METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **18**: 238–249 (2011)
Published online 21 January 2011 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.240

Nowcasting thunderstorms with graph spectral distance and entropy estimation

Sutapa Chaudhuri* and Anirban Middey
Department of Atmospheric Sciences, University of Calcutta, Kolkata 700019, India

ABSTRACT: The aim of the present study is to forecast thunderstorms over Kolkata (22°32′N, 88°20′E), India, during the pre-monsoon season (April–May) with graph spectral distance and entropy analysis. Graph vertices represent points connected by lines or edges, and lifting condensation level, convective condensation level, level of free convection, freezing level, level of neutral buoyancy and the surface level are taken as the input of the graph vertices. The variation in the most probable distance between the vertices is investigated. The result reveals a particular orientation of the vertex distances for thunderstorm days which is significantly different from the non-thunderstorm days. The reference graphs for thunderstorm and non-thunderstorm days are formed using the most probable vertex distances. The spectral distance between the reference graph and the graphs corresponding to thunderstorms are computed with the data collected during the period 1997–2009. The entropies, or the measure of disorderliness or uncertainty, are estimated for the graph distance matrices. The result shows that the thunderstorm days possess lower distance entropy than the non-thunderstorm days. This indicates that the reference graph that has been constructed for thunderstorms is more consistent. The result further depicts that the forecast accuracy through the present method is 98% with 1 h lead time, whereas the accuracy is 93% with 6 h lead time. The forecast is validated with the India Meteorological Department observations for the years 2007, 2008 and 2009. Copyright © 2011 Royal Meteorological Society

KEY WORDS severe thunderstorm forecast; graph theory; statistical probability; entropy; spectral distance

Received 21 April 2010; Revised 25 October 2010; Accepted 28 October 2010

1. Introduction

The most significant weather phenomenon during the pre-monsoon season (April–May) over the northeastern part of India (20–28 °N; 84–93 °E) enclosing Kolkata (22°32′N, 88°20′E) is the outburst of severe local storms. Thunderstorms are reported as light, medium or heavy according to the nature of the lightning and thunder, the type and intensity of the precipitation and hail if any, the speed and gustiness of the wind, and the appearance of the clouds (Byers and Braham, 1949). There are some other specific criteria for defining severe local storms depending on the forecast locations and climatology (Johns and Doswell, 1992). In the present study, thunderstorms with wind gust ≥65 km h^{-1}, cloud mass and convective development as evident from satellite and radar imageries, and lightning activity are considered as severe thunderstorms because the prevalence of hail and tornadoes are very rare over Kolkata. The remaining thunderstorms are considered as ordinary thunderstorms. The purpose of the study is to forecast both severe and ordinary thunderstorms. The large-scale flow over and around the region comprises a shallow layer of moist southerlies and south-westerlies from the Bay of Bengal near the surface and dry north-westerlies aloft. This type of atmospheric conditions prevail over the region almost every day during the pre-monsoon season (April–May). However, thunderstorms do not occur each day. The initiation and development of such storms require reasonably high surface temperature, atmospheric instability, presence of adequate moisture in the lowest part of the troposphere, trigger and pulling mechanisms to raise the warm and moist air upward. Quantification of the requirements is needed. Interactions between convection and large-scale weather phenomena are well established (Mukhopadhyay *et al*., 2005, 2009). The devastation at the surface and the aviation hazards aloft due to such severe local storms demanded extensive research on pre-monsoon thunderstorms to develop a model that can forecast the genesis and severity with considerable accuracy (at least 90%) and adequate lead time (at least 6 h). The operational weather forecast techniques are mainly based on numerical and statistical methods (Litta and Mohanty, 2008; Chaudhuri, 2008a). The numerical and statistical methods of forecasting small scale weather phenomena depict some restrictions due to the inadequate network of observatories. In recent times, sophisticated and precise computational methods have been developed that can ease the study of complex atmospheric processes (Abraham *et al*., 2001; Marzban and Witt, 2001; Benjamin *et al*., 2007; Chaudhuri, 2008b; Chaudhuri, 2010). The high frequency, small scale weather phenomena such

* Correspondence to: Sutapa Chaudhuri, Department of Atmospheric Sciences, University of Calcutta, Kolkata 700019, India.
E-mail: chaudhuri_sutapa@yahoo.com

Copyright © 2011 Royal Meteorological Society

Table I. Fitted normal distribution parameters of the observed distances between the vertices for thunderstorm days.

Events	Mean distances between graph vertices with standard deviations				
	Surface to LCL (m)	Surface to CCL (m)	Surface to LFC (m)	Surface to FL (m)	Surface to LNB (m)
Thunderstorm	552 (±250.8)	1610 (±285.8)	2100.1 (±428)	4458 (±346)	13272.2 (±1320)
No thunderstorm	646.2 (±362.4)	2125 (±348.6)	2689.4 (±1053.2)	4638 (±506)	11 235 (±2503)

as thunderstorms are highly non-linear and chaotic in nature. The deterministic chaos inherent in the occurrence of such weather events has been identified (Chaudhuri, 2006).

The purpose of the present paper is to bring the concept of graph theory to the study of thunderstorms. The method of graph theory sometimes performs better than the other conventional techniques because it can deal with the complexity, non-linearity and inherent chaos of a particular system in its flexible structure (El-Ghoul, 2002; Chaudhuri, 2007; Chaudhuri and Middey, 2009). Nowcasting and very short range forecasts of thunderstorms is one of the major challenges for atmospheric scientists and meteorologists (Charba, 1979). The significance of the altitudes of lifting condensation level (LCL), convective condensation level (CCL), level of free convection (LFC), freezing level (FL) and the level of neutral buoyancy (LNB) from the surface level and their intermediate distances are explored in this study using the upper air radiosonde/rawinsonde data of 0000 and 1200 UTC during the pre-monsoon season (April–May) over the Kolkata station. A graph distance matrix is formed with 10 years (1997–2006) data. The spectral distances and entropies are estimated for various thunderstorm days. The thunderstorm days of 2007, 2008 and 2009 are selected for the validation of the graph theoretical forecast with real time Doppler weather radar output, satellite products and the India Meteorological Department (IMD) forecast.

2. Materials and methods

2.1. Meteorological data

The upper air radiosonde/rawinsonde (RS/RW) data at 0000 and 1200 UTC used in the present study were collected from www.weather.uwyo.edu during the pre–monsoon season (April–May) for the years 1997–2009. The location of the study is Kolkata (22°32′N, 88°20′E) (station no. 42 809). The record of the occurrences of 140 thunderstorms (both severe and ordinary) are collected from the Regional Meteorological Centre, (RMC), Kolkata, India. The records include the time of occurrence, duration, maximum wind speed, direction of squall advancement and available satellite and radar observations.

The input variables used in the present study are the RS/RW sounding data. The altitudes of the lifting condensation level (LCL), convective condensation level (CCL), level of free convection (LFC), freezing level (FL) and level of neutral buoyancy (LNB) from the surface level are computed and taken as the input parameters for the graph spectral distance and entropy estimation (Table I). Twenty thunderstorm days, along with non-thunderstorm days of 2007, 2008 and 2009, were selected for the validation of the graph theoretical forecast with real time Doppler weather radar output, satellite imageries and Indian Meteorological Department (IMD) observations.

2.2. Methodology

Statistical probability distribution function and graph distance analysis with entropy and spectral distance estimation are adopted as the methodology of the present study.

2.2.1. *Graph theory – a brief outline*

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 includes a set of vertices (V_i) and edges (E_j) (Appendix A):

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

$V(G)$ represents the set of vertices whereas $E(G)$ represents the set of edges.

Each graph is composed of an incidence and adjacency matrix. The incidence matrix of a graph G is represented as:

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

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:

$$A = (a_{ij})_{n\times n}. \tag{3}$$

The adjacency matrix of a graph is an ($n \times n$) 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 eccentricity, $\varepsilon(v)$, of a graph vertex (v) in a connected graph G is the maximum graph distance between (v) and any other vertex (u) of G. All vertices are defined to have infinite eccentricity for a disconnected

Copyright © 2011 Royal Meteorological Society

graph (Biggs, 1993). The maximum eccentricity is the graph diameter, whereas the minimum eccentricity is the graph radius. The distance between two vertices in a graph is the number of edges in a shortest path connecting them. This is also known as the geodesic distance. In a weighted or digraph graph, each edge is associated with some values depending on the application.

2.2.2. *Probability distribution – a brief outline*

A probability distribution describes the values and probabilities that a random event can take place (Wilks, 1995). A continuous distribution describes events over a continuous range, where the probability of a specific outcome is zero.

Discrete distributions are characterized by a probability mass function, p:

$$F(x) = P_r[X \leq x] = \sum_{x_i \leq x} p(x_i) \quad (4)$$

A probability distribution is called continuous if its cumulative distribution function is continuous, which means that it belongs to a random variable X for which $P_r[X = x] = 0$ for all x in R. These distributions can be characterized by a probability density function which is a non-negative Lebesgue integral function f defined on the real numbers as:

$$F(x) = P_r[X \leq x] = \int_{-\infty}^{x} f(t)\mathrm{d}t \quad (5)$$

The normal distribution or Gaussian distribution is an important family of continuous probability distributions that is applicable in many fields. Each member of the family may be defined by two parameters, location and scale: the mean (μ) and variance (standard deviation squared) σ^2, respectively to indicate that a real-valued random variable X is normally distributed with mean μ and variance $\sigma^2 \geq 0$ as $X \sim N$ (μ, σ^2). The continuous probability density function of the normal distribution is the Gaussian function:

$$\phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma^2\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sigma}\varphi(\frac{(x-\mu)}{\sigma}), \quad x \in R. \quad (6)$$

2.2.3. *Implementation procedure*

The purpose of the present study is to find the set of most probable vertex distances among the altitudes of the selected levels from the surface level that play significant role in the initiation and severity of thunderstorms (Neumann, 1971; Jacovides and Yonetani, 1990; Anderson, 1996; Manohar *et al*., 1999; Baldi *et al*., 2008). Ten years sounding data are analysed within the period from 1997 to 2006 for the pre-monsoon months of April and May. The probability distribution of various thunderstorm days from the 10 years data has identified a particular range of vertex distances for thunderstorm days which are significantly different from the non-thunderstorm days (Table I).

Graphs with six vertices V_{LCL}, V_{CCL}, V_{LFC}, V_{FL}, V_{LNB} and V_{surface} are constructed. The maximum possible distances (eccentricity) from V_{surface} to other five vertices are estimated (Table II). The normal distributions of the set of vertex distances are plotted to find the most probable combination that is responsible for the genesis of thunderstorms. Each graph corresponding to the thunderstorm days has 15 individual edges (Figure 1): surface to LCL (e_1), surface to CCL (e_2) surface to LFC (e_3), surface to FL (e_4), surface to LNB (e_5), LCL to LFC (e_6), LCL to CCL (e_7), LCL to FL (e_8), LCL to LNB (e_9), LFC to CCL (e_{10}), LFC to FL (e_{11}), LFC to LNB (e_{12}), FL to CCL (e_{13}), FL to LNB (e_{14}) and LNB to CCL (e_{15}). The ranges and patterns of the edges are observed for different thunderstorm days. Two different categories of thunderstorms are considered that occurred on 24 April 2000 (severe thunderstorm) and 30 April 1997 (ordinary thunderstorm) with CAPE values 3055 and 3835 J kg^{-1} respectively (Figure 2). The figure shows that the severe thunderstorm day has less CAPE values than the ordinary thunderstorm day. However, the CIN values for severe and ordinary thunderstorm days are 99 and 178 J kg^{-1} respectively. CAPE is a very important parameter for understanding the character of thunderstorms. However, the release of the available energy (CAPE) to the atmosphere as kinetic energy for generating thunderstorm activity requires some inhibition energy (CIN) to overcome. If that much inhibition energy is absent then CAPE cannot be released in the atmosphere. One of the crucial reasons for selecting different altitudes of the important levels (LCL, CCL, LFC, LNB and FL) is that the convective energies (CAPE and CIN) enclosed within the levels sometimes mislead the forecasters, whereas the distance analysis of the altitudes of the significant levels may provide a more realistic forecast.

The distance graphs are normalized between 0 and 1 (Comric, 1997):

$$X_i = 0 + 1.0\frac{(O_i - O_{\min})}{(O_{\max} - O_{\min})} \quad (7)$$

Table II. Most probable distances between the vertices of the reference graphs for thunderstorm days.

Events	Surface to LCL (m)	Surface to CCL (m)	Surface to LFC (m)	Surface to FL (m)	Surface to LNB (m)
Thunderstorm	550	1600	2000	4450	13 300
No thunderstorm	650	2100	2700	4600	10 000

Copyright © 2011 Royal Meteorological Society

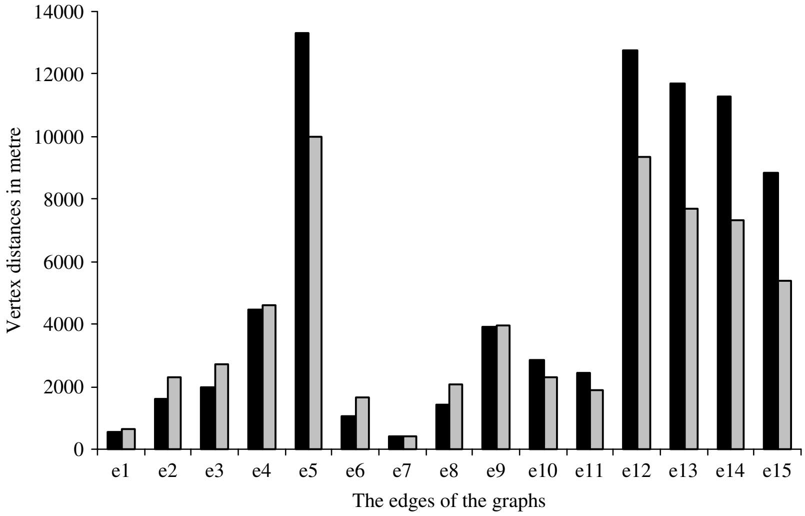

Figure 1. The reference graphs for thunderstorm and no thunderstorm days with vertex distances and edges of the graphs through 10 year (1997–2006) data analyses during pre-monsoon season over Kolkata. ■: reference graph for severe thunderstorm; □: reference graph for no thunderstorm.

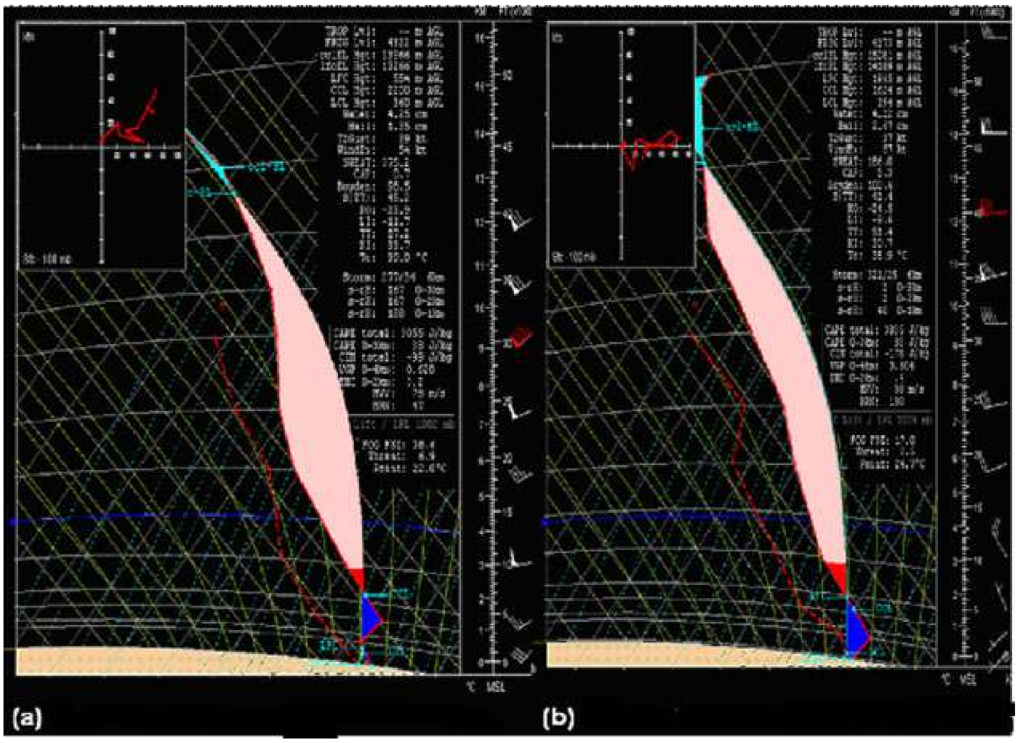

Figure 2. Radiosonde plot showing almost same CAPE value but different CIN values for (a) severe thunderstorm with wind speed 110 km h^{-1} and (b) ordinary thunderstorm with wind speed 40 km h^{-1} over Kolkata. This figure is available in colour online at wileyonlinelibrary.com/journal/met

where,

O_i = actual number of observations
O_{max} = maximum number of observations
O_{min} = minimum number of observations

The distance ≥15 000 m is designated as (O_{max}) and is taken to be 15 000 m and the minimum distance (O_{min}) is taken to be 0.

Equation (7) thus, reduces to:

$$X_i = 0 + 1.0\frac{(O_i)}{15\,000} \quad (8)$$

The entropy measure of the edges is:

$$E = -\sum(w_i \times \ln\ w_i) \quad (9)$$

where, w_i is the normalized weight of the edge e_i.

Copyright © 2011 Royal Meteorological Society

Table III. The cross-validation of the graph entropy and spectral distance patterns with Doppler Weather Radar and IMD forecast.

Dates	Entropy	Spectral distance from reference graph		Forecast with lead time	Actual observation	Maximum wind gusts (km h^{-1})
		Thunderstorm	No Ts			
09 April 2007	1.12	0.12	0.53	Thunderstorm (6 h)	Thunderstorm	70
12 April 2007	1.21	0.022	0.68	Thunderstorm (4 h)	Thunderstorm	71
26 April 2007	1.04	0.11	0.72	Thunderstorm (5 h)	Thunderstorm	59
12 May 2007[a]	1.68	0.52	0.03	No thunderstorm (1 h)	Thunderstorm	52
21 May 2007	1.15	0.02	0.39	Thunderstorm (8 h)	Thunderstorm	81
08 May 2008	1.21	0.20	0.63	Thunderstorm (2 h)	Thunderstorm	48
24 May 2008	0.98	0.03	0.76	Thunderstorm (8 h)	Thunderstorm	72
03 May 2009	0.99	0.09	0.73	Thunderstorm (4 h)	Thunderstorm	82
11 May 2009	1.10	0.11	0.59	Thunderstorm (6 h)	Thunderstorm	87
15 May 2009	1.15	0.03	0.64	Thunderstorm (6 h)	thunderstorm	67

Ts - Thunderstorm.
[a] This is the event missed by this method.

Table IV. Contingency table used for computation of forecast quality.

Forecast	Observed		
	Yes	No	
Yes	8 (a)	1 (b)	9 ($a+b$)
No	2 (c)	9 (d)	11 ($c+d$)
	10 ($a+c$)	10 ($b+d$)	20 ($N=a+b+c+d$)

In information theory, entropy is a measure of the uncertainty associated with a random variable. The distance entropies are estimated for thunderstorm and non-thunderstorm days.

Spectral analyses in graph theory facilitate in fixing the bounds on the distributions of eigenvalues. Some eigenvalues have been referred to as the algebraic connection patterns of a graph (Chung, 1992). Graph eigenvalues have applications in wide areas and in different contexts. The spectral distance between the two graphs G and H are computed using the K largest positive eigenvalue of the corresponding adjacency matrix as:

$$d(G,H)=\sqrt{\sum_{i=1}^{K}(\lambda_i-\mu_i)^2/\min\{\sum_{i=1}^{K}\lambda_i{}^2,\sum_{i=1}^{K}\mu_i^2\}} \tag{10}$$

where, λ_i represents the eigenvalues of graph G and μ_i represents the eigenvalues of graph H. If the two graphs are similar then the spectral distance between them becomes 0. Two reference graphs, one for thunderstorm and the other for non-thunderstorm days, are constructed with the statistical analysis (Table I). The spectral distance between the graph of daily observation and the reference graph for both thunderstorm and non-thunderstorm days are evaluated. The most probable vertex distances are shown in Table II. Validation of the model is made with the observation of 2007, 2008 and 2009 (Table III). To ensure a better quantitative measure, the forecast skills of the model are computed from a contingency table (Table IV) using True Skill Statistic (TSS), Heidke Skill Score (HSS) and Yule's Q.

The formulae of the skill scores are:

$$\text{TSS}=\frac{(ad-bc)}{(a+c)(b+d)} \tag{11}$$

$$\text{HSS}=\frac{2(ad-bc)}{[(a+c)(c+d)+(a+b)(b+d)]} \tag{12}$$

$$\text{Yule's Q}=\frac{(ad-bc)}{(ad+bc)} \tag{13}$$

3. Results and discussion

The vertex distance between V_{surface} and V_{LFC} is observed to be much less on thunderstorm days than that of non-thunderstorm days, indicating that for the genesis of thunderstorms the free convection initiates from a much lower depth of the atmosphere, whereas the vertex distance from V_{surface} to V_{LNB} is greater for thunderstorm days than that of non-thunderstorm days, depicting that the depth of cloud has to be more for severe thunderstorms. However, there is hardly any distinct pattern for the vertex distance V_{surface} to V_{LCL} (Figure 3), whereas the vertex distance V_{surface} to V_{CCL} shows a significant variation between thunderstorm and non-thunderstorm days (Figure 4). The CCL comes down during thunderstorms over Kolkata in the pre-monsoon season (April, May), while it remains much above during non-thunderstorm days (Table I). The

Copyright © 2011 Royal Meteorological Society

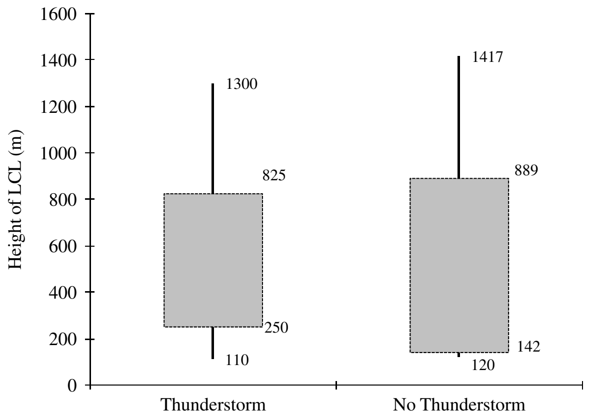

Figure 3. Box plot showing the variation in altitudes of lifting condensation level (LCL) during thunderstorm and non-thunderstorm days over Kolkata.

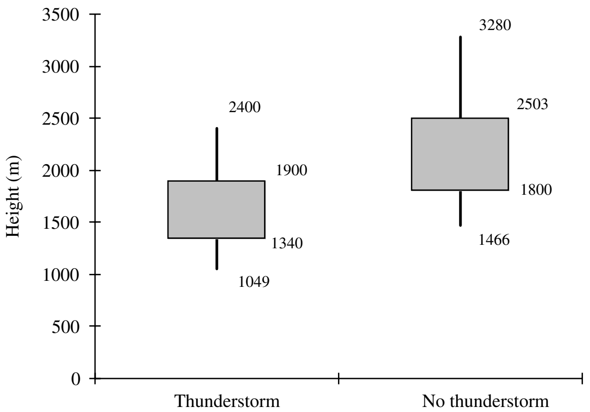

Figure 4. Box plot showing the variation of convective condensation level (CCL) during thunderstorm and non-thunderstorm days over kolkata.

CCL represents the cloud base when air is heated from below to the convective temperature, without mechanical lift. The vertex distance V_{surface} to V_{FL} shows a significant difference between thunderstorm and non-thunderstorm days (Figure 5). The freezing level determines the depth of the atmosphere that is above freezing (0 °C). If the freezing level is high in the atmosphere, hailstones will have more time to melt than if the freezing level is close to the surface. A high freezing level also decreases the vertical depth in which hailstone formation and growth is possible.

The statistical distribution of the vertex distances for both thunderstorm and non-thunderstorm days are shown (Figures 3–7). The box-and-whisker diagrams show that the variations in the vertex distances are more for non-thunderstorm days while are comparatively less for thunderstorm days. The bottom and top of the box represent 25$^{\text{th}}$ and 75$^{\text{th}}$ percentile respectively and the whiskers represent minimum and maximum values. It can thus be surmized that the analyses of the altitudes of LCL, CCL, LFC, FL and LNB from the surface level might aid the forecast of thunderstorms more precisely using the graph distance optimization method.

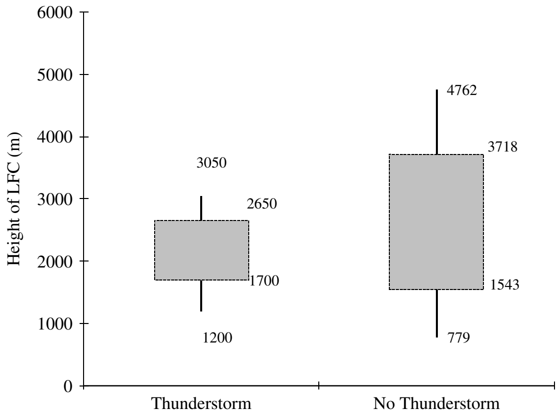

Figure 5. Box plot showing the variation in altitudes of level of free convection (LFC) during thunderstorm and non-thunderstorm days over Kolkata.

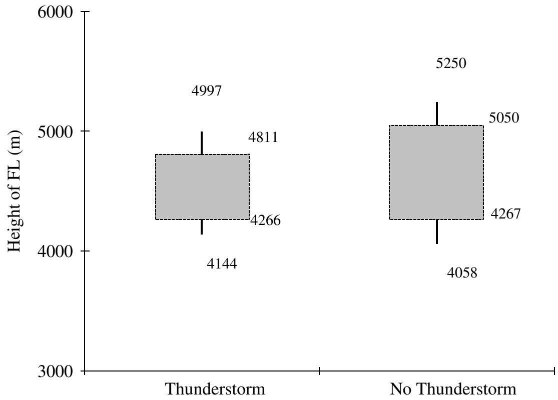

Figure 6. Box plot showing the variation in altitudes of freezing level (FL) during Thunderstorm and non-thunderstorm days over Kolkata.

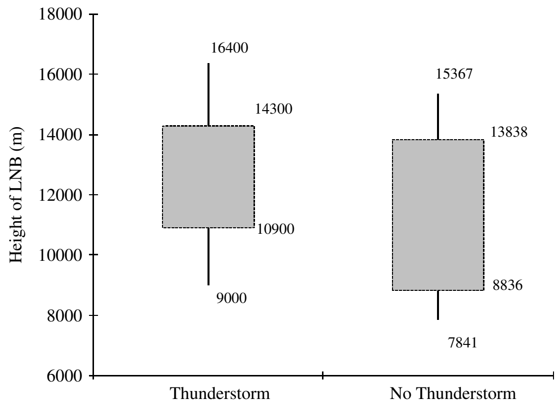

Figure 7. Box plot showing the variation in altitudes of level of neutral buoyancy (LNB) during thunderstorm and non-thunderstorm days over Kolkata.

Normal probability distribution is implemented to the sample set of thunderstorm and non-thunderstorm days. The maximum probability density for thunderstorms is observed to be in a particular range (Table I). However, the probability densities for non-thunderstorm days are observed to be remarkably different from thunderstorm days (Table II). The selected vertex distances of the

Copyright © 2011 Royal Meteorological Society

parameters from the surface ($V_{surface}$) (Table II) are very close to that of the observed vertex distances (Table I). Two distinct patterns are observed for vertex distances, surface to CCL, surface to LFC, surface to FL and surface to LNB (Figures 4–7) for thunderstorms and non-thunderstorm days. However, there is hardly any distinct pattern for the vertex distance, surface to LCL (Figure 3).

Two reference graphs, one for thunderstorms (G_{TS}) and the other for non-thunderstorms (G_{NTS}) are formed with the most probable vertex distances (Table II). Computation and analyses of the spectral distances reveal that the spectral distance [$d(G, H)$] between the reference graph (G_{TS}) of thunderstorms and the graph corresponding to actual thunderstorms (H_{TS}) is minimum. It indicates that the reference graph and the graph for actual occurrence are comparable. Further, the spectral distance [$d(G, H)$] between the reference graph (G_{NTS}) of non-thunderstorm and the graph for an actual non-thunderstorm day (H_{NTS}) is also observed to be minimum.

Estimation of the entropies of the 15 edges of the graph for all the thunderstorm days show a clear distinction between thunderstorms and non-thunderstorms (Figure 8). The entropies of the two categories of the events, thunderstorm and non-thunderstorm from 1997 to 2006 are subjected to normal probability distributaion analyses (Figure 8). It is observed that the maximum probability of the distance entropy before thunderstorms is about 1.09 ($\pm$0.11) and that of non-thunderstorms is 1.72 ($\pm$0.32).The comparison gives a clear idea about the entropy distributions of two distinct categories (Figure 9). For non-thunderstorms, the entropy distributions shows a broad range, that is the standard deviation is higher (0.32) while in case of thunderstorms, the probability distribution of entropy shows a sharp peak with low standard deviation (0.11) (Figure 10). The lower entropy values for thunderstorms reveals that the uncertainity in the distances of the selected altitudes from the surface are less for thunderstorms days than that of the non-thunderstorm days. Thus, the entropy estimation and spectral distance analyses provide better predictability of the prevalence of thunderstorm and its intensity.

4. Validation with observation

The new approach of graph spectral distance and entropy estimation for predicting the occurrence of thunderstorms and its severity is validated with the real time Doppler Weather Radar products, satellite imageries and Indian Meteorological Department (IMD) observations (Table III). Figure 11 represents the INSAT (KALPANA-1) hourly IR images of intense convective cloud development with cloud-top temperature and advancement towards Kolkata on 15 May 2009. The cloud-top temperature was recorded to be −70 °C. Graph spectral distance and entropy analyses provide some clear information about the upcoming event seven hours before the squall appears over Kolkata. The INSAT (KALPANA-1) hourly IR image of intense convective cloud development and advancement towards Kolkata on 19 May 2007 is shown in Figure 12. Figure 13 represents the Doppler imageries of sequential advancement (clockwise from top left) of squall line towards Kolkata on 26 April 2007. It is observed that in all the above 3 days the vertex distances (surface to LCL, surface to CCL surface to LFC, surface to FL and surface to LNB) are within the

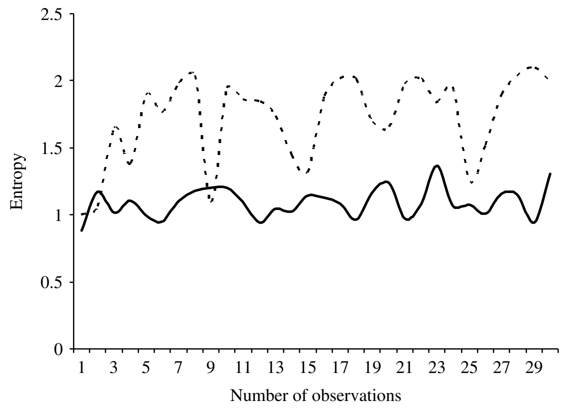

Figure 8. The entropy of distance graph for thunderstorm and non-thunderstorm days over Kolkata. ——: thunderstorm; - - - -: no thunderstorm.

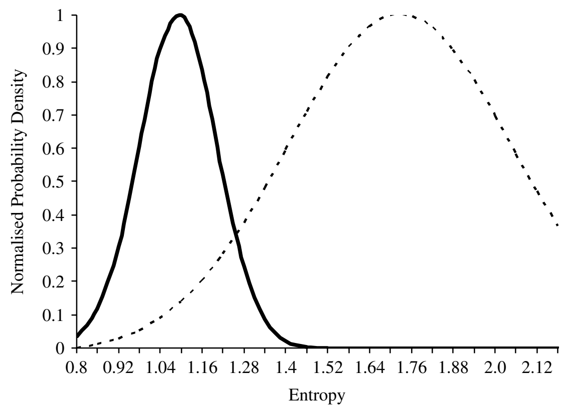

Figure 9. Diagram showing the probability distributions of the entropy values for thunderstorm and non-thunderstorm days. ——: thunderstorm; - - - -: no thunderstorm.

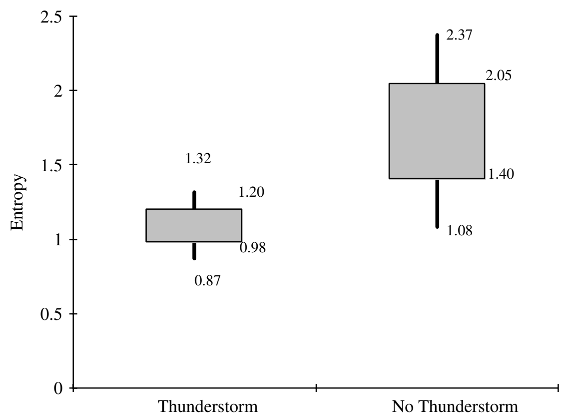

Figure 10. Box plot showing the variation in the distance entropy during thunderstorms and non-thunderstorm days over Kolkata.

Copyright © 2011 Royal Meteorological Society

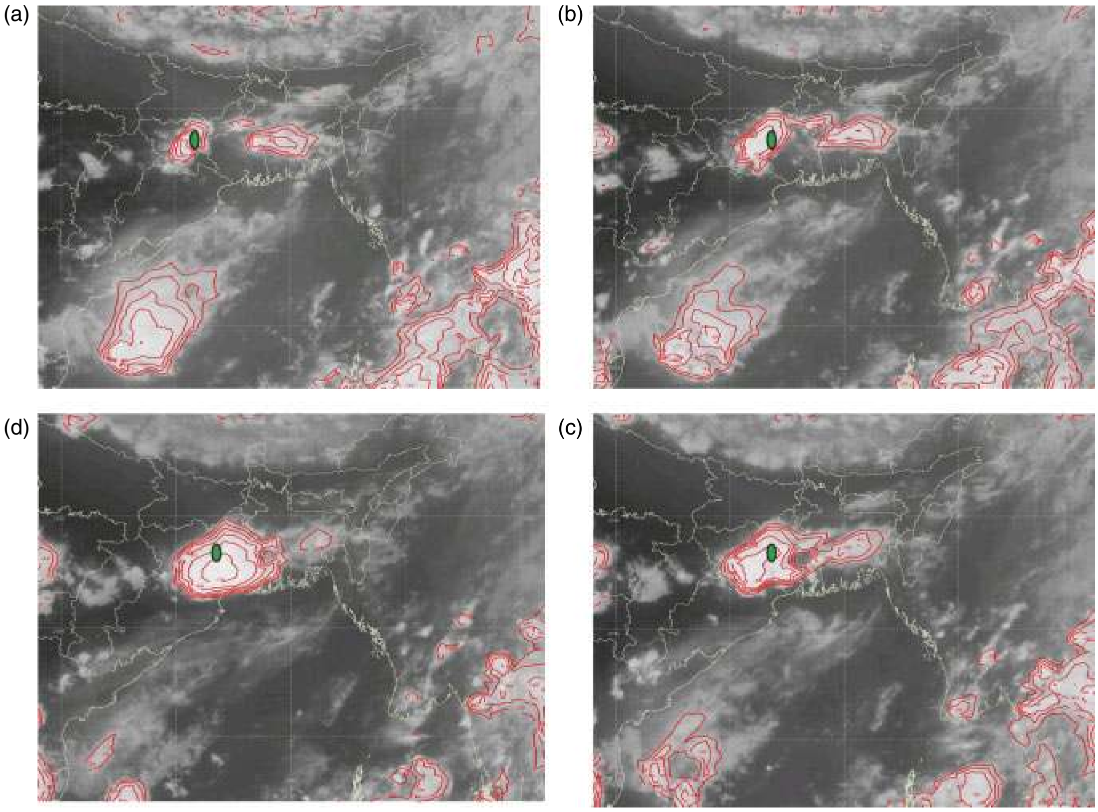

Figure 11. INSAT (KALPANA-1) Satellite IR Images (a, b, c and d) showing the advancement of squall with cloud top temperature contour towards Kolkata (clockwise from top left) on 15 May 2009 (a validation day). This figure is available in colour online at wileyonlinelibrary.com/journal/met

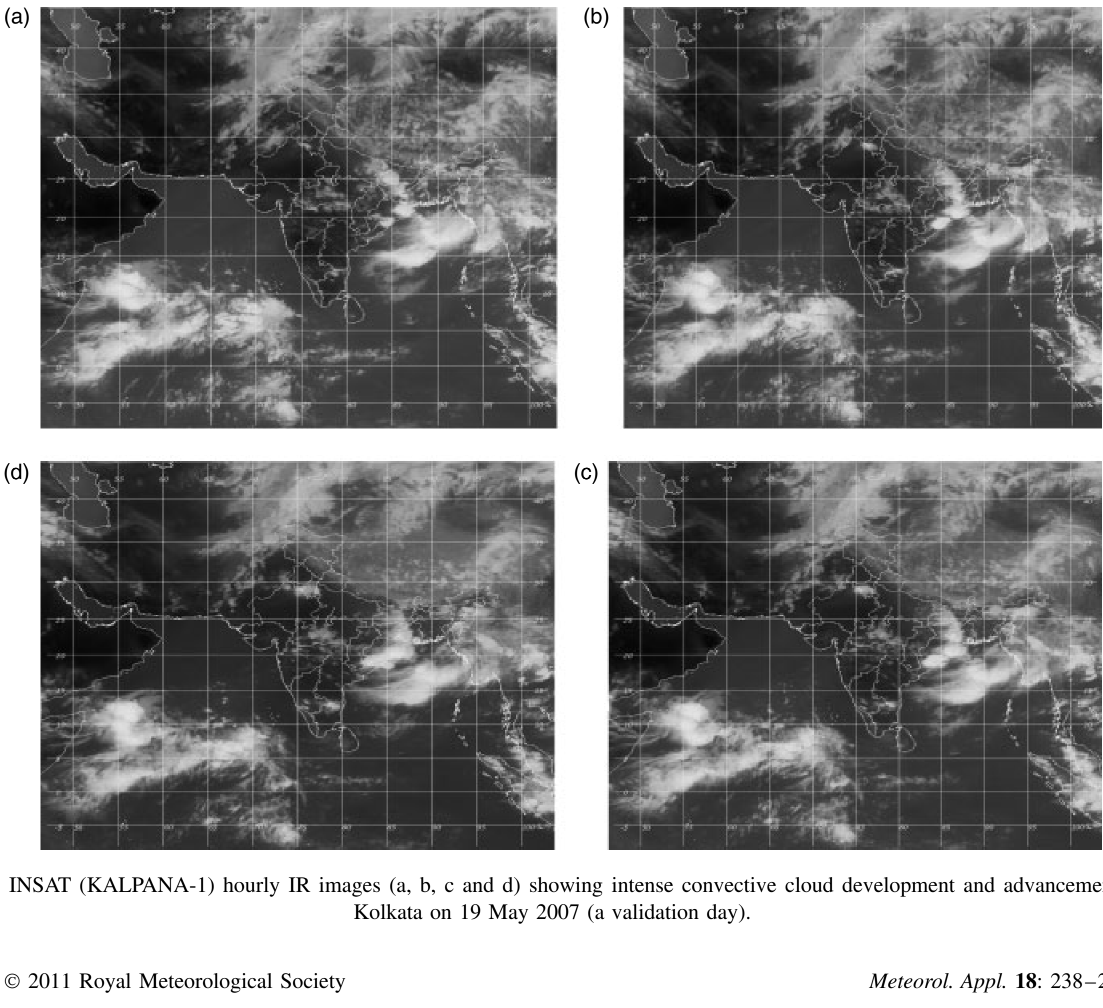

Figure 12. INSAT (KALPANA-1) hourly IR images (a, b, c and d) showing intense convective cloud development and advancement towards Kolkata on 19 May 2007 (a validation day).

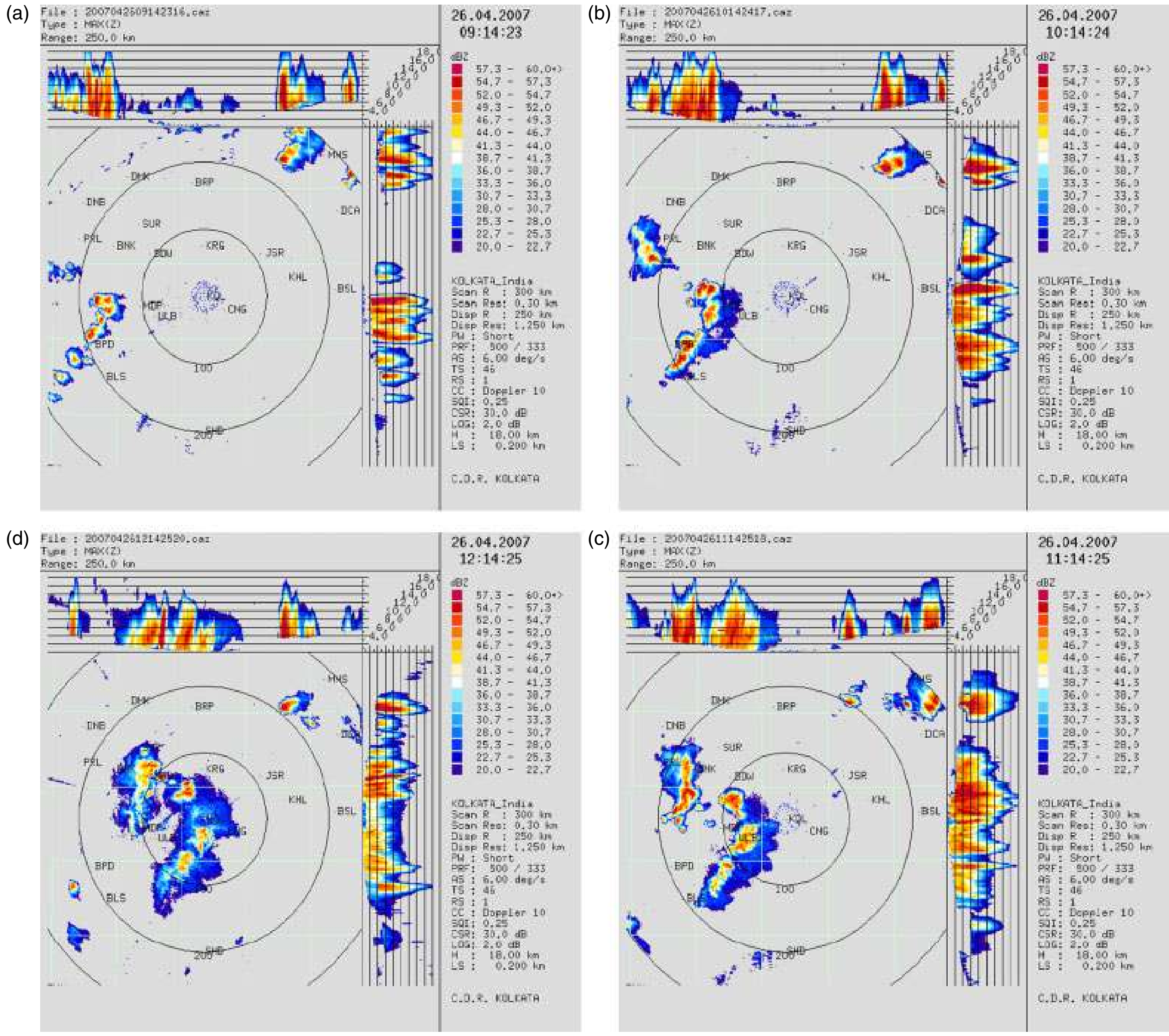

Figure 13. Doppler images (a, b, c and d) of sequential advancement of squall line towards Kolkata on 26 April 2007. This figure is available in colour online at wileyonlinelibrary.com/journal/met

given ranges (Figures 3–7). The result of the graph theoretic approach is validated with the Indian Meteorological Department (IMD) observation of 2007, 2008 and 2009 (Table III). The graph spectral distance and entropy estimation method fails to forecast the event on 12 May 2007. The event was overlooked by the graph spectral distance and entropy estimation method. IMD reports an ordinary thunderstorm on 12 May 2007 with wind speed 52 km h^{-1} and the present study method inferred that there will be no thunderstorm. The thunderstorm days from validation years are taken to compute different skill scores such as True Skill Statistic (TSS), Heidke Skill Score (HSS) and Yule's Q for the present forecast model. Table IV represents the contingency table for the skill score analysis. The TSS, HSS and Yule's Q are found to be 70, 70 and 94% respectively. However, the present graph theoretic approach provides 100% accurate forecast of severe thunderstorms, which is the main endeavour of the present study.

5. Applicability and limitations

The prevelence of thunderstorm over Kolkata can be sensed well before its occurrence with this new approach using graph spectral distance and entropy analysis. The distance analysis of important levels in the atmosphere seems to have more significance than CAPE/CIN analysis (Figure 2). If all the vertex distances and the intermediate distances are categorized well before the thunderstorm events and reference graphs (G_{TS} and G_{NTS}) are constructed, then this approach can be applicable for any geographical location. Figure 14 represents variation of forecast accuracy with lead time. More than 95% accurate forecast is possible with a 6 h lead time. The occurrence of thunderstorms can be predicted with the accuracy of 98% when the lead time is 1 h. Further, with lead time 6 h the accuracy is observed to be 93% through the proposed method of entropy estimation and graph spectral distance computation. If the entropy estimation of the vertex distances is about 1.09 ($\pm$0.11) and the spectral

Copyright © 2011 Royal Meteorological Society

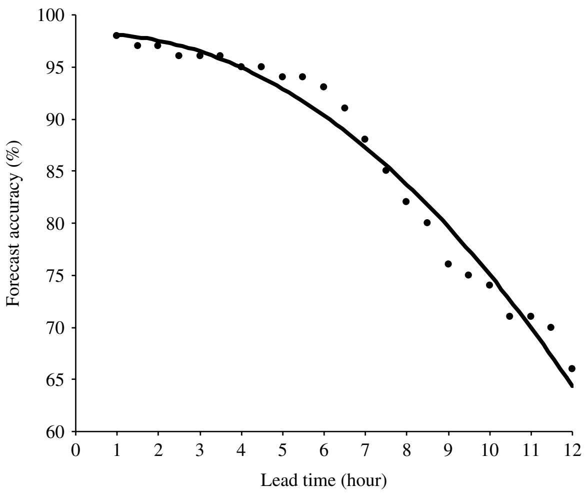

Figure 14. The variation of forecast accuracy with lead time.

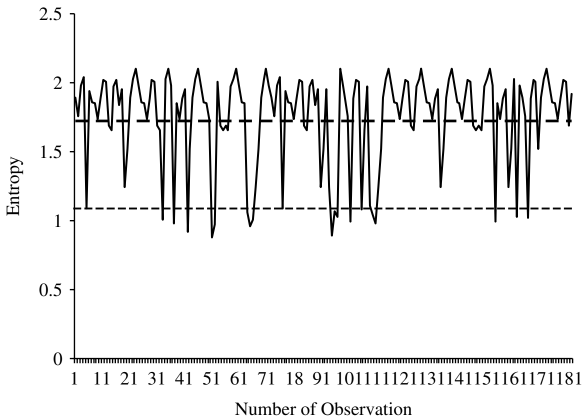

Figure 15. The observed and predicted values of entropy for thunderstorm (TS) and non-thunderstorm (NTS) days during April–May for 2007, 2008 and 2009. ——: observed; - - - ·: predicted entropy of TS; ▬ ▬: predicted entropy of NTS.

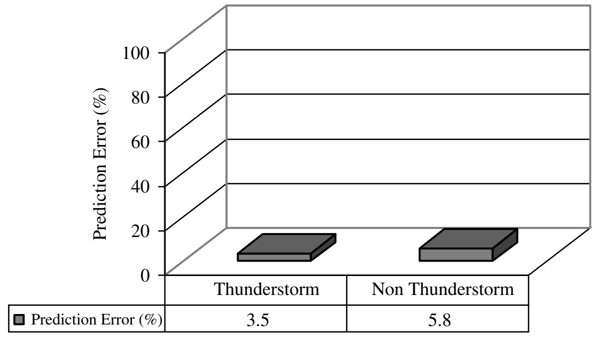

Figure 16. The prediction errors for thunderstorm and non-thunderstorm day forecast.

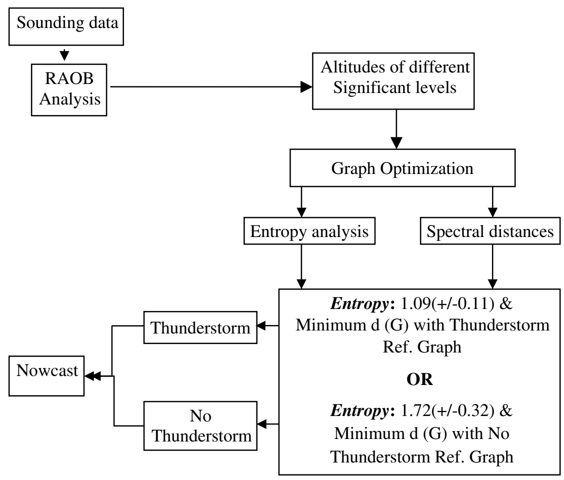

Figure 17. Implementation procedure to forecast thunderstorms over Kolkata with graph spectral distance entropy analysis.

distance from the reference graph of thunderstorm day (G_{TS}) is minimum on a particular day, then a warning for the occurrence of a thunderstorm can be issued. However, as the lead time increases, forecast accuracy decreases. There are several studies on the forecast of pre-monsoon thunderstorm that have been carried out over India. Ghosh *et al*. (2004) used Linear Discriminant Analysis (LDA) thechnique to classify thunderstorm and non-thunderstorm days over Kolkata (formerly Calcutta), India. However, no real time forecast of thunderstorms has been provided in the study. The discriminant analysis led to classify few selected, morning and afternoon thunderstorms with 53.33 and 42.20% accuracy and 20 variables. In the present study, more than 90% accurate forecasts are observed for thunderstorms with lead time 12 to 6 h. Ravi *et al*. (2006) applied a multiple regression method to give a potential forecast in probabilistic terms using nine predictors for pre-monsoon thunderstorms over Delhi, India. This was a potential method for operational use. However, the entropy estimation and spectral distance method have already been validated with the observation of India Meteorological Department for consecutive three years 2007, 2008 and 2009 (Figure 15). The prediction error is computed as:

$$PE(\%) = \frac{\langle |y_{dp} - y_{da}| \rangle}{\langle y_{da} \rangle} \times 100 \qquad (14)$$

where, $\langle\rangle$ denotes the average of test cases. The predicted and actual values of the parameters are denoted by y_{dp} and y_{da} respectively. The results show that prediction errors for thunderstorm days and non-thunderstorm days are 3.5 and 5.8% respectively (Figure 16).

6. Conclusion

The present study using the statistical probability approach on graph vertex distances leads to the conclusion that the empirical range of vertex distances (surface to LCL, surface to CCL, surface to LFC, surface to FL and surface to LNB) obtained from 10 years (1997–2006) data during the pre-monsoon season (April-May) are acceptable. The method of distance entropy estimation and spectral distance analysis seems to be very useful for nowcasting thunderstorms of the pre-monsoon season over Kolkata (Figure 17).

Copyright © 2011 Royal Meteorological Society

Acknowledgements

The first author acknowledges the financial assistance provided by the Department of Science & Technology, Govt. of India for conducting the research and India Meteorological Department for providing the data and the record of thunderstorms. The authors would like to express special gratitude to the anonymous reviewers for their substantial remarks that increased the completeness and clarity of the manuscript. The contribution of the Editor, Dr Peter Burt, during final revision of the paper is greatly appreciated.

Appendix A.

In the present study, different significant levels in the atmosphere like surface (v_1), LCL (v_2), CCL (v_3), LFC (v_4), FL (v_5), and LNB (v_6) are taken as set of vertices and lines joining them are considered as edges. The distance values between the vertices are taken as weight of the edges. 24 April 2000 was a thunderstorm day and the arrangement selected vertices are shown in the diagram below (Figure A1) and its corresponding adjacency matrix (Equation (3)) is shown (Table A1). The adjacency matrix represents actual distance values between vertices.

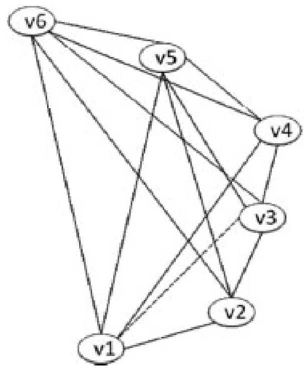

Figure A1.

Table A1.

	Surface	LCL	CCL	LFC	FL	LNB
Surface	0	360	215	545	4414	13 266
LCL	360	0	145	185	4054	12 906
CCL	215	145	0	330	4199	13 051
LFC	545	185	330	0	3869	12 721
FL	4414	4054	4199	3869	0	8852
LNB	13 266	12 906	13 051	12 721	8852	0

The normalized adjacency matrix can be obtained using Equation (8) and is shown below in Table A2.

Table A2.

	Surface	LCL	CCL	LFC	FL	LNB
Surface	0.0000	0.0240	0.0143	0.0363	0.2943	0.8844
LCL	0.0240	0.0000	0.0097	0.0123	0.2703	0.8604
CCL	0.0143	0.0097	0.0000	0.0220	0.2799	0.8701
LFC	0.0363	0.0123	0.0220	0.0000	0.2579	0.8481
FL	0.2943	0.2703	0.2799	0.2579	0.0000	0.5901
LNB	0.8844	0.8604	0.8701	0.8481	0.5901	0.0000

Using the above normalized distance values between graph vertices, distance entropy is calculated (Equation (9)).

References

Abraham A, Philip NS, Joseph B. 2001. Will we have a wet summer? Long term rain forecasting using Soft Computing models. In *Modeling and Simulation 2001, Publication of the Society for Computer Simulation International*, Kerchoffs EJH, Snorek M (eds). Prague: Czech Republic; 1044–1048.

Anderson JL. 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate* **9**: 1518–1530.

Baldi M, Dalu GA, Pielke RA. 2008. Vertical Velocities and Available Potential Energy Generated by Landscape Variability-Theory. *Journal of Applied Meteorology and Climatology* **47**(2): 397–410.

Benjamin R, Paul K, George Y, Steven G, Richard G, Jeremy R. 2007. A fingerprinting technique for major weather events. *Journal of Applied Meteorology and Climatology* **46**(7): 1053–1066.

Biggs N. 1993. *Algebraic Graph Theory*. Cambridge Mathematical Library.

Byers HR, Braham RR Jr. 1949. *The Thunderstorm*. U.S. Government Printing Office; 287.

Chaudhuri S. 2006. Predictability of chaos inherent in the occurrence of severe thunderstorms. *Advances in Complex Systems* **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. *Advances in Complex Systems* **10**: 413–422.

Chaudhuri S. 2008a. Identification of the level of downdraft formation during severe thunderstorms: a frequency domain analysis. *Meteorology and Atmospheric Physics* **102**: 123–129.

Chaudhuri S. 2008b. Preferred type of cloud in the genesis of severe thunderstorms – a soft computing approach. *Atmospheric Research* **88**: 149–156.

Chaudhuri S. 2010. Convective energies in forecasting severe thunderstorms with one hidden layer neural net and variable learning rate back propagation algorithm. *Asia – Pacific Journal of Atmospheric Sciences* **46**(2): 173–183.

Chaudhuri S, Middey A. 2009. Applicability of bipartite graph model for thunderstorms forecast over Kolkata. *Advances in Meteorology* **2009**: 1–12.

Charba JP. 1979. Two to six hour severe local storm probabilities, an operational forecasting system. *Monthly Weather Review* **107**: 268–282.

Chung F. 1992. *Spectral Graph Theory*. American Mathematical Society.

Comric AC. 1997. Comparing neural networks and regression models for ozone forecasting. *Journal of Air and Waste Management Association* **47**: 653–663.

El-Ghoul M. 2002. The most general set and chaos graph, chaos. *Solitons and Fractals, UK* **18**: 833–838.

Ghosh S, Sen PK, De UK. 2004. Classification of thunderstorm and non-thunderstorm days in Calcutta (India) on the basis of linear discriminant analysis. *Atmósfera* **17**: 1–12.

Jacovides CP, Yonetani T. 1990. An evaluation of Stability indices for thunderstorm prediction in Greater Cyprus. *Weather and Forecasting* **5**(4): 559–569.

Johns RH, Doswell CA. 1992. Severe local storms forecasting. *Weather and Forecasting* **7**: 588–612.

Copyright © 2011 Royal Meteorological Society

Litta AJ, Mohanty UC. 2008. Simulation of a severe thunderstorm event during the field experiment of STORM programme 2006, using WRF-NMM model. *Current Science* **95**(2): 204–215.

Manohar GK, Kandalgaokar SS, Tinmaker MIR. 1999. Thunderstorm activity over India and Indian south-west monsoon. *Journal of Geophysical Research* **104**: 4169.

Marzban C, Witt A. 2001. A Bayesian neural network for severe-hail size prediction. *Weather and Forecasting* **16**(5): 600–610.

Mukhopadhyay P, Singh HAK, Mahakur M. 2009. The interaction of large scale and mesoscale environment leading to formation of intense thunderstorms over Kolkata. Part I: doppler radar and satellite observations. *Journal of Earth System Science* **118**(5): 441–466.

Mukhopadhyay P, Singh HAK, Singh SS. 2005. Two severe nor'westers in April 2003 over Kolkata, India, using Doppler radar observations and satellite imagery. *Weather* **60**: 343–353.

Neumann CJ. 1971. The thunderstorm forecasting system at the Kennedy Space Center. *Journal of Applied Meteorology* **10**: 921–936.

Ravi N, Mohanty UC, Madan OP, Paliwal RK. 2006. Forecasting of thunderstorms in the pre-monsoon season at Delhi. *Meteorological Applications* **6**: 29–38.

Wilks DS. 1995. *Statistical Methods in Atmospheric Sciences*. Academic Press: USA.

Copyright © 2011 Royal Meteorological Society