4.1 **OBJECTIVE LIGHTNING FORECASTING AT KENNEDY SPACE CENTER AND CAPE CANAVERAL AIR FORCE STATION USING CLOUD-TO-GROUND LIGHTNING SURVEILLANCE SYSTEM DATA**

Winfred Lambert* and Mark Wheeler
ENSCO, Inc/Applied Meteorology Unit

William Roeder
USAF 45th Weather Squadron

1. INTRODUCTION

The 45th Weather Squadron (45 WS) at Cape Canaveral Air Force Station (CCAFS) in Florida includes a probability of lightning occurrence in their daily 24-hour and weekly planning forecasts. This information is used for general planning of operations at CCAFS and Kennedy Space Center (KSC). These facilities are located in east-central Florida at the east end of a corridor known as 'Lightning Alley', seen in Figure 1 as the red area oriented southwest to northeast across the center of the Florida peninsula. Much of the current lightning probability forecast is based on a subjective analysis of model and observational data and an objective forecast tool developed over 30 years ago. The 45 WS requested that a new lightning probability forecast tool based on statistical analysis of more recent historical warm-season (May-September) data be developed in order to increase the objectivity and repeatability of the daily thunderstorm probability forecast. The resulting tool is a set of statistical lightning forecast equations, one for each month of the warm season, that provide a lightning occurrence probability for the day by 1100 UTC (0700 EDT) during the warm season.

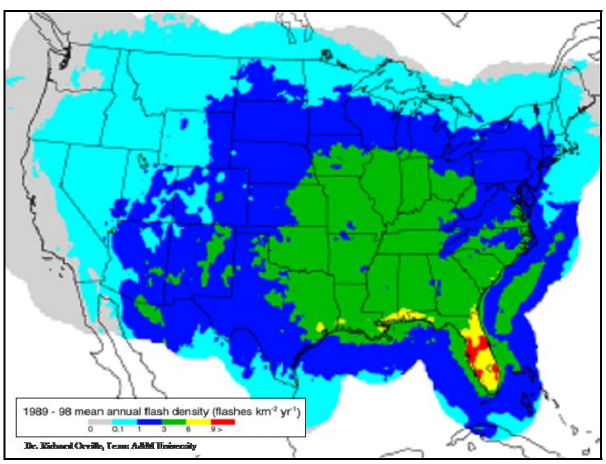

Figure 1. Mean annual lightning flash density across the continental U.S., 1989 – 1998 (Dr. Richard Orville).

* *Corresponding author address:* Winifred C. Lambert, ENSCO Inc., 1980 N. Atlantic Ave., Suite 230, Cocoa Beach, FL 32931, lambert.winifred@ensco.com.

2. BACKGROUND

The 45 WS currently uses the Neumann-Pfeffer Thunderstorm Index (NPTI) as their main objective tool for predicting lightning probability (Neumann, 1971). The NPTI was created to provide the probability of thunderstorm occurrence specifically at CCAFS. However, the NPTI has several shortcomings. The observational data sample size used in its development was relatively small. It was proven to under-forecast lightning occurrence (Wohlwend 1998), though a bias-correction technique was applied to improve performance (Roeder 1998). Howell (1998) and Everitt (1999) have shown that the 1-day persistence forecast performance outperforms NPTI by ~10%. These issues indicated that the NPTI needed to be upgraded or replaced. Since many more years of historical observations are now available and more advanced statistical analysis techniques are possible due to increased computing power, the 45 WS teamed with the Applied Meteorology Unit (AMU, Bauman et al., 2004) to create a new lightning probability tool for the KSC/CCAFS area.

2.1 *Important Factors for Lightning Forecasting*

Several meteorological factors are known to be important in lightning prediction for KSC/CCAFS. They include convective instability, synoptic scale flow regime, persistence, and daily lightning climatology. Previous studies and local experience have shown that the K-Index and Lifted Index derived from the CCAFS sounding are the best predictors of thunderstorm formation in the area (Cetola 1997, Kelly 1998). Lericos et al. (2002) also showed that the synoptic-scale flow regime was important in determining where the highest flash densities would occur over the peninsula. This is due to the influence the synoptic flow has on the propagation and interaction of peninsular-Florida's two sea breezes: the east coast sea breeze from the Atlantic Ocean, and the west coast sea breeze from the Gulf of Mexico. Persistence is also an important contributor (Everitt 1999). Whether lightning was observed the previous day influences the probability that lightning will be observed on the current day. Finally, climatological probability of lightning for each calendar day varies considerably throughout the season, but provides a good starting point when developing a lightning probability forecast.

Another important factor in creating a reliable probability forecast tool is the selection of the statistical regression technique. Linear regression can be used, but has several weaknesses in probability forecasting. The mathematical formulation can allow forecasts of probabilities greater than 100% or less than 0%, which are unrealistic. Linear regression will not calculate the sudden change in probability when a parameter passes beyond a threshold value or range of values, as often happens in the atmosphere. Logistic regression is a more appropriate technique for probability forecast equations (Wilks 1995). It is bounded by 0% and 100% and allows for sudden changes in probability as predictor values exceed a threshold, or it can allow for nearly linear response of the probability to the predictor. Everitt (1999) showed that using logistic regression versus linear regression yielded 48% better skill when using the same predictor variables and data. The gain was solely due to the logistic regression method.

2.2 *Current Work*

The AMU work described herein was based on the results from two earlier research projects already mentioned. Everitt (1999) used hourly surface observations at the Shuttle Landing Facility (TTS) and CCAFS rawinsonde data (XMR) to develop equations that forecast the daily probability of thunderstorm occurrence at KSC/CCAFS. The hourly TTS surface observations of thunder were used as the predictand. The candidate predictors included parameters from the XMR sounding, smoothed climatology of daily thunderstorm probability, and 1- to 6-day persistence. Regression equations were stratified by month and by easterly and westerly flow regimes. The results showed that using the logistic regression method produced a more skillful forecast than multiple linear regression, even when using identical predictors. These equations showed a 48% skill improvement over the NPTI. They also showed a 43% improvement over persistence, which was important since Everitt (1999) also showed that persistence was ~10% more skillful than the NPTI.

Lericos et al. (2002) developed lightning distributions over the Florida peninsula that were stratified by flow regimes. The flow regimes were inferred from the average wind direction in the 1000-700 mb layer from the rawinsondes at Miami (MIA), Tampa (TBW), and Jacksonville (JAX), Florida. The lightning data were from the National Lightning Detection Network (NLDN). The results suggested that the daily flow regime may be an important predictor of lightning occurrence on KSC/CCAFS.

The equations in this study were developed using the XMR sounding variables, daily climatology, 1-day persistence, and the logistic regression method as in Everitt (1999), and the flow regimes as developed in Lericos et al. (2002).

3. DATA

The period of record (POR) for the data in this study included the warm season, May – September, for the 15 year period 1989-2003. Data from three sources were used: 1) the local Cloud-to-Ground Lightning Surveillance System (CGLSS) to determine the dates on which lightning occurred 2) local 1000 UTC XMR sounding for stability parameters, and 3) peninsula-wide 1200 UTC soundings to calculate flow regimes.

The CGLSS is a network of six sensors (Figure 2) that provides date/time, latitude/longitude, strength, and polarity of cloud-to-ground lightning strikes in the local area. The CGLSS data have been found to be more reliable indicators of lightning in the area than the hourly TTS surface observations, as used in Everitt (1999). The CGLSS data also provide greater spatial accuracy and flash detection in the area of interest than the NLDN (Harms et al., 1998).

In the warm season, there are usually three XMR soundings a day at 1000, 1500, and 2300 UTC. The 45 WS typically uses data from the 1000 UTC sounding for issuing the daily planning forecasts at 1100 UTC. Therefore, the 1000 UTC XMR sounding data were used in this work to calculate the stability parameters that are normally available to the 45 WS.

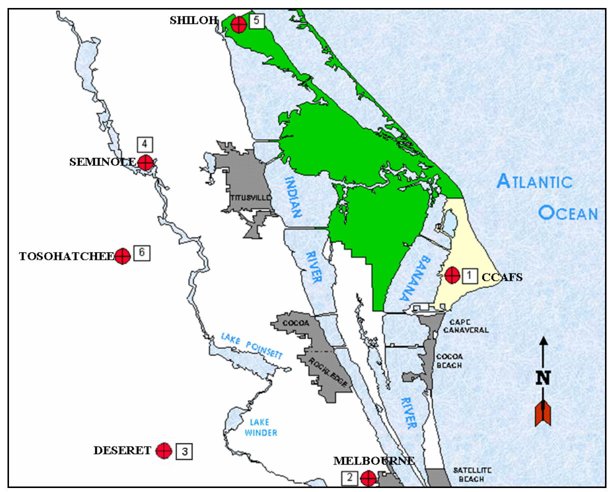

Figure 2. Map of east-central Florida. The locations of the six CGLSS sensors are shown as red circles. The names and numbers of each sensor are to the side of the red circles.

Rawinsonde data from the same stations as in Lericos et al. (2002) were used to develop the daily flow regimes for each day in the POR. Following the procedure in Lericos et al. (2002), the 1200 UTC soundings from MIA, TBW, and JAX were used to determine the large scale flow regime for the day. The current MIA and JAX sites were located at West Palm Beach, FL (PBI) and Waycross, GA (AYS), respectively, prior to 1995. The AYS and PBI data were used as proxies for JAX and MIA, respectively, during the period 1989-1994. All future references to JAX and MIA include the data from AYS and PBI. The map in Figure 3 shows the locations of all the soundings used in the study.

Use of the 1200 UTC sounding may seem inappropriate as it cannot provide data in time for the 1100 UTC briefing. However, the 1000-700 mb flow in Florida warm season 0000 UTC soundings could be contaminated by afternoon convective circulations. For the purpose of determining the flow regimes for each day in the period of record, the 1200 UTC provided the most reliable data. In an operational setting, the 45 WS can use several data sources, including satellite and hourly surface observations, to help determine the flow regime of the day before the briefing. Also, due to the weak synoptic patterns in the Florida warm season, there is likely not to be a flow regime change within 2 hours.

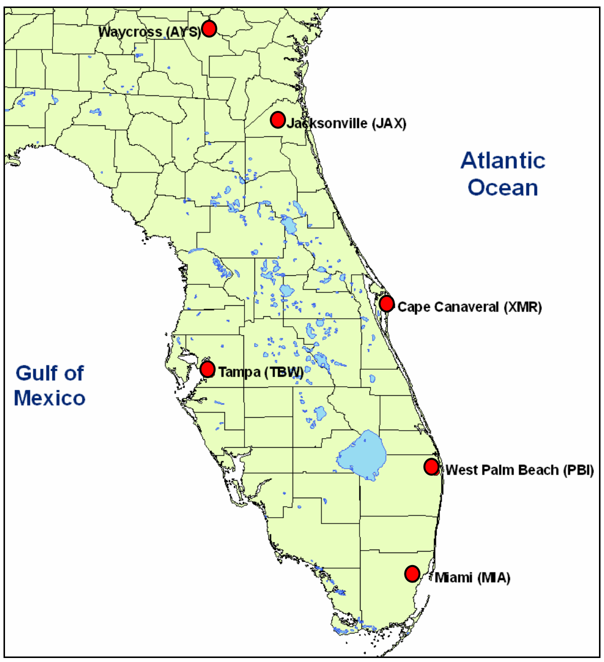

Figure 3. Map of the Florida Peninsula. The red dots show the locations of all soundings used in this task.

4. PREDICTAND/PREDICTOR PREPARATION

Each data set was processed and analyzed to create the variables that would be used in the statistical forecast equation development. The CGLSS data were used as ground truth indicating whether or not lightning occurred. The sounding data were used to calculate the predictors of lightning occurrence.

4.1 *Predictand*

The CGLSS data were used to create a binary predictand for the equations. The analyses hinged only on whether lightning was observed or not during each day. The calculations did not consider how many lightning strikes were detected. Calculation of the predictand was straightforward: the predictand was set equal to '1' if lightning occurred on a specific day, otherwise a '0' was assigned. The data were filtered to include only lightning strikes recorded during the warm season in the time period 1100-0400 UTC (7:00 AM to midnight, EDT) and in the geographic area outlined by the red box in Figure 4. The area is a rectangle defined by the outer-most points of the KSC, CCAFS, and Cape Canaveral Port lightning advisory circles whose radii are 5 n mi. Due to the complexity of computing the area of several intersecting circles, the area for this study is a rectangle defined by the outer-most points of all the circles. Some of the area inside the rectangle is not inside any of the 5 n mi circles, but lightning within the rectangle would be sufficiently close as to likely cause the 45 WS to issue a Phase II lightning warning (Weems et al., 1998).

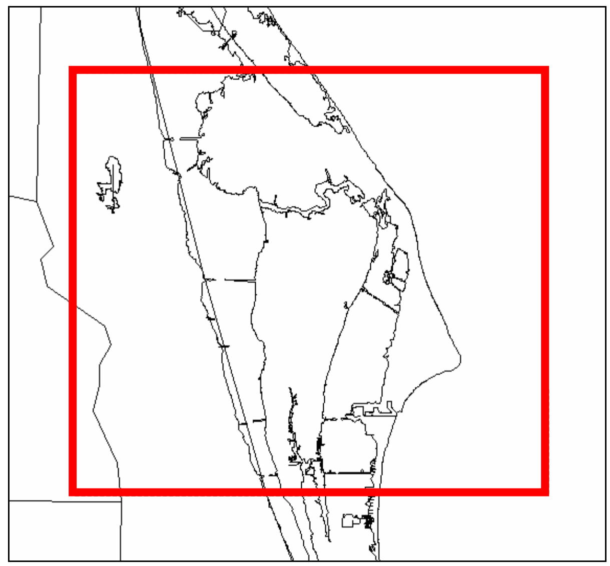

Figure 4. The red box defines the area in which lightning strikes detected by CGLSS were used to indicate whether or not lightning occurred on days in warm season, 1989–2003 between 1100-0400 UTC.

4.2 *Candidate Predictors*

The candidate predictors were tested prior to and during equation development to determine which of them in what combination would provide the best probability forecast of lightning occurrence. The candidate predictors included the climatological daily lightning frequency, 1-day lightning persistence, stability parameters calculated from the XMR rawinsonde, and the flow regimes.

CGLSS Data

The CGLSS predictand was used to develop a climatological daily lightning frequency that would be used as a possible predictor in the equations, as in Everitt (1999). The 'raw' frequency was rather noisy, as evidenced by the light blue curve in Figure 5. To reduce the noisiness, a center-weighted Gaussian smoother with a scale parameter of 3 days was applied to the daily frequency values seven days before and after each day. The result is the smooth dark blue line in Figure 5. The last seven days of April and first seven days of October were used to calculate the smoothed frequencies at the beginning of May and the end of September. The smoothed values were used as candidate predictors for the equations.

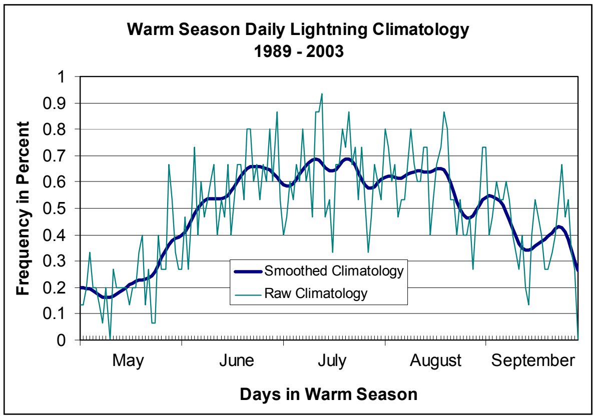

Figure 5. The daily raw (light blue) and smoothed (dark blue) climatological frequency values of lightning occurrence for the warm-season 1989 – 2003.

The CGLSS predictand for each day was also used to create another candidate predictor, the 1-day persistence forecast. If lightning occurred on a particular day, the persistence forecast for the next day was '1'. If lightning did not occur, the persistence forecast was '0' for the next day. The lightning occurrence information for April 30 was used to make the persistence forecast for May 1.

XMR 1000 UTC Rawinsonde Data

The XMR data were used to calculate the stability parameters that are usually available to the 45 WS. The stability parameter candidate predictors include the

- Total Totals (TT),
- K-Index (KI),
- Cross Totals (CT),
- Lifted Index (LI),
- Severe WEAther Threat (SWEAT) Index,
- Showalter Index (SSI),
- Thompson Index (TI)
- Temperature at 500 mb, (T500),
- Mean Relative Humidity in the 800-600 mb layer (RH),
- Precipitable water up to 500 mb (PW),
- Convective Inhibition (CIN),
- Convective Available Potential Energy (CAPE),
- CAPE based on the forecast maximum temperature, and
- CAPE based on the maximum θ_e below 300 mb.

These data were stratified by month, then each month of data was stratified into two subsets by days with and without lightning. A Student's t-test (Wilks 1995) was performed on the mean values of each parameter between the two subsets in each month to determine if they were statistically significantly different for lightning days than for non-lightning days. If the means were statistically different, the parameter would be used as a candidate predictor. The null hypothesis that the means were equal could be rejected at the 99% confidence level for all except CIN and the three CAPE parameters. This meant that the parameters had different means between lightning and non-lightning days and would be used as candidate predictors. For the CAPE and CIN values, the null hypothesis could be accepted (the means were equal) at the 90-99% confidence levels, depending on the month. This was an indication that CAPE in any form and CIN would not be good predictors of lightning occurrence. This agrees with local experience that CAPE is not useful for thunderstorm prediction in central Florida, since CAPE is usually uniformly high during the summer on convective and non-convective days.

Florida Peninsula Rawinsondes

The method outlined in Lericos et al. (2002) used the average wind direction in the 1000-700 mb layer at MIA, TBW, and JAX to determine the peninsular-scale flow regime. The average wind direction in the 1000-700 mb layer at each station was calculated for each 1200 UTC sounding using a depth-weighted averaging method in which the depth for each observation was the distance between the halfway points between adjacent observations. The flow regime for each day depended on the layer-averaged wind direction at each of the three stations. There are eight flow regimes named according to the resulting flow over KSC/CCAFS:

- Southwest flow (SW-1) over KSC/CCAFS occurred when the layer-averaged wind direction at all three stations was 180°-270°, indicating that the ridge associated with high pressure over the Atlantic Ocean was south of the Florida Peninsula.
- Southwest flow (SW-2) also occurred when the ridge was between MIA and TBW, with layer-averaged wind directions of 180°-270° at JAX and TBW and 90°-180° at MIA.
- Southeast flow (SE-1) occurred when the ridge moved north of KSC/CCAFS with the layer-averaged wind directions 180°-270° at JAX and 90°-180° at MIA and TBW.
- Southeast flow (SE-2) also occurred when the ridge was north of the Florida Peninsula and the layer-averaged wind direction at all three stations was 90°-180°.
- Northwest flow regime (NW) occurred when the layer-averaged wind direction at all three stations was 270°-360°.
- Northeast flow regime (NE) occurred when the layer-averaged wind direction at all three stations was 0°-90°.

- When the layer-averaged wind directions at the three stations did not fit any of the above criteria, it was designated as Other.
- When one or more soundings were missing the flow was designated as Missing.

The probability of lightning occurrence for each flow regime were calculated from the CGLSS data. These probabilities were developed as candidate predictors for the forecast equations. They were found to improve the lightning forecast compared to persistence and climatology when used on their own. Six tables containing the probabilities, one for the entire warm season and five for the individual months, were created for the 45 WS. The table for the entire warm season is given in Table 1 as an example.

Table 1. Flow regime lightning probability table for all months in the warm season. The candidate predictors are in the far right column titled 'Probability of Lightning'.

Flow Regime Lightning Statistics
Warm Season (May – September) 1989 – 2003

Probabilities of lightning occurring within a rectangle encompassing all 5 n mi warning rings based on flow regime are shown in the right-most column.

The strikes/day statistical values in the second column are based on lightning days only (fifth column). The median (M) value of strikes per day in each regime is shown with the 1st (Q1) and 3rd (Q3) quantiles in the order Q1, M, Q3. The mean and standard deviation of the strike numbers are shown in parentheses below Q1, M, Q3 (see explanation of M, Q1, and Q3 below).

Flow Regime	***Q1, M, Q3 of Strikes/Day (Mean, Stdev)***	***Total # Days (% of Total)***	***# Non Lightning Days***	***# Lightning Days***	***Probability of Lightning***
SW-1 Ridge S of MIA	68, 248, 507 (396, 496)	271 (12.7)	92	179	**66 %**
SW-2 Ridge between MIA/TBW	37, 169, 528 (357, 435)	218 (10.2)	60	158	**72 %**
SE-1 Ridge between TBW/JAX	4, 18, 110 (117, 223)	283 (13.3)	140	143	**51 %**
SE-2 Ridge N of JAX	3, 8, 41 (61,141)	218 (10.2)	133	85	**39 %**
NW	28, 179, 359 (342, 545)	93 (4.4)	53	40	**43 %**
NE	2, 14, 62 (68, 114)	100 (4.7)	82	18	**18 %**
Other (Regime Undefined)	9, 65, 265 (200, 325)	945 (44.4)	527	418	**44 %**
TOTALS	**10, 75, 324 (238, 381)**	**2128**	**1087**	**1041**	49 %

There is a 6% improvement in the forecast when using the individual flow regime probabilities over the seasonal climatological probability of 49%, and a 23% improvement over 1-day persistence. Forecast improvement was calculated using the Brier Skill Score.

The median is the strike-number value at which 50% of the cases had higher and 50% had lower strike numbers, i.e. the center of the strike-number distribution. It is *not* equal to the mean because the strike-number distributions are not symmetric. The 'middle' 50% of the cases are found between Q1 and Q3. For asymmetric distributions, like lightning strikes/day, the median and inter-quartile range are more representative of the data than the mean and standard deviation.

As expected, the two SW flow regimes were dominant in terms of lightning occurrence in the KSC/CCAFS area. Low-level SW flow impedes the inland progression of the east coast sea breeze, while allowing the west coast sea breeze to propagate eastward. When the two fronts meet near the east coast, low-level convergence is increased and, given sufficient moisture and instability, thunderstorms form. While it has been known anecdotally that SW flow increases the probability of convective development over KSC/CCAFS, the probability values had not been quantified. Note that the 'Other' category contained a large number of cases in the data set. Lericos et al. (2002) attempted to define more flow regimes than the six here. They did not find a sufficient number of cases in each flow category tested to declare any of them as a legitimate flow regime. Nonetheless, the 'Other' regime cannot be ignored as a flow regime in the equations.

The bottom row of Table 1 describes the improvement in skill realized when using the individual flow regime probabilities in the last column over that of climatology and 1-day persistence. This improvement in skill was found in each individual month as well as the full warm season. Given the skill improvement using the flow regime probabilities, these tables provide a reasonable starting point for the lightning forecasts. As such, they were delivered prior to project completion for immediate use by the 45 WS during the 2004 warm season.

5. EQUATION DEVELOPMENT AND TESTING

Once the predictand and candidate predictors were prepared, equation development began. The data were first stratified into development (i.e. dependent) and testing (i.e. independent) data sets, then stratified by month. Of the 15 years in the POR, 13 were used for equation development and two were set aside data for testing the equations. The stratification did not involve choosing individual warm season years, but individual warm season days. There are 153 days in the warm season, and two different years were chosen for each day. The random number generator in Microsoft® Excel© was used to create two sets of 153 numbers between and including 1989 and 2003. The resulting sets of years were assigned to each day in the warm season, such that there were essentially two-years worth of data in the data set. For example, the testing data set contains May 1 1992 and 2000, May 2 1998 and 1999, etc. All other dates were made part of the equation development data set. This random method was chosen to reduce the likelihood that any unusual convective season would bias the results.

The method of choice when creating regression equations for probability forecasts is logistic regression (Wilks 1995), given by the following equation:

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

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. This method was proven by Everitt (1999) to produce superior results when compared to linear regression. There were 13 candidate predictors available for the equations: the daily climatology, 1-day persistence, individual monthly flow regime lightning probabilities, and 10 XMR stability indices. The S-PLUS© v6 statistical software package (Insightful Corporation 2000) was used to develop and test the equations.

5.1 *Equation Development*

One equation was developed for each month in the warm season, for a total of five equations. The final predictors for each equation were selected from the set of candidate predictors using the following method. Each predictor was added one at a time to a logistic regression equation to determine its contribution to the reduction in residual deviance of the forecast. First, each of the predictors was tested as the lone variable in the equation and its contribution to the reduction in residual deviance determined. The variable with the largest contribution to the reduction in the residual deviance was chosen as the first predictor in the equation. Next, the other predictors were added individually with the first in a two-predictor set of equations. The second predictor that reduced the residual deviance by the largest amount in combination with the first was chosen for the equation. This iterative process continued for all 13 predictors. At times, the deviance explained for two or more variables was very similar. In these cases, individual equations were created using each of the predictors. As many as seven equations were created for each month in this manner. While more automatic predictor selection methods, such as principal component analysis (PCA), could have been employed to select an optimal combination of predictors, the manual process used here allowed for more control over understanding exactly how each individual predictor contributed to the residual deviance reduction. It was also facilitated by the small number of predictors available for selection.

Figure 6 shows the plot of reduction in residual deviance as each predictor was added for the August equation. The S-PLUS ANOVA (analysis of variance) function was used to determine the values in Figure 6. This function shows the reduction in residual deviance from that of an equation that produces a probability equal to the monthly climatological value (M Climo in Figure 6). As seen in Figure 6, KI reduced the residual deviance beyond the monthly climatology forecast by the largest amount (~20%), followed by the flow regime lightning probabilities (Flw Reg), TT, the daily climatologies (D Climo), SSI, etc.

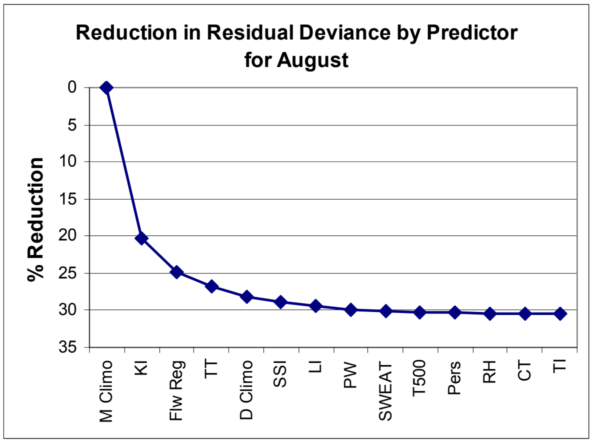

Figure 6. Plot of the reduction in residual deviance from a monthly climatology prediction (M Climo) as each predictor was added for the August equation. The percent reduction is on the y-axis and the names of each predictor are on the x-axis.

Using all the predictors would likely result in over-fitting the regression equations such that they would perform well with the development data but no other data sets. Therefore, only the candidate predictors that reduced residual deviance the most would be chosen for the equations. The final predictors for each equation were chosen in a two-step process. The first was to eliminate the predictors that created a residual deviance reduction of less than 0.5% based on a subjective analysis, close to where the slope of the curve in Figure 6 begins to flatten. Next, the Brier Score (BS) for the probability predictions from each equation were calculated for the development and testing data sets. The BS is calculated using the equation

$$BS = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2,$$

where n is the number of forecast/observation pairs, p_i is the probability forecast from the equation, and o_i is the binary lightning observation (Wilks 1995). Since there were two or more possible equations for each month, the equation that produced the lowest BS values for both the development and testing data sets was chosen as the final equation for the month.

Three predictors stood out in all five equations: the flow regime lightning probabilities, the smoothed daily climatology, and 1-day persistence. The flow regime probabilities and the daily climatology were used in every equation, while persistence was in every equation except for August. The mean RH in the 800-600 mb layer was the next most common predictor. The August equation contains the first five predictors (not including M Climo) in Figure 6: KI, Flw Reg, TT, D Climo, and SSI.

5.2 *Equation Testing*

The first test of the equations was whether or not they showed an improvement in skill over benchmark forecast methods. This involved calculation of the Brier Skill Score (SS) as

$$SS = \frac{BS - BS_{ref}}{BS_{perfect} - BS_{ref}},$$

where BS is the Brier Score of the equation being tested, BS_{ref} is the reference or benchmark forecast, and $BS_{perfect}$ is the Brier Score of a perfect forecast, which is always 0. Four methods were used as benchmark forecasts: the daily climatology, the monthly climatology, the flow regime probabilities, and 1-day persistence. Unfortunately, the NPTI forecasts were not readily available for comparison with these equations. However, it is known a priori that 1-day persistence outperforms NPTI by ~10%. If the equations show an improvement in skill over 1-day persistence, it can be concluded that they outperform NPTI as well.

The results with the testing data are in Table 2. The equations produce an increase in skill over all four forecast methods in all months, although the improvement values are mixed. It appears that the improvement over the daily climatology and flow regime probabilities is minimal in August.

Table 2. The percent (%) improvement in skill of the logistic regression equation forecasts over the benchmark forecasts of persistence, climatology, and flow regime probabilities. These results were calculated using the testing data.

Forecast Method	May	Jun	Jul	Aug	Sep
Persistence	31	53	38	39	43
Daily Climatology	27	18	27	7	21
Monthly Climatology	34	20	27	12	22
Flow Regime	34	13	20	3	21

The next test was to build a reliability diagram, which is used to show the performance of probability forecasts of binary events (Wilks 1995). Figure 7 shows the reliability diagram for the equation probability forecasts using the testing data set. The testing data for each month contained no more than 62 observations, so all months were combined so that the results would be more robust. The forecast probability is along the x-axis and the frequency of lightning occurrence for each probability value is along the y-axis. The pink curve represents perfect reliability and the blue curve is the reliability of the forecast equations. The inset rectangle shows the number of observations in each probability range used to calculate the reliability curve. That the blue line is below the pink line indicates that the equations consistently over-forecast lightning occurrence below probabilities of 0.4. This suggests that a bias correction technique might improve reliability at these lower forecast probabilities. The equations show good reliability at higher probability forecasts, except for 0.8. A detailed examination of the data revealed no clear pattern of why there was such a discrepancy at this value. It could be an artifact of the data set, and a larger data set may not exhibit such behavior.

In the final test, the equation probability forecasts for the testing data were stratified by lightning/non-lightning days, and the distributions of the probability values for each stratification were calculated. Once again, the forecasts for all months were combined to increase the size of the data set. Figure 8 shows the two probability distributions for lightning/non-lightning days. The blue curve for non-lightning days shows a peak of ~40% at forecast probability values of 0.2 then decreases to below 15% at 0.4, followed by a slight rise then a slow decrease to just below 10% at 1. This curve indicates a tendency to over-forecast as noted in the reliability diagram. The pink curve for lightning days shows low frequencies below 5% up to probability values of 0.4, then gradually increasing to 40% at 1, increasing above the non-lightning day curve at ~0.56 probability. This would show that probability forecasts above ~0.56 are more likely to be calculated on lightning days as opposed to non-lightning days.

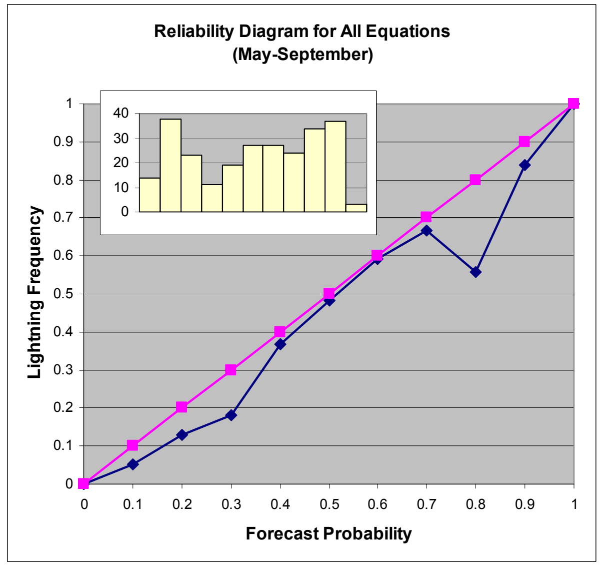

Figure 7. The reliability diagram of the probability forecasts for all months. The pink curve represents perfect reliability and the blue curve represents the reliability of the probability forecasts. The inset rectangle is the histogram showing the number of observations in each probability range.

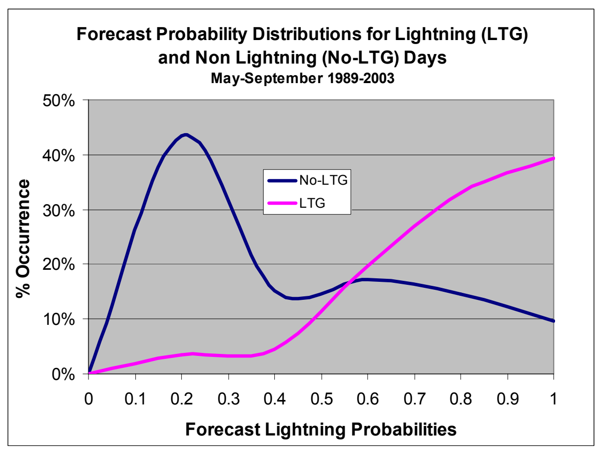

Figure 8. The forecast probability value distributions for lightning (pink) and non-lightning (blue) days in the testing data set. The y-axis values represent the frequency of occurrence of each probability value, and the values on the x-axis represent the forecast probability values output by the equations.

6. FUTURE WORK

The results from this study led to several ideas for future work. One involves using model output in the equations to develop weekly 7-day planning forecasts. Since the equations tend to over-forecast lightning occurrence, a bias-correction technique similar to that in Roeder (1998) could be developed and applied to increase the forecast skill. The 800-600 mb mean RH was one of the more common predictors chosen for the equations. The most appropriate layer for the mean RH may be different or could change from day to day. Future work would include a study on how to choose the most appropriate layer for mean RH. Finally, the role of synoptic-scale flow regime 1-day persistence should be explored.

7. SUMMARY AND CONTINUING WORK

Five logistic regression equations were created that predict the probability of lightning occurrence for the day during each of the five months in the warm season in the KSC/CCAFS area. All of the equations showed an increase in skill over the benchmark forecasts of daily and monthly climatology, persistence, and the flow regime lightning probabilities. As a result, the new equations will be added to the current set of tools used by the 45 WS to determine the lightning probability of occurrence for their daily planning forecast.

In order to use these equations, the 45 WS need an interface that will facilitate user-friendly input and fast output. A graphical user interface (GUI) is being developed using Microsoft® Excel© Visual Basic. The 45 WS is involved in the GUI development by providing comments and suggestions on the design. This will ensure that the final product will address their operational needs.

8. REFERENCES

Bauman, W. H., W. P. Roeder, R. A. Lafosse, D. W. Sharp, and F. J. Merceret, 2004: The Applied Meteorology Unit – Operational Contributions to Spaceport Canaveral. 11th Conference on Aviation, Range, and Aerospace Meteorology, Amer. Met. Soc., Hyannis, MA, 4-8 October 2004, 24 pp.

Cetola, J. D., 1997: A Climatology Of The Sea Breeze At Cape Canaveral, Florida, M. S. Thesis, Florida State University, Jul 97, pp 56

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].

Harms, D. E., B. F. Boyd, R. M. Lucci, and M. W. Maier, 1998: Weather Systems Supporting Launch Operations at the Eastern Range. AIAA 36th Aerospace Sciences Meeting and Exhibit, Reno, NV, 12-15 January 1998, Paper 98-0744, 11 pp.

Howell, C. L., 1998: Nowcasting Thunderstorms At Cape Canaveral, Florida Using An Improved Neumann-Pfeffer Thunderstorm Index, M.S. Thesis, Air Force Institute of Technology, AFIT/GM/ENP/98M-05, Mar 98, pp 93

Insightful Corporation, 2000: *S-PLUS© 6 User's Guide*, Insightful Corp., Seattle, WA, 470 pp.

Kelly, J. L., H. E. Fuelberg, and W. P. Roeder, 1998: Thunderstorm Predictive Signatures For The East Coast Sea Breeze At Cape Canaveral Air Station And The Kennedy Space Center, 19th Conf. on Severe Local Storms, 14-18 Sep 98, 677-680

Lericos, T. P., H. E. Fuelberg, A. I. Watson, and R. L. Holle, 2002: Warm season lightning distributions over the Florida Peninsula as related to synoptic patterns. *Wea. Forecasting*, **17**, 83 – 98.

Neumann, C. J., 1971: Thunderstorm forecasting at Cape Kennedy, Florida, utilizing multiple regression techniques. NOAA Technical Memorandum NWS SOS-8.

Roeder, W. P., 1998: Bias Correction to the Neumann-Pfeffer Thunderstorm Index, 45th Weather Squadron internal document, 45 WS/SYR, 1201 Edward H. White II St., MS 7302, Patrick AFB, FL 32925-3238, 21 Aug 98, 3 pp.

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

Wohlwend, C. S., 1998: Improving Cape Canaveral's Day-2 Thunderstorm Forecasting Using Meso-Eta Numerical Model Output, M. S. Thesis, Air Force Institute of Technology, AFIT/GM/ENP/98M-12, Mar 98, pp 148

NOTICE