MAUSAM, **59**, 4 (October 2008), 433-444

551. 509.326 : 551.553.21(545)(547)

Forecasting of thunderstorms in pre-monsoon season over northwest India

V. B. DHAWAN, AJIT TYAGI* and M. C. BANSAL**

Institute of Technology and Science, Ghaziabad, India

**India Meteorological Department, New Delhi, India*

***M. M. H. College, Ghaziabad, India*

(*Received 19 September 2007*)

***e mail : tyagiajit@hotmail.com**

सार – गर्जवाले भीषण तूफानों से हाने वाली क्षति से बचाव एवं रोकथाम के लिए इन तूफानों का पहले ही पता लगाना और पूर्वानुमान किया जाना आवश्यक है। 1995–1999 तक के वर्षों के मई एवं जून माह के दिल्ली के तापमान के (टेम्प) आँकड़ों को लेते हुए इंडक्टिव मशीन लर्निंग तकनीक का उपयोग करते हुए निचली और ऊपरी स्तर पर होने वाली समस्या को समझते हुए स्वतः ज्ञात करने वाली तकनीक से गर्ज वाले तूफानों के पूर्वानुमान करने की तकनीक का उपयोग करते हुए पश्चिमोत्तर भारत में मुख्य शहर दिल्ली में मानसून पूर्व ऋतु के दौरान गर्ज वाले तूफानों का पूर्वानुमान करने के लिए भारत में पहली बार एक्सपर्ट सिस्टम फॉर थंडरस्ट्राम फारकास्टिंग (ई. एस. टी. एफ.) विकसित की गई है। इसके लिए 0000 यू. टी. सी. पर दिल्ली के सतह के 850, 700, 500, और 300 हे. पा. स्तरों के केवल तापमान (टेम्प) आँकड़ों की प्रविष्टि आवश्यक है। ये नियम स्थायित्व घातांकों एवं अन्य तापीय गति प्राचलों पर आधारित है जिसकी गणना इस शोध–पत्र में बताए गए साउण्डिंग से की गई है। इस प्रणाली से गर्जवाले तूफानों के कारण दिल्ली में होने वाली जलवायविक सूचनाएँ भी प्राप्त हुई है। ई. एस. टी. एफ. की वस्तुनिष्ठ तकनीकों के साथ तुलना करके दिल्ली में मानसून–पूर्व ऋतु के मई और जून के महीनों में गर्जवाले तूफान के पूर्वानुमान के लिए हाँ अथवा नहीं प्रकार की गत्यात्मक–साँख्यिकीय पद्धतियाँ विकसित की गईं है। इन पद्धतियों को ग्राफिकल डिस्क्रिमिनेशन प्रणाली एवं बहु समाश्रयण प्रणाली तथा इसी प्रकार के अन्य आँकड़ा समूहों का उपयोग करते हुए विकसित किया गया है जैसेः– दिल्ली में 1995–1999 तक के वर्षों में मई एवं जून के महीनों के तापमान (टेम्प) आँकड़ों और ई. एस. टी. एफ. के विकास में उपयोग किए गए समान विभव के पूर्वसूचकों का उपयोग बहु समाश्रयण प्रणाली में क्रमागत जाँच प्रक्रिया के अंतर्गत प्राचलों का महत्वपूर्ण स्थान पाया गया है। इस शोध पत्र में वर्ष 2000–2001 के मई और जून माह के अलग–अलग आँकड़ा समूहों से विकसित की गई तीन प्रणालियों की जाँच की गई है। ग्राफिकल डिस्क्रिमिनेशन प्रणाली, बहु समाश्रयण तकनीक तथा ई. एस. टी. एफ. द्वारा जारी किए गए पूर्वानुमानों को प्राचलों के सत्यापन के साथ तुलना करने पर देखा गया है कि बहु समाश्रयण प्रणाली से प्राप्त किए परिणाम ग्राफिकल डिस्क्रिमिनेशन प्रणाली से प्राप्त किए गए परिणाम से बेहतर पाए गए हैं। ई. एस. टी. एफ. के उपयोग से प्राप्त किए परिणाम गत्यात्मक स्थैतिक मॉडल्स के उपयोग से प्राप्त किए गए परिणाम से बेहतर पाए गए हैं।

ABSTRACT. Early detection and forecasting of thunderstorms is important in safeguarding and prevention of damages resulting from violent thunderstorms. An Expert System for Thunderstorm Forecasting (ESTF) during pre-monsoon season over Delhi the representative location over northwest India has been developed for the first time in India by using technique of approaching the problem "bottom up" by using inductive machine learning techniques to automatically acquire the knowledge about thunderstorm forecasting from the weather development data set consisting of TEMP data of Delhi for the months of May and June for the years from 1995-1999. Only input required is the entry of 0000 UTC TEMP data of Delhi at surface, 850, 700, 500 and 300 hPa levels. The rules are based on stability indices and other thermodynamic parameters evaluated from the said sounding. The system also provides climatological information about thunderstorms over Delhi. To compare the ESTF with the objective techniques, Dynamical-statistical methods for yes or no type thunderstorm occurrence forecast over Delhi during pre-monsoon months of May and June have been developed by using graphical discrimination method and multiple regression method and by using the same development data set *i.e*., TEMP data of Delhi for the months of May and June for the years from 1995-1999 and by using the same potential predictors as used in development of ESTF. In multiple regression method the parameters were found to be significant by stepwise screening procedure. The three methods developed were tested with independent data sets of May and June for the years from 2000-2001. Comparison of verification parameters of the forecast issued by Graphical Discrimination method, Multiple Regression Technique and by ESTF indicates that results of multiple regression method are better than those of graphical discrimination method. The results obtained by using ESTF were better than those obtained by using dynamical statistical models.

Key words − Thunderstorm, Expert system, Inductive machine learning, Multiple regression, Graphical discriminant analysis.

1. Introduction

Thunderstorms are the manifestation of convective activity in atmosphere. It is well known that severe thunderstorms are favoured by strong convective instability, abundant moisture at lower levels, strong wind shear and a dynamical lifting mechanism that can release the instability. These large severe storms either develop individually or more typically in groups associated with synoptic scale fronts or mesoscale convergence areas. These lead to severe floods, strong winds, hail, lightning strikes, destruction of property and even loss of life. Thunderstorms and the gust/squalls associated with the thunderstorms pose a serious hazard to aviation as well as to some of other activities such as transportation, agriculture, construction, communication, power transmission etc. Though, this mesoscale phenomenon may occur at any time and over any part of the country in a given year, they are most severe over northwest India during pre-monsoon season. Early detection and forecasting of thunderstorms is important in safeguarding and prevention of damages resulting from these violent thunderstorms.

2. Earlier studies

A number of studies using synoptic, synoptic-objective or purely objective techniques, or techniques using expert systems have been developed for forecasting thunderstorms. Important among these are discussed here.

Surendra Kumar (1972) developed a technique for forecasting pre-monsoon thunderstorm/duststorm activity over New Delhi region by using the parameters Showalter Index, Convective Condensation Level (CCL), mean mixing ratio at 850, 800 and 700 hPa levels, wind direction and the difference in height between the CCL and freezing level. Since the parameters used are based on the experience of the forecaster, the method therefore is semi-objective. Lal (1990) indicated that a Showalter index of – 4 or less, mean relative humidity below the level of 850 hPa of 45% or more and dew point above normal are favourable conditions for the occurrence of thunderstorms over Lucknow.

Reap (1986 & 1990) applied screening regression techniques to relate lightning data to large scale meteorological predictors obtained from numerical forecast models, in order to derive equations for forecasting thunderstorm for different parts of the United States. Collier and Lilley (1994) discussed the occurrence of thunderstorms in northwest Europe, and outlined the conclusion of previous work on the mechanisms for thunderstorm initiation. They tested simple rules for identifying the likely occurrence of thunderstorms using instability indices, which if compared with information derived from satellite imagery, provide the basis of a general alert procedure.

Sahu (1996) studied various thermodynamic parameters such as convective available potential energy (CAPE), static energy profiles, total precipitable water (TPW) alongwith conventional charts to obtain their changes prior, during and after the thunderstorm by using 4 hourly special radio sonde data for Delhi and Jodhpur taken during MONTBLEX-90. He observed that CAPE profile and TPW profiles provide significant clue for forecasting thunderstorm. The changes in the atmosphere at times are available around 6 to 12 hours prior to occurrence of thunderstorm.

The study by Devrani & Mukherjee (1997) on forecasting of pre-monsoon thunderstorm/duststorm over Jodhpur reveals that combination of critical threshold value of lifted index, Showalter index, cross total index, vertical total index, Jefferson's modified index (TMJ), George K index in conjunction with lifting (availability/ non availability) gives good indication of thunderstorm/ duststorm occurrence/non-occurrence over Jodhpur during the months of April and May.

Ravi *et al*. (1999) developed two objective forecasting methods for forecasting of thunderstorms in the pre-monsoon season at Delhi. The first method is based on graphical technique using fifteen different types of stability indices in combinations of different pairs. The second method is based on using nine significant predictors to formulate a multiple regression equation that gives the forecast in a probabilistic term. They found that multiple regression method gives consistently better results and is a potential method for operational use.

3. Expert system approach to forecasting

Expert systems are artificial intelligence computer programs that perform inference processes based on a collection of expertise and a set of known facts about the situation at hand. Researchers have used expert systems to resolve many meteorological problems (Dhawan 2002). Most of the efforts are directed toward the development of expert systems for weather predictions (operational forecasts), when there is no time to provide a more careful analysis, like in thunderstorm forecasting (Conway, 1987). In this case, an expert system simply replaces a human expert, who might not be available at that moment.

Elio *et al*. (1987) developed expert system to predict convective storms. Their system includes heuristics to assist in the interpretation of a station model, and also has an AI interpreter of surface station reports. Colquhoun (1987) developed a decision tree method of forecasting thunderstorms and severe weather that consisted of two main components. The first component determined if thunderstorms were expected. The evaluation process ended if the conclusion was negative. However, given thunderstorm formation, the second component of the algorithm determined the type of thunderstorm based on the wind shear and moisture profiles of the atmosphere. Thunderstorms were listed as either non-severe, dry microburst, wet microburst, or severe. It was also possible to forecast the potential of flooding.

Lee and Passner (1993) developed a rule-based expert system to assist military weather forecaster in the prediction of thunderstorms. Lee and Passner used soundings which resulted in thunderstorm occurrences to establish the critical values of the cross totals, vertical totals, total totals, lifted index and Showalter index. Verification results showed that Thunderstorm Intelligence Prediction System (TIPS) could assist the weather forecaster in predicting thunderstorm occurrence. Kumar *et al*. (1994) used application of machine learning techniques on weather data sets to acquire knowledge automatically for the development of an expert system to predict the occurrence and mean depth of rainfall over Melbourne city and its suburbs in Australia during a 24-hour period. The weather data sets were assembled from the archives of the Australian Commonwealth Bureau of Meteorology. The results were competitive and performance matched that of the human experts in weather forecasting.

4. Present study

The objective of the study is to develop an expert system called "Expert System for Thunderstorm Forecasting (ESTF)" to forecast thunderstorm activities over Delhi during pre-monsoon months by using radio sonde data of Delhi. The system processes 0000 UTC TEMP data of Delhi to make a 14 hours forecast. The rules are based on stability indices and other thermodynamic parameters evaluated from the said sounding. The parameters are so chosen that have relationships with the processes which account for development of thunderstorms.

An attempt has been made to approach the problem of development of expert system "bottom up" by using inductive machine learning techniques to automatically acquire the knowledge about weather forecasting from the weather data sets. ESTF has been developed to forecast occurrences and non-occurrences of thunderstorms. The system does not include severity of thunderstorm activities. The performance of ESTF has been compared with commonly used objective techniques.

5. Expert System for Thunderstorm Forecasting (ESTF)

The expert system has been designed to use data available at 0000 UTC and to forecast the occurrence of thunderstorm over Delhi for a 14 hours period from 0400 to 1800 UTC during the months of May and June. A front-end "T-Phi gram Analysis Program (TPAP)", linked to ESTF, displays computed stability indexes and vertical distributions of various meteorological variables at surface, 900, 850, 700, 500 and 300 hPa levels. Development data set consisted of TEMP data of Delhi for the months of May and June for the years from 1995 - 1999. ESTF compares the input data with the critical values of the parameters stored in the knowledge base and by applying the decision rule decides for or against thunderstorm occurrence.

5.1. *Development of ESTF*

Various stability indices and parameters having relationships with physical processes that are responsible for thunderstorm occurrences were considered as potential predictors for development of knowledge based expert system for forecasting of thunderstorms over Delhi during pre-monsoon months. The factors which favour thunderstorm development *i.e*., moisture, instability and triggering mechanism were kept in mind while selecting the indices and parameters.

5.1.1. *Surface* − Dry bulb and dew point temperatures, dew point depression, relative humidity, surface wind speed and direction, zonal (*u*) and meridional (*v*) components of surface wind.

5.1.2. *Upper Air* − Dry bulb, wet bulb and dew point temperatures, dew point depression, mixing ratios, saturation mixing ratios, relative humidity, wind direction (*D*), wind speed (*F*), zonal (*u*) and meridional (*v*) components of wind, potential and equivalent potential temperatures at levels 900, 850, 700, 500 and 300 hPa, wind shear, lapse rate of temperatures and thickness between various levels.

5.1.3. *Stability Indices* − Showalter Index (SI), Rackliff index (RI), Jefferson's modified index (TMJ), Convective Index of Reap (CIIR), George Index (K), Vertical Total Index (VTI), Cross Total Index (CTI), Total Totals Index (TTI), Modified George index (KMOD), Modified Vertical, Cross and Totals Total Indices (VTIM, CTIM, TTIM), Lifted Index (LI), Potential Instability Index (PII), Severe Weather Threat Index (SWEAT). Computational procedure for few of the indices are given in Appendix B (Showalter, 1953; Galway, 1956; George, 1960; Rackliff, 1962 and Miller 1967).

TABLE 1

Percentage probability of occurrence/non-occurrence of thunderstorms in various ranges of Showalter Index

Range	% probability of occurrences	% probability of non-occurrences
>9	1.9	100
>8 to ≤ -9	1.9	97.3
>7 to ≤ 8	1.9	96.3
>6 to ≤ 7	2.8	95.3
>5 to ≤ 6	4.7	92.7
>4 to ≤ 5	5.7	88
>3 to ≤ 4	5.7	83.9
>2 to ≤ 3	9.4	77.6
>1 to ≤ 2	13.2	70.3
>0 to ≤ 1	19.8	63
>-1 to ≤ 0	29.2	53.6
>-2 to ≤ -1	43.4	44.3
>-3 to ≤ -2	52.8	38
>-4 to ≤ -3	65.1	25
>-5 to ≤ -4	73.6	15.6
>-6 to ≤ -5	82.1	8.9
>-7 to ≤ -6	85.8	5.7
>-8 to ≤ -7	94.3	2.6
>-9 to ≤ -8	100	0.5
≤ -9	100	0

TABLE 2

Percentage probability of occurrence/non occurrence of thunderstorms in various ranges of equivalent potential temperatures at 850 hPa

Range (degree Kelvin)	% probability of occurrences	% probability of non-occurrences
≤ 320	0	100
>320 to ≤ 325	1.9	98.4
>325 to ≤ 330	9.4	90.6
>330 to ≤ 335	22.6	82.3
>335 to ≤ 340	31.1	64.2
>340 to ≤ 345	43.4	43.5
>345 to ≤350	52.8	32.6
>350 to ≤ 355	68.8	20.2
>355 to ≤ 360	84.8	11.4
>360 to ≤ 365	92.3	5.2
>365 to ≤ 370	97	2.6
>370 to ≤ 375	98.9	1
>375 to ≤ 380	100	0
≥ 380	100	0

TABLE 3

Percentage probability of occurrence/non occurrence of thunderstorms in various ranges of Meridional component of wind at 850 hPa

Range (knots)	% probability of occurrences	% probability of non-occurrences
≤ -30	0.0	100
>-30 to ≤ -25	0.9	100.0
≥-25 to ≤ -20	0.9	98.5
>-20 to ≤ -15	0.9	96.4
>-15 to ≤ -10	6.6	91.2
>-10 to ≤ -5	27.4	72.5
>-5 to ≤ 0	54.8	46.6
>0 to ≤ 5	73.7	24.3
>5 to ≤ 10	86.9	8.8
>10 to ≤ 15	93.5	5.2
>15 to ≤ 20	97.3	3.6
>20 to ≤ 25	99.2	0.5
>25 to ≤ 30	99.2	0.5
>30 to ≤ 35	99.2	0.0
>35	100.0	0.0

TABLE 4

Percentage probability of occurrence/non occurrence of thunderstorms in various ranges of dew point temperatures at 850 hPa

Range (degree Celsius)	% probability of occurrences	% probability of non-occurrences
≤-4	0	100
>-4 to ≤ -2	1.9	99.5
≥ -2 to ≤ 0	2.8	97.9
>0 to ≤ 2	3.7	93.3
>2 to ≤ 4	7.5	88.1
>4 to ≤ 6	10.4	80.8
>6 to ≤ 8	18.9	69.9
>8 to ≤ 10	28.3	61.1
>10 to ≤ 11	30.2	50.8
>11 to ≤ 12	33	44.6
>12 to ≤ 13	36.8	39.9
>13 to ≤ 14	47.2	33.7
>14 to ≤ 15	52.8	27.5
>15 to ≤ 16	56.6	19.7
>16 to ≤ 18	70.8	17.6
>18 to ≤ 20	92.5	8.8
>20 to ≤ 22	97.5	2.6
>22 to ≤ 24	100	0.5
> 24	100	0

TABLE 5

Percentage probability of occurrence/non occurrence of thunderstorms in various ranges of relative humidity at 700 hPa

Range (%)	% probability of occurrences	% probability of non-occurrences
≤ 30	6.6	100
>30 to ≤ 40	13.2	76.7
>40 to ≤ 50	25.5	54.9
>50 to ≤ 60	39.6	35.8
>60 to ≤ 70	54.7	25.9
>70 to ≤ 80	72.6	16.1
>80 to ≤ 90	86.8	9.8
>90 to ≤ 100	100.0	5.2

The required data pertaining to above parameters was computed from radio sonde data of Delhi for the relevant period. Present weather (thunderstorm/ cumulonimbus cloud) for the period of study was taken from meteorological station Hindon, situated at about 15 kilometers to the northeast of central Delhi.

Tables were constructed for percentage occurrences and non-occurrences of thunderstorms relating to values of these parameters in various ranges. Probability of occurrences and non-occurrences of thunderstorm for values of each variable and stability index were calculated. This data was then used to find out critical value of various parameters. Critical value taken was the value where probability of occurrence was more than that of non-occurrence. Finally following 5 parameters were chosen: Showalter Index, Equivalent potential temperatures at 850 hPa, Meridional component of wind at 850 hPa, Dew point temperature at 850 hPa and relative humidity at 700 hPa.

For each of the above parameter the cumulative probability values for occurrences and non occurrences of thunderstorms in various value ranges of these parameters are shown in Tables 1 to 5. The critical value of parameters has been taken as the value of the parameter from the respective table as the value where the probability of occurrence is higher than that of non-occurrences. The parameters with their critical values are shown in Table 6. These parameters are easily computed and capture some essential elements needed to describe a potential thunderstorm environment. The threshold index values used to decide if thunderstorms would form are called critical values, the critical values of these parameters were exceeded then the expert system noted that conditions were favourable for thunderstorm formation.

TABLE 6

Selected parameters and their critical values

Parameter	Critical values
Showalter Index	≥ 0 to < –9
Equivalent potential temperatures at 850 hPa	>340 degree Kelvin
Meridional component of wind at 850 hPa	> -10 knots
Dew point temperature at 850 hPa	>13 degree Celsius
Relative humidity at 700 hPa	> 60 %

A 2 × 2 contingency table (Table A in appendix A) was used to calculate various verification parameters. Values of A, B, C, D in Table A of Appendix A calculated by feeding development data set 1995 to 1999 by taking various combination of selected parameters alongwith the calculated values of verification parameters based on development data set are shown in Table 7. Verification parameters are listed in Table 8 with their definitions in Appendix A). In Table 7, EPT850, MC850, SI, TD850, RH700 denote equivalent potential temperature at 850 hPa, meridional component of wind at 850 hPa, Showalter index, dew point temperature at 850 hPa and relative humidity at 700 hPa respectively.

A study of the table shows that best results are achieved when forecast of thunderstorm is issued with 3 parameters out of the 4, *i.e.*, equivalent potential temperatures at 850 hPa, meridional component of wind at 850 hPa, relative humidity at 700 hPa and Showalter Index exceed their critical values.

Table 7 shows that forecast with ≥ 1 parameters out of 5 and 4 respectively exceeding its critical value are the most successful in terms of Probability Of Detection (POD) but the corresponding False Alarm Rate (FAR) is somewhat higher. The lowest FAR is achieved when all 5 or 4 indices indicate thunderstorm, but this situation gives a low POD and Critical Success Index (CSI). Forecast with atleast 3 parameters out of the 4 parameters, *i.e.*, equivalent potential temperatures at 850 hPa, meridional component of wind at 850 hPa, relative humidity at 700 hPa and Showalter index exceeding their critical values is recommended as it results in a POD of 0.71 and FAR of 0.39. It has values of CSI, True Skill Score (TSS), Heidke Skill Score (HSS) as 0.49, 0.46 and 0.44 respectively. The BIAS is 0.83 *i.e.*, it slightly under forecast the storms. Percent correct is 73.58 %. It is interesting that as an increasing number of variables are used the False Alarm falls, but so does the Probability of detection. Use of any single parameter or using exactly 0, 1, 2, 3, 4 or 5 parameters does not give satisfactory results.

TABLE 7

Values of A, B, C and D in Table A of Appendix A alongwith values of verification parameters for forecast by taking various combination of selected parameters

	A	B	C	D	POD	FAR	MR	C-NON	CSI	TSI	HSS	BIAS	PC
Forecast by taking 5 parameters *i.e.*, Equivalent potential temperature at 850 hPa, Meridional component of wind at 850 hPa, Showalter index, Dew point temperature at 850 hPa and Relative humidity at 700 hPa													
Forecast by ≥ 4 parameters	60	46	42	151	0.57	0.41	0.43	0.78	0.41	0.35	0.35	0.78	70.57
Forecast by ≥ 3 parameters	77	29	56	137	0.73	0.42	0.27	0.71	0.48	0.44	0.41	0.90	71.57
Forecast by ≥ 2 parameters	97	9	82	111	0.92	0.46	0.08	0.58	0.52	0.49	0.42	1.07	69.57
Forecast by ≥ 1 parameters	106	0	134	59	1.00	0.56	0.00	0.31	0.44	0.31	0.24	1.49	55.18
Forecast by exactly 5 parameters	38	68	20	173	0.36	0.34	0.64	0.90	0.30	0.25	0.28	0.53	70.57
Forecast by exactly 4 parameters	22	84	22	171	0.21	0.50	0.79	0.89	0.17	0.09	0.11	0.51	64.55
Forecast by exactly 3 parameters	17	89	14	179	0.16	0.45	0.84	0.93	0.14	0.09	0.10	0.38	65.55
Forecast by exactly 2 parameters	20	86	26	167	0.19	0.57	0.81	0.87	0.15	0.05	0.06	0.56	62.54
Forecast by exactly 1 parameters	9	97	52	141	0.08	0.85	0.92	0.73	0.06	-0.18	-0.20	1.03	50.17
Forecast by exactly 0 parameter	0	106	59	134	0.00	1.00	1.00	0.69	0.00	-0.31	-0.34	1.32	44.82
Forecast by relative humidity at 700 hPa	30	76	26	167	0.28	0.46	0.72	0.87	0.23	0.15	0.17	0.58	65.89
Forecast by equivalent potential temperature at 850 hPa	36	70	53	140	0.34	0.60	0.66	0.73	0.23	0.07	0.07	0.94	58.86
Forecast by Showalter index	33	73	47	146	0.31	0.59	0.69	0.76	0.22	0.07	0.07	0.86	59.87
Forecast by meridional component of wind at 850 hPa	34	72	32	161	0.32	0.48	0.68	0.83	0.25	0.15	0.17	0.67	65.22
Forecast by dew point temperature at 850 hPa	29	77	37	156	0.27	0.56	0.73	0.81	0.20	0.08	0.09	0.73	61.87
Forecast by EPT850, MC850,SI	63	43	35	158	0.59	0.36	0.41	0.82	0.45	0.41	0.42	0.72	73.91
Forecast by EPT850, MC850, SI, TD850	52	54	45	148	0.49	0.46	0.51	0.77	0.34	0.26	0.26	0.82	66.89
Forecast by EPT850, MC850, TD850	56	50	43	150	0.53	0.43	0.47	0.78	0.38	0.31	0.31	0.79	68.90
Forecast by EPT850, MC850, RH700	45	61	25	168	0.42	0.36	0.58	0.87	0.34	0.29	0.32	0.60	71.24
Forecast by MC850, SI, RH700	57	49	23	170	0.54	0.29	0.46	0.88	0.44	0.42	0.44	0.60	75.92
Forecast by EPT850, SI, RH700	50	56	27	166	0.47	0.35	0.53	0.86	0.38	0.33	0.35	0.63	72.24
Forecast by taking 4 parameters *i.e.*, equivalent potential temperature at 850 hPa, meridional component of wind at 850 hPa, relative humidity at 700 hPa and Showalter index													
Forecast by exactly 4 parameters	44	62	20	173	0.42	0.31	0.58	0.90	0.35	0.31	0.34	0.55	72.58
Forecast by ≥ 3 parameters	75	31	48	145	0.71	0.39	0.29	0.75	0.49	0.46	0.44	0.83	73.58
Forecast by ≥ 2 parameters	98	8	81	112	0.92	0.45	0.08	0.58	0.52	0.50	0.44	1.06	70.23
Forecast by ≥ 1 parameter	106	0	137	56	1.00	0.56	0.00	0.29	0.44	0.29	0.22	1.52	54.18

5.2. *Rule for forecasting of thunderstorm*

Using 0000 UTC TEMP data of Delhi, forecast thunderstorm during the period from 0400 to 1800 hrs (IST) when at least 3 parameters out of the 4 parameters *i.e.*, equivalent potential temperatures at 850 hPa, meridional component of wind at 850 hPa, relative humidity at 700 hPa and Showalter index exceed their critical values. In all other cases forecast no thunderstorm occurrence.

5.3. *Verification*

Objective evaluations of forecasting quality are carried out to determine the quality of forecasts

TABLE 8

Verification parameters

Abbreviation	Parameter
POD	Probability Of Detection
FAR	False Alarm Rate
MR	Miss Rate
C-NON	Correct Non Occurrence
CSI	Critical Success Index
TSS	True Skill Score
HSS	Heidke Skill Score
BIAS	Bias
PC	Percent Correct

TABLE 9

The mean and standard deviation of selected parameters for occurrence and non-occurrence cases of thunderstorms for development data

Parameters	Occurrence		Non occurrence	
	Mean	S D	Mean	S D
Lifted Index	-2.6	3.8	- 0.29	4.4
Cross Total Index	20.1	4.7	16.8	5.7
Equivalent Potential Temperature at 700 hPa (degree Kelvin)	339.3	9.7	333.5	8.8
George Index (K)	37.0	8.3	29.6	9.5
Jefferson modified index	31.7	5.3	27.2	5.9
Surface dew point temperature (degree Celsius)	20.3	4.4	17.2	7.1

(Wilks, 1995). The verification test of the "ESTF" was done using independent test data set for the months of May and June for the years from 2000 & 2001. The expert system forecast was valid from 0400 to 1800 UTC. By feeding test dataset of May and June for the year 2000 and 2001 to ESTF the values of A, B, C and D in 2 × 2 contingency table (Table A in Appendix A) were worked out. Verification of the technique with independent data set of 2000 and 2001 gave a POD of 0.81, FAR of 0.44, Miss Rate (MR) of 0.19, Correct non occurrences (C-NON) of 0.65, CSI of 0.50, TSS of 0.47, HSS of 0.43, BIAS of 0.58 and Percent Correct (PC) of 71.07 %. Bias indicated that method slightly under predicted thunderstorm occurrence.

5.4. *Validation of forecast by expert system against persistence forecast*

It is a good idea to validate forecasting experiments such as this against a persistence forecast (*i.e.*, a prediction for each day that assumes that the resulting weather that day will be the same as the day before). Using the test data set of May and June for the years from 2000 to 2001, forecast based on persistence were computed. It gave a POD of 0.42, FAR of 0.56, MR of 0.58, C-NON of 0.71, CSI of 0.27, TSS of 0.12, HSS of 0.13, BIAS of 0.52 and PC of 60.33 %.

6. Objective techniques

During pre-monsoon season feeble western disturbances approach northwest India, which are often difficult to be detected by the conventional synoptic analysis. These disturbances create favourable thermodynamical conditions for the development of thunderstorm. Therefore, objective techniques for forecasting of thunderstorms using dynamical and thermodynamical variables have been attempted. Objective 'yes' or 'no' type thunderstorm occurrence forecast valid from 0400 to 1800 UTC based on 0000 UTC data over Delhi during pre-monsoon months of May and June have been developed by using graphical discrimination method and multiple regression method and by using the same development data set, *i.e.*, TEMP data of Delhi for the months of May and June for the years from 1995-1999 and same indices and parameters as used in development of ESTF.

7. Graphical discrimination method

In order to develop an objective method for forecasting of thunderstorms, the above potential parameters were examined for their ability to predict the occurrence of thunderstorms over Delhi. It is found that no single predictor can provide a distinct critical value (critical value is the threshold value used to decide if thunderstorm would be realized or not on that particular day) to forecast the occurrence of a thunderstorm. The mean and standard deviation of potential predictors for the development data are calculated. A suitable parameter is considered to be one whose mean value is distinctly different from occurrence to non-occurrence cases of thunderstorms by at least ±1 standard deviation [Ravi *et al.* (1999)]. It was found that not a single parameter is suitable for use alone as a potential index for prediction of occurrence or non-occurrence cases of thunderstorms. Therefore using the developmental data sample for May and June together for the years from 1995-1999, pairs of parameters, one *versus* the other were tried in the form of a scatter diagram so that distinct groupings of occurrence and non-occurrence of thunderstorms could be achieved. The parameters finally selected for the scatter diagrams were Lifted Index, Cross Total Index, Equivalent Potential Temperature (θe) at 700 hPa, George Index (K), Jefferson's Modified Index (TMJ), Surface Dew Point Temperature. Mean and standard deviation of the parameters for occurrence and non-occurrence cases of thunderstorms for development data are presented in Table 9.

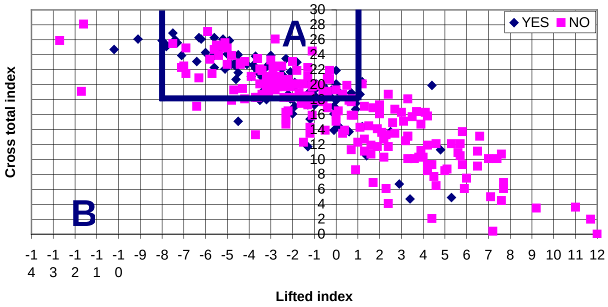

Fig. 1. Scatter Diagram (1)

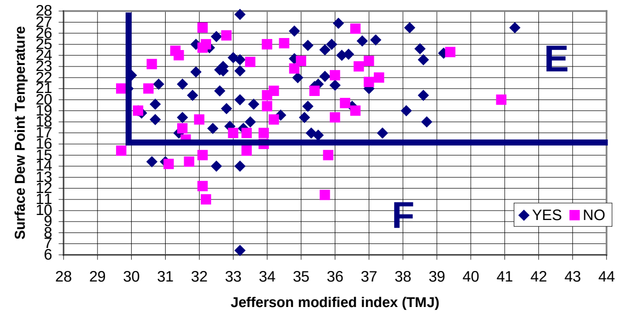

Fig. 3. Scatter Diagram (3)

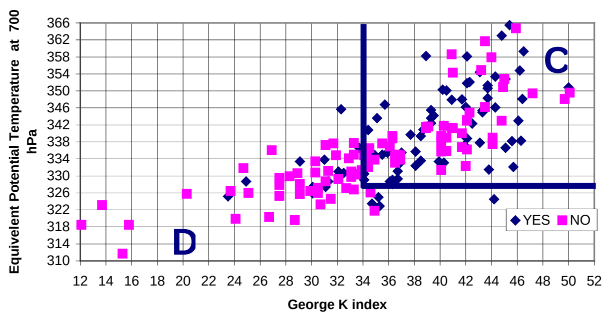

Fig. 2. Scatter Diagram (2)

7.1. *Preparation of scatter diagrams*

From the calculated indices and wind data three scatter diagrams showing strong relationship were prepared:

(*i*) In scatter diagram (1) Lifted Index (LI) was plotted against Cross Total Index (CTI). A diamond in the diagram represents a thunderstorm occurrence and a square represents no thunderstorm occurrence. The diagram was divided into 2 areas by fitting subjectively a curve so that most of the diamonds were included in area A and most of the squares were included in area B. All cases in area B were then excluded and considered as no forecast of the phenomena (Fig. 1).

(*ii*) The cases in area A of scatter diagram (1) were then plotted in scatter diagram (2) with coordinates as equivalent potential temperature (θe) at 700 hPa and George Index (K). A diamond in the diagram represents a thunderstorm occurrence and a square represents no thunderstorm occurrence. This diagram was also divided into 2 areas by fitting subjectively a curve so that most of the diamonds were included in area C and most of the squares were included in area D. Again all cases in area D were then excluded and considered as no forecast of the phenomena (Fig. 2).

(*iii*) The remaining data was plotted on scatter diagram (3) with coordinates as Surface dew point temperature and Jefferson's modified index (TMJ). A diamond in the diagram represents a thunderstorm occurrence and a square represents no thunderstorm occurrence. The diamonds and squares were separated by dividing the curve in to two areas E and F. In this diagram cases falling in E were predominantly diamonds and as such all of them were considered as 'Yes' forecast, while cases falling in F were considered as 'No' forecast (Fig. 3).

To use this technique in practical forecasting the following procedure is adopted:

(*i*) From the radio sonde data of Delhi at 0000 UTC calculate significant predictors and enter diagram (1) with parameters Lifted Index and Cross Total Index. If point falls in area B, forecast No thunderstorm and stop (Fig. 1).

(*ii*) If point falls in area A, enter diagram (2) with parameters equivalent potential temperature (θe) at 700 hPa and George Index (K). If point falls in area D, forecast No thunderstorm and stop (Fig. 2).

(*iii*) If point falls in area C, enter diagram (3) with parameters Jefferson's modified index (TMJ) and surface dew point temperature. If point falls in area F forecast No thunderstorm and if it falls in area E, forecast thunderstorm (Fig. 3).

7.2. *Verification*

Verification of the graphical discrimination method by using a 2 × 2 contingency table (shown as Table A in Appendix A) as in ESTF with the test dataset of May and June for the year 2000 and 2001 gave a POD of 0.60, FAR of 0.50, MR of 0.39, C-NON of 0.67, CSI of 0.37, TSS of 0.27, HSS of 0.48, BIAS of 1.21 and PC of 64%. Bias indicated that method slightly over predicted thunderstorm occurrences.

TABLE 10

Regression method for forecasting thunderstorm at Delhi

Predictor	Level (hPa)	Coefficient	Variance explained (VE)	Cumulative (CVE)	Correlation with predictand (CC)
Constant		-1.3158			
George K index	Stability index	0.018068	14.8	14.8	0.386
Meridional component of wind	850	0.006825	16.9	31.7	0.291
Saturation mixing ratio	500	0.144380	18.6	50.3	0.050
Dry bulb Temperature	500	-0.062890	20.4	70.7	0.028

Multiple correlation coefficient = 0 .4456328

8. Regression method

In the second method, a multiple regression technique is attempted. For the regression model, four significant predictors are selected out of the potential predictors, by use of stepwise screening procedure (Draper & Smith, 1966). An equation of the type:

$$Y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$$

is assumed, where Y the value of the predictand is obtained by a linear combination of various predictors $x_1, x_2, \ldots x_n$, regression coefficients $a_1, a_2 \ldots a_n$ and the regression constant a_0. In the development of the equation, the value of the predictand Y is taken as 1 in the event of occurrence of thunderstorm and 0 if thunderstorm does not occur. The stepwise procedure requires a stopping rule. In this study, selection of predictors is stopped when none of the remaining predictors would reduce the variance by 1 % or more. Out of potential parameters mentioned above George K index, Meridional component of wind at 850 hPa, saturation mixing ratio at 500 hPa and temperature at 500 hPa were found to be significant by stepwise screening procedure.

8.1. *Selection of predictors*

The selected predictor to form the regression model and the variance explained by each of them is given in Table 10. From Table 10 the following is noted:

(*i*) George K index, a stability index is a combination of 850 to 500 hPa thermal lapse rate, 850 hPa dewpoint and moisture at 700 hPa level. This stability index is chosen as first potential predictor and is positively correlated with occurrence of thunderstorm. This indicates that higher values of this predictor are favourable for development of thunderstorms.

(*ii*) Meridional component of wind at 850 hPa is positively correlated with occurrence of thunderstorm. This indicates that higher values of this predictor are favourable for development of thunderstorms. Southerly component of wind indicative of presence of lower level cyclonic circulation over the region is favourable for development of thunderstorms.

(*iii*) The saturation mixing ratio at 500 hPa is positively correlated with occurrence, which indicates that dry air at 500 hPa is favourable for occurrence of thunderstorms.

(*iv*) The temperature at 500 hPa is negatively correlated with occurrence, *i.e.*, the colder temperatures at this level will enhance the occurrence.

All the predictors thus selected are physically significant for occurrence or non-occurrence of thunderstorm.

8.2. *Regression equation*

The following regression equation has been developed:

Y = -1.3158 + 0.018068 (George K index) + 0.006825 (Meridional component of wind at 850 hPa) + 0.144380 (Saturation mixing ratio at 500 hPa) - 0.062890 (Dry bulb temperature at 500 hPa)

The values of Y are re-calculated for all the observations of the development sample by using the developed equation. The re-calculated value of Y is made equal to 0 if it is less than 0 and is made equal to 1 if it is more than 1. The values of Y are grouped with an interval of 0.1 and the corresponding number of cases of occurrence/non-occurrence of thunderstorm and the

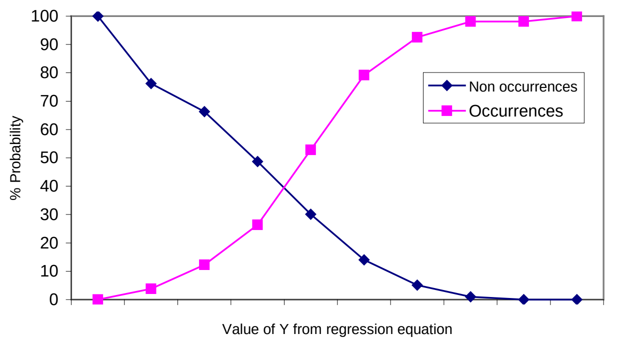

Fig. 4. Percentage probability of occurrence/non occurrence of thunderstorms by multiple regression method

cumulative probability values are calculated. The best fitting curves to these probability values is shown in Fig. 4. The critical value of 0.4 was used as a cut off value for deciding occurrence because it was observed that for *Y* equal to or more than 0.4 the probability of occurrence is higher than that of non-occurrences (Fig. 4).

8.3. *Verification*

Verification of the multiple regression method by using a 2 × 2 contingency table (shown as Table A in Appendix A) as in ESTF and graphical discrimination method with the independent test dataset of May and June for the year 2000 and 2001 gave a POD of 0.65, FAR of 0.48, MR of 0.35, C-NON of 0.67, CSI of 0.41, TSS of 0.32, HSS of 0.30, BIAS of 0.57 and PC of 66.12 %.

9. Conclusions .

Development of expert system "bottom up" by using inductive machine learning techniques to automatically acquire the knowledge about thunderstorm forecasting from the weather dataset have been developed to forecast thunderstorms over Delhi during the pre-monsoon months of May and June. In addition two objective methods, *i.e.*, graphical discrimination and multiple regressions were also developed using the same weather data set to forecast thunderstorms over Delhi during the pre-monsoon months of May and June. The techniques were evaluated using same independent test data. Forecast has been also validated against a persistence forecast. From the verification results the following conclusions are drawn:

A comparison of the verification parameters and skill scores of 3 methods indicate that POD (0.81) is highest and FAR (0.44) & MR (0.19) are lowest in case of ESTF. CSI (0.50), TSS (0.47) and HSS (0.43) are also highest in this case. Percent correct also is highest (71.07 % in this case). With regard to graphical discrimination and multiple regression methods, a comparison of the verification parameters and skill scores of both methods indicate that in case of regression method POD is higher, FAR is lower as compared to graphical discrimination method. CSI, TSS and PC of regression method are better than those of graphical discrimination method. Persistence forecast gave lower POD, higher FAR and lowers values of CSI, TSS and HSS as compared to other 3 methods.

The results given by ESTF are better then other 2 methods and hold promise for application in forecasting.

References

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

Colquhoun, J. R., 1987, "A decision tree method of forecasting thunderstorms and tornadoes", *Wea. Forecasting*, **2**, 337-345.

Conway, B. J., 1987, "Expert system and weather forecasting", *Meteorl. Mag*., **118**, 23-30.

Devrani, A. K. and Mukherjee, 1997, "A forecasting tool for predicting pre-monsoon thunderstorm/duststorm over Jodhpur", *Vatavaran*, **21**.

Dhawan, V. B., 2002, "Study of Thunderstorm and Development of Knowledge Based Expert System for their forecasting over Northwest India", Ph D Thesis, CCS University, Meerut.

Draper, N. R. and Smith, H., 1966, "Applied regression analysis", Wiley and Sons, New York , p407.

Elio, R. J. deHann and Strong, G. S., 1987, "METEOR, An artificial intelligence system for convective storm forecasting", *J. Atmos. Oceanic Technol*., **4**, 19-28.

Galway, J. G., 1956, "The lifted index as predictor of latent instability", *Bull. Am. Meteorol. Soc*., **37**, 528-529.

George, J. J., 1960, 'Weather Forecasting for Aeronautics", Academic Press, New York and London, p673.

Holton, J. R., 1992, "An Introduction to Dynamic Meteorology", Academic Press, San Diego, California.

Kumar, V. R., Chung, C. Y. C. and Lindley, C. A., 1994, "Towards building an expert system for weather forecasting operations", Expert systems with applications , **7**, 2, 373-381.

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

Lee, R. and Passner, J. E., 1993, "The development and verification of TIPS, An expert system to forecast thunderstorm occurrence", *Wea. Forecasting*, **8**, 271-280.

Miller, R. C., 1967, "Notes on analysis and severe storm forecasting procedures of the military weather warning centre", Technical Report No. 20, AWS (MAC), USAF, p139.

Rackliff, P. G., 1962, "Application of an instability index to regional forecasting", *Meteorol. Mag*., **91**, 113-120.

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

Reap, R. M., 1986, "New-6-H thunderstorm probability forecasts for the west", NOAA, National Weather Service Technical Procedures Bulletin, No. 362, p6.

Reap, R. M., 1990, "Six – hour thunderstorms probability forecasts for Kansas – Okalahoma and north east United States", NOAA, National Weather Service Technical Procedures Bulletin, No. 385, p6.

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

Showalter, A. K., 1953, "A Stability index for thunderstorm forecasting", *Bull. Am. Meteorol. Soc*., **34**, 250-262.

Stull Roland, B., 1988, "An introduction to boundary layer Meteorology", Kluwwer Academic Press, 545-551.

Surendra Kumar, 1972, "An objective method of forecasting pre-monsoon thunderstorm/duststorm activity over Delhi and neighbourhood", *Mausam*, **23**, 1, 45-50.

Wallace, J. M. and Hobbs, P. V., 1977, "Atmospheric Science, An introductory Survey", Academic Press, 68-106.

Wilks, D. S., 1995, "Statistical methods in the atmospheric sciences: an introduction", Academic press, New York, p544.

APPENDIX 'A'

Definitions of verification parameters

TABLE A

Contingency table for forecasted and observed phenomena

Observed	*Forecast* Yes	No
Yes	A	B
No	C	D

Verification parameters are listed below

Probability of detection (POD) = A / (A + B)

False alarm rate (FAR) = C / (C + A)

Miss rate (MR) = B / (B + A)

Correct non occurrence (C-NON) = D / (D + C)

Critical success index (CSI) = A / (A + B + C)

True skill score (TSS) = (A/A + B) + (D/D + C) -1

Heidke skill score (HSS) = $2\ (AD - BC) / B^2 + C^2 + 2AD + (B + C)\ (A + D)$

BIAS = (A + C) / (A + B)

Percentage Correct = [(A + D) / (A + B + C + D)] * 100

APPENDIX B

Stability Indices used for forecasting thunderstorms at Delhi

1. *Showalter index (SI)*

The index is given by

$SI = T_{500} - T_{p500}$

Where $Tp500$ is the 500 hPa temperature which a parcel would attain if it is lifted dry-adiabatically from 850 hPa to its condensation level and then moist-adiabatically to 500 hPa.

2. *Jefferson's modified index (TMJ)*

The index is calculated using the relation,
$TMJ = 1.6\ \theta_{w\ 850} - T_{500} - 0.5\ DPD_{700} - 8$

DPD_{700} is the dew point depression at 700 hPa level.
$\theta_{w\ 850}$ is the wet potential temperature at 850 hPa and T_{500} is the temperature at 500 hPa

3. *George index (K)*

This index arithmetically combines the 850-500 hPa temperature differences, the 850 hPa dewpoint (a direct measure of low-level moisture content), and the 700 hPa dewpoint depression (an indirect measure of the vertical extent of the moist layer)

$K = (T_{850} - T_{500}) + Td_{850} - (T\text{-}Td)_{700}$

T and Td are the dry bulb and dew point temperatures at the indicated pressure levels.

4. *Cross Total Index (CTI)*

This is defined as the dew point temperature at 850 hPa (Td_{850}) minus temperature at 500 hPa level (T_{500}).

$CTI = Td_{850} - T_{500}$

5. *Lifted index (LI)*

The lifted index is estimated from

$LI = T_{500} - Tp_{500}$

The method of calculating derived predictor variables used in the study is discussed below. [Holton (1992), Wallace & Hobbs (1977) and Stull (1988)].

Equivalent potential temperature

$\theta_e = \theta \exp (Lr / CpT)$

Saturation mixing ratio

$r_s = 0.622\ \ e_s / p - e_s$

6. *Relative humidity (RH)*

$RH = (r / r_s) * 100\ \%$

Where P is the pressure, T is the temperature in Kelvin and Cp is the specific heat at constant pressure for dry air (1004 J deg^{-1} kg^{-1}). The quantities vapour pressure (e), mixing ratio (r), specific humidity (q) are the variables for saturated air calculated at the pressure and the temperature at the saturation point.