Time series prediction of sea surface temperature in Indian ocean by superensembling of artificial neural network output

Shruthy Govindan[1,*], Akanksha Singh[1]

[1] Department of Computer Science and Engineering
Inderprastha Engineering College,
Ghaziabad, India
[*] Corresponding author email: shruthygovindan21@gmail.com

Abstract: ANN have been used to access the predictability of the sea surface temperature in Indian ocean. Sea surface parameter is used to understand the exchange of momentum, heat,gases and moisture across air sea interface. Its knowledge is necessary to explain and predict important climate and weather processes including the summer monsoon. The objective of the study is to have time series prediction of the sea surface temperature in Indian ocean by superensembling ANN outputs . Forecasting of severe weather conditions are proposed to be carried out. The possibility of accurate weather prediction is still a distant dream for researchers. The sensitive dependence of the weather system on initial conditions does not render the system to be modeled by deterministic methods. Such systems are called chaotic systems. It may, nevertheless, be useful to have a system which may be able to predict if catastrophic events will occur even if the scientific understanding of the subject remains limited. Therefore the creation of a model with dynamical characteristics similar to the unknown model generating the time series produced by these systems could help in the prediction of such systems. Artificial Neural Network is one such system.It remains to find out the actual effect of neural network ensemble forecast on the basis of sensitivity to initial conditions of the ANN and also on the basis of topological organization as the case is with dynamical modeling. Such ensembling is called superensembling in the context of present work. It is not a priori obvious whether or not the concept of multimodelensembling has a natural significance in statistical forecasting. Does superensembling of ANN models can be equated with multimodel havensemble? A variation of the present proposal has been very recently proposed and successfully implemented. The study aimed to do multimodel ensembling by ANN models. In the context of the above study it is interesting to observe the error patterns while doing the superensembling as proposed in the present study wherein topological ensembling has been proposed.

[Keywords: Super ensemble neural networks, sea surface temperature, autocorrelation ,multimodel ensemble]

1. Introduction

Sea Surface Temperature is a measure of the energy due to the motion of molecules at the top layer of the ocean. Depending on the sensor, space borne measurements give us an unprecedented global measurement of sea surface temperatures every few days to a week. To be more precise, Sea surface temperature(SST) is the water temperature close to the ocean's surface. The exact meaning of surface varies according to the measurement method used, but it is between 1 millimeter (0.04 in) and 20 meters (70 ft) below the sea surface. There are a variety of techniques for measuring this parameter that can potentially yield different results because different things are actually being measured.

SST predictions are sought after by the users of coastal communities dealing with fishing and sports. Like the air above it SST changes significantly overtime, although relatively less frequently due to a high specific heat. The changes in water temperature over a vertical are high at the sea surface due to large variations in the heat flux, radiation, and diurnal wind near the surface, and hence SST estimations involve considerable amount of uncertainty. There are a variety of techniques for measuring SST. These include the thermometers and thermistors mounted on drifting or moored buoys and remote sensing by satellites. In case of satellites the ocean radiation in certain wavelengths of an electromagnetic spectrum is sensed and related to SST. Microwave radiometry based on an imaging radiometer called the moderate resolution imaging spectroradiometer is also popularly used to record SST.

A time series is a sequence of data points, measured typically at successive points in time spaced at uniform time intervals. Examples of time series are the daily closing value of the Dow Jones Industrial Average and the annual flow volume of the Nile River at Aswan.

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time.

It has been pointed out by Tripathi and others(2011) that the importance of SST to accurate weather forecasting of both severe events and daily weather has been increasingly recognized over the past several years. Also, it significantly affects the monsoon locally as well as globally. Recent studies indicate that the Indian ocean impacts climate around the globe, including the widespread drought from 1998-2002 in US, southern Europe, and the parts of Asia. The importance of Indian ocean to remote climate variability has renewed interest in the study of SST anomalies of the Indian ocean. A successful simulation of the SST anomalies variation of the Indian ocean can also contribute to the monsoon forecast of the Indian subcontinent.

Using area-average SST twelve different NN models were developed for the twelve months in a year. Typically the model to predict SST for the month of January was based on all past observations of January. On evaluating the performance of the networks the authors found that the models were able to predict the anomalies with a good accuracy and whenever the dependence of present anomalies on past anomalies was nonlinear the NN models worked better than the linear statistical models. Collins and others (2004)have used meteorological variables such as input to predict targeted satellite-derived SST values in the western Mediterranean Sea. The networks trained in this way predicted the seasonal as well as the interannual variability of SST well. The impact of the heat wave that occurred during the summer of 2003 on SST was also reproduced satisfactorily. Kuge and others (2004) have compared prediction skills of different methods based on certain transfer functions, regressions, and artificial neural network. The authors analyzed data of radiolarian faunal abundance from surface sediments observed at the Atlantic and Pacific oceans. The error statistics associated with the NN predictions were found to be more attractive than the other methods, although NN yielded lesser geographic trends than those of the other methods. Gradients of Sea Surface Temperature (SST) are im- portant in determining the position of precipitation over the tropics including monsoon regions. It is assumed that the SST Anomaly (SSTA) over Indian Ocean vitally determines the monsoon rainfall variability.

It has been asserted by K.C Tripathi and others(2006)thatIndian summer monsoon, which is a part of the Asian monsoon system, is a regular annual phenomenon which brings heavy rainfall to India and adjacent countries during summer monsoon season (June to September; JJAS). It contributes about 70% - 90% of rainfall in most parts of country whereas, the rainfall during October-November-December (OND) over south India which is commonly referred as winter monsoon rainfall. It contributes about 50% of annual rainfall in the east cost of Indian Peninsula. The winter monsoon is highly variable both spatially and temporally. During winter monsoon season the prevailing wind becomes north-easterly and the zone of maximum rainfall migrates to southern India.

A better understanding of the monsoon cycle is clearly of scientific and social value. Monsoon prediction studies have utilized indicators of atmospheric circulation, land surface conditions, and Indian and Pacific Oceans SSTs (summarized in Webster et al. 1998). The predictive relationship between Indian Ocean SST and monsoon rainfall remains especially poorly characterized, particularly at lead times greater than 1–2 months before the boreal summer monsoon season. Their study assesses the relationship between SST variations in the tropical Indian Ocean and Indian monsoon rainfall, including the temporal stability of these linkages over the past 50 yr with the goal of improving multivariate monsoon prediction efforts.

1.1 Artificial neural network and forecasting

Intelligent systems are so called because they are "adaptive", in the manner in which we humans are. These systems differ from the "rigid" systems in the sense that the latter, by their inherent design, have poor or zero power of decision-making. As an example there is only one possible regression line through a data set. On the other hand, the intelligent systems "evolve" in time. They search for a better solution (but may not come up with the best solution every time) in the solution space. There may be many solutions to a problem. This, somehow, resemble the behavior of human systems that are capable of decision-making.
It is mentioned in the thesis of K C Tripathi and others (2011) that with the success of the back propagation rule for finding the derivatives . ANN became a popular tool in artificial intelligence, robotics and several authors categorically stated the importance of artifical intelligence techniques for weather forecasting and peer reviews. Comparison of ANN and traditional techniques of forecasts had begun to attract the researchers. Lepedes and Farber (1987) were the first to report that simple neural networks can outperform traditional methods by several orders of magnitude.

1.2 Ensemble forecasting

Neural network ensemble is a learning technique where many neural networks are jointly used to solve a problem. Neural network ensemble helps in improving generalization and also removes the error related to initial conditions. The output of the neural network ensemble is considered to have high correlation and low root mean square error. Aggarwal and others(2013) defined Ensembles represent a natural extension of common neural network practice that undoubtedly owes its roots to earlier works using conventional forecasting tools. Generally to create such an ensemble, several networks are trained out of which best networks are chosen to create an ensemble. Construction of neural network ensemble involves two steps: training the individual networks and then combining their predictions. Individual networks can be trained using different ensemble techniques and these techniques are based on varying parameters related to design and training of ANN such as varying initial weights, varying network type, varying network architecture involving number of hidden neurons, number of hidden layer. To combine the networks, ensemble mean or weighted mean method is used so as each ensemble is assigned weights in such a way that root mean square error is minimized. Network is trained with different random initializations and 3 such models are selected for participating in the ensemble forecasting whose errors are better.The average output of the above 3 models is expected to give better results. We have modified the learning parameters of the network where a number of networks are built with different learning parameters, such as initial weights in an MLP and varying number of hidden neurons, etc. An alternate baseline approach we investigated was the creation of a simple neural network ensemble where each network are trained independently and networks with good correlation are selected then ensemble mean of these networks are taken. Likewise hidden neurons are varied and then again same procedure applied .While theoretical results indicate that, if properly constructed and employed, neural network ensembles can generalize better than any individual model used separately, they do not provide general guidelines about the selection of different models in the ensemble.

1.3 Superensemble neural network

Superensemble architecture for numerical weather prediction was proposed by Krishnamurthy and others. This idea was an extension of the traditional ensemble mean forecasting which could remove errors associated with initial condition. The idea was to first take the ensemble forecast of individual models by using the perturbation theory and then take an ensemble of the ensemble forecasts. Thus the superensemble model was capable of taking into account the errors caused by wrong initializations as well as poor understanding of the concerned deterministic system. This was called the Multi Model Ensemble (MME). Ann superensembling is an extension of ensemble method which further improves the performance of a network by providing high correlation and low root mean square error. Several ensemble models are combined together to form a superensemble. We have used two aspect to create a superensemble that is the initial weights and the architecture of the ANN which essentially means the number of neurons in the hidden layer. Our baseline approach was the creation of a simple superensemble network using two ways first where output of various ensemble networks are combined during training phase to form a superensemble and second where mean of various ensembles are taken. We have also created a superensemble network where weighted mean concept is utilized and comparison between these results are done to see which network performs better .It has been found that weighted mean concept performs better than ensemble mean and gives a significant increase in correlation and root mean square error is decreased. For weighted concept we used network with 12 hidden neurons as it performed better then various weighted mean are statistically combined during training so that skill of ensemble of ensemble is factored into superensemble. In Consolidated ANN ensemble, output of weighted ensemble mean is fed into the network as an input then the network is trained so that the output obtained gives better performance producing good correlation and root mean square error is reduced.

1.4 The present study

Statistical prediction models such as the Artificial Neural Networks (ANN) tend to model the dynamical behavior of the system based on past information. The errors in such techniques can be due to wrong initializations as well due to wrong topology. The present study is aimed to explore the ANN prediction errors in the context of wrong initializations and wrong topology by means of ensembling and Superensembling of the ANN models.
The predictability of dynamical systems are largely governed by two factors – (i) the uncertainty arising out of initial conditions and (ii) poor understanding of the actual process in question. Ensemble forecasting has been in practice since long to account for the errors due to initial conditions. Multi Model Ensembling (MME) has been recently adopted by researchers to take into account the errors due to lesser understanding of the subject.
Super ensemble architecture for numerical weather prediction was proposed by Krishnamurthy and others (2000). This idea was an extension of the traditional ensemble mean forecasting which could remove errors associated with initial conditions (Palmer et al.,

2004, Suzuki et. al., 2004). The idea was to first take the ensemble forecast of individual models by using the perturbation theory and then take an ensemble of the ensemble forecasts. Thus the superensemble model was capable of taking into account the errors caused by wrong initializations as well as poor understanding of the concerned deterministic system. This was called the Multi Model Ensembling (MME).

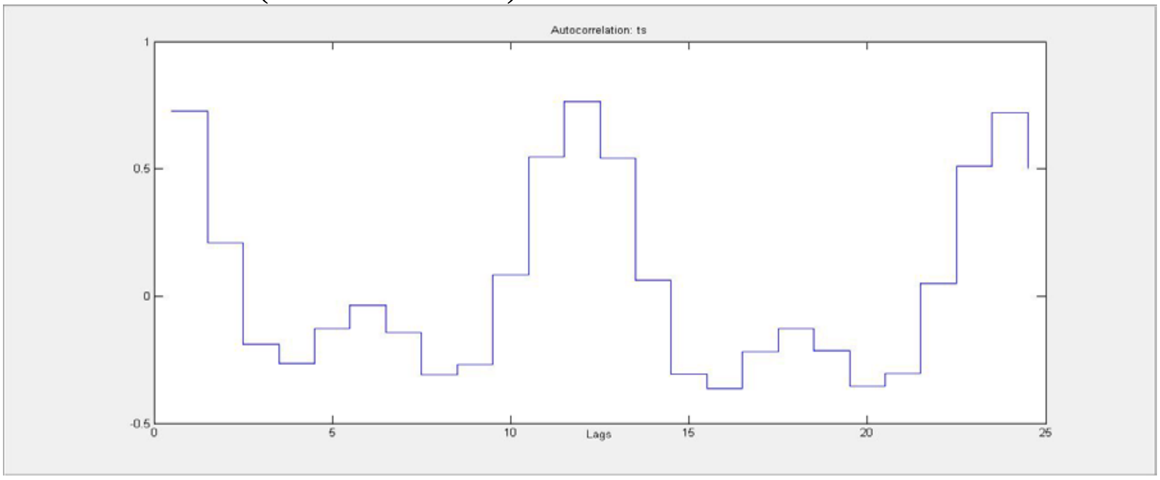
2. Methodology

2.1 Data

The extended reconstructed sea surface temperature from January 1871 to May 2004 (Smith and Reynolds, 2003, 2004, 2005) for the IO region (in the neighborhood of IOD) has been used for the present analysis. The SST for the Central Southern Indian Ocean (CSIO) and Southern Indian Ocean (SIO) has been averaged over the respective areas to obtain the time series for 23 years (January 1982 to December 2004).

2.2 Correlation Analysis

The determination of predictors is an important step as this is a precursor to a good prediction model. Autocorrelation analysis was done for the determination of predictors and also to establish the basis for an attempt to design such a prediction model. Lag values for the autocorrelation analysis have been taken from 0-24 months.Based on the correlation analysis we have selected predictors having highest positive correlations. Figure 2.1 shows the results of the autocorrelation analysis. It can be seen that the maximum correlations are obtained at values of 12 and 24 (0.8691 and 0.8461).



2.3 Partitioning of data

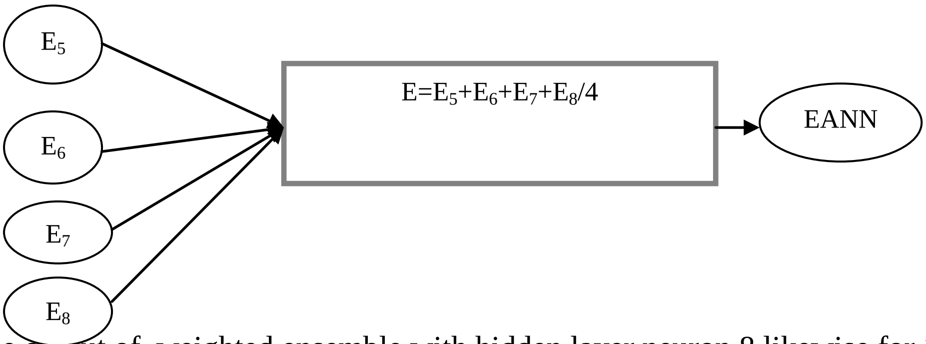
The data is partitioned into three datasets namely training, testing and validation set. The data set comprises of 23-year monthly data (from January 1982 to December 2004). This implies 1596 data points.The test set is obtained by serially taking the last 131 points. The remaining 1179 points are used for training and 262 points for validation.

2.5 Architecture of Superensemble

Multilayer feed-forward neural network models are the most popular network paradigm for forecasting applications. Those factors related to neural network model architecture include the number of input variables, the number of hidden layers and hidden nodes, the number of output nodes, the activation functions for hidden and output nodes, and the training algorithm and process.The activation functions used for all hidden nodes are the logistic function while the identity function is employed in the output layer. The number of input nodes corresponds to the number of past lagged observations used to capture the underlying pattern here which is lag12 and 24. We used four levels of hidden nodes of 5, 6 ,7 and 8 to experiment with in this study. For creating a ensemble network, we have used multilayer feed-forward architecture in which number of neurons in input and output layer is 2 and 1 whereas neurons in hidden layer are varied to create an super ensemble. To form an Consolidated ANN we have used outputs of weighted ensemble mean which is fed into the network as a input. We have used same activation function and architecture, the only difference is in the number of neurons in the input and output layer which is 3 and 1 as weighted ensemble of various networks are calculated with hidden layer neurons 8 ,10 and 12. Networks are trained so that a particular input leads to a specific target output. The training algorithm is the standard BP, which uses the gradient descent technique to minimize error. During training, each desired output is compared with the actual target and calculates error at the output layer. The backward pass is the error back propagation and adjustments of weights. Thus, the network is adjusted based on a comparison of the output and the target until the network output matches the target. When the training process is completed, then the network with adjusted estimated parameters is used to test a set of data, which is different from the training set of data.
We used two approaches for creating a superensemble:

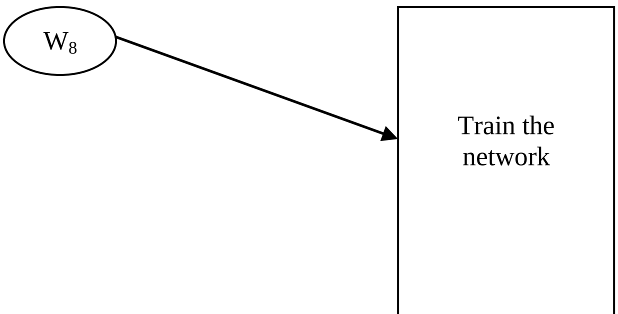
i) Simple average of ensemble forecast then call it EANN
ii) Output of weighted ensembles are fed into the network as input then call it CANN which is consolidated artificial neural networks.

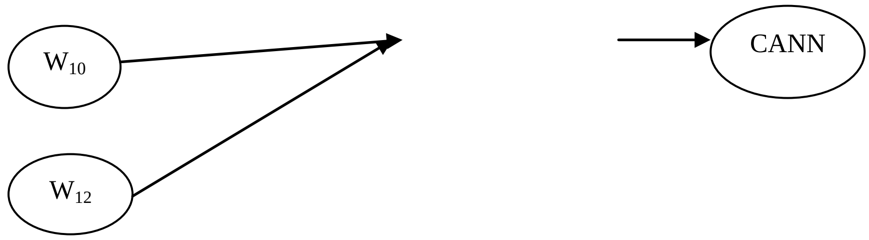
a) E_5: specifies the ensemble output of various networks with 5 neuron in hidden layer likewise for 6,7 and 8
E: Mean of various ensemble output
EANN: final output which is average of various created ensembles



b) W_8: is the output of weighted ensemble with hidden layer neuron 8 likewise for 10 and 12

CANN: output of various weighted ensemble are fed into network as a input then that network is trained to obtain a final output which is called Consolidated Artificial Neural Network .





3 Results and discussions

The performance is measured in terms of its RMSE (Root Mean Square Error), correlation. The correlation and RMSE is calculated and compared. The comparison shows the successful ANN model is developed.
Initially a ANN model with 5 hidden neuron is defined. The RMSE and correlation of the target response with desired response is obtained. The analysis of all neural network with different neuron up to 8 is done and their RMSE and correlation are then compared. The best network that has least RMSE & more correlation between target & output is considered .This is repeated for creating ensembles by varying hidden layer neurons and then taking mean of these ensembles calling it super ensemble. The RMSE and correlation obtained is compared with ANN results and It has been found that super ensemble works better and gives better performance. Then weighted mean of ensembles is taken for hidden layer neuron like 8 ,10 and 12 using RMSE and correlation for measuring performance and obtained results are compared with both ANN and ensemble results . It is found to perform even better by providing significant increase in correlation and RMSE is reduced. Finally a Consolidated ANN is created by taking outputs of previously created weighted ensembles as input . The analysis of these obtained Consolidated ANN are done up to hidden layer neuron 12 and their RMSE and compared with previously obtained results yielding even a better performance by further increase in correlation and decrease in RMSE.

Table 3.1

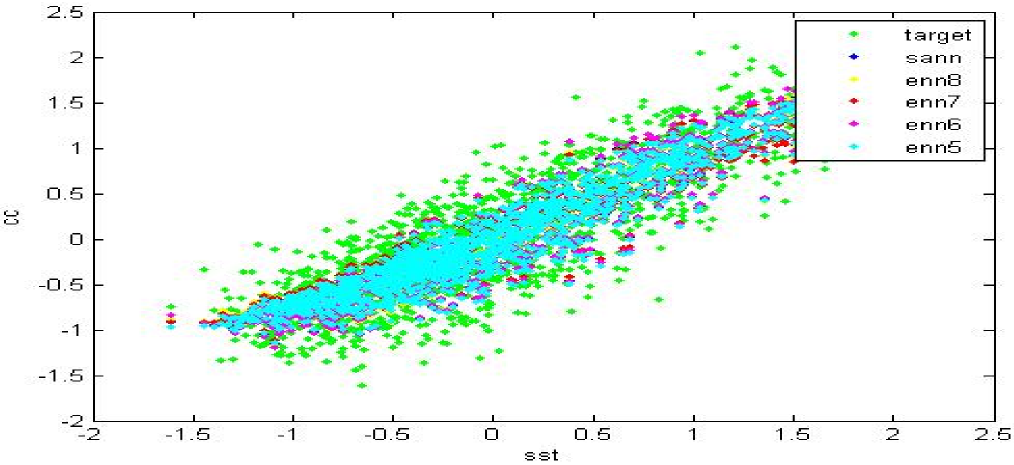
The performance measurement of various ensemble techniques using correlation coefficient and root mean square error.

Ensemble Criterion	**Model details**	**CC**	**RMSE**
Best individual ANN models	ANN5	0.80	0.075
	ANN6	0.79	0.074
	ANN7	0.80	0.074
	ANN8	0.79	0.073
(E1) Simple Ensemble (Ensemble obtained by simple averaging of outputs of individual architectures with different initial conditions)	EANN5	0.79	0.075
	EANN6	0.79	0.074
	EANN7	0.80	0.074

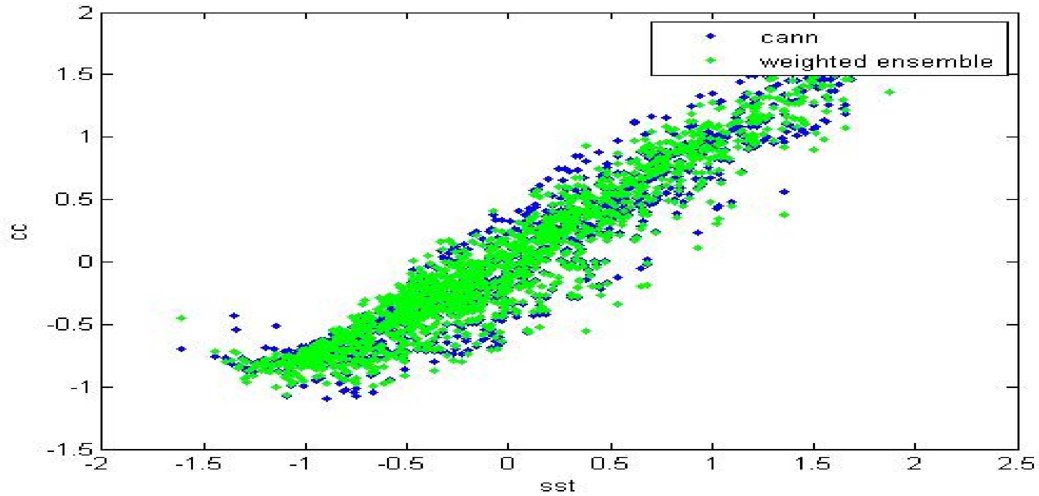
	EANN8	0.80	0.074
(E2) Simple Ensemble (Ensemble obtained by simple averaging of weights of individual architectures with different initial conditions)	WANN8	0.81	0.073
	WANN10	**0.80**	**0.073**
	WANN12	0.81	0.074
(Super ensemble S1) Ensemble of outputs of ensemble ANNs in E1 (ensemble averaging)	EANN	0.80	0.073
(Super ensemble S4) Input of weighted ensemble outputs (E2) used to train a new ANN	**CANN12**	**0.81**	**0.070**

Figure 3.1comparison of various ensemble techniques (a) ensemble averaging output compared with output (b) weighed ensemble compared with consolidated ANN

a)



b)



4 Conclusion

The method of multi model ensemble (MME) forecasting in dynamical scenario has gained considerable attention among the researchers during the last decade with the work of Krishnamurthy and others. It was found that the ensemble of ensemble forecasts leads to better prediction. The logic behind this is that while ensemble forecast removes errors that creep in due to incorrect model initialization the multi model ensembling removes errors due to poor model physics. In both cases the logic behind removal of errors is the concept of averaging. Averaging removes the chaotic factors in the output. Simple ensembling removes errors due to incorrect initializations and MME removes errors due to poor understanding of the problem.

It was, however, further seen that MME does not always lead to better prediction. In some cases ensemble forecast or even individual mode forecasts offer better results. Statistical forecasting depends on the observations made in history. A common factor in statistical and dynamical forecasting is that both utilize the property of the system that there is an underlying dynamics playing behind seemingly random behavior of the parameter. Further, there are two broad categories of statistical prediction models. The regression based models which don't require any learning and the ANN type which evolves in time. The regression based models have no concept of initialization or ensemble. The ANN models, on the other hand, are dependent on initializations. Further, predictions based on these models are highly dependent on the number of hidden layers neurons. Thus there are two sources of errors in the predictions done by ANN models. This raises a pertinent question. Whether the ensemble forecast with ANN correspond to the ensemble forecast of the dynamical models.

The errors due to incorrect model initialization are reduced by taking several initializations and averaging the outputs. This can be called ensemble forecast. The errors due to incorrect model development resulting by taking wrong number of neurons in the hidden layer can be reduced by taking average of models with different number of neurons in the hidden layer. This is called MME forecast in statistical scenario. In the present study we call such ensemble as a superensemle.

The present study investigated the effect of enesmbling and superensemling in context of ANN predictions. SST time series in a region of Indian ocean was analysed for predictability by ANN, ANN ensemble and ANN superensemble. Two approaches to ANN superensembling were discussed. It was observed that ensembling and superensembling of ANN models does not lead to better predictability of the SST. It can thus be concluded that it is better to take individual ANN forecast than taking ensemble and superensemble forecast. In past several studies have been undertaken to demonstrate the better prediction capability of ensemble ANN models than individual models. However, there are no theoretical justifications provided for the same. In light of the past and present results it can be concluded that ensembling of ANN models needs further investigation. Conclusive decisions shall either be provided by sound theoretical justifications for the use of superenesmble models or by carrying out more such studies either to prove or disprove the claim.

5 References

Aggarwal, Rashi, Kumar Rajendra ,"A comprehensive review of Numerical Weather Prediction Models", International Journal of Computer Applications,2013,vol.47

Collins D. C. , Reason C. J. C. andTanggang F. , "Predictability of Indian Ocean sea surface temperature using canonical correlation analysis," Climate Dynamics, 2004, vol22(5), pp.481-497.

Gheyas, I. A. and Smith, L. S., A novel neural network ensemble architecture for time series forecasting, Neurocomputing, **2012** (in press) doi:10.1016/j.neucom.2011.08.005
Hansen LK, Salamon P," Neural network ensembles".IEEE Trans Pattern Anal 1990,12(10):993-1001.
Hsieh, W.W,& Tang, B.,"Applying neural networks model to prediction and data analysis in meteorology and oceanography", Bull. Amer. Meteorol Soc.,1998,79,1855-1870
Imran Maqsood, Muhammad RiazKhan ,AjithAbaraham . "An ensemble of neural networks for weather forecasting", Springer-Verlag, Neural Comput&Applic ,vol. 13 pg.112-122,2004.
Kang I. S. , Lee J. Y. , and Jhung J. G. , "A statistical approach to Indian Ocean sea surface temperature prediction using a dyamical ENSO prediction," Geophysical Reasearch Letters, 2004.vol.31(9), article ID L09212, pp.1-5.
Krishnamurti, T. N. and C. M. Kishtawal, T. LaRow, et, al., Improved weather and seasonal climate forecasts from multimodelsuperensemble, Science, 285 **(1999)**, 1548–1550.
Lynch P." The Origins of Computer Weather Prediction and Climatic Modelling". Jounal on Computational Physics, 2008,227: 3431–3444.
Maqsood Imran, Khan Muhammad Riaz ,Abaraham . "An ensemble of neural networks for weather forecasting", Springer-Verlag, Neural Comput&Applic ,2004,vol. 13, pg.112-122.
McCulloh W.S. & Pitts W.,"A logical calculus of the ideas immanent in neural nets" ,Bull Math. Biophys.,5(1943) ,133-137
Palmer, T.N. and others, Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER), Bulletin of American Meteorological Society, 85 **(2004)** 853-872.
Shukla ,R. P, Tripathi, K. C, Pandey, A. C. and Das, I. M. L, Prediction of Indian summer monsoon rainfall using Niño Indices: a neural network approach, Atmospheric Research, DOI information: 10.1016/j.atmosres.2011.06.013, 102**(2011)**, 99–109
Smith T.M. and Reynolds R.W., Extended Reconstruction of Global Sea Surface Temperatures Based on COADS Data: (1854-1997), Journal of Climate, 16 (2003) 1495-1510.
Smith T.M. and Reynolds R.W., Improved Extended Reconstruction of SST (1854-1997), Journal of Climate, 17 (2004) 2466-2477.
Smith T.M. and Reynolds R.W., A global merged land-air-sea surface temperature reconstruction based on historical observations (1880 – 1997). Journal of Climate, 18 (2005) 2021-2036.
Tripathi K.C, Agarwal Rashi,Hrisheekesha P N, "Global Prediction algorithms and predictability of anomalous points in a time series", International journal of computers & technology, 2013.
Tripathi ,K.C. and Das ,I.M.L, and Sahai ,A.K. "Predictability of sea surface temperature anomalies in the Indian ocean using Artificial neural network" ,Indian Journal of Marine Sciences ,vol.35(3), September 2006
Yun W.T , L.Stefanova, A.K Mitra, T.S.V. Vijaya Kumar, W. Dewar and T.N. Krishnamurti, "Multi-Model Synthetic superensemble Algorithm for seasonal climate prediction using DEMETER Forecasts", Tellus, DEMETER (2004)
Zhou ZH, Wu J, Tang W," Ensembling neural networks: many could be better than all", Artificial Intelligence 137(1-2):239-263,2002.
Zurada, J.M., "Introduction to artificial neural systems",(Jaico Publishing House, Mumbai, India)2002,pp. 30-400 .