Estimation of Thunderstorm Days from the Radio-sonde Observations at Kolkata (22.53³N, 88.33³E), India during Pre-monsoon Season: an ANN Based Approach

Pijush Basak, Dipu Sarkar and A. K. Mukhopadhyay
Narula Institute of Technology, 81, Nilgunj Road, Kolkata 700109, India.
E-mail: pijushbasak@rediffmail.com

Abstract

The focus of the present work aims at the forecast of formation of thunderstorm using the effective 20 parameters at Kolkata (22.53³N, 88.33³E), India during the pre-monsoon season (March-May). One hidden layer Artificial Neural Network (ANN) with variable learning rate back propagation algorithm is utilized for the purpose.

In the first phase of the work, in the region of active convective development, that is, from surface to the 500 hPa pressure level, the parameters selected are derived from wind speed, pressure, temperature, dew-point temperature, pressure at lifting condensation level of radio-sonde observations of 12 years (1985-1996). Thereafter, thunderstorm or otherwise days are predicted with ANN model with the observations of 3 years (1997-1999).

In the second phase of the analysis, the skill scores such as Probability of detection (POD), Heidke skill score (HSS), False alarm ratio (FAR) etc. are computed and the scores endorses the efficiency of the model. The squared errors are observed as 10⁻⁵. A 60-70% success rate is achieved in the analysis.

Keywords: Artificial Neural Network (ANN), Fair weather (FW), False alarm ratio (FAR), Heidke skill score (HSS), Multi-layer network (MLP), Probability of detection (POD), Thunderstorm (TS).

Introduction

Thunderstorm (TS) is a natural phenomenon that causes damage to the life and property and also has socio-economic impact on the society. Thunderstorm in the pre-monsoon period (Mar-May) is of great concern in the Gangatic plains of West Bengal including Kolkata. It is the source of water in the region in one hand and on the other hand, it has adverse effects on the inhabitants whenever it is severe.

During pre-monsoon season, the eastern and North Eastern parts of India, i.e., Assam, Orissa, Gangatic West Bengal and other parts are affected by high frequency of severe thunderstorms.
The moist warm southerly level flow from the Bay of Bengal and the cool dry westerly and North-Westerly upper level flow existing over the region favor exact climatological balance for the thunderstorms (Desai and Rao, 1954). The severity of thunderstorm is estimated by strong surface wind speed, heavy rain huge cloud mass and sometimes hail (Glickman, 2000; Chawdhury, 2006). Sometimes, the thunderstorms are associated with moderate squalls with wind velocity of 70-150 Km/h, low frequency of lightening and small cloud patches.

It may be emphasized that TSs are strongly favored by convective instability, abundant moisture at lower levels, strong wind shear and a dynamical lifting mechanism that can release instability (Kessler, 1982). Also, the vertical shear of the environmental winds has to match the value of convective instability for proper development of a large convective cloud (Asnani, 2005).

Atmospheric phenomenon such as TS is a complex system and thus no single technique can be most efficient for its occurrence (Lorenz, 1963). In order to assess the occurrence of thunderstorm, scientists used stability indices and/or atmospheric parameters in various cases (Ravi et al., 1999; Mukhopadhyay et. al., 2003; Dasgupta and De, 2001).

Many previous researchers utilized different multivariate techniques in different situations of atmosphere. The principal components based on covariance matrix and correlation matrix has been applied for a comparison of given data set of cyclone frequencies (Overland and Preisendorfer, 1982). To study principal anomaly in winter temperature, eigenvector methods have been applied by Diaz and Fulbright (1981). Sánchez et. al. (2008) studied short term thunderstorm in Argentina. To describe a multivariate statistical model for forecasting anomalies of surface pressure over Europe and North America, a composite Empirical orthogonal function (EOF) of monthly sea surface temperature (SST) and also of precipitation in the tropical Pacific Ocean region was performed (Weare, 1987). Cluster analysis (CL) and Linear discriminant analysis (LDA) have been comprehensively used for prediction of mean surface area (Maryon and Storey, 1985). Ward and Folland (1991) utilized both multiple linear regressions and LDA to forecast the rainfall and SST in northeast Brazil. A number of attempts were made to establish empirical models for prediction of atmospheric stability in connection with thunderstorm (Showalter, 1953; Darkow, 1968). Thiagarajan et al. (1995) relied on Markov chain model for daily rainfall prediction in Tamilnadu. Iyenger and Basak (1994) utilized Principal component analysis (PCA) to classify regions of India with respect to rainfall.

In Indian sub-continent, investigations connected with TS were undertaken by different scientists. Rao and Raman (1961) examined the frequency of thunder days in India. Chauhdhury (1961) and Sen and Basu (1961) discussed pre-monsoon TS in Assam, Tripura and Manipur. Mukherjee (1964) studied the thunderstorm pattern around Guwahati airport whereas Sahu (1964) examined thermodynamic conditions for TS over North-East India. Ravi et al. (1999) and Surendra Kumar (1972) utilized objective methods with parameter skills to pre-monsoon TS over Delhi and its neighborhood that yield reasonably good forecast. Dhawan et al. (2008) utilized atmospheric parameters and its skills to forecast TS in pre-monsoon season over North-East India. Lal (1990) forecasted severe pre-monsoonal convective activity over Lucknow. Hoddinot (1986) and Koteswaram and Srinivasan (1958) discussed different synoptic factors favorable for the formation of TS in Gangatic West Bengal.
Specifically, in the region of Gangatic West Bengal, namely, Kolkata and its suburbs, attempts were made to understand pre-monsoon TS pattern. Dasgupta and De (2001) utilized Markov-chain model for TS at Kolkata. Mukhopadhyay et. al. (2003) and Tyagi et. al. (2010) performed a significant job of forecasting thundery days utilizing the atmospheric parameters and a comprehensive study of its skills in the region. Ghose et. al (1999) applied LDA for prediction of thunderstorm days. Chatterjee et al. (2009) reduced the number of parameters responsible for TS over Kolkata utilizing LDA.

Scientists adopted sometimes computer technologies such as Artificial Neural Network (ANN) for the study of meteorological phenomenon to achieve perfection in forecasting. In the field of oceanography and climatology, one may refer to the works of Marzban and Witt (2001), Heish and Tang (1998) and Gardner and Dorling (1998). Chaudhuri (2006) and Ghanfarzadeh and Noghrehabadi (2009) predicted wind speed using the ANN network model. Specially, in the field of thunderstorm, McCann (1992) with multiple parameters as input utilized an ANN model for Thunderstorm prediction in USA. Collins and Tissot (2007) used ANN to forecast thunderstorm location (USA). In the region of Kolkata, the works of Chaudhury (2010) has showed efficiently how two atmospheric parameters, namely, convective available potential energy (CAPE) and convective inhibition (CINE) has been efficiently utilized to predict thunderstorm in the region.

In the present paper work, a neural network model with one hidden layer and a variable learning rate back-propagation algorithm is developed to forecast severe TSs in Kolkata. The model is trained with the set of thermodynamic and dynamic parameters obtained from the radio-sonde pre-monsoonal observations (Mar-May) of 12 years (1985-1996) and the validations are tested with the pre-monsoonal observations of 3 years (1997-1999).

**Formulation of the Problem**

The work of forecasting of thunderstorm consists of three sub-sections. Those are data, development of Artificial Neural Network (ANN) and skill score evaluation.

**Data:**

A thunderstorm (TS) occurring within the next 12 hours of the RS/RW observation taken at 0000 UTC (0530 IST) is considered as TS related to the ‘morning’ RS/RW. If not, it is treated as Fair-weather (FW) at the same RS/RW. Identical consideration for RS/RW observation taken at 1200 UTC is processed for classification levels linked to ‘afternoon’.

In the present experiment, only the air-mass between surface and 500 hPa is considered; it is due to several scientists mention this level as the level of cloud formation (Fujita et. al., 1970; Galway, 1956 and Miller, 1970). Between 500 hPa and surface levels, the other standard levels such as 850 hPa, 700 hPa and 600 hPa are also considered.
On many occasions, the data, either at one or more of the significant levels, that is approximately 1000 hPa (surface), 850 hPa, 700 hPa, 600 hPa or 500 hPa are not available. Consequently, those occasions are omitted from the data-set. These limitations have considerably reduced the data-size. The statistical indices (morning and afternoon) for forecasting the TS at Kolkata have been constructed utilizing all the available radio-sonde data for 12 years (1985-1996). The radio-sonde observations for 3 years (1997-1999) have been used to check the validation of indices. The corresponding number of observations are counted and presented in Table-1.

Table-1: Data-size RS/RW observations for the period 1985-1999 for 20 parameters.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of observations</td>
<td>Total</td>
</tr>
<tr>
<td>Morning (0000-1200 UTC)</td>
<td>TS</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>FW</td>
<td>372</td>
</tr>
<tr>
<td>Afternoon (1200-0000 UTC)</td>
<td>TS</td>
<td>197</td>
</tr>
<tr>
<td></td>
<td>FW</td>
<td>284</td>
</tr>
</tbody>
</table>

In the first stage, 20 parameters $P_i$ (i=1, 2, ..., 20) are formed from the radio-sonde observation of 12 years (1985-1996) for morning (0000-1200 UTC) and those have been involved in ANN. Those indices have been utilized for forecast of another 3 years (1997-1999) for the morning through the network.

The 20 parameters $P_i$ (i=1, 2, ..., 20) are constructed as follows:

- $P_1 = (\theta_{es} - \theta_e)$ at surface
- $P_2 = (P - P_{LCL})$ at surface
- $P_3 = \partial u / \partial z$ for the layer (surface – 850) hPa
- $P_4 = \partial \theta_e / \partial z$ for the layer (surface – 850) hPa
- $P_5 = \partial \theta_e / \partial z$ for the layer (surface – 850) hPa
- $P_6 = (\theta_{es} - \theta_e)$ at 850 hPa level
- $P_7 = (P - P_{LCL})$ at 850 hPa level
- $P_8 = \partial u / \partial z$ for the layer (850-700) hPa
- $P_9 = \partial \theta_e / \partial z$ for the layer (850 – 700) hPa
- $P_{10} = \partial u / \partial z$ for the layer (850 – 700) hPa
- $P_{11} = (\theta_{es} - \theta_e)$ at 700 hPa level
- $P_{12} = (P - P_{LCL})$ at 700 hPa level
- $P_{13} = \partial \theta_e / \partial z$ for the layer (700-600) hPa
- $P_{14} = \partial \theta_{es} / \partial z$ for the layer (700-600) hPa
- $P_{15} = \partial u / \partial z$ for the layer (700-600) hPa
- $P_{16} = (\theta_{es} - \theta_e)$ at 600 hPa level
- $P_{17} = (P - P_{LCL})$ at 600 hPa level
- $P_{18} = \partial \theta_e / \partial z$ for the layer (600-500) hPa
- $P_{19} = \partial \theta_{es} / \partial z$ for the layer (600-500) hPa
- $P_{20} = \partial u / \partial z$ for the layer (600-500) hPa
where $z$ and $u$ are the geo-potential height in meters, resultant wind speed expressed in $\text{ms}^{-1}$ respectively. Also, $\partial \theta_e / \partial z$ stands for conditional instability, $\partial \theta_e / \partial z$ for convective instability and $\partial u / \partial z$ for the vertical shear of horizontal wind. It is to be mentioned that values of $(\theta_{es} - \theta_e)$ and $(P - P_{LCL})$ at the lower level of each year have been treated as their respective values for that layer ($P$ and $P_{LCL}$ denoting pressure and pressure at lifting condensation level in hPa respectively).

It also should be noted that conditional instability is an essential criteria for supporting electrification (Williams and Renno, 1993). The thermodynamic parameter $(\theta_{es} - \theta_e)$ (Betts, 1974) is a measure of the instauration of the atmosphere. Moreover, it should be emphasized that the parameter $(P - P_{LCL})$ may be considered as the forcing factor for the saturation of the parcel (Kuo, 1965).

In the second stage of analysis, the 20 parameters ($P_i$; $i = 1, 2, \ldots, 20$) are formed from the radio-sonde observations of 12 years (1985-1996) for afternoon (1200-0000 UTC) and ANN network is performed for the afternoon. Consequently, the forecast for another 3 years (1997-1999) for afternoon are obtained.

**Development of ANN:**

The field of ANN has been very much versatile since its introduction in the 1980’s and has been widely used in almost every field in the last few years. A number of literatures are available in the area of different architectures (Hertz et al., 1991; Hecht-Nielsen, 1991 and Caudill, 1990); the most popular are the multi-layer feed-forward network. A feed-forward neural network can be considered as a transformation which maps a set of input variables into a set of output variables.

For a single layer network, these transformations can be expressed as:

$$v_k = \sum_{j=1}^{n} x_j \cdot w_{kj} + b_k$$

$$C_k = f(v_k)$$

where $x_j$ is the jth component of input vector $X$; $C_k$ is the kth component of the output vector $C$ while $W\{w_{kj}\}$ and $f(.)$ are called weight matrix and activation function respectively.

In case of prediction of the time evolution of a system, the components of input vector represent the value of a given variable of the system for a number of past measurements. The problems treated by means of feed-forward multi-layer networks are reasonably fast and a number of environmental problems has been analyzed (thunderstorm prediction McCann,1992; Collins and Tissot, 2007 and Chaudhury, 2010); daily solar radiation (Elizondo et al.,1994); El-Niño prediction (Derr and Slutz,1993); classification of solar radiation data (Liu and Xiao 1991; Foody et al.,1995), resource management (Gimblett and Ball, 1995), rainfall field (Kuligowski and Barros,1998 and Nath et al. 2008), wind speed (Ghanfarzadeh and Noghrehabadi, 2009) and many others.
A multi-layer network consists of one or more neurons inter-connected, with links among all the nodes in adjacent layers (Stanley, 1988). Each of the nodes was characterized by a sigmoidal or linear transfer functions on the hidden and output layers (Fig. 1). Each layer consists of one or more nodes represented in diagram by the small circles. The lines between the nodes indicate the flow of information from one node to the next. The links are characterized by connection strength of weights which are adjusted during the training of network and stored once network is trained. Two parameters may be varied easily to produce an optimum network solution, namely, the size of the hidden layer and the structure of transfer function. It may be mentioned that the weights linked with the hidden layer hold the ANN’s ‘knowledge’. If it is too small, it will not be able to identify; conversely, if it is too large, it would move towards generalizing the issue (Stanley, 1988). The number of neurons in the first layer is equal to the number of input variables and the numbers of neurons in the hidden layers may be varied. However, finding appropriate architecture (number of neurons in the hidden layer and transfer function type) needs trial and error method.

The training of the network is done using the error back-propagation training algorithm. It means that the error between the desired result and the result computed by the network is back-propagated through the network to adjust its weight. A gradient descend technique (Hertz et. al., 1991) is utilized for the purpose.

**Analysis with ANN:**

ANN utilized in our work has been formed in three layers. The first layer is the input layer where the selected parameters comes into ANN and varies as the number of parameters in the analysis; it is followed by a hidden layer and an output layer (Fig.1) where the output is produced. The input for ANN (1st layer) is the meteorological parameters prepared in earlier section named ‘data’ (20 neurons) and the output layer is the thunderstorm occurrence or non-occurrence (1 neuron). If the output neurons produce some value between 0 and 0.5, then TS is expected. Otherwise, Fair-weather is expected. In our network, number or neurons in the hidden layer are being varied and different combinations of transfer functions (namely, sigmoidal such as Tansigmoidal or ‘Tansig’, Logsigmoidal or ‘Logsig’ and purely linear such as ‘Purelin’) are also being utilized. For different number of neurons in the hidden layer as well as different transfer functions, the results are analyzed and the best one is presented in Table-2. The squared error is maintained at 10^-5.

In the first stage of the analysis, the parameters constructed from the radio-sonde observations of the morning (0000-1200 UTC) are analyzed. It has been noted that the same nature of transfer function for all the three layers gives the best result. The abbreviated designations TR1, TR2 and TR3 are being used when the transfer function for the three layers are ‘Tansig’, ‘Logsig’ and ‘Purelin’ respectively. The percentages of success are obtained for the designations TR1, TR2 and TR3. As an example, for TR1, the specific combination of neurons in the first, hidden and output layers, namely 20, 40 and 1 yields best percentage of success and is thus considered. The process is repeated for TR2 or TR3 respectively. The results thus obtained are detailed in Table-3.

In the second stage of the analysis, the parameters built from radio-sonde observation of afternoon (1200-0000 UTC) are considered and the process for the morning is repeated. The result so developed is presented in Table-4.
**Fig. 1:** Schematic Diagram of ANN.

**Table-2:** Properties of Developed ANN for Thunderstorm Forecasting.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Time</th>
<th>Number of parameters</th>
<th>Designation of transfer function</th>
<th>Nature of transfer function in layers 1, 2 and 3</th>
<th>Number of neurons in layers 1, 2 and 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Morning (0000-1200 UTC)</td>
<td>20</td>
<td>TR1</td>
<td>Tansig, Tansig, Tansig</td>
<td>20, 40, 1</td>
</tr>
<tr>
<td>1</td>
<td>Morning (0000-1200 UTC)</td>
<td>20</td>
<td>TR2</td>
<td>Logsig, Logsig, Logsig</td>
<td>20, 40, 1</td>
</tr>
<tr>
<td>1</td>
<td>Morning (0000-1200 UTC)</td>
<td>20</td>
<td>TR3</td>
<td>Purelin, Purelin, Purelin</td>
<td>20, 80, 1</td>
</tr>
<tr>
<td>2</td>
<td>Afternoon (1200-0000 UTC)</td>
<td>20</td>
<td>TR1</td>
<td>Tansig, Tansig, Tansig</td>
<td>20, 10, 1</td>
</tr>
<tr>
<td>2</td>
<td>Afternoon (1200-0000 UTC)</td>
<td>20</td>
<td>TR2</td>
<td>Logsig, Logsig, Logsig</td>
<td>20, 10, 1</td>
</tr>
<tr>
<td>2</td>
<td>Afternoon (1200-0000 UTC)</td>
<td>20</td>
<td>TR3</td>
<td>Purelin, Purelin, Purelin</td>
<td>20, 20, 1</td>
</tr>
</tbody>
</table>
Table-3: ANN analysis for TR1, TR2, TR3 for combination for morning (0000-1200 UTC).

<table>
<thead>
<tr>
<th>Nature of Trans. function</th>
<th>Nature of days</th>
<th>Number of days involved</th>
<th>Number of correct results</th>
<th>Percentage</th>
<th>Average percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR1</td>
<td>TS</td>
<td>47</td>
<td>23</td>
<td>48.93</td>
<td>72.66</td>
</tr>
<tr>
<td></td>
<td>FW</td>
<td>92</td>
<td>72</td>
<td>78.26</td>
<td></td>
</tr>
<tr>
<td>TR2</td>
<td>TS</td>
<td>47</td>
<td>23</td>
<td>48.94</td>
<td>68.34</td>
</tr>
<tr>
<td></td>
<td>FW</td>
<td>92</td>
<td>72</td>
<td>78.26</td>
<td></td>
</tr>
<tr>
<td>TR3</td>
<td>TS</td>
<td>47</td>
<td>22</td>
<td>46.81</td>
<td>72.66</td>
</tr>
<tr>
<td></td>
<td>FW</td>
<td>92</td>
<td>79</td>
<td>85.86</td>
<td></td>
</tr>
</tbody>
</table>

Table-4: ANN analyses for TR1, TR2, TR3 combination for afternoon (1200-0000 UTC).

<table>
<thead>
<tr>
<th>Nature of Trans. function</th>
<th>Nature of days</th>
<th>Number of days involved</th>
<th>Number of correct results</th>
<th>Percentage</th>
<th>Average percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR1</td>
<td>TS</td>
<td>58</td>
<td>30</td>
<td>51.72</td>
<td>61.90</td>
</tr>
<tr>
<td></td>
<td>FW</td>
<td>68</td>
<td>48</td>
<td>70.50</td>
<td></td>
</tr>
<tr>
<td>TR2</td>
<td>TS</td>
<td>58</td>
<td>28</td>
<td>48.28</td>
<td>61.90</td>
</tr>
<tr>
<td></td>
<td>FW</td>
<td>68</td>
<td>50</td>
<td>73.53</td>
<td></td>
</tr>
<tr>
<td>TR3</td>
<td>TS</td>
<td>58</td>
<td>33</td>
<td>56.89</td>
<td>60.39</td>
</tr>
<tr>
<td></td>
<td>FW</td>
<td>68</td>
<td>43</td>
<td>63.23</td>
<td></td>
</tr>
</tbody>
</table>

Skill – Score evaluation:

The forecast having been achieved, skill scores are computed to check up the efficiencies of the results. For each stage of the analysis, namely, stage 1 (morning 0000-1200 UTC) and stage 2 (afternoon 1200-0000 UTC) (Tables-3 and 4 respectively), the results are presented in the form of a 2 X 2 contingency table. The entries of the table are ‘correctly forecasted events (A)’, ‘events not correctly forecasted (B)’, ‘events forecasted but not observed (C)’ and ‘events not forecasted and also not observed (D)’. The presentation is shown in Table 5.

Based on these, nine skill scores, namely, Probability of Detection (POD), False Alarm Ration (FAR), Critical Success Index (CSI), True Skill Score (TSS), Heidke Skill Score (HSS), Miss Rate (MR), Correct Non-occurrence (C-NON), BIAS and Percentage of correct result (PC) are computed. Brief description of the skill scores are presented in Table 6.

It may be emphasized that a perfect forecast will show a HSS score of 1, a set of random forecast will be 0 and a lesser hits compared to the forecast by chance will have negative score. TSS and HSS both are seldom used in the literature as thunderstorm forecast skill parameters (Mukhopadhyay et al., 2003; Tyagi, et al., 2010); however, there seems to be a quite difference between their characteristics, namely, TSS pursues a high POD, HSS
attempts to reduce FAR to reasonable rate (Donaldson, et al., 1975). Also, TSS and HSS consider correct non-event forecast, whereas CSI does not. The limitations of TSS and HSS is that, if the number of correct forecast (A) and number of correct non-event forecasts (D) are inter-changed and number of misses (B) and also number of false alarms (C) are inter-changed, scores remain unchanged. But, CSI would change. Thus, no single forecast would give a complete picture (Mukhopadhyay et al., 2003; Tyagi et al., 2010). Whereas, we highly rely on POD and FAR, it is desirable to include CSI, POD, FAR, MR C-NON, BIAS, and PC in addition for broader and useful forecast.

Results of ANN Model and Discussion

One should note that we are using the parameters derived from RS/RW data obtained at 0000 UTC (morning) and 1200 UTC (afternoon). The occurrence/non-occurrence output is taken to be valid for the next 12 hours only (0000 UTC to 1200 UTC of the current day for morning and 1200 of the current day to 0000 UTC of the next day for afternoon). The performance of the ANN model for three cases is given in the subsequent sections.

Performance of ANN model for the morning:

The results of the stage 1 are given in Table-3. With the selected 20 parameters (thermodynamic and dynamic), the ANN analysis in the morning (0000-1200 UTC) yields 48.93%, 48.94% and 46.81% correct prediction for thunderstorm (TS) for TR1, TR2 and TR3 respectively and 78.56%, 78.26% and 85.86% for Fair-weather (FW) for the respective cases for verification of 3 years. On the whole, the overall percentages of correct prediction for morning are 72.66%, 68.34% and 72.66% for TR1, TR2 and TR3 respectively. In all, the ANN model can sense better the FW situation and sensing ability of TR1 and TR3 seems reasonably high.

Performance of ANN model for the afternoon:

In the next stage (afternoon: 1200-0000 UTC), the verification figures for correct prediction for TS are 51.72 %, 48.28% and 56.89% for TR1, TR2 and TR3 respectively and for FW, the corresponding figures are 70.50%, 73.53% and 63.23% respectively (Table-4) and the respective overall percentage of correct prediction for afternoon are 61.90%, 61.90% and 60.31% respectively. In this stage, the performances of TR1 and TR2 are comparatively better. Also, the sensing ability of FW is slightly high.

Both in the stages of morning (0000-1200 UTC) and afternoon (1200-0000 UTC), the ANN analysis points out to TR1 for comparatively higher percentage of correct prediction.

Results of Skill Scores

The overall forecast skills are reasonably consistent. The POD are reasonable good both for morning and afternoon (>0.5) (Table-7). The skills TSS and HSS for stages 1-2 (morning and afternoon) are also consistent (Table-7); almost equal scores occurred in case
of morning and afternoon (approx. 0.20-0.34 for both). The accuracy of MR and C-NON are also more or less same both for morning and afternoon (approx. 0.43-0.53 and 0.63-0.78 respectively) indicating that the model is not too aggressive in predicting thunderstorm. The CSI is reasonably same for all the cases (approx.0.36-0.43). Also, FAR is almost same for morning and afternoon combination (approx. 0.40) which is reasonable. However, PCs for the morning are of slightly higher side in the morning (around 70%) compared to 61% correct result in the afternoon.

**Table-5:** Contingency table of skill-scores.

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Events predicted</th>
<th>Events not predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>A (Hits)</td>
<td>B (Misses)</td>
</tr>
<tr>
<td>Events not observed</td>
<td>C (False Alarm)</td>
<td>D (Nonevent Hits)</td>
</tr>
</tbody>
</table>

**Table-6:** Description of different skill scores.

<table>
<thead>
<tr>
<th>Skill score</th>
<th>Code</th>
<th>References</th>
<th>Equation</th>
<th>Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>POD</td>
<td>Donaldson et al. (1975)</td>
<td>POD=A/(A+B)</td>
<td>0≤POD≤1</td>
</tr>
<tr>
<td>False alarm ratio</td>
<td>FAR</td>
<td>Donaldson et al. (1975)</td>
<td>FAR=C/(A+C)</td>
<td>0≤FAR≤1</td>
</tr>
<tr>
<td>Critical success index</td>
<td>CSI</td>
<td>Donaldson et al. (1975)</td>
<td>CSI= A/(A+B+C)</td>
<td>0≤CSI≤1</td>
</tr>
<tr>
<td>True skill statistics</td>
<td>TSS</td>
<td>Hanssen and Kuippers (1965)</td>
<td>TSS =(AD-BC)/(A+B)(C+D)</td>
<td>-1≤TSS≤1</td>
</tr>
<tr>
<td>Hiedke skill score</td>
<td>HSS</td>
<td>Brier and Allen (1952)</td>
<td>HSS= 2(AD-BC) / (B²+C²+ 2AD + (B+C).(A+D))</td>
<td>-1≤HSS≤1</td>
</tr>
<tr>
<td>Miss Rate</td>
<td>MR</td>
<td>-</td>
<td>B / (B+A)</td>
<td>0≤MR≤1</td>
</tr>
<tr>
<td>Correct Non-Ocurrence</td>
<td>C-Non</td>
<td>Dhawan et. al. (2008)</td>
<td>D / (D+C)</td>
<td>0≤C-Non≤1</td>
</tr>
<tr>
<td>Bias</td>
<td>BIAS</td>
<td>Dhawan et. al. (2008)</td>
<td>(A+C)/(A+B)</td>
<td>-</td>
</tr>
<tr>
<td>Percent Correct</td>
<td>PC</td>
<td>-</td>
<td>((A+D)/(A+B+C+D)) x 100</td>
<td>0≤PC≤100</td>
</tr>
</tbody>
</table>
Table 7: Table of Different Skill Scores.

<table>
<thead>
<tr>
<th></th>
<th>Morning (0000-1200 UTC)</th>
<th>Afternoon (1200-0000 UTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TR1</td>
<td>TR2</td>
</tr>
<tr>
<td>POD</td>
<td>0.4894</td>
<td>0.4895</td>
</tr>
<tr>
<td>FAR</td>
<td>0.4348</td>
<td>0.4348</td>
</tr>
<tr>
<td>MR</td>
<td>0.5106</td>
<td>0.5106</td>
</tr>
<tr>
<td>C-NON</td>
<td>0.7826</td>
<td>0.7826</td>
</tr>
<tr>
<td>CSI</td>
<td>0.3453</td>
<td>0.3433</td>
</tr>
<tr>
<td>TSS</td>
<td>0.2721</td>
<td>0.2721</td>
</tr>
<tr>
<td>HSS</td>
<td>0.2777</td>
<td>0.2717</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.9149</td>
<td>0.9149</td>
</tr>
<tr>
<td>Percent Correct</td>
<td>68.34%</td>
<td>68.34%</td>
</tr>
</tbody>
</table>

Conclusion

The above analysis reveals that, overall percentage of correct prediction of the ANN model under consideration are around 68-72% in the morning and around 60% in the afternoon which is consistent with the contemporary relevant works (Chatterjee et al., 2009; 58% for morning and 72% in the afternoon) and (Chaudhury et al., 2010; around 65% (with CAPE) and 80% (with CINE) overall correct result). Particularly, the model is most effective in predicting fair-weather (FW) situation amounting to around 70% (Tables 3 and 4). In the ANN model, the Tansigmoideal transfer function is prone to yield better result both for morning and afternoon. If we combine with the skill-score ability, the present analysis leads to the conclusion that TR1 (that is ‘Tansig’ in three layers) and TR3 (that is ‘Purelin’ in three layers) is best suited for predicting developments in the next morning/afternoon.

References

Estimation of Thunderstorm Days from the Radio-sonde Observations at Kolkata (22.53°N, 88.33°E), India during Pre-monsoon Season: an ANN Based Approach: Basak et al.


Gimblett, R. H. and Ball, G. L. (1995) Neural network architectures for monitoring and simulating changes in forest resource management. AI Appl., v. 9, 2, pp. 103-123.


