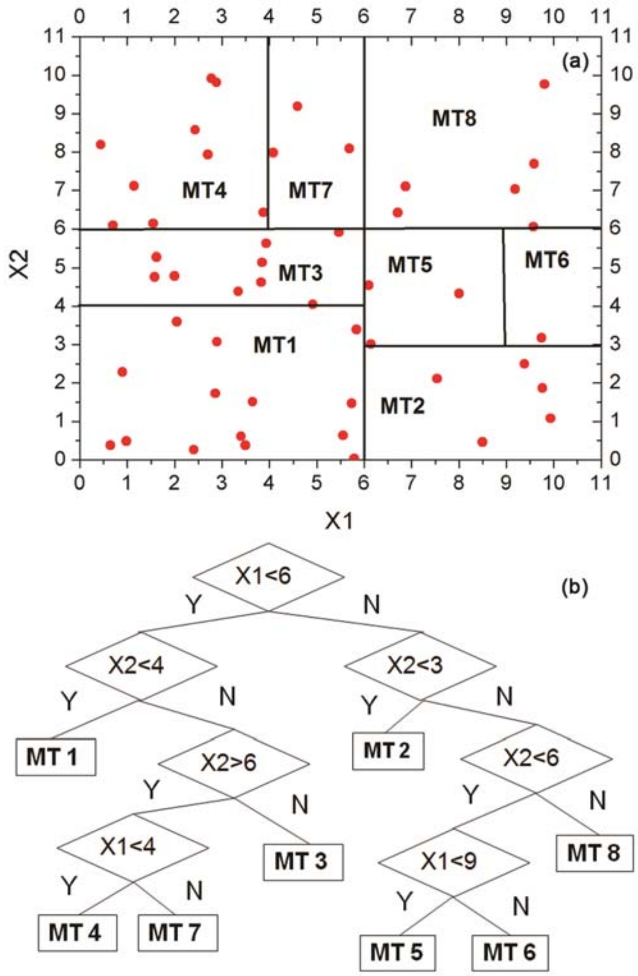

Fig. 1 — Schematic depiction of a model tree: (a) Domain splitting in model tree; (b) Decision tree corresponding to domain splitting in part (a), in which Y = yes and N = no

$$\sigma_R = \sigma(N) - \sum_{i=1}^{N} \frac{N_i}{N} \cdot \sigma(N_i) \qquad \ldots (1)$$

Where, σ_R : standard deviation reduction;

σ (N) : standard deviation of all training samples (N numbers);

N_i = i^{th} sub-domain of N,

σ (N_i) = standard deviation of the i^{th} sub-domain.

During the calibration or training process, many possible input division options are explored and the one that results in the maximum value of the standard deviation reduction is selected for domain division. Subsequently, linear models are built for each sub domain thus generated. The typical linear model could be:

$$o = a_o + a_1x_1 + a_2x_2 \quad \ldots (2)$$

Where, o: output variable;

a_0, a_1, a_2, …: coefficients derived using, say, a least squares fit;

x_1, x_2, ……: input variables.

Now, the termination of the process of domain splitting might be done when the class values of all data in a sub-domain do not vary much, say by 5 %, or when very few data remains.

For initiation of the domain sub-division, an attribute, say, $x_1 < 6$, (Fig. 1b) is taken as the root node and two branches separate out as per 'Yes' or 'No' to the criterion ($x_1 < 6$) and such processes continues thereafter as depicted in Figure 1 (a & b). As mentioned earlier, the selection of the attribute as the root node or other decision nodes, is performed such that the standard deviation reduction is maximized. Certain measures to reduce the number of splits and to avoid large discontinuities between neighboring models might be essential for development of efficient model trees, as discussed in literature[10]. After completion of the domain splitting process, the result is the collection of leaf nodes, or the various sub-domains, for each of which linear regression models are developed, as indicated by boxes: MT1 to MT8 in Figure 1(b).

There are many applications of MT in forecasting and prediction problems in the domain of hydraulics and water resources reported in literature. These include prediction of flood discharge in a river based on antecedent discharges with one day lead time, rainfall and evaporation[6], flow prediction based on past flows and rainfall records, with lead times of a few hours[13], temporal and spatial prediction of ground water levels in Delft[14], and time series forecasting of low flows[15] among others.

The numerical model

The operational prediction model: HYCOM (Hybrid Coordinate Ocean Model) solves the governing differential equations based on a variant of finite difference formulations. The various governing equations are obtained from the principles of continuity, momentum and advection-diffusion[16]. In HYCOM, the ocean–atmosphere exchange of momentum is also incorporated. The Arakawa-C type of grids is adopted for the model and it uses the relaxation or sponge and open boundary conditions. At INCOIS, this numerical model: HYCOM was set-up for the Indian Ocean between longitudes 20° E to 125° E and latitudes 35° S to 31° N. The open boundary conditions at the eastern, southern and south-eastern edges are derived from global HYCOM simulations. Twenty-eight vertical layers are used in the model and the horizontal resolution at mid-latitudes is around 25 km. Data assimilation was not performed while running the model presently. In this study, the daily current predictions so made at two deepwater sites in Indian Ocean: Site 1: with coordinates of 8° N and 90° E (~ 3000 m) and Site 2: with coordinates of 12° S and 80.5° E (~5000 m) have been used. The time periods for which the records were available for this work were different for the two sites, being 30 months (November 4, 2009–April 1, 2012) for Site 1 and 24 months (May 18, 2010–Apr. 2, 2012) for Site 2, respectively. The locations for the study are depicted in Figure 2: Site 1 – T1 and Site 2 – T2.

The current observations

The daily current measurements at the same two sites belonged to those under the Project: RAMA (Research Moored Array for African-Asian Australian Monsoon Analysis and Prediction)[17]. The current observations were for 10 m depth and were recorded using a Sontek Current Meter. The resolution of the current meter for speed was 0.1 cm/s and for direction, the resolution was 0.1°. The range of speed measured by the current meter was 0–600 cm/s and it had an accuracy of + 5 cm/s for speed and + 5° for direction[18]. These observations were extracted for the same time period as of the above numerical data. The current measurements as well as the numerical current data were earlier collected by the Indian National

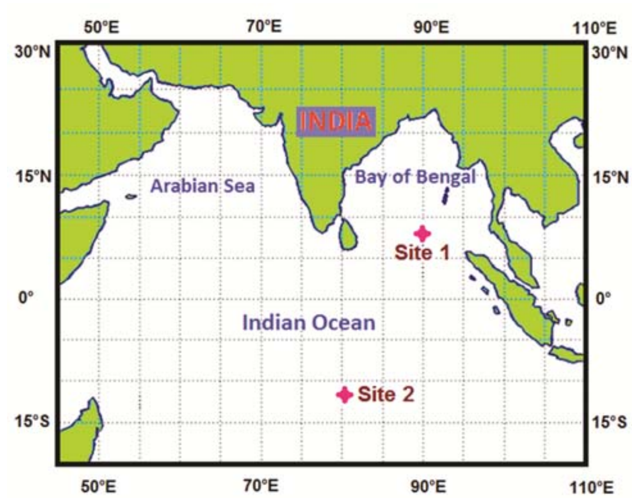

Fig. 2 — Map showing the Study Locations: Site 1 and Site 2

Centre for Ocean Information Service (INCOIS), Hyderabad.

Performance of the numerical model

The performance of the numerical model for the two sites for the entire period of observation was first evaluated by comparing with the current observations at the same time step. This is presented in Table 1 from where it can be noted that the R value for both components of current, namely, “u” or zonal or east-west and “v” or meridional or south-north is low and the RMSE and MAE are high, indicating unsatisfactory performance of the numerical model at both Site 1 and Site 2. This can be understood considering that they are targeted more towards spatial information rather than location-specific one.

It was therefore decided to work for the improvement of the prediction accuracy using the data driven technique of model tree. This tool was selected considering that unlike other machine learning tools such as neural network and genetic algorithm the solutions here are transparent and portable at other sites.

Methodology

For this purpose, first an error time series was constructed in which every term (for each day) represented the difference between the numerical current prediction and concurrent observation. This represented the error time series, based on which the correction to the numerical current prediction at the subsequent time step (one or more days ahead) was predicted. While doing so, the previous pattern in the occurrence of currents is recognized and the next time step value is predicted in continuation and by sliding the same pattern forward by that time step. It is to be noted that this particular process of time series forecasting would be equivalent to the one based on the causal relationship, as per Takens’ theorem[19] and hence no input, other than the previous current values (numerical model prediction and observations), is necessary in the prediction exercise. The error time series was thus modelled with MT[20] with the help of a running sequence of preceding values whose number was fixed by trials (in this case up to 5 preceding values for different lead times) and the error for the next time step (varying from 1 to 5 days ahead) was predicted. For training or calibrating the model tree (MT) the initial 70 % of the data (or training data) was employed and subsequently, the performance of the model thus developed was evaluated with the latter 30 % of the data (or testing data). The predicted error was added to the numerical model evaluation at that current time step used to arrive at the updated prediction for the desired time step. The performance evaluation was performed with various quantitative error metrics, namely, correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE). Qualitative indication of the performance of the developed models was provided by scatter plots and time history plots.

For each location (Site 1 and Site 2), and for each lead time (1, 2, 3, 4, and 5 days), separate MT was developed with the initial 70 % of the data to predict the error at the desired time step (1, 2, 3, 4, and 5 steps ahead). The error prediction from the MT was added to the numerical model prediction to arrive at the combined numerical-MT prediction. The combined numerical-MT model employed in this study is schematically represented in Figure 3.

Table 1 — Performance of the numerical model for Site 1 (8° N 90° E) and Site 2 (12° S 80.5° E); Observation period: 30 months (November 4, 2009–April 1, 2012) at Site 1 and 24 months (May 18, 2010–Apr. 2, 2012) at Site 2

Site	U-velocity			V-velocity		
	R	RMSE (cm/sec)	MAE (cm/sec)	R	RMSE (cm/sec)	MAE (cm/sec)
1	0.47	23.21	17.94	0.09	22.62	18.18
2	0.25	20.91	15.29	0.04	16.42	10.69

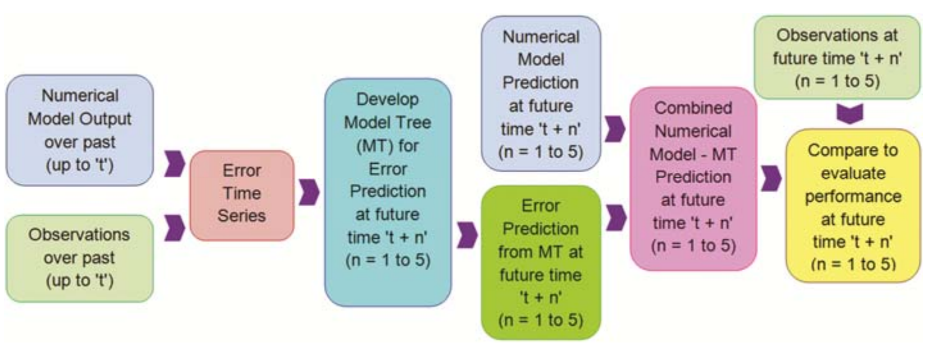

Fig. 3 — Schematic methodology for development and evaluation of combined numerical-MT model for current prediction

Subsequently, the developed MT was employed to predict the error for the remaining 30 % of the data designated as testing data and the combined numerical-MT predictions were obtained for the testing data. The performance evaluation was done for the testing data: by comparing the observations from RAMA buoys and the output of the combined numerical-MT model for the respective lead time. The results are discussed in the following section.

Results and Discussion

Combined Numerical-MT model: Site 1

The comparison of the current predictions from the combined numerical-MT model with corresponding actual observations at Site 1 is given in Figure 4 for the u-velocity component and in Figure 5 for the v-velocity component. The figure represents the results obtained for the lead time of 1 day. The updated predictions are quite closer to the observations when compared to the numerical estimations. The time required for development of each MT (separate MT was developed for each lead time) was typically less than a minute, which is quite efficient for application.

A quantitative comparison of the current predictions from the combined numerical-MT model with corresponding actual observations at Site 1 is given in Table 2 (indicated as N-MT in Table 2). The results for the testing period and up to the prediction horizon of 5 days are indicated. The performance of the estimates of

Table 2 — Performance of the combined numerical-MT model and numerical-RW model: Site 1; Testing period

Lead Time (days)	Tool	U-velocity			V-velocity		
		R	RMSE (cm/sec)	MAE (cm/sec)	R	RMSE (cm/sec)	MAE (cm/sec)
	N	**0.50**	**25.28**	**19.00**	**0.22**	**22.98**	**18.99**
1	N-MT	0.74	15.57	12.22	0.66	15.94	12.33
	N-RW	*0.63*	*20.97*	*16.67*	*0.61*	*20.51*	*16.21*
2	N-MT	0.66	18.55	15.09	0.48	20.16	15.63
	N-RW	*0.56*	*23.98*	*18.75*	*0.44*	*24.70*	*19.54*
3	N-MT	0.69	17.28	13.93	0.50	19.12	14.80
	N-RW	*0.60*	*21.95*	*17.56*	*0.46*	*24.49*	*19.50*
4	N-MT	0.62	18.94	15.21	0.42	22.41	17.93
	N-RW	*0.51*	*24.05*	*18.98*	*0.45*	*27.78*	*22.28*
5	N-MT	0.52	22.02	16.90	0.33	25.01	20.13
	N-RW	*0.44*	*27.19*	*21.87*	*0.34*	*30.07*	*23.73*

N: Numerical only; N-MT: Combined Numerical-Model Tree; N-RW: Combined Numerical-Random Walk

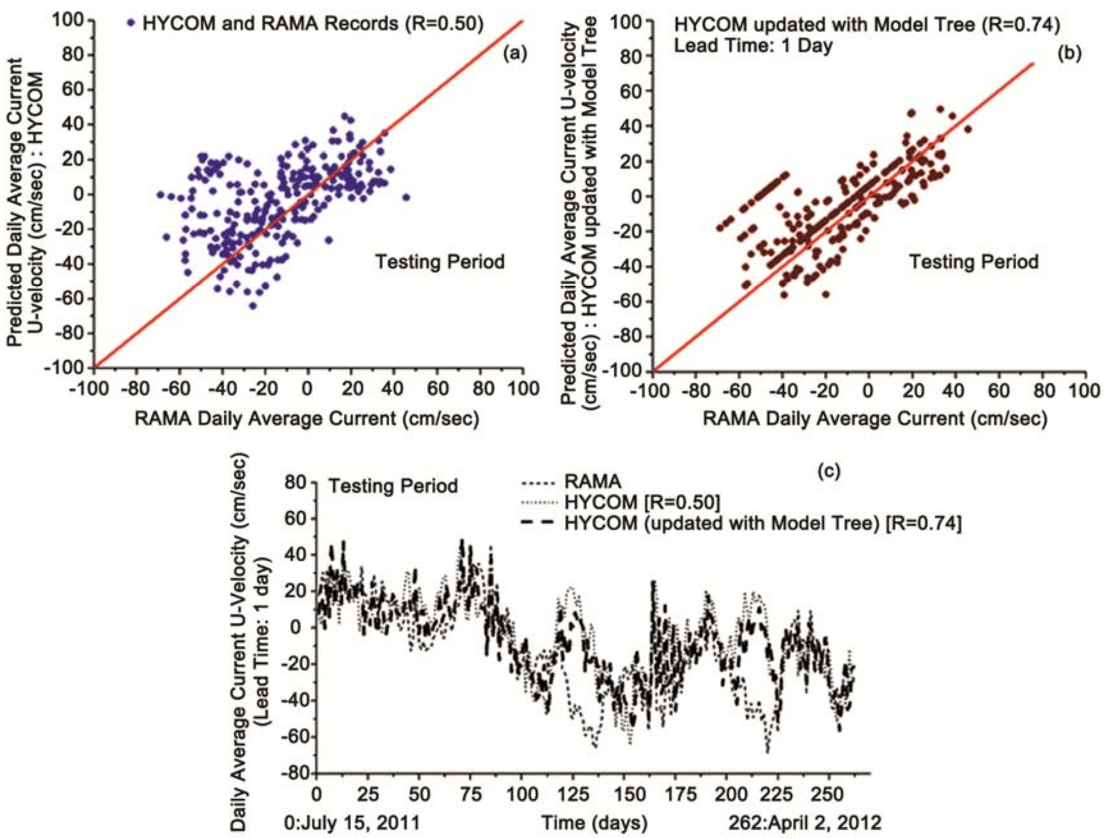

Fig. 4 — Scatter plot for Site 1 for testing period: u-velocity (a) Numerical versus observed at the same time step; (b) Combined numerical-MT versus observed - lead time: 1 day; (c) Time history plot

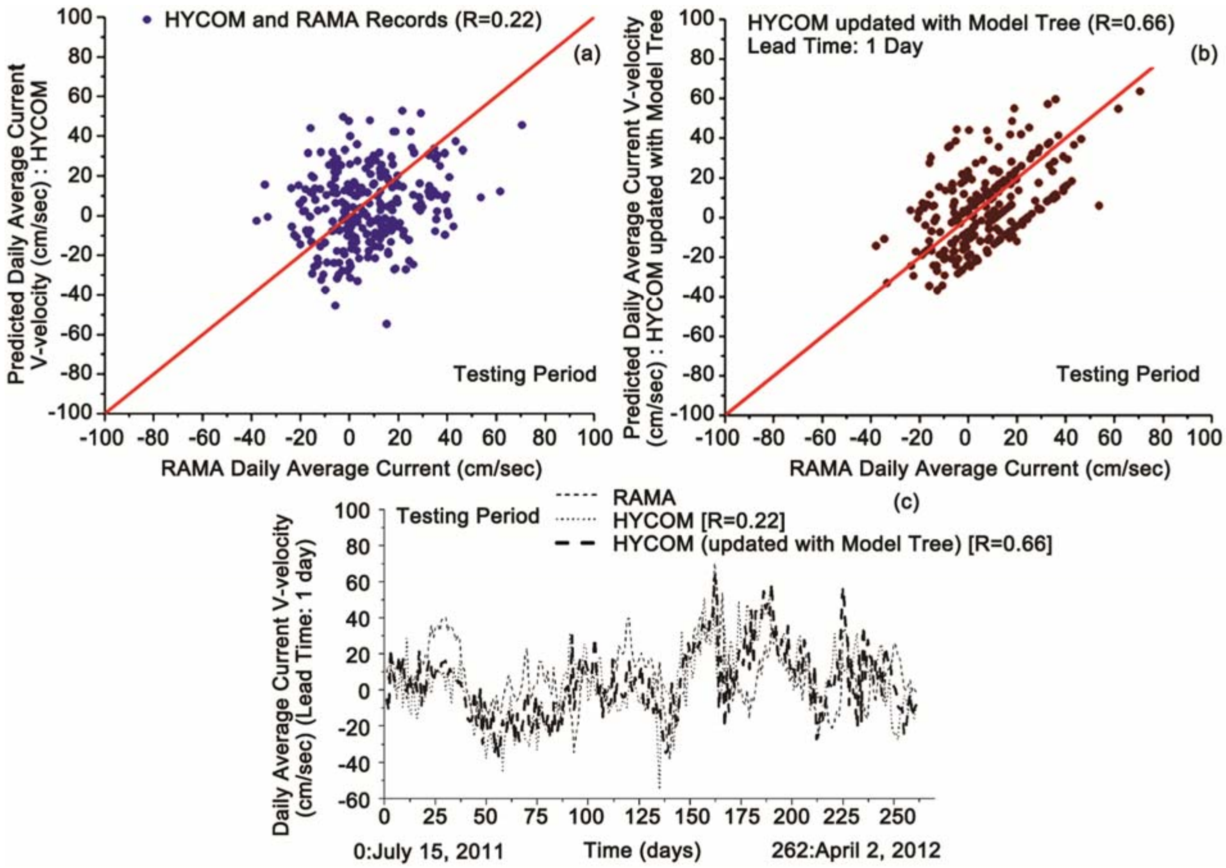

Fig. 5 — Scatter plot for Site 1 for testing period: v-velocity (a) Numerical versus observed at the same time step; (b) Combined numerical-MT versus observed - lead time: 1 day; (c) Time history plot

the numerical model during the same time period is included for comparison in the table (in boldface). It may be noticed that the use of combined numerical-MT model is successful at this location since the R values are much higher and RMSE and MAE are lower than those from the numerical evaluations for lead time of five days. With increasing lead time beyond five days, the developed model tree was incapable of capturing the dependencies involved and the accuracy of combined numerical-MT model did not improve over the numerical model. The structure of the model tree was again limited by the available length of data. The results might be different when a longer dataset would be available for developing a more accurate model tree for this application.

We have also used an alternative error correction scheme in which the MT was replaced by the basic level model or random walk model. In the random walk model[21], a random error term is added to the observation at a certain time step to obtain the predicted value at the next time step. Mathematically,

$$Y_{t+1} = Y_t + \xi_{t+1} \qquad \dots (3)$$

Where, Y_{t+1} is the prediction for time (t+1), Y_t is the observation at time (t), and ξ_{t+1} is the zero-mean stationary random process (the standard normal variate).

As can be understood, the model represented by Eq. 3 can be directly applied only for prediction performed for unit time step ahead, which in this case, is one day. It was intended to use this model for multiple day ahead predictions and hence, it was extended for n-day horizon as follows: the prediction at the $(t + n)^{th}$ time step was made equal to sum of the observation at the t^{th} time step and n random error terms, as was earlier done by the authors[22]. Mathematically, the n-day ahead prediction, at the $(t + n)^{th}$ time step, is represented as follows:

$$Y_{t+n} = Y_t + \sum_{i=1}^{n} \xi_{i+1} \qquad \dots (4)$$

The results of this exercise is included in Table 2 (indicated as N-RW in Table 2), where it can be seen that the random walk based error update provides an improvement over the numerical model but works less efficiently compared to the MT based scheme. Compared to the error updating scheme based on random walk model, as reported in Table 2(N-RW) for Site 1, the performance of the MT (Table 2, N-MT) are distinctly superior.

Combined Numerical-MT model: Site 2

Similar to Site 1, comparisons for Site 2 of the current predictions from the combined numerical-MT model with corresponding actual observations are depicted in Figure 6 for the u-velocity component and in Figure 7 for the v-velocity component. This particular result represents the current prediction for the lead time of 2 days. As can be observed from the figures, the updated predictions are closer to the observations when compared to the numerical estimations.

Like the earlier case of Site 1 a quantitative comparison of the current predictions from the combined numerical-MT model with corresponding actual observations at Site 2 is given in Table 3 (indicated as N-MT in Table 3). The results are shown

Table 3 — Performance of the combined numerical-MT model, numerical-RW model, and numerical-ANN model[5]: Site 2; Testing period

Lead Time (days)	Tool	U-velocity			V-velocity		
		R	RMSE (cm/sec)	MAE (cm/sec)	R	RMSE (cm/sec)	MAE (cm/sec)
	N	**0.75**	**15.86**	**12.67**	**-0.01**	**26.99**	**21.54**
1	N-MT	0.91	9.34	7.30	0.74	12.55	8.98
	N-RW	*0.72*	*19.11*	*16.04*	*0.73*	*13.27*	*10.72*
	N-ANN[5]	*0.93*	*8.00*	*6.30*	*0.77*	*11.73*	*8.21*
2	N-MT	0.92	8.32	6.69	0.62	15.01	11.27
	N-RW	*0.74*	*18.26*	*15.54*	*0.72*	*14.91*	*10.98*
	N-ANN[5]	*0.92*	*8.24*	*6.67*	*0.68*	*13.56*	*9.67*
3	N-MT	0.85	12.35	10.06	0.44	18.49	14.01
	N-RW	*0.71*	*19.61*	*16.11*	*0.61*	*17.58*	*14.02*
	N-ANN[5]	*0.88*	*10.40*	*8.21*	*0.59*	*16.41*	*11.87*
4	N-MT	0.84	13.06	10.77	0.34	19.66	14.49
	N-RW	*0.69*	*19.56*	*16.29*	*0.52*	*19.37*	*14.63*
	N-ANN[5]	*0.87*	*11.15*	*8.88*	*0.50*	*16.88*	*12.70*
5	N-MT	0.81	14.27	11.54	0.30	21.25	16.11
	N-RW	*0.64*	*21.85*	*17.77*	*0.49*	*20.74*	*16.20*
	N-ANN[5]	*0.85*	*12.34*	*9.86*	*0.37*	*19.63*	*14.63*

N: Numerical only; N-MT: Combined Numerical-Model Tree; N-RW: Combined Numerical-Random Walk; N-ANN[5]: Combined Numerical-Artificial Neural Network

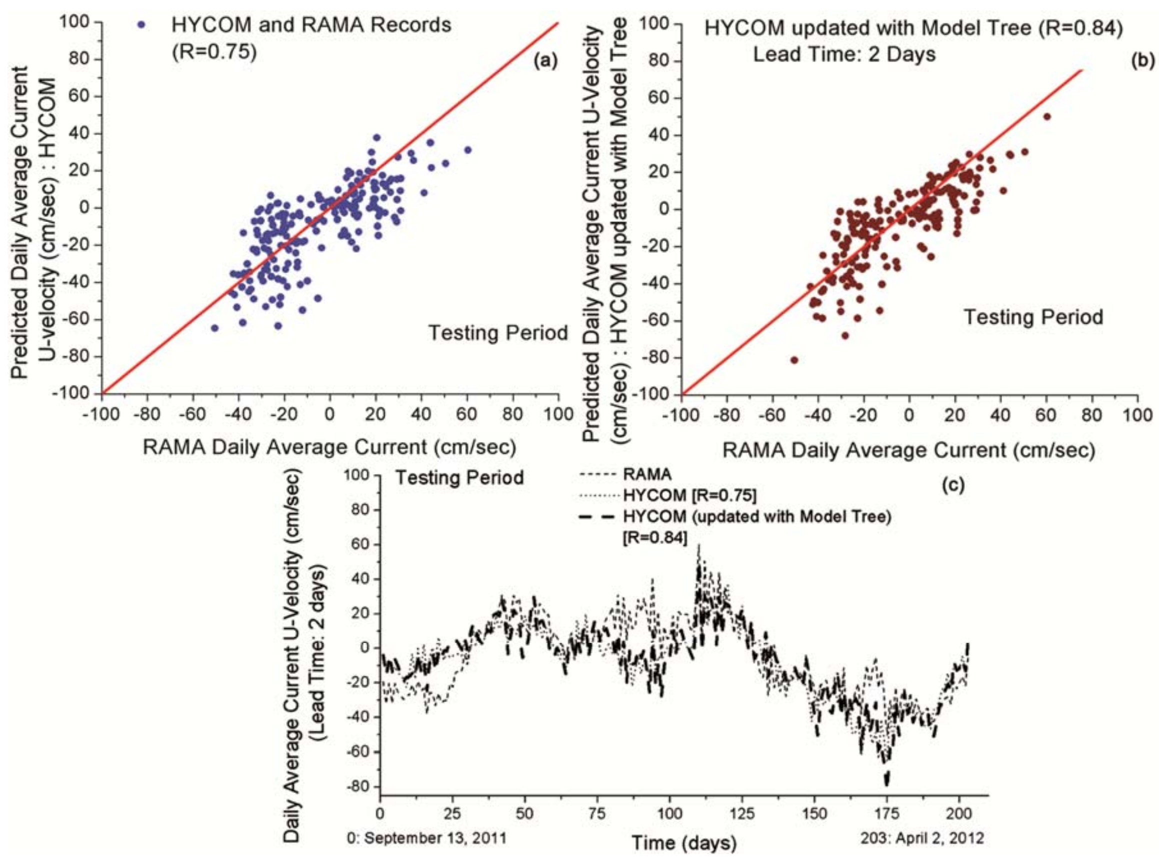

Fig. 6 — Scatter plot for Site 2 for testing period: u-velocity (a) Numerical versus observed at the same time step; (b) Combined numerical-MT versus observed - lead time: 2 days; (c) Time history plot

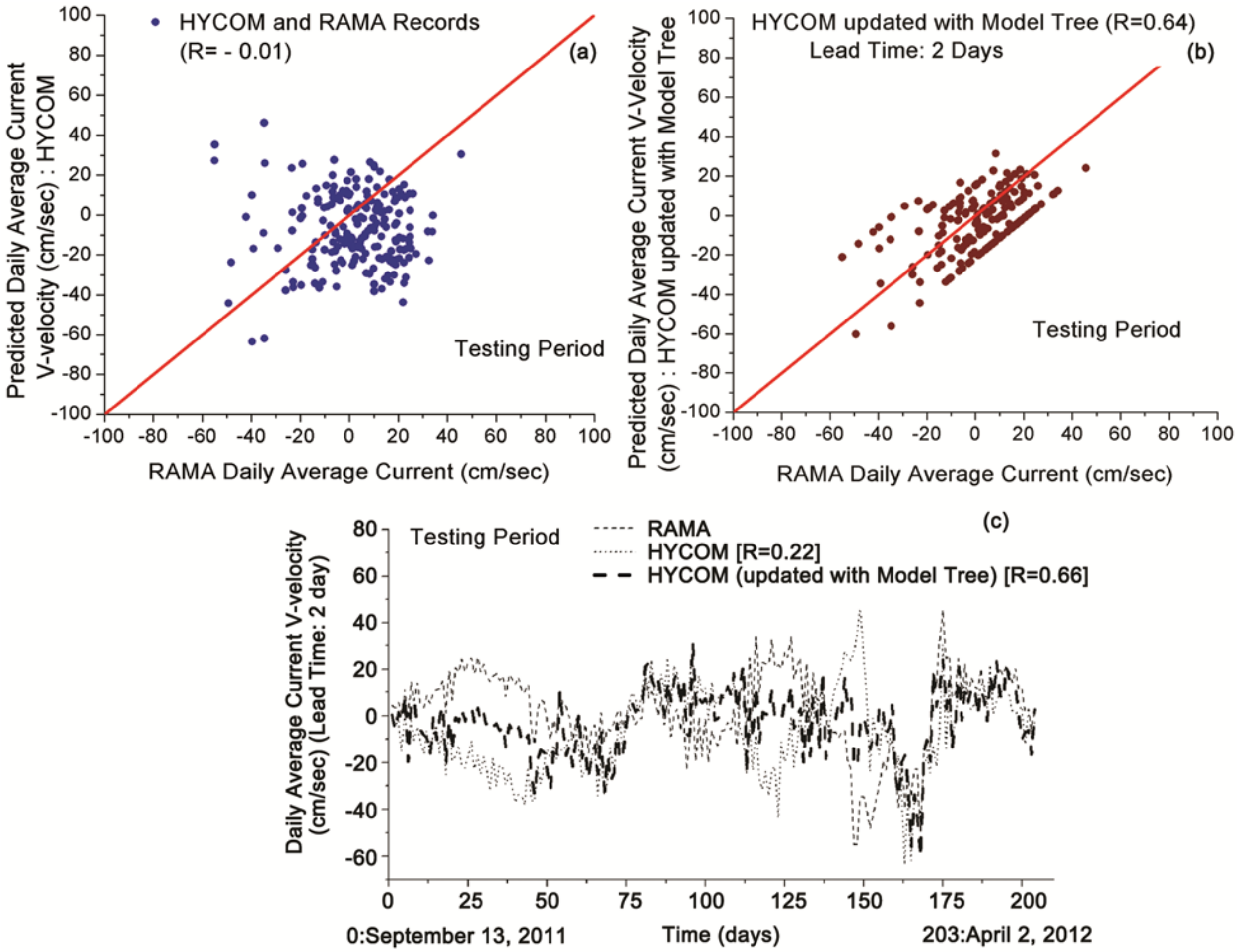

Fig. 7 — Scatter plot for Site 1 for testing period: v-velocity (a) Numerical versus observed at the same time step; (b) Combined numerical-MT versus observed - lead time: 2 days; (c) Time history plot

for the testing period and up to the prediction horizon of 5 days are indicated. In addition, performance metrics for the numerical model estimates are included in the Table in boldface, for the similar period of time. It may be noticed that the use of combined numerical-MT model is successful at this location since the R-values are much higher and RMSE and MAE are lower than those from the numerical evaluations.

Table 3 (indicated as N-RW in Table 3) also gives the comparison of the numerical-MT model with the baseline random walk model. It can be seen that the random walk based error update provides an improvement over the numerical model (Table 3, N-RW)but works less efficiently compared to the MT based scheme (Table 3, N-MT).

As a comparison of the effectiveness of MT for improving the currents obtained from numerical models, vis-à-vis that of ANN, the performance of combined numerical-ANN model for Site 2, as reported earlier by the authors[5], is also reproduced in Table 3 (indicated as N-ANN[5] in Table 3). It can be seen from the Table 3 that the MT performs reasonably well when compared to the ANN for this application.

Conclusions

The preceding sections presented a study in which numerically evaluated daily values of ocean surface current was made more accurate by combining it with the data driven technique of model tree. The case study was made at two deepwater locations in Indian Ocean. For both locations examined in this study, the accuracy of the numerical model current predictions was improved by subtracting the errors obtained from model tree. This was achieved up to five times steps (days) into the future. Both meridional and zonal components were improved, although the zonal component improvement was relatively less probably due to complex wind or tidal forcing from that direction. The accuracy of combined numerical-model tree based prediction was better than that of the combined numerical-random walk model indicating that the complexity of the former is necessary in this application.

Acknowledgements

The authors express their gratitude to Dr. Sudheer Joseph and Dr. M. Ravichandran, INCOIS, Hyderabad, India for sharing the numerical model results; and TAO Project Office of NOAA/PMEL for the buoy data, from Indian Ocean which taken from their website and used for this study. The authors are greatly indebted to the anonymous reviewers for their critical review comments and constructive suggestions for improving clarity of the manuscript.

Conflict of Interest

To the knowledge of authors, there is no competing or conflicting interest involved in the research presented in this article.

Author Contributions

Both authors contributed to design, analysis, data compilation, and, interpretation of results. Both authors read and approved the final manuscript.

References

1 Sivakumar B & Berndtsson R, *Advances in Data-based Approaches for Hydrologic Modeling and Forecasting*, (World Scientific, Singapore), 2010.
2 INCOIS, Numerical models for current forecast employed at INCOIS. http://www.incois.org version (08/2013).
3 HYCOM, Technical details of numerical model HYCOM.http://www.hycom.org, version (08/2013).
4 Dauji S, Deo M C, Joseph S & Bhargava K, Neural network based data assimilation to improve numerical current predictions, *Proc. of 11th International Conference in Hydrodynamics, HIC 2014*, (New York City, USA), 2014.
5 Saha D, Deo M C, Joseph S & Bhargava K, A combined numerical and neural technique for short term prediction of ocean currents in the Indian Ocean, *Environ Syst Res*, 5 (4) (2016) 1-14. doi: 10.1186/s40068-016-0057-2.
6 Solomatine D P & Xue Y, M5 Model Trees Compared to Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China, *ASCE J Hydrol Eng*, 9 (4) (2004) 491-501.
7 Kim J W & Pachepsky Y A, Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT stream flow simulation, *J Hydrol*, 394 (2010) 305-314.
8 Gharaei-Manesh S, Fathzadeh A & Taghizadeh-Mehrjardi R, Comparison of artificial neural network and decision tree models in estimating spatial distribution of snow depth in a semi-arid region of Iran, *Cold Reg Sci Technol*, 122 (2016) 26-35.
9 Garg N K, Deo M C & Kumar V S, Short term prediction of coastal currents using model trees, *Proc. of Indian National Conference on Advances in Hydraulic Engineering: Hydro 2008* (India), 2008, pp. 250–256.
10 Witten I H & Frank E, *Data Mining: Practical Machine Learning Tools and Techniques*, (Morgan Kaufmann, San Francisco), 2000.
11 Solomatine D P, Computational intelligence techniques in modeling water systems: Some applications*, IEEE*, (2002) 1853-1858.
12 Quinlan J R, *C4.5: Programs for Machine Learning*, (Morgan Kaufmann, California, USA), 1993.
13 Solomatine D P & Siek M B L A, Flexible and Optimal M5 Model Trees with Applications to Flow Predictions, *Proc. 6th International Conference on Hydroinformatics, Liong et al. (eds)*, (World Scientific, Singapore), 2004, pp. 1719-1725.
14 Bhattacharya P, Van-Geer F, Van-Der-Veer P & Cser J, Neural Networks and Model Trees in Spatio-Temporal Modeling of Ground Water Head, *Proc. 6th International Conference on Hydroinformatics, Liong et al. (eds)*, (World Scientific, Singapore), 2004, pp. 1303-1310.
15 Stravs L & Brilly M, Application of the M5 Machine Learning Method for the Development of the Low Flow Forecasting Model, *Proc. 7th International Conference on Hydroinformatics, HIC 2006* (Nice, France), Vol 2, 2006, pp. 863-878.
16 Joseph S & Ravichandran M, Validation of 0.25 X 0.25 Indian Ocean HYCOM.” INCOIS Report (2013). http://www.incois.gov.in/documents/hycom/hycom_0.25x0.25_tech_rep.pdf, version (08/2013).
17 NOAA, Current observation data from RAMA buoys in Indian Ocean. http://www.pmel.noaa.gov/tao/data_deliv/version (08/2013).
18 McPhaden M J, Meyers G, Ando K, Masumoto Y, Murty V S N, *et al.*, Supplement to RAMA: the research moored array for African—Asian—Australian monsoon analysis and prediction, *Bull Am Meteorol Soc*, 90 (2009) 459–480. doi:10.1175/2008B AMS2608.2 (ES5–ES8).
19 Takens F, *Detecting strange attractors in turbulence in dynamical systems and turbulence*, In: Springer Lectures Notes in Mathematics, Volume 898, (Springer-Verlag, Berlin), 1981.
20 Jekabsons G, M5PrimeLab: M5' regression tree, model tree, and tree ensemble toolbox for Matlab/Octave .http://www.cs.rtu.lv/jekabsons/version (05/2016).
21 Thomakos D D & Guerard J B Jr, Naïve, ARIMA, nonparametric, transfer function and VAR models: a comparison of forecasting performance, *Int J Forecast*, 20 (1) (2004) 53–67.
22 Dauji S, Deo M C & Bhargava K, Prediction of ocean currents with artificial neural networks, *ISH J Hydraul Eng*, (2014) 1-14. doi: 10.1080/09715010.2014.938133