J5.6

A REAL-TIME LEARNING TECHNIQUE TO PREDICT CLOUD-TO-GROUND LIGHTNING

V. Lakshmanan[1,2*] and Gregory J. Stumpf[1,3]

[1]*Cooperative Institute for Mesoscale Meteorology Studies, Univ. of Oklahoma, Norman, OK*
[2]*NOAA/National Severe Storms Laboratory, Norman, OK*
[3]*NOAA/National Weather Service Meteorological Development Laboratory, Silver Spring, MD*

1. INTRODUCTION

A short-term (0-1 hour) warning for intense cloud-to-ground (CG) lightning has the potential to be a very valuable new National Weather Service (NWS) operational product. A variety of forecast techniques and "rules-of-thumb" have been developed by several local NWS Weather Forecast Offices (WFO) which manually combine Weather Surveillance Radar – 1988 Doppler (WSR-88D) data with thermodynamic information from local rawinsondes to predict CG lightning occurrence (Vincent et al. 2002). Several WFOs are also issuing "excessive lightning alerts" based on the frequency of CG lightning in storm cells. The NWS Meteorological Development Laboratory (MDL) partnered with the NSSL to develop a prototype application to predict intense CG lightning and perhaps add guidance for issuance of current and future operational lightning warning products.

Although a 0-3 hour forecast of CG lightning is available in AWIPS, this product provides no guidance on the likelihood of very high CG flash rates (Kitzmiller et al. 1999). Other lightning prediction applications have also been developed to forecast the occurrence of convection and lightning in the time frame beyond the typical 0-1 hour severe weather warning periods (Bright et al. 2005, Keller 2004). However, our intent is to concentrate on the development of an intense CG lightning warning application with emphasis on the 0-1 hour time frame.

Using the National Severe Storms Laboratory (NSSL) Warning Decision Support System – Integrated Information (WDSSII; Hondl 2002) as a prototyping environment, a multiple-sensor application which predicts the initiation and the advection of CG lightning is being developed. This application first combines radar data from multiple WSR-88D locations into a three-dimensional (3D) Cartesian grid (Lakshmanan 2002). The radar grid is then integrated with near-storm environment information extracted from the 20 km Rapid Update Cycle model (RUC20) initial analysis grids. The multi-sensor data are finally integrated with CG lightning data from the National Lightning Detection Network (NLDN) to predict the onset of CG lightning. A statistical clustering scheme can also be used to advect the 3D radar data, as well as any existing CG lightning density grids (Lakshmanan et al. 2003).

The application uses a radial basis function (RBF) to form a relation between past-observed multi-sensor meteorological variables to current CG lightning activity. The weights and standard deviations of the RBF are optimally determined by training it on historical data. The RBF relationship matrix is constantly re-calibrated in real-time, and used to predict the onset of CG lightning activity in the future based on current observations of multiple-sensor variables. This training is continually repeated at regular intervals (e.g., every 5, 15, or 30 minutes) based on the latest CG lightning flash density information, and the radar and NSE data from a time prior to the current time. This time difference becomes the prediction interval (e.g., 30 minutes).

A number of multiple-sensor derived variables, with a 3D radar reflectivity grid as a "base variable" can be tested as predictors for future CG lightning activity. Tested are input variables used to diagnose radar reflectivity information within the thermodynamic levels in the mixed-phase (liquid and ice) region of storms, typically between 0° and -20°C, where charge separation and subsequent lightning is known to occur. These can include reflectivity values at 0°C, -10°C, and -20°C constant temperature altitudes as determined by integrating 3D gridded reflectivity data with RUC zero-hour analysis grids. Also tested are various vertically-integrated and layer-averaged reflectivity parameters in the mixed-phase region (e.g., Severe Hail Index, Witt et al. 1998). In this paper, we report on the current status of the application (which is relatively new). We hope to provide more information at the conference, including a preliminary evaluation of the application.

2. RADIAL BASIS FUNCTIONS

Radial Basis Functions (RBFs) are, like other neural networks, a non-parametric regression technique. Unlike in the more common multi-layer perceptron networks, the activation of a hidden unit in a RBF is determined by the distance between the input vector and a prototype vector (Bishop 1995). Intuitively, one can think of a RBF as forming a set of examples (along with their desired outputs), and formulating the output for any new input as a linear combination of these examples, with the examples closer in distance to the new input providing higher weights.

Mathematically, a linear mapping between the input and output can be assumed to be of the form:

$$h(x) = \Sigma_n w_n \phi_n(||x - x_n||) \quad (1)$$

where x represents the input vector, x_n the nth prototype vector, and || represents a distance measure, usually the Euclidean distance. A common choice for the function φ_n is a Gaussian function:

$$\phi_n(r) = e^{\frac{-r^2}{2\sigma_n^2}} \quad (2)$$

because of its highly localized nature. Since the Gaussian is a non-linear function, the RBF is a non-linear function. However, if we keep the number and values of the hidden prototypes (x_n) fixed, then the RBF is a linear model and can be solved by least squares methods (involving effectively just a matrix inversion and several multiplications). For the necessary derivations, the interested reader is referred to Broomhead and Lowe (1988) and Bishop (1995).

A RBF where the prototypes are kept fixed can be solved in a deterministic manner without resorting to numerical methods such as gradient descent. Therefore, such a RBF is particularly amenable to real-time learning tasks.

3. METHOD

We formulate this problem as a spatial-temporal prediction problem. At a particular location, we seek to estimate the probability that there will be a lightning strike at that position in the next 30 minutes, for example. Since lightning is an almost instantaneous event, the probability of lightning is also estimated in a spatial-temporal sense: a particular location is said to have experienced lightning if there is a lightning strike within a given distance of that location within a given time period. This spatial-temporal definition of lightning activity is represented by a lightning density grid.

3.1 Lightning Density

The lightning density grid is a two-dimensional grid that has a resolution of 0.01° latitude and longitude (approximately 1km × 1km). The remapping of lightning source data into lightning density grids is achieved using temporal averaging and spatial smoothing. All the source data that impacts a grid cell over a given time period are used to determine the lightning density at a grid cell. We experimented with time periods ranging from 1 minute to 15 minutes. Spatially, we let each source impact not just the grid cell into which it falls, but all grid cells within a given radius (using a triangular neighborhood function to determine the weight of impact). In a later study, we will experiment using a Gaussian or some other smooth weighting function, but we don't expect to see much of an impact. The triangular weighting function was chosen mainly for computational speed. Figure 1 depicts an example gridded CG density field for a 5 minute time period from a case in Central Florida. Overlaid are actual CG strike points.

To avoid numerical problems resulting from the very small dynamic range of lightning density data, the lightning density data are remapped using an exponential function to a more linear range before training the RBF network. Forecasts are unmapped using a logarithmic function to their original range before being presented.

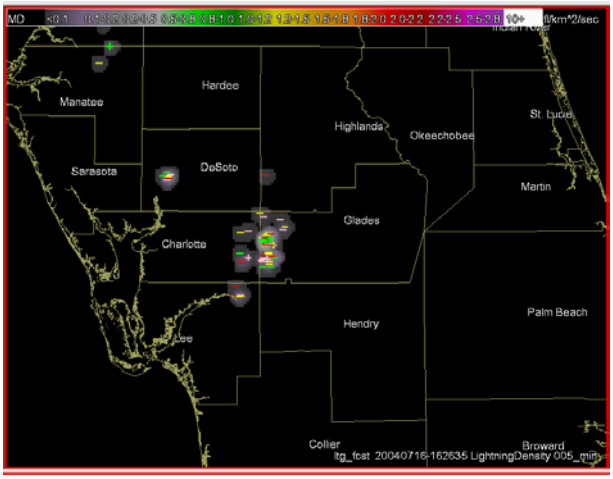

Fig. 1: Gridded lightning density and cloud-to-ground (CG) lightning locations. Minus (Plus) symbols represent negative (positive) CG strikes. Data are for the 5 minute period ending at 162635 UTC 16 July 2004. Location is Central Florida, and county names and boundaries are shown.

3.2 Predictors

When creating 3D grids of reflectivity from multiple radars (Lakshmanan 2002), we usually map the reflectivity values to constant altitudes above mean sea level. By integrating numerical model data zero-hour analysis grids, it is possible to obtain an estimate of reflectivity at constant temperature altitudes – at time intervals of less than an hour, this information is quite reliable. Thus, it is possible to compute the reflectivity value from multiple radars and interpolate it to points not on a constant altitude plane, but on a constant temperature surface. This information, updated every 60 seconds in real-time, is valuable for forecasting lightning.

Following Hondl and Eilts (1994) and Watson et al. (1995), as well as our own selection of "good predictors" of CG lightning, we can use determine the following multi-sensor attributes as potential predictors of CG lightning activity:

- Reflectivity at constant temperature altitudes of 0°C, -10°C, and -20°C.
- Echo top heights, from multiple radars to minimize radar geometry problems.
- Vertically Integrated Liquid (VIL; Green and Clark 1972), estimated from multiple radars.

- Vertically integrated reflectivities relative to constant temperature altitudes, such as a Severe Hail Index (Witt et al. 1998), or layer reflectivity differences (LRD), averages (LRA) and maxima (LRM) (e.g., average reflectivity between 0°C and -20°C).

- Cell echo areas at constant temperature altitudes.

A variety of these fields from the same Central Florida case are shown in Figure 2.

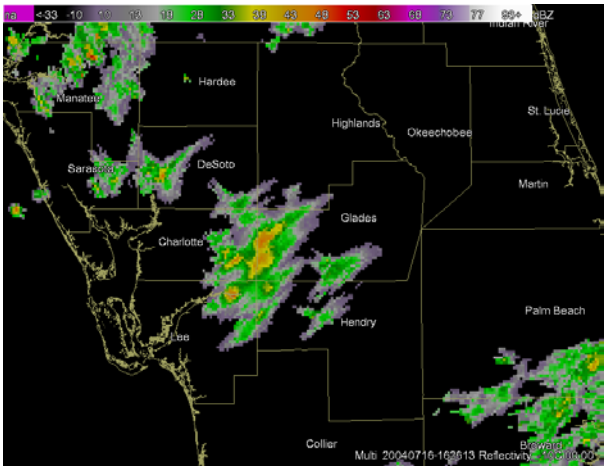

Fig. 2a: Reflectivity at the -10°C temperature altitude. Data are derived from 3D grid of reflectivity data from multiple radars at KMLB, KTBW, and KAMX. Data are at 162613 UTC 16 July 2004. Location is Central Florida, and county names and boundaries are shown.

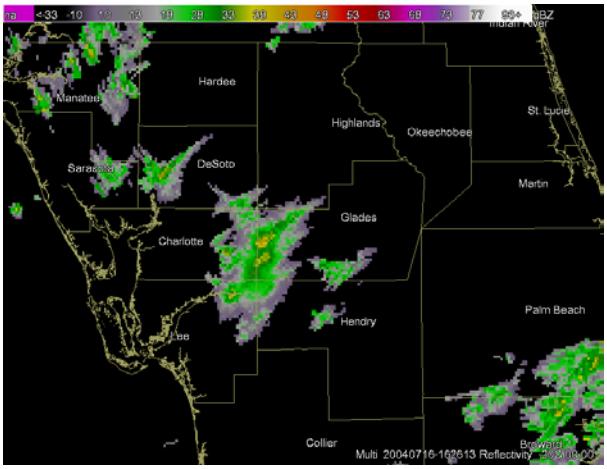

Fig. 2b: Same as Fig. 2a except for reflectivity at the -20°C temperature altitude.

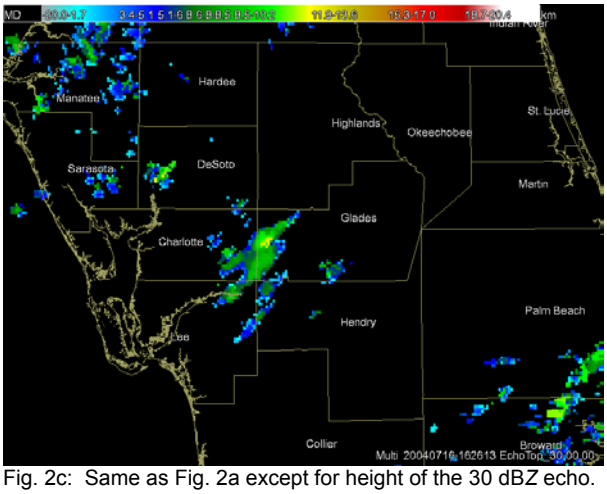

Fig. 2c: Same as Fig. 2a except for height of the 30 dBZ echo.

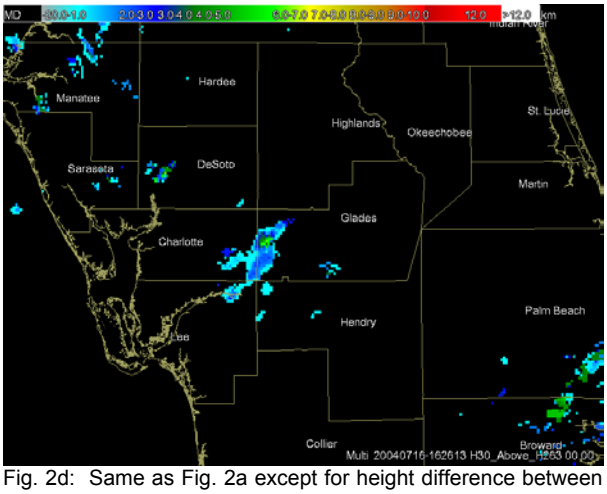

Fig. 2d: Same as Fig. 2a except for height difference between the height of the 30 dBZ echo and the height of the -10°C altitude.

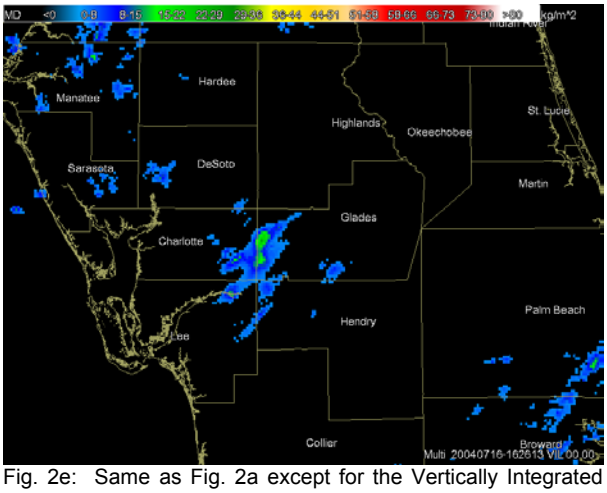

Fig. 2e: Same as Fig. 2a except for the Vertically Integrated Liquid (VIL).

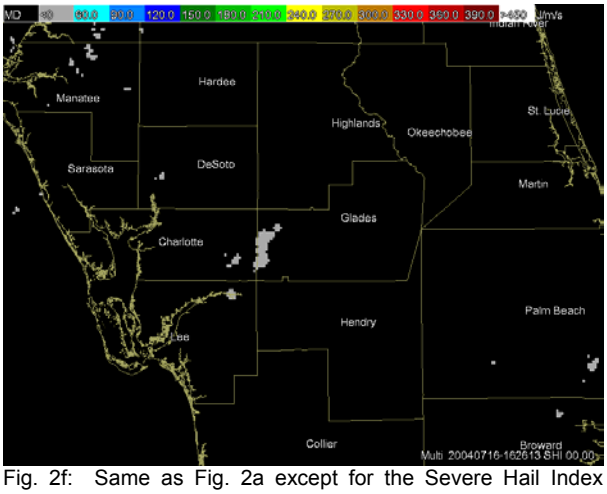

Fig. 2f: Same as Fig. 2a except for the Severe Hail Index (SHI).

3.3 Training and Forecast

One of the major concerns in RBF training is choosing the prototype vectors. Recall that for the RBF to be a linear model, the prototype vectors have to be kept fixed. To obtain a reasonable mapping between input and output, we try to ensure that there are enough pixels with lightning activity, and an approximately equal number of pixels without lightning activity in the data set. We maintain a historical queue of pixels and their attributes, along with the current CG lightning activity. As new data arrives, older training data is aged off. We will experiment with a variety of weighting schemes design to favor a larger proportion of more current data versus older data in the queue.

The size of the historical queue is set to ensure that the training constantly adapts to evolving conditions, but that there is enough data for training even in low-weather conditions. Every hour or so, the current training data is saved to disk, such that a statistical climatology can be developed for future runs of the program. The climatological information can be used as an initial default "background field" prior to the development of any convective activity.

Older CG lightning activity and the input training attributes can be advected to their expected location before being used for RBF training. The motion is estimated using K-means clustering of reflectivity data from multiple radars and tracking the clusters using a minimum mean-absolute error in combination with a Kalman filter (Lakshmanan et al. 2003).

For example, to train the RBF for 30-minute lightning prediction, we use the input attributes from 30 minutes ago, advected to their current expected positions. The current lightning density is used as the training output. The prototypes, x_n, for the RBF are chosen through K-means clustering of the available prototypes (without regard to the desired output). Then, using these prototypes, the RBF matrix equations are solved for the weights (w_n) and sigmas (σ_n). New forecasts of lightning density 30 minutes in advance are then made using the computed RBF equation and the current values for all the inputs, advected to their expected locations 30 minutes later.

4. RESULTS

Using the same Central Florida case, Figure 3 depicts a 30-minute forecasted lightning density field, with actual CG lightning strike points for the forecast time overlaid. This is using three multi-sensor inputs into the RBF, namely the multi-radar reflectivity at constant temperature altitudes of 0°C, -10°C, and -20°C. Figure 4 depicts the relationship between the three multi-sensor inputs to lightning density for the particular RBF equation trained on these data.

We anticipate more results to be reported at the conference as the evaluation of the application continues.

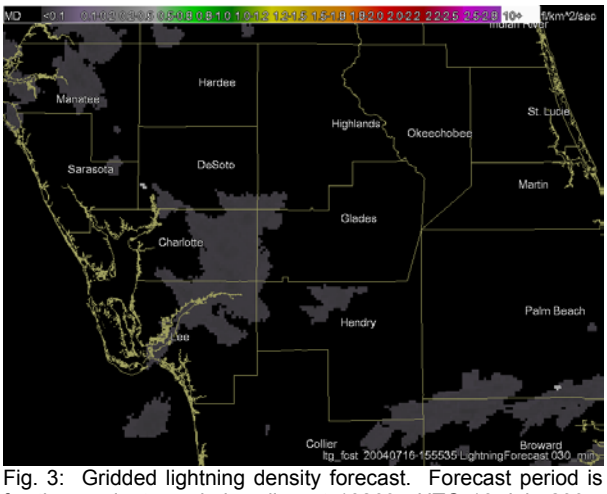

Fig. 3: Gridded lightning density forecast. Forecast period is for the 5 minute period ending at 162635 UTC 16 July 2004. Location is Central Florida, and county names and boundaries are shown.

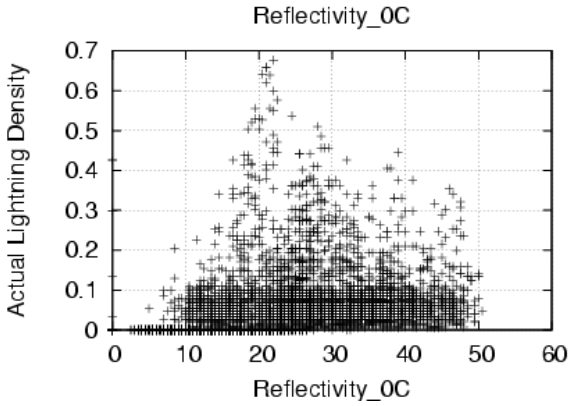

Fig. 4a: Distribution of values of the reflectivity at the 0°C altitude versus lightning density observations.

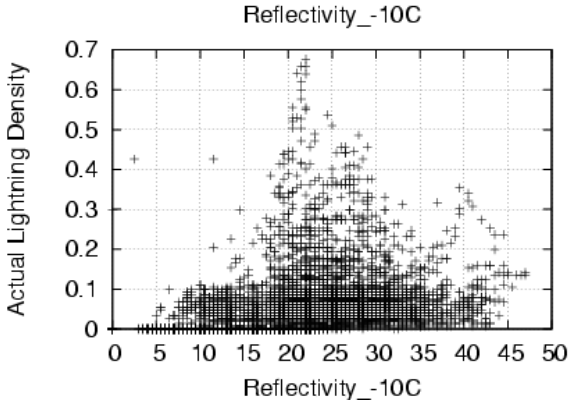

Fig. 4b: Distribution of values of the reflectivity at the -10°C altitude versus lightning density observations.

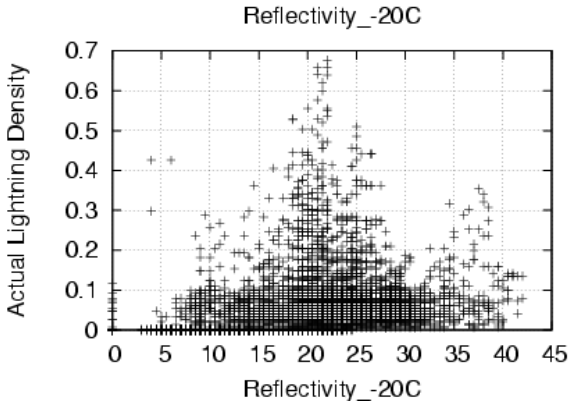

Fig. 4c: Distribution of values of the reflectivity at the -20°C altitude versus lightning density observations.

5. ACKNOWLEDGEMENTS

Funding for this research was provided under NOAA-OU Cooperative Agreement NA17RJ1227 and the National Science Foundation Grant 0205628. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Severe Storms Laboratory, the National Weather Service, or the U.S. Department of Commerce.

6. REFERENCES

Bishop, C., 1995: *Neural Networks for Pattern Recognition*. Oxford.

Bright, D. R., M. S. Wandishin, R. E. Jewell, and S. J. Weiss, 2005: A physically-based parameter for lightning prediction and its calibration in ensemble forecasts. *Preprints, Conf. on Meteo. Appl. Of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., CD Preprints.

Broomhead, D. and D. Lowe, 1988: Multivariable functional interpolation and adaptive networks. *Complex Systems*, **2**.

Greene, D. R., and R. A. Clark, 1972: Vertically integrated liquid water – A new analysis tool. *Mon. Wea. Rev.*, **100**, 548-552.

Hondl, K. and M. Eilts, 1994: Doppler radar signatures of developing thunderstorms and their potential to indicate the onset of cloud-to-ground lightning. *Mon. Wea. Rev.*, **122**, 1818–1836.

Hondl, K. D., 2002: Current and planned activities for the Warning Decision Support System – Integrated Information (WDSS-II). *Preprints, 21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 146-148.

Keller, D. L., 2004: Forecasting cloud-to-ground lightning data with AFWA-MM5 model data using the "Bolt Of Lightning Technique" (BOLT) algorithm. *Preprints, 22nd Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., CD preprints.

Kitzmiller, D. H., M. A. R. Lilly, and S. D. Vibert, 1999: The SCAN 0-3 hour rainfall and lightning algorithms. Available online at: http://www.nws.noaa.gov/tdl/radar/03h_doc.htm

Lakshmanan, V., 2002: Real-time merging of multisource data. *Preprints, 21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 225-228.

Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *Atmos. Research*, **66**, 367–380.

Vincent, B. R., L. D. Carey, D. Schneider, K. Keeter and R. Gonski, 2002: Using WSR-88D reflectivity for the prediction of cloud-to-ground lightning: A Central North Carolina study. *National Weather Digest*, (submitted).

Watson, A. I., R. L. Holle, and R. E. Lopez, 1995: Lightning from two national detection networks related to vertically integrated liquid and echo-top information from WSR-88D radar. *Wea. Forecasting*, **10**, 592–605.

Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286-303.