Superensembling of Artificial Neural Network Models for Investigating The Effect of Polar Sea Ice on Sea Surface Temperature in Indian Ocean Region

Shruthy Govindan\textsuperscript{1}, K. C. Tripathi\textsuperscript{2}

Associate professor\textsuperscript{1, 2}
Department of Computer Science and Engineering
Inderprastha Engineering College, Ghaziabad, India

Abstract:
There are two broad sources of errors when prediction of any atmospheric parameter is made by dynamical models - one due to errors in model initializations and two due to limited understanding of the physical phenomenon at hand. The error introduced by model initialization is reduced by executing the model with different initial conditions. The error due to model design has been recently attempted to be reduced by taking the multi model ensemble. Apart from dynamical forecasting, statistical forecasting contributes significantly to the research of this area. The statistical prediction models fall under two categories – one the rigid models such as the regression and two the Artificial Neural Networks (ANN) models that employ learning rules for prediction. The regression models being rigid and hard coded have neither an issue of initialization and design nor do they have a scope of improvement. On the other hand, the ANN models have, like dynamical models, issues of initialization and design. It becomes a pertinent question if the ensembling in the context of ANN has same effect as the dynamical ensembling. The present study aims to answer this question in the context of effect of Antarctic sea ice concentration on the sea surface temperature in the Indian Ocean as the effect of former on the latter has been of interest. The concept of Super ensemble ANN architecture and various approaches to ensembling and superensembling have been discussed. The super ensemble forecast is compared with normal and ensemble forecasts. It is observed that superensembling in the context of prediction with ANN appears to have the same effect as in the context of prediction with dynamical models.

Keywords: Super ensemble neural networks, sea ice concentration, correlation, Indian summer monsoon rainfall

1. INTRODUCTION
Sea ice concentration is both an indicator and driver of high latitude climatic change with strong social and economic importance. Sea ice is simply frozen ocean water (D. N. Thomas and G. S. Dieckmann, 2009). It forms, grows and melts in the ocean. Sea ice only develops around Antarctica (Shailendra Rai et al. 2008), in the southern hemisphere occurring as far north as 55 degrees south latitude. On average it covers about 25 million square kilometers of the earth. Although it may not directly affect us, it is a key component of our planet because it influences climate and wildlife. Sea ice regulates the exchange between Earth’s surface and the atmosphere, or the Earth’s energy balance (Tripathi and Das, 2008). It is a key boundary condition for atmospheric models (Jaiser et al., 2011). Sea ice also affects the movement of ocean waters (Aagaard and Carmack, 2012). When sea ice forms, most of the salt is pushed into the ocean water below the ice. Water below sea ice has a higher concentration of salt and is more dense than surrounding ocean water and so it sinks (Barber et al., 2012). Sea ice, thus, contributes to the ocean’s global conveyor belt circulation. Changes in the amount of sea ice can disrupt normal ocean circulation thereby leading to changes in the global climate. India being primarily agricultural economy depends largely on agricultural yields. Further the farmers mostly rely on rain water for their crops. Hence advance prediction of precipitation over the Indian Ocean region plays a crucial role in the economic development of Indian and neighboring countries. Sea Surface Temperatures in the Indian Ocean region play a critical role in the overall precipitation over the Indian subcontinent (Shukla et al. 2013; Sahai et al., 2008). Any predictor which helps in early prediction of the Indian Ocean SST is thus investigated at length. Investigation of long-term impact of Antarctic sea ice concentration on Indian Ocean SST is important, not only because of the potential importance of sea ice in global climate, but also for the practical purpose making advance predictions. This effect is thus investigated at length by the researchers of the community (Shailendra Rai et al., 2008). The present author (Tripathi and Das, 2009) has demonstrated that AnSIC is a prospective predictor of southern Indian Ocean SST indices and also of All India Rainfall (Tripathi and Das, 2009). Recently it has been found by Tripathi et al., 2008 that Antarctic sea ice concentration has a possible relationship with the sea surface temperature. In our study, We investigated the possibility of prediction of effect of polar sea ice on Sea surface temperature in the Indian ocean using superensemble technique. We used super ensemble technique to predict the effect of polar sea ice on sea surface temperature in the Indian Ocean by training network with different random initializations like by varying number of hidden neurons in the hidden layer.

1.1 Artificial neural network and forecasting
Artificial Neural Networks are statistical models which are used to discover the underlying behavior of the dynamical system. It has been successfully utilized by the researchers of diverse disciplines such as meteorology, oceanography, remote sensing for non linear regression. Zurada, 2002 defined Artificial neural network as a mathematical model that...
simulates the behavior of human brain and composed of simple processing units called synthetic neurons which were formally defined by McCulloh and Pitts, 1943 where each unit has small amount of local memory. ANNs provide a methodology for solving many types of non-linear problems that are difficult to solve by the traditional techniques. With ANNs, they are capable of extracting the relationship between the inputs and outputs of a process, without the physics being explicitly provided. Thus, these properties of ANNs are well suited to the problem of weather forecasting under consideration. Neural network technique is used as it works as universal approximator capable of approximating any continuous function or unknown function when given a representative set for training. They are also referred to as intelligent systems because they are adaptive like we humans are. Using mathematical models of the atmosphere to predict future weather based on current weather conditions is called numerical weather prediction. This needs full knowledge of atmospheric dynamics and involves calculations with a large number of variables and huge datasets. Although this process requires a lot of computational resources due to the advancement of modern computer hardware there have been many improvements in numerical weather prediction (Lynch, 2008). We are using statistical model for weather prediction where behavior of a system is modeled using historical information and are computationally inexpensive.

1.2 Ensemble forecasting
The error in forecasts arises as a result of errors in observations and also because of the limitation of the model. To compensate for these shortcomings there has been a trend in recent years toward ensemble forecasting, which is simply the realization of a number of model integrations using perturbed initial conditions. Hansen et al.,1990 defined Neural network ensemble as a learning technique where many neural networks are combined together to solve a problem. Neural network ensemble helps in improving generalization and prediction ability 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. Ensembles represent a natural extension of common neural network practice that 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 which are based on random initializations 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. It has been stated by Vladimir et al., 2012 that weighted ensemble mean is better than conservative ensemble that is arithmetic ensemble mean. This study evaluates the effectiveness of several methods of creating an ensemble in predicting the rainfall.

1.3 Superensemble neural network
Krishnamurthy et al., 1994 stated that Superensemble architecture for numerical weather prediction is an extension over traditional ensemble mean forecasting and is capable of taking into account the error caused by wrong initializations as well as poor understanding of the system. 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 aspects 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 by 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 and then compared where former provided better results. 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 by Vladimir et al., 2012 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 8 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 a superensemble.

![Diagram](https://via.placeholder.com/150)

- Super ensemble network
- To create
- Mean of these ensembles taken
- Best resultant ensemble with hidden neuron 5
- Best resultant ensemble with hidden neuron 6
- Best resultant ensemble with hidden neuron 7
- Best resultant ensemble with hidden neuron 8

1.4 ANN Ensemble
In 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 and then
curve fitting is applied to establish a quantitative relationship between a group of predictor variables and output. We also created a ANN ensemble where output of various created ensembles is fed into the network as a input then the network is trained so that the provided output enhances the performance and gives better results in comparison to superensemble that we obtain by simply averaging all the various created ensembles.

2. METHODOLOGY

2.1 Data
The monthly AnSIC data for 23 years (January 1982 to December 2004) downloaded from http://www.cdc.noaa.gov has been used. For the sea surface temperature, The extended reconstructed sea surface temperature from January 1871 to May 2004 for the IO region (in the neighborhood of IOD) has been used for the present analysis. We used data for 22 years that is from January 1982 to December 2003 to investigate the effect of polar sea ice on sea surface temperature.

2.2 Correlation Analysis
The correlation analysis has been done between the time series of AnSIC and the ERSST, the ERSST time series being ‘ahead’ by 1 to 24 months. It has been found that the AnSIC has a very strong correlation (nearly 0.62) with the SST when the former lags behind by 1 month. Hence, the AnSIC is used as the predictor for the Sea Surface Temperature in the region one month ahead. The anomalies of SST are calculated. The determination of predictors is an important step as this is a precursor to a good prediction model. Cross-correlation analysis was done for the determination of predictors and also to establish the basis for an attempt to design such a prediction model. Cross correlation value of an AnSIC series with a lagged series of ERSST gives an insight of the level of dependence of the future values in the series on the present value.

2.3 Normalization
The predictor and the predictand series has been normalized using the following scheme:

\[ X_{N} = \left( \frac{X - X_{MIN}}{X_{MAX} - X_{MIN}} \right)^{0.6} + 0.2 \]

Where \( X \) is the sea ice concentration, \( X_{MAX} \) is the maximum concentration for the month, \( X_{MIN} \) is the minimum concentration for the month, \( X_{N} \) is the normalized concentration for the month. The factors of 0.6 and 0.2 are included so that the normalized lies within 0.8.

2.4 Partitioning of data
The data set comprises of 22-year monthly data (from January 1982 to December 2003). This implies 264 data points. Since we are predicting 1 month in advance, there are effectively 263 input-output relationships. The normalized data is partitioned into three datasets namely training, testing and validation set. The data set comprises of 22-year monthly data (from January 1982 to December 2003). This implies 264 data points. Since we are predicting 1 month in advance, there are effectively 263 input-output relationships. The test set is obtained by serially taking the last 22 points. The remaining 197 points are used for training and 44 points for validation.

2.5 Architecture of ANN Ensemble
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 lag1. 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 1 whereas neurons in hidden layer are varied to create a super ensemble. To form an ANN Ensemble we used output of various created ensembles which are fed into the network as input then that network is trained to obtain the output. We then used weighted ensemble output as input to form an weighted ANN ensemble. 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. Networks are trained so that a particular input leads to a specific target output. The training algorithm is the standard Back-propagation, which uses the Gradient descent technique to minimize error. During training, each desired output is compared with the actual target and then error is calculated 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, the network with adjusted estimated parameters is used to test a set of data, which is different from the training set of data.

3 RESULTS AND DISCUSSIONS
Table 1 shows the performance measures of various superensemble models. The mode of superensembling is mentioned in short. We have also compared the performance of superensembling with simple ensemble forecasts as well as individual models. We have used correlation coefficient and RMS errors as the criteria for performance evaluation. The best results under all the categories are highlighted in red. It can be seen that the best individual model is the ANN with 6 hidden neurons. This is demonstrated by best in class correlation (0.80) and RMS error (0.074). The best ensemble model, in the two categories of simple ensembling, also turns out to be the same with the same performance measures. However, the results show no significant improvement over
the individual model ANN6. Hence simple ensembling does not lead to significant improvement in the performance. Looking at the superensembling results (last four rows in the table) we see that the ensemble method S4 does not lead to any significant improvement in performance. Thus the results obtained indicate that the ANN ensemble (S4) gives better performance than its counterparts and can be effectively used in predicting the effect of polar sea ice on rainfall. However using the method of training ANN with outputs of ensemble ANN gives significantly better results (correlation coefficient=0.88 and RMSE=0.054). We then applied polynomial fitting to various created ensembles and of which ensemble with hidden neuron 8 is found to give better results so ensemble with hidden neuron 8 is chosen and polynomial fitting is applied to weighted mean of this ensemble and also to the ANN ensemble and their results are compared where ANN ensemble performed better and Weighted ANN ensemble giving the best results. Curve fitting generally that is linear, quadratic and cubic are applied to find the best fitting line for the data.

Table 1. The performance measurement of various ensemble techniques using correlation coefficient and root mean square error.

<table>
<thead>
<tr>
<th>Ensemble Criterion</th>
<th>Model details</th>
<th>CC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best individual ANN models</td>
<td>ANN5</td>
<td>0.76</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>ANN6</td>
<td>0.70</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>ANN7</td>
<td>0.63</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>ANN8</td>
<td>0.65</td>
<td>0.093</td>
</tr>
<tr>
<td>(E1) Simple Ensemble (Ensemble obtained by simple averaging of outputs of individual architectures with different initial conditions)</td>
<td>EANN5</td>
<td>0.86</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>EANN6</td>
<td>0.76</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>EANN7</td>
<td>0.83</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>EANN8</td>
<td>0.73</td>
<td>0.084</td>
</tr>
<tr>
<td>(E2) Simple Ensemble (Ensemble obtained by simple averaging of weights of individual architectures with different initial conditions)</td>
<td>WANN5</td>
<td>0.84</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>WANN6</td>
<td>0.84</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>WANN7</td>
<td>0.73</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>WANN8</td>
<td>0.83</td>
<td>0.075</td>
</tr>
<tr>
<td>(Super ensemble S1)</td>
<td>EANN(AVG)</td>
<td>0.73</td>
<td>0.084</td>
</tr>
<tr>
<td>(Super ensemble S2)</td>
<td>WANN(AVG)</td>
<td>0.80</td>
<td>0.072</td>
</tr>
<tr>
<td>(Super ensemble S3)</td>
<td>EANN</td>
<td>0.85</td>
<td>0.077</td>
</tr>
<tr>
<td>(Super ensemble S4)</td>
<td>CANN6</td>
<td>0.88</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Table 2. The performance evaluation of the various ensemble techniques using curve fitting that is linear fit, quadratic fit and cubic fit and correlation coefficient and root mean square error is used for measuring performance of these ensemble networks.

<table>
<thead>
<tr>
<th></th>
<th>LINEAR FIT</th>
<th>QUADRATIC FIT</th>
<th>CUBIC FIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>RMSE</td>
<td>CC</td>
</tr>
<tr>
<td>EANN5</td>
<td>0.7945</td>
<td>0.0603</td>
<td>0.7928</td>
</tr>
<tr>
<td>EANN6</td>
<td>0.7991</td>
<td>0.0582</td>
<td>0.7994</td>
</tr>
<tr>
<td>EANN7</td>
<td>0.8004</td>
<td>0.0601</td>
<td>0.8005</td>
</tr>
<tr>
<td>EANN8</td>
<td>0.8033</td>
<td>0.0583</td>
<td>0.8036</td>
</tr>
<tr>
<td>EANN(AVG)</td>
<td>0.8008</td>
<td>0.0589</td>
<td>0.8011</td>
</tr>
<tr>
<td>EANN</td>
<td>0.8051</td>
<td>0.0582</td>
<td>0.8053</td>
</tr>
<tr>
<td>WANN8</td>
<td>0.8058</td>
<td>0.0579</td>
<td>0.8066</td>
</tr>
<tr>
<td>CANN8</td>
<td>0.8246</td>
<td>0.0573</td>
<td>0.8268</td>
</tr>
</tbody>
</table>
The results thus obtained show that ANN ensemble performs better than simple ensemble averaging and weighted ensemble mean better than ensemble averaging and Weighted ANN ensemble even performs better than ANN ensemble. ANN ensemble shows better correlation and root mean square error is further minimized when we applied curve fitting techniques to the ANN ensemble output obtained

![Image](http://ijesc.org/)  

**Figure 1.** Comparison of various ensemble techniques (a) ensemble averaging output compared with target and individual ensembles output (b) ANN ensemble compared with target and ensemble averaging output (c) Weighted ensemble mean output compared with target and Weighted ANN ensemble

4 CONCLUSION

Antarctic Sea Ice Concentration, AnSIC, is a factor which significantly affects the amount of precipitation that the Indian subcontinent receives. The impact of AnSIC on the Indian monsoon is required to be investigated. This is achieved by mathematical modeling of the processes that connects the two. However, numerical prediction models require huge computational resources and manpower. This incurs a huge cost in such modeling. Statistical modeling is used since long to achieve the objective of the numerical weather prediction models at a low cost. After the advancements in the research of expert systems and other computational models such as the ANN, statistical prediction and modeling has gained significant attention of the researchers. The quality of forecasts by statistical and dynamical methods has been found to be of almost same accuracy. This is because both methods model the underlying deterministic laws that govern the actual system. There are two sources of errors in prediction by dynamical systems. Firstly due to wrong initialization of the model and secondly due to wrong equations being incorporated in the model. Different researchers employ different equations for the modeling of same phenomena depending upon their understanding of the problem. Both types of errors are bound to creep in. The error due to wrong initialization cannot be eliminated because of finite capacity of recording instruments. Weather being a chaotic system deviates from that of the model fast because of this error. The error due to wrong equations cannot be eliminated because weather is not a deterministic system as of now. Hence these errors are bound to be present. During the last decade researchers of the dynamical modeling community have devised the technique of Multi Model Ensemble, MME, to get reduce the errors due to the two sources. It is ensemble of ensembles of individual models initialized with slightly different initialization conditions. It has been found that MME forecast is better than individual forecasts. Since statistical and dynamical modeling works on the same principle of modeling the underlying deterministic system and so far the quality of forecasts with the two types of models has been found to be similar, we are tempted to put up a question if the concept of MME is applicable in statistical modeling as well. If yes, what is the principle and how does the quality of forecasts compare. We have taken the ANN model to be our statistical model for the analysis. We call the two layer of ensemble of models a superensemble ANN model. As a moderate first step we first wish to investigate if such a system would give better forecast than the individual model forecast. In the present study we have investigated the predictability of SST in a region of Indian Ocean using the AnSIC as predictor in order to see the effect of AnSIC on the Indian climate. Owing to the importance of MME forecast, we have designed a superensemble ANN architecture to carry out the predictability studies. Three approaches to superensembling were outlined and discussed. The results show no difference between the first two. Further, the results show no significant improvement over best individual forecast. However, the third approach to superensembling gives encouraging results. It was found that the forecast obtained by connecting two separate ANN layers is better than the individual forecast, their ensemble forecast and their superensemble forecast when first two approaches to superensembling were used. We have used the correlation coefficient and RMS error as the measures of quality of forecasts. Although the improvements were slight but since they match with the improvements obtained by MME forecast, it can be prima facie concluded that superensembling of ANN models bears resemblance to the idea of MME forecast used by numerical models. However, further research is needed to consolidate our findings. We plan to carry out such research in future for time series forecasting of various meteorological parameters.
5 REFERENCES


