University of Florida
Outreach Engineering Management Program

Master's Project

Global Positioning System (GPS) Precipitable Water in Forecasting Lightning at Spaceport Canaveral

Kristen Kehrer
Brian Graf

April 17, 2005

Mentor: Bill Roeder, Chief Staff Meteorologist, 45th Weather Squadron
Advisor: Dr. Edwin Romeijn, University of Florida

Table of Contents

I. Executive Summary

II. Background/Introduction
- a. Lightning in Central Florida
- b. Lightning impacts to Spaceport Canaveral
- c. Weather Predictions at Spaceport Canaveral
- d. Description of Global Positioning System (GPS) Precipitable Water (PW)
- e. Mazany Model

III. Data
- a. Sources
- b. Validation
- c. Synchronization

IV. Categorical Verification of the Mazany Model

V. 2-Hr Forecasting Tool
- a. Data Structure
- b. Model Selection Process
- c. Lightning Index
- d. New Logistic Regression Equation
- e. Independent Test

VI. 9-Hr Forecasting Tool
- a. Data Structure
- b. Model Selection Process
- c. Lightning Index
- d. New Logistic Regression Equation
- e. Independent Test

VII. Recommendations

VIII. References

IX. List of Figures
- Figure 1: Lightning by Year
- Figure 2: Lightning by Month
- Figure 3: Lightning Throughout the Day
- Figure 4: Linear Regression of the Data Set
- Figure 5: Logistic Regression of the Data Set
- Figure 6: Scale for Interpretation of K-Index
- Figure 7: Sample Scatter Plot of Precipitable Water
- Figure 8: Sample X-Bar Chart of Precipitable Water
- Figure 9: Contingency Table

Figure 10: Sample Timeline for the 2-Hr Forecasting Tool
Figure 11: Operational Utility Index vs. Lightning Index for the 2-Hr Forecasting Tool
Figure 12: Contingency Tables for the 2-Hr Forecasting Tool
Figure 13: KSS vs. Lightning Index for the 9-Hr Forecasting Tool
Figure 14: Contingency Tables for the 9-Hr Forecasting Tool

X. List of Tables

Table 1: Accuracy Measures and Skill Scores of Forecasts
Table 2: Comparison of Accuracy Measurements of GPS Lightning Index
Table 3: Comparison of Accuracy Measurements for lead times of 1.5 and 6 hours for the Mazany Model
Table 4: Candidate Independent Variables
Table 5: Comparison of Goodness-Of-Fit Statistics for all models for the 2-Hr Forecasting Tool
Table 6: Comparison of Accuracy Measurements and Skill Scores for the 2-Hr Forecasting Tool
Table 7: Comparison of Accuracy Measurements and Skill Scores at Thresholds of 0.50 and 0.32 for the 2-Hr Forecasting Tool
Table 8: Comparison of Model Data and Test Data for the 2-Hr Forecasting Tool
Table 9: Comparison of Goodness-Of-Fit Statistics for all models for the 9-Hr Forecasting Tool
Table 10: Comparison of Accuracy Measurements and Skill Scores for the 9-Hr Forecasting Tool
Table 11: Comparison of Accuracy Measurements and Skill Scores at Thresholds of 0.35 and 0.37 for the 9-Hr Forecasting Tool
Table 12: Comparison of Accuracy Measurements and Skill Scores at Thresholds of 0.50 and 0.37 for the 9-Hr Forecasting Tool
Table 13: Comparison of Model Data and Test Data for the 9-Hr Forecasting Tool

XI. Appendix: Course Documentation

Executive Summary

Using meteorology data, focusing on precipitable water (PW), obtained during the 2000-2003 thunderstorm seasons in Central Florida, this paper will, one, assess the skill and accuracy measurements of the current Mazany forecasting tool and, two, provide additional forecasting tools that can be used in predicting lightning.

Kennedy Space Center (KSC) and Cape Canaveral Air Force Station (CCAFS) are located in east Central Florida. KSC and CCAFS process and launch manned (NASA Space Shuttle) and unmanned (NASA and Air Force Expendable Launch Vehicles) space vehicles. One of the biggest cost impacts is unplanned launch scrubs due to inclement weather conditions such as thunderstorms. Each launch delay/scrub costs over a quarter million dollars, and the need to land the Shuttle at another landing site and return to KSC costs approximately $1M. Given the amount of time lost and costs incurred, the ability to accurately forecast (predict) when lightning will occur can result in significant cost and time savings.

All lightning prediction models were developed using binary logistic regression. Lightning is the dependent variable and is binary. The independent variables are the Precipitable Water (PW) value for a given time of the day, the change in PW up to 12 hours, the electric field mill value, and the K-index value.

In comparing the Mazany model results for the 1999 period B against actual observations for the 2000-2003 thunderstorm seasons, differences were found in the False Alarm Rate (FAR), Probability of Detection (POD) and Hit Rate (H). On average, the False Alarm Rate (FAR) increased by 58%, the Probability of Detection (POD) decreased by 31% and the Hit Rate decreased by 20%. In comparing the performance of the 6 hour forecast period to the performance of the 1.5 hour forecast period for the Mazany model, the FAR was lower by 15% and the Hit Rate was higher by 7%. However, the POD for the 6 hour forecast period was lower by 16% as compared to the POD of the 1.5 hour forecast period. Neither forecast period performed at the accuracy measures expected.

A 2-Hr Forecasting Tool was developed to support a Phase I Lightning Advisory, which requires a 30-minute lead time for predicting lightning. This tool resulted from forward stepwise model selection processes and contains four independent variables, specifically the 0.5-hr and 7.5-hr change in PW, the K-Index, and the current PW reading. These independent variables were considered to be significant in predicting lightning. A Lightning Index was established for this model at 0.32 because it provided the highest Operational Utility Index. Establishing the threshold at 0.32 increased the Operational Utility Index by 52.1% to 46.3%. Independent testing of the model showed minimal differences in the output, therefore validating the model.

A 9-Hr Forecasting Tool was developed to support lightning predictions for major extended outdoor activities, such as roll-out of the space shuttle to the launch pad. This tool resulted from forward stepwise model selection processes and contains five independent variables, specifically the 3.5-hr, 8.5-hr, and 12-hr change in PW, the K-Index, and the current PW reading. These independent variables were considered significant in predicting lightning. A Lightning Index was established for this model at 0.37 because it provided the highest Kuipers Skill Score (KSS) and best performance regarding FAR and Probability of False Detection (POD_f). Establishing the threshold at 0.37 increased the KSS by 29.2% to 36.8%. Independent testing of the model showed minimal differences in the output, therefore validating the model.

The lightning indices were selected based on maximizing specific skill scores. For the 2-Hr Forecasting Tool, the Operational Utility Index was determined to be most important. For the 9-Hr Forecasting Tool, the KSS was determined to be most significant. In both cases, the Lightning Index can be changed to adjust the mix of the amount of lightning forecasted to the amount of lightning not forecasted. This will adjust the various other accuracy measurements, including H, POD, and FAR.

Recommend further assessment of other independent variables that may be useful in predicting lightning with the necessary lead time to perform the forecasts. Use the forecasting tools developed and continually update actual observations are made available. Increase the frequency at which the values of the independent variables are collected to provide more accurate forecasts. Reassess the lightning indices based on the other accuracy measures and skill scores.

Background/Introduction

Lightning in Central Florida

During the months of May through September, the Central Florida area is known as the lightning capital of the North American Continent. During this period the density of lightning flashes per square kilometer per year ranges from 8 to 12. Since Central Florida encompasses approximately 11 counties with over 46,000 square kilometers, Central Florida experiences almost half a million lightning strikes a year. The period that most lightning occurs is during the month of July. With the amount of people and businesses now in the Central Florida area, the risk to people's lives and business resources is high. Kennedy Space Center (KSC) and Cape Canaveral Air Force Station (CCAFS) (to be referred to as Spaceport Canaveral), which are located in Central Florida, alone have 25,000 people, over $17 billion in facilities and flight and payload vehicle systems supporting the manned and unmanned Space programs. Given the high risk to people's lives and resources, the ability to predict when and where lightning will occur is extremely helpful and cost effective.

The graph below shows the amount of lightning that occurred each year during the 2000-2003 thunderstorm seasons in Central Florida. More lightning occurred in 2001 and 2003, with the least amount occurring in 2002.

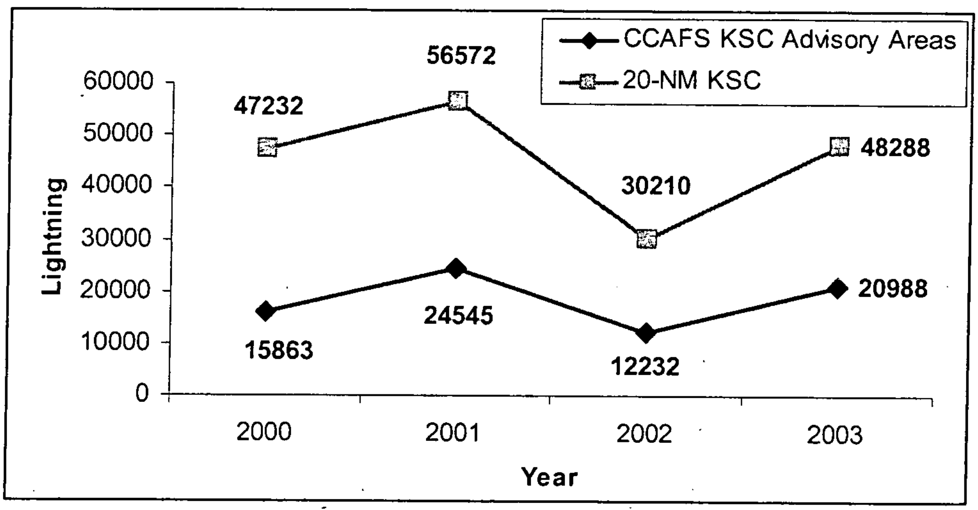

Figure 1: Lightning by Year

The graph below shows the lightning frequency by month during the thunderstorm season in Florida. There is more lightning in the 20-NM within KSC, because it is a much larger area than the KSC/CCAFS Advisory Area, although the lightning patterns are the same. As noted above, July is the month with the most lightning.

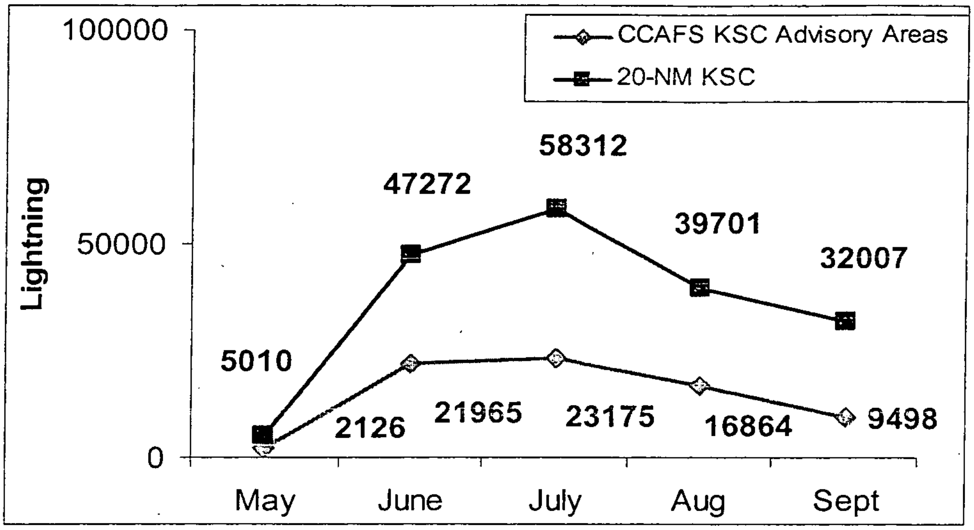

Figure 2: Lightning Distribution by Month

The graph below shows how lightning varies throughout the day. About 93% of lightning occurs between 1600 and 0200.

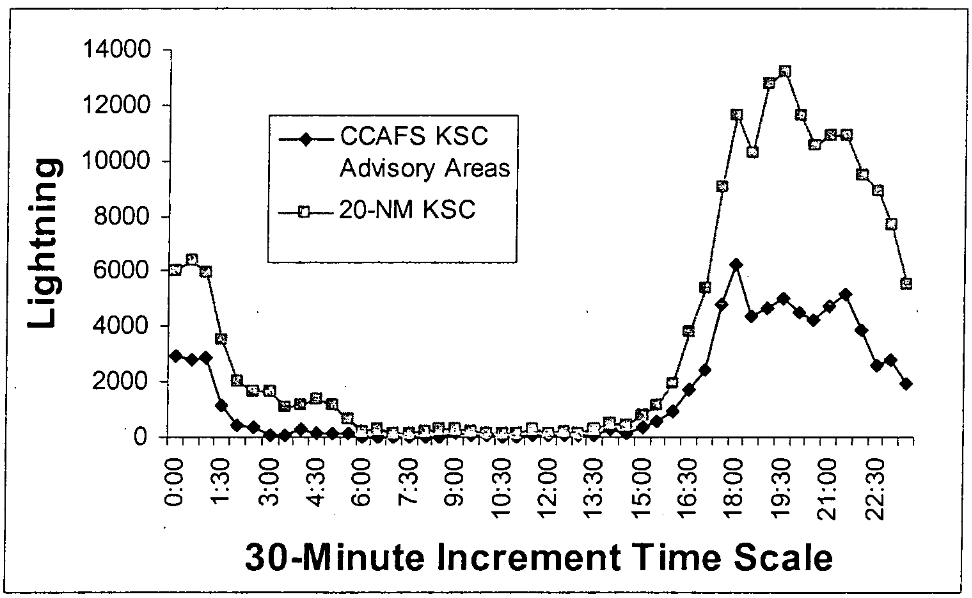

Figure 3: Lightning Throughout the Day

Lightning begins with the water molecule. The water molecule has an asymmetric distribution of charge that results in a permanent dipole moment. This unique structure results in a large latent heat associated with changes in the phase, making possible the phenomena of lightning (1). Specifically, lightning is a transient discharge of static electricity that serves to re-establish electrostatic equilibrium within a storm environment. Strong updrafts and down drafts occur with regularity, even within small thunderstorms. The updrafts transport water droplets up into the cloud, while ice particles descend from the frozen upper regions of the cloud. As they do, they bump and collide with each other. Through this process, electrons shear off of the ascending water droplets and collect on the descending ice particles. This generates an electric field within the cloud, with the top having a positive charge, and the bottom having a negative charge. An electric field is also generated between the bottom of the cloud and the surface of the earth.

In a developing storm cloud such as the example above, there is an electric attraction (i.e. electric field) between its top and bottom. As the charges separate, the field strength grows. The greater the magnitude of separation, the stronger the field, and the stronger the attraction is between the positively charged top and the negatively charged bottom. However, the negative charge at the base of the cloud induces a positive charge on the earth beneath it. When the electrical potential between two clouds or between a cloud and the earth reaches a sufficiently high value (about 10,000 volts (V) per centimeter (cm), the air becomes ionized along a narrow path and lightning flash results (2). The electric field mills located at Spaceport Canaveral measure this potential.

Lightning Impacts to Spaceport Canaveral

As mentioned above, Spaceport Canaveral has many personnel and expensive flight vehicles. Since the flight vehicle and payload systems are not common, the majority of the facilities and ground support equipment (GSE) are unique and therefore expensive and not easily replaceable. Spaceport Canaveral is located on the eastern side of Central Florida where the average density of lighting strikes per year is approximately 8 strikes per square kilometer. Even though the lightning density is not as high as other areas of Central Florida, given the value of people and the many unique facilities and GSE, the risk for loss of life or damage to facilities and GSE is high. Because of the high risk and the high potential for lightning occurring during the day hours from the months of May to September, operations such as the move of the Shuttle Launch Vehicle from the Vertical Assembly Building (VAB) to the launch pad during this period is performed during the early morning hours. This requirement to perform operations of this nature requires personnel in during third shift hours which results in significant costs. At least this is planned.

One of the biggest cost impacts is unplanned launch scrubs due to inclement weather conditions such as thunderstorms. Weather is the single greatest cause for launch delays and scrubs. Since the Shuttle began launch operation in 1980 and until 2004, about 30 percent of weather delays are related to lightning avoidance rules (3). Each launch delay/scrub costs over a quarter million dollars, and the need to land the Shuttle at another landing site and return to KSC costs approximately $1M. Given the amount of time lost and costs incurred, if the weather could have been forecast many hours before the actual launch was initiated, a significant cost savings and potential undue exercise of the flight and ground systems could have been avoided.

Current method for forecasting lightning

The Air Force 45th Weather Squadron operates from Range Weather Operations (RWO) at Cape Canaveral Air Force Station (CCAFS). The RWO is the center for the forecasting and detection of thunderstorms for Spaceport Canaveral. The RWO houses the Meteorological Interactive Data Display System (MIDDS), which analyzes data from the National Centers for Environmental Prediction, weather satellite imagery and local weather sensors (4). The data collected is applied to various numerical weather prediction models (1) to assist in putting Spaceport Canaveral area weather forecasts together.

The current method of forecasting thunderstorms does not directly predict lightning. The means for forecasting lightning is more accurately described as an early warning system which is based upon the occurrence of lightning within a certain area or the electrical charge in the atmosphere rising to levels significant to trigger lightning. Hence, the lightning prediction is based upon lightning detection. The RWO current methods of collecting data for atmospheric electrical activity are the Launch Pad Lightning Warning System (LPLWS), Lightning Detection and Ranging (LDAR) system, and the LLP Lightning Detection System (4).

The LPLWS is made up of 31 electric-field mills uniformly distributed throughout Spaceport Canaveral. They serve as an early warning system for electrical charges building aloft or approaching as part of a storm system. These instruments are ground-level electric field strength monitors. Information from the LPLWS gives forecasters information on trends in electric field potential and the locations of highly charged clouds capable of supporting natural or triggered lightning. The data are valuable in detecting early storm electrification and the threat of triggered lightning for launch vehicles (4).

The LDAR detects and locates lightning in three dimensions using a "time of arrival" computation on signals received at seven antennas. Each part of the stepped leader of lightning sends out pulses which LDAR receives at a frequency of 66 MHz. By knowing the speed of light and the locations of all of the antennas, the position of individual steps of a leader can be calculated to within 100-meter accuracy in three dimensions. LDAR provides between 1 and 1,500 points per flash. This is the only system currently able to provide detailed information on the vertical and horizontal extent of a lightning flash rather than just the location of its ground strike. LDAR detects all lightning including cloud-to-cloud and in-cloud as well as cloud-to-ground (4).

The LLP detects, locates and characterizes cloud-to-ground lightning within approximately 60 miles of the RWO. Electromagnetic radiation emitted from lightning is first detected by the system's three direction finder antennas located at Melbourne, Orlando, and in the northern area of KSC. Lightning positions are computed using triangulation from two of the sites, and relayed to a color display video screen in the RWO. Once lightning-producing cells are identified and located, it becomes easier for the forecaster to predict just where the next lightning bolts will hit (4).

These primary lightning detection systems, LPLWS, LDAR and LLP, along with other atmospheric condition and weather prediction systems and numerical prediction models, are the primary Range Weather Operations thunderstorm surveillance tools for evaluating weather conditions that lead to the issuance of lightning warnings.

With the lightning warnings available from the RWO, Spaceport Canaveral created a two-phase (Phase I and Phase II) lightning warning policy for operations performed at Spaceport Canaveral. Specifically, Phase I is deemed an advisory and is issued when lightning is forecast within five miles of the designated site and within 30 minutes from the issuance of the Advisory. The 30-minute warning gives personnel in unprotected areas time to get to protective shelter and gives personnel working on lightning sensitive tasks time to secure operations in a safe and orderly manner. Phase II is identified as a Warning and is issued when lightning is imminent or occurring within five miles of the designated site. All outdoor and lightning-sensitive operations are terminated until the Phase II Warning is lifted. This two-phase policy provides adequate lead time for sensitive operations without shutting down less sensitive operations until the hazard becomes immediate (4).

Description of Global Positioning System (GPS) Precipitable Water (PW)

An important water vapor parameter currently being obtained from satellite and radiosonde measurements is precipitable water. Precipitable water is the total atmospheric water vapor contained in a vertical column of unit cross-sectional area extending from the surface of the earth to the top of the atmosphere. Precipitable water is commonly expressed in terms of the height to which that water substance would stand if completely condensed and collected in a vessel of the same unit cross section. In thunderstorms the amount of rain very often exceeds the total precipitable water vapor of the overlying atmosphere. This results from the action of convergence that brings into the rainstorm the water vapor from a surrounding area that is often quite large. Nevertheless, there is general correlation between precipitation amounts in given storms and the precipitable water vapor of the air masses involved in those storms. Climatologies of PW are currently being compiled using measurements from the Global Positioning System (GPS) and the operational network of radiosondes (6).

Since the inception of space geodesy, the tropospheric delay of signals propagating through the atmosphere of the Earth has affected geodetic estimations of coordinates of points on the surface of the Earth. The amount of precipitable water vapor (PW) contained in the neutral atmosphere can be inferred from the propagation delay of Global Positioning System (GPS) signals passing through the troposphere. Recent research has shown that the estimates of the wet tropospheric delay from very long baseline GPS observations agree closely with estimates from radiosonde launches. Mathematical techniques have been developed to map the delay at any elevation to delay in the zenith (or vertical) direction, and the removal of the tropospheric delay by estimation has become an integral part of precise GPS analyses (7).

The Global Positioning System consists of a constellation of satellites which transmit on two *L*-band frequencies (1575.42 MHz for *L*1 and 1227.6 MHz for *L*2). These two signals are delayed as they propagate through the atmosphere due to the presence of atmospheric water vapor. This "wet delay" is detectable in geodetic analyses of GPS phase observations and can be transformed into an estimate of the PW present in the troposphere. Recent studies from small-scale networks

(~50 km) have demonstrated that PW can be estimated from GPS observations with an accuracy of better than 2 mm relative to a fixed GPS station (7).

Mazany Model

A study was conducted in 2000 that sought to develop a lightning prediction index that utilizes GPS PW. The model was developed and paper was written by representatives from the University of Hawaii, National Oceanic and Atmospheric Administration (NOAA), and Patrick Air Force Base. The paper is entitled "A Lightning Prediction Index that Utilizes GPS Integrated Precipitable Water Vapor" (1).

The Mazany lightning prediction model was developed using binary logistic regression. Ideally, a linear regression model could have been used to model the relationship between the explanatory variable (PW) and the dependent variable (lightning). Furthermore, if linear regression was able to be applied, the amount of linear relation between the variables could have been additionally measured by determining the covariance, or correlation. However, since the observed outcome of lightning is restricted to two values (i.e., Yes Lighting did Occur, 0 or No Lightning Occurred, 1), the dependent variable is binary. With a binary dependent variable, it would be very difficult to determine the correlation coefficient since the coefficient has values between -1 and 1. Hence, the logistic regression model is the most viable model to use in predicting lightning. The graph below represents why a linear regression model does not apply in this case.

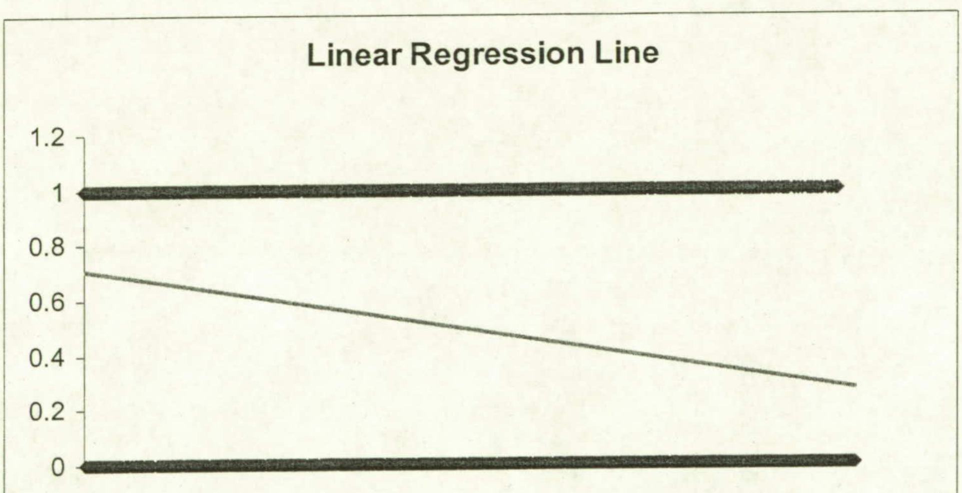

Figure 4: Linear Regression of the Data Set

Changing from a linear regression to a logistic regression provides a better fit of the binary data set.

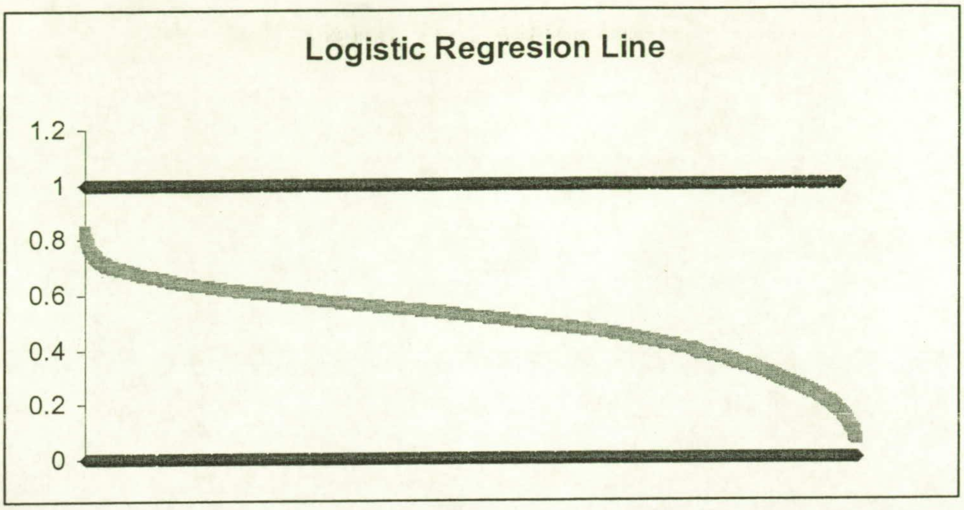

Figure 5: Logistic Regression of the Data Set

The Mazany model contains four independent variables, specifically the PW value for a given time of the day, a 9 hr change in PW, electric field mill value, and the K-index as significant independent variables. The dependent variable is a binary variable, with zero indicating a lightning strike and a one indicating no lightning. The output of the model is a lightning index between zero and one that indicates conditions for lightning. The lightning index is compared to thresholds to determine if and when lightning will occur. The model is below:

$$\hat{y} = \frac{1}{1 + \exp(-6.7866 + 0.0011359x_1 + 0.06063x_2 + 0.32341x_3 + 0.06728x_4)}$$

Where,
x_1 = Electric Field Mill Reading (V m^{-1})
x_2 = PW (mm)
x_3 = 9-h ΔPW (mm)
x_4 = K-index
$\hat{y}$ = Lightning Index

The lightning index is then compared against established thresholds for interpretation as shown below:

0.7 – 1.0: No lightning in the next ~ 6 hrs
0.6 – 0.7: Lightning expected in the next ~6 hrs
0.0 – 0.6: Lightning expected in the next ~1.5 hrs

Data

Data Sources

Data from four thunderstorm seasons, May 1 – September 30, 2000 – 2003 was used in the analysis. Electric field mill values are obtained from 31 electric field mills located throughout Spaceport Canaveral. Electric field mills measure the electric potential of the atmosphere in volts per meter every 5 minutes. Precipitable water level data was obtained from a GPS receiver site located on the Cape Canaveral Air Force Station.

Weather balloons carry instrument packages called radiosondes high into the atmosphere to gather essential upper-air data needed to forecast the weather. Typically, these balloons/instruments are released twice a day at the many sites around the world. Since Spaceport Canaveral is the "lightning capital" of the North American continent, the weather balloon is released three times a day during the lightning season. Temperature, humidity and air pressure are measured at various altitudes and transmitted via radio waves to a receiving station. Radio navigation supplies wind speed and direction at each altitude (5).

These weather balloons provide the data necessary to calculate the K-index. The K-index is a measure of the thunderstorm potential, according to the scale and color fill scheme shown in Figure 6. The K-Index is determined on a skew-t thermodynamic diagram. K-index represents the thunderstorm potential as a function of vertical temperature lapse rate at 850mb temperature and 500mb temperature, low level moisture content at 850mb dew point, and the depth of the moist layer at 700mb dew point. With the temperatures determined at these data points the K-index is calculated using the following linear equation: KI= (T850mb- T500mb) + Td850mb - (T700mb -Td700mb). The K-Index is interpreted using the chart below:

0-15	No thunderstorms
18-19	Thunderstorms unlikely
20-25	Isolated thunderstorms
26-30	Widely scattered thunderstorms
30-35	Numerous thunderstorms
36-39	Thunderstorms very likely
40+	100% chance of thunderstorms

Figure 6: Scale for Interpretation of K-Index

Two sets of lightning data were used in the analysis. The first set of lightning data included lightning observations that fell within 20 nautical miles from the Vertical Assembly Building (VAB) at the Kennedy Space Center. This is the data used to validate the Mazany lightning prediction model as well as to develop the 9-Hour Forecasting Tool. The Mazany model was developed using thunder heard by the weather observer at the KSC Shuttle Landing Facility, which is near the VAB. Since thunder is typically heard 10-15 miles away, lightning detected within 20 NM of KSC is a close proxy to ground truth for verifying the Mazany model. This data is also used for the 9-Hour model, because this model will mainly be used to forecast lightning during shuttle roll-out from the VAB to the launch pad. A second set of lightning data was used which indicated lightning from within a rectangular area enclosing the 45 WS CCAFS/KSC lightning warning circles. This data is currently used to support the current Phase I/II Lightning Advisory System, and therefore will be used to support the development of the new 2-Hour Forecasting Tool that will support lightning advisories in the future.

Data Validation

Scatter plots of the precipitable water and K-index data were developed to identify patterns and outliers in the data. Figure 7 below shows a sample scatter plot of precipitable water data from the 2003 thunderstorm season. A significant number of values at or below zero for precipitable water were flagged as errors in the data set and eliminated from consideration in the development and validation of the models. Similar scatter plots were built for the remaining years in the PW data set, as well as for the K-index data set, to identify outliers and errors.

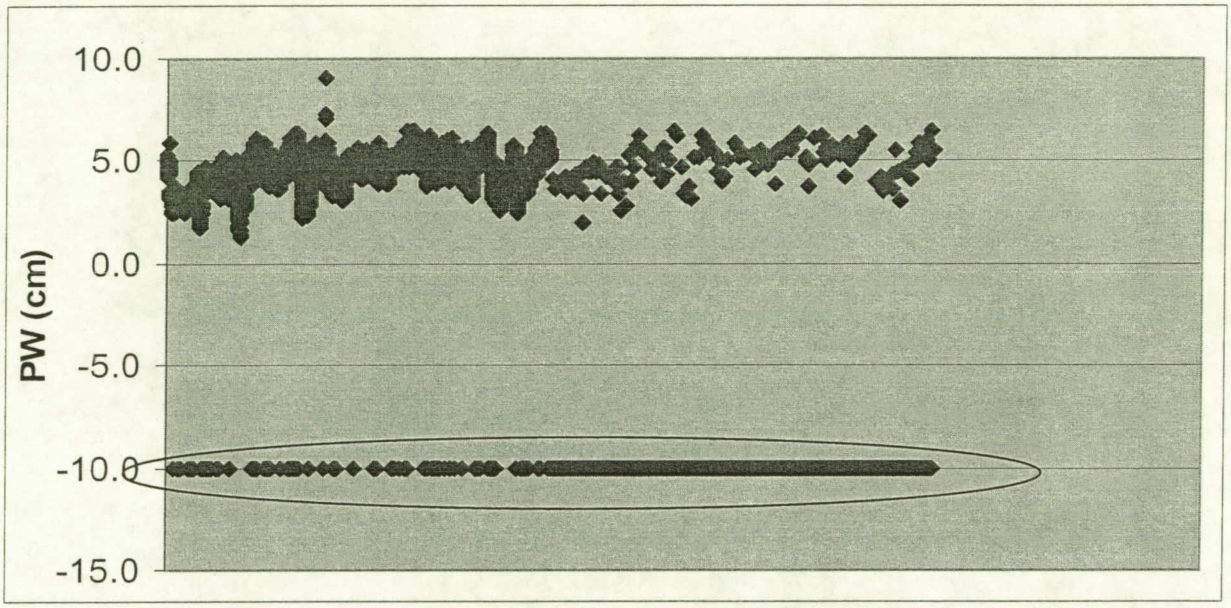

Figure 7: Sample Scatter Plot of Precipitable Water

After initial data points were eliminated using scatter plots, Statistical Process Control (SPC) charts were used to identify points outside three standard deviations of the mean of the data set. Specifically, X-bar charts were used to identify X-bar, Upper Control Limits (UCL), and Lower Control Limits (LCL) for each year. SPC X-bar charts were created for both precipitable water and K-Index data sets for each year. Below is an SPC X-bar chart for the precipitable water data set from the 2003 thunderstorm season. Similar SPC charts were built for the remaining years in the PW data set, as well as for the K-index data.

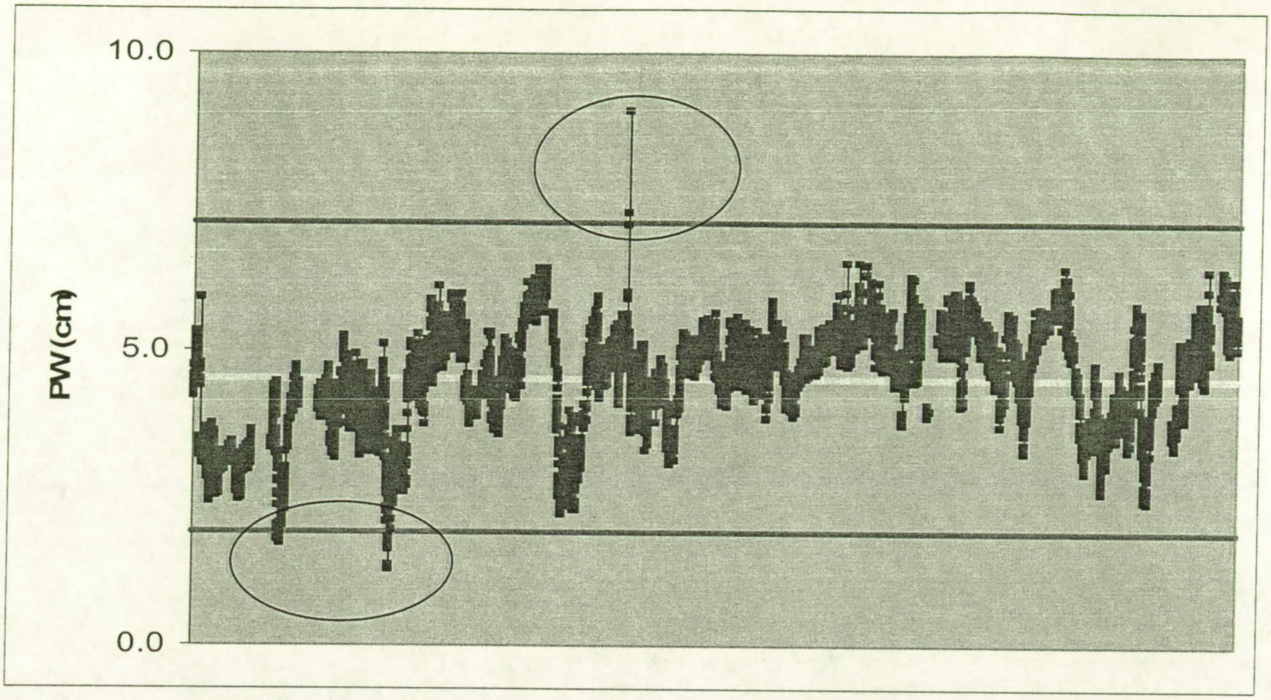

Figure 8: Sample SPC X-bar Chart of Precipitable Water

Points that fell below the LCL and above the UCL were flagged as outliers and further examined against meteorological conditions surrounding the data points. In some cases, abnormal weather conditions could drive unusually low or high K-index and precipitable water data points that should be considered in the model. However, if no unusual meteorological conditions surrounded the data point, it was considered an error and eliminated from consideration in the development and validation of the models. In total, 34 points were eliminated from the precipitable water data set and none were eliminated from the K-index data set.

The K-Index variable showed extensive variability about the mean, and therefore a substantial number of points fell outside of the control limits and were flagged as outliers. Upon further examination of the K-Index values collected between 1989 and 2003, a range of -33.8 to 44.3 had been observed. A meteorological team at the 45^{th} Weather Squadron reviewed the data and determined that the outliers were actually valid data points that should be considered in the development and validation of the models. The team noted the only data points to remove, or not consider in the model, were those data points having a value of -99.9. This value was a default value for times in which no data was collected.

Data Synchronization

Time differences in the collection of data drove the requirement to synchronize the data with the PW collection points. Electric field mill data is collected every five minutes; PW values are collected on the half hour each day; K-Index values are collected three times per day at varying intervals; and lightning data occurs randomly throughout the thunderstorm season.

The PW data point collected represented a centered time stamp, which means that the reported data time is the middle of the half-hour that the sensor spends looking at the atmosphere. For example, if the GPS-PW time stamp for a certain day is 1215, the data represents the sensor looking at the atmosphere and averaging the results from 1200-1230. To synchronize the lightning data with the PW data, the lightning was categorized into 30-minute groupings that aligned with the corresponding 30-minute GPS-PW interval. In our previous example, if lightning occurred between 1200 and1230, it was mapped to the 1215 GPS-PW data point.

The K-Index was synchronized with the GPS-PW data using two methods. To verify the existing model, the K-Index was interpolated between existing observations. Therefore, the K-Index was assumed to climb or fall at a continuous rate between the two observations. The K-Index was matched with the corresponding GPS-PW observation timeframe. To create the new predictive models, interpolation of the K-Index could not be used because the lag in data collection points would not provide the predictive capability necessary. In this instance, the K-Index was assumed to be constant between data collection points. In this case again, the K-Index was matched with the corresponding GPS-PW observation timeframe.

The electric field mill data was not synchronized with the PW, K-Index, or lightning data. A constant value of 300 V/m for the electric field mill value was used to validate the Mazany model. The electric field mill value is only important less than 30-60 minutes prior to a lightning strike, which falls inside the forecast intervals in which the new models perform. Also, the Mazany model is very insensitive to the electric fields and thus has little impact on the outcome.

Categorical Verification of Mazany Model

Actual observations of the independent and dependent variables from the 2000-2003 thunderstorm seasons were used to validate the Mazany Model. The process for verifying the model used the categorical forecasts of discrete predictands. The term categorical means the forecast consists of a flat statement that one and only one of a set of possible events occurs. In this model the categorical forecast "yes" lighting will occur and "no" lightning will occur. A discrete predictand is an observable variable that takes on one and only one of a finite set of possible values (8). In this model there are two discrete predictands, or possible outcomes, lightning did occur, "yes", and lightning did not occur, "no".

The categorical verification data is displayed in a 2 x 2 contingency table of the forecasts and observed events (8). Figure 9 shows the contingency table of counts of the possible combinations of the forecast and observed events pairs, or forecast/event pairs.

		Observed	
		Yes	No
Forecast	Yes	a	b
	No	c	d

Figure 9: Contingency Table

The possible forecast/event pairs a, b, c, and d are defined as follows:

a = Event Predicted to Occur and did Occur = Lightning Predicted to Occur and did Occur
b = Event Predicted to Occur but did Not Occur = Lightning Predicted to Occur but did Not Occur
c = Event Predicted to Not Occur but did Occur = Lightning Predicted to Not Occur but did Occur
d = Event Predicted to Not Occur and did Not Occur = Lightning Predicted to Not Occur and did Not Occur

The total number of forecast/event pairs in the data set is $n = a + b + c + d$ (8).

There are several measures to determine the accuracy of the categorical forecasts. The first measure is the hit rate. The hit rate determines the proportion of forecasts that were correct. The categorical forecast correctly forecast the event ("yes" lightning did occur) or nonevent ("no" lightning did occur). The hit rate is determined as follows:

$$H = (a + d)/n$$

The scale of the hit rate is from zero to one. The most accurate forecast is when the hit rate, H, is equal to one (8).

The second measure typically used for categorical verification data is the POD. The POD is the fraction of those occasions when the forecast event occurred on which the event was also forecast. For the event of lightning occurring, lightning was forecasted to occur and lightning occurred. Hence, the probability of detection is determined as follows:

$$POD = a/ (a + c)$$

The scale of the POD is from zero to one. The most accurate forecast is when the POD is equal to one (8).

Complementary to the POD is the POD_f which is the probability of false detection. This measures the probability that lightning did not occur given that lightning was predicted.

$$POD_f = b/ (b + d)$$

The scale of POD_f is from zero to one. The most accurate forecast is when POD_f is equal to zero.

The final measure commonly used to measure the accuracy of the forecasts is the FAR. The FAR is the proportion of forecast events that fail to occur. For the event of lightning occurring, lightning was forecasted to occur but lightning did not occur. Hence, the false-alarm rate is determined as follows:

$$FAR = b/ (a + b)$$

The scale of the FAR is from zero to one. The most accurate forecast is when the FAR is equal to zero (8).

The means of assessing, or measuring, the performance of the forecast model is to compute relative accuracy measures. These types of measures are called skill scores and are scalar measures of the forecast model performance. One such measure of skill is the Hansen-Kuipers skill score (KSS) (also identified as the true skill statistic (TSS)). The KSS is based upon the basic accuracy measure of the hit rate obtained by random forecasts that are constrained to be unbiased and referenced in the denominator. Based upon these assumptions, the Hansen-Kuipers skill score is determined as follows:

$$KSS = (ad - bc)/ (a + c) (b + d)$$

The most accurate forecast is a score of one, random forecasts score zero and forecasts inferior to the random forecasts receive negative scores (8).

Another means of measuring the accuracy of the forecasts is the Operational Utility Index (OUI). The OUI was developed by the 45th Weather Squadron to measure the skill of their forecasts. The Operational Utility Index is calculated as shown below:

$$OUI = (3*POD + 2*KSS - FAR) /6$$

The index places the most importance, and therefore weight, on POD and KSS.

Table 1 shows the results of the predictive capability of the current statistical model for the years 2000 through 2003 from May 1 to September 30. Based upon the Lightning Index ($\hat{y}$) value of 0.6 – 0.7 indicating lightning expected in the next, approximately 6 hours, and the value of 0.0 – 0.6 indicating lightning expected in the next approximately 1.5 hours, the categorical verification was performed on lighting forecast to occur in the next 6 hours and lighting forecast to occur in the next 1.5 hours, respectively. (Note, all values ranges are between zero and one. For ease of review, the percentage value was used.)

Lightning Forecasts in next 6 and 1.5 Hours

Year	Forecast period	Forecast	Observed Yes	Observed No	Hit	POD	POD_f	FAR	KSS	OUI
2000	6 Hours	Yes	186	232	68%	29%	56%	56%	14%	10%
		No	447	1281						
	1.5 Hours	Yes	370	946	63%	51%	72%	72%	17%	19%
		No	353	1796						
2001	6 Hours	Yes	344	559	69%	35%	62%	62%	15%	12%
		No	646	2291						
	1.5 Hours	Yes	682	2004	62%	57%	75%	75%	20%	23%
		No	508	3332						

Year: 2002									
		Observed							
6 Hours		Yes	No	Hit	POD	POD_f	FAR	KSS	OUI
Forecast	Yes	241	446	75%	58%	65%	65%	36%	30%
	No	178	1584						
		Observed							
1.5 Hours		Yes	No	Hit	POD	POD_f	FAR	KSS	OUI
Forecast	Yes	345	1240	64%	63%	78%	78%	28%	28%
	No	202	2248						
Year: 2003									
		Observed							
6 Hours		Yes	No	Hit	POD	POD_f	FAR	KSS	OUI
Forecast	Yes	381	582	69%	48%	60%	60%	24%	22%
	No	406	1773						
		Observed							
1.5 Hours		Yes	No	Hit	POD	POD_f	FAR	KSS	OUI
Forecast	Yes	587	2026	59%	62%	78%	78%	20%	25%
	No	354	2788						

Table 1: Accuracy Measure and Skill Scores of Forecasts

The results of the accuracy measures for the 1.5 hour forecast period can now be compared to the accuracy measures of the tests results of the Mazany Model forecasts performed for the time period of June 10 to September 26, 1999, which was identified as Period-B. The Period-B forecasts were based upon the index value falling below 0.7 and 1.5 hours prior to the first strike. The accuracy measures applied to the initial tests results of the Mazany Model forecasts were FAR, POD and H. Table 2 compares 1999 Period-B with the four thunderstorm seasons from the year 2000 to 2003 from May 1 to September 30.

1.5 Hour Forecast Period	1999 Period B	2000	2001	2002	2003
FAR	16.7 %	71.9 %	74.6 %	78.2 %	77.5 %
POD	89.3 %	51.2 %	57.3 %	63.1 %	62.4 %
H	82.6 %	62.5 %	61.5 %	64.3 %	58.6 %

Table 2: Comparison of Accuracy Measurements of GPS Lighting Index

The GPS lightning index accuracy measure results for all four, full period thunderstorm seasons yield results not well compared to the results of the 1999 Period B. On average, the FAR increased by 58%, the POD decreased by 31% and the Hit Rate decreased by 20%. Even though the accuracy measures indicate the Mazany model for forecasting lightning events 1.5 hours prior to the first strike is not a reliable forecasting method, the encouraging aspect of the model results were the consistency of the measures. This may provide future related projects the focus on the types of predictands to use and the lead time period to use.

Regarding the lead time for forecasting, as was determined in the Mazany paper, the performance of the GPS lightning index lead time with regard to the timing of the first strike follows approximately a normal distribution. With the range of lead time before the first strike varying between 0 to 12 hours, the nominal lead time is approximately 6 hours. Based upon this finding, the GPS lighting index was also applied to the 6 hour forecast period. When compared to the 1.5 Hour forecast period the Mazany model did better overall. Table 3 compares the average 1.5 hour forecast period accuracy measures with the 6 hour forecast period accuracy measures.

6 Hour Forecast Period	Average 1.5 Hour Forecast for 2000-2003	2000	2001	2002	2003
FAR	75.6 %	55.5 %	61.9 %	64.9 %	60.4 %
POD	58.5 %	29.4 %	34.7 %	57.5 %	48.4 %
H	61.7 %	68.4 %	68.6 %	74.5 %	68.6 %

Table 3: Comparison of Accuracy Measurements for lead times of 1.5 and 6 hours for the Mazany Model

On average, the FAR decreased by 15%, the POD decreased by 16% and the Hit Rate increased by 7%. The decrease of the FAR and increase of the Hit Rate were encouraging results. The decrease of the POD was not a good indicator of the overall accuracy of the model. Even though the results were not as expected, the encouragement is the consistent results of the accuracy measures.

2-Hr Forecasting Tool

The 2 Hr Forecasting Tool seeks to predict the occurrence or non-occurrence of lightning to support Phase 1 lightning advisories. A 2- hour period was chosen to support the current process for issuing lightning advisories. This accounts for a 15-minute lag-time in receiving the PW readings, 45 minutes to process the information, and 30 minutes desired lead time to support Phase 1 Lightning Advisories. When the forecaster receives the PW reading, it is already 30-minutes behind the time-stamp or 15-minutes behind the dwell time. After receiving the PW reading, the operator will use the model to predict the occurrence or non-occurrence of lightning. That information is then communicated to the 45th Weather Squadron Weather Officer, who will then compare that data with other current weather information. The process allows 45 minutes for performing this analysis. Finally, the Weather Officer will make a decision as to whether or not to issue a Phase 1 Lightning Advisory with the desired 30-minute lead time.

If, for example, the PW is time-stamped at 1215, the forecaster will receive the PW reading at 1245. When the operator receives the PW information, the target period will be between 1400 and 1430. The operator will process the data and communicate the information to the 45th Weather Squadron Weather Officer at or before 1315. The Weather Officer will compare the data with other current weather conditions, which will take approximately 15 minutes. At 1330, the Weather Officer will make the decision as to whether or not to issue a Phase 1 Lightning Advisory at 1400 for lightning between 1400 and 1430. Below is a timeline depicting this example:

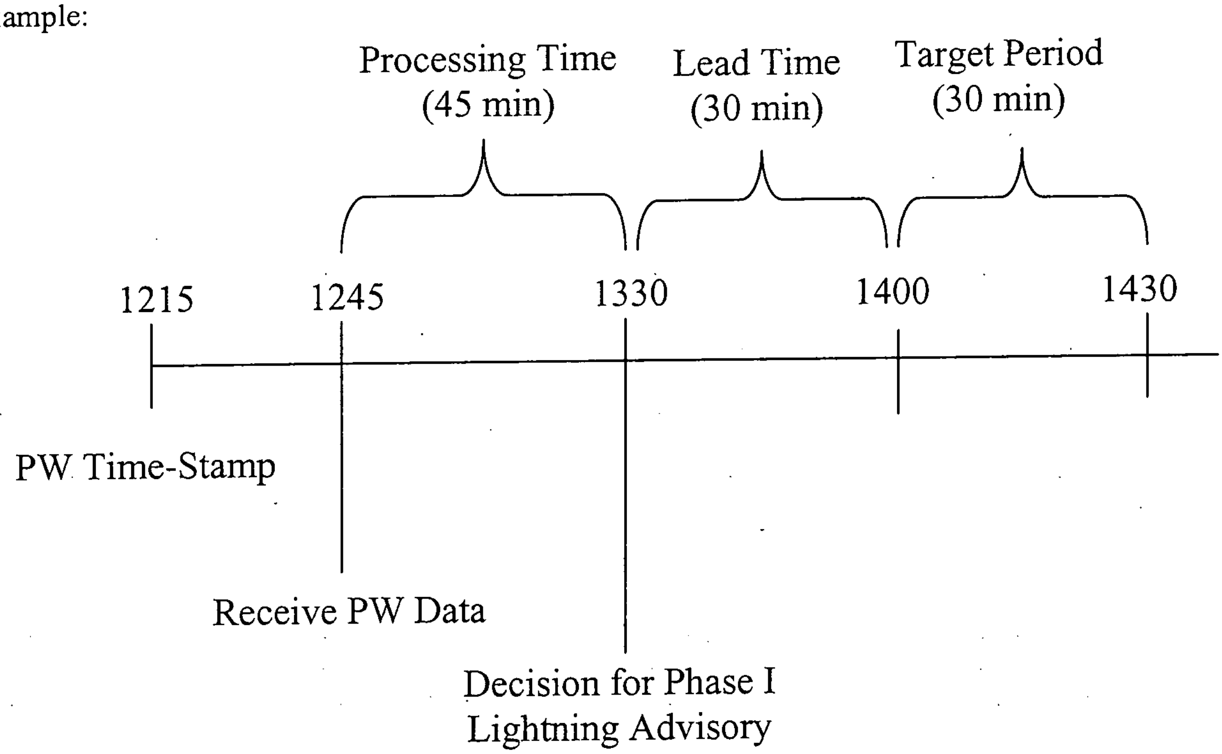

Figure 10: Sample Timeline for 2-Hr Forecasting Tool

Data Structure

The data set consisted of 26 candidate independent variables that are shown in the table below.

Precipitable Water (PW) (cm)	6.5 Hr Δ PW (cm)
0.5 Hr Δ PW (cm)	7.0 Hr Δ PW (cm)
1.0 Hr Δ PW (cm)	7.5 Hr Δ PW (cm)
1.5 Hr Δ PW (cm)	8.0 Hr Δ PW (cm)
2.0 Hr Δ PW (cm)	8.5 Hr Δ PW (cm)
2.5 Hr Δ PW (cm)	9.0 Hr Δ PW (cm)
3.0 Hr Δ PW (cm)	9.5 Hr Δ PW (cm)
3.5 Hr Δ PW (cm)	10.0 Hr Δ PW (cm)
4.0 Hr Δ PW (cm)	10.5 Hr Δ PW (cm)
4.5 Hr Δ PW (cm)	11.0 Hr Δ PW (cm)
5.0 Hr Δ PW (cm)	11.5 Hr Δ PW (cm)
5.5 Hr Δ PW (cm)	12.0 Hr Δ PW (cm)
6.0 Hr Δ PW (cm)	K-Index (step)

Table 4: Candidate Independent Variables

To assess the probability of lightning in 2 hours, the data set had to be modified. The 2-Hr Forecasting Tool seeks to determine the probability that lightning will occur in two hours. All independent variables were shifted 2 hours ahead and aligned with the value of the dependent variable at that point in time. The dependent variable lightning was not shifted. This associates the conditions for the latest observed precipitable water level, the change in precipitable water levels, and the latest K-Index observation with the occurrence or non-occurrence of lightning in 2 hours. After the data was structured to properly reflect the occurrence of lightning, the data sets that did not have a value for each variable were removed. This meant a reduction in data points from 29,376 to 13,450.

Model Selection Process

Six methods were used for model selection: Forward Stepwise Conditional, Forward Stepwise Likelihood Ratio, Forward Stepwise (Wald), Backward Stepwise Conditional, Backward Stepwise Likelihood Ratio, and Backward Stepwise (Wald). When the initial models were generated, none of the models predicted the occurrence of lightning, yet it was correct 93.3% of the time. This occurred because out of 13,450 total data points, lightning only occurred 890 times. Therefore, by never predicting lightning, the model was correct 93.3% of the time. This model produced a favorable FAR and POD_f of 0%, because lightning was never predicted. However, POD=0%, KSS =0%, and the Operational Utility Index = 0%. To account for this instance, the data set once again had to be modified.

Only a subset of the data was used to include half of the data points with the value of the independent variable at one and the other half with the value of the independent variable at zero. This reduced the size of the data set from a total of 13,450 data points to a total of 1802 data points, of which 890 represented a one for lightning.

Goodness-of-Fit statistics were initially calculated to evaluate the model fit. Specifically, Cox & Snell R-squared, Nagelkerke R-Squared, and the Hosmer-Lemeshow goodness-of-fit $\chi 2$ were used as an assessment of model fit. A comparison of model outputs is shown below:

Model Selection Method	Cox & Snell R Square	Nagelkerke R Square	Hosmer-Lemeshow Chi-square
Forward Conditional	0.0773	0.1031	19.0867
Forward LR	0.0773	0.1031	19.0867
Forward Wald	0.0773	0.1031	19.0867
Backward Conditional	0.0807	0.1076	15.1618
Backward LR	0.0807	0.1076	16.3335
Backward Wald	0.0807	0.1076	15.1617

Table 5: Comparison of Goodness-Of-Fit Statistics for all models for the 2-Hr Forecasting Tool

As you can see, there appears to be a lack of fit based on the values of Cox & Snell R-squared, Nagelkerke R-Squared, and the Hosmer-Lemeshow goodness-of-fit $\chi 2$. However, this was not considered to be an issue for several reasons. First, the R-Squared values from a logistic regression are not the same as the R-Squared values calculated in a linear regression model. Therefore, they are not proven to be good measures of model fit. Second, these values are not relevant in measuring the model's utility, because they are evaluated against a threshold of 0.5. This means that outputs above 0.5 were predicted as lightning, while values below 0.5 were predicted as no-lightning. The utility of the model will be evaluated based on an optimized threshold.

In selecting the model, the most weight was given to the values of the accuracy measurements and skill scores discussed above. These ratios include the Hit Rate (H), False Alarm Rate(FAR), Probability of Detection (POD), Probability of False Detection (POD_f), Hansen Kuipers or True Skill Score (KSS), and an Operational Utility Index. The Operational Utility Index is considered to be the most critical factor, because this ratio applies more weight to the POD, which is critical when there are a large number of personnel that are affected. The purpose of a Phase 1 Lightning Advisory is to ensure personnel performing outdoor operations have time to seek shelter.

These accuracy measurements were calculated based on varying the lightning index from 0.0 to 1.0. Note that all forward methods of model development process yielded the same model, and all backward methods of model selection yielded the same model. They are grouped into two categories in the table below: forward methods and backward methods. Again, the objective is to maximize the Operational Utility Index.

Model Selection Method	Index	Hit	POD	FAR	KSS	Operational Utility Index	POD_f
Forward Methods	0.0	49.7%	100.0%	50.3%	0.0%	41.6%	100.0%
Backward Methods	0.0	49.7%	100.0%	50.3%	0.0%	41.6%	100.0%
Forward Methods	0.1	50.2%	100.0%	50.1%	0.9%	41.9%	99.1%
Backward Methods	0.1	50.0%	100.0%	50.2%	0.5%	41.8%	99.5%
Forward Methods	0.2	52.6%	99.5%	48.8%	5.8%	43.6%	93.7%
Backward Methods	0.2	53.1%	99.4%	48.5%	6.8%	43.9%	92.6%
Forward Methods	0.3	57.4%	96.9%	46.0%	15.2%	45.8%	81.7%
Backward Methods	0.3	57.3%	96.5%	46.1%	15.0%	45.6%	81.5%
Forward Methods	0.4	59.8%	87.8%	43.9%	19.9%	43.3%	67.9%
Backward Methods	0.4	60.1%	87.2%	43.6%	20.6%	43.2%	66.7%
Forward Methods	0.5	58.3%	64.2%	42.9%	16.6%	30.5%	47.6%
Backward Methods	0.5	60.8%	66.5%	40.5%	21.7%	33.7%	44.9%
Forward Methods	0.6	55.5%	28.9%	39.1%	10.6%	11.5%	18.3%
Backward Methods	0.6	57.4%	32.6%	35.8%	14.6%	15.2%	18.0%
Forward Methods	0.7	51.5%	4.8%	33.3%	2.4%	-2.4%	2.4%
Backward Methods	0.7	52.6%	7.6%	28.2%	4.7%	0.7%	3.0%
Forward Methods	0.8	50.4%	0.4%	25.0%	0.3%	-3.9%	0.1%
Backward Methods	0.8	50.5%	0.6%	28.6%	0.4%	-4.3%	0.2%
Forward Methods	0.9	50.3%	0.0%	N/A	0.0%	N/A	0.0%
Backward Methods	0.9	50.3%	0.0%	N/A	0.0%	N/A	0.0%
Forward Methods	1.0	50.3%	0.0%	N/A	0.0%	N/A	0.0%
Backward Methods	1.0	50.3%	0.0%	N/A	0.0%	N/A	0.0%

Table 6: Comparison of Accuracy Measurements and Skill Scores for the 2-Hr Forecasting Tool

The table shows how the various accuracy measurements change as the threshold changes. Setting the threshold at zero means that when the model outputs the probability of lightning at greater than zero, the model predicts that lightning will occur. Because the output of a logistic regression equation is between zero and one, lightning will always be predicted at a threshold of zero. Conversely, setting the threshold at one means that when the model outputs the probability of lightning at greater than one, the model predicts that lightning will occur. The latter case is impossible, because the output of a logistic regression is always between zero and one. Therefore, lightning will never be predicted when the threshold equals one. For example, POD is at its highest when the threshold equals zero because lightning is predicted every time. This results in a very high Operational Utility Index, because the Operational Utility Index places the most weight on POD. A lower threshold drives a higher FAR and POD_f because lightning is falsely predicted more often. At a threshold of 1.0, the POD and POD_f become zero, because lightning is never detected. This also produces a 0% FAR, because lightning is never falsely predicted.

Both the forward and backward model selection methods performed similarly at all levels of the threshold, however, the models developed using the forward selection processes produced the highest Operational Utility Index. The Operational Utility Index is maximized at 45.8% at a lightning threshold of 0.3. However, indexes of 0.2 or 0.4 both provide a good Operational Utility Index ranging from 43.2% - 43.9%.

The Operational Utility Index is at its highest at thresholds of 0.2, 0.3, and 0.4, therefore, the threshold was further refined around these three points by adding an additional decimal place. Further refinement of the threshold will provide a more specific threshold. The graph below shows how the Operational Utility Index varies with the Lightning Index for both forward and backward models.

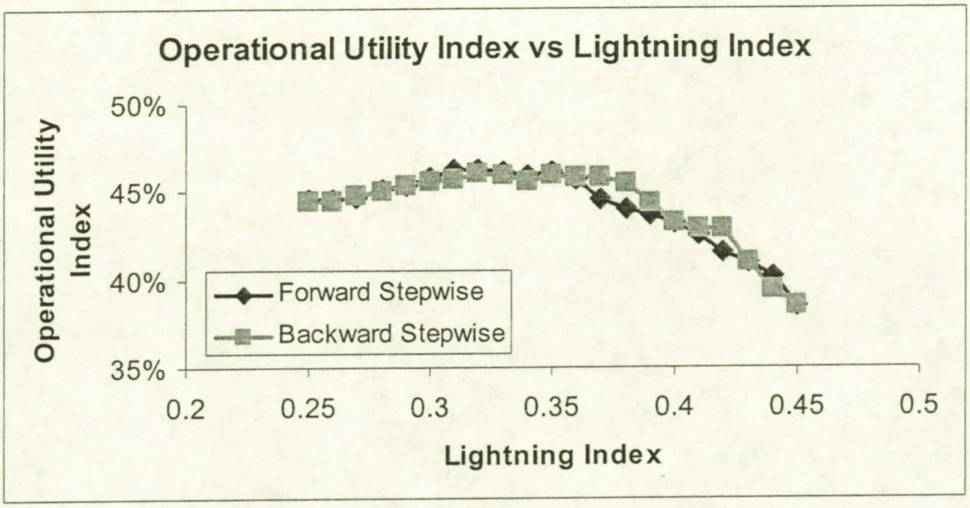

Figure 11: Operational Utility Index vs. Lightning Index for the 2-Hr Forecasting Tool

The graph shows that the Operational Utility Index increases slightly up to a peak of 0.32 and then begins to fall steadily at a threshold of 0.38. The highest Operational Utility Index results from the forward selection model process at a Lightning Threshold of 0.32. This Operational Utility Index is 46.3%. The graph also shows that there is not much sensitivity in the Operational Utility Index in this range of thresholds. The OUI ranges between 38.5% and 46.3% when the threshold is varied between 0.25 and 0.45

Lightning Index

Lowering the lightning threshold from 0.5 to 0.32 will adjust the mix of lightning forecast/not forecast and lightning observed/not observed.

Threshold = 0.5

		Observed	
		Yes	No
Forecast	Yes	512	384
	No	286	423

Threshold = 0.32

		Observed	
		Yes	No
Forecast	Yes	768	637
	No	30	170

Figure 12: Contingency Tables for the 2-Hr Forecasting Tool

Decreasing the threshold increases the amount of forecasted lightning from 896 to 1405, and decreases the amount of not-forecasted lightning from 709 to 200. The number of missed lightning events decreases from 286 to 30, yet the amount of falsely predicted lightning events increases from 384 to 637. Changing from a threshold of 0.5 to 0.32 changes accuracy measurements as shown in the table below:

Threshold	**Hit**	**POD**	**FAR**	**KSS**	**Operational Utility Index**	**POD_f**
0.5	58.3%	64.2%	42.9%	16.6%	30.5%	47.6%
0.32	58.4%	96.2%	45.3%	17.3%	46.3%	78.9%
% Increase	0.3%	50.0%	5.8%	4.4%	52.1%	65.9%

Table 7: Comparison of Accuracy Measurements and Skill Scores at Thresholds 0.5 and 0.32 for the 2-Hr Forecasting Tool

Lowering the threshold means that more lightning is detected. This results in a higher Hit Rate, POD, KSS, and Operational Utility Index. However, this also increases the FAR and POD_f. This result is acceptable, because detecting lightning is much more important that falsely warning of a lightning strike when there are a large number of lives at stake. While a threshold of 0.32 maximizes the Operational Utility Index, other thresholds provide a higher HR and KSS and a lower FAR and POD_f while still maintaining a relatively good Operational Utility Index.

New Logistic Regression Equation

The models perform differently at different threshold levels; therefore the model selected will vary based on the threshold. At a threshold of 0.32, the model generated using forward model selection processes was selected as the new logistic regression equation for several reasons. First, this model maximized the Operational Utility Index. Second, this model has only four independent variables while the backward regression model has six. The logistic regression takes the form of

$$f(z) = 1 / (1+e^{-z})$$

where $z = \alpha + \Sigma\beta_i x_i$

and
$\alpha = -2.366$
$\beta_1 = 2.053$ $x_1 = \Delta$ 0.5 Hr PW
$\beta_2 = -0.538$ $x_2 = \Delta$ 7.5 Hr PW
$\beta_3 = 0.031$ x_3 = K-Index
$\beta_4 = 0.322$ x_4 = Precipitable Water (PW) (cm)

This translates to:

$$1/\left(1+e^{-(-2.366 + 2.053x_1 + -0.538x_2 + 0.031\,x_3 + 0.322x_4)}\right)$$

The most significant independent variable in the model is the 0.5 hr change in PW. A change in the 0.5 hr PW will have the most impact the outcome of the model.

Independent Test

A random sample of 10% of the data set was selected to validate the 2-Hr Forecasting Tool. The sample was selected based on a random number generator that assigned a value between 0 and 1 to each data point. After the random number was assigned, all values that fell below .10 were selected and removed from the model data set. The test data set consisting of 197 data points was used to validate the 2- Hr Forecasting Tool. The model data set consisted of 1,605 data points.

The values for the independent variables were plugged into the new regression equation, and new predictions were made. At all levels of the threshold, the independent data closely matched the model performance. This independent verification of the model indicates that the model generated is valid. The table below provides a comparison of the model and test data sets at a threshold of 0.32 that maximizes the Operational Utility Index, and at the original threshold of

0.5. The test data set performed very closely to the model data set at both thresholds, indicating that the logistic regression equation is valid.

	Threshold	**Hit**	**POD**	**FAR**	**KSS**	**Operational Utility Index**	**POD_f**
Model	0.32	58.4%	96.2%	45.3%	17.3%	46.3%	78.9%
Test	0.32	56.9%	94.6%	47.9%	18.4%	45.4%	76.2%
Delta		.016	.017	.026	.011	.009	.027
Model	0.5	58.3%	64.2%	42.9%	16.6%	30.5%	47.6%
Test	0.5	62.9%	64.1%	40.4%	26.0%	34.0%	38.1%
Delta		.047	.001	.025	.095	.035	.095

Table 8: Comparison of Model Data and Test Data for the 2-Hr Forecasting Tool

9-Hr Forecasting Tool

The 9-Hr Forecasting Tool seeks to predict the probability of lightning in the next nine hours to support major outdoor operations, such as shuttle roll-out from the VAB to the launch pad. Prior to beginning an extended outdoor activity, it is essential to know the probability of lightning. If lightning has a high probability of occurrence, the outdoor operations will be postponed until weather conditions are more favorable.

Data Structure

The same candidate independent variables were to develop in the 9-Hr Forecasting Tool as were used in the 2-Hr Forecasting Tool as shown in Table 4. To assess the probability of lightning in the next nine hours, the data set had to be modified. If lightning occurred at any time within nine hours of a precipitable water data point, a value of one was assigned to the dependent variable. Likewise, if lightning did not occur within nine hours of a precipitable water data point, a value of zero was assigned to the dependent variable. After the data was structured to properly reflect the occurrence of lightning, the data sets that did not have a value for each variable were removed. This meant a reduction in data points from 29,309 to 13,426.

Model Selection Process

Six methods were used for model selection: Forward Stepwise Conditional, Forward Stepwise Likelihood Ratio, Forward Stepwise (Wald), Backward Stepwise Conditional, Backward Stepwise Likelihood Ratio, and Backward Stepwise (Wald).

Goodness-of-Fit statistics were initially calculated to evaluate the model fit. Specifically, Cox & Snell R-squared, Nagelkerke R-Squared, and the Hosmer-Lemeshow goodness-of-fit $\chi 2$ were used as an assessment of model fit. A comparison of model outputs is shown below:

Model Selection Method	Cox & Snell R Square	Nagelkerke R Square	Hosmer –Lemeshow Chi-square
Forward Conditional	0.181	0.246	99.416
Forward LR	0.181	0.246	99.416
Forward Wald	0.181	0.246	99.416
Backward Conditional	0.181	0.246	98.296
Backward LR	0.181	0.246	90.723
Backward Wald	0.181	0.246	98.296

Table 9: Comparison of Goodness-Of-Fit Statistics for all models for the 9-Hr Forecasting Tool

While these values show an improvement over the 2-Hr Forecasting Tool, they still do not represent a good model fit. However, for the same reasons as described earlier, these values were determined not to be good indicators of model fit.

In selecting the model, the most weight was given to the value of several accuracy measurements and skill scores. These ratios include the Hit Rate (H), False Alarm Rate(FAR), Probability of Detection (POD), Probability of False Detection (POD_f), Hansen Kuipers or True Skill Score (KSS), and an Operational Utility Index. The KSS is considered to be the most critical factor in the 9-Hr Forecasting Tool. In the case of the 9-Hr Forecasting Tool, fewer personnel will be impacted because there are fewer employees outside for the space shuttle roll-out. Therefore, while POD is still critical, it is also important not to falsely detect lightning and cause a launch delay costing millions of dollars.

The accuracy measurements and skill scores were calculated based on varying the lightning index from 0.0 to 1.0. Note that all forward methods of model development process yielded the same model, and all backward methods of model selection yielded the same model. They are grouped into two categories in the table below: forward methods and backward methods. Again, the objective is to maximize the KSS.

Model Selection Method	Threshold	Hit	POD	FAR	KSS	Operational Utility Index	POD_f
Forward Methods	0	38.4%	100.0%	61.6%	0.0%	39.7%	100.0%
Backward Methods	0	38.4%	100.0%	61.6%	0.0%	39.7%	100.0%
Forward Methods	0.1	51.8%	98.8%	55.7%	21.3%	36.6%	77.5%
Backward Methods	0.1	51.9%	98.8%	55.7%	21.4%	36.5%	77.4%
Forward Methods	0.2	59.2%	95.8%	51.5%	32.2%	34.0%	63.6%
Backward Methods	0.2	59.3%	95.7%	51.5%	32.4%	33.9%	63.3%
Forward Methods	0.3	63.1%	89.1%	48.9%	35.9%	30.4%	53.2%
Backward Methods	0.3	63.1%	88.9%	48.9%	35.8%	30.3%	53.1%
Forward Methods	0.4	66.2%	75.6%	45.7%	35.9%	24.2%	39.7%
Backward Methods	0.4	66.2%	75.4%	45.7%	35.9%	24.1%	39.5%
Forward Methods	0.5	67.1%	51.8%	41.9%	28.5%	14.2%	23.3%
Backward Methods	0.5	67.2%	51.5%	41.8%	28.5%	14.1%	23.1%
Forward Methods	0.6	64.9%	24.8%	39.5%	14.7%	3.4%	10.1%
Backward Methods	0.6	64.8%	24.5%	39.5%	14.5%	3.3%	10.0%
Forward Methods	0.7	62.7%	5.9%	33.7%	4.0%	-3.3%	1.9%
Backward Methods	0.7	62.7%	6.0%	33.7%	4.1%	-3.3%	1.9%
Forward Methods	0.8	61.6%	0.1%	44.4%	0.1%	-7.4%	0.1%
Backward Methods	0.8	61.6%	0.1%	57.1%	0.0%	-9.5%	0.1%
Forward Methods	0.9	61.6%	0.0%	N/A	0.0%	N/A	0.0%
Backward Methods	0.9	61.6%	0.0%	N/A	0.0%	N/A	0.0%
Forward Methods	1	61.6%	0.0%	N/A	0.0%	N/A	0.0%
Backward Methods	1	61.6%	0.0%	N/A	0.0%	N/A	0.0%

Table 10: Comparison of Accuracy Measurements and Skill Scores for the 9-Hr Forecasting Tool

The KSS varies significantly with changes in the threshold, although the values at thresholds of 0.3 and 0.4 are similar. As you can see, the KSS is maximized at 35.9% at a lightning threshold of 0.3 for the forward methods and 0.4 for both the forward and backward methods. Both forward and backward selection models performed similarly at all levels of the threshold.

The KSS is at its highest at thresholds of 0.3 and 0.4, therefore the threshold was further refined around these two points by adding an additional decimal place. Further refinement of the threshold will provide a more specific threshold. The graph below shows how the KSS varies with the Lightning Index for both forward and backward models between 0.25 – 0.45.

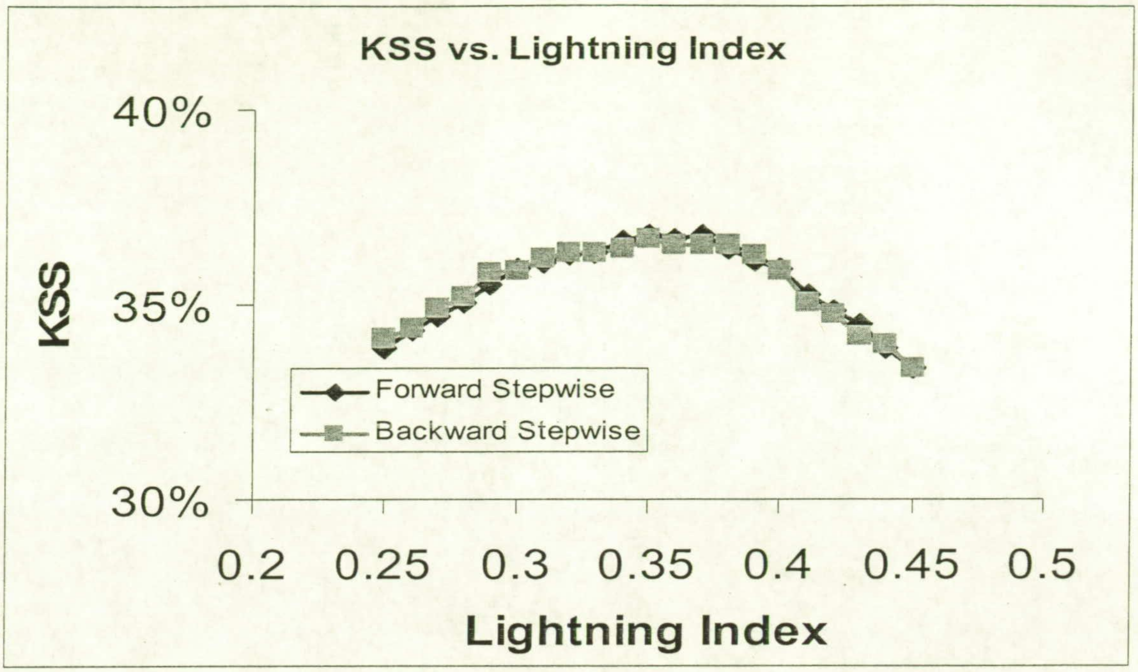

Figure 13: KSS vs. Lightning Index for the 9-Hr Forecasting Tool

The graph shows that KSS increases slightly up to a maximum of 0.35 and begins to fall steadily at a threshold of 0.38. The maximum KSS of 36.8% results from the forward selection model processes at lightning thresholds of 0.35 and 0.37. In this range of the Lightning Index, the KSS changes minimally, with the largest value being 36.8% and the smallest being 33.4%.

Lightning index

The KSS was maximized at thresholds of both 0.35 and 0.37. Further comparison of accuracy measurements and skill scores is required to select the optimal threshold. The table below compares all accuracy measurements and skill scores of the forward selection models at thresholds of 0.35 and 0.37.

Threshold	Hit	POD	FAR	KSS	Operational Utility Index	POD_f
0.35	64.9%	83.5%	47.3%	36.8%	27.7%	46.7%
0.37	65.5%	80.9%	46.6%	36.8%	26.6%	44.1%

Table 11: Comparison of Accuracy Measurements and Skill Scores at Thresholds of 0.35 and 0.37 for the 9-Hr Forecasting Tool

Increasing the threshold from 0.35 to 0.37 results a lower POD and Operational Utility Index. However, a higher HR, lower FAR and lower POD_f result from increasing the threshold. Over long periods of time, such as the 9-Hr Forecasting Tool, more weight is given to FAR and POD_f. Therefore, a threshold of 0.37 is superior to a threshold of 0.35.

The output of the logistic regression model, or the lightning index, is the probability that the outcome is equal to one. In this case, that is interpreted as the probability of lightning. The model is set up to predict lightning when the probability of lightning is greater than 50%. Lowering the lightning index from 0.5 to 0.37 will increase the amount of lightning that is detected, which will adjust the mix of lightning forecast/not forecast and lightning observed/not observed.

Threshold = 0.5

		Observed	
		Yes	No
Forecast	Yes	2393	1727
	No	2227	5677

Threshold = 0.37

		Observed	
		Yes	No
Forecast	Yes	3738	3268
	No	882	4136

Figure 13: Contingency Tables for the 9-Hr Forecasting Tool

Decreasing the threshold increases the amount of forecasted lightning from 4,120 to 7,006, and decreases the amount of not-forecasted lightning from 7,904 to 5,018. The number of missed lightning events decreases from 2,227 to 882, yet the amount of falsely predicted lightning events increases from 1,727 to 3,268. Changing from a threshold of 0.5 to 0.37 changes the skill and accuracy measurements of the model as shown in the table below:

Threshold	Hit	POD	FAR	KSS	Operational Utility Index	POD_f
0.5	67.1%	51.8%	41.9%	28.5%	14.2%	23.3%
0.37	65.5%	80.9%	46.6%	36.8%	26.6%	44.1%
% Increase/ Decrease	-2.4%	56.2%	11.3%	29.2%	87.4%	89.2%

Table 12: Comparison of Accuracy Measurements and Skill Scores at Thresholds of 0.5 and 0.37 for the 9-Hr Forecasting Tool

Lowering the threshold means that more lightning is detected. This results in a higher POD, KSS, and Operational Utility Index. However, this also increases the FAR and POD_f. While a threshold of 0.37 maximizes the KSS, other thresholds provide improvements to FAR and POD_f while still providing good KSS.

New Logistic Regression Equation

The model generated using forward model selection processes was selected as the new logistic regression equation for several reasons. First, this model produced the highest KSS. Second, this model has only five variables while the backward regression model has six. The logistic regression takes the form of

$$f(z) = 1 / (1+e^{-z})$$

where $z = \alpha + \Sigma\beta_i x_i$

and

$\alpha = -4.885$
$\beta_1 = 0.541$ x_1 = Precipitable Water (PW) (cm)
$\beta_2 = 0.346$ $x_2 = \Delta$ 3.5 Hr PW
$\beta_3 = -0.446$ $x_3 = \Delta$ 8.5 Hr PW
$\beta_4 = 0.235$ $x_4 = \Delta$ 12 Hr PW
$\beta_5 = 0.071$ x_5 = K-Index

This translates to:

$$1/(1+e^{-(-4.885 + 0.541x_1 + 0.346x_2 + -0.446x_3 + 0.235x_4 + 0.071x_5)})$$

The most significant independent variable in predicting lightning is the current PW level. The second most significant variable is the 8.5-hr change in the PW level in the atmosphere. This is similar to the Mazany model that determined that the 9-hr change in PW was most significant in predicting lightning.

Independent Test

A random sample of 10% of the data set was selected to validate the 9-Hr Forecasting Tool. The sample was selected based on a random number generator that assigned a value between 0 and 1 to each data point. After the random number was assigned, all values that fell below .10 were selected and removed from the model data set. The test data set consisting of 1,402 data points was used to validate the 9- Hr Model. The model data set consisted of 12,024 data points.

The values for the independent variables were plugged into the new regression equation, and new predictions were made. At all levels of the threshold, the independent data closely matched the model performance. This independent verification of the model indicates that the model generated is valid. The table below provides a comparison of the model and test data sets at the threshold of 0.37 that maximizes KSS, and at a threshold of 0.5 which was the default threshold. The test data set performed very closely to the model data set, indicating that the logistic regression equation is valid.

Model vs. Test Data	Threshold	Hit	POD	FAR	KSS	Operational Utility Index	POD_f
Model	**0.37**	65.5%	80.9%	46.6%	36.8%	26.6%	44.1%
Test	**0.37**	64.8%	80.2%	47.7%	35.6%	29.3%	44.7%
Delta		.007	.007	.011	.012	.027	.005
Model	**0.5**	67.1%	51.8%	41.9%	28.5%	14.2%	23.3%
Test	**0.5**	65.8%	49.9%	44.6%	25.5%	13.0%	24.5%
Delta		.013	.019	.026	.030	.012	.011

Table 13: Comparison of Model and Test Data for the 9-Hr Forecasting Tool

Recommendations

The forecasting tools developed show promise in improving on the daunting task of trying to predict lightning. As the validation of the Mazany model shows, these tools sometimes do not prove to be as useful when put into practice. Therefore, these two new forecasting tools should be used; however validation of these models should be done on a regular basis to ensure that they are effective. As more data becomes available, the models should be consistently tested and updated when necessary. This research focused on PW, the change in PW over a 12 hour period, and the K-Index. Other factors should be tested for significance in forecasting lightning that could improve the effectiveness of the models. These models could be improved by increasing the frequency at which data is collected. Several assumptions had to be made regarding the K-Index and PW because of the data collection intervals. For example, the average PW over a 30 minute timeframe was used to represent the latest PW observation. Also, the latest K-Index value was used, which could sometimes be hours from the current time. Increasing the frequency of PW and K-Index data collection will improve the model's ability to accurately forecast lightning.

The lightning index for both models was established based on maximizing the OUI for the 2-Hr Forecasting Tool and the KSS for the 9-Hr Forecasting Tool. These lightning indices can be established at any level, and the various accuracy measures and skill scores will change. Establishing the threshold at 0.32 for the 2-Hr Forecasting Tool causes the POD_f to be almost 80%. This level may be unacceptable and therefore the threshold would need to be increased. Therefore, the lightning indices should be established such that all accuracy measures and skill scores are at an acceptable level.

References

(1) Mazany, R. A., Businger, S., Gutman, S. I., and W. Roeder, 2002: A Lightning and Prediction Index that Utilizes GPS Integrated Precipitable Water Vapor. *Weather and Forecasting,* **17**, 1034-1047
(2) Energy: The Ultimate Resource?; "Lightning," http://encarta.msn.com, 1997-2004, accessed December 19, 2004.
(3) Lightning and Launches. NASA, Educator Features, April 22, 2004
(4) Lightning and the Space Program. NASA Facts, KSC Release No. 72-90, August 1998.
(5) Erck, A. Weather balloons take a close look at the sky. *USA TODAY,* March 24, 2001
(6) IR Group, "Precipitable Water", http://www.ghcc.msfc.nasa.gov/irgrp/pw_studies.html, November 2, 1999, accessed December 21, 2004.
(7) "Accuracy of absolute precipitable water vapor estimates from GPS observations" http://wwwrses.anu.edu.au/geodynamics/gps/papers/pw_jgr.html, 1998, accessed December 19, 2004.
(8) Wilks, D., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 238-243, 248–249 pp.

Appendix

Applied Statistics and Probability for Engineers contributed most in the research for this project. Specifically, scatter plots, Statistical Process Control (SPC) charts, regression methods, and model selection processes were applied throughout the project. Deterministic Method in Operations Research contributed to the project in understanding maximization of the lightning indices. Quality Management and Engineering also provided the background for supporting the process that would be used to develop both the 2-Hr and 9-Hr Forecasting Tool.