3.11 STATISTICAL MODELS FOR LIGHTNING PREDICTION USING CANADIAN LIGHTNING DETECTION NETWROK OBSERVATIONS

William R. Burrows*
Atmospheric Environment Service, Downsview, Ontario, Canada

1. INTRODUCTION

Previously to February 1998, information on thunderstorm occurrence in Canada was available only from manned surface stations and a few provincial networks with limited coverage. Since February 1998 the Canadian Lightning Detection Network (CLDN) provides continuous lightning detection over all Canada to about 65°N in the west and 55°N in the east. Coverage is melded with the U.S. network to 35°N in eastern North America and about 40°N in western North America. CLDN flash data for 1998-2000 was analyzed by Burrows et. al. (2002) to understand the "climatological" characteristics of lightning occurrence in Canada and adjacent United States. An analysis of lightning detected for North America by the combined U. S. and Canadian networks for 1998-2000 was done by Orville et al. (2002). A complex pattern of lightning occurrence was revealed, showing strong latitudinal, seasonal, and diurnal dependencies, and significant influences by local elevated terrain features and major land-water boundaries.

At the time of writing work on designing and developing statistical lightning prediction models for the CLDN region is just underway by the author, Dr. Colin Price, and Patrick King. The ultimate goal is to develop algorithms that run diagnostically to give lightning prediction in real time using variables output by the Canadian Meteorological Center (CMC) GEM numerical weather prediction model (Cote et al., 1997).

2. METHOD

The lightning predictand is "lightning report density", formed by giving each flash report a weight of 1 if it is within 0-10 km of a grid point and a decreasing weight 1-0 if it is within 10-20 km. Currently we are working with a three-hour total lightning density at each GEM grid point for the eight 3-hour periods each day. The predictand is matched with potential predictors generated at GEM grid points from 00-hr to 12-hr GEM forecasts by 0000 UTC and 1200 UTC runs. Resolution is 22 km, going to 16 km later this year. Basic potential predictors are: CAPE, convective inhibition, convective stability indices (lifted, Showalter, sweat), severe storm index, storm relative helicity, elevation, convective cloud height, predicted flash rate (Price and Rind, 1992), tropopause height and temperature, precipitable water (total in column, and above 700 mb), geopotential height difference in 4 layers (500-1000 hPa, 700-1000 hPa, 850-1000 hPa, 700-850 hPa), rainfall (total and convective), maximum column wet bulb potential temperature, temperature (500 hPa, 700 hPa, 850 hPa), 700 hPa vertical motion, dew point (surface, 850 hPa, 700 hPa), geopotential height (500 hPa, 1000 hPa), binary land/water designation, and vegetation type designation. For each three-hour period the average, absolute maximum value, and 3-hour change of each potential predictor is found. Currently predictor fields are derived diagnostically from archived 00-hr, 6-hr and 12-hr GEM forecasts, so linear interpolation is used for each 3-hour interval. When the next version of GEM is implemented in 2002, two-dimensional fields such as CAPE, cloud liquid water content, and cloud ice content that are calculated in the convection scheme will be output as the model runs and archived at each time step (about 5 minutes). When we have developed a lightning prediction algorithm from these, the model will output and archive lightning prediction as well.

Price and Rind (1992) developed simple statistical relations for lightning flash frequency occurrence based on the electrical power of a cloud volume, which is proportional to the fifth power of the storm dimension:

$$F_c = 3.44 \times 10^{-5} H^{4.9}$$

$$F_m = 6.4 \times 10^{-4} H^{1.73} \qquad (1)$$

where F_c is the flash frequency in continental clouds (flashes/minute), F_m is the flash frequency in maritime clouds that are at least 500 km offshore, and H is the vertical cloud dimension. We plan to tune these relations for the area of the CLDN.

A problem in direct application of Eqn. (1) is that cloud convective processes are parameterized in the current GEM model. Thus there are no clouds per se, so convective cloud location and height must be inferred by the convection parameterization scheme at each grid point. This will be true for several years to come until computer power allows a model resolution of only a few kilometers over large areas and clouds can be directly forecast. Even then there is no guarantee that cloud forecasts will be correct. At this time there is no archived output from the convection scheme. We must infer the likelihood of lightning from other atmospheric variables and the lightning observations themselves and apply relationships such as Eqn. (1) in a conditional, or Bayesian, context, since it will always predict lightning if there is sufficient depth of cloud. Most lightning in Canadian latitudes is associated with fronts or other upper air features that produce vertical motion. We have already discovered that lightning occurs in only a small portion of convective clouds generated by active fronts, thus we expect the relationships we find will still be needed in the next version of GEM, even though we will know the location of convective cloud inferred by the convective parameterization scheme.

A few days with substantial lightning activity were chosen to begin the study. The predictand and predictors were matched for each 3-hour period at each grid point. For the 22 km GEM grid over the CLDN area this results in more than 50,000 records for each period. Since lightning occurs at only a small fraction of grid points the first stage of the statistical modeling procedure involves finding where lightning is *unlikely* to occur on any given day. For this the predictand is transformed to 0 where no lightning occurs and 1 where lightning occurs. The CLDN region was split at 100°W. Analysis is done with the decision-tree data mining procedure CART (Brieman et. al, 1984). An application of CART to another atmospheric problem is in Burrows (1997). The next stage is to study the conditions leading to various levels of lightning activity in the area identified by CART where lightning was observed, to identify the relevant predictors. At that point the flash density can be modelled with those predictors using a non-linear data modelling method such as a neural network or CANFIS (Burrows and Montpetit, 2000).

At the time of writing this study is at a very early stage. So far only the eastern region has been studied, and only 2 days examined thus far (May 12, 2000 and June 14, 2001). Based on the variables chosen by CART to split the data sets for these 2 days it appears that lightning is much less likely if the column precipitable water is less than 28 mm. Lightning is a very sensitive field. Even though we are only using GEM forecast data to 12 hours we find we must assess how to deal with frontal position error on some days. A probable strategy is to coordinate the study with satellite pictures and to use only those regions where the GEM model forecast of frontal position agrees fairly well the satellite pictures. Areas where agreement is not close are attributed to model error and are not used for algorithm calculation.

ACKNOWLEDGMENTS

Global Atmospherics, Inc. of Tucson, Arizona processes the North American Lightning Detection Network Canadian lightning data and provides telecommunications services to Environment Canada. The data are the property of Environment Canada. CART software is available from Salford Systems, 8880 Rio San Diego Drive, Suite 1045, San Diego, CA 92108; tel. 619-543-8880.

REFERENCES

Brieman, L., Friedman, J. H., Olshen, R. A., and C. J. Stone, 1984: *Classification and Regression Trees*. CRC Press, ISBN 0412048418, 358 pp.

Burrows, W. R., P. King, P. J. Lewis, B. Kochtubajda, B. Snyder, V. Turcotte, 2002: Lightning occurrence patterns over Canada and adjacent United States from lightning detection network observations. *Atmos.-Ocean*, (in press).

Burrows, W. R. and J. Montpetit: A procedure for neuro-fuzzy dynamic-statistical data modeling with predictor selection. *Preprints, Second AMS Conf. on Artificial Intelligence*, January 10-14, 2000, Long Beach, CA, 8-13.

Burrows, W. R., 1997: CART regression models for predicting UV radiation at the ground in the presence of cloud and other environmental factors. *J. Appl. Meteor.,* **36**, 531-544.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1997: The operational CMC/MRB Global Environmental Multiscale (GEM) model: Part I - Design considerations and formulation, *Mon. Wea. Rev.,* **126**, 1373-1395.

Price, C. P. and D. Rind, 1992: A simple lightning parameterization for calculating global lightning distributions. *J. Geophys. Res.*, **97-D9**, 9919-9933.

Orville, R. E., G. R. Huffines, W. R. Burrows, R. L. Holle, and K. L. Cummins, 2002: North American Lightning Detection Network (NALDN); First results: 1998-2000. *Mon. Wea. Rev.*. (In review: submitted Sep 05 2001).