Contents lists available at ScienceDirect

Comptes Rendus Geoscience

www.sciencedirect.com

External geophysics, climate

Multivariate forecast of winter monsoon rainfall in India using SST anomaly as a predictor: Neurocomputing and statistical approaches

Prévision multivariée relative à la pluviosité de mousson d'hiver en Inde, utilisant l'anomalie de température de surface océanique (SST) en tant que prédicteur : calcul neuronal et approches statistiques

Goutami Chattopadhyay [a], Surajit Chattopadhyay [b,*], Rajni Jain [c]

[a] Institute of Radiophysics and Electronics, 92, Acharya Prafulla Chandra Road, Kolkata 700009, India
[b] Department of Computer Application, Pailan College of Management and Technology, Bengal Pailan Park, Kolkata 700104, India
[c] National Centre for Agricultural Economics and Policy Research, Pusa, DPS Marg, New Delhi 110012, India

A B S T R A C T

In this article, the complexities in the relationship between rainfall and sea surface temperature (SST) anomalies during the winter monsoon over India were evaluated statistically using scatter plot matrices and autocorrelation functions. Linear, as well as polynomial trend equations were obtained, and it was observed that the coefficient of determination for the linear trend was very low and it remained low even when polynomial trend of degree six was used. An exponential regression equation and an artificial neural network with extensive variable selection were generated to forecast the average winter monsoon rainfall of a given year using the rainfall amounts and the SST anomalies in the winter monsoon months of the previous year as predictors. The regression coefficients for the multiple exponential regression equation were generated using Levenberg-Marquardt algorithm. The artificial neural network was generated in the form of a multilayer perceptron with sigmoid non-linearity and genetic-algorithm based variable selection. Both of the predictive models were judged statistically using the Willmott's index, percentage error of prediction, and prediction yields. The statistical assessment revealed the potential of artificial neural network over exponential regression.

© 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

R É S U M É

Dans cet article, les complexités de relation entre anomalies de pluviosité et de température de surface océanique (SST) pendant la mousson d'hiver en Inde ont été évaluées en utilisant les matrices de dispersion et les fonctions d'auto-corrélation. Des équations à tendance linéaire, de même que des équations à tendance polynomiale ont été obtenues et il a été observé que le coefficient de détermination dans le cas de la tendance linéaire était très bas et restait bas, même quand la tendance polynomiale de degré 6 était utilisée. Une équation de régression exponentielle et un réseau neuronal artificiel avec sélection largement variable ont été générés, pour prédire la pluviosité moyenne de mousson d'hiver, lors d'une année donnée, en utilisant en tant que prédicteurs les anomalies de précipitations et de température de surface océanique pendant les mois de

* Corresponding author.
   E-mail addresses: goutami15@yahoo.co.in (G. Chattopadhyay), surajit_2008@yahoo.co.in (S. Chattopadhyay), jainrajni@hotmail.com (R. Jain).

mousson d'hiver de l'année précédente. Les coefficients de régression pour l'équation de régression exponentielle multiple ont été générés en utilisant l'algorithme Levenberg-Marquardt. Le réseau neuronal artificiel a été généré sous la forme d'une perception multicouche à non-linéarité sigmoïde et algorithme basé sur une sélection variable. Les deux modèles prédictifs sont jugés statistiquement en utilisant l'index de Willmott, le pourcentage d'erreur de prédiction et le rendement de bonne prédiction. L'évaluation statistique révèle que le réseau neuronal artificiel a un potentiel supérieur à celui de la régression exponentielle.

1. Introduction

Sea surface temperature (SST) in the tropical Indian Ocean varies significantly at interannual time scales. This variability is dominated by an Indian Ocean basin mode and dipole mode. The Indian Ocean basin accounts for about 30% of the total variance of the tropical Indian Ocean SST anomaly (Yang et al., 2010). Murtugudde and Busalacchi (1999) simulated the interannual variability of the tropical Indian Ocean SST. The tropical Indian Ocean is influenced by two contrasting air masses (continental and marine) associated with the Indian monsoon system (Li and Ramanathan, 2002). The summer monsoon (southwest monsoon) usually starts by the end of May and continues until November. The winter or northeast monsoon rainfall, which occurs mainly from October to December, is dominant over southern peninsular India, consisting of the six meteorological subdivisions, coastal Andhra Pradesh, Rayalaseema, south interior Karnataka, coastal Karnataka, Tamilnadu, and Kerala (Rao, 1999) (Fig. 1). The winter circulation prevails for the six months from November to April. May and October are the transitional months (Ananthakrishnan, 1977). During summer monsoon, winds are mainly southwesterly or westerly and during winter monsoon, the prevailing winds are mostly northeasterly. The southwest (summer) monsoon (SWM) and the northeast (winter) monsoons (NEM) influence weather and climate between 30°N and 30°S over the African, Indian, and Asian land-masses (Reddy and Salvekar, 2003). Accurate long lead prediction of monsoon rainfall can improve planning to mitigate the adverse impacts of monsoon variability and to take advantage of the beneficial conditions (Reddy and Salvekar, 2003). The variability in the monsoon rainfall depends heavily upon the SST anomaly over the Indian Ocean (Clark et al., 2000). Several authors, for example Barsugli and Sardeshmukh (2002), Branstor (1985), Clark et al. (2000), Guo et al. (2002), Hartmann et al. (2007), Lau and Weng (2001), Nicholls (1983) have discussed the association between SST anomaly and rainfall. The SST anomalies cause local changes in the low-level temperatures, winds, and humidity, usually leading to precipitation anomalies in the vicinity of the SST anomaly (Barsugli and Sardeshmukh (2002)). Gradients of SST within the oceans are important in determining the location of precipitation over the Tropics, including the monsoon regions (Clark et al., 2000).

As the extra tropical circulation anomalies display energy dispersion away from the region of anomalous tropical convection, they are interpreted as a Rossby wave response to the latent heat release associated with the tropical convection (Ferranti et al., 1990). In regions of anomalous tropical heating, there is a dynamical response with anomalous large-scale ascent and upper tropospheric divergence, which acts as a Rossby wave source (Sardesh-mukh and Hoskins, 1988) for extra tropical waves. Conversely, in regions of reduced convection and anomalous cooling, the tropical responses are the anomalous descent and upper-tropospheric convergence (Matthews et al., 2004). Positive (negative) SST anomalies lead the enhanced (suppressed) MJO convection by approximately 10–12 days (a quarter cycle), consistent with the atmosphere responding to the ocean forcing (Matthews et al., 2004). The SST anomalies themselves have been simulated in thermo dynamical ocean models as the response to the observed anomalous surface fluxes of latent heat and short-wave radiation (Matthews et al., 2004; Shinoda and Hendon, 1998) without the need to a mass flux approach and closed on buoyancy. SST anomalies has been addressed by Woolnough et al. (2000), who examined the equilibrium response to an idealized, eastward-propagating equatorial SST dipole anomaly in an "aquaplanet" atmospheric general circulation model (AGCM).

The SST anomalies influence the atmosphere by altering the flux of latent heat and sensible heat from the ocean (Holton, 1972). The efficiency of such anomalies in exciting global scale responses depends on their ability to generate
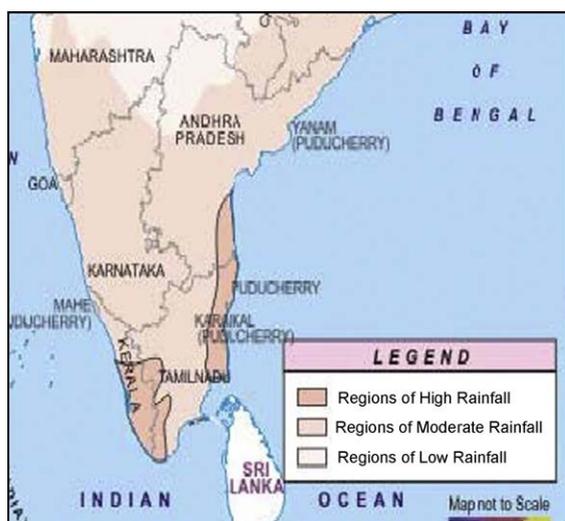

Fig. 1. Map of southern India where winter or northeast monsoon rainfall is prominent.

Fig. 1. Carte du Sud de l'Inde où les précipitations de la mousson d'hiver ou du nord-est prévalent.

Rossby waves. In the extra tropics SST, anomalies primarily generate low-level heating and this is balanced by horizontal temperature advection. In the tropics, positive SST anomalies are associated with enhanced convection and the resulting heating is balanced by adiabatic cooling. SST anomalies also play an important role in producing rainfall (Reddy and Salvekar, 2003). Linkage between Asian winter monsoon and SST anomalies over tropical Indian Ocean is discussed in the papers like Annamalai et al. (2005), Yang et al. (2010) and the linkage between Asian winter monsoon and SST anomaly in the tropical Pacific is discussed in Wen et al. (2000).

El-Niño-Southern Oscillation (ENSO) is a coupled Ocean-atmosphere phenomenon that has worldwide impact on climate in general and Indian monsoons in particular and the association between ENSO and Indian monsoon rainfall has been thoroughly studied in Krishnamurthy and Goswami (2000), Wu and Kirtman (2004) and Xavier et al. (2007).

The long-recognized negative correlation between Indian monsoon rainfall and ENSO (Webster and Palmer, 1997) in which a weak (strong) monsoon is related to a warm (cold) event through an anomalous Walker cell driven by tropical East Pacific SST anomalies, has weakened rapidly since the late 1970s (Chang et al., 2001). This weakened relationship is defined by the correlation between June and September all-India rainfall and Niño-3 (5°S–5°N, 150°–90°W) SST (Chang et al., 2001; Qiaoping et al., 2007). A handful of significant studies are available where the relationship between SST and rainfall has been studied (e.g. Nobre and Srukla, 1996; Uvo et al., 1998). Chongyin (1990) investigated the relationship between El-Nino event and NEM over South-East Asia.

The oscillations in wind stress owing to the Southern Oscillation are associated with changes in the circulation of the ocean and the SST anomaly that are referred to as El-Nino. This term was originally applied to a warming of the coastal waters of Peru and Ecuador, which occurs annually near Christmastime. ENSO's maximum SST anomalies occur in the equatorial eastern and central Pacific, ENSO affects the global ocean (Wang et al., 2006). Outside the tropical Pacific significant ENSO related SST anomalies are found over many places, such as in the tropical North Atlantic, the tropical Indian Ocean, the extratropical North and South Pacific, and the South China Sea (Deser et al., 2004; Klein et al., 1999). The strength of ENSO is measured by two principle indices namely Southern Oscillation Index (SOI) and SST index over Nino-3.4 region, which refers to, the anomalous SSTs within the region bounded by 5°N–5°S and 170°E–120°W. Warm ENSO episodes are characterised by increased number and intensity of tropical storms over the Bay of Bengal and hence enhanced NEM rainfall. An analysis of mean monthly data of 124 years by Bhanu Kumar et al. (2004a) reveals that the relationship between the SOI in April and May and the NEM rainfall during October through December over South India is variable and non stationery.

Present paper attempts to develop a predictive model for NEM over India using SST anomaly as one of the predictors. Some significant studies in the field of NEM over East Asia and its association with SST anomalies have been reviewed. The literatures include Qiaoping et al. (2007), Wen et al. (2000) and Zhang et al. (1997). A handful of works are available on the prediction of SWM over India (e.g. Guhathakurta, 2008; Sahai et al., 2000; Venkatesan et al., 1997 and many others). However, limited studies could be identified where prediction of NEM rainfall over India has been discussed. During the NEM season, southern peninsular India receives dominant rainfall. Nayagam et al. (2009) studied spatial and temporal variability of rainfall over Peninsular India during NEM. Bhanu Kumar et al. (2004b) established a probable link between mean September upper-air temperatures at Indian stations and the ensuing winter monsoon rainfall over eight individual meteorological subdivisions of southern India where the winter monsoon accounts for a large percentage of mean annual rainfall. During the season of NEM, India receives about 11% of its annual rainfall, while many of the above subdivisions receive 17–49% of their annual rainfall. Major agricultural operations in this southern part of the country are normally undertaken during this season. The winter monsoon rainfall is of considerable economic importance for this region, which constitutes about 15% of the Indian subcontinent (Rao, 1999). Rao (1963) provided a detailed description of the Indian NEM. Dhar and Rakhecha (1983) examined the association between the SWM and NEM rainfall over Tamilnadu for the 100-years period (1877–1976) and found that the SWM rainfall is negatively correlated with that of the northeast monsoon. Singh (1995) studied the influence of Bay of Bengal on NEM rainfall in two contrasting winter monsoon years 1987 and 1988. Similarly, Raj (1996) studied the thermodynamic structure of the atmosphere over coastal Tamilnadu during the NEM season. Deviating from these works, Rao (1999) examined the sub divisional and time variations of the SO relationship with the rainfall over India during summer as well as winter monsoon seasons, by using long time data for the 122-years period from 1872 to 1993.

In the present work, we decided to generate a predictive model for rainfall during NEM over India. From the above discussion and the literatures mentioned in the earlier paragraphs we felt the necessity of incorporating SST anomaly as a predictor. Since artificial neural network (ANN) is being attempted in several papers dealing with complex non-linear meteorological events (e.g. Elsner and Tsonis, 1992), we decided to attempt an ANN in this problem. The subsequent sections are organized as follows: section 2 discusses some literatures in the application of ANN in rainfall prediction. Section 3 describes the structure of the time series considered in the study. Section 4 describes theory, implementations details, and statistical assessment of the ANN and non-linear regression based models. The conclusion is presented in Section 5.

2. Artificial Neural Network in rainfall prediction – a brief overview

Accurate rainfall predictions are essential for planning day-to-day activities. Several papers are available where the rainfall time series have been dealt with statistically

(Bell and Suhasini, 1994; Singh and Kripalani, 2007; Ward and Folland, 1991; Wilks, 1991). Present paper endeavours to develop an ANN model to forecast average winter shower in India. The ANN technique is widely accepted as a potentially useful way of modeling complex non-linear and dynamic systems for which a large amount of data is used (Hornik, 1991; Hsieh and Tang, 1998). It is particularly useful, where the underlying physical processes or relationships are not fully understood or may display chaotic properties (Sivakumar, 2000). Conventional weather forecasting models are highly data specific and based on complex and expensive to maintain mathematical models that are built many months in advance of the events. ANN offers real prospects for an effective, more flexible, less assumption dependent adaptive methodology well suited for modeling weather forecasting which by its nature are inherently complex because of non-linearity and chaotic effects (Maqsood et al., 2002). The ANN is based on biological neural systems (Kartalopoulos, 2000). It is highly vital with respect to underlying data distribution and no assumptions are made about relationships between variables. The basic unit of ANN is the artificial neuron, which stimulates the four basic functions of natural neurons. The artificial neurons are much similar than biological neuron. The input layers consist of neurons that received input from the external environment. The outputs of the system layer communicate to the user about the external environment. There are usually numbers of hidden layers between these two layers. To determine the number of hidden neurons the network should have performed its best, are often left out to the method trial and error. Usually the available data are separated into training set and test set. The optimal weights are obtained by using 'Back Propagation'. The quality of the prediction is obtained from the performance with the test set of data. The weights are determined by iteration to produce the lowest error in the output. Initial weights are randomly assigned in subsequent iterations, individual weights are incrementally adjusted to reduce error (Kamarthi and Pittner, 1999).

Since the last few decades, ANN has opened up new avenues to the forecasting task involving atmosphere related phenomena (Gardner and Dorling, 1998; Hsieh and Tang, 1998). Michaelides et al. (1995) compared the performance of ANN with multiple linear regressions in estimating missing rainfall data over Cyprus. Kalogirou et al. (1997) implemented ANN to reconstruct the rainfall time series over Cyprus. Lee et al. (1998) applied ANN in rainfall prediction by splitting the available data into homogeneous subpopulations. Wong et al. (1999) constructed fuzzy rule bases with the aid of SOM and Back propagation neural networks and then with the help of the rule base developed predictive model for rainfall over Switzerland using spatial interpolation. Application of ANN in forecasting monsoon rainfall over India is not very new. We have surveyed the literature and could get some significant papers in this direction. Chattopadhyay (2007), Chattopadhyay and Chattopadhyay (2010a); Guhathakurta (2008); Sahai et al. (2000); Venkatesan et al. (1997) implemented ANN based methodologies in prediction of SWM rainfall over India. Iyengar and Raghu Kanth

(2005) decomposed the Indian monsoon rainfall time series into six intrinsic mode functions, and handled the non-linear part by ANN technique. In a study by Singh and Deo (2007) alternative forms of neural networks have been applied to forecast daily river flows on a continuous basis with the purpose of understanding how recent architectures like adaptive neuro fuzzy inference system (ANFIS), generalized regression neural network (GRNN) and radial basis function (RBF) and compared with traditional back propagation ANN, when monsoon-fed rivers involving significant statistical bias are involved. Banik et al. (2008) used ANN, ANFIS, and genetic algorithm (GA) for forecasting Bangladeshi monsoon rainfall and their findings suggest that ANFIS and GA approaches can be used more accurately than the other two selected approaches to forecast the Bangladeshi monsoon rainfall. Chakraverty and Gupta (2008) modelled southwest monsoon rainfall over India using ANN with different network configurations and predicted Indian southwest monsoon rainfall 6 years in advance. Karmakar et al. (2009) developed 3-layer perceptron feed forward back propagation ANN model on mean rainfall variable of 102-rainguage station of Chhattisgarh, India.

3. Autocorrelation structure of the time series

The existence of autocorrelation in meteorological and climatological data has important implications regarding the applicability of some standard statistical methods to atmospheric data. In particular, uncritical application of classical methods requiring independence of data within a sample will often give badly misleading results when applied to strongly persistent series (Wilks, 2006, page 59). The present article deals with the monthly average winter shower data and the SST anomaly data from 1950 to 1998. The rainfall and SST anomaly data have been collected from www.tropmet.res.in and http://jisao.washington.edu/. The winter shower (mm) and tropical SST anomaly (°C) has been examined for their autocorrelation pattern. We have calculated autocorrelation function (ACF) for the average (over November-January) rainfall time series up to 25 lags and displayed in Fig. 2a. This figure shows that the ACF does not follow any specific sinusoidal pattern and the autocorrelations fall below 0.3 in numerical value. Moreover, the ACF is not decaying to 0 with increase in the lag number. These features of the ACF indicate that the time series does not have any persistence and it is not stationary in nature. The ACF of the average NEM rainfall has been examined because this is the prediction of the modeling problem. The monthly rainfall amounts in November, December, and January are also examined for their ACF and similar nature is discerned (Fig. 2b-d). Similarly, the ACF are calculated for tropical SST anomaly in the said months (Fig. 3a-c) and it is revealed that ACF for the SST time series are remaining at positive level up to more than 10 lags. However, the ACF is lying below 0.5 in magnitude. This indicates that despite more positive association with past values than the NEM rainfall no persistence exists within the time series. Like the rainfall time series, the ACF for SST anomaly is not showing any sinusoidal pattern and hence it can be said that stationarity
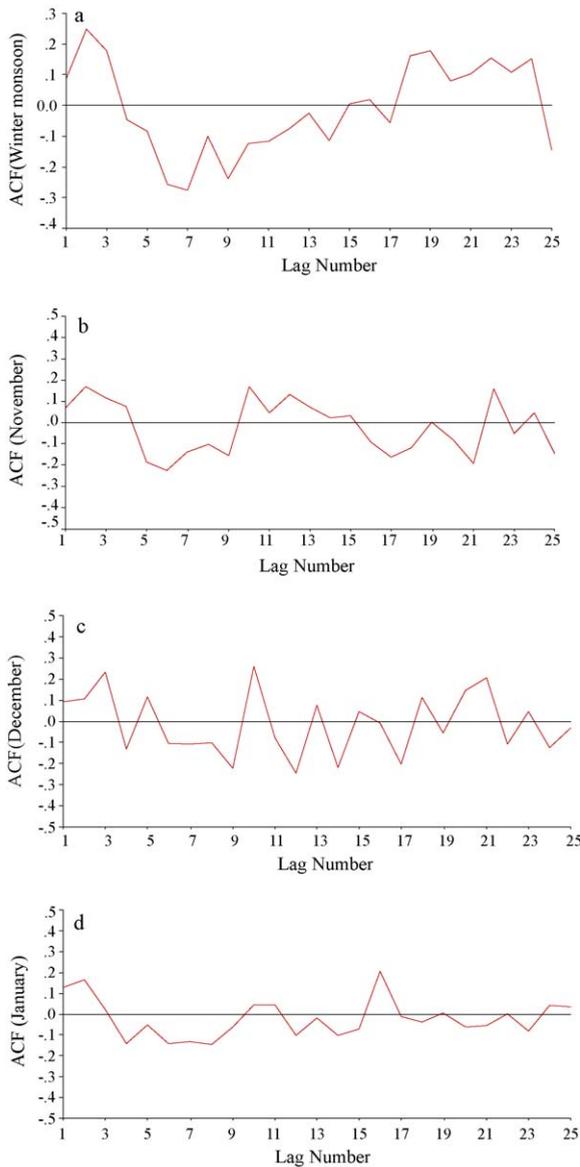
Fig. 2. (a). Autocorrelation function (ACF) for rainfall averaged over November, December, and January in India and during November (b), December (c), and January (d) over India.

Fig. 2. (a). Fonction d'auto-corrélation (ACF) pour la pluviosité moyennée sur novembre, décembre et janvier en Inde et pendant les mois de novembre (b), décembre (c) et janvier (d) sur l'Inde.
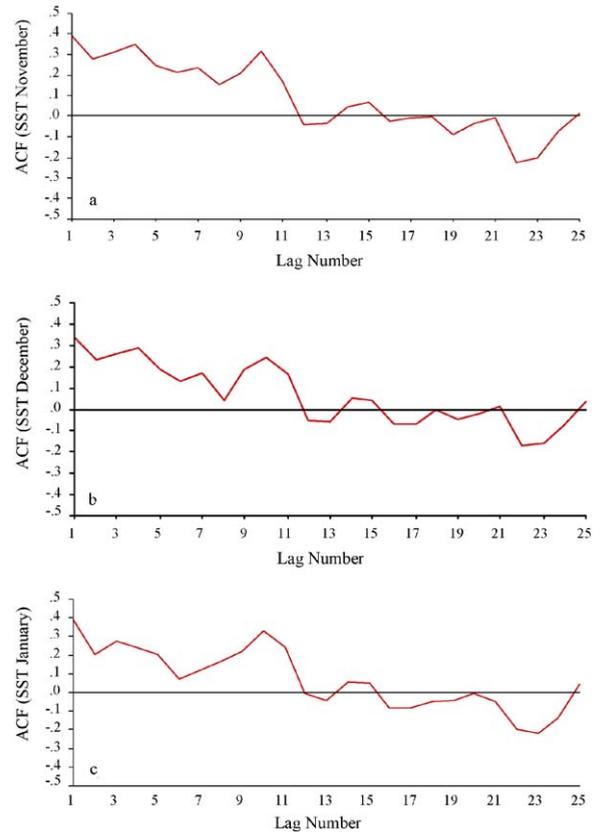

Fig. 3. Autocorrelation function (ACF) for tropical sea surface temperature (SST) anomaly during November (a), December (b), and January (c).

Fig. 3. Fonction d'auto-corrélation pour l'anomalie de température de la surface océanique (SST) tropicale pendant les mois de novembre (a), décembre (b) et janvier (c).

is not there with the SST anomaly time series. Thus, it can be said that neither the time series of the predictors (i.e. rainfall amounts and SST anomaly in November, December, and January for year n) nor the time series of the predictand (i.e. average NEM rainfall for year $(n+1)$) exhibits stationarity. The non-linear association between the predictors and the predictand is also apparent from the widely distributed data clouds in the scatterplot matrix presented in Fig. 4. The scatterplot matrix, which shows the arrangement of individual scatterplots, is a graphical extension of the correlation matrix. The physical arrange-

ment of the correlation coefficients in a correlation matrix is convenient for quick comparisons of relationships between pairs of variables, but distilling these relationships down to a single number such as a correlation coefficient inevitably hides important details (Wilks, 2006, page 65). In the scatterplot matrix presented in Fig. 4, it is found in all rows and columns that there is no linear association between the SST anomalies and the NEM rainfall amounts. It is apparent from the absence of any straight line like clouds of data pairs in the scatter diagrams. Even in the case of rainfall amounts or SST anomalies, there are no linear associations between the monthly observations. In Fig. 5, we have calculated the cross-correlation functions (CCF) between average SWM rainfall time series and SST anomalies for the three winter months under consideration. We have calculated by taking the predict and (average NEM rainfall) as fixed time series and time shifting the SST anomaly data set relative to the predictand. It is observed that there is no symmetry in the CCF plot about the horizontal (lag k) axis. This indicates that there is stable relationship between the two time series under consideration. Moreover, none of the cross coefficients are showing any high value, which indicates
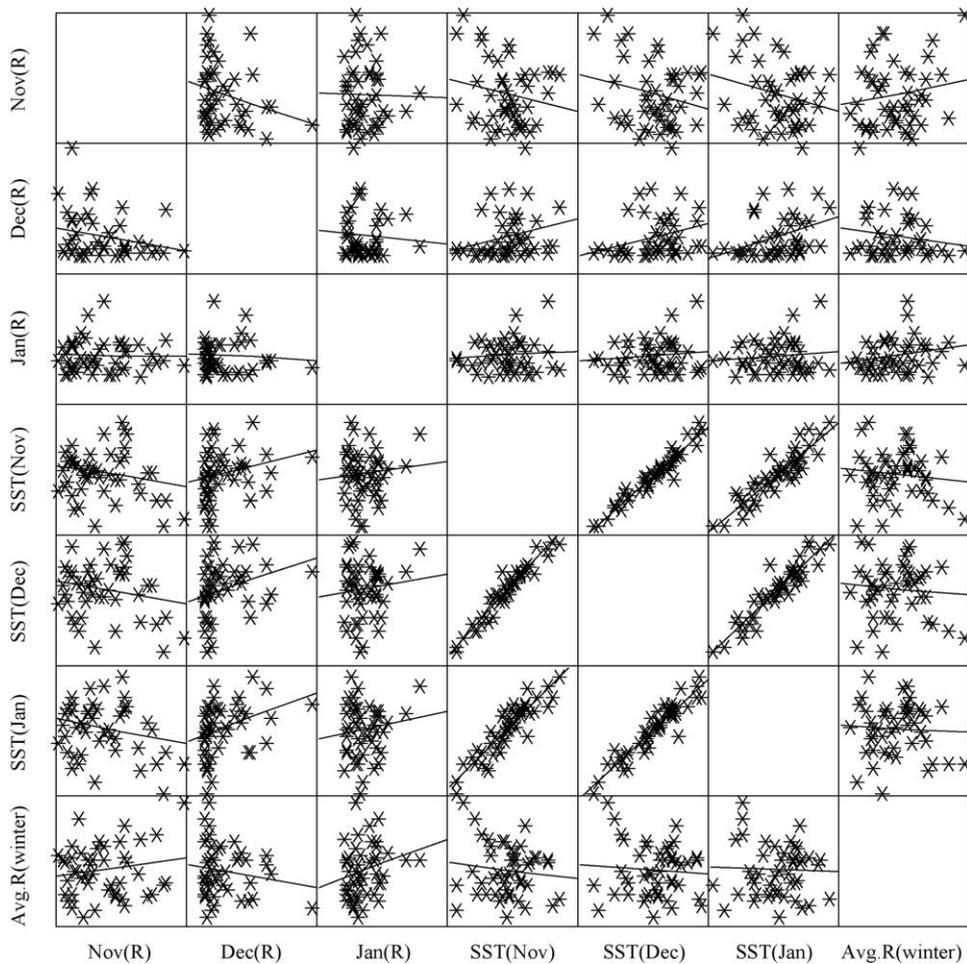
Fig. 4. Scatterplot showing the association between pairs of data considered in the study.

Fig. 4. Diagramme de points montrant l'association entre paires de données considérées lors de l'étude.

that there is no strong association between SST anomalies and NEM rainfall amounts. Computation of cross correlation has been performed using the software citable as Wessa (2010). The ACF, scatterplot, and CCF strongly indicate the requirement of ANN in the present prediction problem.

4. Theory

4.1. Multiple non-linear regression

Initially, all of the six predictors are tested over time for their trend. Linear as well as non-linear (polynomial of degree 6) trend equations (Wilks, 2006) are examined and the corresponding coefficients of determination (R^2) are calculated. We have presented the values of R^2 in Table 1, which shows that in all of the cases the values are small. This indicates the absence of any significant trend within the time series of all of the predictors. In the previous section, the non-stationarity of the time series under study has been revealed. It is also understandable from the ACF patterns that there are no seasonality and trend in the time

series. Non-linear regression equations are now generated to discern the relationship between the predictors and predictands. A non-linear regression equation involving exponential terms is generated as follows (Chattopadhyay and Chattopadhyay, 2008a):

$$\hat{y} = \sum_{i=1}^{6} a_i \exp(b_i x_i). \tag{1}$$

where, a_i, b_i are the regression parameters, and $x_i (i = 1, 2, ..., 6)$ denote the six predictors i.e. rainfall in November, December, January, and the tropical SST anomalies in November, December, and January respectively for the year n; in the left hand side $\hat{y}$ denotes the estimated average NEM rainfall in the year (n+1).

The regression parameters are estimated using Levenberg-Marquardt algorithm and iterating up to 500 steps. Finally, the regression parameters come out to be 0.355 (a_1), 0.031 (b_1), 4.705 (a_2), 0.018 (b_2), 2.505 (a_3), 0.056 (b_3), 4.730 (a_4), 0.001 (b_4), 3.343 (a_5), 0.001 (b_5), 3.480 (a_6), and 0.002 (b_6). The results would be analyzed statistically in the subsequent sections.
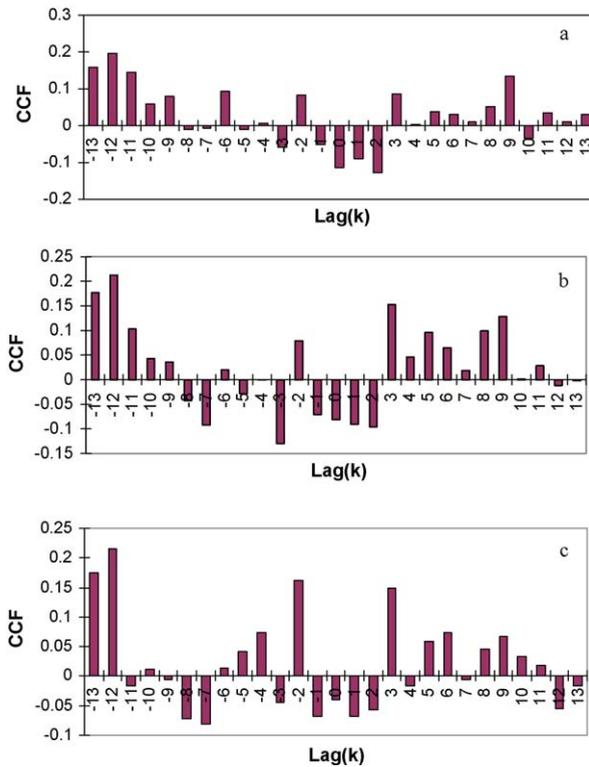
Fig. 5. The cross correlation function (CCF) between the average monsoon rainfall time series and SST anomalies for November (a), December (b) and January (c).

Fig. 5. Fonction de corrélation croisée (CCF) entre les séries temporelles moyennes des précipitations de mousson et les anomalies de température de la surface océanique (SST) pour les mois de novembre (a), décembre (b) et janvier (c).

Table 1
Coefficient of determination (R^2) of linear and polynomial trends.

Tableau 1
Coefficient de détermination (R^2) des tendances linéaire et polynomiale.

Parameters	R^2 for linear trend	R^2 for polynomial trend (degree 6)
November-SST	0.0029	0.1725
December-SST	0.0026	0.1592
January-SST	0.0052	0.1241
November-shower	0.0057	0.1837
December-shower	0.0822	0.1574
January-shower	0.024	0.0466

SST: sea surface temperature.

4.2. Artificial Neural Network

In the present paper, the data have been transformed to $[-1, +1]$ before applying them to ANN formation using the transformation

$$y_i = -1 + 2(x_t - x_{min})(x_{max} - x_{min})^{-1} \qquad (2)$$

This transformation is performed to get rid of the asymptotic effect arising from the sigmoid activation function to be used in the ANN model. A thorough discussion on the usefulness of scaling of data prior to ANN model generation is presented in Section 5 of Maier and Dandy (2000). An exhaustive variable selection procedure is adapted to find a good subset of the full set of input variables created as a result of data analysis and transformation. This selection has been done by means of genetic algorithm, where the cross over probability and mutation factor has been selected as 0.7 and 1, respectively, while generating the ANN model in the form of multilayer perceptron (MLP), using a modified cascade method together with an adaptive gradient learning rule (Lundin et al., 1999). The cascade mode of construction involves adding hidden nodes, one or more than one at a time, and always connecting all the previous nodes to the current node. The Cascade-Correlation architecture has several advantages over existing algorithms: it learns very quickly, the network determines its own size and topology; it retains the structures it has built even if the training set changes (Fahlman and Lebiere, 1990).

The adaptive gradient learning rule uses back-propagated gradient information to guide an iterative line search algorithm. Sigmoid function ($f(x) = (1 + e^{-x})^{-1}$) (Widrow and Lehr, 1990) has been used as activation function for both hidden and output layer. From the entire dataset under consideration, we have chosen 70 percent of the original data as training set. The training set formation has been done using the method of Round Robin (O'Neill and Song, 2003), which is an arrangement of choosing all elements in a group equally in some rational order, usually from the top to the bottom of a list and then starting again at the top of the list and so on. The root mean squared error (RMSE) has been used to evaluate the model. After training, the network has been validated over the entire dataset. It should be mentioned that there is no strict rule to decide the ratio of training and test cases. In a survey on the ANN literatures it is found that the ratios 1:1 (Chattopadhyay and Chattopadhyay, 2008b), 7:3 (Lundin et al., 1999), and 3:1 (Perez et al., 2000, Perez and Reyes, 2001) are frequently used in ANN applications. In the present paper the approach similar to that of Lundin et al. (1999) has been adopted after examining the other approaches. The ANN has been validated over the entire dataset under consideration. We have started with 30 nodes in the hidden layer and after training through adaptive gradient learning, we have obtained the final structure of the MLP as 5-12-1. It should be noted that this is the optimal architecture of the ANN. Number of hidden nodes has reduced to 12 because this has led to the minimum mean squared error. The error curve showing the evolution of the mean squared error with the number of epochs is shown in Fig. 6. This implies that there are five units (it is reduced from six due to exhaustive variable selection) in the final input set, twelve nodes in the final hidden layer, and there is only one output node. Results obtained from this network architecture are explained in the subsequent section.

4.3. Statistical assessment

In the present section, the results obtained in the last two sections would be analyzed. A confusion matrix
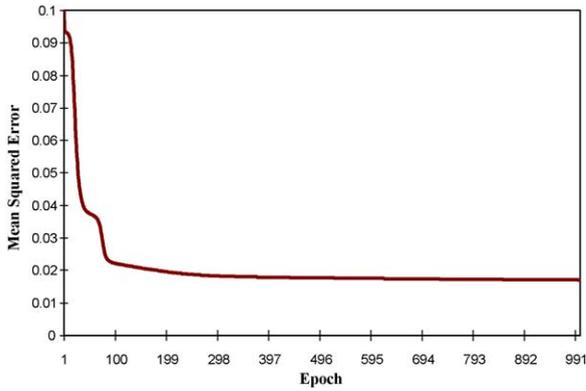
Fig. 6. Evolution of the mean squared error with epochs while training the artificial neural network (ANN).

Fig. 6. Évolution de l'erreur quadratique moyenne pendant le déroulement du réseau neuronal artificiel (ANN).

(Fielding and Bell, 1997), whose rows and columns represent the sub-ranges for the real world target and model outputs respectively, has been generated to view the prediction capacity of the ANN model. The value in the (i,j) position of the matrix is the number of records for which the real world target output is in the ith sub-range and whose real world model output is within the jth sub-

range. The confusion matrix for the ANN prediction comes out to be:

		Actual sub ranges					
	Confusion matrix	4.23	13.13	22.02	30.91	39.81	Totals
Predicted sub ranges	4.23	2	8	0	0	0	10
	13.13	2	10	3	1	0	16
	22.02	0	4	8	0	0	12
	30.91	0	1	2	2	1	6
	39.81	0	0	2	0	1	3
	Total	4	23	15	3	2	47

From the above matrix, it is found that for the sub range (4.23 mm to 13.13 mm) there are actually 23 observations and 16 predictions fall in this sub range. Thus, 69.57% accuracy is there for this sub range. For the sub range 13.13 mm to 22.02 mm, there are actually 15 observations and 12 predictions fall in this sub range. This means that there is 80% accuracy for this sub range. Thus, it can be interpreted that for higher NEM rainfall amounts there is more possibility of accurate forecast by ANN than in the case of lower rainfall amounts. Following Kuligowski and Barros (1998), we call the error of underforecast by dry bias and subsequently over forecast by wet bias. It is observed that there is lower possibility of dry bias than wet bias,
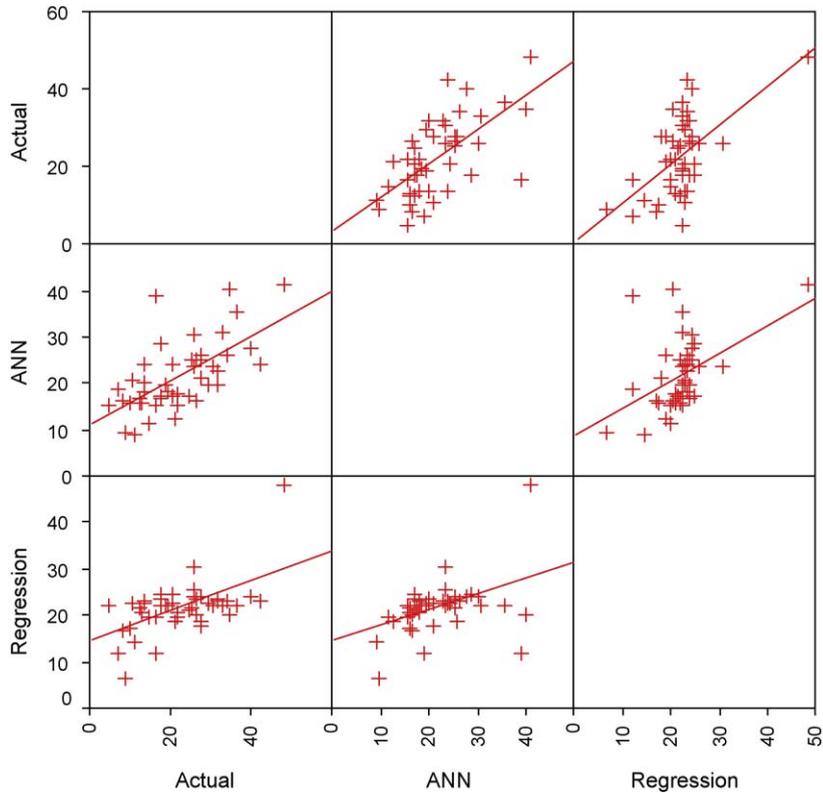

Fig. 7. Scatter plot showing the associations between actual rainfall amounts and those predicted by the artificial neural network (ANN) and non-linear regression.

Fig. 7. Diagramme de dispersion de points montrant les associations entre teneurs en précipitations actuelles et celles prédites par le réseau neuronal artificiel (ANN) et la régression non linéaire.
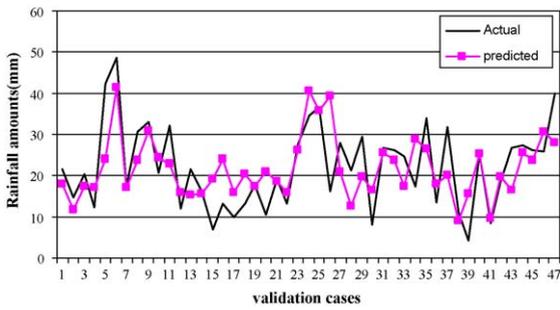
Fig. 8. Line diagram of the actual and predicted rainfall amounts for the validation cases of the artificial neural network (ANN).

Fig. 8. Tracé linéaire des teneurs en précipitations actuelles et prédites pour les cas de validation du réseau neuronal artificiel (ANN).

which indicates that the model tends to overforecast precipitation amount in some validation cases. This may be due to the presence of some heavy rainfall events during training of the ANN.

In the next step, the predictions are judged using the scatter plot matrix presented as Fig. 7, where it is apparent that the data cloud is well distributed around the linear trend line, thus it is understood that there is a good linear association between the actual and the ANN predicted values, which is further supported by the correlation coefficient of 0.641. However, the values deviated significantly from the trend line indicate that there are some cases where the prediction performance of the ANN is not satisfactory. Now, the predictions from ANN are judged case wise over the entire validation set. Line diagram presented in Fig. 8 shows the comparison between actual and predicted rainfall amounts for the validation cases. Computing the percentage of errors in the validation cases we find that if maximum 15% error of prediction is allowed, then prediction yield is 0.30. If 20, 25, and 30% errors of prediction are allowed, then the prediction yields are 0.45, 0.55, and 0.62 respectively.

Now we consider the predictions from the exponential regression presented in Eq. (1). Observing the scatter plot presented in Fig. 4, we find that there is a near-linear association between actual and regression-predicted values of the rainfall. To test the results statistically we calculate the Willmott's index (WI) for the regression as well as ANN. The Willmott's index is given by (Willmott, 1982):

$$WI = 1 - \left[\sum_{i=1}^{N}(P_i - O_i)^2 / \sum_{i=1}^{N}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2 \right] \quad (3)$$

where, O_i, P_i, and $\bar{O}$ denote the observed value in the ith case, predicted value in the ith case, and mean observed value respectively.

The utility of WI in judging the predictions has been discussed thoroughly in Chattopadhyay and Chattopadhyay (2008a). In the present case, the WI for non-linear regression and ANN come out to be 0.672 and 0.724 respectively. As the higher value of WI indicates better predictive model, it can be said that the ANN based predictive model performs better than non-linear regression. Further, we calculate percent error of the prediction (PE) given by (Chattopadhyay, 2007):

$$PE = \frac{\langle |P_i - O_i| \rangle}{\langle O_i \rangle} \times 100 \quad (4)$$

where, $\langle \rangle$ implies the average over the whole test set.

The predictive model is identified as a good one if the PE is sufficiently small. In the present case, PE for non-linear regression and ANN come out to be 30.96 and 27.16 respectively. This further proves that ANN is performing better than regression in predicting NEM rainfall. Like ANN, the prediction yields are also calculated for the non-linear regression based prediction. It is found that if 15, 20, 25, and 30% errors of prediction are allowed, then the prediction yields are 0.23, 0.30, 0.45, and 0.51 respectively, and each of the prediction yields is less than the corresponding values in the case of ANN. This further proves the better prediction potential of ANN than regression. For better viewing, the values of the statistical parameters are presented in Table 2.

It should be noted that there may be atmospheric problems where regression may perform better than ANN (e.g. Chattopadhyay and Chattopadhyay, 2010b). The physical mechanism that is behind the better performance of ANN than the exponential regression in the present problem may be the following:

- there is an unstable relationship (indicated by the cross-correlation function) between the SST anomalies and NEM rainfall, and it is well-established fact that ANN performs better than conventional regression in such cases where underlying processes are difficult to understand;
- use of cascade mode while training the MLP has its own advantages over the regression approach;
- as the predictor-predictand relationship is highly non-linear (indicated by the very small correlation coefficient), ANN has proved its well-established prediction potential over a non-linear regression model.

Table 2
Values of different statistical parameters used to assess the prediction potential of regression and artificial neural network (ANN) based models.

Tableau 2
Valeurs des différents paramètres statistiques utilisés pour évaluer le potentiel de régression prédit et les modèles basés sur le réseau neuronal artificiel (ANN).

Models	Willmott's index	Percentage error of prediction	Prediction yield (for 15% error)	Prediction yield (for 20% error)	Prediction yield (for 25% error)	Prediction yield (for 25% error)
Regression	0.67	30.96	0.23	0.30	0.45	0.51
ANN	0.72	27.16	0.30	0.45	0.55	0.62

5. Conclusion

The present article dealt with the prediction of NEM rainfall time series over India. The study compared Machine-learning approach namely ANN and a statistical approach namely exponential regression method for forecast of NEM rainfall using SST anomaly as a predictor. The purpose was to predict the average winter-monsoon rainfall using six predictors. To do the same, we considered the rainfall amounts and the SST anomalies for the NEM months of a given year to predict the average NEM rainfall of the next year. Consequently, we had six predictors and one predictand. The study also justified the use of non-linear methods for prediction of rainfall by analyzing autocorrelation structure of the data series. The results showed that ACF neither follows any specific sinusoidal pattern nor the autocorrelations fall below 0.3 in numerical value. These features of the ACF indicate that the time series does not have any persistence and it is not stationary in nature. Thus, highly non-linear trends and absence of any serial correlation in the time series under study prompted us to go for a non-linear regression equation in the form of exponential regression. To compare the results of the non-linear methods i.e. exponential regression and ANN, Willmott's index, Prediction Errors, Yield of Prediction and scatter plots were used. Higher value of Willmott's index for ANN (0.724) as compared to exponential regression (0.672) favours the ANN method for rainfall prediction. Lower value of Prediction Error for ANN (27%) as compared to exponential Regression (30%) recommends ANN approach over the other non-linear methods of prediction of rainfall. Average Yield of Prediction was 11% higher for ANN than regression method. Further, visual inspection of linearity in scatter plots (actual vs. predicted values) for each method using all the cases shows that there is apparently more linearity in ANN method as compared to the regression method hence ANN method should be used for predicting the rainfall. The results obtained in the paper are specifically valid for the data collected for India. In future, there is a scope to apply the similar methodology to other regions in the world and suggest general suitability of ANN over other methods for prediction of rainfall using SST anomaly.

Acknowledgements

References

Ananthakrishnan, R., 1977. Some aspects of the monsoon circulation and monsoon rainfall. Pure and Applied Geophysics 115, 1209–1249.

Annamalai, H., Liu, P., Xie, S.-P., 2005. Southwest Indian Ocean SST variability: its local effect and remote influence on Asian monsoons. Journal of Climate 18, 4150–4167.

Banik, S., Chanchary, F.H., Khan, K., Rouf, R.A., Anwer, M., 2008. Neural network and genetic algorithm approaches for forecasting bangladeshi monsoon rainfall. 11th International Conference on Computer and Information Technology, 735–740, doi:10.1109/ICCITECHN.2008.4802997.

Barsugli, J.J., Sardeshmukh, P.D., 2002. Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific basin. Journal of Climate 15, 34427–43442.

Bhanu Kumar, O.S.R.U., Naidu, C.V., Rao, S.R.L., 2004a. Influence of southern oscillation and SSTs over Nino-3.4 region on the winter monsoon rainfall over coastal Andhra Pradesh. Journal of Earth System Science 113, 313–319.

Bhanu Kumar, O.S.R.U., Naidu, C.V., Rao, S.R.L., Rao, B.R.S., 2004b. Prediction of southern Indian winter monsoon rainfall from September local upper-air temperatures. Meteorological Applications 11, 189–199.

Branstor, G., 1985. Analysis of general circulation model sea-surface temperature anomaly simulations using a linear model. Part I: forced solutions. Journal of Atmospheric Sciences 42, 2225–2241.

Bell, T.L., Suhasini, R., 1994. Principal modes of variation of rain-rate probability distribution. Journal of Applied Meteorology 33, 1067–1078.

Chakraverty, S., Gupta, P., 2008. Comparison of neural network configurations in the long-range forecast of southwest monsoon rainfall over India. Neural Computing and Applications 17, 187–192.

Chang, C.P., Harr, P., Ju, J., 2001. Possible Roles of Atlantic Circulations on the Weakening Indian Monsoon Rainfall–ENSO Relationship. Journal of Climate 14, 2376–2380.

Chattopadhyay, S., 2007. Feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India. Acta Geophysica 55, 369–382.

Chattopadhyay, S., Chattopadhyay, G., 2008a. Comparative study among different neural net learning algorithms applied to rainfall time series. Meteorological Applications 15, 273–280.

Chattopadhyay, S., Chattopadhyay, G., 2008b. Identification of the best hidden layer size for three layered neural net in predicting monsoon rainfall in India. Journal of Hydroinformatics 10, 181–188.

Chattopadhyay, S., Chattopadhyay, G., 2010a. Univariate modelling of summer-monsoon rainfall time series: comparison between ARIMA and ARNN. C. R. Geoscience 342, 100–107.

Chattopadhyay, G., Chattopadhyay, S., 2010b. Univariate approach to the monthly total ozone time series over Kolkata, India: autoregressive integrated moving average (ARIMA) and autoregressive neural network (AR-NN) models. International Journal of Remote Sensing 31, 575–583.

Chongyin, L., 1990. Interaction between anomalous winter monsoon in East Asia and El Niño events. Advances in Atmospheric Sciences 7, 36–46.

Clark, O.C., Cole, J.E., Webster, P.J., 2000. Indian Ocean SST and Indian summer monsoon rainfall: predictive relationships and their decadal variability. Journal of Climate 14, 2503–2519.

Deser, C., Phillips, A.S., Hurrell, J.W., 2004. Pacific Interdecadal Climate Variability: linkages between the Tropics and the North Pacific during Boreal Winter since 1900. Journal of Climate 17, 3109–3124.

Dhar, O.N., Rakhecha, P.R., 1983. Foreshadowing northeast monsoon rainfall over Tamilnadu India. Monthly Weather Review 111, 109–112.

Elsner, J., Tsonis, A., 1992. Nonlinear prediction, chaos, and noise. Bulletin of American Meteorological Society 73, 49–60.

Fahlman, S. E., Lebiere, C., 1990. The cascade-correlation learning architecture. In Advances in Neural information Processing Systems 2. In:Touretzky, DS (Ed.). Morgan Kaufmann Publishers, San Francisco, CA, 524–532.

Ferranti, L., Palmer, T.N., Molteni, F., Klinker, E., 1990. Tropical-extratropical interaction associated with the 30–60 day oscillation and its impact on medium and extended range prediction. Journal of Atmospheric Science 47, 2177–2199.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of rediction errors in conservation presence/absence models. Environmental Conservation 24, 38–49.

Gardner, M.W., Dorling, S.R., 1998. Artificial Neural Network (Multilayer Perceptron) – a review of applications in atmospheric sciences. Atmospheric Environment 32, 2627–2636.

Guhathakurta, P., 2008. Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic artificial neural network model. Meteorology and Atmospheric Physics 101, 93–108.

Guo, Y., Zhao, Y., Wang, J., 2002. Numerical simulation of the Relationships between the 1998 Yangtze river valley flood and SST anomalies. Advances in Atmospheric Sciences 19, 391–404.

Hartmann, H., Becker, S., King, L., 2007. Predicting summer rainfall in the Yangtze River basin with neural networks. International Journal of Climatology (on-line first), doi:10.1002/joc.1588.

Holton, J.R., 1972. An Introduction to Dynamic Meteorology. Academic Press, San Diego.

Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Network 4, 251–257.

Hsieh, W.W., Tang, T., 1998. Applying Neural Network Models to Prediction and Data Analysis in Meteorology and Oceanography. Bulletin of the American Meteorological Society 79, 1855–1869.

Iyengar, R.N., Raghu Kanth, S.T.G., 2005. Intrinsic mode functions and a strategy for forecasting Indian monsoon rainfall. Meteorology and Atmospheric Physics 90, 17–36.

Kalogirou, S.A., Constantinos, C.N., Michaelides, S.C., Schizas, C.N., 1997, A time series construction of precipitation records using Artificial Neural Networks, EUFIT'97, September 8–11, 2409–2413.

Kamarthi, S.V., Pittner, S., 1999. Accelerating neural network training using weight extrapolation. Neural Networks 12, 1285–1299.

Karmakar, S., Kowar, M. K., Guhathakurta, P., 2009. Spatial interpolation of rainfall variables using artificial neural network, Proceedings of the International Conference on Advances in Computing, Communication and Control, Mumbai, India, 547–552, ISBN:978-1-60558r-r351-8.

Kartalopoulos, S.V., 2000. Understanding Neural Networks and Fuzzy Logic – Basic Concepts and Applications. Prentice Hall, New Delhi.

Klein, S.A., Soden, B.J., Lau, N.C., 1999. Remote Sea Surface Temperature Variations during ENSO: evidence for a Tropical Atmospheric Bridge. Journal of Climate 12, 917–932.

Krishnamurthy, V., Goswami, B.N., 2000. Indian Monsoon – ENSO Relationship on Interdecadal Timescale. Journal of Climate 13, 579–595.

Kuligowski, R.J., Barros, A.P., 1998. Experiments in Short-Term Precipitation Forecasting Using Artificial Neural Networks. Monthly Weather Review 126, 470–482.

Lau, K.M., Weng, H., 2001. Coherent modes of global SST and summer rainfall over China: an assessment of the regional impacts of the 1997–98 El Niño. Journal of Climate 14, 1294–1308.

Lee, S., Cho, S., Wong, P.M., 1998. Rainfall prediction using Artificial Neural Network. Journal of Geographic Information and Decision Analysis 2, 233–242.

Li, F., Ramanathan, V. 2002. Winter to summer monsoon variation of aerosol optical depth over the tropical Indian Ocean. J. Geophys. Res. (Atmospheres) 107, pp. AAC 2-1, CiteID 4284, doi:10.1029/2001JD000949.

Lundin, M., Lundin, J., Burke, H.B., Toikkanen, S., Pylkkanen, L., Joensuu, H., 1999. Artificial Neural Networks applied to survival prediction in breast cancer. Oncology 57, 281–286.

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling and Software 15, 101–124.

Maqsood, I., Muhammad, R.K., Abraham, A., 2002. Neurocomputing Based Canadian Weather Analysis. Computational Intelligence and Applications. Dynamic Publishers Inc., USA, pp. 39–44.

Matthews, A.J., Hoskins, B.J., Masutani, M., 2004. The Global Response to Tropical Heating in the Madden-Julian Oscillation during Northern Winter. Q.J. Royal Met. Soc. 130, 1–20.

Michaelides, S.C., Neocleous, C.C., Schizas,C.N., 1995. Artificial Neural Networks and multiple linear regression in estimating missing rainfall data, Proc. DSP95 Intern. Confer. "Digital Signal Processing", Limassol, Cyprus. 668–673.

Murtugudde, R., Busalacchi, A.J., 1999. Interannual variability of the dynamics and thermodynamics of the tropical Indian Ocean. Journal of Climate 12, 2300–2326.

Nayagam, L.R., Janardanana, R., Mohana, H.S.R., 2009. India Variability and teleconnectivity of northeast monsoon rainfall over. Global and Planetary Change 69, 225–231.

Nicholls, N., 1983. Predicting Indian monsoon rainfall from sea surface temperature in the Indonesia-North Australia area. Nature 306, 576–577.

Nobre, P., Srukla, J., 1996. Variations of sea surface temperature, wind stress, and rainfall over the tropical Atlantic and South America. Journal of Climate 9, 2464–2479.

O'Neill, M.C., Song, L. 2003. Neural network analysis of lymphoma microarray data: prognosis and diagnosis near-perfect. BMR Bioinformatics 4, published online, http://www.biomedcentral.com/1471-2105/4/13 [Accessed 1 February 2009].

Perez, P., Reyes, J., 2001. Prediction of particulate air pollution using neural techniques. Neural Computing and Application 10, 165–171.

Perez, P., Trier, A., Reyes, J., 2000. Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago Chile. Atmospheric Environment 34, 1189–1196.

Qiaoping, L., Yihui, D., Wenjie, D., Guanhua, Y., 2007. A numerical study on the winter monsoon and cold surge over East Asia. Advances in Atmospheric Sciences 24, 664–678.

Raj, Y.E.A., 1996. Inter and intra-seasonal variation of thermodynamic parameters of the atmosphere over coastal Tamilnadu during northeast monsoon. Mausam 47, 259–268.

Rao, K.V., 1963. A study of the Indian northeast monsoon season. Indian Journal of Meteorology and Geophysics 14, 143–155.

Rao, G.N., 1999. Variations of the SO Relationship with Summer and Winter Monsoon Rainfall over India: 1872–1993. Journal of Climate 12, 3486–3495.

Reddy, P.R.C., Salvekar, P.S., 2003. Equatorial East Indian Ocean sea surface temperature: a new predictor for seasonal and annual rainfall. Current Science 85, 1600–1604.

Sahai, A.K., Soman, M.K., Satyan, V., 2000. All India summer monsoon rainfall prediction using an artficial neural network. Climate Dynamics 16, 291–302.

Sardeshmukh, P.D., Hoskins, B.J., 1988. The generation of global rotational flow by steady idealized tropical divergence. J. Atmos. Sci. 45, 1228–1251.

Shinoda, T., Hendon, H.H., 1998. Mixed layer modeling of intraseasonal variability in the tropical Pacific and Indian Oceans. Journal of Climate 11, 2668–2685.

Sivakumar, B., 2000. Chaos theory in hydrology: important issues and interpretations. Journal of Hydrology 227, 1–20.

Singh, O.P., 1995. Influence of Bay of Bengal on winter monsoon rainfall. Mausam 46, 307–312.

Singh, P., Deo, M.C., 2007. Suitability of different neural networks in daily flow forecasting. Applied Soft Computing 968–978.

Singh, S.V., Kripalani, R.H., 2007. Analysis of persistence in daily monsoon rainfall over India. International Journal of Climatology 6, 625–639.

Uvo, C.B., Repelli, C.A., Zebiak, S.E., Kushnir, Y., 1998. The Relationships between Tropical Pacific and Atlantic SST and Northeast Brazil Monthly Precipitation. J. Climate 11, 551–562.

Venkatesan, C., Raskar, S.D., Tambe, S.S., Kulkarni, B.D., Keshavamurty, R.N., 1997. Prediction of All India Summer Monsoon Rainfall Using Error-back-propagation Neural Networks. Meteorology and Atmospheric Physics 62, 225–240.

Wang, C., Wang, W., Wang, D., Wang, Q., 2006. Interannual variability of the South China Sea associated with El Nino. J. Geophys. Res. III, 1–19.

Ward, M.N., Folland, C., 1991. Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea-surface temperature. International Journal of Climatology 11, 711–745.

Webster, P.J., Palmer, T.N., 1997. The past and the future of El Niño. Nature 390, 562–564.

Wen, C., Graf, H.F., Ronghui, H., 2000. The nnterannual variability of East Asian winter monsoon and its relation to the summer monsoon. Advances in Atmospheric Sciences 17, 48–60.

Wessa, P., 2010. Free Statistics Software, Office for Research Development and Education, version 1.1.23-r5, URL: http://www.wessa.net/.

Widrow, B., Lehr, M.A., 1990. 30 years of Adaptive Neural Networks: perceptron, Madaline and Backpropagation. Proceedings of IEEE 78, 1415–1442.

Wilks, D.S., 2006. Statistical Methods in Atmospheric Sciences, Second Edition. Elsevier Inc.

Wilks, D.S., 1991. Representing serial correlation of meteorological events and forecasts in dynamic decision – analytic models. Monthly Weather Review 119, 1640–1662.

Willmott, C.J., 1982. Some Comments on the Evaluation of Model Performance. Bulletin of American Meteorological Society 63, 1309–1313.

Woolnough, S.J., Slingo, J.M., Hoskins, B.J., 2000. The Relationship between Convection and Sea Surface Temperature on Intraseasonal Timescales. Journal of Climate 13, 2086–2104.

Wong, K.W., Wong, P.M., Gedeon, T.D., Fung, C.C. 1999. Rainfall prediction using neural fuzzy technique, URL: www.it.murdoch.edu.au/~wong/publications/SIC97.pdf, pp. 213–221.

Wu, R., Kirtman, B.P., 2004. Impacts of the Indian Ocean on the Indian Summer Monsoon – ENSO Relationship. Journal Climate 17, 3037–3054.

Xavier, P.K., Marzin, C., Goswami, B.N., 2007. An objective definition of the Indian summer monsoon season and a new perspective on the ENSO-monsoon relationship. Quarterly Journal of Royal Meteorological Society 133, 749–764.

Yang, J, Liu, Q. and Liu, Z., 2010. Linking Asian monsoon to Indian Ocean SST in the observation: Possible roles of Indian Ocean Basin Mode and Dipole Mode, Journal of Climate, Early Online Release, doi:10.1175/2010JCLI2962.1, accessed on May 20, 2010.

Zhang, Y., Sperber, K.R., Boyle, J.S., Dix, M., Ferranti, L., Kitoh, A., Lau, K.M., Miyakoda, K., Randall, D., Takacs, L., Wetherald, R., 1997. East Asian winter monsoon: results from eight AMIP models. Climate Dynamics 13, 797–820.