

# Development of a perfect prognosis probabilistic model for prediction of lightning over south-east India

M RAJEEVAN<sup>1,\*</sup>, A MADHULATHA<sup>1</sup>, M RAJASEKHAR<sup>2</sup>, JYOTI BHATE<sup>1</sup>,  
AMIT KESARKAR<sup>1</sup> and B V APPA RAO<sup>2</sup>

<sup>1</sup>National Atmospheric Research Laboratory, Department of Space, Gadanki 517 112, India.

<sup>2</sup>Satish Dhawan Space Centre, SHAR, ISRO, Sriharikota, India.

\*Corresponding author. e-mail: rajeevan@narl.gov.in rajeevan61@yahoo.co.in

A prediction model based on the perfect prognosis method was developed to predict the probability of lightning and probable time of its occurrence over the south-east Indian region. In the perfect prognosis method, statistical relationships are established using past observed data. For real time applications, the predictors are derived from a numerical weather prediction model. In the present study, we have developed the statistical model based on Binary Logistic Regression technique. For developing the statistical model, 115 cases of lightning that occurred over the south-east Indian region during the period 2006–2009 were considered. The probability of lightning (yes or no) occurring during the 12-hour period 0900–2100 UTC over the region was considered as the predictand. The thermodynamic and dynamic variables derived from the NCEP Final Analysis were used as the predictors. A three-stage strategy based on Spearman Rank Correlation, Cumulative Probability Distribution and Principal Component Analysis was used to objectively select the model predictors from a pool of 61 potential predictors considered for the analysis. The final list of six predictors used in the model consists of the parameters representing atmospheric instability, total moisture content in the atmosphere, low level moisture convergence and lower tropospheric temperature advection. For the independent verifications, the probabilistic model was tested for 92 days during the months of May, June and August 2010. The six predictors were derived from the 24-h predictions using a high resolution Weather Research and Forecasting model initialized with 00 UTC conditions. During the independent period, the probabilistic model showed a probability of detection of 77% with a false alarm rate of 35%. The Brier Skill Score during the independent period was 0.233, suggesting that the prediction scheme is skillful in predicting the lightning probability over the south-east region with a reasonable accuracy.

---

## 1. Introduction

Severe thunderstorms are capable of producing baseball-sized hail, strong winds, intense rain, flash floods, and tornadoes. Thunderstorms produce heavy rain. They also pose serious hazards to aviation and satellite launch activities. Cloud to Ground (CG) lightning is one of the leading

causes of weather-related fatalities in India (De *et al* 2005). Improved forecasts of CG lightning would have many potential societal benefits. Skillful probabilistic guidance 3–24 hours in advance would allow the public to better assess the CG lightning threat and thereby support better decision making with regard to the protection of life and property.

**Keywords.** Thunderstorm; lightning; mesoscale convective systems; probabilistic prediction.

Forecasting a thunderstorm is one of the most difficult tasks in weather prediction, due to their rather small spatial and temporal extension and the inherent nonlinearity of their dynamics and physics. Very short period forecasting of the future location of convective storms has historically been based primarily on the extrapolation of radar reflectivity echoes (Wilson *et al* 1998). The majority of individual thunderstorms have lifetimes less than 20 minutes, therefore, forecast techniques based on the extrapolation of existing conditions are limited. For forecast periods beyond 20 minutes, techniques for forecasting the initiation, growth and dissipation of convective storms are essential (Wilson *et al* 1998). Two methods used for forecasting storm evolution are: knowledge based expert systems including statistical models and explicit numerical forecast models that are initialized with radar data. However, with the numerical weather prediction models, the main problems are associated with the fact that storm data are not included in the initial conditions. Low-level convergence in these simulations takes some time to spin up from the large-scale circulation and hence the models are not generally reliable for the first 6 hours or so. When storms do develop, there are often significant errors in timing and location that cannot be corrected unless new storm data are incorporated. However, numerical models are successful in cases where the large scale synoptic forcing is well marked. The improvement in thunderstorm prediction is also highly handicapped due to lack of meso-scale observations and insufficient understanding. The development of a lightning forecast procedure is a more challenging problem. Lightning is governed by cloud microphysical processes that are poorly resolved by numerical models.

During the recent years, a variety of statistical techniques have been used to develop forecast models for thunderstorms and lightning. Some of the statistical models that have been used include multiple linear regression, binary logistic regression, and classification and regression trees (CART) (Lambert *et al* 2005). For continuous predictands, the most common method is multiple linear regression (MLR). However, when the predictand is 'yes' or 'no', binary logistic regression (BLR) is often employed (Mazany *et al* 2002; Lambert *et al* 2005; Shafer and Fuelberg 2006). These methods attempt to quantify the relationship between a set of predictors and thunderstorm probability or lightning frequency (Reap 1994a). Shafer and Fuelberg (2006) developed a statistical scheme to forecast warm season lightning over portions of the Florida peninsula. For this purpose, they have used 16 years of CG lightning data from the National Lightning Detection Network

(NLDN) and morning radiosonde derived parameters. Logistic regression techniques are used to develop equations predicting lightning activity. Shafer and Fuelberg (2008) used a perfect prognosis scheme for forecasting warm-season lightning over Florida. Analysis data from the Rapid Update Cycle (RUC) and lightning data have been used to develop a high-resolution, gridded forecast guidance product for cloud-ground lightning over Florida. A forecast example using the high resolution WRF model reveals that exact timing and placement of forecast lightning are not perfect, there is generally a good agreement between the forecasts and their verification. Statistical prediction models for lightning over Canada and the northern United States also have been developed using CART (Burrows *et al* 2005).

Dasgupta and De (2007) developed binary logistic regression models for short term prediction of pre-monsoon convective developments over Kolkata (India). Ghosh *et al* (1999) examined the significant meteorological parameters for predicting thunderstorms at Kolkata using statistical methods. Chatterjee *et al* (2008) used the multivariate technique for predicting pre-monsoon thunderstorms. Bhowmik Roy *et al* (2007) have examined the thermodynamics of the atmosphere in relation to occurrence of convective rainfall over the Indian region using various thermodynamical and kinematic parameters. Their results showed that presence of strong thermodynamic environment is not sufficient for the occurrence of deep convection. Other factors like dynamical conditions also play a very important role in controlling the occurrence of deep convection. Numerical simulations of thunderstorm cases observed over India were made by Chatterjee *et al* (2008), Litta and Mohanty (2008), Srivastava *et al* (2008), Mukhopadhyay *et al* (2009) and Rajeevan *et al* (2010).

There are few studies on the climatological features of lightning activity over India. Manohar *et al* (1999), Tinmaker *et al* (2009) and Nath *et al* (2009) have studied the climatological features of lightning over India. Ranalkar and Chaudari (2009) examined the seasonal variation of lightning activity over the Indian subcontinent using TRMM LIS data. Figure 1 shows the monthly climatology of lightning flashes (flashes  $\text{km}^{-2} \text{yr}^{-1}$ ) over India, derived from TRMM LIS data. Over south peninsula, lightning activity peaks in April and May. Once the summer monsoon arrives in June, lightning activity drastically reduces over India, except along the foothills of Himalayas and north-east India. Over south-east India, moderate lightning activity however continues till October. The Satish Dhawan Space Centre, SHAR, Sriharikota (13.7°N, 80.2°E), which is about 70 km north of Chennai (shown in figure 1) is the responsible

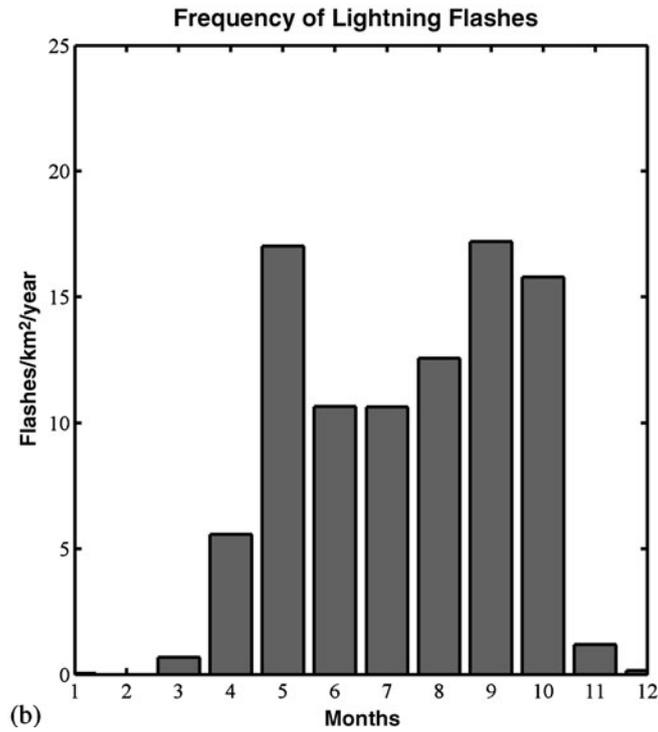
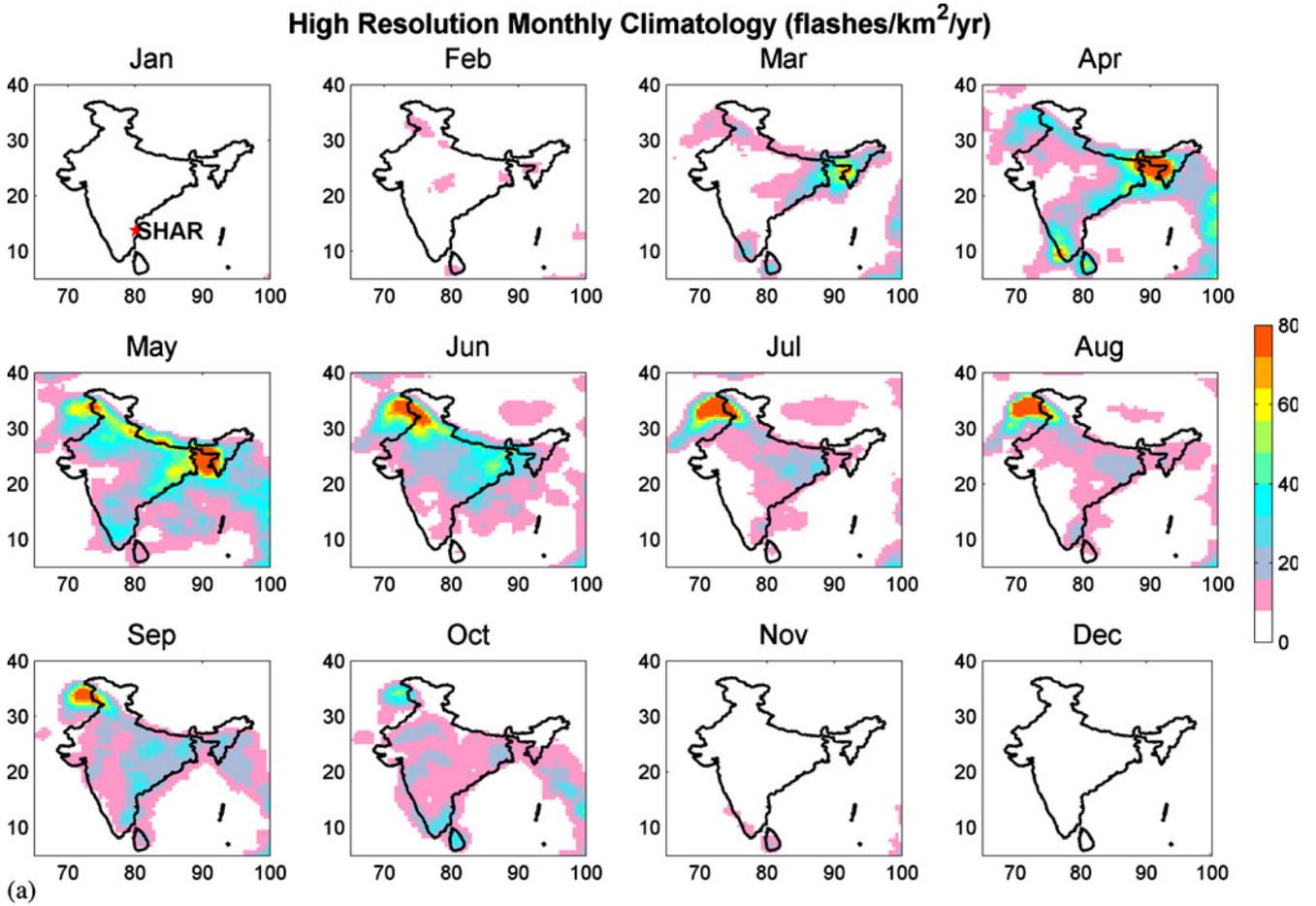


Figure 1. (a) Monthly climatology of lightning flashes over India derived from TRMM LIS data (1998–2005) in flashes/km<sup>2</sup>/year and (b) the same for the SHAR region.

organization of Indian Space Research Organization (ISRO) for satellite launch operations. On an average, 4–5 launches are scheduled in a year. The SHAR Space Centre requires accurate weather forecasts and warning on a wider time scale for their satellite launch missions. Thunderstorm/lightning is a major risk factor for the rocket launch operations, as they may affect communication links, fuelling operations and even complex electronic circuits installed in the launch vehicle. For careful planning, the mission team requires prediction of thunderstorm probability over the SHAR region at least 1–24 hours in advance. They need forecasts of probability of occurrence and also probable time of its occurrence.

In this study, we report the results of our efforts to develop a probabilistic model for predicting lightning occurrence over the SHAR Space Centre. For this purpose, we have used the perfect prognosis approach in numerical weather prediction (Shafer and Fuelberg 2008). The basic premise of this study is that a numerical weather prediction model such as the Weather Research and Forecast (WRF) model is capable of predicting large scale synoptic features associated with lightning activity, at least 9–18 hours in advance. In this approach, we have established reliable statistical relationships of lightning occurrence (probability) with large scale thermodynamic and dynamic parameters derived from the observed data. Since the predictand is an event (or probability), we have used the Binary Logistic Regression (BLR) method to develop regression equation. The regression equation thus developed was tested for independent days of lightning and non-lightning using the parameters derived from the WRF predictions as the predictors. Lightning and non-lightning cases have been derived from the time records of the electric Field Mill (an instrument measuring electrical discharge from cloud to ground) data available at the SHAR Space Centre. In this study, we have considered only the cases of severe thunderstorms with lightning activity in which the CG electric discharge is more than 4000 V/m.

Data used in this study are described in section 2, methodology is discussed in section 3. The results are given in section 4 and the conclusions are drawn in section 5.

## 2. Data

For identifying lightning cases, we have used the Field-Mill data available at the SHAR Space Centre. Field-Mill instruments are installed at five places within 30 km of the Centre. The Field-Mill instrument records all CG lightning discharges within 30 km radius. We have considered the

Field-Mill data of the SHAR Space Centre for the period 2006–2009 to find out the lightning and non-lightning cases. Using the Field-Mill data, we have identified 115 cases of thunderstorm/lightning cases during the period, 2006–2009. For identifying the cases, we have considered only the months March–October with reasonable lightning activity. The 115 lightning cases identified are the cases in which the CG electric discharge was more than 4000 V/m. To develop the probabilistic model, we require non-lightning cases also. Therefore, we have identified additional 190 cases of non-lightning days. We tried to keep the lightning–non-lightning ratio as close to the observed climatology of lightning over the region. At the same time, for obtaining a reasonable estimate of the regression coefficients, we have kept the ratio as 115:190 cases of lightning/non-lightning. The non-lightning cases were identified during the same months carefully using the Field-Mill data. Thus, we have considered 305 cases of lightning and non-lightning cases for training the probabilistic model.

For identifying thermodynamic predictors for developing probabilistic models, ideally radiosonde profiles of the SHAR Space Centre can be used. Unfortunately, there are no regular radiosonde measurements at the SHAR Space Centre as they are made only during the satellite launch campaign periods. Therefore, we have considered the NCEP Final Analysis data (FNL) for deriving the temperature, humidity and wind profiles for identifying potential predictors. We are aware that the NCEP FNL analysis is not really observed data. From the NCEP FNL data, many parameters were investigated for possible inclusion as the predictors for the development of the model. In total, 51 thermodynamic variables and 10 dynamical variables (total 61 parameters) were considered as the predictors. These predictors are known thermodynamic and dynamic parameters which have bearing on the development of thunderstorms (Shafer and Fuelberg 2008). The initial list of the predictors considered for the study is given in table 1. The parameters for the 305 cases were calculated from the FNL temperature, dew point, wind, height and surface pressure fields from the nearest analysis (mostly 1200 and 1800 UTC analyses). We have taken care to use the analysis data just before the storm occurred so that the influence of storm development is not seen in the predictor dataset. An important assumption made in this study is that the model analyses provide the best estimate of the state of the atmosphere at the analysis time. Therefore, they can be treated as ‘observations’ for purposes of developing the PP equations (Shafer and Fuelberg 2008). While selecting the potential predictors, we focused on those parameters that are produced reasonably

Table 1. List of thermodynamic and dynamic parameters considered for the study.

Sl. no	Abbreviation	Name	Description/levels
1	LCL	Pressure at LCL	Lifting condensation level
2	LFC	Pressure at LFC	Level of free convection
3	LCL-LFC	Pressure at LCL-LFC	Difference between lifting condensation level and level of free convection
4	EL	Pressure at EL	Equilibrium level
5	CINE	Convective inhibition energy	Negative area between the surface and 700 hPa by lifting the surface parcel
6	MUCAPE	Most unstable CAPE	Largest CAPE obtained when each parcel between the surface and 700 hPa is lifted
7	LCAPE1	MUCAPE in various layers	Cloud base to cloud top
8	LCAPE2		Cloud base to $-20^{\circ}\text{C}$ (Bothwell 2002)
9	LCAPE3		Mixed phase region: $0^{\circ}$ to $-40^{\circ}\text{C}$
10	LCAPE4		Charging zone: $-10^{\circ}$ to $-25^{\circ}\text{C}$
11	LCAPE5		Between $-15^{\circ}$ and $-20^{\circ}\text{C}$ (Bothwell 2002)
12	NCAPE1		Layer CAPE divided by the geometric thickness of the layer
13	NCAPE2		
14	NCAPE3		
15	NCAPE4		
16	NCAPE5		
17	CCTHGT	Convective cloud-top height	Geometric height of equilibrium level
18	CCTHICK	Cold cloud thickness	Thickness between $0^{\circ}\text{C}$ level and cloud top (equilibrium level)
19	PRFREQ	Price and Rind frequency	Price and Rind function (Appendix 1)
20	CPTP	Cloud physics thunder parameter	Appendix 1
21	LI	Lifted index	Appendix 1
22	KI	K index	Appendix 1
23	TT	Total–Total index	Appendix 1
24	SWEAT	Severe weather threat index	Appendix 1
25	WINDEX	Wind index	McCann (1994)
26	LTHICK1	Layer thickness at different levels	1000–850 hPa
27	LTHICK2		850–500 hPa
28	LTHICK3		700–400 hPa
29	LTHICK4		500–300 hPa
30	MEANRH		1000–850 hPa
31	RHFRZL	Mean relative humidity	
32	LAYRH	Relative humidity at $0^{\circ}\text{C}$ level	
33	LAYRH	Layer mean relative humidity	21 levels between 1000 and 100 hPa
34	MEANU	Layer averaged $u$ component	21 levels between 1000 and 100 hPa
35	MEANV	Layer averaged $v$ component	21 levels between 1000 and 100 hPa
36	MEANSP	Layer average speed	21 levels between 1000 and 100 hPa
37	SHEAR1	Mean wind shear	21 levels between 1000 and 100 hPa
38	SHEAR2	Wind shear	1000–100 hPa
39	SHEAR3	Wind shear	925–500 hPa
40	THETAES-THETA1	Difference of saturated equivalent potential temp (EPT) and EPT at different levels	1000 hPa
41	THETAES-THETA2		850 hPa
42	THETAES-THETA3		500 hPa
43	TLAPSE1	Temperature lapse rate	1000–700 hPa
44	TLAPSE2		700–400 hPa
45	TLAPSE3		400–100 hPa
46	THELAPSE1	Theta lapse rate	1000–700 hPa
47	THELAPSE2		700–400 hPa
48	THELAPSE3		400–100 hPa

Table 1. (Continued).

Sl. no	Abbreviation	Name	Description/levels
48	CCL	Cloud condensation level	
49	TCONV	Convective temperature	
50	PWC	Precipitable water content	
51	WBZP	Wet bulb zero pressure	
52	DIV	Wind divergence	21 levels between 1000 and 100 hPa
53	VOR	Vorticity	21 levels between 1000 and 100 hPa
54	MFC	Moisture flux convergence	21 levels between 1000 and 100 hPa
55	TEMPA	Temperature advection	21 levels between 1000 and 100 hPa
56	THETAEA	Theta-e advection	21 levels between 1000 and 100 hPa
57	VORTA	Vorticity advection	21 levels between 1000 and 100 hPa
58	MVORT	Mean vorticity	Mean between 850 and 800 hPa levels
59	MMFC	Mean moisture flux convergence	Mean between 950 and 850 hPa levels
60	MTEMPA	Mean temperature advection	Mean between 950 and 925 hPa levels
61	MTHETAEA	Mean Theta-e advection	Mean between 925 and 900 hPa levels

by the present NWP models. This aspect will be discussed later in this paper.

### 3. Methodology

#### 3.1 Statistical model

Past climatological data reveals that at the SHAR Space Centre, lightning activity is more pronounced during late evening or night (0900–2100 UTC). Many of the previous statistical studies utilized parameters derived from morning soundings to forecast lightning occurrence during the afternoon. However, this approach sometimes can produce large forecast errors if morning conditions change, or if the sounding is not representative of the entire forecast area (Shafer and Fuelberg 2008). An alternate to soundings is data from Numerical Weather Prediction (NWP) models. Since NWP models provide input data that are more location- and time-specific than soundings, they may produce more skillful forecasts. Using the NWP model forecasts, prediction models for lightning probability can be derived in two ways, Model Output Statistics (MOS) and Perfect Prognosis Method (PPM).

MOS is an objective forecasting technique in which a statistical relationship is determined between a predictand and variable forecasts by an NWP model. The primary advantage of MOS is that model biases and local climatology are automatically built into the equations. Reap (1994b) developed MOS equations predicting the spatial distribution of CG lightning over Florida during different low-level regimes using predictors from the Nested Grid Model (NGM). MOS has several drawbacks that can limit its forecast skill. Since

NWP models are constantly changing, it is often difficult to obtain a long archive of forecasts from the same model that will be used to develop the MOS equations. Any modifications to the NWP model that change systematic model errors require redevelopment of the MOS equations (Wilks 2006).

An alternative to MOS is the Perfect Prognosis (PP) method. This approach develops statistical relationships between observed atmospheric parameters and observations of the predictand (Klein 1971). Once the statistical relations are determined, forecasts of the predictand are obtained by inserting NWP model forecasts of the predictors into the PP equation (Wilks 2006). The perfect prognosis forecast system to predict probabilistic CG lightning (Bothwell 2002) was first implemented at the Storm Prediction Center (SPC) in 2003. Bothwell (2002, 2005, 2008) and Bothwell and Buckley (2009) used the PP method to develop lightning guidance for the western United States on a  $40 \times 40$  km grid using analyses from the NCEP 40-km Rapid Update Cycle (RUC 40). Later, they have developed similar schemes for Alaska also. When models are upgraded to newer version, no changes to the predictive equations are necessary using the PP method. A drawback to the PP method is that it assumes a ‘perfect’ forecast of the predictors by the NWP model and thus does not account for model biases. Conversely, one of the significant advantages is the stability of the equations. Since PP equations are developed without NWP information, any changes to the driving NWP models do not require redevelopment of the PP equations. In fact, improving random or systematic errors in the NWP model should improve the statistical forecasts (Wilks 2006).

Multi-linear Regression (MLR) has been used in the majority of previous statistical lightning

studies (Neumann and Nicholson 1972; Reap and MacGorman 1989; Reap 1994b; Hughes 2001). However, unless the assumptions of constant variance and Gaussian residuals are met (which is rarely the case with count data), these methods can lead to undesirable results. Thus, we considered an alternative regression method, called the Binary Logistic Regression (BLR).

Logistic regressions are fit to binary predictands according to the non-linear equation

$$\ln\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (1)$$

or

$$p_i = \frac{\exp(b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k)}{1 + \exp(b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k)} \quad (2)$$

where  $p_i$  is the predicted probability resulting from the  $i$ th set of predictors ( $x_1, x_2, \dots, x_k$ ). The quantity on the left (1) is the logit link function, which relates the log of the odds ratio ( $p/1-p$ ) to a linear combination of predictors. In BLR, the regression parameters ( $b_0, b_1, \dots, b_k$ ) are estimated by maximizing a log likelihood function using iterative methods (Wilks 2006). Unlike MLR, equation (2) guarantees that the probabilities are bounded within the interval (0,1). BLR does not assume a direct linear relationship between the predictors and the response and accommodates the non-Gaussian distributions of the regression residuals.

### 3.2 Selection of predictors

The favourable conditions for triggering thunderstorm activity are atmospheric instability, adequate moisture content, especially in the lower and middle troposphere and a physical mechanism to lift the air parcel from the surface. Figure 2(a) shows the composite vertical profile of temperature, humidity and winds for 115 lightning cases and 190 non-lightning cases derived from NCEP FNL data. Figure 2(b) shows the difference of temperature, relative humidity and zonal and meridional winds between lightning and non-lightning cases. The profiles of lightning cases compared to non-lightning cases show a slightly cooler lower troposphere. Similarly, lightning cases show more moisture content, especially in the mid-troposphere. The difference in RH in the middle troposphere is of the order of 20%. Wind flow profiles show the presence of stronger westerlies in the lower and middle troposphere, and stronger easterlies in the upper troposphere. This suggests that vertical wind shear in the lightning cases is larger compared to non-lightning cases.

The potential predictors considered for the study are shown in table 1. These are known predictors for lightning prediction considered in previous studies (e.g., Shafer and Fuelberg 2008). However, the number of potential predictors is too large to include in the model with 305 cases. Therefore, we have short-listed the predictors using two different criteria. The first one is based on Spearman Rank Correlation, which is similar to Pearson correlation coefficient, but for ranked variables. The statistical significance of the correlation can be tested using a Student's  $t$  test. The  $t$  value is calculated as:

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

which is distributed approximately as Student's  $t$  distribution with  $n-2$  degrees of freedom under the null hypothesis.  $r$  is the correlation coefficient. The Spearman rank correlations were calculated using the statistical package, STATISTICA.

First of all, we have short-listed the predictors with Spearman correlation more than 0.22, which is statistically significant at the 99% level or more. The second level screening is done by calculating cumulative frequency distribution of the predictors for lightning and non-lightning cases. To use the predictors in the probabilistic model, there should be significant differences in the cumulative frequency distribution between the lightning and non-lightning cases. We used a criterion that the difference in cumulative distribution should be more than 20% to shortlist the predictors. Figure 3 shows the cumulative frequency plots of two predictors, NCAPE2 and RHFRZL. There is significant difference in the cumulative frequency distribution of RHFRZL between the two cases. However, such difference is not appreciable for the variable NCAPE2. Therefore, NCAPE2 may not be a good predictor to distinguish between lightning and non-lightning cases. When these two criteria (Spearman rank correlation and cumulative probability distribution) were applied, the number of potential predictors has been reduced to 12 from the original list of 61. The list of 12 potential predictors thus identified is shown in table 2. The list of potential predictors is a mixture of parameters representing atmospheric instability, moisture content and low level moisture convergence and vorticity. Out of the 12 predictors short-listed, there are four dynamical variables representing low level vorticity, moisture convergence, temperature advection and  $\theta_e$  advection (MVORT, MMFC, MTEMPA and MTHETAEA). Banacos and Schultz (2005) discussed the historical and operational perspectives using moisture flux convergence in forecasting convective initiation. There are three thermodynamic indices (KI, TTI and SWEAT), generally

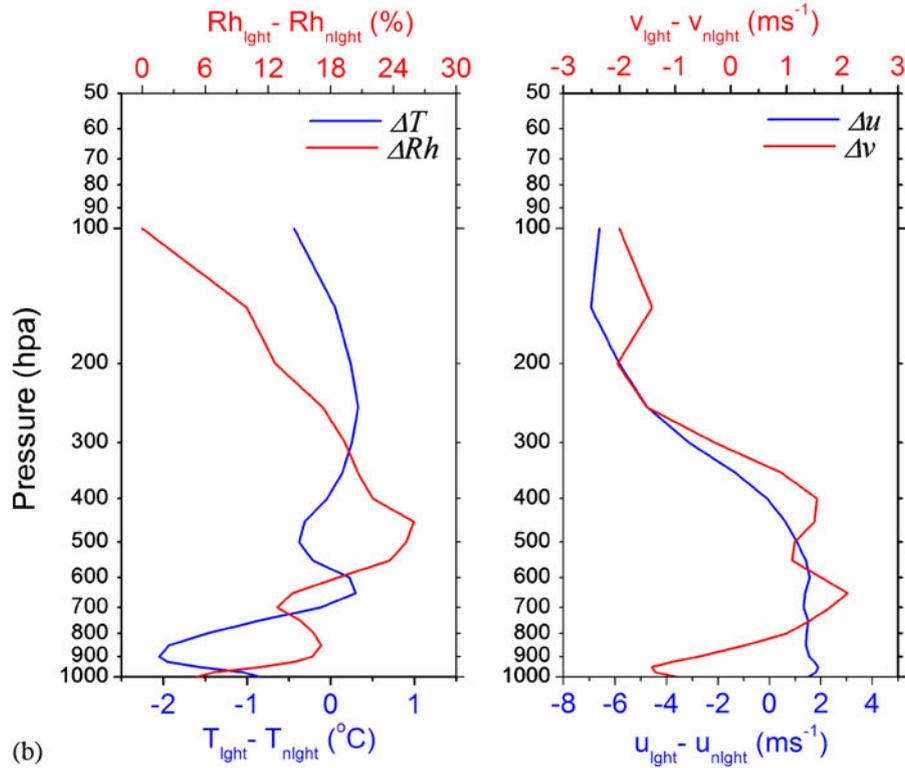
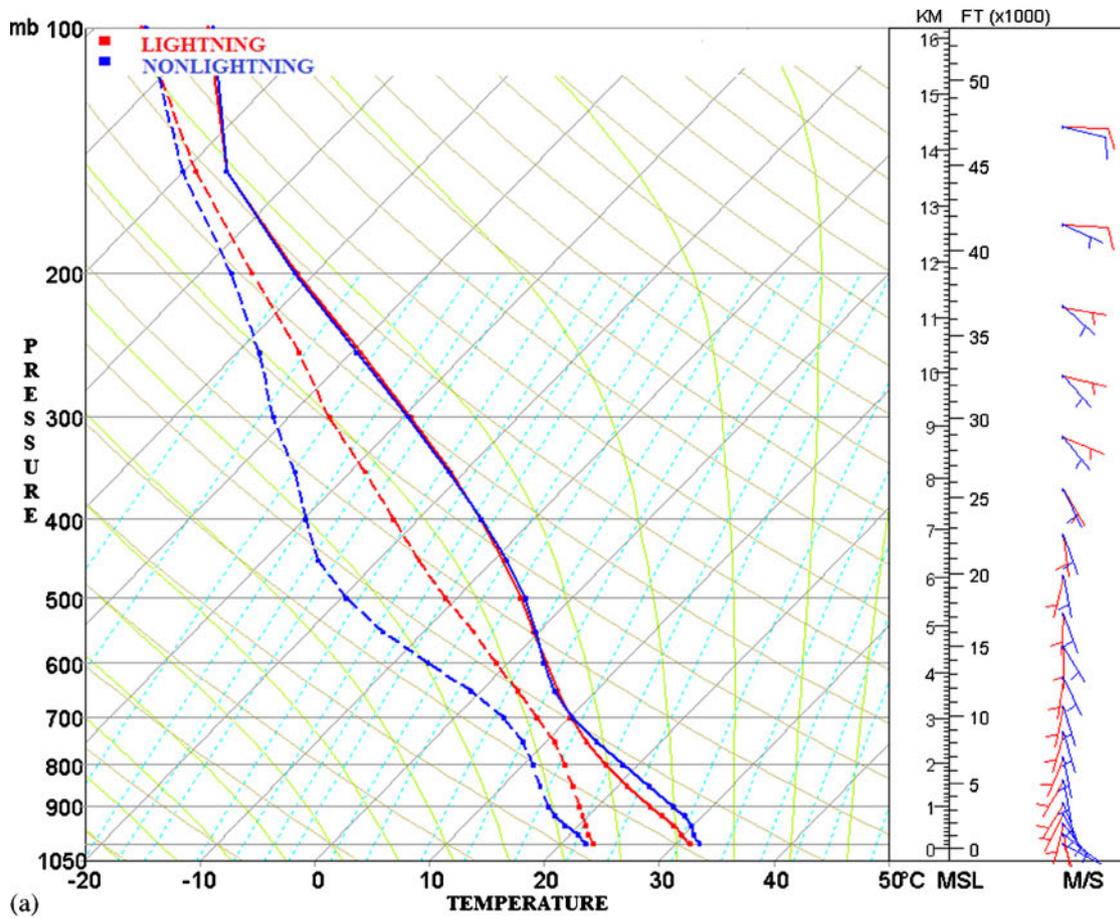


Figure 2. (a) Composite dry bulb temperature and dew point temperature profiles for lightning (red) and non-lightning cases (blue). (b) Difference between composite temperature and relative humidity profiles between lightning and non-lightning cases (left) and the same but for zonal and meridional winds (right).

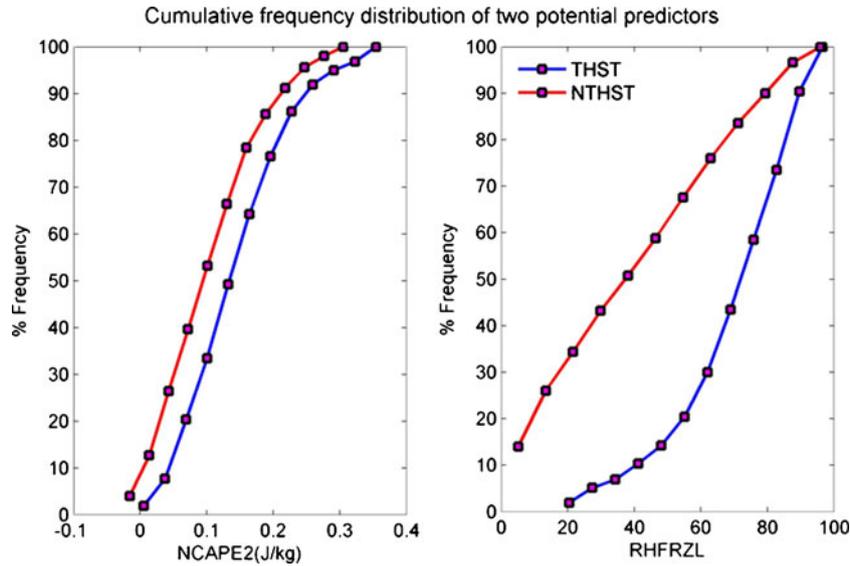


Figure 3. Cumulative frequency plots of the potential predictors, NCAPE2 and RHFRZL.

Table 2. List of 12 parameters short listed for developing probabilistic model.

Sl. no	Abbreviation	Name
1	KI	K index
2	TTI	Total–Total index
3	SWEAT	Severe weather threat index
4	WINDEX	Wind index (see Appendix 1)
5	RHFRZL	Relative humidity at 0°C level
6	MEANRH	Mean relative humidity (1000–850 hPa)
7	LAYRH	Layer mean relative humidity, 21 levels between 1000 and 300 hPa
8	PWC	Precipitable water content
9	MVORT	Mean vorticity (850–800 hPa)
10	MMFC	Mean moisture flux convergence (950–850 hPa)
11	MTEMPA	Mean temperature advection (950–925 hPa)
12	MTHETAEA	Mean Theta-e ( $\theta_e$ ) advection (925–900 hPa)

associated with thunderstorm activity. RHFRZL, MEANRH, LAYRH and PWC are the parameters indicating moisture content in the atmosphere. WINDEX (McCann 1994) is a parameter indicating moisture content and atmospheric instability.

Some of the predictors given in table 2 are inter-correlated and they may contain redundant information. For example, precipitable water content (PWC) has information from the moisture content at freezing level also (RHFRZL). Including predictors with strong mutual correlation in a prediction equation can lead to poor estimates of the regression parameters (Wilks 2006) and the model performance can be adversely affected. This problem was addressed by performing a principal

component analysis (PCA) to examine inter-correlations among the predictors and to aid in choosing a smaller subset to retain for the regression analysis. PCA is a mathematical procedure that transforms a number of correlated variables called principal components (PCs). In this study, the PCs were used as a classification method to cluster the highly correlated predictors into groups having some physical meaning. Only components with eigen values more than 1 were considered as suggested by Wilks (2006). The parameters having the greatest weights (or loadings) on each component were grouped together. The principal component groupings were used as an objective method to select a subset of the most physically relevant predictors containing less mutual

correlation. By performing the PCA, we could group the 12 predictors into four different groups. Further details are given below.

### 3.3 Forecast verification

The most commonly used measure of accuracy for probabilistic forecasts is the Brier score (Brier 1950) given by:

$$BS = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2$$

where  $N$  is the number of forecast-observation pairs,  $f_i$  is the forecast probability, and  $o_i$  is the observation (set to 1 if the event occurred or zero if the event did not occur). The Brier score essentially is the mean of the squared differences between the forecast probabilities and the binary (0 or 1) observations. Perfect forecasts exhibit  $BS = 0$ , while less accurate forecasts have  $0 < BS < 1$ . We also calculated the Brier skill score, given by:

$$BSS = 1 - BS_{\text{model}}/BS_{\text{ref}}$$

where the  $BS_{\text{model}}$  is the Brier score for the model and  $BS_{\text{ref}}$  is the Brier score for a reference forecast. Forecasts with a lower Brier score than the reference will have  $BSS > 0$  (or positive skill) while forecasts with higher Brier scores than the reference forecasts have  $BSS < 0$  (or negative skill).

Unfortunately, we do not have daily climatological occurrence for calculating the reference skill based on climatology. Therefore, we have used the reference skill based on persistence.

## 4. Results

### 4.1 BLR model performance

With the 12 potential predictors, we have tried different combination of predictors to develop the BLR model. By performing PCA of 12 predictors, we could identify four different groups. The first group contains KI and TTI. The second group contains the predictors, SWEAT and WINDEX. The third group contains MEANRH, RHFRZL, LAYRH and PWC and the fourth group contains the remaining dynamic parameters MVOR, MMFC, MTEMPA and MTHETAEA. We have tried different combinations using at least one parameter each from the four groupings.

Another important factor which we need to consider is the fidelity of the NWP model in predicting these parameters on real time. It is necessary to test the skill of NWP model to predict these parameters. We have considered the Weather Research and Forecast (WRF) model to derive these predictors for all 31 days of August 2010. The details of the model configuration are given in the next section. Using 00 UTC initial conditions, we have derived 12 UTC and 18 UTC predictions of all the

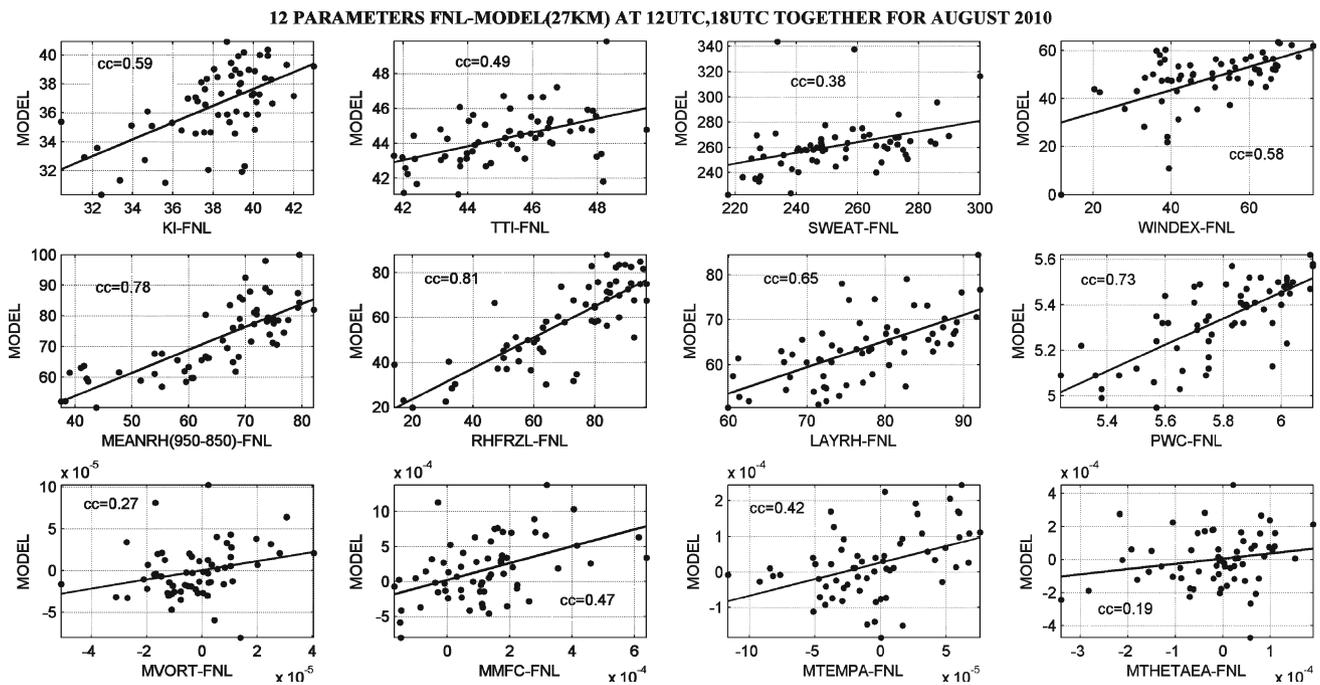


Figure 4. Scatter plot of 12 parameters identified for the model development derived from WRF model prediction and NCEP FNL for the month of August 2010 (12 and 18 UTC combined).

12 parameters. Figure 4 shows the scatter plots showing the 12 parameters derived from the WRF predictions (12 UTC and 18 UTC combined) and the corresponding values derived from NCEP FNL data. It shows that the WRF model has some ability to predict all the parameters.

Before the analysis, we have standardized all the 12 predictors using their respective mean and standard deviation. The standardization was required as different parameters have different mean values. The final model with the following six parameters was found to have the best performance during the training period. The BLR model thus arrived is given below:

$$\ln \left( \frac{p_i}{1 - p_i} \right) = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (3)$$

where,  $x_1$  to  $x_6$  are the predictors, TTI, WINDEX, RHFRZL, LAYRH, MEANMFC and MTEMPA respectively. The binary logistic regression coefficients thus calculated are  $b_0 = -0.9249$ ,  $b_1 = -0.0686$ ,  $b_2 = 0.7110$ ,  $b_3 = 0.7476$ ,  $b_4 = 0.5504$ ,  $b_5 = 0.4199$  and  $b_6 = 0.5669$ .

The model contains a parameter, TTI representing atmospheric instability. WINDEX is a parameter representing atmospheric instability and moisture content. RHFRZL and LAYRH represent availability of moisture content in the atmosphere. MEANMFC and MTEMPA are the dynamical variables representing low level moisture convergence and temperature advection. The plots of

cumulative frequency of thunderstorm for each of the six predictors used in the model are shown in figure 5.

The BLR equation (3) provides a probability ranging between 0 and 1. To forecast whether a lightning will occur, a threshold probability needs to be determined. If the calculated probability is more than the threshold, then the forecast indicates that a thunderstorm will occur, if not, no thunderstorm will occur.

The optimum threshold was determined using verification scores from a  $2 \times 2$  contingency table using varying thresholds. The verification scores considered are the Probability of Detection (POD), BIAS, False Alarm Rate (FAR) and the Critical Success Index (CSI). The POD is the ratio of the number of events correctly predicted by the model to the total number of observed events in the sample. The FAR is a measure of the forecast events that fail to occur. The bias B indicates the degree of over-forecasting ( $B > 1$ ) or under-forecasting ( $B < 1$ ) an event. Finally the CSI combines attributes of the POD and FAR and can be viewed as a hit rate (HR) for the event being forecast after removing correct no forecasts from consideration (Wilks 2006).

Figure 6 shows the variation of these verification statistics using different probability thresholds. The value of the scores decreases as the threshold is increased. As done by Reap (1994b) and Shafer and Fuelberg (2006), we sought to maximize the CSI and POD while minimizing the FAR

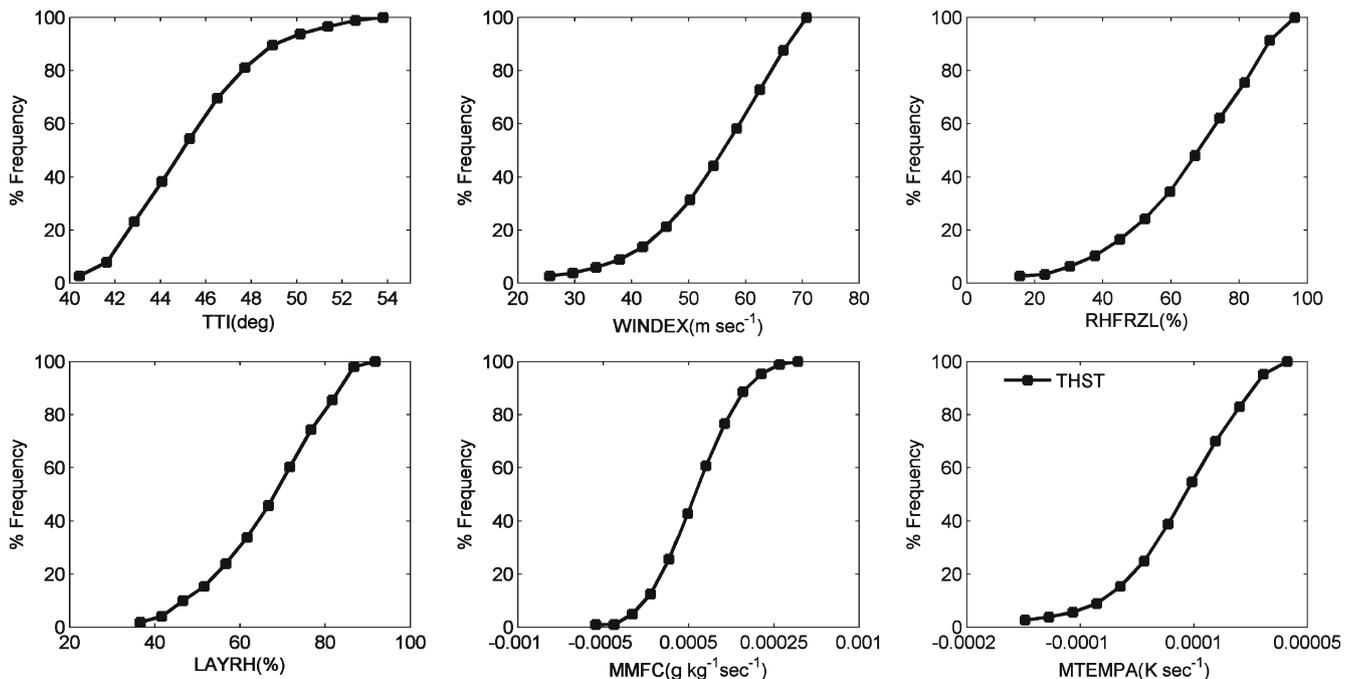


Figure 5. Cumulative frequency distribution plots of the six predictors used for the development of BLR model.

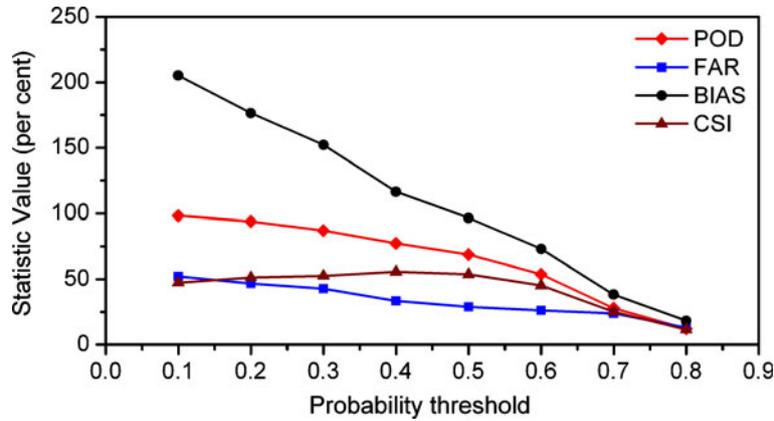


Figure 6. Values of CSI, POD, FAR and BIAS for varying probability thresholds. The statistics are derived from the BLR model giving the probability of lightning.

and capturing as many of the nonevents as possible. We find that the value of POD decreases sharply just after the 50% threshold, while FAR does not improve much. Therefore, a 50% threshold was chosen for the SHAR region to determine the probability of thunderstorm. At the 50% threshold, the model has a bias just below one which suggests the model has a tendency of predicting less number of lightning cases compared to the observations.

The contingency table showing the performance of the model with 50% threshold for the 305 cases is shown in table 3. From the table, the verification scores measuring the skill of the model are calculated and given in the same table. The probability of detection is 69% and false alarm rate is around 29%. The model has a bias towards predicting less frequent lightning compared to the observed frequency. The CSI is 53.7%. The model also has a reasonable overall hit rate of 77.7%.

Table 3. A  $2 \times 2$  contingency table for the number of cases with lightning was observed vs. the number of cases predicted by the BLR model over the SHAR region with 50% as the threshold. These results are for the model development period (2006–2009).

Observed	Predicted		Total
	Yes	No	
Yes	79	36	115
No	32	158	190
Total	111	194	305
Probability of detection	POD = $79/115 = 68.6\%$		
Overall hit rate	HR = $(79 + 158)/305 = 77.7\%$		
False alarm ratio	FAR = $32/111 = 28.8\%$		
BIAS	B = $(79 + 32)/(79 + 36) = 96.5\%$		
Critical success index	CSI = $79/(79 + 36 + 32) = 53.7\%$		
Hit rate non-events	HRNE = $158/190 = 83.1\%$		
Brier skill score	BSS = 0.20		

The Brier score calculated for the 305 cases is 0.222 and the Brier skill score (BSS) is 0.20 which is very encouraging.

#### 4.2 Independent verifications

For validating the performance of the probabilistic prediction scheme with independent data, we have considered 92 independent days in the months of May, June and August 2010. Lightning activity was very strong over the SHAR region during this period with 39 days of lightning activity.

For testing the performance of the model for these independent cases, we need values of the six predictors used in the BLR model. As the perfect prognosis method envisages, these values are derived from a NWP model. For this purpose, we have considered the WRF model (V 3.1.1) for generating forecasts of these predictors. The details of the WRF model configuration are given in table 4. Due to the constraints in computing facility, the model was run in a single domain with 27 km resolution. We assume, the spatial resolution of 27 km is adequate to resolve the large scale environmental conditions triggering the thunderstorm activity and also to derive the predictors used in the BLR model. However, to predict three-dimensional structure of thunderstorm, we need to run the model with very high resolution (1–2 km). Even in this configuration, many unresolved issues remain to predict the genesis and life cycle of a thunderstorm (e.g., Rajeevan et al 2010).

For making predictions with the WRF model, the initial and boundary conditions for 0000 UTC initializations are taken from the NCEP GFS (<http://ftp.cgd.noaa.gov/pub/data/nccf/com/gfs/prod/>). To improve the forecasts, local meteorological observations observed by automatic

Table 4. Details of the WRF model configuration used for prediction.

Horizontal resolution	27 km
Vertical levels	38
Dynamical core	ARW
Cumulus parameterization	Betts Miller and Janjic Scheme
Microphysics	Thompson Scheme
Planetary boundary layer	Yonshing University PBL Model
Shortwave radiation	RRTM
Longwave radiation	Lacis and Hansen
Land surface process model (LSM)	Noah LSM
Time step for integration	120 s

weather stations of Indian Space Research Organization (ISRO), Kalpana satellite derived atmospheric motion vectors, upper air observations recorded by ISRO GPS sonde, winds from the MST Radar installed at Gadanki were also assimilated. The observations during the period 00–03 UTC were only considered for the assimilation. Before assimilation, the duplicate observations are ignored and the quality and consistency checks are also performed. The assimilation is done using the observation nudging method described by multi-quadratic scheme. It uses hyperboloid radial basis functions to perform the objective analysis. If observations are not sufficient to evaluate multi-quadratic function, the Cressman method is used for nudging these observations.

Using the 24-h forecasts of the WRF model, the data of the six predictors used in the BLR model were derived for all the days of May, June and August 2010. The first 6 hours were used for spin up of the model and therefore not considered for calculating probability. Since lightning activity over the SHAR region is maximum during the late evening and early night, we have considered the period 09–21 UTC for testing the model. The BLR model was tested for the period May, June and August 2010 by predicting lightning probability for the period 09–21 UTC period. If the model predicted probability during this 12-h period exceeded 0.5 for at least in one hour period, then it is assumed that the model has predicted a lightning. If a lightning case was observed during this 12-h period (confirmed by the field mill data at the SHAR Space Centre and Chennai DWR data), then the forecast is assumed as correct. Otherwise, the model forecast was assumed as incorrect.

The results for the independent period, May, June and August 2010 are shown in table 5. During this period, there were 39 days in which a lightning was observed over the SHAR region during the 09–21 UTC period. The model was able to predict the lightning activity (probability exceeding 0.5) in 30 days out of these 39 days with 77% of success rate. However, the model also showed a large bias of false alarm with 35% of false alarm rate. Out

of 53 non-thunderstorm days, the model was able to correctly indicate non-lightning activity only in 37 days. For the remaining days, the model gave a false alarm. The model has shown a bias more than 1, suggesting the model has a tendency of predicting more lightning cases compared to the observations. The overall hit rate is 73%, which is smaller than the value during the model development period. CSI is 54.5%, which is slightly more than the value during the model development period.

We have also calculated Brier Skill Score (BSS) for the independent period. The Brier score for the probabilistic model was 0.25 and for the persistence model, it was 0.326. Thus the BSS of the model for the independent period was 0.233, which suggests the probabilistic model has positive and useful skill in predicting lightning cases over the SHAR region.

The model coefficients were calculated with the FNL data pertaining to the SHAR region. Therefore, the model as it is, cannot be applied for thunderstorm prediction over any parts of the country. However, we believe that the model with the same coefficients can be used for south-east India, the region close to the SHAR Space Centre. This model was recently tested on real-time application for a severe thunderstorm activity occurred over Gadanki, about 100 km west of the SHAR

Table 5. Performance of the BLR model during the independent period, May–June and August 2010.

Observed	Predicted		Total
	Yes	No	
Yes	30	9	39
No	16	37	53
Total	46	46	92
Probability of detection	POD = 30/39 = 76.9%		
Overall hit rate	HR = (30 + 37)/92 = 72.8%		
False alarm ratio	FAR = 16/46 = 34.7%		
BIAS	B = (30 + 16)/(30 + 9) = 117.9%		
Critical success index	CSI = 30/(30 + 9 + 16) = 54.5%		
Hit rate non-events	HRNE = 37/46 = 69.8%		
Brier skill score	BSS = 0.233		

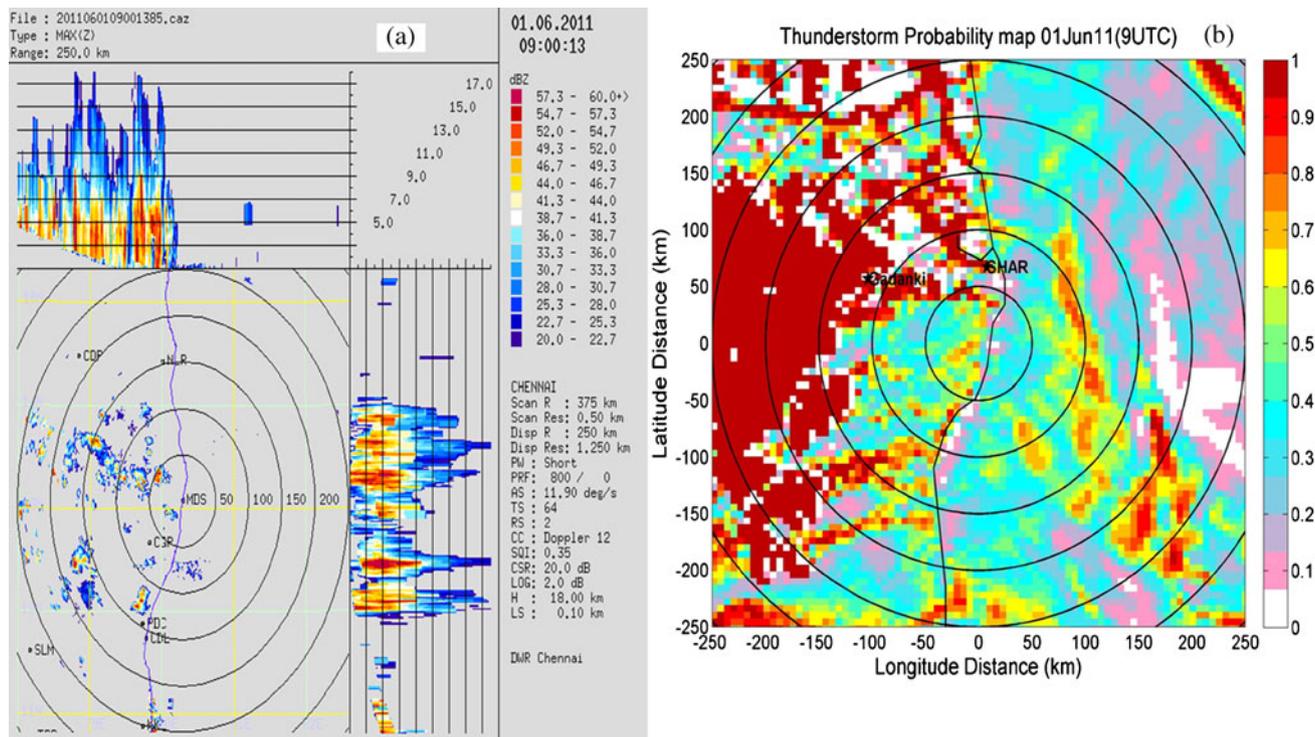


Figure 7. (a) Doppler Weather Radar data at 1000 UTC of 1 June, 2011 showing thunderstorm activity over the Gadanki region and (b) prediction of probability of lightning at 0900 UTC of 1 June, 2011 using 00 UTC initial conditions. Probability threshold of 0.5 is used for identifying lightning activity.

region on 1 June 2011. For deriving the predictors, we have used the WRF model with two nested domains of 18 and 6 km, respectively. We have used the NCEP GFS initial conditions of 00 UTC of 1 June 2011. Probability of lightning occurrence has been calculated using the BLR model. The predictors are derived from the WRF model forecasts of 6 km resolution. Figure 7 shows the probability forecasts of the severe thunderstorm event occurred over Gadanki using the prediction scheme developed in this study. The Doppler weather radar image from Chennai DWR station is also shown in the same figure. It can be seen that the model has shown useful skill in predicting the evolution of large scale convective activity over the region on 1 June over south-east India. The probability map shows an area of high probability oriented southwest–northeast direction, west of the SHAR region, across Gadanki.

## 5. Conclusions

A prediction scheme based on the perfect prognosis method was developed to predict probability of thunderstorm/lightning and probable time of occurrence over the SHAR region. For this purpose, 115 cases of lightning observed over the SHAR region and 190 non-lightning cases were considered.

The probability model is based on the BLR technique, in which the probability of thunderstorm (yes or no) was the predictand. Using the NCEP FNL analysis, data of 61 potential predictors were derived. The final list of predictors used in the model was objectively selected using a three-stage strategy (Spearman ranking correlation, cumulative probability distribution and principal component analysis). The 50% probability was identified as the threshold for predicting lightning probability. The model developed showed encouraging results. For the independent verifications of the prediction scheme, the model was tested with independent data for the months of May, June and August 2010. The six predictors used in the BLR were derived using the WRF model with 27 km resolution with 00 UTC initial conditions. The probability of detection was 77% with the false alarm rate of 35%. The critical success rate was 55%. The BSS of the probabilistic model during the independent period was 0.233. This suggests that the model has a positive and useful skill in predicting lightning probability over the SHAR region with reasonable accuracy.

The present study is an exhaustive one within the limitations of data and computing system resources. However, we believe that there is good scope for further improvement of the model performance. The false alarm rate is about 35%, which is a major concern and needs to be reduced.

Similarly, the prediction model has a positive bias, a tendency of predicting more lightning frequency than observed. The present results can therefore be further improved in many ways. The number of lightning cases (115 cases) considered for the analysis is small for the model development using BLR technique. Increasing the number of lightning cases appreciably for training the model will help us to obtain more robust regression coefficients. In the next phase, we plan to include more lightning cases by consulting the weather records at nearby stations like Chennai, Tirupati, etc., also. In this study, we have combined the lightning cases during different seasons due to the availability of limited lightning cases. If we have adequate cases for different seasons, we can develop prediction schemes for different seasons separately. For want of required computing facility, we have considered the model resolution only as 27 km with the assumption that this resolution is sufficient for resolving large scale synoptic environment responsible for lightning genesis and also deriving the model predictors. However, it is proposed to test whether the present results will improve if we consider better resolution (say 9 km) for the WRF predictions. Similarly, the model results also may depend on the physical parameterization scheme used in the model. Some sensitivity studies are required to understand the dependence on parameterization schemes, especially the convective parameterization schemes. The predictions of large scale synoptic features associated with the thunderstorm occurrence could be sensitive to the physics used in the model. Similarly, the WRF predictions can be greatly improved by assimilating more additional local data, especially the DWR data. This will be a major agenda for our future research.

The present study further highlights many gap areas where institutions like the India Meteorological Department needs to put extra efforts to improve the quality of monitoring and predicting mesoscale systems like thunderstorms. For example, the present observational system for monitoring thunderstorm genesis and its propagation is not adequate. We need to develop a large observational network for detecting the thunderstorm genesis and life cycle, especially over the area where thunderstorm frequency is high. More observational and modelling studies are also required to understand the physical mechanisms of thunderstorm genesis and its life cycle.

### Acknowledgements

The authors are thankful to the Director, NARL for his encouragements for this study and to Dr. Mohan Satyanarayana, NARL for some fruitful

discussions. They also thank the two anonymous reviewers for their constructive and critical comments, which helped them to improve the quality of the paper. The authors extend their thanks to Prof. Fuelberg, Florida State University and Dr Shaffer, NOAA for useful discussions on the development of probabilistic model. They also thank Dr S B Thampi, Scientist, DWR Chennai for providing DWR data and images.

### Appendix 1

Formulae of some of the predictors used in the study

- 1) Convective available potential energy (CAPE) is calculated as:

$$CAPE = \int_{LFC}^{EL} (T_{vp} - T_{ve}) R_d d(\ln P)$$

- 2) Convective inhibition energy (CINE) is given by:

$$CINE = \int_{ps}^{LFC} (T_{vp} - T_{ve}) R_d d(\ln P)$$

where  $T_{vp}$  = virtual temperature of parcel,  $T_{ve}$  = virtual temperature of environment,  $g$  = acceleration due to gravity,  $LFC$  = level of free convection,  $EL$  = equilibrium level and  $ps$  = surface pressure.

- 3) Price and Rind Frequency (PRFREQ) (Price and Rind 1992) is given by:

$$F = (3.44 \times 10^{-5}) \times CCTHGT^{4.9}$$

where  $CCTHGT$  is the Convective cloud-top height.

- 4) Cloud physics thunder parameter (CPTP) (Bright *et al* 2005) is given by:

$$CPTP = \frac{(-19^\circ\text{C} - T_{EL})(CAPE_{-20} - K)}{K}$$

where  $K = 100$  J/kg (constant),  $CAPE_{-20}$  is cape from 0 to  $-20^\circ\text{C}$ , and  $T_{EL}$  is temperature at equilibrium level.

- 5) Lifted index is given by:

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

where  $T_p$  = parcel temperature and  $T$  = air temperature.

- 6) K index (KI) is given by:

$$KI = (T_{850} - T_{500}) + T_{d850} - (T_{700} - T_{d700})$$

where  $T$  is air temperature and  $T_d$  is dew point temperature.

7) Total Total index (TT) is given by:

$$TT = T_{850} + T_{d850} - 2T_{500}$$

8) Severe weather threat index (SWEAT) is given by:

$$12 \times T_{d850} + 20 \times (TT - 49) + 2 \times ws_{850} \\ + 2 \times ws_{500} + 125 \\ \times (\sin(wd_{500} - wd_{850})) + 0.2$$

where  $T_d$  is dew point temperature, TT is Total Total index,  $ws$  is wind speed,  $wd$  is wind direction.

9) Wind index is given by:

$$WINDEX = 5 \times (HM \times RQ \\ \times (G \times G - 30 + QL - 2 \times QM))^{0.5}$$

where  $HM$  is the height at melting level in km above surface,  $QL$  is the mixing ratio in lowest 1 km above surface,  $RQ = QL/12$  and  $G =$  mean lapse rate from surface to melting level.

## References

- Banacos P C and Schultz D M 2005 The use of moisture flux convergence in forecasting convective initiation: Historical and operational perspectives; *Wea. Forecasting* **20** 351–366.
- Bhowmik Roy S K, Soma Sen Roy and Kundu P K 2007 Analysis of large-scale conditions associated with convection over the Indian monsoon region; *Int. J. Climatol.*, doi: 10.1002/joc.1567.
- Bothwell P D 2002 Prediction of cloud-to-ground lightning in the western United States; Ph.D. thesis, University of Oklahoma, 178p.
- Bothwell P D 2005 Development of an operational statistical scheme to predict the location and intensity of lightning; *Conference on Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., 6p.
- Bothwell P D 2008 Predicting the location and intensity of lightning using an experimental automated statistical methods; *Third Conference on Meteorological Applications of Lightning Data*, New Orleans, LA, Amer. Meteor. Soc., 6p.
- Bothwell P D and Buckley D R 2009 Using the perfect prognosis technique for predicting cloud-to-ground lightning in mainland Alaska; *Fourth Conference on Meteorological Applications of Lightning Data*, Phoenix, AZ, Amer. Meteor. Soc., 10p.
- Bright D R, Wandishin M, Jewell R and Weiss S 2005 A physically based parameter for lightning prediction and its calibration in ensemble forecasts; Preprints, *Conference on Meteorological Applications of Lightning Data*, Amer. Meteor. Soc., San Diego, CA (CD-ROM 4.3).
- Brier G W 1950 Verification of forecasts expressed in terms of probability; *Mon. Wea. Rev.* **75** 1–3.
- Burrows W R, Price C and Wilson L J 2005 Warm season lightning probability prediction for Canada and the northern United States; *Wea. Forecasting* **20** 971–988.
- Chatterjee P, Pradhan D and De U K 2008 Simulation of local severe storm by mesoscale model MM5; *Ind. J. Radio Space Phys.* **37** 419–433.
- De U S, Dube R K and Prakasa Rao G S 2005 Extreme weather events over India in the last 100 years; *J. Ind. Geophys. Union* **9(3)** 173–187.
- Dasgupta S and De U K 2007 Binary logistic regression models for short term prediction of premonsoon convective developments over Kolkata (India); *Int. J. Climatol.* **27** 831–836.
- Ghosh S, Sen P K and De U K 1999 Identification of significant parameters for the prediction of pre-monsoon thunderstorms at Calcutta; *Int. J. Climatol.* **19** 673–681.
- Hughes K K 2001 Development of MOS thunderstorm and severe thunderstorm forecast equations with multiple data sources; Preprints, *18th Conf. on Weather Analysis and Forecasting*, Fort Lauderdale, FL, Amer. Meteor. Soc., pp. 191–195.
- Klein W H 1971 Computer prediction of precipitation probability in the United States; *J. Appl. Meteor.* **10** 903–915.
- Lambert W C, Wheeler M and Roeder W 2005 Objective lightning forecasting at Kennedy Space Center and Cape Canaveral Air Force Station using cloud-to-ground lightning surveillance system data; Preprints, *Conf. on Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., 4.1.
- Litta A J and Mohanty U C 2008 Simulation of a severe thunderstorm event during the field experiment of STORM programme 2006 using WRF-NMM model; *Curr. Sci.* **95(2)** 204–215.
- Manohar G K, Kandalgaonkar S S and Tinmaker M I R 1999 Thunderstorm activity over India and the Indian southwest monsoon; *J. Geophys. Res.* **104** 4169–4188.
- Mazany R A, Businger S, Gutman S I and Roeder W 2002 A lightning prediction index that utilizes GPS integrated precipitable water vapor; *Wea. Forecasting* **17** 1034–1047.
- McCann D W 1994 Windex – a new index for forecasting microburst potential; *Wea. Forecasting* **9** 532–541.
- Mukhopadhyay P, Singh H A K and 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; *J. Earth Syst. Sci.* **118** 441–466.
- Nath A, Manohar G K, Dani K K and Devara P C S 2009 A study of lightning activity over land and oceanic regions of India; *J. Earth Syst. Sci.* **118** 467–481.
- Neumann C J and Nicholson J R 1972 Multivariate regression techniques applied to thunderstorm forecasting at the Kennedy Space Center; Preprints, *Int. Conf. on Aerospace and Aeronautical Meteorology*, Washington, DC, Amer. Meteor. Soc., pp. 6–13.
- Price C and Rind D 1992 A simple lightning parameterization for calculating global lightning distributions; *J. Geophys. Res.* **97** 9919–9933.
- Rajeevan M, Kesarkar A, Thampi S B, Rao T N, Radhakrishna B and Rajasekhar M 2010 Sensitivity of WRF cloud microphysics to simulations of a severe thunderstorm event over south-east India; *Ann. Geophys.* **28** 603–619.
- Ranalkar M R and Chaudari H S 2009 Seasonal variation of lightning activity over Indian subcontinent;

- Meteor. Atmos. Phys.* **104** 125–134, doi: 10.1007/500703-009-0026-7.
- Reap R M and MacGorman D R 1989 Cloud-to-ground lightning: Climatological characteristics and relationships to model fields, radar observations, and severe local storms; *Mon. Weather Rev.* **117** 518–535.
- Reap R M 1994a Analysis and prediction of lightning strike distributions associated with synoptic map types over Florida; *Mon. Weather Rev.* **122** 1698–1715.
- Reap R M 1994b 4-h NGM based probability and categorical forecasts of thunderstorms and severe local storms for the contiguous U.S; *NWS Technical Procedures Bulletin* **419** 14p.
- Shafer P E and Fuelberg H E 2006 A statistical procedure to forecast warm season lightning over portions of the Florida peninsula; *Wea. Forecasting* **21** 851–868.
- Shafer P E and Fuelberg H E 2008 A perfect prognosis scheme for forecasting warm-season lightning over Florida; *Mon. Weather Rev.* **136** 1817–1846.
- Srivastava K, Roy Bhowmik S K, Hatwar H R, Das A K and Kumar A 2008 Simulation of mesoscale structure of thunderstorm using ARPS model; *Mausam* **59** 1–14.
- Tinmaker M I R, Kaushar A and Beig G 2009 Relationship between lightning activity over peninsular India and sea surface temperature; *J. Appl. Meteor.* **49** 828–835.
- Wilks D S 2006 Statistical methods in the Atmospheric Sciences, *International Geophysics Series*, Academic Press, **91**, 627p.
- Wilson J W, Crook N A, Mueller C K, Sun J and Dixon M 1998 Nowcasting thunderstorms: A status report; *Bull. Am. Meteor. Soc.* **79** 2079–2099.

*MS received 22 April 2011; revised 7 November 2011; accepted 9 November 2011*