P1.9

PREDICTING LIGHTNING STRIKES FOR THE ENHANCEMENT OF FIRE WEATHER FORECASTS

Randall P. Benson*
South Dakota School of Mines and Technology, Rapid City, South Dakota

1. INTRODUCTION

Lightning is a safety risk to anyone working outdoors and is the source of ignition for nearly 60% of all wildfires each year in the Black Hills of South Dakota and Wyoming (Andrews and Bradshaw, 1997). The prediction of lightning has previously been an elusive aspect of a fire weather forecast. Improvements in forecasting the daily probability of lightning in fire weather forecasts could potentially greatly improve general planning operations of fire management officials.

Anticipating the electrification of a thunderstorm has proven to be difficult. Radar has been a tool used to aide in the assessment of lightning production by analyzing storm growth rate, formation and location of precipitation, and storm height. Using non-electric storm characteristics has also proven troublesome in predicting lightning production due to local differences in geography and storm-to-storm variability. One of the reasons for the complication involved in forecasting lightning production probably lies in the fact that storm electrification requires interactions of numerous storm parameters. However, an understanding of one of the leading mechanisms for lightning production (noninductive charging which relies on interactions and/or collisions of ice crystals and cloud graupel particles) may prove useful in prediction methods. Specifically, non-electric pre-storm parameters related to updraft speed, the structure of the wind field at different levels, influences of previous convection, and the nature of the microphysical properties of the cloud, particularly above the lower boundary of the mixed phase region, all may be useful predictors of lightning occurrence (MacGorman and Rust 1998).

Thunderstorms and associated lightning activity is currently forecasted by meteorologists using the Lightning Activity Level (LAL) to convey the likelihood of lightning for a specific area or region.

* *Corresponding author address:* Randall P. Benson, South Dakota School of Mines and Technology, Department of Atmospheric Science, 501 East Saint Joseph Street, Rapid City, South Dakota, 57701, email: randall.benson@sdsmt.edu

While serving a useful purpose, the LAL is an index based on forecasted meteorological variables. Specifically, maximum radar echo height, cumulus cloud development, predicted rainfall, and the anticipated frequency of lightning strikes during a finite period (5 minutes) comprise the LAL (Fuquay 1980). This research provides a statistical method to forecast lightning based on actual observations that removes much of the subjectivity and uncertainty involved in current lightning prediction for fire weather forecasts.

A variety of statistical models have been developed to predict thunderstorm and lightning occurrence. Reap and Foster (1979) used Model Output Statistics (MOS) to develop probability forecasts of lightning for 12-36 hours by incorporating digitized radar data with large scale meteorological parameters. Burrows *et al.* (2004) developed statistical guidance to predict lightning occurrence in Canada and the northern United States. Their method utilized a tree-based regression algorithm combining input from the Canadian Meteorological Center's Global Environmental Multiscale (GEM) numerical forecast model. Lambert *et al.* (2005) developed a set of statistical lightning forecast equations through multiple logistic regression to provide a daily probability of lightning occurrence for the Cape Canaveral Air Force Weather Station in Florida. Logistic regression techniques were utilized by Neumann and Nicholson (1972) using radiosonde-derived predictors and Mazany *et al.* (2002) developed an index for short-term lightning occurrence at the Kennedy Space Center in Florida.

Wilks (1995) determined that the logistic regression method is a more appropriate technique than linear regression for probability-based forecast equations. Everitt (1999) showed that using logistic regression versus linear regression provided 48% better skill when using the same predictor variables and data.

The current study uses multiple logistic regression to determine the predictors that best explain the daily variation in lightning occurrence. Predictors used for this statistical model are radiosonde-derived parameters that are commonly used by meteorologists to forecast convection. The radiosonde parameters provide the foundation for which the statistical model was developed. The cloud-ground lightning strike database (Vaisala, Inc.) for the Black Hills is from the period 1 January 1994 through 31 December 2003 and the domain for the observed lightning activity is shown in Figure 1. The 1200 UTC Rapid City radiosonde was used exclusively from the period 1 January 1994 through 31 December 2003. The radiosonde predictors were derived from the dataset and are available on the "Radiosonde Data of North America" CD-ROM distributed by the National Climatic Data Center (NCDC) and the Forecast Systems Laboratory (FSL) (NCDC and FSL 1999). Data from the years 2000 to 2003 were obtained directly from the FSL website (http://raob.fsl.noaa.gov). The statistical equations determine the probability of one or more lightning strikes occurring during the day of the observed upper air sounding.

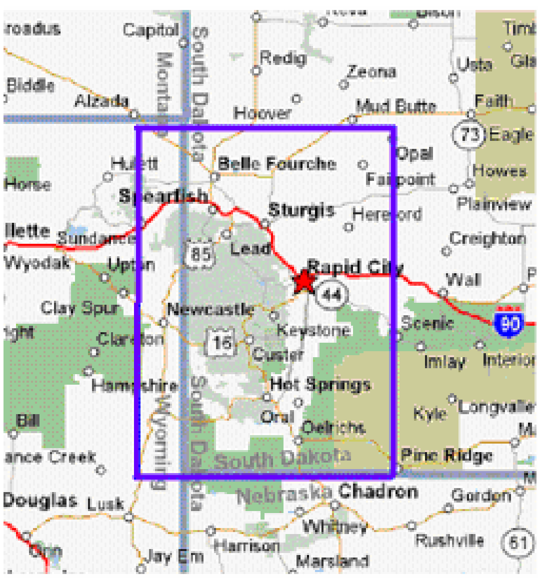

Figure 1. Cloud-ground lightning strike domain. National Lightning Detection Network data courtesy of Vaisala, Inc.

2. PREDICTAND, PREDICTORS, METHODS

A binary "1" was used to indicate a day with one or more lightning strikes and a "0" was used to indicate days in which there was no occurrence of lightning within the Black Hills domain. The radiosonde and lightning database was grouped into two-month periods beginning in March and ending at the end of the convective season in October.

Forty-five possible predictors were used to estimate the likelihood of convection and lightning. The potential predictors in the radiosonde database were an assortment of temperature, stability, moisture, pressure, and wind parameters. Several potential predictors were developed by calculating a parameter through different levels of the daily radiosonde and by different lifting mechanisms (Bolton 1980). Three distinct thermodynamic lifting mechanisms were employed for the development of the possible predictors: most unstable (MU); mixed layer (ML); and surface-based (SB). The MU parameters were calculated through the layer from the surface to 300 hPa above ground level with the largest equivalent potential temperature. The ML parameters are determined by lifting a layer with average potential temperature and mixing ratio through the depth of the mixed layer with average quantities derived by interpolation every 250 meters. The SB parameters are determined by lifting the surface parcel whereby when the SB parameter is the most unstable it will equal the MU parameter. Likewise, when the ML is equal to zero, the SB parameter will equal the ML parameter. Lastly, five wind shear terms were used over the lowest 1, 3, 4, 6 and 8 km of the radiosonde and were interpolated at 500 meter steps.

Additionally, the height of the 500 hPa surface was used as a potential predictor as well as the daily change in the height of the 500 hPa surface. The persistence of lightning was also used as a predictor. This was developed in a similar fashion as the daily occurrence of lightning. If lightning was observed the previous day in the domain area, then a "1" was entered for the persistence value and a "0" was entered if there was no observed lightning. A computer program was used to calculate both the historic and current values of the predictors used in the equations.

The historical radiosonde and cloud-ground lightning strike data (total daily lightning strikes within the domain) were combined into a single database and adjusted for any days of missing or bad data. Next, a natural log transformation of the total number of lightning strikes was used in the original data to eliminate the largest outliers through the use of simple linear regression. Specifically, a histogram plot of the residuals from the linear regression was developed and major outliers were removed from the radiosonde and lightning occurrence database until a normally distributed histogram was obtained.

The condensed database was then grouped into two-month periods beginning in March and continuing through October to obtain the seasonal equations. A statistical software program (MINITAB) was used to formulate the binary logistic regression equations and a backward stepwise regression method was used to screen the possible predictors. This method utilizes the p-value (Pearson 1900), which is also known as the rejection level, to determine which of the predictors should be eliminated. The p-value is a probability ranging from 0 to 1 and indicates by a small value that the relationship between the predictor and the predictand is not likely to be a coincidence and that the predictor may have statistical significance. In the backward stepwise regression technique, the predictor with the largest p-value was deleted first and this procedure was continued until only the smallest p-values (< .05) remained. The final logistic regression equation (Wilks 1995) takes the following form:

$$y = \frac{e^{\left(b_0 + b_1 x_1 + \ldots + b_k x_k\right)}}{1 + e^{\left(b_0 + b_1 x_1 + \ldots + b_k x_k\right)}}$$

where *y* is the predicted probability of occurrence, b_0 is the intercept, b_k are the coefficients for the predictors, x_k and *k* is the number of predictors.

Once the final set of equations was produced, the probability of a lightning strike was calculated for each day of the historical record. Then a verification was performed designed to analyze the days with lightning activity and the total number of days for each probability range. For each set of equations, the calculated historical probabilities were divided into quartiles based on the probability range (Figure 2). Optimally, the highest percentage of lightning days versus total days should be demonstrated in the top quartile of the probability range and the lowest percentage of actual lightning days versus total days in the bottom quartile.

3. RESULTS, DISCUSSION, CONCLUSIONS

The results from Figure 2 indicate that by and large for each month the upper two quartiles demonstrated a high percentage of lightning days to total days with a correspondingly high probability of a lightning strike. Especially encouraging are the results for the months of July and August, which correspond climatologically (1994-2003) to the highest occurrence of lightning in the Black Hills area each year (Vaisala, Inc). For the 10-year historical record, there were 341 lightning days in the July and August record and the percentage of lightning days (LD) to total days (TD) in the top two quartiles or in the calculated probability range from 50% to 99% was just over 86%. This result suggests that the logistic regression equation for July and August predicts with a high degree of accuracy (86%) the actual occurrence of one or more lightning strikes when the calculated probability is 50% or greater. Correspondingly, when the calculated probability was less than 50% then the occurrence of lightning was only 24%. Similar promising results are found for the months of May and June, the second most active convective period of the year.

PROB LTNG STRIKE					
MAR-APR	LDAYS	TDAYS	LD/TD	MNPLS	MXPLS
1	15	16	0.94	0.76	1.00
2	14	20	0.70	0.52	0.72
3	7	29	0.24	0.26	0.49
4	15	275	0.05	0.00	0.25
PROB LTNG STRIKE					
MAY-JUN	LDAYS	TDAYS	LD/TD	MNPLS	MXPLS
1	134	145	0.92	0.76	1.00
2	66	94	0.70	0.52	0.72
3	42	94	0.45	0.30	0.50
4	35	143	0.24	0.01	0.29
PROB LTNG STRIKE					
JUL-AUG	LDAYS	TDAYS	LD/TD	MNPLS	MXPLS
1	284	312	0.91	0.75	1.00
2	32	55	0.58	0.50	0.74
3	19	48	0.40	0.26	0.50
4	6	55	0.11	0.00	0.24
PROB LTNG STRIKE					
SEP-OCT	LDAYS	TDAYS	LD/TD	MNPLS	MXPLS
1	84	167	0.50	0.75	0.99
2	23	148	0.16	0.50	0.74
3	4	120	0.03	0.26	0.50
4	0	89	0.00	0.02	0.26

Figure 2. Probability of a lightning strike from the historical record. In the first column the numbers 1, 2, 3, and 4 represent each quartile of the probability range for that period (Mar-Apr, etc.). LDAYS is the number of actual days with one or more lightning strikes. TDAYS is the total number of days within the quartile. LD/TD is the percentage of lightning days to total days. MNPLS is the minimum calculated probability for that quartile. MXPLS is the maximum calculated probability for that quartile.

Not as encouraging, however, were the results from the months of September and October. Climatologically, this is a transition period from the warm season to the cold season and thunderstorm activity begins to wane, especially toward the end of the period. For reasons not clearly understood, the percentage of lightning days to total days when the calculated probabilities were above 50% or for the top two quartiles was near 35%. This percentage is clearly lower compared to any of the other months for which lightning probabilities were calculated for the top two quartiles. Further, the September and October period contains the largest amount of total days in the historical database for any of the months in the study and the expectation is that more data should yield a more representative result for the entire study.

Results from Summer 2005 indicate that from a limited number of cases (20) in June and July the equations performed adequately. After dividing these cases into two groups depending on the ranked probability range, the upper half (>50% PLS) containing the highest values of PLS forecast lightning 88% of the time for 17 possible days. In the lower half of the ranked PLS values (<50%), there were no lightning days of the 3 possible days.

In conclusion, this study offers promise for the enhancement of traditional fire weather forecasts and the prediction of lightning compared to traditional methods utilizing the LAL. The final equations continue to be tested in real-time during the 2005 wildfire season. Initial real-time results are reassuring albeit they originate from a small data sample. This statistical tool increases the confidence of the meteorologist during the development of the lightning activity portion of the fire weather forecast. Future work is necessary to improve the robustness of the monthly equations by further predictor development. The calculated probabilities of a lightning strike are being used as a predictor for a model designed to estimate the probability of a lightning-caused wildfire occurring in the Black Hills National Forest.

4. ACKNOWLEDGEMENTS

I would like to thank Dr. Frank Matejcik from the department of Industrial Engineering at the South Dakota School of Mines and Technology for his help in the statistical equation development. Also, I would like to express my appreciation for the help of Matt Bunkers, Science and Operations Officer for the National Oceanic and Atmospheric Administration's National Weather Service office in Rapid City for supplying the radiosonde data and computer program to formulate the thermodynamic parameters used in this study. The position of fire meteorologist is funded by the State of South Dakota.

5. REFERENCES

Andrews K., and L. S. Bradshaw, 1997: "Fires: Fire information retrieval and evaluation system – a program for fire danger rating analysis." USDA Forest Service, Intermountain Research Station General Technical Report INT-367. 64 pp. (Ogden, UT)

Bolton, D., 1980: The computation of equivalent potential temperature. *Mon. Wea. Rev.,* **108,** 1046-1053.

Burrows, W. R., C. Price, and L. J. Wilson, 2004: Warm season lightning probability prediction for Canada and the northern United States, *Wea. Forecasting,* (submitted 29 June 2004).

Everitt, J. A., 1999: An improved thunderstorm forecast index for Cape Canaveral, Florida. M. S. Thesis, AFIT/GM/ENP/99M-06, Department of Engineering Physics, Air Force Institute of Technology, 98 pp. [Available from AFIT, Wright-Patterson Air Force Base, OH 45433].

Fuquay, D. M., 1980: Forecasting lightning activity level and associated weather. USDA Forest Service Research Paper INT-244, 30 p. Intermountain Forest and Range Experiment Station, Ogden, Utah 84401.

Lambert, W. C., M. Wheeler, and W. Roeder, 2005: Objective lightning forecasting at Kennedy Space Center and Cape Canaveral Air Force Station using cloud-to-ground lightning surveillance system data. *Conference on Meteorological Applications of Lightning Data*, 9-13 January, 2005, San Diego, California.

MacGorman, D. R., and W. D. Rust, 1998: *The Electrical Nature of Storms.* Oxford University Press, 422 pp.

Mazany, R. A., S. Businger, S. I. Gutman, and W. Roeder, 2002: A lightning prediction index that utilizes GPS integrated precipitable water vapor. *Wea. Forecasting,* **17,** 1034-1047.

Minitab, Incorporated, State College, PA, U.S.A.

NCDC and FSL, 1999: *Radiosonde Data of North America 1946-1999.* CD-ROM, Version 1.0 [Available from DOC/NOAA/OAR, Forecast Systems Laboratory, R/FSL, 325 Broadway, Boulder, CO 80305.]

Neumann, C. J., and J. R. Nicholson, 1972: Multivariate regression techniques applied to thunderstorm forecasting at the Kennedy Space Center. Preprints, *International Conf. Aerospace and Aeronautical Meteor.,* Washington D. C., Amer. Meteor. Soc., 6-13.

Pearson, K., 1900: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5, **50**, 157–175.

Reap, R. M., and D. S. Foster, 1979: Automated 12-36 hour forecasts of thunderstorms and severe local storms. *J. Appl. Meteor.,* **18**, 1304-1315.

Vaisala, Incorporated. Remote Sensing Division, Tucson, AZ, U.S.A.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.