MAUSAM, **63**, 3 (July 2012), 423-432

551.558.1 : 551.510 : 551.553.21

Convective development at Kolkata (22.53° N, 88.33° E), India during pre-monsoon season using linear discriminant analysis technique

P. BASAK

Department of Mathematics, Narula Institute of Technology, Kolkata, India

(*Received 13 July 2010, Modified 3 May 2011*)

e mail : pijushbasak@rediffmail.com

सार – इस शोध पत्र में मानसून पूर्व के मौसम के समय (मार्च – मई) भारत के कोलकाता (22.53 डिग्री उ., 88.33 डिग्री प.) में 12 वर्षों (1985 – 1996) के आँकड़ों का उपयोग करते हुए संवहनीय विकास हेतु भिन्न–भिन्न प्रकार के अत्यधिक प्रभावी 22 एवं 20 ताप गतिकीय तथा गतिकीय प्राचलों की सांख्यिकीय अनुक्रमणिकाएँ तैयार की गई हैं।

0000 यू.टी.सी. पर रेडियो सौदें प्रेक्षणों से प्राप्त किए गए 22 आरंभिक चुनिंदा प्राचलों का उपयोग करके लीनियर डिस्क्रीमिनेन्ट एनालिसिस (एल. डी. ए.) नामक एक बहुचर सांख्यिकीय तकनीक द्वारा अगामी तीन वर्षों (1997 से 1999) तक आगामी 12 घंटों मे होने वाले संवहनीय विकास (सी. डी.) और साफ मौसम (एफ. डब्ल्यू.) के क्रमश 59.57 प्रतिशत और 58.70 प्रतिशत तक सही पूर्वानुमान दिए गए। 1200 यू. टी. सी. पर रेडियो सौदें से प्राप्त प्रेक्षणों के भी इसी प्रकार के विश्लेषण के द्वारा आगामी 12 घंटों में होने वाली सी.डी. और एफ.डब्ल्यू. के लिए क्रमशः 63.79 प्रतिशत एवं 50 प्रतिशत सही पूर्वानुमान दिए गए।

उपयुक्त अवधि के आँकड़ों के 20 प्राचलों आरंभिक सेट के मिलर (1972) और जॉर्ज (1960) आँकड़ों को छोड़कर का उपयोग करते हुए इसी प्रकार के अन्य एल. डी. ए. विश्लेषण किए गए जो 0000 यू. टी. सी. पर रेडियो सौदें प्रेक्षणों के आधार पर तैयार किए गए थे और आगामी 12 घंटों के प्रेक्षणों से पता चला है कि आगामी 3 वर्षों में सी. डी. और एफ. डब्ल्यू. के क्रमशः 63.83 प्रतिशत और 56.21 प्रतिशत सही पूर्वानुमान रहे थे। इसी प्रकार के परिणाम 1200 यू. टी. सी. पर आगामी 12 घंटों के क्रमशः 54.41 प्रतिशत और 67.34 प्रतिशत रहे।

संवहनीय उपलब्ध स्थितिज ऊर्जा (सी. ए. पी. ई.) और संवहनीय अवरोधन (सी. आई. एन.) नामक प्राचलों के प्रभाव को समझने के लिए इसी प्रकार का एल. डी. ए. विश्लेषण 17 प्राचल सेटों पर तैयार किया गया है इसमें प्रत्येक स्तर पर सी. ए. पी. ई. और सी. आई. एन. के अनुपात को शामिल किया गया है और θ_{es} - θ_e को छोड़कर यहाँ θ_{es} एवं θ_e क्रमशः संतृप्त समतुल्य स्थितिज तापमान और समतुल्य स्थितिज तापमान हैं। 0000 यू. टी. सी. पर लिए गए रेडियो सौंदे प्रेक्षणों से आगामी 3 वर्षों में 12 घंटों के लिए सी. डी. और एफ. डब्ल्यू. के क्रमशः 68.29 प्रतिशत एवं 54.43 प्रतिशत सही पूर्वानुमान किए गए ।अगले 1200 यू. टी. सी. पर रेडियो सौंदे प्रेक्षणों से आगामी 12 घंटों के लिए सी.डी. और एफ. डब्ल्यू. के लिए दिए गए सही पूर्वानुमानों का प्रतिशत क्रमशः 77.08 प्रतिशत तथा 57.44 प्रतिशत रहा।

प्राचलों के सभी 22, 20 एवं 17 सेटों के (0000 यू. टी. सी. तथा 1200 यू. टी. सी. दोनों प्रेक्षणों के आगामी 12 घंटों के लिए)दक्षता कौशल अंकों की गणना की गई है। जिन्हे ट्रू स्किल स्कोर (टी. एस. एस.), हाईडक स्किल स्कोर (एच. एस. एस.), क्रिटिकल सक्सेस इंडेक्स (सी. एस. आई.) के नाम से जाना जाता है। इसमें 17 प्राचल युग्म अत्यधिक उपयुक्त है और इनका उपयोग स्टीक पूर्वानुमान के लिए कुशलता के साथ किया जा सकता है।

आगामी 12 घंटों के लिए 0000 यू. टी. सी और 1200 यू. टी. सी. पर लिए गए प्रेक्षणों के अन्वेषण से सी. ए. पी. ई./सी. आई. एन. प्राचलों (17 प्राचल युग्म) विशेष रूप से 1200 यू. टी. सी. के प्रेक्षणों के शामिल किए जाने से 22 प्राचल, 20 प्राचल और 17 प्राचल सेटों की स्थिति में भी सुबह के पूर्वानुमानों की अपेक्षा अपराह्न के पूर्वानुमान अधिक सही पाए गए है।

ABSTRACT. In the present work, statistical indices are formed using the most effective different combinations of 22 and 20 thermodynamic and dynamic parameters for the convective development at Kolkata (22.53° N, 88.33° E), India during pre-monsoon season (Mar-May) utilizing the data of 12 years (1985-1996).

A multivariate statistical technique, namely Linear discriminant analysis (LDA) has been utilized to 22 primarily selected parameters derived from the radiosonde observations of 0000 UTC for next 12 hours yields respectively 59.57% and 58.70% correct prediction for convective development (CD) and fair-weather (FW) in next three years (1997 to 1999). A similar analysis for radiosonde observations of 1200 UTC for next 12 hours yields 63.79% and 50% for CD and FW respectively.

Another similar LDA analysis with the above data period utilizing 20 parameters [excluding Miller's (1972) & George's (1960) from the earlier set] built from the radiosonde observations of 0000 UTC for next 12 hours observation yield 63.83% & 56.21% correct prediction for CD and FW respectively in the next 3 years. The corresponding figures for 1200 UTC for next 12 hours are 54.41 % & 67.34% respectively.

With a view to understand the effect of the parameters, namely convective available potential energy (CAPE) and Convective inhibition (CIN), a similar LDA analysis has been applied to 17 parameter set (including the ratio of CAPE and CIN and excluding (θ_{es} -θ_e) at each level constructed form the above; θ_{es} and θ_e being saturated equivalent potential temperature and equivalent potential temperature respectively). The radiosonde observations of 0000 UTC for next 12 hours yield 68.29% and 54.43% correct prediction for CD and FW respectively in the next 3 years. Next, 1200 UTC radiosonde observations for next 12 hours yield 77.08% and 57.44% correct prediction for CD and FW respectively.

For all the 22, 20 and 17 set of parameters (both for 0000 UTC and 1200 UTC observations for next 12 hours), the efficient skill scores, namely, True skill score (TSS), Heidke skill score (HSS), Critical success index (CSI) are computed. The 17 parameter combination is most efficient and may be utilized in an effective manner for prediction purpose.

The investigations reveal that both for 0000 UTC and 1200 UTC observations for next 12 hours, correct prediction improves with the inclusion of the parameter CAPE/CIN (17 parameter combination), especially for 1200 UTC observations. Also, afternoon predictions are more effective than morning predictions for 22 parameter, 20 parameter and 17 parameter cases.

Key words – Linear disriminant analysis (LDA), Convective development (CD), Fair-weather (FW), Convective available potential energy (CAPE), Convective inhibition (CIN).

1. Introduction

The convective development (CD) (thunderstorm) occurring during March, April and May in Kolkata is a very important phenomenon in the Gangetic plains of West Bengal, India covering Kolkata. It is the source of natural water in the region and also has adverse effect on the inhabitants when the event becomes severe in nature in a localized region. So, prediction of any atmospheric phenomenon is always of ultimate interest to the researches.

The CD during March, April and May in Kolkata (22.53° N, 88.33° E), is usually known as pre-monsoon thunderstorm. Ghose *et al*. (1999) utilized Principal component analysis (PCA) and Linear discriminant analysis (LDA) to identify the significant parameters responsible for occurrence of CD at Kolkata.

In fact, many previous researchers utilized different multivariate techniques in different situations of atmosphere. To study principal anomaly in winter temperature, eigenvector methods have been applied by Diaz and Fulbright (1981). The principal components based on covariance matrix and correlation matrix have been applied for a comparison of given data set of cyclone frequencies (Overland and Preisendorfer 1982). To describe a multivariate statistical model for forecasting anomalies of surface pressure over Europe and North America, Cluster analysis (CL) and LDA have been comprehensively used (Maryon and Storey 1985). 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). Ward and Folland (1991) utilized both multiple linear regressions and LDA to forecast the rainfall and SST in north-east Brazil. A number of attempts were made to establish empirical models for prediction of atmospheric stability (Showalter 1953 and Darkow 1968).

Several multivariate statistical methods have been used by several researchers to establish different phenomenon in India. A good number of attempts have been made to predict the occurrence of rainfall by two-state Markov-chains (Dasgupta and De 2001; Pant and Shivhare 1998 and Thiagarajan *et al.*, 1995). Complex EOF was used to determine vertical wind profiles over Indian Ocean (Kistwal *et al.*, 1996). PCA has been applied by several scientists to understand the monsoon

TABLE 1

Data-size RS/RW observations for the period 1985-1999 for 22 and 20 parameters

Time		For construction of discriminant index (1985-1996)		For validation test (1997-1999)	
		No. of observations	Total	No. of observations	Total
Morning	FW	392	544	47	139
	CD	152		92	
Afternoon	FW	284	481	58	126
	CD	197		68	

rainfall (Iyenger and Basak 1994 and Sengupta and Basak 1998) where they have identified specified regions of India with respect to rainfall.

The CD has been comprehensively studied by several researchers. Chowdhury *et al.* (1996) analyzed pre-monsoon CD at Dhaka for 1983-1992. Haklander and Deldon (2003) studied CD predictors and then skills for the Netherlands. Huntrieser *et al.* (1997) compared traditional and newly developed atmospheric indices connected to CD for Switzerland: whilst Sãnchez *et. al.* (2008) successfully predicted short-term thunderstorm in Argentina. In North-East Europe, the CD was forecasted by thermodynamic indices and satellite data by Collier and Lille (1994). Six-hourly probability forecasts for Kansas (USA) and pre-convective summer thunderstorm over the Florida (USA) was investigated by Reap (1990) and Fuelberg and Bigger (1994).

In Indian sub-continent, investigations were undertaken by different scientists. Rao and Raman (1961) examined the frequency of thunder days in India. Chowdhury (1961) and Sen and Basu (1961) discussed pre-monsoon CD in Assam, Tripura and Manipur. Mukherjee (1964) studied the thunderstorm pattern around Guwahati airport whereas Sahu (1996) examined thermodynamic conditions for thunderstorm over North-East India. Lal (1990) forecasted severe pre-monsoonal convective activity over Lucknow. Ravi *et al.* (1999) and Kumar (1972) utilized objective methods with parameter skills to pre-monsoon thunderstorm over Delhi and its neighborhood that yield reasonably good forecast. Hoddinot (1986) and Koteswaram and Srinivasan (1958) discussed different synoptic factors favorable of CD in Gangetic West Bengal.

Particularly, in Gangetic West Bengal, around the city of Kolkata, several attempts made to understand convective development pattern. Mukhopadhyay *et al.* (2003) and Tyagi *et al.* (2010) performed a pioneering job of forecasting thundery days utilizing the atmospheric parameters and a comprehensive study of its skills in the region. Chatterjee *et al.* (2009) reduced the number of parameters responsible for Kolkata utilizing Linear discriminant technique.

It may be emphasized that CDs are accelerated by convective instability, abundant moisture at lower levels, strong wind shear and a dynamical lifting mechanism that can release instability (Kessler 1982). Moreover, the vertical shear of the environmental winds has to match the value of convective instability of proper development of a large convective cloud (Asnani 2005). It also should be noted that conditional instability is an essential criteria for supporting electrification (Williams and Renno 1993). Apart from that, the thermodynamic parameter (θ_{es} - θ_e) (Betts 1974) is a measure of the insaturation of the atmosphere and equivalent potential θ_{es} and θ_e denote the saturated equivalent potential temperature and equivalent potential temperature respectively. Lastly, 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) (where P and P_{LCL} are level pressure and pressure at corresponding lifting condensation level respectively).

In this paper, an attempt has been made to identify the set of thermodynamic and dynamic parameters that can most successfully forecast the occurrence of CD in Kolkata, India.

2. Data

A convective development (CD) occurring within the next 12 hours of the RS/RW observation taken at 0000 UTC (0530 IST) is considered as CD related with 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 afternoon.

In the present experiment, only the air-mass between surface and 500 hPa is considered. The justified reason is that several scientists mention this level as the level of cloud formation (Galway 1956; Fujita *et al.*, 1970 and

TABLE 2

Data-size RS/RW observations for the period 1985-1999 for 17 parameters

Time		For construction of discriminant index (1985-1996)		For validation test (1997-1999)	
		No. of observations	Total	No. of observations	Total
Morning	FW	314	445	41	120
	CD	131		79	
Afternoon	FW	228	393	48	102
	CD	165		54	

Miller 1972). 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 CD at Kolkata have been constructed utilizing all the available radiosonde data for 12 years (1985-1996). The radiosonde 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.

In the first stage, 22 parameters, P_i (i = 1, 2,…, 22) are formed from the radiosonde observations of 12 years (1985-1996) and those have been utilized for Linear discriminant analysis (LDA) both for morning and afternoon separately to the radiosonde observations for 3 years (1997-1999).

The 22 parameters P_i (i = 1, 2,….., 22) are constructed as follows:

P_1 = $(\theta_{es} - \theta_e)$ at surface
P_2 = $(P - P_{LCL})$ at surface
P_3 = $\partial\theta_e/\partial z$ for the layer (1000 – 850) hPa
P_4 = $\partial\theta_{es}/\partial z$ for the layer (1000 – 850) hPa
P_5 = $\partial u/\partial z$ for the layer (1000 – 850) hPa
P_6 = $(\theta_{es} - \theta_e)$ at 850 hPa level
P_7 = $(P - P_{LCL})$ at 850 hPa level
P_8 = $\partial\theta_e/\partial z$ for the layer (850-700) hPa
P_9 = $\partial\theta_{es}/\partial z$ 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
P_{21} = $(T_{850} - T_{500}) + (T_{d850} - T_{500})$ (Miller's total index)
P_{22} = $(T_{850} - T_{500}) + T_{d850} - (T_{700} - T_{d700})$ (George's K-index)

where z, u, T and T_d are the geo-potential height in meters, resultant wind speed expressed in ms^{-1}, dry bulb temperature and dew point temperature in Kelvin respectively. Also, $\partial\theta_{es}/\partial z$ 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 worth mentioning that values of $(\theta_{es}-\theta_e)$ and $(P-P_{LCL})$ at the lower level of each layer have been treated as their respective values for that layer.

In the second stage of analysis, 20 parameters (P_i; i = 1, 2,…, 20) are considered. For construction of the parameters at this stage, P_{21} and P_{22} are removed from the parameter set of stage 1 and the remaining 20 parameters are retained.

In the last stage of the analysis, for all layers, the parameters $(\theta_{es}-\theta_e)$, the forcing factors of the layers are removed from the parameter set of second stage; in addition, the ratio of two parameters, namely, Convective available potential energy (CAPE) and Convective inhibition (CIN) are introduced to identify their effect, if any. Consequently, 17 parameters are retained in this last stage. Due to non-availability of few CAPE and CIN data, the data-size of 17 parameter case is inadvertently reduced. It is given separately in Table 2.

It may be emphasized that CAPE indicates the amount of energy (or positive buoyancy) of a parcel of air if lifted a certain distance vertically and CIN indicates the amount of energy that would prevent an air parcel from arising from the surface to the level of free convection (Donaldson *et. al.*, 1975).

TABLE 3

Discriminant functions for 22, 20 and 17 parameters

Time	Nature of days for auto-verification	Number of parameters	Number of days involved	Discriminant functions
Morning	FW	22	392	MD_X = -82.122
	CD	22	152	MD_Y = -82.943
	FW	20	392	MD_X = -2.255
	CD	20	152	MD_Y = -1.682
	FW	17	314	MD_X = -2.144
	CD	17	131	MD_Y = -2.942
Afternoon	FW	22	284	AD_X = -3.594
	CD	22	197	AD_X = -2.278
	FW	20	284	AD_X = -0.025
	CD	20	197	AD_X = -1.568
	FW	17	228	AD_X = -2.094
	CD	17	165	AD_X = -2.274

3. Methodology

In this section, the basics of the multivariate of technique, namely, Linear discriminant analysis (LDA) has been discussed in short. It is followed by the result of the analysis.

3.1. *Linear discriminant analysis (LDA)*

We consider the two sets of observations $X = [X_{ij}]$, (i = 1, 2,…,k and j = 1, 2,…,m) and $Y = [Y_{ij}]$, (i = 1, 2, …k and j = 1, 2, …, n) where i, j stands for number of parameters (k in each case) and number of days (m and n in two cases) respectively. X and Y are the group-symbols of convective development (CD) and fair-weather (FW) respectively.

The groups X and Y are arranged as follows:

$$X = \left[\overline{X}_1, \overline{X}_2, ..., \overline{X}_k\right]; \quad \overline{X}_i = \left(\frac{1}{m}\right)\sum_{j=1}^{m} X_{ij}$$

$$Y = \left[\overline{Y}_1, \overline{Y}_2, ..., \overline{Y}_k\right]; \quad \overline{Y}_i = \left(\frac{1}{n}\right)\sum_{j=1}^{n} Y_{ij}$$

The covariance matrices of each group are as follows:

$$S_x = \left[S_x(i, j)\right]_{k\times k}$$

where $S_x(i, j) = [1/(m-1)]\sum_{p=1}^{m} \left(X_{ip} - \overline{X}_i\right)\left(X_{jp} - \overline{X}_j\right)$

$$S_y = \left[S_y(i, j)\right]_{k\times k}$$

where $S_y(i, j) = [1/(n-1)]\sum_{p=1}^{n} \left(Y_{ip} - \overline{Y}_i\right)\left(Y_{jp} - \overline{Y}_j\right)$

In the analysis, without any loss of generosity, it is assumed that the population in each of the groups have same covariance matrix; the pooled estimate of the dispersion of data around their means are

$$S = [1/(m + n-2)].[(m-1)S_x + (n-1)S_y]$$

We now verify the nature of the unknown group $U = [U_{ij}]$ (i = 1, 2,..,k and j = 1, 2, ..,l) ; i, j standing for number of parameters and number of days respectively. The discriminant functions of X, Y and U are

$$D_x = \overline{X}' S^{-1}\left(\overline{X} - \overline{Y}\right)$$

$$D_y = \overline{Y}' S^{-1}\left(\overline{X} - \overline{Y}\right)$$

$$D_u = \overline{U}' S^{-1}\left(\overline{X} - \overline{Y}\right)$$

where dash denotes the transpose of the matrix.

If $| D_x - D_u | < | D_y - D_u |$, then U belongs to the X-group, that is, the nature of unknown days to be of the nature of CD day.

If $| D_x - D_u | > | D_y - D_u |$, then U belongs to Y-group, that is, the nature of unknown day is identical to the nature of FW day.

TABLE 4.1

LDA analysis for stage 1 for 22 parameter combination (morning)

Time	Nature of days	Number of variables	Number of days involved	Number of correct results	Percentage of success	Average percentage of success
Morning	CD	22	47	28	59.57	58.99
0000-1200 UTC	FW	22	92	54	58.70	

TABLE 4.2

LDA analysis for stage 2 for 20 parameter combination (morning)

Time	Nature of days	Number of variables	Number of days involved	Number of correct results	Percentage of success	Average percentage of success
Morning	CD	20	47	30	63.83	58.99
0000-1200 UTC	FW	20	92	52	56.21	

TABLE 4.3

LDA analysis for stage 3 for 17 parameter combination (morning)

Time	Nature of days	Number of variables	Number of days involved	Number of correct results	Percentage of success	Average percentage of success
Morning	CD	17	41	28	68.29	59.16
0000-1200 UTC	FW	17	79	43	54.43	

4. Analysis

The analysis has been performed in 3 stages: namely, 22 parameters, 20 parameters and 17 parameters respectively.

In the first stage, the LDA technique is applied to the matrices X and Y where X and Y contain the original 22 parameters P_i (i = 1, 2,.., 22) as mentioned in the earlier section. X consists of the parameters of Convective development (CD) and Y consists of those of fair-weather (FW) days. The work has been done for morning and afternoon separately.

The discriminant function for the morning is denoted by MD_X and MD_Y (Table 4.1) for the convective (CD) and fair-weather (FW) respectively. The indices MD_X and MD_Y are constructed utilizing RS/RW data of 0000 UTC for Kolkata for pre-monsoon period of 12 years (1985-1996). The dimensions of X and Y matrices are 392 × 22 and 152 × 22 (*e.g.*, $x.y$; x = number of days; y = number of parameters (Table 3). These MD_X and MD_Y are used to predict the nature of days of Unknown System (US) from the set of 3 years (1997-1999). The results are presented in Table 4.1.

The same procedure has been applied to the data-set of 1200 UTC for CD and FW situations respectively during 12 years (1985-1996) (data-matrix sizes are 284 × 22 and 197 × 22 respectively). The corresponding discriminant functions have been denoted by AD_X and AD_Y for CD days and FW days respectively (Table 3). Utilizing those AD_X and AD_Y, the result of the prediction from the nature of days of the US are presented in Table 5.1.

In the second stage of the analysis, the procedures discussed in the first stage are repeated with the number of parameters now being reduced to 20. From the first stage 22 parameters, the parameters P_{21} and P_{22} are eliminated with some perception of better result. The data matrix sizes are presented in Table 1. The corresponding MD_X, MD_Y, AD_X and AD_Y are presented in Table 3 and the result of US are presented in Table 4.2 (morning) and Table 5.2 (afternoon).

In the third stage of the analysis, the number of parameters is further reduced. At each level, namely, approximately 1000 hPa (surface), 850 hPa, 700 hPa, 600 hPa and 500 hPa, the forcing factor parameters ($\theta_{es} - \theta_e$) are eliminated; instead, ratio of the parameters convective

TABLE 5.1

LDA analysis for stage 1 for 22 parameter combination (afternoon)

Time	Nature of days	Number of variables	Number of days involved	Number of correct results	Percentage of success	Average percentage of success
Afternoon	CD	22	58	37	63.79	56.34
1200-0000 UTC	FW	22	68	34	50.00	

TABLE 5.2

LDA analysis for stage 2 for 20 parameter combination (afternoon)

Time	Nature of days	Number of variables	Number of days involved	Number of correct results	Percentage of success	Average percentage of success
Afternoon	CD	20	58	39	54.41	60.31
1200-0000 UTC	FW	20	68	37	67.34	

TABLE 5.3

LDA analysis for stage 3 for 17 parameter combination (afternoon)

Time	Nature of days	Number of variables	Number of days involved	Number of correct results	Percentage of success	Average percentage of success
Afternoon	CD	17	48	37	77.08	70.00
1200-0000 UTC	FW	17	54	31	57.44	

available potential energy (CAPE) and convective inhibition (CIN) is introduced. It may be emphasized that CAPE is the positive buoyancy of an air parcel and is an indicator of atmospheric instability; whereas CIN is the measure that indicates the amount of energy that will prevent an air parcel from rising from surface to the level of free convection (Hanssen and Kuippers 1965). The count of parameters is now 17. In this particular stage, due to non-availability of a few CAPE and CIN data, the data-size is further reduced (Table 2). The concerned discriminant functions and the corresponding MD_X, MD_Y, AD_X and AD_Y are presented in Table 3. The result of the US system is presented in Tables 4.3 (morning) and 5.3 (afternoon).

5. Results and discussion

5.1. *Morning (0000-1200 UTC)*

In the morning, with the 22 parameters (thermodynamic and dynamic), the LDA analysis yields 59.57% correct prediction (28 out of 47 cases) for CD and 58.70% (54 out of 92 cases) for FW (Table 4.1) for verification of 3 years for the next 12 hours (stage1). The corresponding figures for 20 parameters (stage 2) are 63.83% (30 out of 47 cases) and 56.21% (52 out of 92 cases) respectively (Table 4.2; stage 2). In the third stage with 17 parameters (with CAPE/CIN), the verification for CD and FW respectively yield 68.29% (28 out of 41 cases) and 54.43% (43 out of 79 cases) correct prediction (Table 4.3). The overall percentages of correct prediction for the 22, 20 and 17 parameter combination are 58.99, 58.99 and 59.16 respectively. The analysis revealed that 17 parameter combinations which replace the parameter (θ_{es} - θ_e) at each level with the ratio of CAPE and CIN (connected with the ratio of positive buoyancy and negative energy value) show the maximum occurrence of correct result. Thus, without neglecting the contribution of the other parameters, it may be stated that, the above mentioned 17 parameters may be considered sufficient for operational purpose in the morning.

TABLE 6

Contingency table of skill-scores

Observation	Prediction	
	Events predicted	Events not predicted
Event observed	A (Hits)	B (Misses)
Events not observed	C (False Alarm)	D (Non-event Hits)

TABLE 7

Description of different skill scores

Skill score	Code	References	Equation	Limits
Probability	POD	Donaldson *et al*. (1975)	POD = A/(A+B)	0≤POD≤1
False alarm ratio	FAR	Donaldson *et al*. (1975)	FAR = C/(A+C)	0≤FAR≤1
Critical success index	CSI	Donaldson *et al*. (1975)	CSI = A/(A+B+C)	0≤CSI≤1
True skill statistics	TSS	Hanssen and Kuippers (1965)	TSS = (A/A+B)-(C/C+D) = (AD-BC)/(A+B)(C+D)	-1≤TSS≤1
Hiedke skill score	HSS	Brier and Allen (1952)	HSS = (CF-E)/(N-E) = 2(AD-BC)/(A+B) (B+D) + (A+C)(C+D)	-1≤HSS≤1
Miss Rate	MR	-	B / (B+A)	0≤MR≤1
Correct Non-Occurrence	C-Non	Dhawan *et. al*. (2008)	D / (D+C)	0≤C-Non≤1
Bias	BIAS	Dhawan *et. al*.(2008)	(A+C) / (A+B)	-
Percent Correct	PC	-	[(A+D) / (A+B+C+D)] × 100	0≤PC≤100

5.2. *Afternoon (1200-0000 UTC)*

In the afternoon, the 22 parameter combination shows 63.79% correct prediction (37 out of 58 cases) for CD and 50% correct prediction (34 out of 68 cases) for FW for the verification of 3 years for the next 12 hours (stage 1, Table 5.1). The corresponding figures in case of 20 parameters are 67.24% (39 out of 58 cases) and 54.41% (37 out of 68 cases) respectively (Table 5.2, stage 2). In the third stage consisting of 17 parameters, the figures are 77.08% (37 out of 48 cases) and 57.41% (31 out of 54 cases) respectively (Table 5.3, stage 3). The overall percentages in 3 stages are 56.34, 60.31 and 70 respectively. Thus, in the afternoon, it again reveals that combination of 17 parameters show the maximum percentage of correct result. Hence, by the similar analogy above, it again can be stated that these 17 parameters may be used for operational prediction in the afternoon for next 12 hours, instead of original 22 or 20 parameters.

6. Skill – Scores

For each stage of the analysis and also for both morning (0000 UTC) and afternoon (1200 UTC) (Tables 4.1 – 4.3 for morning and Tables 5.1 – 5.3 for afternoon), the results are presented in the form of a 2 × 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 Table 6.

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), Percentage of correct result (PC) and others are computed (Table 7). Brief description of the skill scores are presented in Table 7.

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 being used in the literature as CD 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 (Haklander and Deldon 2003). The limitations of TSS and HSS is that, if the number of correct forecast (A) and number of correct non-event forecast (D) are interchanged and number of misses (B) and also number of false alarms (C) are interchanged, scores remain unchanged. But, CSI would change. Thus, no single forecast would give complete picture. (Mukhopadhyay *et al*., 2003 and Tyagi *et al*., 2010). However, it is desirable to include CSI, POD, FAR, MR, C-NON, BIAS, and PC in addition to HSS for boarder and useful forecast.

6.1. *Results of skill score*

For each stage of the analysis and also for each morning and afternoon, the skill scores are presented in Table 8.

6.1.1. *Morning (0000 – 1200 UTC)*

The overall forecast skills are almost consistent. The TSS and HSS for stages 1-3 (22, 20 and 17 parameters for morning) are also consistent. The highest score occurred

TABLE 8

Table of different skill scores

	Morning 0000 – 1200 UTC			Afternoon 1200 – 0000 UTC		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
POD	0.5957	0.6382	0.6829	0.6379	0.6724	0.7708
FAR	0.5757	0.5714	0.4814	0.4594	0.4428	0.3835
MR	0.4042	0.3617	0.3170	0.3621	0.3276	0.2292
C-NON	0.5869	0.5652	0.6239	0.5000	0.5441	0.5741
CSI	0.3294	0.3448	0.4179	0.4022	0.4382	0.5211
TSS	0.1826	0.2034	0.3068	0.1379	0.2165	0.3448
HSS	0.1807	0.1818	0.2876	0.1357	0.2133	0.3402
BIAS	1.4042	1.4892	1.3171	1.2241	1.2069	1.2500
Percent Correct	58.99%	58.99%	66.54%	56.35%	60.32%	66.67%

in case of 17 parameter combination (0.3 and .28). The POD, PC and C-NON are also highest for 17 parameters case (0.68, 64.54 and 0.62 respectively). The CSI is reasonably high for 17 parameters case (0.42). However, FAR and MR is lowest for 17 parameter case (0.48 and 0.32). The discussion clearly highlights the efficiencies of 17 parameter combination.

6.1.2. *Afternoon (1200 – 0000 UTC)*

The almost consistent overall forecast skills are prominent. The TSS and HSS for stages 1-3 (22, 20 and 17 parameters for morning) are reasonably consistent. The highest score occurred in case of 17 parameter combination (0.34 for both). The POD, PC and C-NON are also highest for 17 parameters cases which are 0.77, 66.67 and 0.57 respectively. The CSI is considerably high for 17 parameters case (0.52). Also, FAR and MR are lowest for 17 parameter case which are 0.32 and 0.22 respectively. The details clearly points out the effectiveness of 17 parameter case.

7. Conclusion

The above analysis reveals that, both in the morning and afternoon but especially in the afternoon, 17 parameter combinations (with CAPE/CIN) have profound impact on CD formation in comparison to 20 and 22 parameter combinations. These 17 parameter set may be utilized in an effective manner for prediction purpose. It ultimately helps to reduce the computational time and handling of huge volume of data. Thus, the methodology may be very useful for operational purpose.

References

Asnani G. C., 2005, *Tropical Meteorology,* **2**, 829-833.

Betts, A. K., 1974, "Thermodynamics of tropical convective soundings", *Mon. Wea. Rev.,* **102**, 760-764.

Brier, G. W. and Allen, R. A., 1952, "Verification of weather forecasts", Compendium of Meteorology, *Amer. Meteor. Soc.*, 841-848.

Chatterjee, S., Ghosh, S. and De, U. K., 2009, "Reduction of number of parameters and forecasting convective developments at Kolkata (22.53ºN, 88.33ºE), India during pre-monsoon season: An application of multivariate techniques", *Ind. J. Radio & Space Phys.*, **38,** 5, 275-282.

Chowdhury, A. M., Ghose, A. and De, U. K., 1996, "Analysis of pre-monsoon thunderstorm occurrence at Dhaka from 1983 to 1992 in terms of (θ_{es} - θ_e) and convergence/divergence at surface", *Ind. J. Phys.,* **70B**, 357-363.

Chowdhury, A. K., 1961, "Pre-monsoon thunderstorm in Assam", Tripura and Manipur, *Ind. J. Met. & Geophys.*, **3**, 319-321.

Collier, C. G. and Lilley, R. B. E., 1994, "Forecasting thunderstorm initiation in north-west Europe using thermodynamic indices, satellite and radar data", *Met. Appls.*, **1**, 75-84.

Darkow G. L., 1968, "The total energy environment of storms", *J. Appl. Meteorol.,* **7**, 199-205.

Dasgupta, S. and De, U. K., 2001, "Markov chain models for pre-monsoon thunderstorm in Calcutta", *Ind. J. Radio & Space Phys.*, **31**, 138-142.

Dhawan, V. B., Tyagi, A. and Bansal, M. C., 2008, "Forecasting of thunderstorms in pre-monsoon season over north-west India", *Mausam*, **59**, 4, 433-444.

Diaz, H. F. and Fulbright, D. C., 1981, "Eigenvector analysis of seasonal temperature, precipitation and synoptic scale system frequency over contiguous United States, Part.1 (winter)", *Mon. Wea. Rev.,* **109**, 1267-1284.

Donaldson, R., Dyer, R. and Kraus, M., 1975, “An objective evaluator of techniques for prediction of severe weather events, Reprints, ninth Conf. on Severe Local Storms, Norman, OK”, *Amer. Meteor. Soc.*, 321-326.

Fuelberg, H. E. and Bigger, D. G., 1994, “The pre-convective environment of summer thunderstorm over Florida Panhandle”, *Wea. Forecasting,* **9**, 316-326.

Fujita, T. T., Bradbury, D. L. and Thullenar, C. F. Van Palm, 1970, “Sunday Tornadoes of April 11, 1965”, *Mon. Wea. Rev.,* **98,** 26-69.

Galway, J. G., 1956, “The lifted index as a predictor of latent instability”, *Bul. Am. Meteorol. Soc.*, **37**, 528-529.

George, J. J., 1960, “Weather forecasting for aeronautics”, Academic Press, New York, p673.

Ghose, S., Sen, P. K. and De, U. K., 1999, “Identification of significant parameters for the prediction of pre-monsoon thunderstorm at Calcutta”, *Int. J. Climatol.*, **19**, 673-681.

Haklander, A. J. and Delden, A. V., 2003, “Thunderstorm predictors and their forecast skill for the the Netherlands”, *Atmos. Res.***, 67-68**, 273-299.

Hanssen, A. W. and Kuippers, W. J. A., 1965, “On the relationship between the frequency of rain and various meteorological parameters”, *Verhand. K. Nederlands. Meteor. Inst*., **81**, 2-15.

Hoddinot, M. H. O., 1986, “Thunderstorm observations in West Bengal 1945-46”, *Weather*, **41**, 2-5.

Huntrieser, H., Schiesser, H., Schmid, W. and Waldvogel, A., 1997, “Comparison of traditional and newly developed thunderstorm indices for Switzerland”, *Weather Forecasting*, **12**, 108-123.

Iyenger, R. N. and Basak, P., 1994, “Regionalization of Indian Monsoon rainfall and long term variability signals”, *Int. J. Climatol*., **14**, 1095-1114.

Kessler, E., 1982, “Thunderstorm morphology and dynamics”, US Department of Commerce, **2**, 5-7, 93-95, 146-149.

Kistwal, C. M., Basu, S. and Pandey, P. C., 1996, “An algorithm for retrieving vertical wind profiles from satellite-observed winds over the Indian Ocean using complex EOF analysis”, *J. Appl. Meteorl*. , **35**, 673-678.

Koteswaram, P. and Srinivasan, V., 1958, “Thunderstorms over Gangetic West Bengal in the pre-monsoon season and the synoptic factors favorable for their formation”, *Ind. J. Met. & Geophys.,* **9**, 301-312.

Kumar, Surendra, 1972, “An objective method of forecasting pre-monsoon thunderstorm/dust storm activity over Delhi and neighborhood”, *Mausam*, **23**, 1, 45-50.

Kuo, H. L., 1965, “On formation and intensification of tropical cyclones through latent heat release by cumulus convection”, *J. Atm. Sci.*, **22**, 40-63.

Lal, R., 1990, “Forecasting of severe convective activity over Lucknow in pre-monsoon season”, *Mausam*, **41**,455-458.

Maryon, R. H. and Storey, A. H., 1985, “A multivariate statistical model for forecasting anomalies of half-monthly mean surface pressure”, *Int. J. Climatol.,* **5**, 561-578.

Miller, R. G., 1972, “Notes on analysis and severe storm forecasting procedure of the military weather warming centre”, Technical Report No. **20**, AWS(MAC), USAF, p139.

Mukherjee, A. K., 1964, “Study of thunderstorms around Guwahati Air Port”, *Ind. J. Met. & Geophys.*, **3**, 425-430.

Mukhopadhyay, P., Sanjay, J. and Singh, S. S., 2003, “Objective forecast of thundery/non-thundery days using conventional indices over three northeast Indian stations”, *Mausam*, **54**, 4, 867-880.

Overland, J. E. and Preisendorfer, R. W., 1982, “A significance test for principal components applied to a cyclone climatology”, *Mon. Wea. Rev*., **110**, 1-4.

Pant, B. C. and Shivhare, R. P., 1998, “Markov chain model for study of wet/dry spells at A. F. station, Sarsawa during SW monsoon season”, *Vatavaran*, **22**, 37-50.

Rao, K. N. and Raman, P. K., 1961, “Frequency of days of thunder in India”, *Ind. J. Met. & Geophys.*, **12**, 103-108.

Ravi, N., Mohanty, U. C., Madan, O. P. and Paliwal, R. K., 1999, “Forecasting of thunderstorm in the pre-monsoon season at Delhi”, *Meteorl. Appl.*, **6**, 29-38.

Reap, R. M., 1990, “Six-hour thunderstorms probability forecasts for Kansas – Oklahoma and north east United States, NOAA”, National Weather Service Technical Bulletin, No. **385**, NOAA, US dept. of commerce, p6.

Sahu, J. K., 1996, “A study on the thermodynamics conditions of atmosphere for forecasting thunderstorms over north-west India”, *Vatavaran*, **19**, 27-35.

Sẩnchez, J. L., Lộpez, L., Bustos, C., Marcos, J. L. and Garcỉa, O. E., 2008, “Short-term forecast of thunderstorms in Argentina”, *Atmos. Res.*, **88**, 36-45.

Sen, S. ẳ . and Basu, S. C., 1961, “Pre-monsoon thunderstorm in Assam and synoptic conditions favourable for their occurrence”, *Mausam*, **12**, 15-20.

Sengupta, P. R. and Basak, P., 1998, “Some studies of southwest monsoon rainfall”, *Proc. Ind. Nat. Sci. Acad*., **64, A,** 737-745.

Showalter, A. K., 1953, “A stability index for forecasting thunderstorm”, *Bull. Am. Meteorol. Soc.*, **34**, 250-252.

Thiagarajan, R., Ramadoss and Ramaraj, 1995, “Markov chain model for daily rainfall occurrences at east Thanjavur district”, *Mausam*, **46**, 383-388.

Tyagi, B., ẳ aresh Krishna,V. and Satyanarayana, A. ẳ . V., 2010, “Study of thermodynamic indices in forecasting pre-monsoon thunderstorm over Kolkata during STORM pilot phase 2006-2008”, *Nat. Hazards, online pub.*, August, 2010.

Ward, M. ẳ . and Folland, C. K., 1991, “Prediction of seasonal rainfall in the ẳ ordeste of Brazil using eigenvector of sea-surface temperature”, *Int. J. Climatol.*, **11**, 711-743.

Weare, B. C., 1987, “Relationship between monthly precipitation and SST variation in the Tropical Pacific region”, *Mon. Wea. Rev.,* **115**, 2687-2698.

Williams, E. and Renno, ẳ ., 1993, “An analysis of the conditional instability of the tropical atmosphere”, *Mon. Wea. Rev.,* **121**, 21-36.