# P2.7 DYNAMICAL-STATISTICAL MODELS FOR LIGHTNING PREDICTION TO 48-HR OVER CANADA AND THE UNITED STATES

William R. Burrows

Environment Canada - Science and Technology Branch -
Cloud Physics and Severe Weather Research Section, Toronto, ON
and
Hydrometeorology and Arctic Lab, Edmonton, AB

## 1. INTRODUCTION

The Canadian portion of the North American Lightning Detection Network (NALDN) has given continuous lightning detection since 1998 over all of Canada up to the middle north. This allowed determination for the first time of a spatially continuous climatology over this vast area. Complex patterns of lightning occurrence were revealed, with strong latitudinal and seasonal dependency and significant influences by topography and land-water boundaries (Burrows et al., 2002; Orville et al., 2002). Lightning is not specifically predicted in the Canadian numerical weather prediction model (GEM) (Coté et al., 1998) yet there is a requirement for thunderstorm prediction in public and aviation forecasts and fire weather prediction.

Dynamical- statistical models for real-time lightning prediction have run at the Canadian Meteorological Centre (CMC) since 2003 (Burrows et al., 2005). Since these models were developed new predictors from the GEM convection parameterization schemes became available, and archived GEM data became available hourly. New dynamical-statistical models were developed with 2005 data and began twice-daily prediction in April 2006. Increased computer capacity allowed for improvement in model design that enabled extension of the prediction area to include all of Canada and the United States except Hawaii, and extension to year-round predictions. An important feature of the new models is only a few days' training data are needed because the geographical region for training is very large. This means that model updates can be done frequently to keep abreast of changes in the driving NWP model. The intent of this paper is to describe the new models and provide some preliminary verification.

* *Corresponding author address*: William Burrows, Environment Canada - Hydrometeorology and Arctic Lab, Twin Atria Building - Room 200, 4999-98 Avenue, Edmonton, Alberta, T6B 2X3
email: william.burrows@ec.gc.ca

## 2. MODELING METHOD

### 2.1 Predictands and Predictors

The predictors, predictands, calculation methods, and model design are different from the original models described in Burrows et al. (2005). New predictors were added from the GEM parameterization schemes for deep convection (Kain and Fritsch, 1993) and shallow convection (Kuo, 1974), while the number of environmental predictors was decreased to include only those deemed to be most important in the original models. The new basic predictor set is shown in Table 1. The first column is a designated predictor symbol following the CMC naming convention where possible. All predictors are available hourly unless otherwise specified. Data for lightning predictands came from lightning flashes detected by the NALDN and received in Canada. On average more than 96% of detected flashes are cloud-ground. Only the portion of flashes north of 35°N to the east of 100°W and north of 40°N to the west of 100°W are received in Canada. Flashes per 3-hr were counted on a grid of 15 km resolution to match the GEM resolution.

Since the predictand is 3-hr lightning and most predictors are available hourly it was decided to use predictor data for t, $t_{+1hr}$, $t_{+2hr}$, $t_{+3hr}$ to cover a three-hour interval. When lightning flash counts were matched with the predictors it became evident that rarely does predictand data coincide exactly with predictor data even if we were modeling hourly lightning flashes with hourly predictor data from a forecast of a few hours. This is a common type of problem encountered in many environmental spatial statistical modeling endeavours. The problem is illustrated with Figure 1. The colored shaded areas are 3-hr lightning flash count, the black

| PRED | DESCRIPTION |
|---|---|
| | **Environment:** |
| SH | Showalter index of convection (°K) |
| LI | lifted index of convection (°K) |
| NCP | net convective available potential energy: (CAPE – convective inhibition (CIN)) (Joules $kg^{-1}$) |
| BH | lifted parcel cloud top, allowing entrainment ($10^3$ ft) |
| IH | precipitable water, surface to top of atmosphere (mm) |
| IY | precipitable water, 700 hPa to top of atmosphere (mm) |
| SW | severe weather threat index (SWEAT) |
| PN | mean sea level pressure (hPa) |
| HE | boundary layer helicity ($m^2s^{-2}$) |
| SI | CMC severe storm index (no units) |
| WW | 700 hPa vertical motion ($Pas^{-1}$) |
| TH | maximum wet bulb potential temperature ($\Theta_w$) in lowest 50 hPa (°K) |
| SHR | wind speed shear, ($\sigma$=.7161 – $\sigma$=.9958) (kt) where $\sigma \equiv p/p_{surface}$ , (p=pressure) |
| DH | three-hour change of (500-1000) hPa layer thickness (dam) |
| | **GEM Convection parameterization:** |
| RY | deep convection precipitation rate ($ms^{-1}$) |
| K6 | deep convection maximum updraft velocity ($ms^{-1}$) |
| K4 | deep convection cloud top (m) |
| U9 | deep convection cloud base (m) |
| DEP | deep convection cloud depth (K4 – U9) (m) |
| L8 | deep convection vertical integral of cloud-ice mixing fraction (kg/kg) |
| RZ | Shallow convection precipitation rate ($ms^{-1}$) |
| PC | accumulated 3-hr implicit precipitation (m) |

**Table 1**. Basic predictors for new lightning forecast models. First column shows the letter symbol assigned to a predictor described in column 2.

contours are maximum updraft velocity triggered by the Kain- Fritsch deep convection parameterization in the 12-hr GEM forecast, and the green contours are the same, triggered in the 15-hr GEM forecast. The overall locations of GEM convection are in the general neighborhood of observed lightning. However there are regions in Fig. 1 where GEM convection only partially coincides with the observed lightning, regions where convection was triggered near observed lightning but does not does not coincide with it, regions where lightning was observed but convection was not triggered, and regions where convection was triggered but no lightning was observed. NWP models are known to resolve meteorological details to about $8\Delta x$, where $\Delta x$ is the model grid spacing. In order to increase the chances of matching the predictand with meaningful values of predictors it was decided to smooth the predictors and to smooth the predictand as well. A 9 by 9 point smoothing grid was applied at each of the 4 forecast times in each 3-hr period ($t$, $t_{+1hr}$, $t_{+2hr}$, $t_{+3hr}$). This gave a cloud of 324 data points at every grid point with which to formulate new predictands and derive new predictors from the basic set in Table 1. Two predictands were defined: (1) the fraction of 324 points where lightning was observed, and (2) the average flash rate *for points where lightning was observed*. The first predictand can be called the "time-area coverage" and is similar to probability, the second predictand is the "average flash rate". They are referred to below by the acronyms LCHA and LFLS respectively.

Figure 2 illustrates these predictands. Consider an area of lightning moving by grid point (i,j) at forecast times $t$, $t_{+1hr}$, $t_{+2hr}$, $t_{+3hr}$, where lightning is observed at the red points. Lightning is observed at red points.. The total number of points is 324, of which 135 are red,

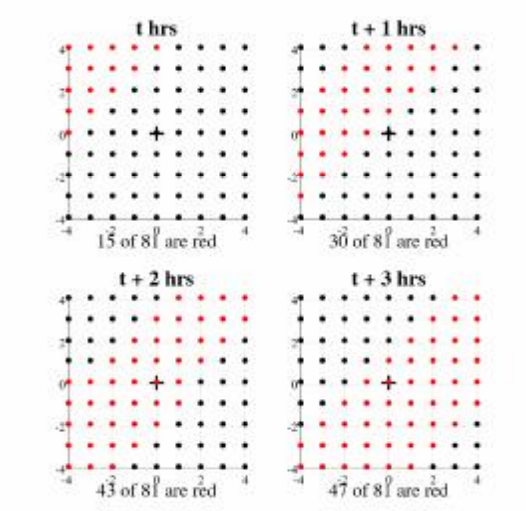



**Figure 2**. The 9*9 point smoothing grid surrounding grid point (i,j), indicated with "+", and applied at forecast times $t$, $t_{+1hr}$, $t_{+2hr}$, $t_{+3hr}$.

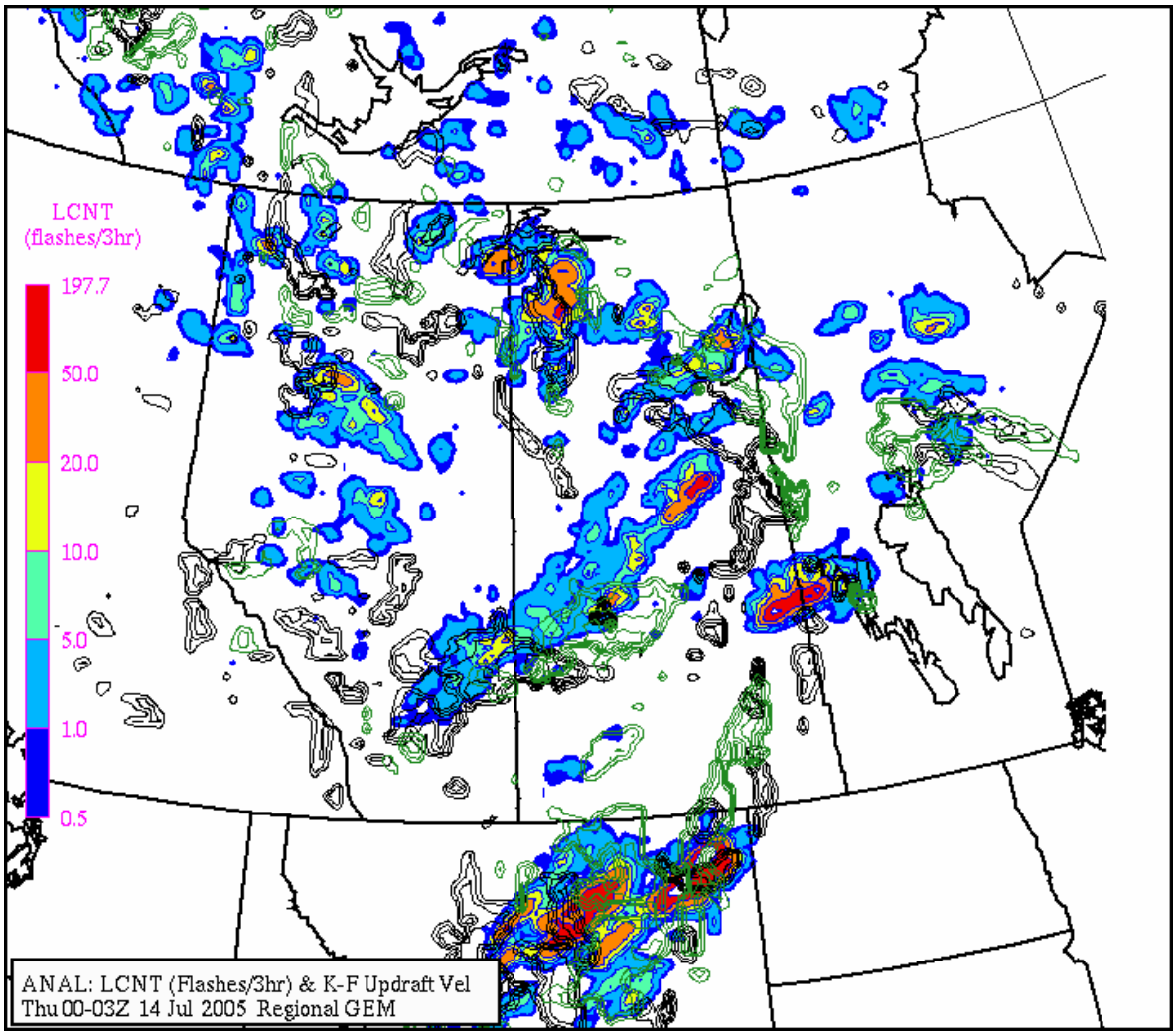



**Figure 1**. Flash count 0000-0300 UTC 14 July 2005 (indicated by colorbar). The 12-hr forecast of GEM deep convection maximum updraft velocity valid 0000 UTC 14 July 2005 is contoured in black, the 15-hr forecast valid 0300 UTC 14 July 2005 is contoured in green.

giving a time-area coverage of 135/324, or 0.4167. The flash rate predictand is defined as the average of total flash counts at red points. Figure 3 shows the time-area coverage predictand calculated for the same grid-point flash count shown in Fig. 1 with the same GEM deep convection updraft velocity fields overlain. Nearly all the GEM forecast convection areas are now associated with lightning occurrence. The effect of smoothing the flash count is evident, in that the area of lightning is spread out. The majority of GEM forecast deep convection fields are now covered by positive time-area coverage values, however maxima and minima do not coincide, or coincide partially. This shows that other predictors will be required to build a successful model to fit this data and the model is likely to be complex. Figure 4 shows the flash rate predictand overlain with MSL pressure at 0300 UTC 14 July 2005 to show the synoptic situation. The positive flash count regions seen in Fig. 1 are smoothed and spread out. The MSL pressure shows a general fact found during this study, namely that the vast majority of lightning occurs when the MSL pressure is less than 1025 hPa. Physically when MSL pressure is higher than this the atmospheric dynamics are unlikely to support convection because deep convective instability is unlikely and vertical motion will be primarily downwards, suppressing any convection that forms.

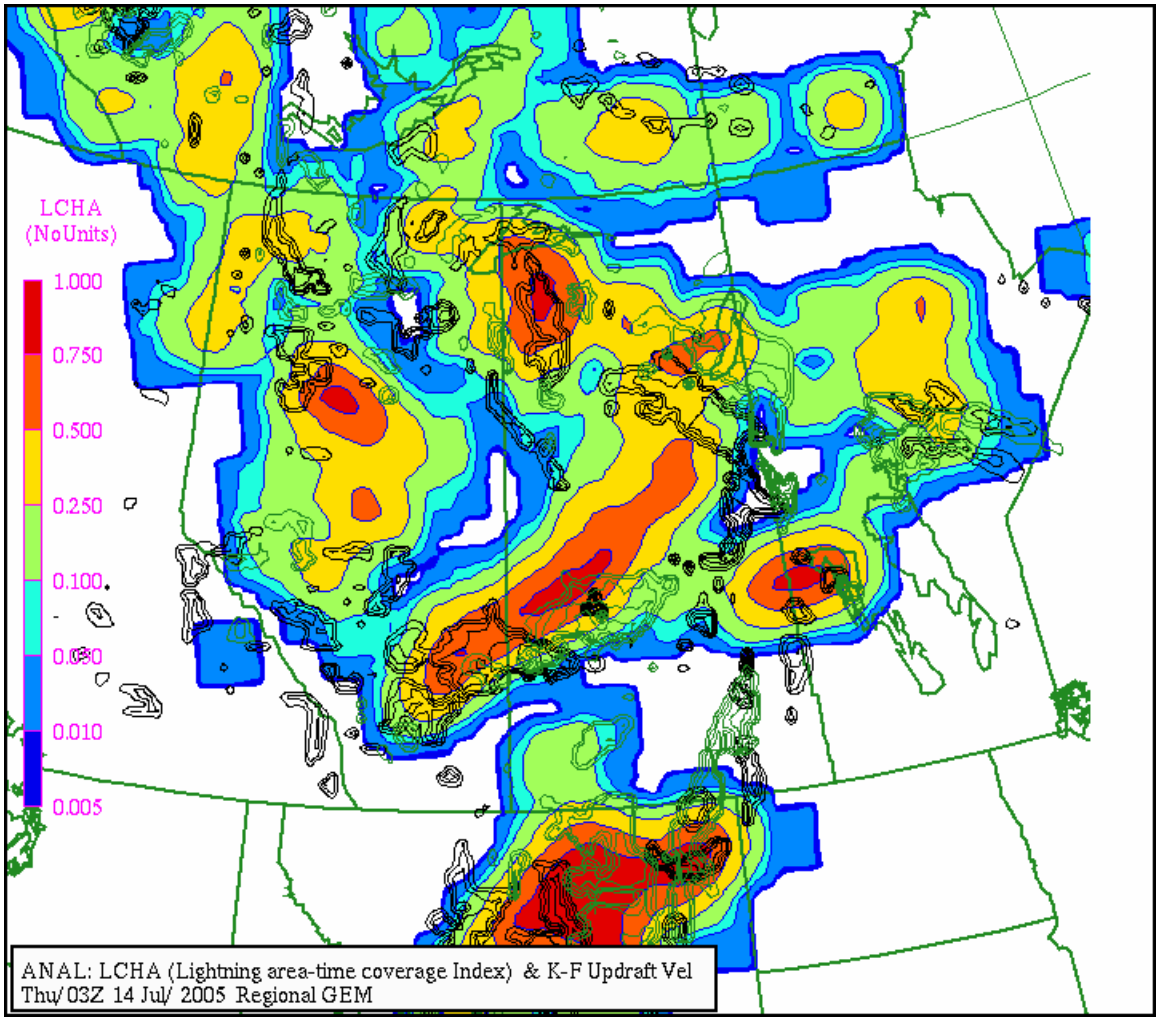



**Figure 3**. The time-area coverage predictand formed from the flash count data in Fig. 1 (shaded), with the same GEM Kain-Fritsch deep convection areas overlain.

Smoothing the predictands over the 324 point data cloud suggested forming new predictors from statistics of the basic predictors shown in Table 1. Table 2 shows new predictors formed from the basic predictors shown in Table 1. Mean, maximum, (or minimum in some cases) were calculated over all 324 points and/or over only those points where a condition is true, e.g. the mean Showalter index calculated for those points where it is negative.

### 2.2 Training Data

The area of forecast interest covers all of the USA and Canada and offshore for a considerable distance. However the region where lightning is available in Canada is much smaller. Figure 5 shows detection efficiency for lightning on the NALDN. Lightning detected north of 35°N east of 100°W, and north of 40°N west of 100°W is available in Canada (shown as solid black lines in Fig. 5). The region within roughly the 80% detection efficiency contour for data available in Canada was chosen for gathering training data for dynamical – statistical lightning prediction models. Alaska was not included in the training data because lightning detected there became available in Canada in 2006.

The region where training data were available covers a large geographical region where several meteorological situations exist on any given day. Because of this only a few days were needed for training the models. One day per month was selected from March to September 2005. Dates varied somewhat for models built from 00 UTC and 12 UTC data due to data availability. Dates are, for models built from 00 UTC GEM runs: 23 March , 22 April, 20 May, 6 June, 26 July, 26 August, 20 September; and for

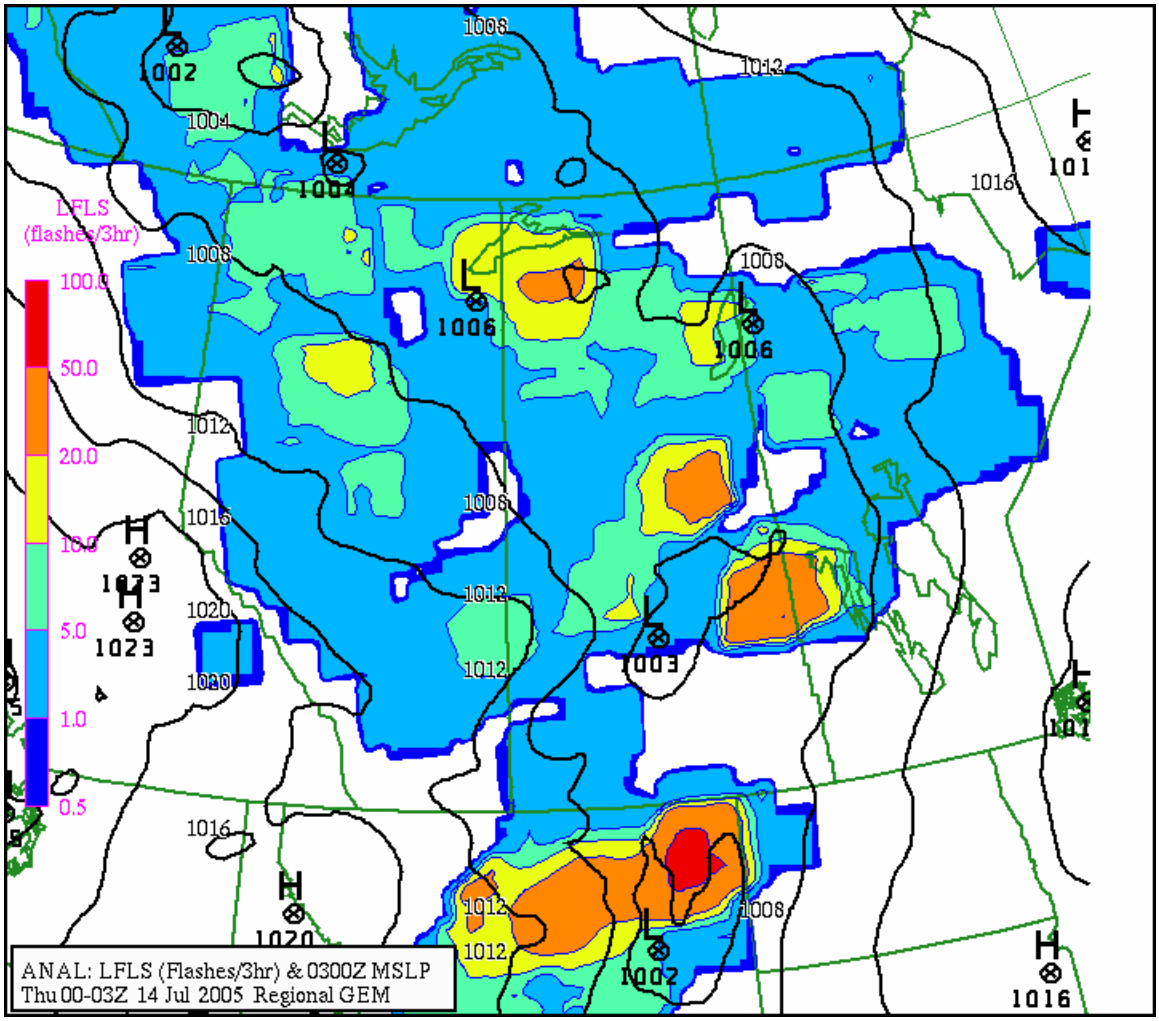



**Figure 4.** The flash rate predictand formed from the flash count data in Fig. 1 (shaded), and MSL pressure (hPa) contours at 0300 UTC 14 July 2005 (solid black lines).

models built from 12 UTC GEM runs: 22 March, 21 April, 19 May, 9 June, 25 July, 26 August, 19 September. These dates were selected because each had considerable lightning activity.

**2.3 Model Production**

To forecast lightning in real time models covering a 24-hr diurnal period are built from a training data consisting of archived model output and predictand data, then the models are applied in real time with NWP output covering 0-24 hours and 24-48 hours, using the same models in both periods. This is a variation on the time-offset MOS method put forward in Burrows (1985). GEM suffers from model spin-up error for the first few hours, so 1-5 hr forecasts were not used. 6-18 hr forecasts were used for all models. Lightning forecast models valid 0600-0900, 0900-1200, 1200-1500, and 1500-1800 UTC were built with forecasts with GEM 0000 UTC runs, and models valid 2100-2400, 0000-0300, 0300-0600, 0600-0900, and 0900-1200 UTC were built with 1200 UTC GEM runs.

The predictands and predictors were calculated for the geographical region inside approximately the NALDN 80% detection efficiency contour in Fig. 5 at all grid points for the 7 training days using the 15 km GEM grid space resolution. This resulted in about 766,000 records of matched predictand – predictor data for training models. Data was eliminated at a grid point if, over the 324 data point cloud, both the mean of the Showalter Index was 5°K or greater, and the mean of MSL pressure was 1025 hPa or greater, since lightning is very unlikely under these conditions. This left training data sets ranging in size from about 228,000 cases in the 2100-2400 UTC time period to 168,000 cases

| PRED | DESCRIPTION |
|---|---|
| | **Environmental Predictors:** |
| SH | mean , min SH at points where SH < 0 °K;<br>fraction of points where SH < (1, 0, -1, -2, -4, -6, -8, -10) °K |
| LI | mean, min LI over all points;<br>fraction of points where LI < (2, 1, 0, -1, -2, -4, -6, -8) °K |
| ECP | mean, max ECP at points where ECP > 0 $Jkg^{-1}$;<br>fraction of points where ECP > (0, 500, 1000, 1500, 2000, 3000, 4000) $Jkg^{-1}$ |
| BH | mean, max BH of points where BH > 0;<br>fraction of points where BH > (0, 20, 30, 40, 50, 60 ) * $10^3$ ft |
| IH | mean, max IH over all points;<br>fraction of points where IH > (10, 15, 20, 25, 30, 40, 50) mm |
| IY | mean, max IY over all points;<br>fraction of points where IY > (5, 10, 15, 20, 25) mm |
| SW | mean, max SW at points where SW > 0;<br>fraction of points where SW < 50;<br>fraction of points where SW > (50, 100, 200, 300, 400) |
| PN | mean, min PN over all points ;<br>fraction of points where PN < (1020, 1015, 1010, 1005, 1000, 995, 990) hPa |
| HE | mean , max HE at points where HE >0; mean, min at points where HE < 0;<br>fraction of points where HE < (-100, -200) $m^2s^{-2}$;<br>fraction of points where HE > (100, 200, 400, 600) $m^2s^{-2}$ |
| TH | mean , max TH over all points;<br>fraction of points where TH > (280, 285, 290, 295) °K |
| WW | mean, min WW at points where WW < 0;<br>fraction of points where WW < (0, -.25, -.50, -.75, -1.0, -1.5, -2.0, -3.0) $Pas^{-1}$ |
| SI | mean, max at points where SI > 0;<br>fraction of points where SI > (50, 100, 125, 150, 175);<br>fraction of points where SI < 50 |
| SHR | mean, max SHR over all points;<br>fraction of points where SHR > (10, 20, 30, 50, 75) kt |
| DH | max, min DH over all points; mean where DH <0;<br>fraction of points DH < (0, -2, -5, -10, -15) dam;<br>fraction of points where DH > (0, 2) dam |
| | **GEM Convection parameterization:** |
| RY | mean, max RY at points where RY > 0;<br>fraction of points where RY > (0, 0.1, 0.25, 0.5, 0.75, 1.0, 2.0) * $10^{-6}$ $ms^{-1}$ |
| K6 | mean, max K6 at points where K6 > 0;<br>fraction of points where K6 > (5, 10, 15, 20, 25, 30, 35) $ms^{-1}$ |
| K4 | mean, max K4 at points where K4 > 0;<br>fraction of points where K4 > (2000, 4000, 6000, 8000, 10000, 12000, 14000) m |
| U9 | mean, max U9 at points where U9 > 0;<br>fraction of points where U9 > (500, 1000, 1500, 2000, 2500, 3000) m |
| DEP | mean, max DEP at points where DEP > 0;<br>fraction of points where DEP > (0, 2000, 4000, 6000, 8000, 10000, 12000, 14000) m |
| L8 | mean, max L8 at points where L8 > 0; fraction of points where L8 > (1, 2, 4, 6, 8) $kgkg^{-1}$ |
| RZ | mean, max RZ at points where RZ > 0;<br>fraction of points where RZ > (0, 0.1, 0.2, 0.25, 0.3) * $10^{-6}$ $ms^{-1}$ |
| PC | mean, max PC at points where PC> 0; fraction of points where<br>PC > (0, 0.1, 0.5, 1.0, 5.0, 10.0) mm |

**Table 2.** Predictors derived from the basic predictors shown in Table 1, to be applied to the 324 point smoothing data cloud at each grid point. “max” stands for maximum, “min” for minimum.

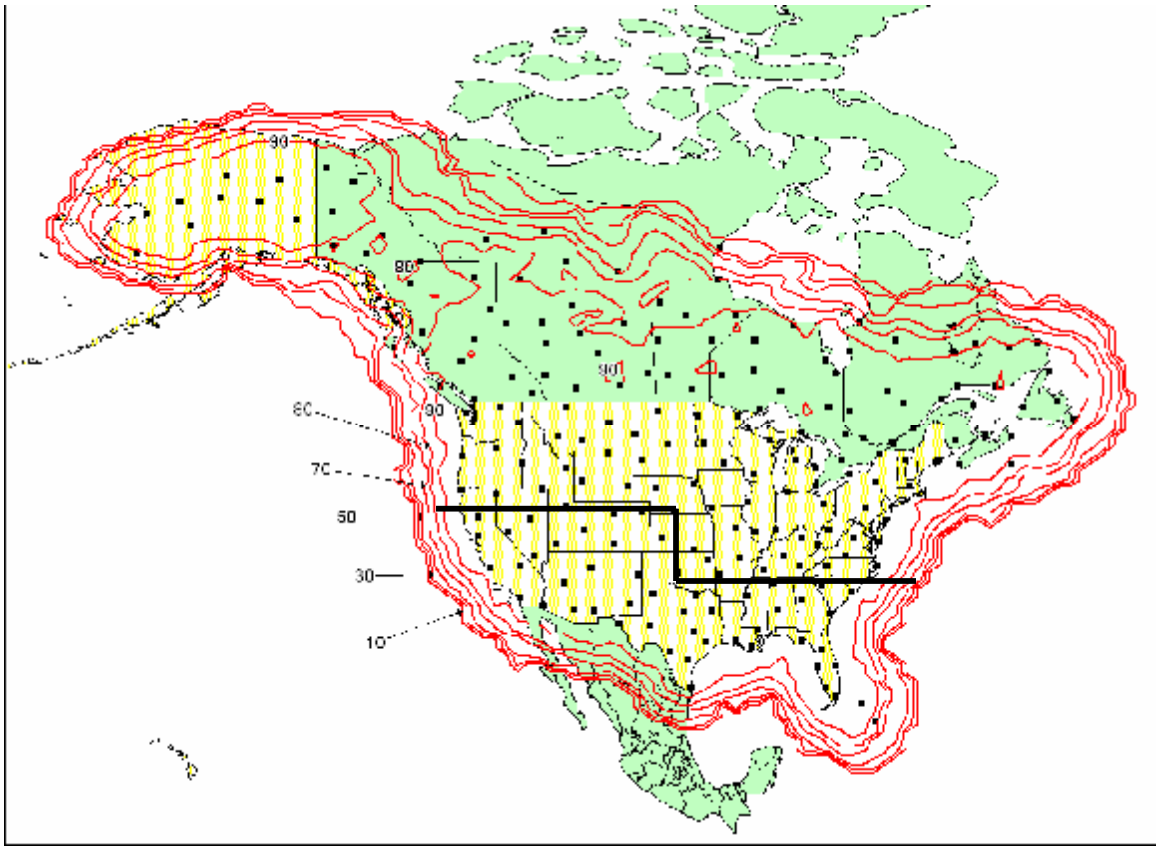



**Figure 5.** Average lightning detection efficiency (%) for the NALDN for 2006 up to 16 August 2006. Lightning detected in the region north of the solid black lines is available in Canadian weather offices.

in the 1200-1500 UTC time period. Models were generated for both predictands for each three-hour time segment, giving a total of 16 models.

There are 189 potential predictors in Table 2. To reduce data size for each model generation run, predictors with correlation .2 or less with the predictand in the whole dataset were eliminated. Of the surviving predictors, groups of predictors correlated more than .9 or more with each other were identified, and only the predictor correlated highest with the predictand was kept. To check that predictors that might be needed to fit predictand values greater than the 95 percentile value were not eliminated, the same data reduction procedure was repeated for the set of cases where the predictand value was equal to or higher than its 95 percentile value. Any predictors eliminated when the procedure was applied to the whole data set were added back in to the final predictor set offered to the modeling algorithm. These procedures reduced the number of predictors to less than 49 for all models, and to 30-40 for 11 of the 16 models.

**2.3.1 Tree Structured Regression**

Models were built with tree structured regression (Venables and Ripley, 2002), also known as CART (Brieman et al., 1984). A detailed theoretical discussion can be found in Brieman et al. (1984). A brief heuristic explanation of how tree-structured regression works follows, with important concepts in italics.

Essentially a tree-structure is found that recursively partitions the training data into subsets of cases based on threshold predictor values, thus *partitioning the data with a lattice of intersecting planes in feature space*. The tree consists of a recursive series of node branches where the data in each internal node is split into left and right child nodes based on the predictor value that gives *maximum reduction of predictand error* after the split. Node splitting continues until predictand error cannot be further reduced or until a user-specified minimum allowable *complexity parameter* $\alpha$ is reached. $\alpha$ can be thought of as a tree complexity cost per *terminal node*. Terminal nodes are nodes which are not split further. Then the tree is pruned from

the bottom up by sequentially removing weakest links (effectively, node branches whose contribution to predictand error reduction is least).

The "*prediction*" assigned to a terminal node is the mean value of the predictand in cases residing in the terminal node, although other definitions are possible. After each newly pruned structure in the tree sequence is derived its value $\alpha_T$ *is noted and* N-fold *cross-validation relative error* is calculated in order to estimate the *relative error* that would occur if that tree were applied to independent data. *Relative error* is defined as the ratio: error when the predictand is fit by the tree structure corresponding to $\alpha_T$ divided by error when the predictand is fit by its mean value (i.e. the initial variance). Cross validation error is found by ordering the training data by predictand value, randomly dividing the ordered training data £ into N subsets of approximately equal size $£_1$, $£_2$, …, $£_N$, repeating the same tree growing and pruning procedure using $\alpha_T$ for each of the N remainder data sets £ - $£_n$ (n=1,2,..N), calculating the error of each fit, and averaging the N errors. A common specification for N is 10. *The "best" tree structure is one whose relative error is least or is within one standard error of the minimum cross-validated relative error.*

When the tree is applied to an independent case, the case will run down the tree according to its predictor values until it reaches a terminal node. The set of values in the terminal nodes of the tree is a piecewise-continuous fit of the data. *When a tree is large the fit to the training data becomes quasi-continuous.* Predictions on independent data using a large tree are likely to be as low or lower in error compared to predictions by other methods that generate a continuous fit to the data because error reduction is optimal and all relevant predictors are used.

For continuous predictands such as wind where a fully continuous prediction is desired tree-structured regression can be used to *select relevant predictors from a larger set of potential predictors* in the training data, then a model giving continuous prediction can be built with a second method such as a neural network or support vector machines. Examples of this approach can be seen in Burrows (1998), Faucher at al. (1999), and Walmsley et al. (1999).

**2.2.2 Model Selection**

Figure 6 shows a tree-structured regression model for the time-area coverage predictand (LCHA) for the diurnal period 2100-2400 UTC. Predictors found for node-splitting decisions in Fig. 6 are defined in Table 3.

| PRED | DESCRIPTION |
|---|---|
| RY0 | maximum RY over all points ($ms^{-1}$) |
| SH2 | mean SH over all points (°K) |
| K64 | fraction of points where K6 > 20 $ms^{-1}$ |
| IH0 | minimum IH over all points (mm) |
| SW9 | mean SW over all points |
| IY2 | fraction of points where IY > 10 mm |
| SR9 | mean SHR at points where SHR > 0 (kt) |

**Table 3**. Definitions of predictor acronyms shown in Fig. 6.

Tree growth was limited for this example by setting a high complexity parameter value of .01. 228230 cases reside in the training data with an average predictand value of 0.054 represented in the initial internal node. The data is initially split by RY (deep convection rain rate). *The direction of the splits make physical sense*. 4124 cases where the maximum deep convection rain rate is at least 5.05 * $10^{-7}$ $ms^{-1}$ (or about 1.8 mm/hr) go to the right branch while the remainder go to the left branch. The cases sent right have an average predictand value of 0.21. These cases are very near or overlap deep convection triggered in GEM. The 186806 cases going left are likely those where deep convection triggered in GEM was either weak or too far away to coincide with observed lightning. Indeed, the second leftmost split there occurs with Showalter Index, which is an environment predictor. Following along the series of rightmost splits, the second rightmost split sends to the right 7199 cases where the fraction of the 324 point smoothing data cloud with deep convection maximum updraft velocity greater than 20 $ms^{-1}$ is at least.06327. The mean time-area coverage of these cases is 0.44. The third rightmost split sends 4477 cases to the right whose mean SWEAT index over the 324 point smoothing data cloud is 289.9 or greater. The predictand mean for these cases is 0.54. The fourth rightmost split sends 2354 cases to the right whose fraction of the 324 point smoothing data cloud with minimum IY greater than 10 mm is 0.1003. The mean time-area coverage for these cases is 0.64.

The cross-validated relative error of the tree in Fig. 6 is .60 but the tree-structured regression algorithm is capable of reducing the relative

error much lower for this predictand. The tree was allowed to grow much larger by setting the minimum allowable complexity parameter α at .0001. Figure 7 shows cross-validated error versus the number of terminal nodes. α is shown on the bottom horizontal axis. As the number of terminal nodes increases the cross-validated error decreases sharply at first then the rate of decrease becomes less and is nearly zero when the complexity parameter of .0001 is reached. If the tree were allowed to grow larger the cross-validated error would eventually begin to increase again as the predictand becomes over-fitted. The zone of low cross-validated error is unusually broad in this example, meaning that several tree structures would suffice as a tree to use and would deliver comparable results. (In the majority of problems this author has worked on the zone of minimum error is not broad). The tree selected had 712 terminal nodes, fit the training data with relative error .112, and had cross-validated error of .151.

Figure 8 shows a tree-structured regression model for the three-hour flash rate predictand (LFLS) for the diurnal period 2100-2400 UTC. Tree growth was limited for this example by setting the complexity parameter value to .01. Predictors found for node-splitting decisions are defined in Table 4. The deep convection rain rate predictor RY is also the first predictor chosen for this predictand. This reiterates the not surprising view that predictors formed from the GEM deep convective parameterization scheme are indeed important for lightning prediction. However, as shown in Fig.1 the deep convection regions triggered in GEM do not coincide exactly with observed convection, thus other information is needed to quantify lightning flash rate. The remaining predictors in Fig. 8 are environmental predictors known to be important for thunderstorm production. 228230 cases reside in the initial internal node with a 3-hr flash rate of 3.2 averaged over all the cases. A relatively large group of 221590 cases with RY6 less than .03549 goes left. Of these cases, 35617 whose 324-point average lifted parcel top (BH9) was greater than 41.18 thousand feet were sent right in the next split. The case-average 3-hr flash rate for these cases was 8.7, considerably higher than the 0.95 case-average flash rate of the 185973 cases sent left whose BH9 value was less than 41.18 thousand feet. The tree in Fig. 8 has 6 terminal nodes and a cross-validated relative error of .55. Reducing the complexity parameter to .0001 produced a tree with 396 nodes and cross-validated relative error of .157 which was used as final model. Its error reduction curve had the same appearance as Fig.7.

| PRED | DESCRIPTION |
|---|---|
| RY6 | fraction of points where RY > $1.0*10^{-6}$ $ms^{-1}$ |
| BH9 | mean BH for points where BH > 0 |
| TH0 | maximum TH for all points (°K) |
| IH6 | fraction of points where IH > 40 mm |

**Table 4.** Predictor acronyms in Fig. 8.

Table 5 shows the resubstitution relative error, cross-validated relative error, and number of nodes for the final trees selected for all 8 three-hour diurnal periods. All trees were built with the specification $\alpha$ =.0001. ***Resubstitution relative error*** is found by running all the training cases down the tree which was developed from them and calculating the error of the tree's fit to the predictand, and is always lower than cross-validated relative error. Overall the trees with lowest error were for the period 0300 UTC to 1200 UTC and the trees with highest error were for the period 1500 UTC to 2100 UTC.

| | **LCHA** | | | **LFLS** | | |
|---|---|---|---|---|---|---|
| Period | RSerr | XVerr | Nodes | RSerr | XVerr | Nodes |
| 0003 | .112 | .143 | 664 | .119 | .168 | 440 |
| 0306 | .076 | .106 | 428 | .119 | .180 | 408 |
| 0609 | .073 | .104 | 462 | .105 | .154 | 414 |
| 0912 | .083 | .120 | 450 | .096 | .136 | 324 |
| 1215 | .088 | .129 | 433 | .121 | .194 | 295 |
| 1518 | .127 | .185 | 491 | .118 | .210 | 288 |
| 1821 | .129 | .180 | 729 | .139 | .195 | 448 |
| 2124 | .112 | .151 | 712 | .112 | .157 | 396 |

**Table 5**. Column 1: 3-hr time period (e.g. 0003 is 0000 to 0300 UTC). For the LCHA and LFLS predictands respectively: columns 2 and 5: resubstitution relative error (RSerr); columns 3 and 6: cross-validated relative error (XVerr); columns 4 and 7: number of terminal nodes in tree.

The importance of predictors chosen to split internal nodes in the final trees were ranked on a scale of 0-100 by their contribution to reduction of total error. Table 6 shows the top 10 predictors for the LCHA tree for the 2100-2400 UTC period. The deep convection rain rate and fraction of 324 points with updraft vertical velocity greater than 20 ms-1 were the first two predictors, followed by the total column precipitable water. Table 7 shows the same

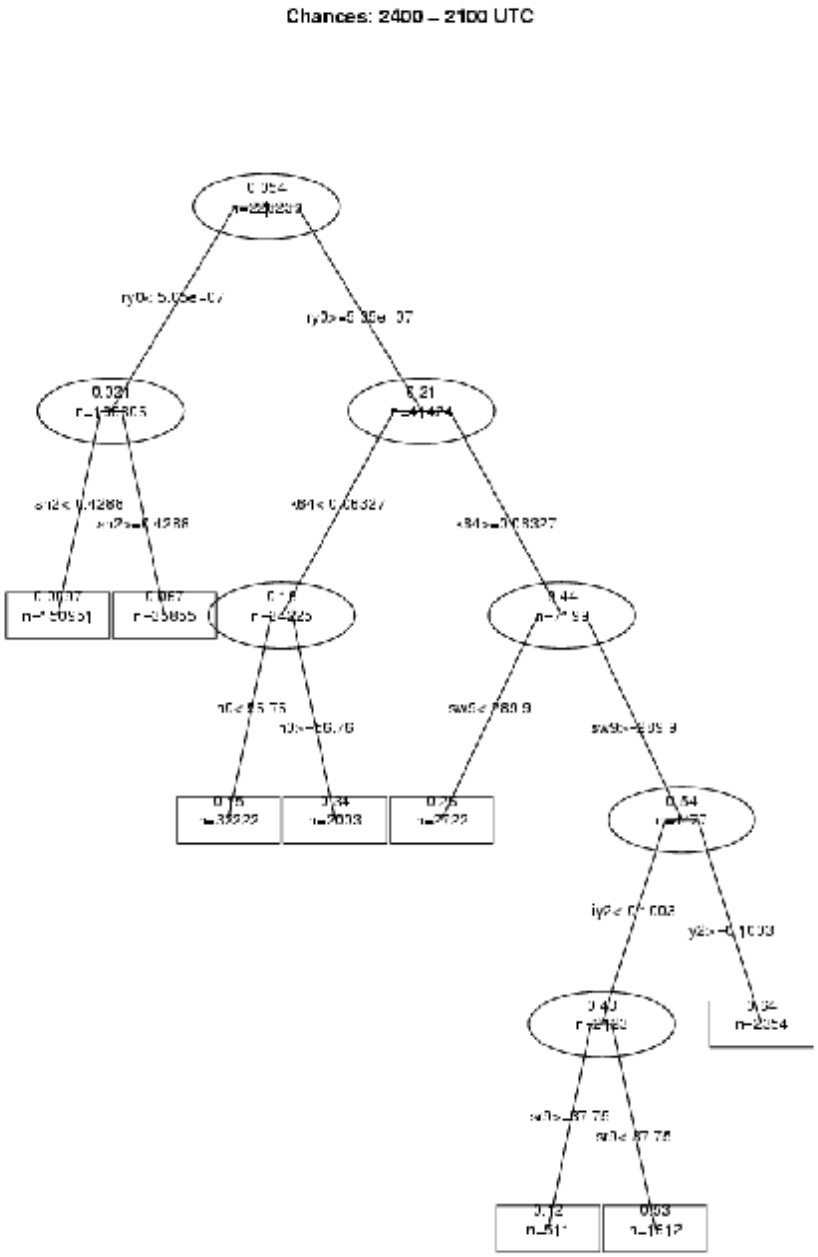



**Figure 6.** Tree-structured model with complexity parameter $\alpha = .01$ for the lightning time-area coverage (LCHA) predictand, valid for 2100-2400 UTC. A circle denotes an internal node, a square denotes a terminal node. Inside each circle or square, the number of cases in the node is the lower number, e.g. n=228230 in the root node; the mean predictand value of these cases is the upper number, e.g. 0.054 in the root node. Left and right child nodes descend from each internal node and are connected to it with solid lines. The node-splitting predictor and its value used to split a node is shown below each internal node.

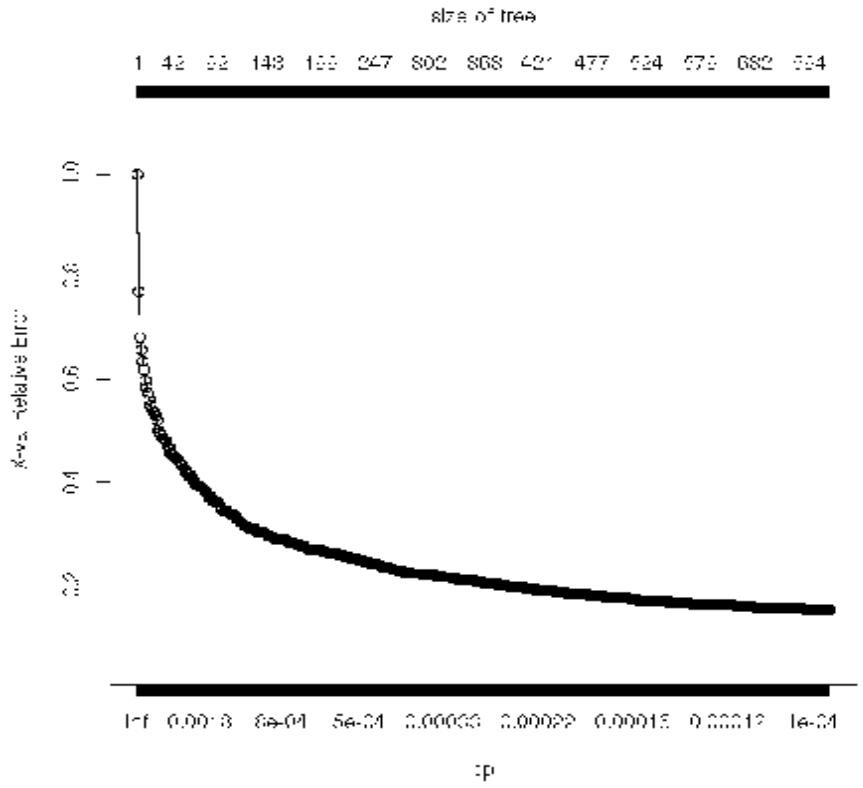


**Figure 7.** The cross-validated error versus number of terminal nodes in the sequence of tree structures for the lightning time-area coverage predictand, valid for 2100-2400 UTC as the tree is pruned upward from a complexity parameter of .00010.

| PRED | DESCRIPTION | RANK |
|---|---|---|
| RY0 | maximum RY at points where RY > 0 | 100.0 |
| K64 | fraction of points where K6 > 20 $ms^{-1}$ | 37.2 |
| IH0 | maximum IH | 21.0 |
| SR9 | mean SR where SR >0 $s^{-1}$ | 18.3 |
| DH9 | minimum DH | 13.5 |
| SH0 | minimum SH where SH < 0 | 13.1 |
| SW9 | mean SW at points where SW > 0 | 12.7 |
| SH2 | fraction of points where SH < 0 °C | 11.8 |
| U90 | maximum U9 where u9>0 | 9.9 |
| SR4 | fraction of points where SR > 20 $ms^{-1}$ | 9.7 |

**Table 6.** Top 10 predictors in tree for LCHA for the period 2100-2400 UTC, ranked on a scale of 0-100 according to their contribution to the total reduction of error.

| PRED | DESCRIPTION | RANK |
|---|---|---|
| RY6 | fraction of points where RY > $1.0*10^{-6}$ $ms^{-1}$ | 100.0 |
| BH9 | mean BH for points where BH > 0 | 59.6 |
| IH6 | fraction of points where IH > 40 mm | 58.9 |
| TH0 | maximum TH | 58.5 |
| SW0 | maximum SW | 32.7 |
| IH0 | maximum IH | 24.2 |
| WW0 | maximum WW | 18.9 |
| U90 | maximum U9 | 17.3 |
| SH0 | minimum SH at points where SH < 0 °C | 14.4 |
| IY0 | maximum IY | 13.4 |

**Table 7.** Top 10 predictors in tree for LFLS for the period 2100-2400 UTC, ranked on a scale of 0-100 according to their contribution to the total reduction of error.

| LCHA PERIOD | PRED | DESCRIPTON |
|---|---|---|
| 0300 | K69 | mean K6 at points where K6 > 0 |
| 0603 | K69 | ditto |
| 0906 | SW0 | maximum SW |
| 1209 | SW4 | fraction of points where SW > 300 |
| 1512 | SW9 | mean SW at points where SW > 0 |
| 1815 | K69 | mean K6 at points where K6 > 0 |
| 2118 | K60 | maximum K6 |
| 2421 | RY0 | maximum RY at points where RY > 0 |
| | | |
| LFLS PERIOD | PRED | DESCRIPTION |
| 0300 | RY9 | mean RY at points where RY > 0 |
| 0603 | K69 | mean K6 at points where K6 > 0 |
| 0906 | SW0 | maximum SW |
| 1209 | SW4 | fraction of points where SW > 300 |
| 1512 | PN0 | minimum PN |
| 1815 | PN0 | ditto |
| 2118 | CD6 | fraction of points with cloud depth > 10000 ft |
| 2421 | RY6 | fraction of points where RY > $1.0*10^{-6}$ $ms^{-1}$ |

**Table 8.** Top ranked predictor for LCHA and LFLS for all diurnal periods.

analysis for the LFLS tree for the 2100-2400 UTC period. The top predictors are fraction of points with deep convection rain rate greater than $1.0*10^{-6}$ $ms^{-1}$and lifted parcel cloud top followed by fraction of points with total precipitable water greater than 40 mm. Table 8 shows the top ranked predictor for all three hour periods for LCHA and LFLS. Deep convection predictors dominate in diurnal times of greater convective activity while environmental predictors dominate in diurnal times of less convective activity.

## 3. VERIFICATION

The models have run in real time twice daily since April 2006 and provide predictions in three-hour intervals from 6-9 hrs to 45-48 hrs. Forecasts are available to all Canadian forecast offices via an internal website. A example of the forecasts follows.

Figure 9 shows observed 1-hour lightning flash rates for the entire North American Lightning Network for 2100 UTC 01 August 2006 to 0000 UTC 02 August 2006. The observed LCHA predictand for the same period is shown in Figure 10 for the portion of the NALDN reports received by Environment Canada (described above). Figures 11 and 12 show the 9-12 hr forecasts of LCHA and LFLS, respectively, for the same period. Comparing these with the observations in Figs 9 and 10 shows the forecasts worked out well overall, except for a tendency to over forecast LCHA in the southeastern USA north of the Gulf coast and south of the large front in mid-continent, and over the Florida peninsula. However, in the former region forecast LCHA values were small and the LFLS values were below .5 flashes per 6 hr. Intensive lightning in the large frontal zone extending from Texas to Quebec and in the Gulf coast region is well forecast in Fig. 12, as are the regions of lower lightning activity over western Canada, the Yukon, Alaska, and central Quebec.

The LCHA forecast in Fig. 11 includes all values down to .005, which is only 1 or 2 points out of the 324 point data cloud at each grid point. This is very low and may be actually in the noise range. Even LCHA = .01 is only about 3 points out of the 324. This corresponds to an area of "popcorn towering cumulus" where an isolated cumulus cloud grows a bit larger than the others and produces occasional lightning. Relatively high values of LCHA represent either widespread thunderstorm activity or slow moving systems. Relatively low values of LCHA represent scattered to isolated thunderstorms, or possibly very fast moving systems, although the latter is less likely. Looking at Figs. 11 and 12 it is natural to ask what the LFLS forecasts are for various threshold values of LCHA. Figure 13 shows the LFLS forecast in Fig. 12 for areas where LCHA >= .01, and Figure 14 shows the forecast in Fig. 12 for areas where LCHA >= .05. Isolated areas of lightning seen over the north in Fig. 9 and 10 are largely filtered out at the LCHA = .01 level but elsewhere over Canada and the northern half of the USA the LFLS forecast filtered by LCHA >= .01 is in overall good agreement with the observed lightning in Figs 9 and 10. Over the southern half of the USA the LFLS forecast filtered by LCHA >= .05 is in generally good agreement with observed

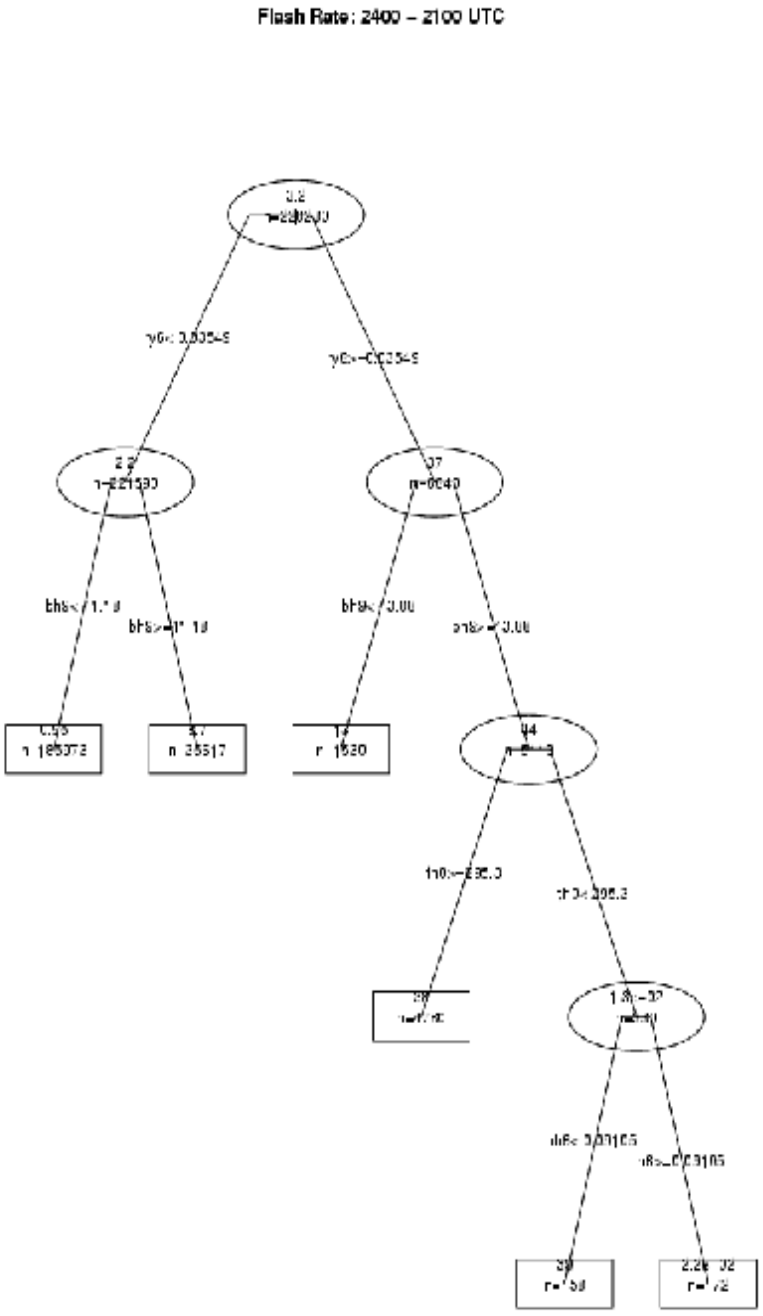



**Figure 8.** Tree structured regression model with complexity parameter set at .01 for the three-hour flash rate predictand valid 2100-2400 UTC. See Fig. 6 for explanation.

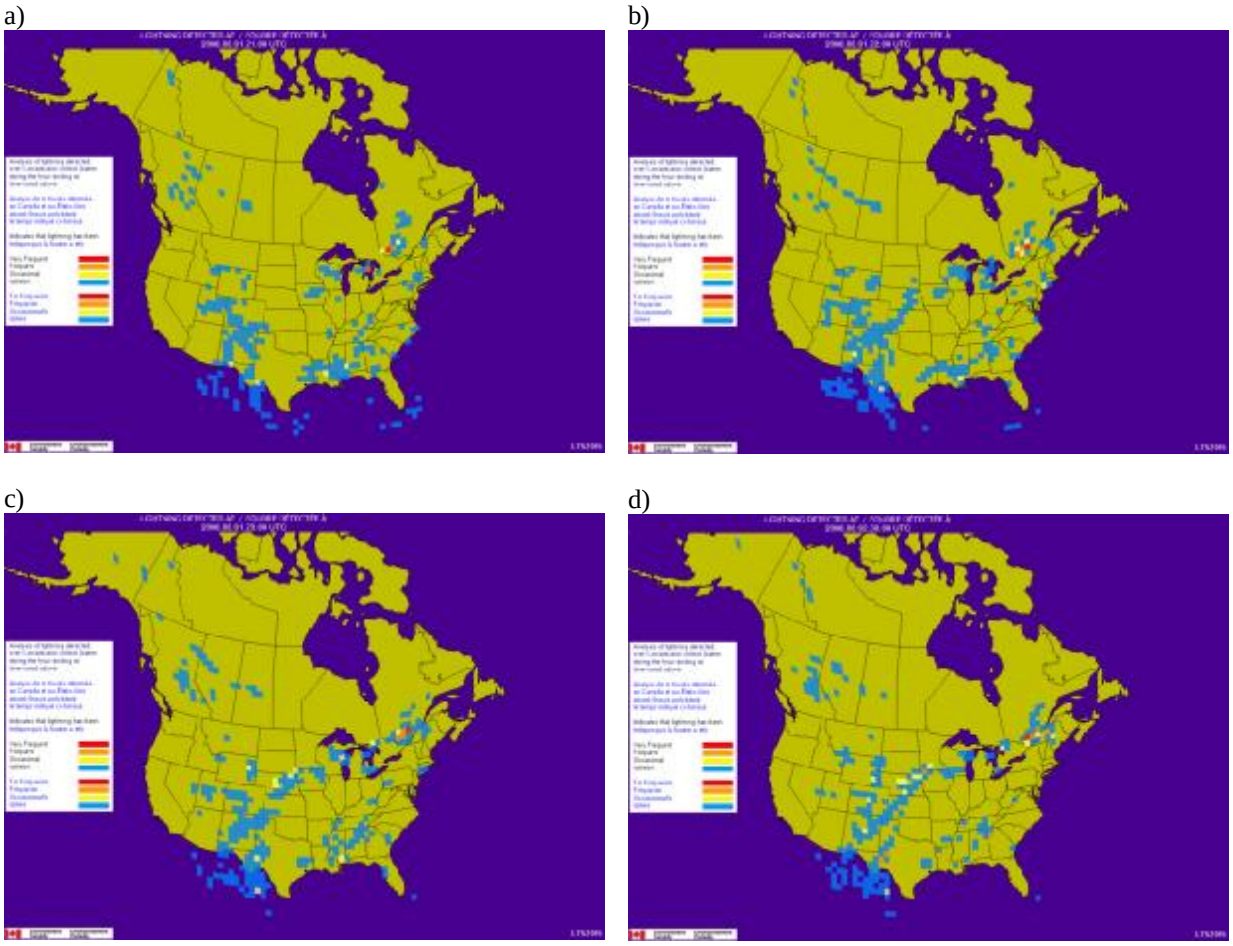



**Figure 9**. Observed 1-hr lightning flashes for a) 2100 UTC 01 August 2006; b) 2200 UTC 01 August 2006; c) 2300 UTC 01 August 2006; d) 0000 UTC 02 August 2006. Images courtesy of Vaisala Inc.

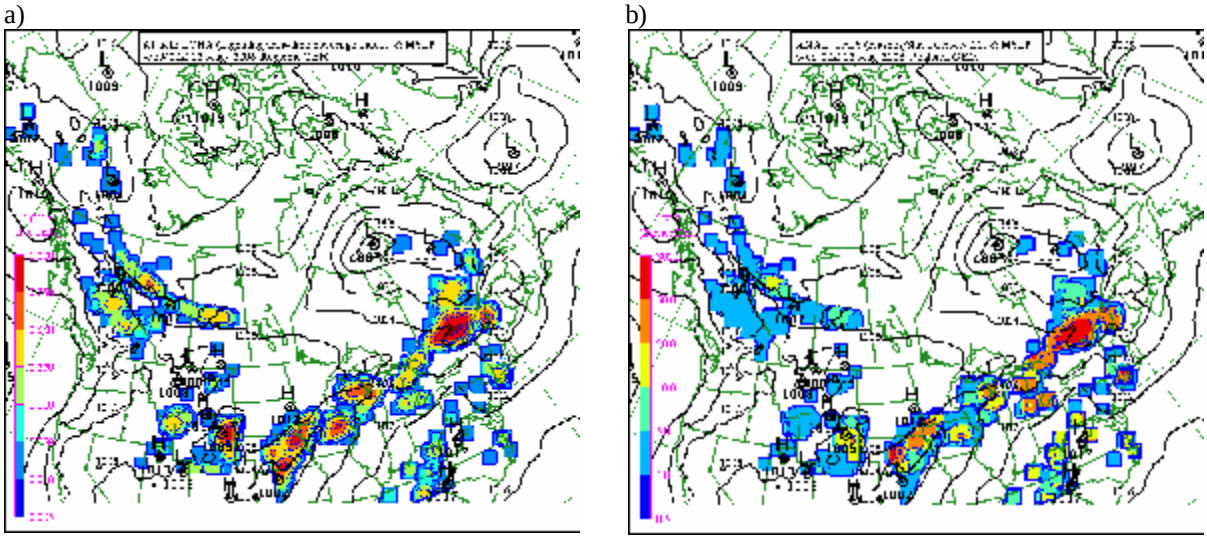



**Figure 10.** Observed a) time-area coverage index (LCHA); b) three-hour flash count (LFLS) for 2100 – 2400 UTC 01 August 2006. MSL pressure is overlain. Only the lightning reports received by Environment Canada are shown.

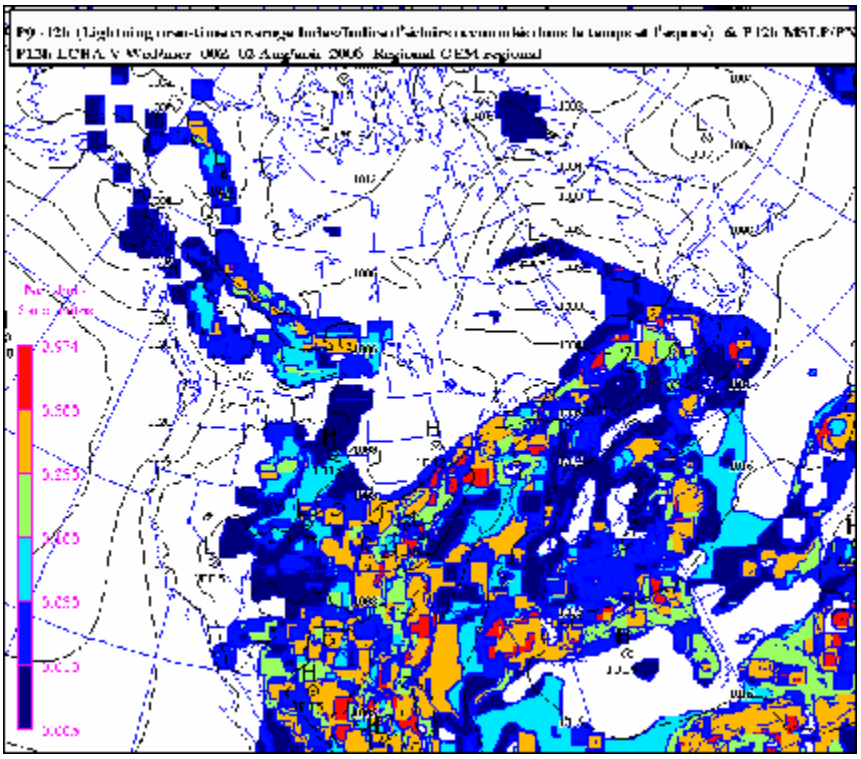


**Figure 11.** 9-12 hour forecast of the time-area coverage index (LCHA) valid for 2100 UTC 01 August 2006 to 0000 UTC 02 August 2006.

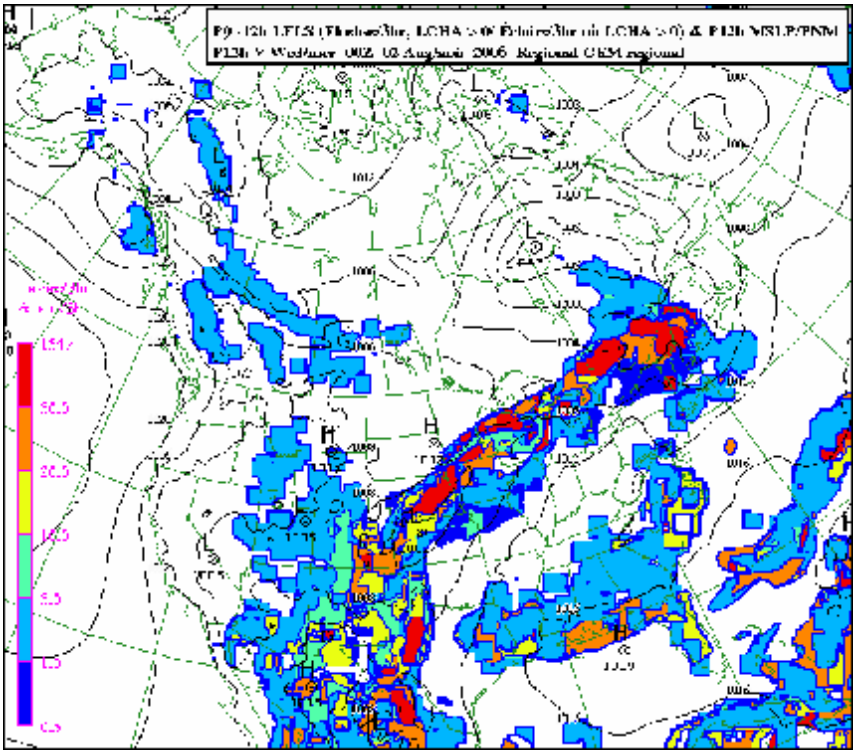


**Figure 12.** 9-12 hour forecast of the three-hour flash count (LFLS) valid for 2100 UTC 01 August 2006 to 0000 UTC 02 August 2006.

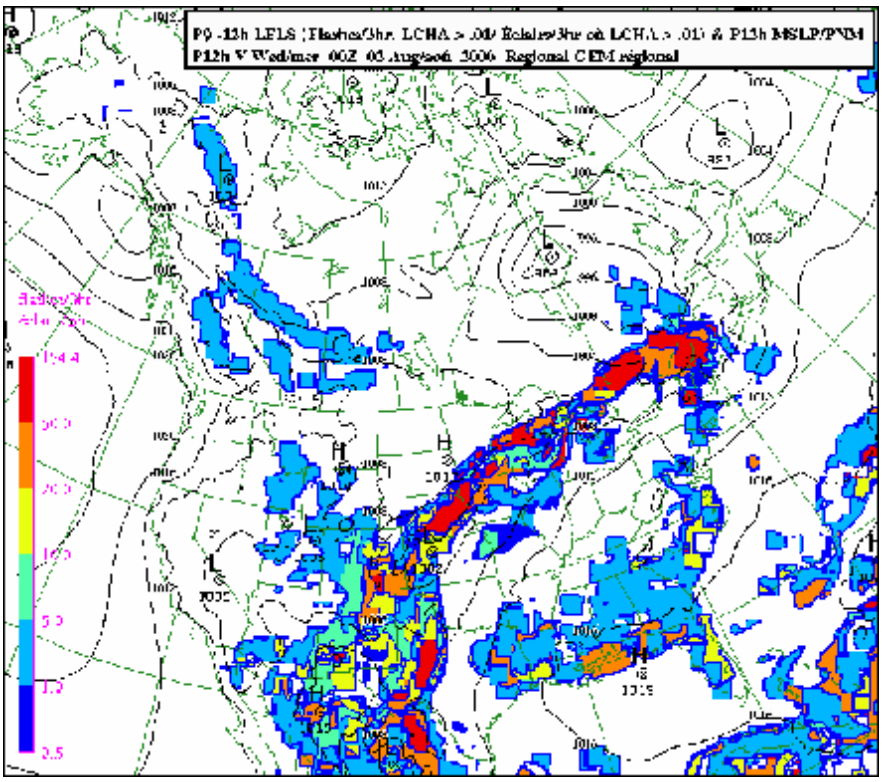


**Figure 13.** The LFLS forecast filtered by values of LCHA > .01 for the same period as Fig. 12.

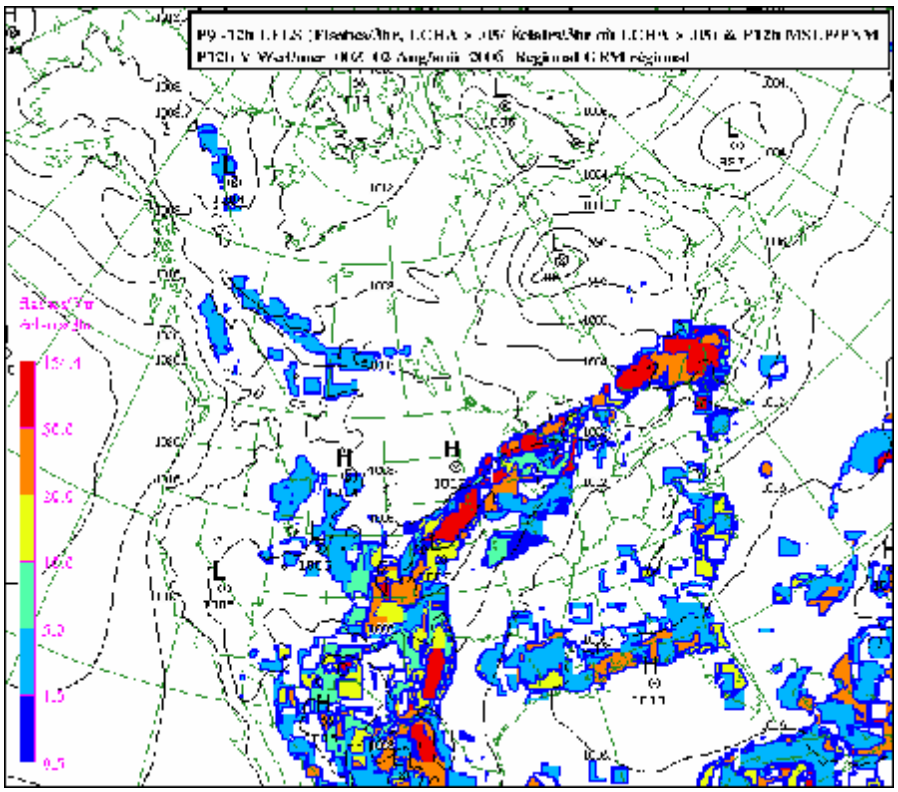


**Figure 14.** The LFLS forecast filtered by values of LCHA > .05 for the same period as Fig. 12

lightning there. LFLS forecasts filtered at the LCHA >= .10 level and LCHA >= .25 level are not shown here. The LFLS forecast filtered at the LCHA >= .10 level forecast the intensive lightning areas and the Florida peninsula very well. The LFLS forecast filtered at the LCHA >= .25 level was found to be too strongly filtered for all but high flash rates in most areas, in fact this proved true on most days. These forecasts are provided in the suite of forecasts available to Canadian weather offices.

The forecasts were verified against observations for the portion of the NALDN region seen in Environment Canada. Table 9 shows results for July 2006. Total numbers of forecasts of lightning and no lightning are shown for periods of relatively large lightning activity (0000 - 0300 UTC) and relatively little lightning activity (1200 – 1500 UTC). FL01, FL05, FL10, and FL25 are the LFLS forecast where LFLS is set to 0 where LCHA < .01, .05, .10, and .25 respectively. The LCHA and LFLS forecasts of lightning have greater accuracy at times of greater lightning activity and least accuracy at times of least lightning. The accuracy of forecasts of no lightning is generally greater than 90%, while the accuracy of forecasts of lightning is 45-50% or higher at diurnal times of greater lightning activity (2100-0600 UTC) and 30-40% or higher at times of lesser activity (0900-1800 UTC). Not surprisingly the accuracy of forecasts of lightning > 0 decreases for projections greater than 24 hours. When LFLS forecasts that lightning will occur are filtered by successively greater values of LCHA their overall accuracy increases, however their number drops drastically below what is observed. Thus there is a trade-off of overall accuracy against number of forecasts that lightning will occur. For July 2006 the FL01 forecasts were found to give the greatest overall accuracy at times of greater lightning activity, while the unfiltered LFLS forecasts had greatest accuracy for the greatest number of forecasts at times of least lightning activity.

Table 10 shows the same verification results as Table 9 for selected forecasts for the 0000-0300 UTC and 1200-1500 UTC periods for May, June, August, and September 2006. The FL10 and FL25 columns are dropped since the conclusion from results is the same as for July. Conclusions for June and August are similar to July. For May and September, when lightning activity is diminished, the unfiltered LFLS forecasts appear to give the best overall accuracy at all times.

The previous results are for verification of the predictands with observations calculated exactly as the predictands were calculated. However if the forecasts of the three-hour time-area coverage index and flash count are verified against observations of their ***maximum*** in the 324 point data cloud the forecasts of lightning to occur are seen to have better accuracy. Figures 15 and 16 show this for 24 hour forecasts valid 2100-2400 UTC and 0900-1200 UTC, respectively. The black lines in Figs. 15 and 16 show the fraction correct LCHA and LFLS forecasts of "any" lightning occurrence (LCHA ≥ .01 or LFLS ≥ 1 flash per 3 hours) and "sig (significant)" lightning" (LCHA ≥ .05 or LFLS ≥ 5 flashes per 3 hours), where the verifying observation is calculated from the 324 data point cloud in the same way as the predictand. Note that for LCHA and LFLS values less than the threshold number for any lightning and significant lightning that the values shown are the fraction correct for ***no lightning*** forecasts. The red lines show the same quantities where the verifying observation is calculated as the ***maximum value*** in the 324 data point cloud. The fraction of correct forecasts for LCHA ≥ 1 and LFLS ≥ 1 is seen to be greater when the verifying observation is calculated from the maximum value for both lightning designations. The forecasts for LCHA < .01 and LFLS < 1 flash/3-hr are less accurate when the verifying observation is calculated from the maximum value because there is a greater chance of at least 1 of the 324 points having LCHA ≥ .01 or LFLS ≥ 1. The above results held true for all other projection times.

In Figs. 15 and 16 the fraction of correct LFLS forecasts increases slightly as LFLS increases beyond 1 flash/3-hr for forecasts valid in the 2100-2400 UTC period, but the fraction of correct LFLS forecasts decreases as LFLS increases beyond 1 flash/3-hr for forecasts valid in the 0900-1200 UTC period.

In general there are too many forecasts that lightning will occur at both low and high levels of activity (around 1 flash/3-hr) in afternoon and evening and in very warm air masses, where greater convective activity can occur. Evidence of this can be seen in Fig. 11. Filtering the forecasts may improve this situation. Fig 17 shows statistics for 12-hr July LFLS forecasts valid 2100-2400 UTC, where the forecasts are filtered by LCHA values 0, .01, .05 .10, and .25. A filter of .25 means those LFLS forecasts where LCHA ≥ .25. The ratio of forecasts to observations is above 1 for lightning 1-5 fls/3-hr

and 20-50 fls/3-hr or more for unfiltered LFLS forecasts. A filtering value between .01 and .05 will reduce the ratio to about 1 for 1-5 fls/3hr and 20-50 fls/3-hr, and a filtering value near .25 will reduce the ratio to about 1 for forecasts of 50+ fls/3-hr. This filtering would increase accuracy. Forecasts verified with the maximum LCHA and LFLS values in the 324 data point cloud, shown in Fig. 18, showed the same results for low levels of lightning activity and that any filtering would increase accuracy.

## 4. REMARKS

New dynamical-statistical models to forecast lightning in 3-hr intervals to 48 hrs for all of Canada and the United States were developed and have run in real time since late April 2006. Output is available to all Canadian forecast offices. The forecast domain in includes large areas for which either training data was not available in Canada, or lightning is not detected. The forecasts have become widely used for production of public and aviation forecasters. The most common use thus far is for daily convective assessment by forecasters and for defining areas of convection in aviation area forecasts. Many forecasters have remarked that while regions of strong convective development are usually recognized by other methods, these lightning forecasts often alerted them to areas of weaker convective development that were overlooked. Since lightning activity is relatively low over most of Canada and the area is huge, help in forecasting lower lightning activity is much appreciated by forecasters. The forecasts should also prove useful eventually in forecast generation software at CMC. There has been considerable interest from forest fire weather forecasters.

## REFERENCES


Brieman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 1984: *Classification and Regression Trees*. ISBN 0412048418, Chapman & Hall / CRC Press, Boca Raton, FL, (www.crcpress.com), 368 pp.

Burrows, W. R., C. Price, and L. J. Wilson, 2005: Warm season lightning probability prediction for Canada and the northern United States. *Wea. And Forecasting*, **20**, 971-988.

____, 2002:P. King, P. J. Lewis, B. Kochtubajda, B. Snyder, V. Turcotte: Lightning occurrence patterns over Canada and adjacent United States from lightning detection network observations. *Atmosphere.-Ocean*, **40(1)**, 59-80.

____, 1998: CART-neuro-fuzzy statistical data modeling, part 1: method. *Preprints, First Conf. on Artificial Intelligence,* 11-16 January, 1998, Phoenix, AZ, J33-40.

____, 1985: On the use of time-offset model output statistics for production of surface wind forecasts. *Mon. Wea. Rev.*, **113**, 2049-2054

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.

Faucher, M., W. R. Burrows, and L. Pandolfo, 1999: Empirical-statistical reconstruction of marine winds along the western coast of Canada. *Clim. Res.* **11**, 173-190.

Kain, J. S. and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain-Fritsch scheme. The representation of cumulus convection in numerical models. *Meteor. Monogr.*, **27**, Amer. Meteor. Soc., 165-170.

Kuo, H. L., 1974: Further studies on the parameterization of the influence of cumulus convection on large scale flow. *J. Atmos. Sci.*, **31**, 1232-1240.

Orville, R. E., G. R. Huffines, W. R. Burrows, R. L. Holle, and K. L. Cummins: North American Lightning Detection Network (NALDN); First results: 1998-2000. *Mon. Wea. Rev..*, **130**, 2098-2109.

Venables, W. N. and Ripley, B. D., 2002: *Modern Applied Statistics with S*. New York, Fourth Edition, Chap. 9. *Springer-Verlag*, ISBN 0-387-95457.

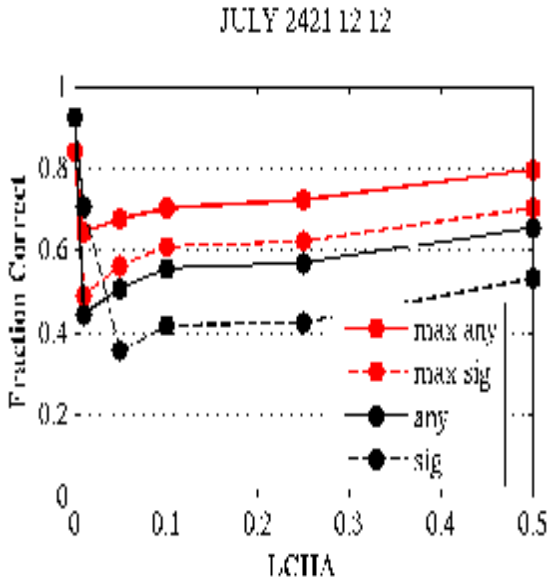
Walmsley, J. L., W. R. Burrows, and R. S. Schemenauer, 1999: The use of routine weather observations to calculate liquid water content in summertime high-elevation fog. *J. Appl. Meteor.*, **38**, 369-384.

| Time | Proj | LCHA | LFLS | FL01 | FL05 | FL10 | FL25 |
|---|---|---|---|---|---|---|---|
| **0300** | **15** | 1952006<br>799461<br>.93 / .45 | 2148876<br>602591<br>.91 / .51 | 2221526<br>529941<br>.90 / .52 | 2490122<br>261345<br>.86 / .54 | 2587571<br>163896<br>.85 / .60 | 2665316<br>86151<br>.84 / .63 |
| | **27** | .92 / .44 | .91 / .50 | .90 / .51 | .86 / .56 | .85 / .59 | .83 / .60 |
| | **39** | .92 / .43 | .90 / .48 | .89 / .49 | .86 / .54 | .85 / .58 | .84 / .59 |
| **0603** | **18** | .94 / .37 | .91 / .47 | .90 / .48 | .87 / .56 | .87 / .59 | .87 / .59 |
| | **30** | .93 / .37 | .90 / .46 | .89 / .47 | .87 / .55 | .87 / .57 | .87 / .58 |
| | **42** | .93 / .35 | .90 / .43 | .89 / .44 | .87 / .50 | .87 / .54 | .87 / .56 |
| **0906** | **09** | .93 / .43 | .91 / .44 | .91 / .47 | .90 / .50 | .90 / .50 | .91 / .47 |
| | **21** | .93 / .41 | .92 / .42 | .91 / .45 | .90 / .46 | .90 / .47 | .90 / .47 |
| | **33** | .93 / .40 | .91 / .39 | .91 / .42 | .90 / .44 | .90 / .45 | .90 / .45 |
| | **45** | .93 / .38 | .92 / .38 | .91 / .40 | .90 / .42 | .90 / .43 | .90 / .44 |
| **1209** | **12** | .93 / .39 | .93 / .41 | .92 / .44 | .92 / .44 | .92 / .46 | .92 / .45 |
| | **24** | .94 / .36 | .93 / .39 | .92 / .41 | .92 / .42 | .92 / .43 | .92 / .44 |
| | **36** | .93 / .35 | .93 / .37 | .92 / .40 | .92 / .39 | .92 / .41 | .92 / .41 |
| | **48** | .93 / .32 | .93 / .36 | .93 / .37 | .92 / .37 | .92 / .38 | .92 / .39 |
| **1512** | **15** | 2557675<br>193792<br>.95 / .33 | 2611387<br>140080<br>.94 / .31 | 2676291<br>75176<br>.94 / .34 | 2704390<br>47077<br>.93 / .36 | 2710887<br>40580<br>.93 / 36 | 2725628<br>25839<br>.93 / .37 |
| | **27** | .95 / .32 | .94 / .32 | .94 / .35 | .94 / .39 | .94 / .40 | .93 / .42 |
| | **39** | .95 / .29 | .94 / .28 | .94 / .30 | .93 / .33 | .93 / .33 | .93 / .34 |
| **1815** | **18** | .93 / .37 | .92 / .35 | .92 / .36 | .91 / .36 | .91 / .36 | .91 / .35 |
| | **30** | .93 / .35 | .92 / .32 | .92 / .34 | .91 / .36 | .91 / .36 | .91 / .38 |
| | **42** | .92 / .32 | .92 / .30 | .92 / .31 | .91 / .32 | .91 / .32 | .91 / .34 |
| **2118** | **09** | .92 / .48 | .91 / .50 | .90 / .53 | .87 / .58 | .86 / .57 | .85 / .57 |
| | **21** | .91 / .48 | .90 / .49 | .89 / .52 | .86 / .55 | .85 / .53 | .84 / .51 |
| | **33** | .91 / .46 | .90 / .47 | .89 / .50 | .87 / .53 | .86 / .51 | .85 / .50 |
| | **45** | .90 / .46 | .89 / .46 | .88 / .49 | .86 / .52 | .85 / .50 | .85 / .50 |
| **2421** | **12** | .93 / .50 | .93 / .53 | .91 / .56 | .86 / .61 | .85 / .62 | .83 / .62 |
| | **24** | .92 / .50 | .92 / .51 | .90 / .56 | .86 / .61 | .84 / .61 | .83 / .61 |
| | **36** | .92 / .48 | .92 / .49 | .90 / .53 | .86 / .60 | .85 / .60 | .83 / .61 |
| | **48** | .91 / .47 | .91 / .47 | .89 / .51 | .85 / .57 | .84 / .57 | .83 / .59 |

**Table 9.** Fraction of correct forecasts of no lightning and of lightning, respectively, for July 2006. First number in the first column is the UTC valid time of the forecast, second number is the projection hours, e.g. 0300 15 is a 15 hour forecast valid 0000 – 0300 UTC. For the 0300 15 and 1512 15 rows, the total number forecasts of no lightning is the first number shown, the second number shown is the total number of forecasts of lightning. FL01 is the LFLS forecast where LFLS is set to zero where LCHA < .01. FL05, FL10, and FL25 are the LFLS forecast where LFLS is set to 0 where LCHA < .05, .10, and .25 respectively.

| Time Proj | LCHA | LFLS | FL01 | FL05 |
|---|---|---|---|---|
| **May 0300 15** | 2439445<br>312022<br>.96 / .49 | 2474450<br>277017<br>.96 / .50 | 2517881<br>233586<br>.95 / .53 | 2631605<br>119862<br>.93 /.58 |
| **39** | .95 / .44 | .95 / .46 | .94 / .48 | .93 / .51 |
| **1512 15** | 2581816<br>80894<br>.98 / .33 | 2618407<br>44303<br>.98 / .35 | 2635884<br>26826<br>.97 / .41 | 2645667<br>17043<br>.97 / .41 |
| **39** | .98 / .29 | .98 / .29 | .97 / .34 | .97 / .36 |
| **June 0300 15** | .93 / .49 | .92 / .53 | .91 / .55 | .88 / .61 |
| **39** | .93 / .46 | .91 / .50 | .91 / .53 | .88 / .58 |
| **1512 15** | .96 / .30 | .95 / .34 | .95 / .41 | .95 / .50 |
| **39** | .96 / .28 | .95 / .32 | .95 / .34 | .95 / .40 |
| **August 0300 15** | .95 / .45 | .94 / .48 | .93 / .50 | .90 / .58 |
| **39** | .94 / .34 | .93 / .46 | .93 / .48 | .90 / .54 |
| **1512 15** | .96 / .36 | .95 / .35 | .95 / .39 | .95 / .44 |
| **39** | .96 / .29 | .95 / .29 | .95 / .32 | .95 / .36 |
| **September 0300 15** | .97 / .37 | .97 / .43 | .97 / .45 | .96 / .51 |
| **39** | .97 / .35 | .97 / .39 | .96 / .41 | .95 / .46 |
| **1512 15** | .98 / .24 | .98 / .30 | .98 / .33 | .98 / .38 |
| **39** | .98 / .22 | .98 / .26 | .96 / .43 | .98 / .34 |

**Table 10.** Same as Table 9 for May, June, August, and September 2006, for selected forecasts.



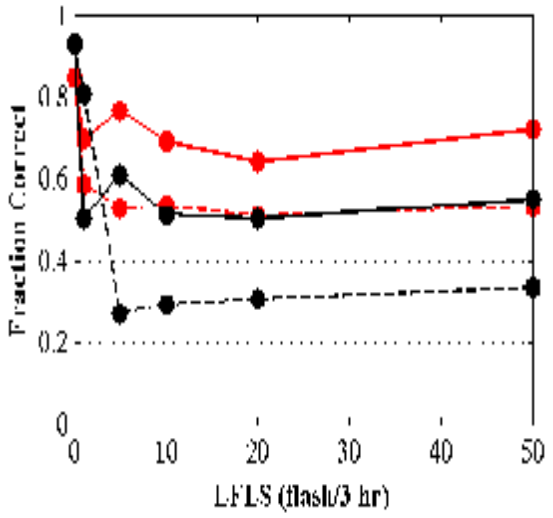
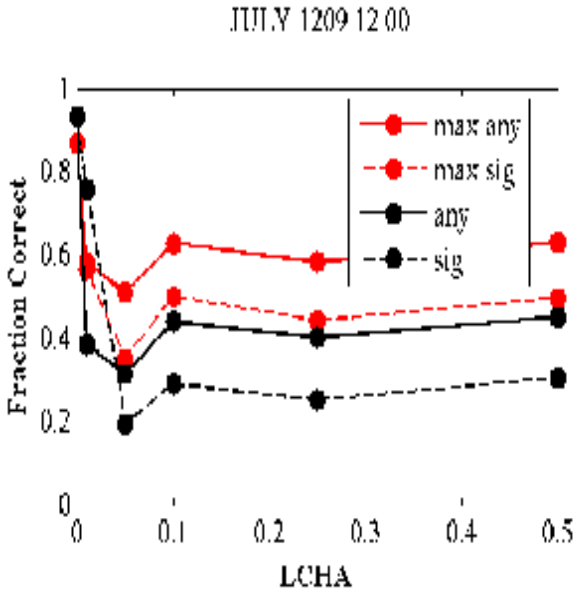




**Figure 15**. The fraction of correct LCHA and LFLS forecasts of "any" lightning and "significant" lightning (defined above) as LCHA and LFLFS increase for 12-hr July 2006 forecasts valid 2100-2400 UTC. "max" refers to using the maximum LCHA and LFLS value in the 324 data point cloud for the verifying observation. Note that for LCHA and LFLS values less than the threshold number for "any" lightning and "significant" lightning that the values shown are the fraction correct for *no lightning* forecasts.



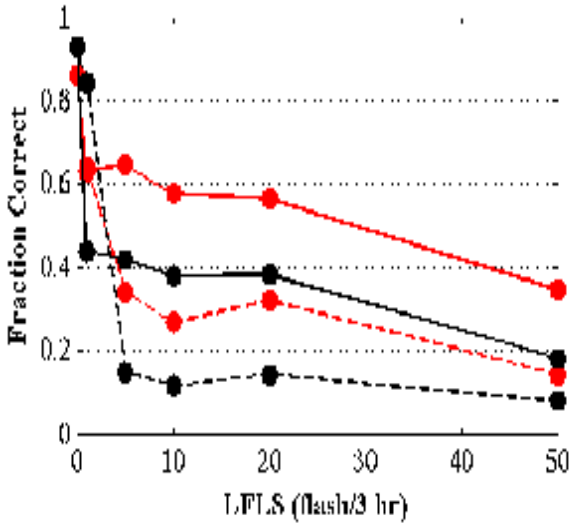




**Figure 16.** The fraction of correct LCHA and LFLS forecasts of "any" lightning and "significant" lightning (defined above) as LCHA and LFLFS increase for 12-hr July 2006 forecasts valid 0900-1200.

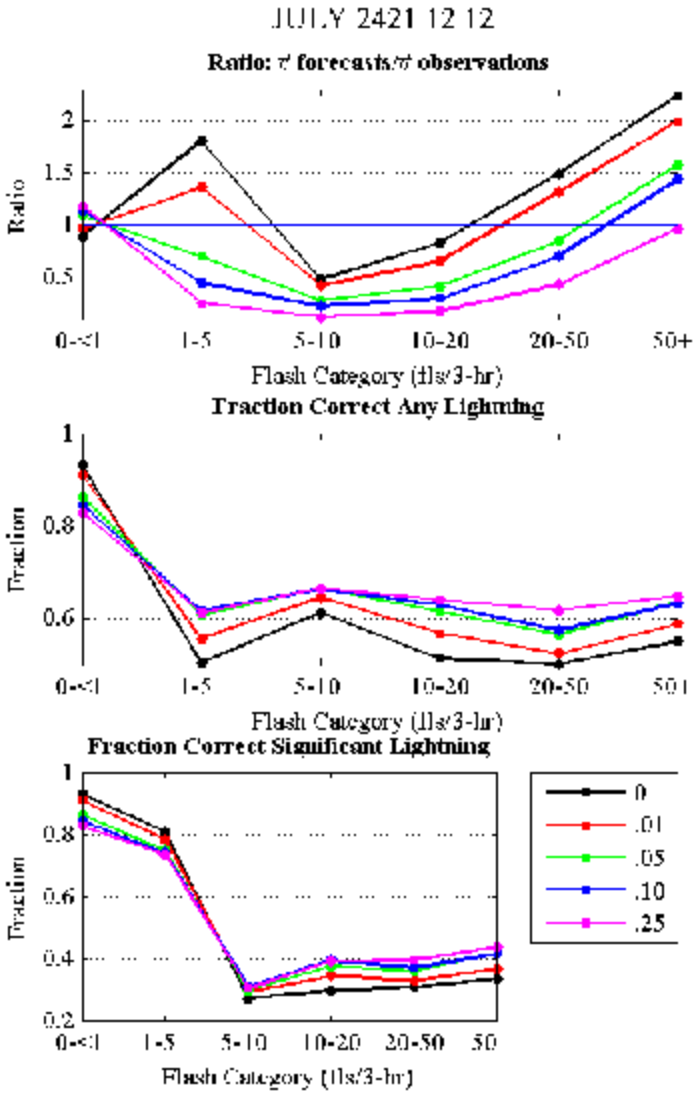



**Figure 17.** The ratio of forecasts to observations and the fraction correct for any lightning and significant lightning for 12-hr forecasts valid 2100-2400 UTC July 2006.

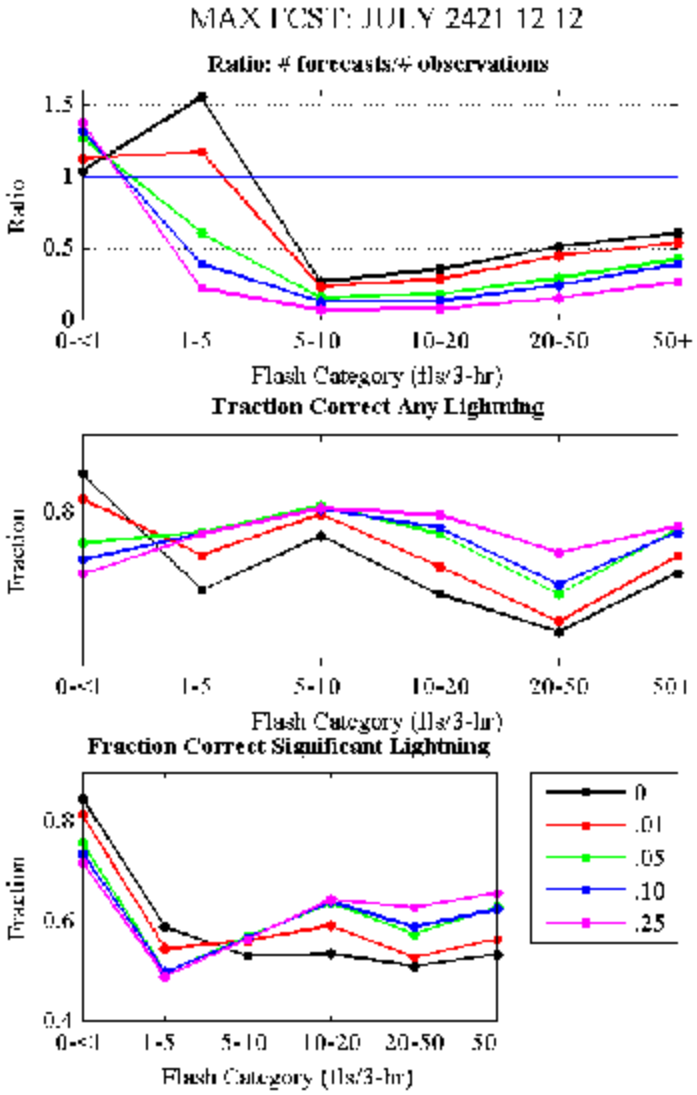



**Figure 18.** The ratio of forecasts to observations and the fraction correct for any lightning and significant lightning for 12-hr forecasts valid 2100-2400 UTC July 2006, where the verifying observations of LCHA snd LFLS are the maximum value in the 324 data point cloud.