Asian Journal of Water, Environment and Pollution, Vol. 9, No. 4 (2012), pp. 79–86.

Appraisal of Stability Indices for Forecasting Severe Thunderstorms over Kolkata Using Weighted Tree-graph Analysis

Sutapa Chaudhuri and Anirban Middey

Department of Atmospheric Sciences, University of Calcutta, 51/2 Hazra Road
Kolkata – 700 019, India
✉ chaudhuri_sutapa@yahoo.com

Received May 21, 2010; revised and accepted September 14, 2012

Abstract: In the present study the stability indices that are relevant and useful for convective development are taken to assess the significance in forecasting severe thunderstorms over Kolkata (22° 34'N, 88° 22'E). A tree-graph is constructed using the stability indices and the record of severe thunderstorm as vertices. Weights of the edges of different indices are attributed according to two forecast skills. Probability of Detection (POD) and False Alarm Rate (FAR). Two tree-graphs are thus constructed. The most significant and relevant stability indices are obtained using the sub-graph matching analysis. The results reveal that among all the stability indices the lifted index (LI) and the convective inhibition energy (CIN) with the estimated ranges are the most significant stability indices for forecasting severe thunderstorm over Kolkata whereas dew point temperature (Td), convective available potential energy (CAPE) and bulk Richardson number (BRN) are important for indicating the convective development over the region but the occurrence of severe thunderstorms are not assured by these three indices with the estimated ranges.

Key words: Tree-graph, severe thunderstorm, stability index, probability of detection, false alarm rate, sub-graph matching.

Introduction

The pre-monsoon thunderstorms are one of the most hazardous cloud-scale weather phenomena over North Eastern part of India (20°N to 28°N latitude; 84°E to 93°E longitude) enclosing Kolkata (22° 34′ N, 88° 22′ E). Destruction on the surface and the aviation hazards due to such weather phenomena demanded extensive research to develop a model that can forecast the occurrence and severity of such high frequency, chaotic weather phenomena with sufficient accuracy and adequate lead time. The purpose of the present study is to figure out the most adept stability indices with the threshold ranges that can provide the accurate forecast of severe thunderstorm over Kolkata. Broad spectrums of stability indices are now available for thunderstorm study (Jacovides and Yonetani, 1990). The statistical skill score analysis aids to select the significant stability indices for the present study from the broad spectrum which include Lifted Index (LI), Showalter Index (Sw), Total Total Index (TT), K-Index (KI), Convective Available Potential Energy (CAPE), Convective Inhibition Energy (CIN), Bulk Richardson Number (BRN), Boyden Index (BI), Severe Weather Threat (SWEAT), Storm Relative Helicity (SrH), Vorticity Generation Parameter (VGP) and Surface Dew Point Temperature (Td).

The threshold ranges of the indices are computed using the raw data obtained from the radiosonde observation and are then processed and analyzed (Anderson et al.,

*Corresponding Author

1989). The statistical probabilities of the occurrence of severe thunderstorm using the stability indices are estimated (Charba, 1979). A wide variety of forecast verification procedures are available but there can be different views of what constitutes a good forecast. The statistical parameters computed and analyzed for the accuracy measurement of binary forecast are Probability of Detection (POD) and False Alarm Rate (FAR). The accurate forecast can be obtained by proper understanding of the relationship between the stability indices and the convective weather events (Schultz, 1989). The present study adopts a technique to attain the objective that is different from the conventional methods. The concept of graph theory is introduced in the study of thunderstorms. The advantage of graph theoretic approach over other conventional methods is that it can adopt all the complexity, non-linearity and inherent chaos of a system in its heuristic, flexible framework (Bron and Kerbosch, 1973; El-Ghoul, 2006; Chaudhuri, 2007; Chaudhuri and Middey, 2009).

A tree-graph is constructed in this paper using the selected stability indices and the prevalence of severe thunderstorm as the vertices. The weights of the edges of different indices are attributed according to the forecast skills: Probability of Detection (POD) and False Alarm Rate (FAR). Thus two tree-graphs are constructed, one having the weighted edge of probability of detection values and the other with the values of false alarm rate. POD should approach to 100 whereas FAR should approach towards 0 for a good forecast. The central leaf of the tree graph is the severe thunderstorm (STS) and the other leaves of the tree are the stability indices connected to the central leaf (STS) (Figures 1 and 2).

Dichotomous Yes/No variables are used for categorical forecast of the discrete predictands like occurrence and non-occurrence of severe thunderstorms. While there is a long history of research on graph matching, most of the studies have focused on exact sub-graph matching, that is, the sub-graph isomorphism problem, which is known as NP complete. The sub-graph matching of the two trees are chosen for optimum ranges of the weights. The sub-graph matching of the weighted trees reveals that the accuracy of forecast of severe thunderstorms with the combination of LI and CIN is more than 90%. Other stability indices provide some information regarding the convective development over the region but are not so reliable for forecasting severe thunderstorm with high-quality accuracy.

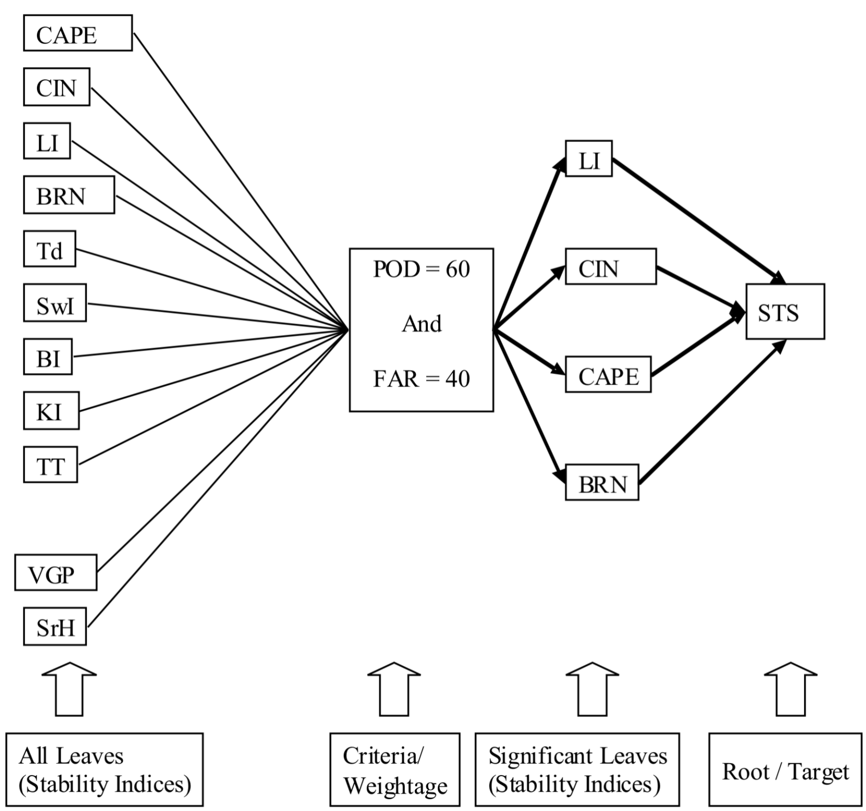

Figure 1: The weighted tree of the stability indices with POD and FAR as their weighted edge.

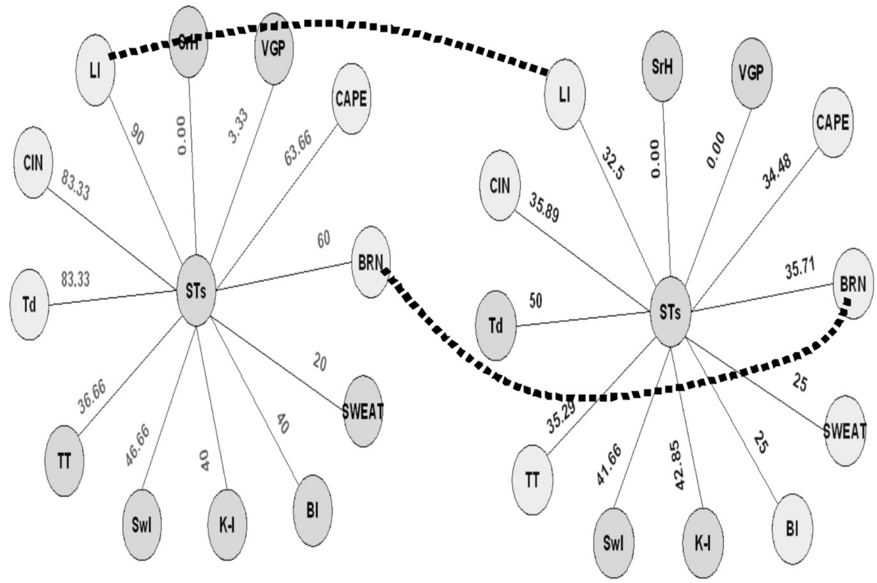

Figure 2: The sub-graph matching of two trees with POD ≥ 60 and FAR ≤ 40.

Materials and Methods

Meteorological Data

The data used in the present study are collected from http://www.weather.uwyo.edu for the pre-monsoon season (April-May) during the period from 1997 to 2006. The location of the study is Kolkata (22° 32′ N, 88° 20′ E). The record of the occurrence of severe thunderstorms is collected from Regional Meteorological Centre, Kolkata, India at 00 UTC and 12 UTC.

The input variables used in the present study are the RS/RW sounding data. The raw data are processed for the computation of the stability indices. The stability indices are then categorized using the threshold values (Table 1). The severe and ordinary thunderstorm days over Kolkata are considered for the present study. The thunderstorm days are considered to be severe when the surface wind speed is greater than or equal to 72 km/h following the Indian Meteorological Department standards.

Table 1: The threshold values of the selected stability indices

Stability indices	*Ranges*
Lifted Index (LI)	≤ –5 °C
Convective Inhibition Energy (CIN)	≤ –150 J/kg
Convective Available Potential Energy (CAPE)	≥ 2000 J/kg
Bulk Richardson No. (BRN)	≥ 45
Showalter Index (SwI)	≤ –4 ºC
Boyden Index (BI)	≥ 99
K-Index (KI)	≥ 35
Total Total Index (TT)	≥ 55
Severe Weather Threat (SWEAT)	≥ 500
Vorticity Generation Parameter (VGP)	≥ 0.6
Storm Relative Helicity (SrH)	≥ 250 m^2/s^2
Surface Dew Point Temperature (Td)	≥ 22 °C

Methodology

In the present study the statistical technique are implemented to examine the efficiency of the forecast of severe thunderstorms, using the computed values of the stability indices and graph theory is adopted for analysis and interpretation.

Statistical Method

The threshold values of the selected stability indices (Table 1) are taken to forecast the severity of the thunderstorms. Forecast verification can easily be understood with reference to categorical forecasts of discrete predictands. Categorical means that the forecast consists of a flat statement that one and only one of the set of possible events can occur. The probability of detection (POD) is the fraction of the occasions when the events selected to be predicted have actually occurred and are predicted too. That is, the POD is the likelihood that the event would be predicted, given that it occurred. The POD for perfect forecast is one, and worst POD is zero.

The False Alarm Rate (FAR) is that proportion of forecast events that fails to materialize. It is the fraction of the occasions when the forecast event does not occur but has been predicted. So the smaller values of FAR is preferred. The best possible FAR is zero and for the worst case it is one.

Familiarization with Graph Theory

Graph theory is a very important branch of theoretical mathematics and computer science that has immense application potential in real world problems. The applicability of graph theory is established in the study of complex atmospheric processes (Chaudhuri and Middey, 2009). A graph is comprised of a set of vertices (V_i) and edges (E_j):

$$G = \{V(G), E(G)\} \quad (1)$$

$V(G)$ represents the set of vertices whereas $E(G)$ represents the set of edges. Every graph is composed of an incidence and adjacency matrix. The incidence matrix of a graph G is represented as:

$$M(G) = [m_{ij}] \quad (2)$$

where m_{ij} is the number of times the vertex (v_i) or edge (e_j) are incident. A graph (G) can also be represented by its adjacency matrix:

$$A = (a_{ij})_{nxn} \quad (3)$$

The adjacency matrix of a graph is a ($n \times n$) matrix $A = (a_{i,j})$ in which the entry $a_{i,j} = 1$, if there is an edge from vertex i to vertex j and is 0 if there is no edge between them. The connectivity of a graph is an important measure of its robustness as a network (Carre, 1979).

An acyclic graph is one that contains no cycles. A tree is a connected acyclic graph. A graph G is a tree when $e = v - 1$. In a tree any two vertices are connected by a unique path. The vertices of degree 1 in a tree are called its leaves. Every non-trivial tree has leaf (Diestel, 2005).

Implementation Procedure

In the present paper the attempt is to find the most effective stability indices which can give accurate forecast of severe thunderstorm over Kolkata. The stability indices are computed with the data obtained before the occurrence of thunderstorm. Thus assigning some critical values to the selected stability indices (Table 1) the severe thunderstorms can be predicted. Statistical parameters—False Alarm Rate (FAR) and Probability of Detection (POD)—are computed for the accuracy measurement of binary forecast. Thus for all the stability indices the POD and FAR are computed with the ten years (1997 to 2006) data during the pre-monsoon months (April and May) (Table 2).

Two identical tree graphs are constructed. These trees have the same "root" that is severe thunderstorm (STS) and the same leaves with stability indices. Every tree contains 13 vertices and 12 edges. The only difference between these two trees is the weights of their edges. In the first tree the weights are assigned according to the POD and FAR of the stability indices (Figure 1). Some criteria are imposed to obtain the effective stability indices with good forecast skill and accuracy for the connectivity of the leaves to the roots. For the first graph where POD is the weight of the edges, the major significant leaves are considered where weights are more than 60. For the second graph where FAR is the weight of the edges, the significant leaves are considered where weights are less than 40. The newly attributed trees are then matched with each other (Figure 2). This is the simple vertex matching technique to find out the order of the stability indices according to their forecast quality and accuracy. In the graph matching process, the purpose is to figure out the stability indices which satisfy both the conditions POD $\geq$ 60 and FAR $\leq$ 40. It is observed that few stability indices are satisfying both the criteria and those stability indices are the most effective for forecasting severe thunderstorm over Kolkata. A new tree is thus formed and the connectivity is assigned according to the conditions POD $\geq$ 60 and FAR $\leq$ 40 (Figure 2). Only few leaves are remained connected to the root (STS) of the trees.

Table 2: The POD and FAR of the selected stability indices

Stability indices	*POD*	*FAR*
Lifted Index (LI)	90	32.5
Convective Inhibition Energy (CIN)	83.33	35.89
Surface Dew Point Temperature (Td)	83.33	50
Convective Available Potential Energy (CAPE)	63.33333	34.48276
Bulk Richardson No. (BRN)	60	35.71429
Showalter Index (SwI)	46.66667	41.66667
Boyden Index (BI)	40	25
K-Index (KI)	40	42.85714
Total Total Index (TT)	36.66667	35.29412
Severe Weather Threat (SWEAT)	20	25
Vorticity Generation Parameter (VGP)	6.666667	50
Storm Relative Helicity (SrH)	3.333333	75

Results and Discussion

Among the various existing formulae for validating the success of the Yes/No forecast, the statistical parameters used for the present study are the Probability of Detection (POD) and False Alarm Rate (FAR). The results show that Lifted Index (LI) with the estimated range provides nearly accurate and quality forecast (Francis et al., 2006). The probability of detection (POD) for LI is 90% (Figure 1) and the false alarm rate (FAR) is 32% (Figures 3 and 4). It is observed that if LI ≤ (-) 5, the coverage factor for the occurrence of severe thunderstorms is 90%. The efficiency of LI is, thus, undeniable in forecasting the severe thunderstorm over Kolkata. The reason lies within the formula of LI; it is expressed as LI = $(T_E - T_P)_{500hPa}$; where T_E is the environmental temperature at 500 hPa and T_P is the parcel temperature at 500 hPa. The Lifted Index is a good indicator of thunderstorm prevalence. Higher the difference between environmental and parcel temperature at 500 hPa, higher will be the negative value of LI and higher will be the available potential energy (APE) for a thunderstorm to become a severe thunderstorm.

Convective Inhibition Energy (CIN) is another index that gives accurate and realistic forecast of severe thunderstorms. The probability of detection (POD) for CIN is 83% and the false alarm rate (FAR) is 35% (Figure 4). CIN represents the amount of energy that must be supplied to a parcel to rise to the Level of Free Convection (LFC). Convective Inhibition is non-zero only if an LFC exists. Lowering of CIN indicates more potential energy available for thunderstorms to generate. Low values of CIN are required to release CAPE and the low value is possible when the Level of Free Convection (LFC) is near the earth surface. This is possible when the surface temperature and moisture is very high over the surface. Thus in the temperate region like Kolkata, when there is adequate amount of moisture and temperature in the atmosphere, CIN can be a good diagnostic tool for the detection of severe thunderstorms.

Dew points (Td) indicate the amount of moisture in the air. The higher the dew points, the higher the moisture content in the air at a given temperature. Dew point temperature is defined as the temperature to which the air would have to cool at constant pressure and constant water vapour content in order to reach the saturation. A

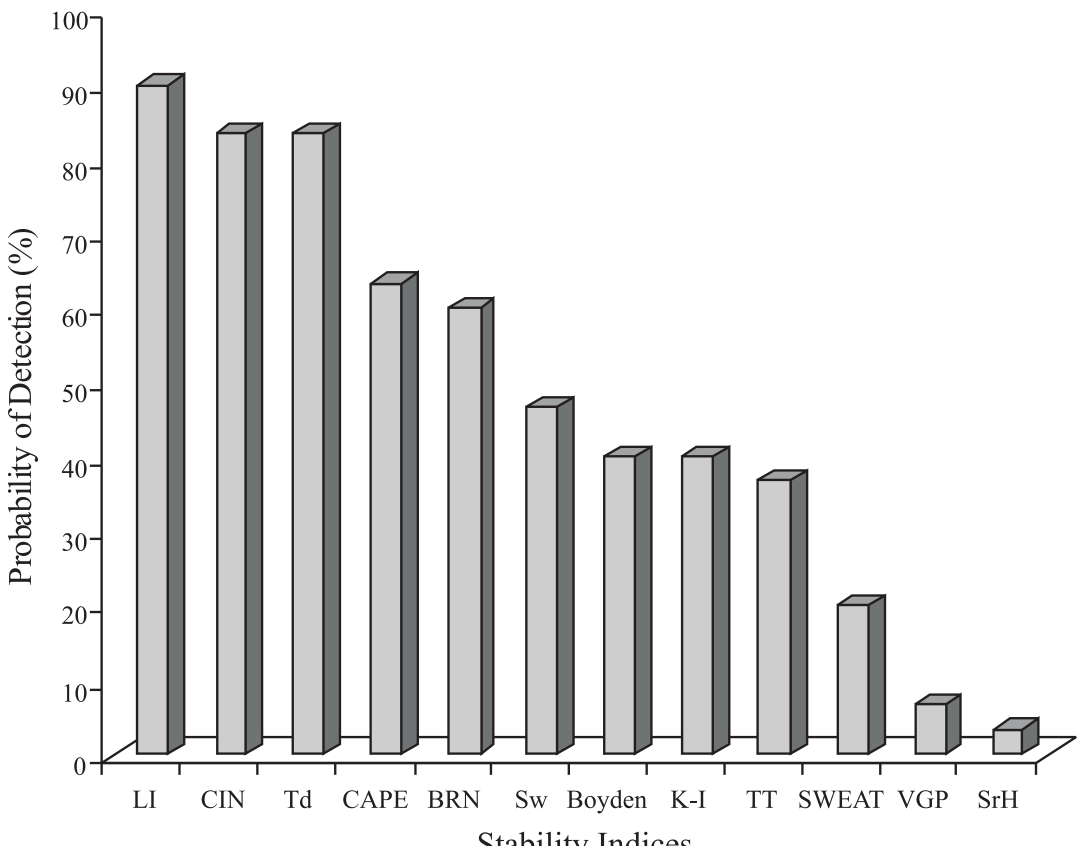

Figure 3: The probability of detection (POD) of different stability indices.

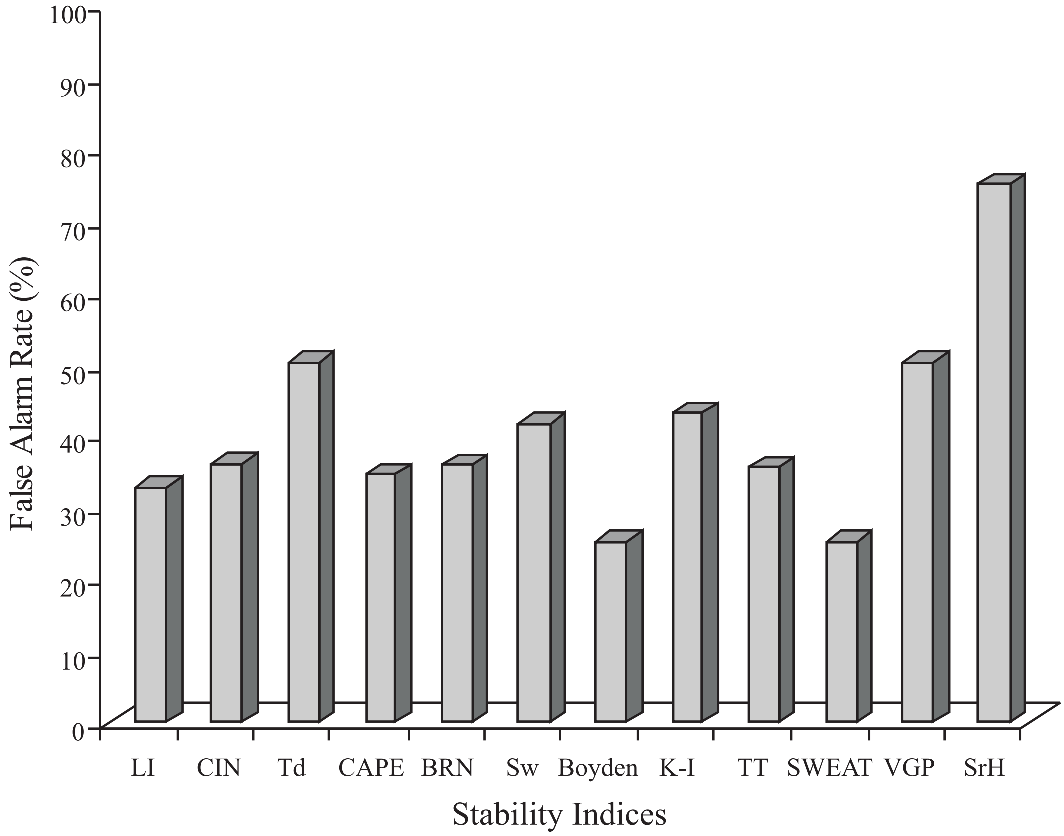

Figure 4: The false alarm rate (FAR) of different stability indices.

state of saturation exists when the air holds the maximum amount of water vapour possible at a certain temperature and pressure. For Td ≥ 22 °C, the probability of detection (POD) is 83% (Table 2) but the FAR is quite high, which is 50%.

Convective Available Potential Energy (CAPE) and Bulk Richardson Number (BRN) are found to be the next significant indices to forecast the severe thunderstorms over Kolkata. The POD for CAPE and BRN are 63% and 60% respectively; whereas the FAR for CAPE and BRN are 34% and 36% respectively (Table 2). Convective Available Potential Energy (Positive Area) is the amount of energy available to a parcel as it freely rises between the Level of Free Convection (LFC) and the Equilibrium Level (EL). CAPE ≥ 2000 J/kg indicates the possibility of severe thunderstorm. However, the range of CIN is more persistent than the range of CAPE for the genesis of severe thunderstorms over Kolkata. Lifted index along with CIN provides better prediction potential of severe thunderstorm over Kolkata.

Bulk Richardson Number (BRN) is an index that assesses the balance between instability (CAPE) and wind shear in a convective environment. This is another predictor of severe thunderstorm with an acceptable degree of forecast quality.

SRH is a measure of the potential for cyclonic updraft rotation in right-moving super cells, and is calculated for the lowest 1 and 3 km layers above ground level. There is no clear threshold value of SRH when forecasting super cells, since the formation of super cells appears to be related more strongly to the deeper layer vertical shear. However, larger values of 0-3 km SRH (greater than 250 m^2s^{-2}) and 0-1 km SRH (greater than 100 m^2s^{-2}) do suggest an increased threat of tornadoes with super cells. However, over Kolkata there are very rare occurrences of tornado during severe thunderstorm events.

The POD of this stability index is very low (3.33%) and FAR is very high (75%). The remaining stability indices, Showalter Index (Sw), Total Total Index (TT), K- Index (KI), Boyden Index (BI), and Severe Weather Threat (SWEAT), Storm Relative Helicity (Srh), Vorticity Generation Parameter (VGP) and Surface Dew Point Temperature (Td) are not significant for the prevalence of severe thunderstorms over Kolkata.

A conventional statistical model, Multiple Linear Regression (MLR) analysis is done to compare the prediction quality with the selected stability indices as input. The output is the surface wind speed during severe thunderstorms. The MLR equation is expressed as

$$Y_P = 42 - 2 \times LI + 0.13123 \times CIN + 0.0005488 \times CAPE + 0.001343 \times BRN \quad (4)$$

where Y_P is the predicted wind speed obtained from MLR analysis.

The prediction using MLR is quite close to the actual observation (Figure 5). The R^2 value of the scatter plot between observed and predicted wind speed using MLR analysis is found to be 0.845 (Figure 6). Percent errors

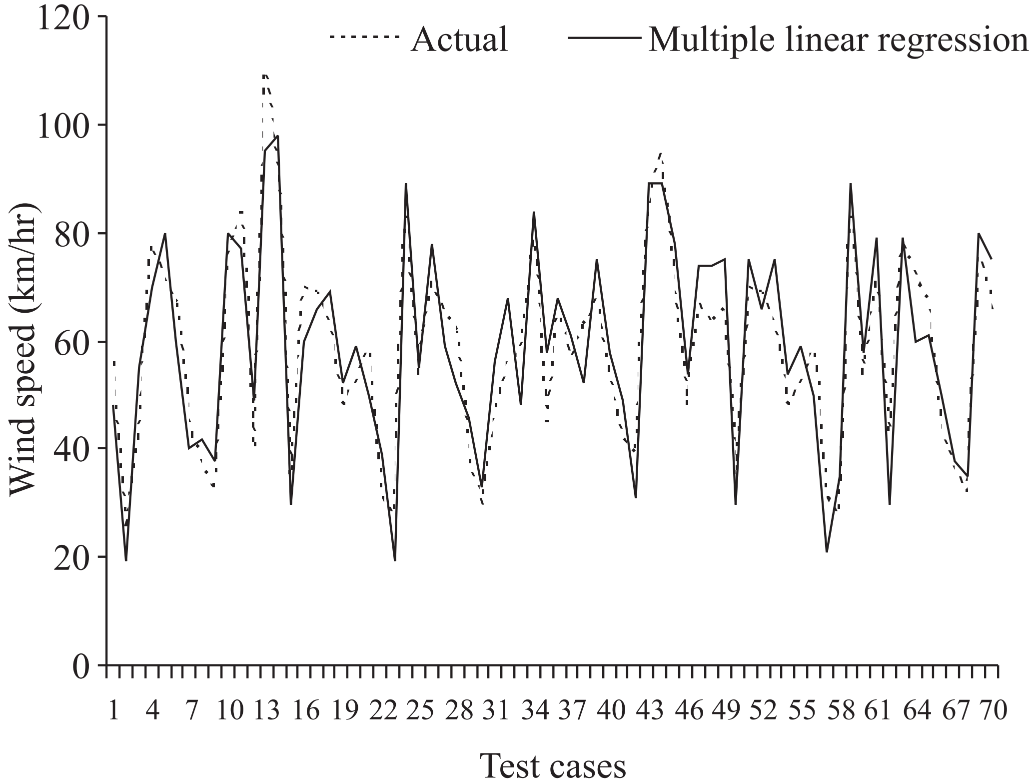

Figure 5: The actual versus forecast through multiple linear regression analysis of wind speed during thunderstorms.

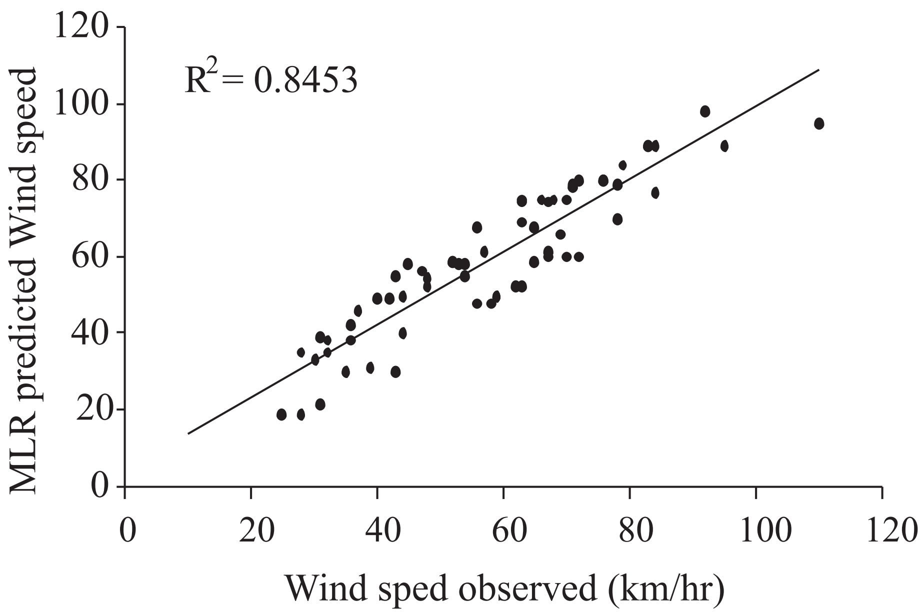

Figure 6: The scatter plot of observed wind speed and multiple linear regression (MLR) forecast of wind speed.

of prediction (PE) are calculated as (Perez and Reyes, 2001):

$$PE = \frac{\langle |y_{dp} - y_{da}| \rangle}{\langle y_{da} \rangle} \quad (5)$$

where $\langle\ \rangle$ denotes the average of the test cases. The predicted and actual values of the parameters are denoted by y_{dp} and y_{da} respectively.

The overall prediction error for MLR analysis is 0.1251 or 12.51%.

Conclusions

The graph matching analysis of categorical forecast of severe thunderstorms with the stability indices shows better skill scores with the estimated ranges of LI and CIN. Forecasting severe thunderstorm over Kolkata, LI and CIN can thus provide effective information and are good predictors. The other two indices—CAPE and BRN—can give us some information about the instability in the atmosphere before the occurrence of severe thunderstorms. The surface dew point temperature (Td) can give useful information about surface moisture content which is very important for lifting mechanism. However, as the false alarm rate for Td is too high, it could thus be misleading when only surface dew point temperature is concerned.

Acknowledgements

The first author acknowledges the financial assistance rendered by the Department of Science & Technology, Govt. of India for conducting the research and grateful to India Meteorological Department for helping us with their data archives.

References

Anderson, T.M., Anderson, C. Jacobson and S. Nilsson (1989). Thermodynamic indices for forecasting thunderstorms in southern Sweden. *Meteor. Mag*., **118:** 141-146.

Baldi, M., Dalu, G.A. and R.A. Pielke (2008). Vertical Velocities and Available Potential Energy Generated by Landscape Variability Theory. *J. of Appl. Met. and Clim*., **47(2):** 397-410.

Bron, C. and J. Kerbosch (1973). Algorithm 457: Finding all cliques of an undirected graph. *Comm. Assoc. Comput. Machinery*, **16:** 575-577.

Carre, B. (1979). Graphs and Networks. Oxford University Press, Oxford, England.

Charba, J.P. (1979). Two to six hour severe local storm probabilities: An operational forecasting system. *Mon. Wea. Rev*., **107:** 268-282.

Chaudhuri, S. (2006). Predictability of chaos inherent in the occurrence of severe thunderstorms. *Adv. Complex Syst.*, **9:** 1-9.

Chaudhuri, S. (2007). Chaotic Graph Theory Approach for Identification of Convective Available Potential Energy (CAPE) Patterns Required for the Genesis of Severe Thunderstorm. *Adv. Complex Syst*., **10:** 413-422.

Chaudhuri, S. and A. Middey (2009). Applicability of Bipartite Graph Model for Thunderstroms Forecast over Kolkata. *Advances in Meteorology*, **2009:** 1-12.

Diestel, R. (2005). Graph theory. Springer-Verlag, New York.

El-Ghoul, M. (2002). The Most General Set and Chaos Graph. *Chaos, Solitons and Fractals*, UK, **18:** 833-838.

El-Ghoul, M., El-Ahmady, A.E. and T. Homoda1 (2006). Retraction of Simplicial Complexes. *International Journal of Applied Mathematics & Statistics*, **4:** 54-67.

Francis, L., Douglas, K. Millar and Shawn G. Gallaher (2006). Evaluating a Hybrid Prognostic-Diagnostic Model that improves Wind Forecast Resolution in Complex Coastal Topography. *J. of Appl. Met. and Clim.*, **45(1):** 155-177.

Jacovides, C.P. and T. Yonetani (1990). An evaluation of Stability indices for Thunderstorm prediction in Greater Cyprus. *Weather and Forecasting*, **5:** 559-569.

Schultz, P. (1989). Relationship of several stability indices to convective weather events in Northeast Colorado. *Wea. Forecasting*, **4:** 73-80.