

THESIS

AUTOMATED TROPICAL CYCLONE EYE DETECTION USING
DISCRIMINANT ANALYSIS

Submitted by

Robert DeMaria

Department of Computer Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2015

Master's Committee:

Advisor: Charles Anderson

Bruce Draper
Wayne Schubert

Copyright by Robert DeMaria 2015

All Rights Reserved

ABSTRACT

AUTOMATED TROPICAL CYCLONE EYE DETECTION USING DISCRIMINANT ANALYSIS

Eye formation is often associated with rapid intensification of tropical cyclones, so this information is very valuable to hurricane forecasters. Linear and Quadratic Discriminant Analysis (LDA and QDA) were utilized to develop a method for objectively determining whether or not a tropical cyclone has an eye. The input to the algorithms included basic storm information that is routinely available to forecasters, including the maximum wind, latitude and longitude of the storm center, and the storm motion vector. Infrared imagery from geostationary satellites in a 320 km by 320 km region around each storm was also used as input. Principal Component Analysis was used to reduce the dimension of the IR dataset. The ground truth for the algorithm development was the subjective determination of whether or not a tropical cyclone had an eye made by hurricane forecasters. The input sample included 4109 cases at 6 hr intervals for Atlantic tropical cyclones from 1995 to 2013.

Results showed that the LDA and QDA algorithms successfully classified about 90% of the test cases. The best algorithm used a combination of basic storm information and principal components from the IR imagery. These included the maximum winds, the storm latitude and components of the storm motion vector, and 10 PCs from eigenvectors that primarily represented the symmetric structures in the IR imagery. The QDA version performed a little better using a Peirce Skill Score, which measures the ability to correctly classify cases. The LDA and QDA algorithms also provide the probability that each case contains an eye. The LDA version performed a little better using the Brier Skill Score, which measures the utility of the class probabilities.

The high success rate indicates that the algorithm can reliably reproduce what forecasters are currently doing subjectively. This algorithm would have a number of applications, including providing forecasters with an objective way to determine if a tropical cyclone has or is becoming more likely to form an eye. The probability information and its time trends could be used as input to other algorithms, such as existing operational forecast methods for estimating tropical cyclone intensity changes.

ACKNOWLEDGEMENTS

Thanks to Greg DeMaria for helping to make a number of the figures as well as Julia Pillard for editing and moral support.

This thesis is typeset in L^AT_EX using a document class designed by Leif Anderson.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iv
List of Tables	vii
List of Figures	viii
Chapter 1. Introduction	1
1.1. Current Approach to TC Eye Detection	1
1.2. Objective	4
1.3. Proposed Approach	4
1.4. Overview	6
Chapter 2. Data	8
2.1. CIRA/RAMMB Tropical Cyclone Satellite Image Archive	8
2.2. TAFB Classifications	8
2.3. Best Track	9
2.4. Algorithm Input	10
Chapter 3. Mathematical Background	11
3.1. Quadratic and Linear Discriminant Analysis	11
3.2. Principal Component Analysis	14
3.3. Verification	15
Chapter 4. Algorithm Development	20
4.1. Baseline Algorithm	20
4.2. Satellite Data Processing	29

4.3. Satellite Data Based Algorithm	40
4.4. Combined Satellite Data and Ancillary Data Algorithm	44
Chapter 5. Discussion	50
5.1. Case Studies	50
5.2. Independent Testing Data	54
5.3. Alternate Sensitivity Vector	55
Chapter 6. Conclusions	58
Bibliography	61

LIST OF TABLES

3.1	Table of confusion defining how estimates are considered correct or errors. Estimates that are considered true positives and true negatives are considered correct, while false positives and false negatives are considered errors.	16
4.1	Steps used for the baseline algorithm.....	25
4.2	Steps used for the satellite data based algorithm.....	41

LIST OF FIGURES

1.1	High resolution visible VIIRS image of Hurricane Edouard.	2
1.2	Hurricane Gonzalo over its lifetime.	7
4.1	Probability distributions as functions of the seven baseline model inputs.	23
4.2	Average error statistics generated from 1000 runs of the Baseline algorithm.	25
4.3	Sensitivity vector for the baseline algorithm.	27
4.4	Boundary of the “eye-present” and “eye-absent” classes produced from all 4109 available samples. Only maximum winds and latitude were used so that a 2D plot could be examined. Probability of each class can be seen in the color intensity. Correct classifications appear as large filled dots, incorrect classifications appear as small dots. Large black dots indicate the location of the means of each class. ...	28
4.5	Example IR images from Hurricane Katrina. Boxes show the selection of pixels used with the algorithm. Image classified as eye absent (left). Image classified as eye present (right)	29
4.6	The mean image and standard deviation image produced from the entire set of IR data samples available for use with this project.	30
4.7	Eigenvectors 1-15.	32
4.8	Eigenvectors 16-25.	33
4.9	Normalized variance vs the principal component number. The first few principal components explain the vast majority of the variance of the data. The sum of the normalized variance with 25 PCs is roughly 95%.	34

4.10	Reconstruction of a Hurricane Katrina image at 08/28/2005 06:15 UTC using 24 eigenvectors	36
4.11	Reconstruction of a Hurricane Danielle image at 08/24/2010 23:45 UTC using 24 eigenvectors	37
4.12	Reconstruction of a Tropical Storm Arlene image at 06/10/2005 11:45 UTC using 24 eigenvectors	38
4.13	Wide view of Hurricane Katrina, Hurricane Danielle, and Tropical Storm Arlene. Orange box indicates pixels selected for use with the eye detection algorithm.	39
4.14	PC value amplitudes for Hurricane Katrina, Hurricane Danielle, and Tropical Storm Arlene.	40
4.15	25 EOFs with the highest variance explained sorted from largest to smallest sensitivity.	42
4.16	Comparison of satellite data algorithm runs using 25 EOF's sorted by variance explained, 10 EOF's sorted by variance explained, and 10 EOF's sorted by significance. All metrics are averages from 1000 runs using different shufflings of training and testing data.	44
4.17	Performance of the satellite data based algorithm run with increasing number of EOF's where the order of included EOF's was determined by the magnitude of its significance.	45
4.18	Comparison of combined satellite data and ancillary data algorithm runs. LDA and QDA were run using a randomly selected shuffling of the training and testing set as well as using the entire dataset for training and testing. All metrics are averages from 1000 runs using different shufflings of training and testing data.	46

4.19	Sensitivity vector generated from the 14 predictor combined model.	47
4.20	LDA and QDA boundaries using EOF 9 and VMAX. Generated using entire dataset for both training and testing.	49
5.1	Algorithm probabilities/classifications vs truth for the lifetime of Hurricane Danielle.	51
5.2	Algorithm probabilities/classifications vs truth for the lifetime of Hurricane Katrina.	52
5.3	IR images from Hurricane Danielle used as input to algorithm. The algorithm incorrectly classified cases 18 and 26 as “eye-absent” cases while case 9 was mis-classified as “eye-present”	54
5.4	IR images from Hurricane Katrina used as input to algorithm. The algorithm incorrectly classified the images as “eye-absent” cases.	54
5.5	Comparison between average LDA/QDA performance among 1000 runs of the “combined” model using randomly shuffled training/testing sets and LDA/QDA trained using all available data and tested against 131 images from the 2014 hurricane season in the Atlantic basin.	56
5.6	Comparison between the 25 element sensitivity vector generated using the standard deviation term and the vector generated without the standard deviation term. Elements are sorted by their magnitude within the sensitivity vector.	57

CHAPTER 1

INTRODUCTION

1.1. CURRENT APPROACH TO TC EYE DETECTION

Tropical cyclones (TC) are rotating weather disturbances that form in the tropics and subtropics [1]. TC circulation extends over hundreds of miles, with surface winds that can exceed 80 m/s near the center. Because of the Coriolis effect, the TC circulation rotates counter-clockwise in the northern hemisphere and clockwise in the southern hemisphere. TCs derive their energy from the vast amounts of heat stored in the ocean, and have a different structure than the cold season storms associated with fronts in higher latitudes. TCs typically form near, but not on the equator, move westward in the trade winds, and then gradually move north around the large high pressure systems in the subtropical regions. If they get far enough north, they often move eastward under the influence of the mid-latitude westerly jet stream. TCs routinely occur in all the major ocean basins except the Arctic Ocean, and are responsible for loss of life and property through high winds, storm surge, and flooding from heavy rain.

Because of the extraction of heat from the ocean in the high wind regions of TCs, strong thunderstorms form a ring-like structure near the center, which is referred to as the eye wall. Inside the eye wall is the eye, which is a region of fairly light winds and few thunderstorms. The eye of strong TCs is very striking in satellite imagery. This can be seen in Figure 1.1 where the tropical cyclone eye clearly stands out from the surrounding cloud-tops of hurricane Edouard. As air rises within the eye wall of a tropical cyclone it cools due to the decreasing pressure with height. This causes the water vapor in the air to condense into liquid and eventually freeze. The phase change from vapor to liquid or ice adds heat to the

rising air, causing it to cool more slowly and continue to accelerate upwards. The rising air eventually reaches a level of the atmosphere known as the tropopause. Within this layer, temperature outside the eye wall either remains constant with height or increases. This acts as a ceiling for rising warm air in the eye wall, which then moves to the side (outward and inward from the eye wall) and downward into the low pressure center of the tropical cyclone. As the air descends in the center, its temperature increases due to compression, preventing condensation and precipitation from forming, and creating the circular, fairly cloud-free region which distinguishes the feature as an eye[2].

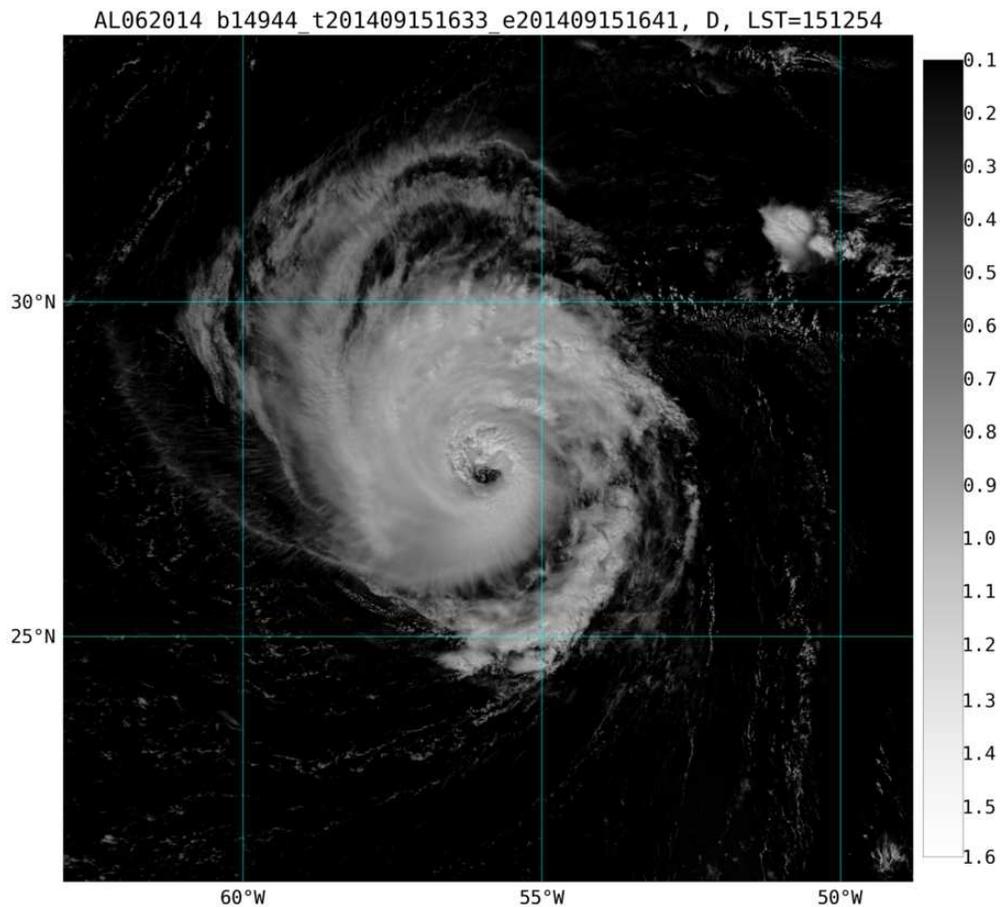


FIGURE 1.1. High resolution visible VIIRS image of Hurricane Edouard.

Formation of a TC eye is often associated with the beginning of rapid intensification, as long as the atmospheric environment around the storm remains favorable [2]. Thus,

determining the onset of eye formation is of paramount importance for intensity forecasts. By the same token, eye formation is also important for tropical cyclone location estimation, as the cyclones center becomes much more obvious when an eye is present.

There are several ways that tropical cyclone eyes can be detected [2]. In the western part of the North Atlantic, Hurricane Hunter aircraft routinely monitor these storms. These aircraft are equipped with radar, which the flight scientists use to determine if an eye is present if a ring-like structure appears when the aircraft is near the storm center. Visual inspection of the cloud patterns is also sometimes used. However, one limitation of this approach is that aircraft data are not routinely available for the majority of TCs around the world. Even in the Atlantic region, the aircraft missions can be 12 or more hours apart. For this reason, other TC detection methods use satellite data, which are much more commonly available.

There are two operational forecast agencies that routinely perform TC analysis using satellite data. These include the National Oceanic and Atmospheric Administration (NOAA), National Environmental Satellite Data and Information Service (NESDIS), Satellite Analysis Branch (SAB), and the NOAA National Hurricane Center (NHC) Tropical Analysis and Forecast Branch (TAFB). SAB provides satellite analysis for all TCs around the globe, while TAFB performs analysis for storms in the Atlantic and Northeast Pacific region. Both agencies determine the center location of the TCs and make an estimate of the intensity of the storms using the Dvorak method [3]. This method uses visible and infrared (IR) data as input to a pattern matching technique that estimates the maximum surface winds. Although the Dvorak method is fairly structured, it is still partially subjective. Center locations and intensities are estimated four times per day, though this number can increase if a TC changes quickly between 6 h periods.

1.2. OBJECTIVE

As stated above, the detection of the presence of a TC eye is important for intensity forecasting as well as TC location estimation. Currently, the determination of eye formation from satellite imagery is performed subjectively as part of the Dvorak intensity estimate and/or official warning/advisory/discussion processes. At present, little investigation has been made into the use of objective techniques, and, as a consequence, much of the satellite imagery available to depict eye-formation is not used. The goal of this study is to do objectively what the forecasters now do subjectively. Once a method for objectively determining eye presence is developed, it can easily be run on all available imagery rather than just selected images, and could be used as input to other statistical methods such as the Rapid Intensification Index (RII) used by NHC to forecast hurricane intensity changes [4]. Additionally, the use of image processing techniques to perform eye detection represents novel work in the field of tropical meteorology. For these reasons, an algorithm to perform automated eye detection using IR satellite data was developed.

1.3. PROPOSED APPROACH

As described in Kidder and Vonder Harr[5], weather satellites are in two basic kinds of orbits. Geostationary (Geo) satellites are in an orbit about 36,000 km above the equator, and remain at the same location relative to a point on the earth. They detect radiation in the visible and IR portions of the electromagnetic spectrum. Low earth orbiting (LEO) satellites are in much lower orbits (typically about 800 km) and pass over the same point on the earth roughly twice per day. LEO satellites tend to have higher spatial resolution compared with Geo, and include measurements in the visible, IR, and microwave portions of the electromagnetic spectrum. Due to the size of the atmospheric particles in clouds

relative to the wavelengths of the radiation, clouds are nearly opaque to visible and IR radiation. Thus, when looking from space, only the tops of clouds can be viewed with IR and visible data. In contrast, microwave radiation can pass through clouds, and so information below cloud top can be obtained from measurements in that portion of the spectrum. As a compromise between spatial, temporal, and spectral resolution, only measurements from Geo satellites will be used in this study.

There is a constellation of Geo satellite around the globe, with the U.S. Geostationary Observational Environmental Satellites (GOES) available over the Atlantic (GOES-east) and eastern Pacific (GOES-west). Data is available at half hour intervals over the Atlantic and Pacific areas. The current generation of GOES satellites include 5 imager channels, four of which are in the IR region and one in the visible region. Since nearly all of the visible radiation that the satellite detects comes from sun, this data is only available during the day. Therefore, only IR data will be used. Atmospheric molecules such as carbon dioxide and water vapor absorb IR radiation in certain parts of the spectrum, and the channels on GOES were selected to exploit those regions for various purposes. A region of the IR spectrum with wavelengths near 10.7 microns is fairly transparent to IR radiation, and one of the GOES channels measures radiation in that region. This is channel 4 on GOES, and is sometimes called the “window” channel. Since the window channel radiation is primarily emitted from the atmosphere, it is available day and night, and is one of the most common satellite data sources used for TC analysis. For this reason, the IR window channel data from GOES will be utilized in this study.

Figure 1.2 shows an example of the IR window channel data during the formation of Hurricane Gonzalo from the 2014 Atlantic hurricane season. The images are color-enhanced so that the highest clouds with the coldest temperatures are highlighted. This figure shows

that the eye of Gonzalo formed between about 12 UTC and 18 UTC on October 14, 2014 and lasted through 18 UTC on October 17.

While there are a plethora of statistical algorithms available that may be used for pattern recognition given a set of images, the task of performing eye-detection was framed as a classification problem, given the IR imagery as well as commonly available TC information. That is, given a satellite image, the algorithm would classify the image as “eye-present” or “eye-absent”. The ground truth of whether or not an eye was present came from the TAFB subjective estimates, which are produced as part of the Dvorak technique. Fairly standard statistical methods such as Principal Component Analysis (PCA) as well as Linear and Quadratic Discrimination Analysis (LDA and QDA) will be utilized. These methods were chosen due to their relative simplicity as well as their use in other TC forecast algorithms such as the RII, and the NESDIS tropical cyclone formation probability product [6].

1.4. OVERVIEW

The data for this study is described in Chapter 2. The mathematical background behind the techniques used in this study is presented in Chapter 3. The development of the algorithm is described in Chapter 4. Discussion of the algorithm is presented in Chapter 5. Conclusions and future work are discussed in Chapter 6.

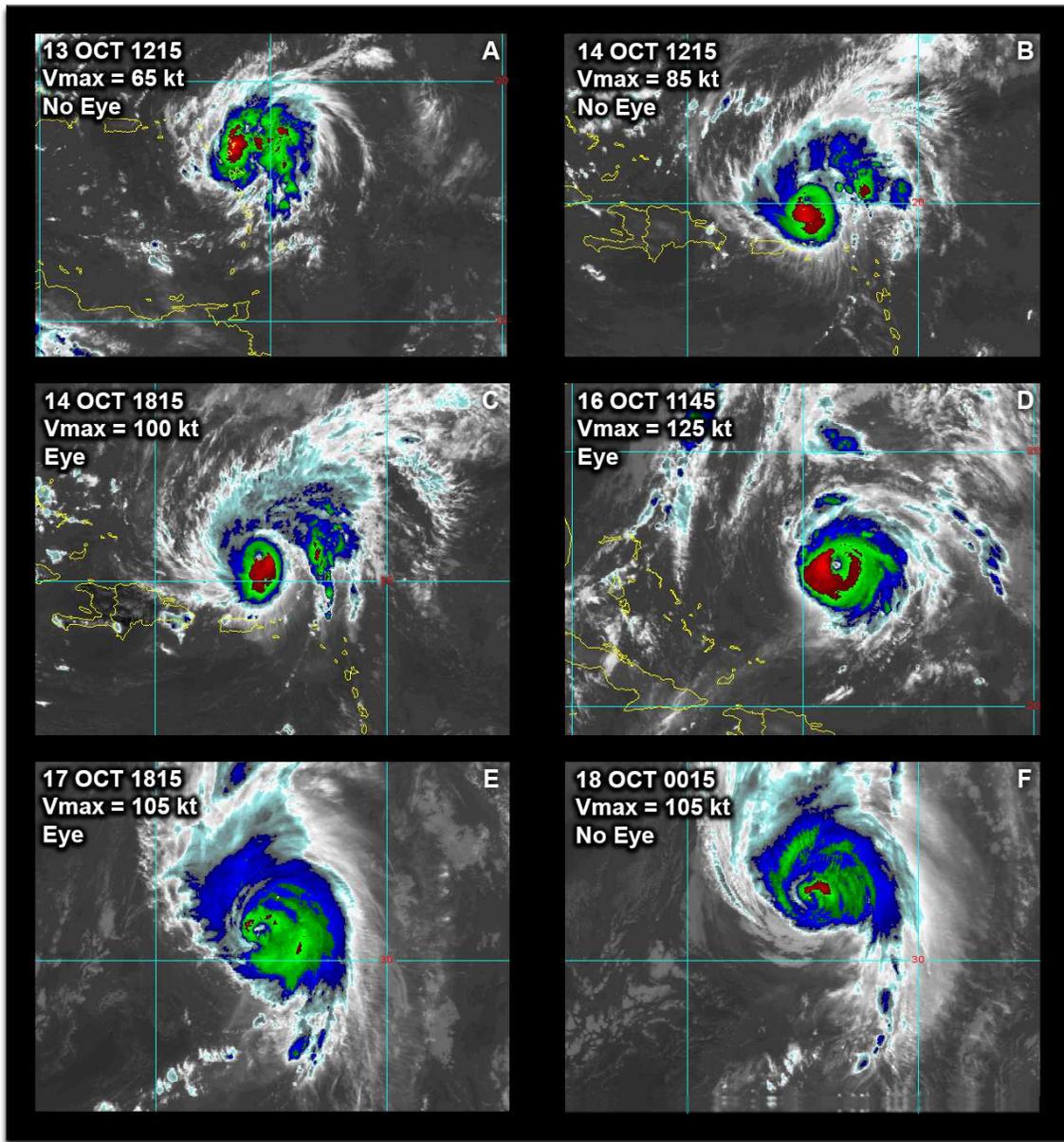


FIGURE 1.2. Hurricane Gonzalo over its lifetime.

CHAPTER 2

DATA

The databases used in this study are described in this section. These include the satellite imagery, the ground truth for algorithm development, and basic tropical cyclone information, such as center position and maximum wind.

2.1. CIRA/RAMMB TROPICAL CYCLONE SATELLITE IMAGE ARCHIVE

Available for use with this project is an archive of geostationary satellite data from the CSU Cooperative Institute for Research in the Atmosphere (CIRA) [7]. The archive contains $10.7\ \mu\text{m}$ infra-red (IR) data from 1995-2013 for tropical cyclones around the globe, and contains more than 100,000 images. Each image is associated with a particular tropical cyclone at roughly 30 minute intervals. While the archive also contains visible imagery, only IR imagery was considered for this project since the visible channel is not available at night. Each pixel within an IR image consists of a temperature value as well as the latitude and longitude location of the pixel. The horizontal resolution of the IR data is 4 km.

2.2. TAFB CLASSIFICATIONS

Currently, when a tropical cyclone is present in the Atlantic basin, satellite analysts at the Tropical Analysis and Forecast Branch (TAFB) at the NHC generate a Dvorak intensity fix [3] which is used as input to the TC forecasts. In order to generate a Dvorak intensity fix, the analyst will determine if an eye is currently present, utilizing IR and other satellite data available at the time of the fix. These classifications are typically generated every six hours at 00, 06, 12 and 18 UTC, which are referred to as synoptic times. The Dvorak intensity fixes have been recorded and made available in digital form since 1996. The IR image closest to

each synoptic time was obtained from the CIRA satellite archive and matched with the TAFB records associated with its “eye-present” or “eye-absent” classification. These classifications are considered truth for the algorithm development described here. The images are within 30 minutes of each synoptic time. Only one image is used for each synoptic time.

The classifications made by the TAFB satellite analysts represent the best estimate of truth available for use with this algorithm. It should be noted that the classifications generated as part of the Dvorak method are subject to the interpretation of the analyst that performed the classification. To get an estimate of how often different forecasters will “agree” on the eye classification of an image, the TAFB classifications were compared to classifications performed by satellite analysts at the NESDIS Satellite Analysis Branch. The classifications between TAFB and SAB agreed approximately 95% of the time. By using these subjective classifications as truth, the algorithm will be replicating any biases present in the set, but the close agreement between SAB and TAFB analysts provides some confidence in the ground truth for the algorithm.

2.3. BEST TRACK

Whenever a tropical cyclone is present in the Atlantic or eastern North Pacific basins, the NHC generates estimated locations for the center of the tropical cyclone every 6 hours, using all available information, including satellite and aircraft reconnaissance data. After the season has ended, the NHC returns to these tracks and refines the estimate. These refined tracks are known as the “best track” [8]. As will be described in Chapter 4, the algorithm presented here uses a square subset of pixels available from each IR satellite image. This subset is centered on the best track location interpolated to the time of the image. Although an algorithm that used a refined center estimate might be more desirable

since some smoothing is introduced by the best track process, the best track was selected in order to reduce the complexity of the problem being solved. Future work may involve developing methods to reduce the impact of center position error on the algorithm. Included with the best track is an estimate of the maximum wind speed. This estimate represents the maximum sustained wind speed from a one-minute average at 10 meters above the surface [9]. The latitude and longitude positions in the best track are rounded to the nearest 1/10th deg, and the maximum wind values are rounded to the nearest 5 kt.

2.4. ALGORITHM INPUT

The final input for the algorithm development is the IR image, the date and time of the image, the latitude, longitude and maximum wind of the TC at the time of the image, and the determination by TAFB of whether or not an eye was present. TCs in the Atlantic and eastern North Pacific can have different characteristics. For example, [7] have shown that Atlantic TCs tend to be larger than east Pacific TCs. For the sake of simplification, only the Atlantic basin was considered in this study. Further work may involve adapting the algorithm for other basins.

The formation of a TC eye is related to the organization of the thunderstorm activity near the storm center. As described by [2], the formation of an eye does not usually occur until the maximum surface wind reaches about 50 kt. In the TAFB dataset, a few cases where the TC maximum wind was less than 50 kt were identified as having eyes. Therefore, all cases where the TCs were of at least tropical storm intensity (maximum winds of 34 kt or greater) were included. The final dataset includes 4109 cases from 265 different Atlantic tropical cyclones. All cases are 6 hr or greater apart in time, and 991 (24%) were considered to have an eye by TAFB.

CHAPTER 3

MATHEMATICAL BACKGROUND

As described in the Introduction, QDA and LDA will be used in the eye classification method using the storm information from the best track and IR satellite imagery. A basic overview of QDA and LDA are provided in this section.

The data sample includes 4109 cases. As can be seen in Fig. 1.2, the cloud structure associated with tropical cyclones is several hundred kilometers across. Assuming a 320 km square is needed to capture the storm structure, which corresponds to 80 by 80 pixels from the 4 km resolution satellite imagery, for a total of 6400 pixels. Thus, the dimension of a single IR image (6400) is greater than the number of available samples of data (4109). Therefore, the dimensions of the IR data will need to be greatly reduced. Principal Component Analysis (PCA) will be used to extract basic patterns from the imagery. The PCA development is also described in this section.

The eye classification method will provide an estimate of which group (eye or no eye) each case belongs to. Several verification metrics will be used to evaluate the accuracy of the method. These metrics are also described in this chapter.

Rather than writing and debugging new software to perform PCA, LDA and QDA, open source python libraries were used. The PCA implementation was provided by the Modular Toolkit for Data Processing [10]. Additionally, the LDA and QDA implementation was provided by scikit-learn [11].

3.1. QUADRATIC AND LINEAR DISCRIMINANT ANALYSIS

QDA is a machine learning technique for performing classification. In the problem under discussion here, each case belongs to either the “Eye Present” or “Eye Absent” class. A

vector of input parameters from the best track and IR satellite images is used to estimate to which class each case belongs. The algorithm is developed using cases from a training set, and then evaluated using a testing set. LDA is a special case of QDA.

Let x represent the vector of length d containing the input variables. Then, we wish to know the probability that a case belongs to class k , given x , or $P(C = k|x)$. From the training set, the class of each case is known, so it is possible to estimate the probability distributions of x for each class, which are given by $P(x|C = k)$. Bayes Rule provides the link between the above two distributions and can be written as:

$$(1) \quad P(C = k|x) = \frac{P(x|C = k)P(C = k)}{P(x)}$$

The denominator of Equation 1 can be calculated by summing the contributions from each class so that it can also be written as:

$$(2) \quad P(C = k|x) = \frac{P(x|C = k)P(C = k)}{\sum_{k=1}^K P(x|C = k)P(C = k)}$$

For QDA it is assumed that the probability distribution $P(x|C = k)$ can be represented by a multivariate Gaussian distribution given by:

$$(3) \quad P(x|C = k) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1}(x-\mu_k)}$$

where Σ_k is the covariance matrix of x for class k , and μ_k is a vector containing the means of the elements of x for class k . Once the Gaussian distributions are fitted, they can be substituted into Bayes Rule so $P(C = k|x)$ can be determined. For a given vector x , the class the vector belongs to can be assigned by examining $P(C = k|x)$ for all k classes and selecting the class with the largest value of P . The form of $P(C = k|x)$ can be simplified by taking the natural log of Equation 2 after Equation 3 is substituted, which gives the discriminant function $\delta_k(x)$

$$(4) \quad \delta_k(x) = -\frac{1}{2} \ln |\Sigma_k| - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k) + \ln P(C = k)$$

where

$$(5) \quad P(C = k) = \frac{N_k}{N}$$

and N is the total number of samples and N_k is the total number of samples belonging to class k . For the two class problem, the boundary between classes is defined by $\delta_1(x) = \delta_2(x)$, which is a quadratic function of x .

Depending on the size of the training sample relative to the number of input variables d , it may not always be possible to accurately estimate the covariance matrices for each class. Thus, an additional assumption can be made that the covariance matrices for each class are the same $\Sigma_k = \Sigma$, where Σ is calculated as the class-sample weighted average of the covariance matrices for each class. This method is called LDA, and the discriminant function reduces to

$$(6) \quad \delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \ln P(C = k)$$

Equation 6 is linear in x , so this method is called LDA. In some cases, the surfaces represented by the linear discriminant functions may perform better classification than the quadric surfaces. It is common to use both LDA and QDA and compare performance.

The discriminant function for LDA in Equation 6 also provides a way to estimate the relative contributions of each of the inputs to the classification in the two-class eye algorithm. Consider a single component of the vector x , denoted by z . If we define the sensitivity of the difference of the LDA discriminant functions to z as $\frac{\partial}{\partial z}[\delta_2(x) - \delta_1(x)]\Delta z$ where Δz is the standard deviation of component z , then a sensitivity vector can be defined as:

$$(7) \quad S(x) = \sigma \Sigma^{-1} (\mu_2 - \mu_1)$$

where σ is a $d \times d$ diagonal matrix where the diagonal elements are the standard deviations of each component of x .

3.2. PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis is a statistical technique routinely used in the field of image processing and other domains where high dimensional data needs to be analysed. Given a matrix of data such as a collection of images where each row represents an image and each column represents a pixel within an image, PCA finds a set of basis vectors that explain variance within the data. To perform PCA, first the data is normalized by subtracting the mean from the data and dividing by the standard deviation. Then, the covariance matrix

of the normalized data is computed. The eigenvectors and eigenvalues are computed from the covariance matrix. There will be one eigenvector and one eigenvalue for each dimension of the original data. The magnitude of the eigenvalue indicates how much variance can be explained by each eigenvector. The set of eigenvectors and eigenvalues have the useful property of being able to perfectly reconstruct the original data by multiplying the matrix of eigenvectors by a vector containing the eigenvalues. If only a subset of the eigenvectors and eigenvalues are used in this operation then the output will be an approximation of the original dataset. This property can be used to reduce the dimension of the original dataset. By constructing a matrix of only the most significant of the eigenvectors and multiplying the original data by this matrix, the data will have its dimension reduced while still retaining most of its information. Reducing the dimension of data may assist other statistical pattern recognition techniques.

In the field of meteorology, eigenvectors are also known as Empirical Orthogonal Functions (EOF).

3.3. VERIFICATION

3.3.1. ERROR METRICS. To evaluate the quality of the eye detection, the data samples were randomly shuffled and partitioned so 70% of the data would be used for training and 30% would be used for testing. Once the algorithm is trained, it can be used to generate estimated classifications for each sample in the testing set. These estimated classifications can then be compared to the “true” classifications associated with each sample. As defined in Table 3.1, each estimate will be considered one of the four categories based on whether it agrees with the true value. While true positives and true negatives are considered to be correct, false positives and false negatives are considered errors.

TABLE 3.1. Table of confusion defining how estimates are considered correct or errors. Estimates that are considered true positives and true negatives are considered correct, while false positives and false negatives are considered errors.

	Estimated “eye-present”	Estimated “eye-absent”
True “eye-present”	True Positive	False Negative
True “eye-absent”	False Positive	True Negative

Once the estimated classifications have been categorized for the testing data, the total number of samples belonging to each category in the confusion matrix can be counted. From these sums, various statistics can be calculated and used to analyze the performance of the algorithm. The following measures were used in this project.

3.3.1.1. *Fraction Correct.*

$$(8) \quad \textit{FractionCorrect} = \frac{N_{\textit{TruePositives}} + N_{\textit{TrueNegatives}}}{N_{\textit{Samples}}}$$

The fraction of the testing data that was correctly classified offers a simple view of the accuracy of the algorithm, but is not enough to analyze the algorithm’s performance. Since the distribution of the data consists mostly of “eye absent” cases, a scheme with no skill could easily have a high fraction of correct classifications by simply classifying samples as “eye absent” most or all of the time. Therefore, other statistics are necessary to get a clear view of the algorithm’s accuracy.

3.3.1.2. *True Positive Fraction.*

$$(9) \quad \textit{TruePositiveFraction} = \frac{N_{\textit{TruePositives}}}{N_{\textit{ObservedEyePresentSamples}}}$$

The true positive fraction was used to examine the probability that the algorithm would correctly classify an “eye-present” sample.

3.3.1.3. *True Negative Fraction.*

$$(10) \quad \text{TrueNegativeFraction} = \frac{N_{\text{TrueNegatives}}}{N_{\text{ObservedEyeAbsentSamples}}}$$

The true negative fraction was used to examine the probability that the algorithm would correctly classify an “eye-absent” sample.

3.3.1.4. *False Positive Rate.*

$$(11) \quad \text{FalsePositiveRate} = \frac{N_{\text{FalsePositives}}}{N_{\text{ObservedEyeAbsentSamples}}}$$

The false positive rate was used to examine the probability that the algorithm would incorrectly classify an “eye-absent” sample.

3.3.1.5. *False Negative Rate.*

$$(12) \quad \text{FalseNegativeRate} = \frac{N_{\text{FalseNegatives}}}{N_{\text{ObservedEyePresentSamples}}}$$

The false negative rate was used to examine the probability that the algorithm would incorrectly classify an “eye-present” sample.

3.3.1.6. *Peirce Skill Score.*

$$(13) \quad \text{PeirceSkillScore} = \frac{N_{\text{TruePositives}}N_{\text{TrueNegatives}} - N_{\text{FalsePositives}}N_{\text{FalseNegatives}}}{(N_{\text{TruePositives}} + N_{\text{FalseNegatives}})(N_{\text{FalsePositives}} + N_{\text{TrueNegatives}})}$$

Using the Peirce Skill Score (PSS) [12], the skill of the algorithm can be compared to that of a hypothetical no skill scheme that would randomly classify samples based on the distribution of the data. If the algorithm performed classification perfectly, the output of the formula would be 1. An algorithm that performed equally to the no skill scheme would

receive a PSS of 0. Finally, an algorithm that performed worse than the no skill scheme would receive a negative score. The PSS is especially useful for comparing different versions of the algorithm (for example, LDA versus LDA) since it includes all four elements of the confusion matrix in a single parameter.

3.3.1.7. *Brier Score.*

$$(14) \quad \text{BrierScore} = \frac{1}{n} \sum_{k=1}^n (y_k - o_k)^2$$

The Brier Score [12] is defined in Equation 14 where n is the number of samples, y_k is the algorithm’s estimated probability that a sample k belongs to the “eye-present” class, and o_k is 0 if no eye was observed for the sample and 1 if an eye was observed for the sample. This measure is essentially equivalent to the mean squared error. However, it is important to note that this operates on the probabilities produced by LDA and QDA rather than just the classifications they produce. A valuable use for the algorithm would be to produce these probabilities rather than a classification. Therefore, a measure of error for the probabilities is necessary to evaluate the quality of the algorithm. These can be obtained from the multivariate probability distributions utilized in the QDA and LDA derivations as described above in Section 3.1.

It is useful to compare the BS to a BS from a reference probability forecast. This comparison produces a Brier Skill Score (BSS). If no other information was available, the probability of an eye in the training set would be a reasonable choice. Letting the BS with the training set probability as the estimate be given by BS_t , then the BSS is calculated from:

$$(15) \quad \text{BSS} = \frac{BS_t - BS}{BS}$$

With the definition in Equation 15, a positive BSS is the fraction reduction (improvement) of the BS from a model relative to a constant probability prediction. If the BSS is negative, then the model had worse performance than the constant probability estimate.

3.3.1.8. *Accounting For Shuffling.* As described above, the data is randomly shuffled and split between a testing set and a training set. It is possible that the performance metrics described above may be unusually high or low for a particular shuffling of the data. For example, it is possible, though unlikely, that a shuffling could place the vast majority of the “eye-present” samples inside the testing set and therefore skew the error metrics towards unusually poor performance. Therefore, many shufflings are required in order to evaluate the performance of the algorithm while minimizing the impact of any particular shuffling of data. To do this, typically the algorithm is run 1000 times. Each time, the data is randomly shuffled. The above error metrics are calculated for each shuffling and then averaged among the 1000 runs.

CHAPTER 4

ALGORITHM DEVELOPMENT

To better understand the behavior of LDA and QDA for the eye classification algorithm, three models of increasing complexity were developed. First, a simple baseline model was developed using only information that is available from the best track. The baseline model is also used to illustrate the verification metrics and analyze how different input parameters influence the classification. In meteorological applications, simple baselines using only best track input are commonly used to evaluate the performance of more general methods. For example the CLImatology and PERsistence (CLIPER) track forecast model [13], and Statistical Hurricane Intensity FORcast (SHIFOR) model [14] are routinely used to determine the skill of more general prediction systems. If the errors of a given model are lower than those from the corresponding baseline for the same forecast cases, that model is considered skillful [15]. The baseline eye detection classification algorithm will be used to measure the skill of the two more general eye classification models. The second eye detection model will only include input from the IR satellite data and the third model will include data from the best track and satellite data. The details of the development of these algorithms are described here.

4.1. BASELINE ALGORITHM

Similar to the CLIPER and SHIFOR models, the input to the eye baseline algorithm consists of restricted parameters from the best track or those that can be derived from it. Those include the Julian day, the latitude and longitude of the storm center and the maximum winds. The change in maximum wind, and storm motion vector are also used, and are described below.

To calculate the change in the maximum wind speed, the following formula was used:

$$(16) \quad \Delta_{vmax} = vmax_t - vmax_{t-12}$$

where $vmax_t$ is the maximum wind speed interpolated to the time of the sample from the best track. Similarly $vmax_{t-12}$ is the maximum wind speed interpolated to the time of the sample minus twelve hours. If this time is before the track starts, then the value for the $vmax_t = 12$ is used instead.

To calculate the translational speed in knots, the following formula was used for the components of motion:

$$(17) \quad u = \frac{(lon_t - lon_{t-12}) \cos(lat_t) 60}{12}$$

$$(18) \quad v = \frac{(lat_t - lat_{t-12}) 60}{12}$$

where u is the eastward component of motion and v is the northward component. Again, the lon_t, lat_t are interpolated to the time of the sample and the lon_{t-12}, lat_{t-12} are interpolated to the time of the sample minus twelve hours. If this time is before the track starts, then the motion components are calculated from the latitude and longitude at $t=0$ and $t=12$ hr, similar to the maximum tendency calculation.

The ability to determine whether or not an eye is present for a given case relies on differences in the probability distributions of the two classes as a function of the input parameters. Figure 4.1 shows the PDFs of the eye and no-eye cases as functions of the seven

inputs to the baseline model. This figure shows that there are large differences in the PDFs for maximum wind, with eye cases being much more likely for stronger storms. There are some differences in the right tails of the distributions as a function of maximum wind change. A storm is more likely to have an eye if it has been getting stronger during the past 12 hr. The PDFs for the eye and no eye cases as functions of latitude and longitude do not show large differences, although there is some tendency for an eye to be more likely in the latitude band from 20 to 25 degrees north and in the longitude band from 50 to 70 degrees west. There is not much difference in the PDFs for the northward component of storm motion, but TCs moving westward (negative values of the eastward component) are more likely to have eyes. The peak of the Atlantic hurricane season is in September. TCs are more likely to have an eye when they occur near the peak of the season.

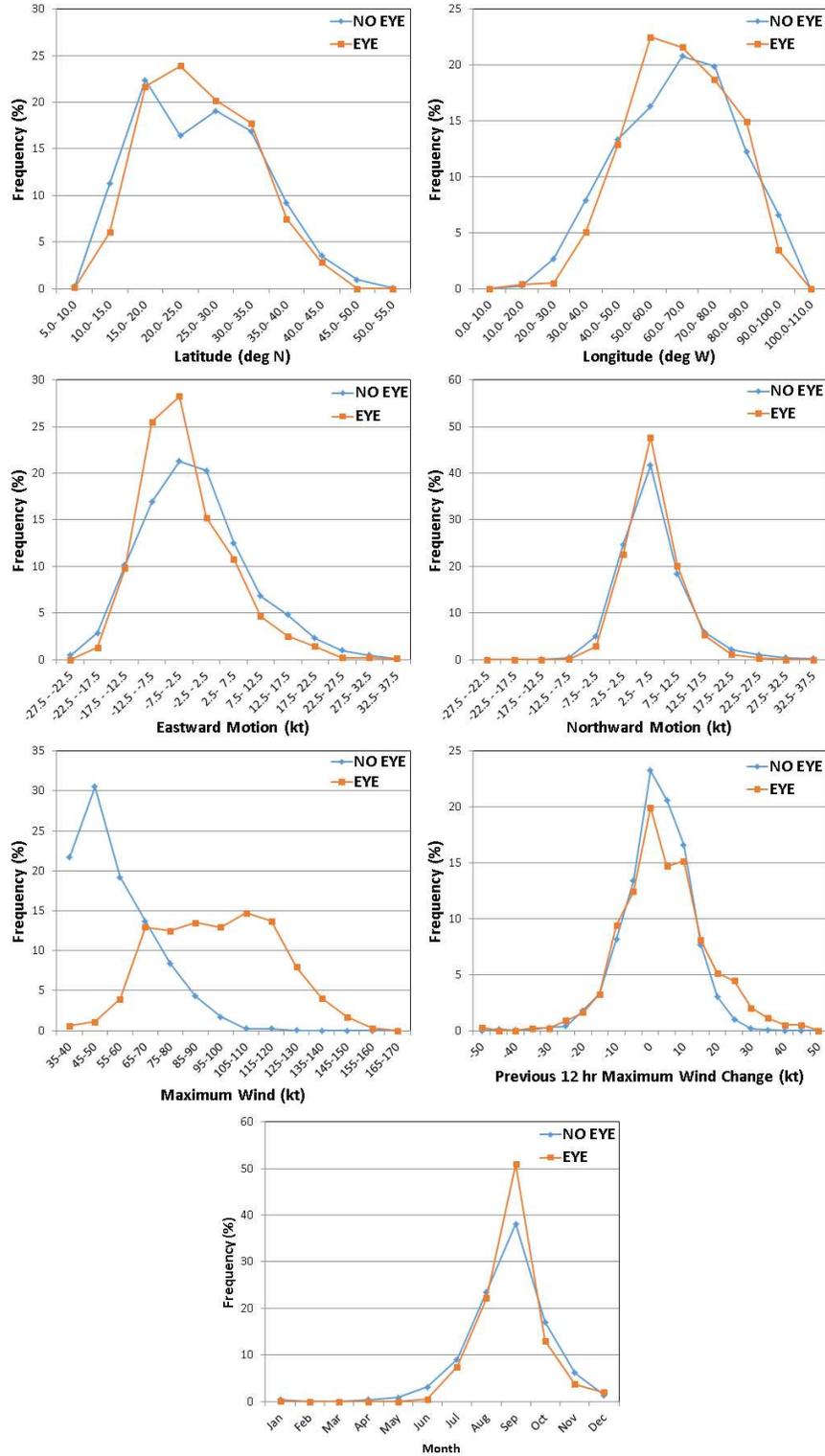


FIGURE 4.1. Probability distributions as functions of the seven baseline model inputs.

Using the steps described in Table 4.1, the algorithm was run with these seven inputs using both LDA and QDA. As described in Chapter 3, the average error statistics were found using 1000 runs of randomly selected testing samples. These error statistics can be seen in Figure 4.2. As can be seen in the figure, the overall performance was very good, with both LDA and QDA having an average accuracy of about 88%. QDA seemed to favor classifying samples as “eye-present” more than LDA, and achieved a higher true positive fraction as well as a higher false positive rate. The true positive fraction of roughly 0.68 for LDA and 0.69 for QDA indicates that the algorithm was able to successfully separate the two classes since only 19% of the data are “eye-present” cases. The BSS and PSS of roughly 0.5 and 0.6 also indicate that the algorithm has predictive skill better than naive models based purely on the distributions of classes within the training data. This represents an excellent starting point for the algorithm as the baseline is already fairly accurate.

The verification results illustrate why the PSS, which includes all the elements of the confusion matrix, is more useful for comparing different versions of the algorithm than some of the metrics that only include a subset of the elements. For example, Figure 4.2 shows that QDA has a higher true positive fraction, but a lower true negative fraction. The PSS shows that the QDA has slightly more skill than LDA.

TABLE 4.1. Steps used for the baseline algorithm.

- (1) A matrix of N samples and D dimensions is prepared with a companion vector of N corresponding “true” classifications.
- (2) The matrix of samples and companion classifications are randomly shuffled and split into a “training” set and a “testing” set.
- (3) The mean value and standard deviation are found for each predictor using the training data.
- (4) Each predictor in both the training data and testing data are normalized by subtracting its mean and dividing by its standard deviation.
- (5) Either LDA or QDA is trained against the training data and corresponding truth values.
- (6) Classification is performed on the testing data.
- (7) The classifications produced by the trained LDA or QDA implementation are compared to the “true” classifications.

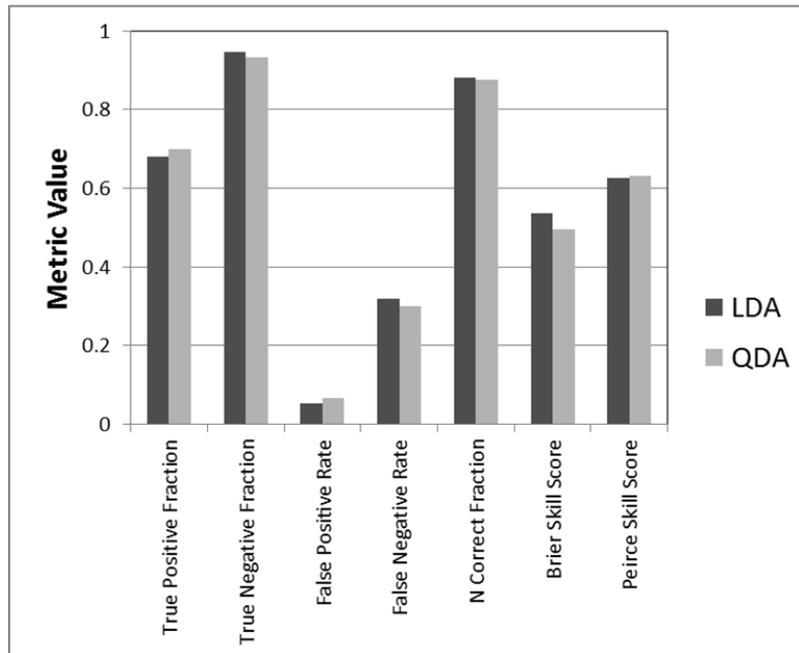


FIGURE 4.2. Average error statistics generated from 1000 runs of the Baseline algorithm.

Using the entire dataset with Equation 7, the sensitivity vector was generated to analyze which of the seven input variables have the most impact on the classification. Examining Figure 4.3, it can be seen that maximum wind speed and latitude have the most effect on the classification. This is consistent with the histograms in Figure 4.1.

The sensitivity vector was defined in terms of the eye class discriminant function minus the no-eye class function. With that definition, when a variable is larger, an eye is more likely. Figure 4.3 shows that an eye is more likely when the storm is stronger, has intensified in the past 12 hr, is further north, is further west (longitude was input as degrees west positive), and is later in the year. The negative values for the motion vector components show that an eye is less likely for storms moving east and north.

To get a clear view of how the algorithm was performing its classification, just the maximum wind speed and latitude were fed into the algorithm. These were the two most important inputs in Fig. 4.2. Using the entire data sample, the algorithm was trained and a classification was produced for each case. Figure 4.4 was produced from these classifications. It can be seen that the boundary in both QDA and LDA was primarily placed horizontally across the figure. This is consistent with the sensitivity vector as well as the histograms as the maximum windspeed information dominates the classification. As a result, QDA and LDA included almost the same set of samples in each region. An interesting feature of the data that can be seen here is that the region between 60kt and 90kt has many “eye-present” samples overlapping with large number of “eye-absent” samples. Within this range, data appears to be difficult to separate by maximum windspeed or latitude. Given that QDA has a higher true positive fraction, it is likely that the other five predictors are being used to create a non-linear surface that more effectively separates the data. However, it does not appear that six of seven predictors contribute a large amount to the separability of the classes. Additional data, such as the brightness temperatures provided by IR imagery, are clearly needed in order to more effectively separate the classes.

The LDA boundary shows that the maximum wind threshold between the eye and no eye cases decreases with increasing latitude. However, the QDA boundary has some curvature,

which shows that for very high latitudes, the maximum wind threshold that divides the eye and no eye cases increases again.

The verification statistics were calculated for the two input version of the algorithm. Results show that they were nearly the same as those for the seven input version. The PSS did decrease by about 0.01 for both the LDA and QDA. The similar values of the verification metrics for the two input version compared with the seven input version is consistent with the sensitivity matrix components in Figure 4.3, which showed most of the variability of the linear discriminant function difference is due to just one variable.

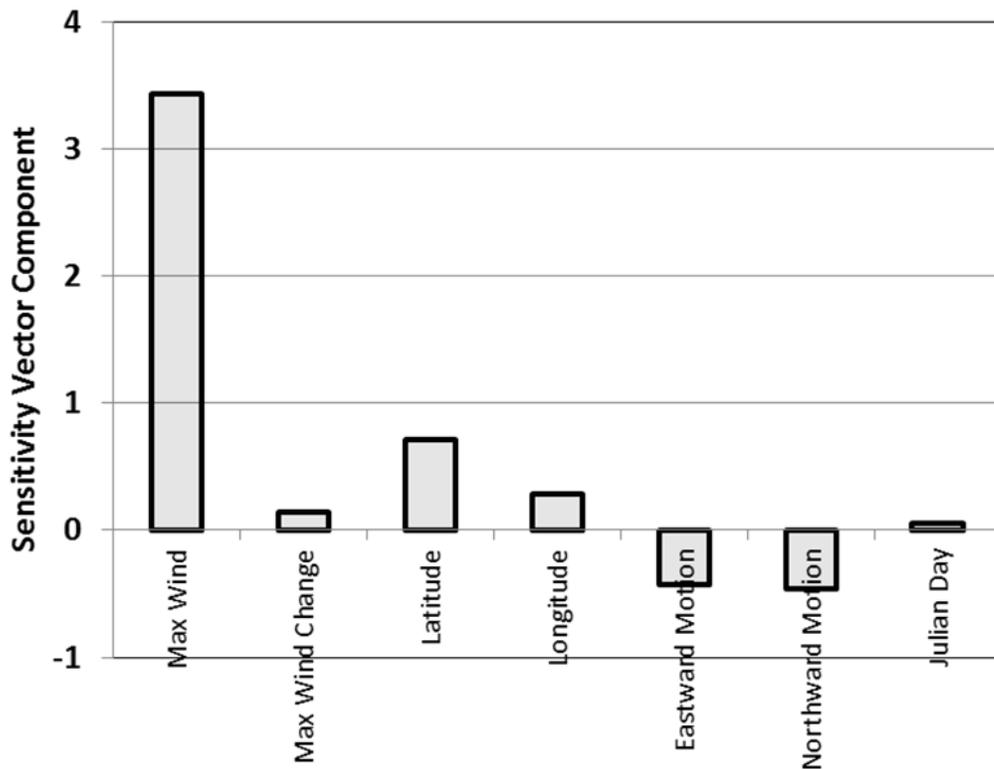


FIGURE 4.3. Sensitivity vector for the baseline algorithm.

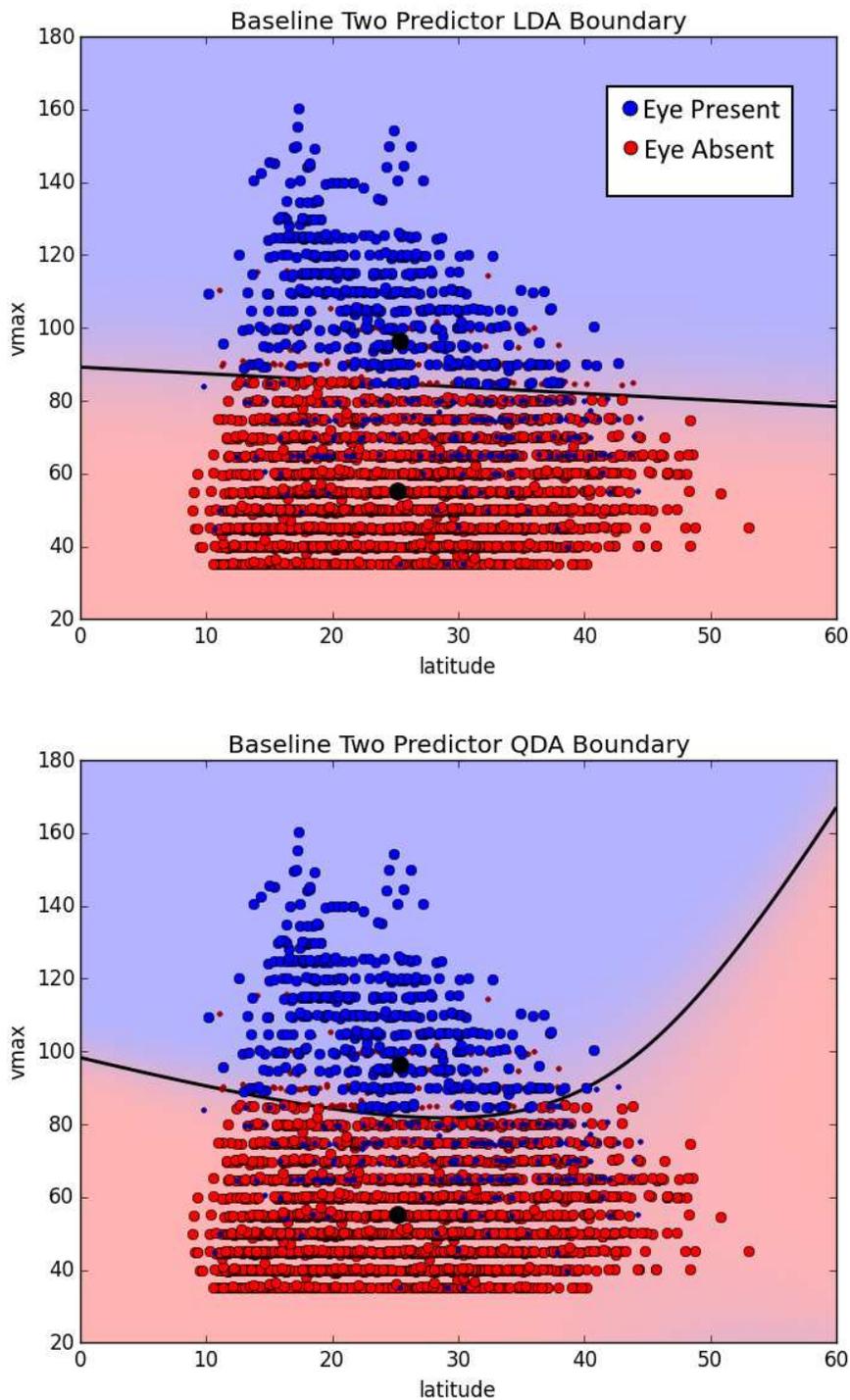


FIGURE 4.4. Boundary of the “eye-present” and “eye-absent” classes produced from all 4109 available samples. Only maximum winds and latitude were used so that a 2D plot could be examined. Probability of each class can be seen in the color intensity. Correct classifications appear as large filled dots, incorrect classifications appear as small dots. Large black dots indicate the location of the means of each class.

4.2. SATELLITE DATA PROCESSING

The next step in the eye classification algorithm is to determine if the IR satellite data can improve upon the baseline model. As described in Section 2.4, a total of 4109 samples remained after removing the cases in which TCs had maximum winds less than 35 kt. The IR imagery such as that shown in Figure 1.2 typically includes $800 \times 600 = 480,000$ pixels, which is two orders of magnitude larger than the number of training samples. The inner part of the tropical cyclones that depict the eye structure covers a small fraction of each image. To reduce the dimension of the IR data, the input was restricted to 80×80 pixels centered on the best track storm positions. Figure 4.5 shows examples of the sub-area for Hurricane Katrina. Even with the restricted area, each image contains more pixels than the size of the training set. For this reason, PCA was applied to the imagery to further reduce its dimension so that a small number of parameters can be used as input to the LDA and QDA classification algorithms.

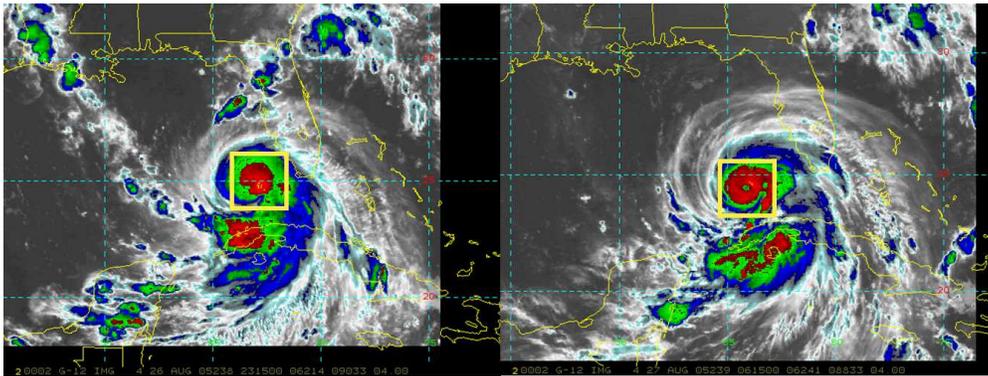


FIGURE 4.5. Example IR images from Hurricane Katrina. Boxes show the selection of pixels used with the algorithm. Image classified as eye absent (left). Image classified as eye present (right)

The first step in PCA is to calculate “mean” and “standard deviation” images. That is, using all the samples in the training data, the mean and standard deviation are calculated for each of the 6400 pixels. The mean image is then subtracted from each sample, and similarly

each sample is divided by the standard deviation image. The subtraction of the mean image and the division by the standard deviation image is performed on both the training data and the testing data. For illustrative purposes, the mean and standard deviation images were generated from the entire dataset. These images can be seen in Figure 4.6. The mean image shows a fairly circular temperature pattern, with some evidence of a local warm region near the center of the colder area covering most of the middle of the area. This suggests an eye structure. The standard deviation is largest near the center, which is probably due to the mixture of eye and no eye images, so the temperature variations are largest near where they would be located.

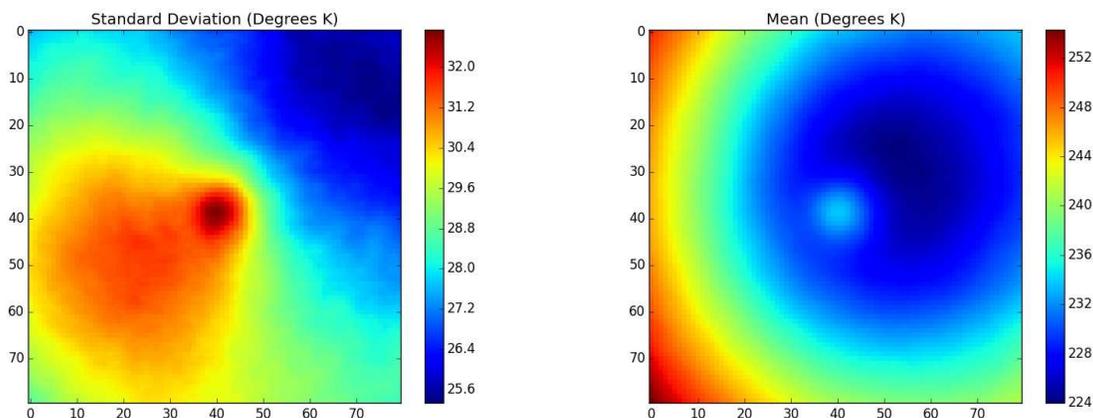


FIGURE 4.6. The mean image and standard deviation image produced from the entire set of IR data samples available for use with this project.

Once the mean has been subtracted and the images are normalized by the standard deviation image, the PCA outlined in section 3.2 is applied to find a set of eigenvectors. From these eigenvectors, a subset is selected which explain the majority of the variance in the original dataset. The 25 most important eigenvectors can be seen in Figure 4.7 and Figure 4.8. The variance explained by each of these eigenvectors as well as the accumulated variance can be seen in Figure 4.9. Figure 4.9 shows that the first 25 eigenvectors explain

about 95% of the variance of the IR imagery. Thus, the dimension of the IR data has been greatly reduced by PCA.

Some of the patterns in the EOFs are very symmetric and indicative of an eye structure such as 1,4, 9, 15 and 25. Others show asymmetric patterns. In the subjective version of the Dvorak technique, images are first classified into four basic types: curved band, central dense overcast (CDO), shear, and eye patterns. The curved band pattern has spiral cloud patterns away from the storm center, somewhat similar to EOFs 7 and 8, and are associated with the weakest TCs. The CDO pattern has cold clouds near the center, but without an eye, similar to EOF 4, and is related to storms that are usually a little stronger than those with a curved band pattern. The strongest TCs are associated with the eye pattern. The shear pattern is a special case where a storm is in an area with strong upper level winds, which moves the thunderstorm clouds away from the center. These are usually associated with TCs that will not get much stronger, and are somewhat similar to the structure in EOFs 5 and 6. These patterns suggest that it might be possible to use a machine learning algorithm to automate the scene classification part of the Dvorak technique.

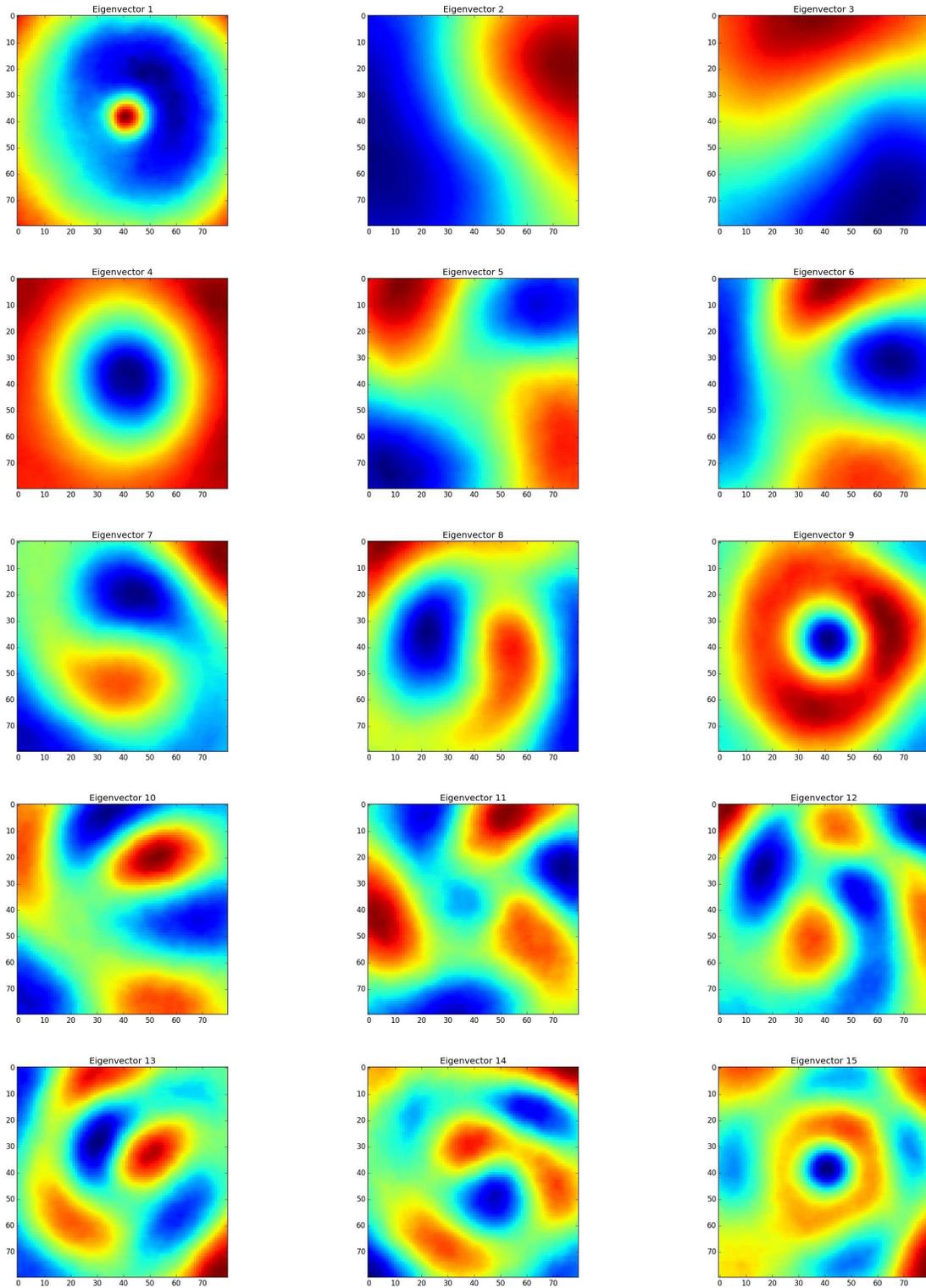


FIGURE 4.7. Eigenvectors 1-15.

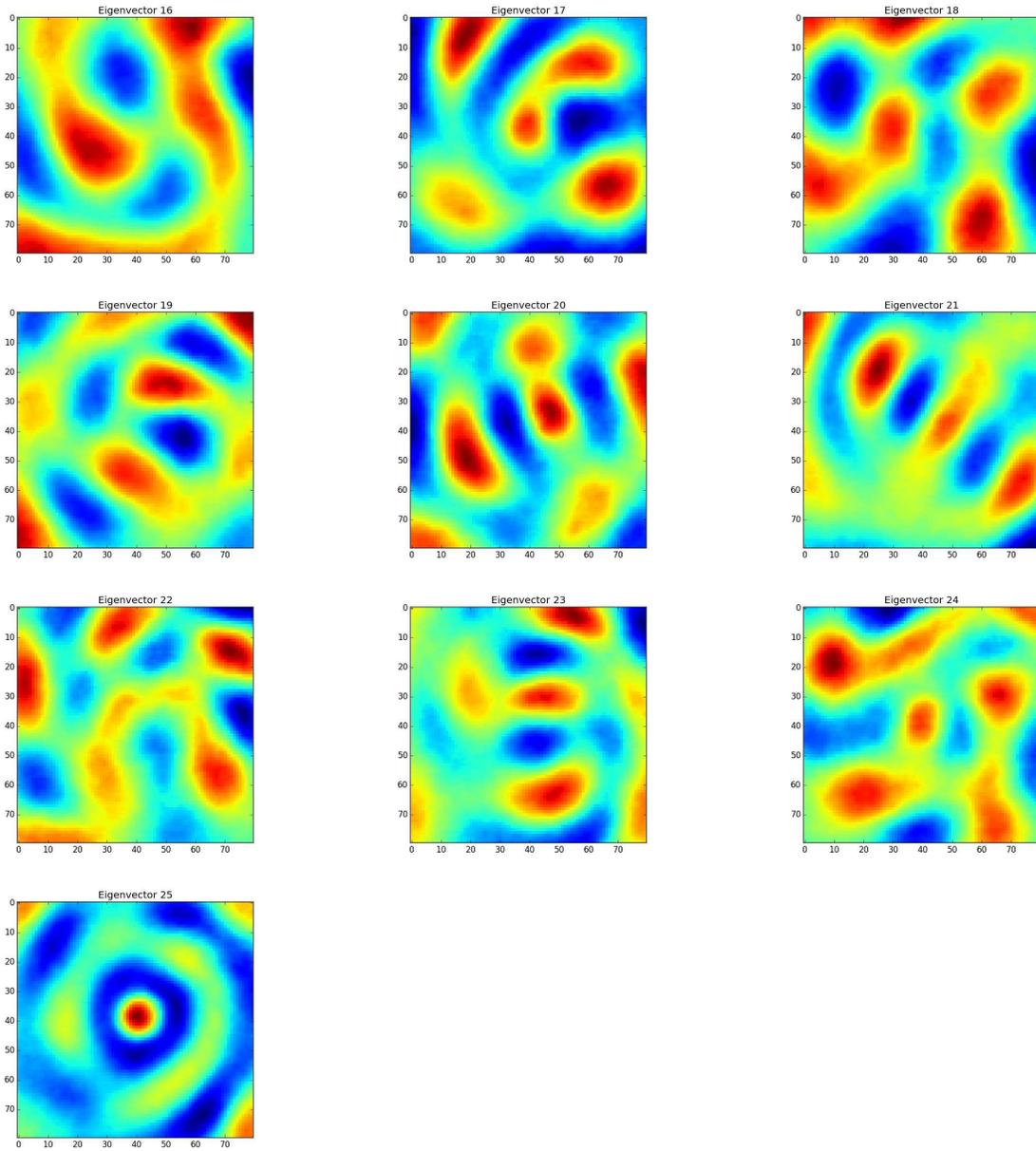


FIGURE 4.8. Eigenvectors 16-25.

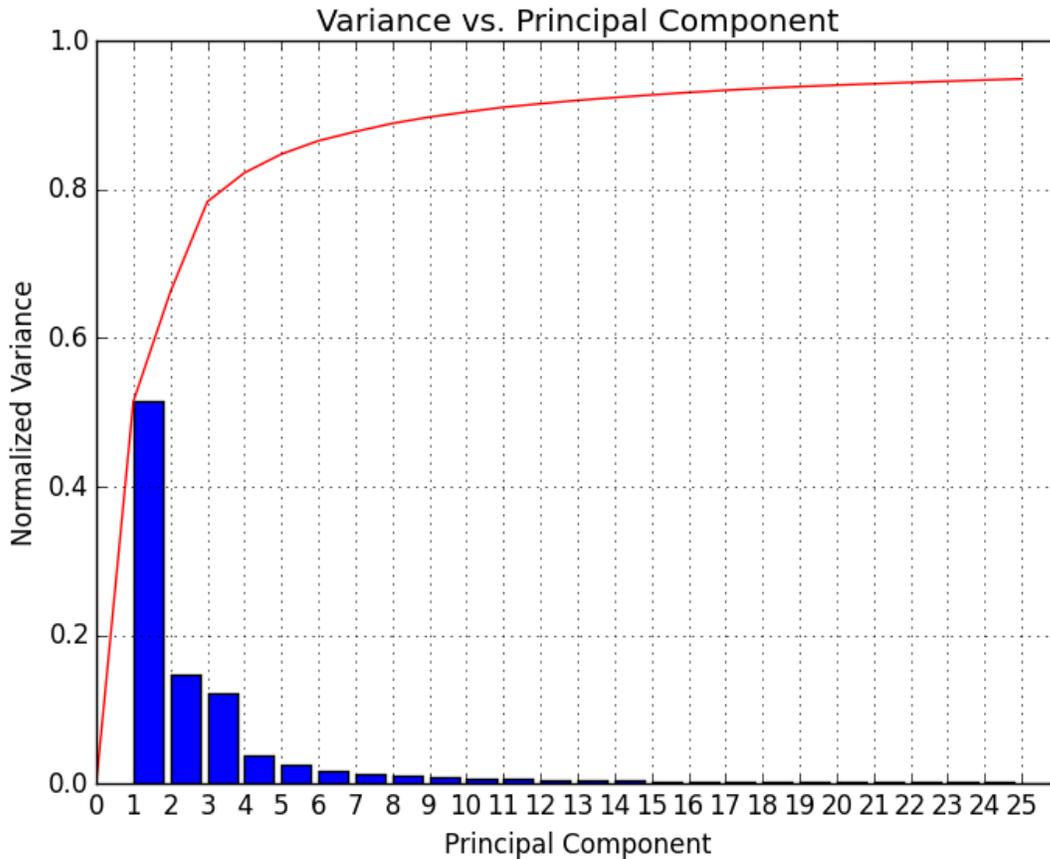


FIGURE 4.9. Normalized variance vs the principal component number. The first few principal components explain the vast majority of the variance of the data. The sum of the normalized variance with 25 PCs is roughly 95%.

In order to verify that the first 25 eigenvectors can accurately reconstruct a variety of patterns that appear in the data, three example cases were selected to be reconstructed using a subset of the eigenvectors. These cases were Hurricane Katrina at 08/28/2005 06:15 UTC, Hurricane Danielle at 08/24/2010 23:45 UTC, and Tropical Storm Arlene at 06/10/2005 11:45 UTC. Katrina was selected to represent cases with clearly visible eyes, Danielle was selected to represent weaker cases with no eye but with cold clouds near the center (a CDO pattern of the Dvorak technique), and Arlene was selected to represent cases in a shear pattern, where most of the thunderstorms were away from the center. The maximum winds

for Arlene, Danielle, and Katrina at these times were 50, 70, and 125 kt, respectively. Figure 4.13 shows the IR imagery for these cases from a wide view, with the 80x80 pixel area indicated by the yellow rectangle. Danielle and Katrina have a fairly circular region of cold cloud tops in the small area, with noticeable warm spot in the center for Katrina. The Arlene case shows cold clouds in the upper left of the small area, with most of the domain covered by temperatures that increase from west to east across the domain.

Figure 4.14 shows the PC values for these three cases. Katrina has a large positive value for PC1 and a large negative value for PC4. As seen in Figure 4.7, these two EOFs have a symmetric structure. The negative PC4 value for Katrina reverses the pattern so the warm area is in the center. In contrast, the Danielle case has a large positive value for PC4, since it has cold temperatures near the center. The Arlene case has a large negative value for PC2. The EOF2 pattern in Figure 4.7 captures the east-west temperature gradient, but with the sign reversed. The negative PC2 value reverses the sign of the gradient.

The reconstruction of these three cases can be seen in Figure 4.10 (Katrina), Figure 4.11 (Danielle), Figure 4.12 (Arlene) with 3, 6, ..., 24 EOFs along with the original images. These figures show that, with only a small number of EOFs, the overall patterns of the original images are reproduced, but with some of the final details smoothed out. The eye structure of Katrina and the CDO pattern of Danielle can clearly be seen in the reconstructed images. The eye pattern in Katrina starts to show up with just a few EOFs.

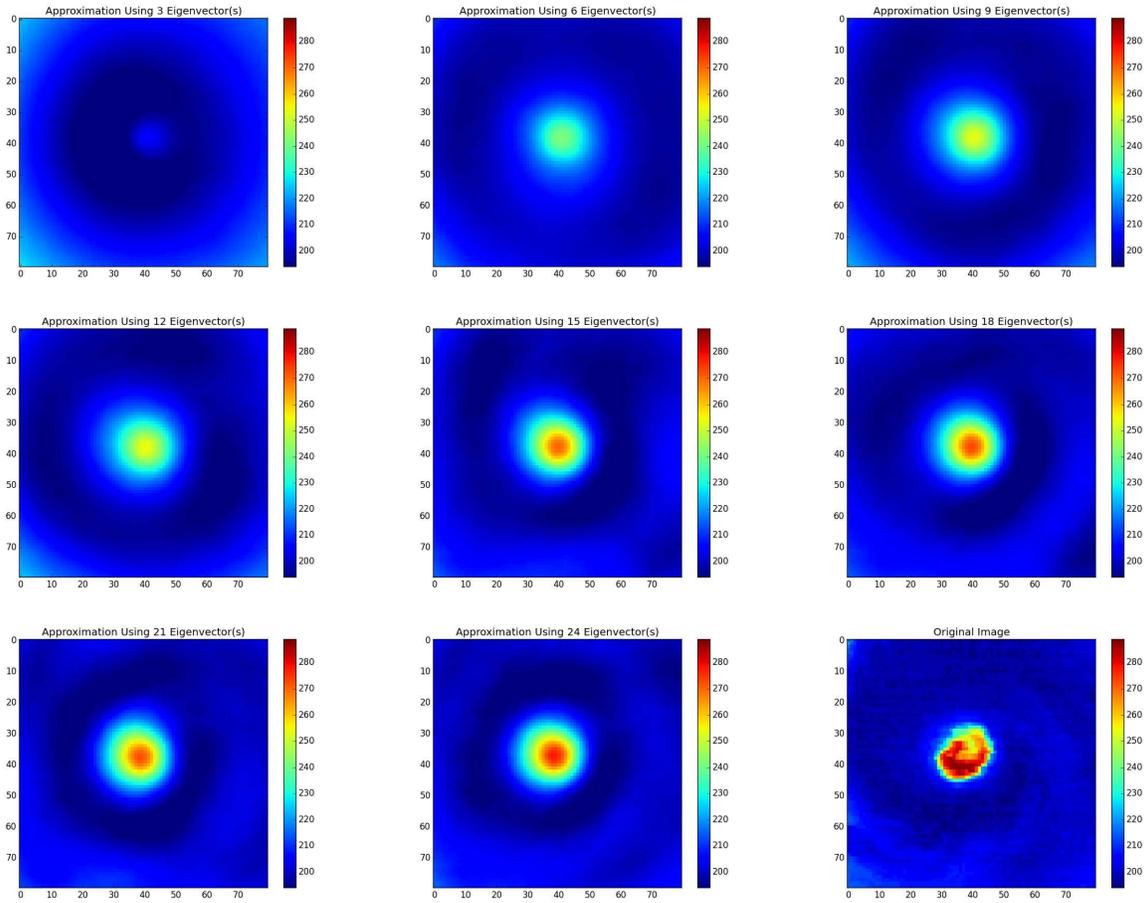


FIGURE 4.10. Reconstruction of a Hurricane Katrina image at 08/28/2005 06:15 UTC using 24 eigenvectors

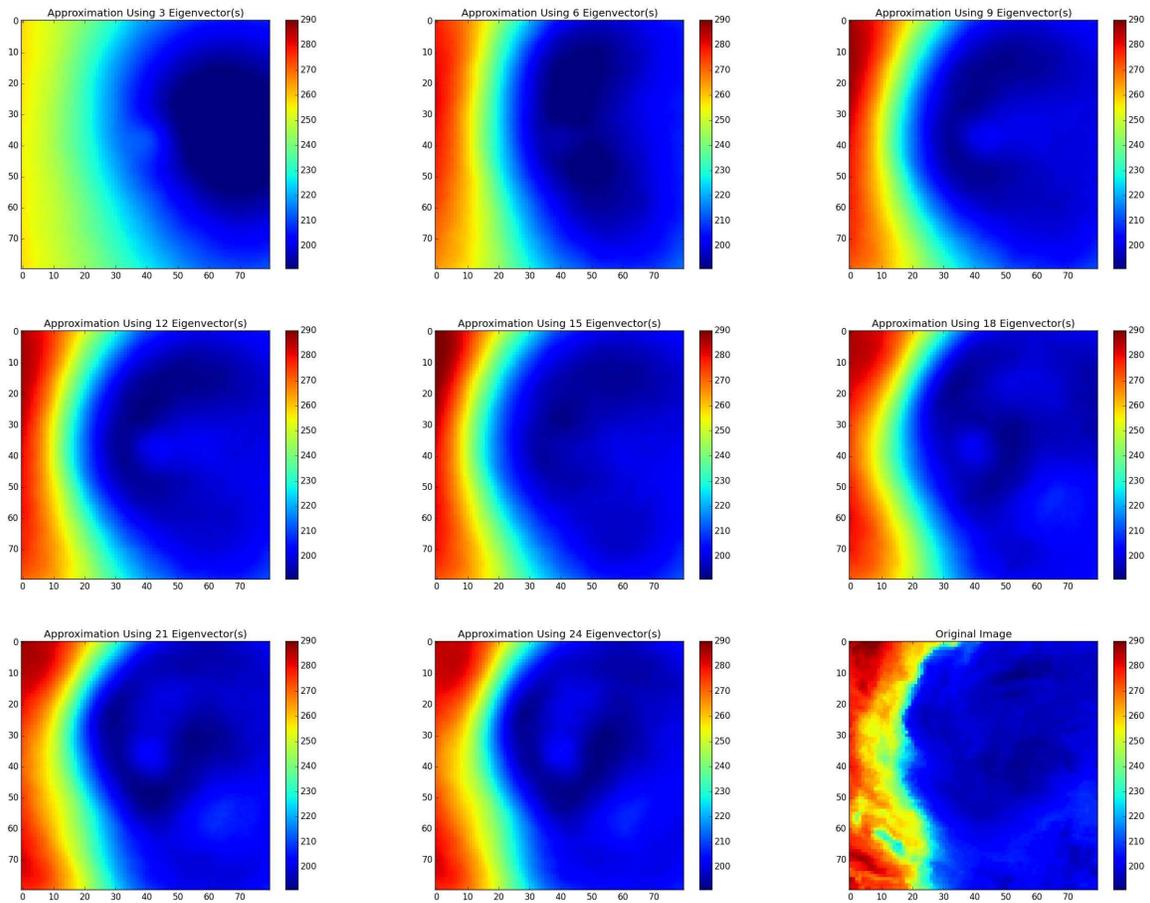


FIGURE 4.11. Reconstruction of a Hurricane Danielle image at 08/24/2010 23:45 UTC using 24 eigenvectors

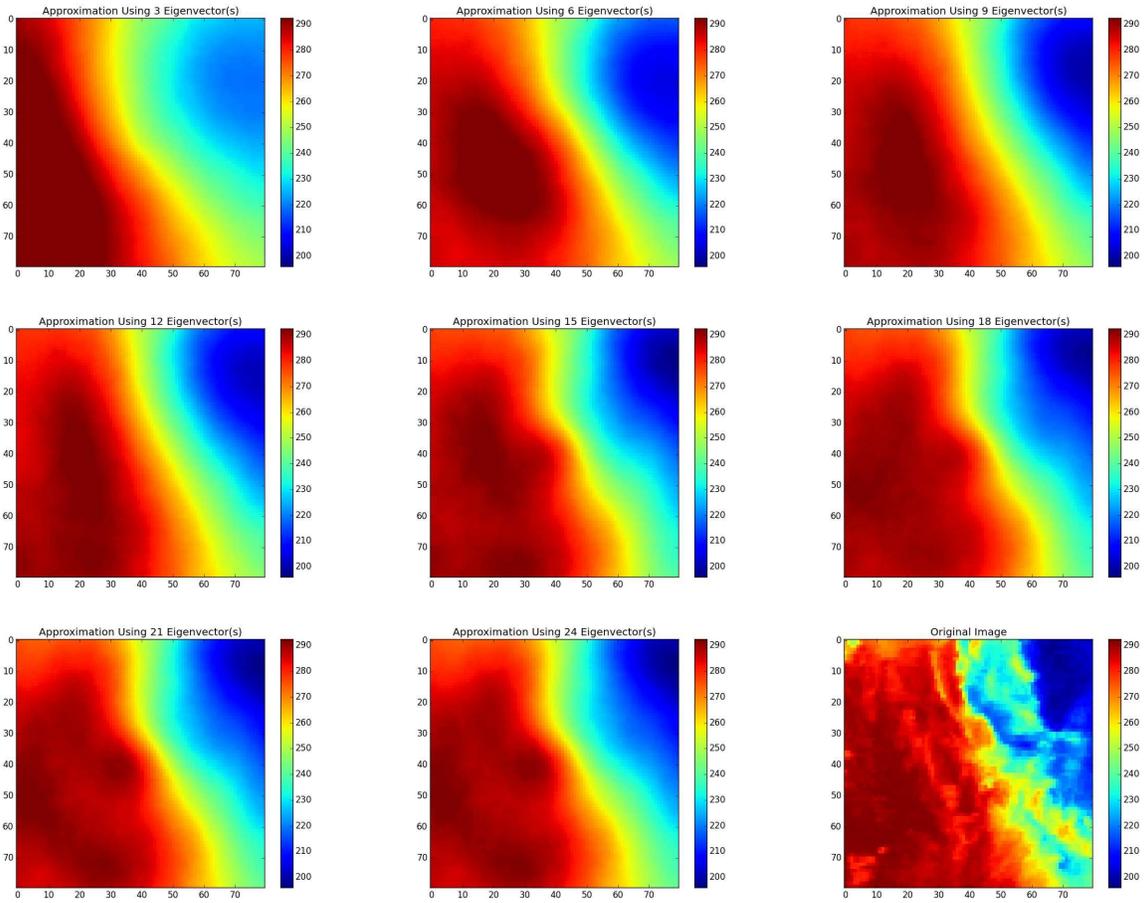


FIGURE 4.12. Reconstruction of a Tropical Storm Arlene image at 06/10/2005 11:45 UTC using 24 eigenvectors

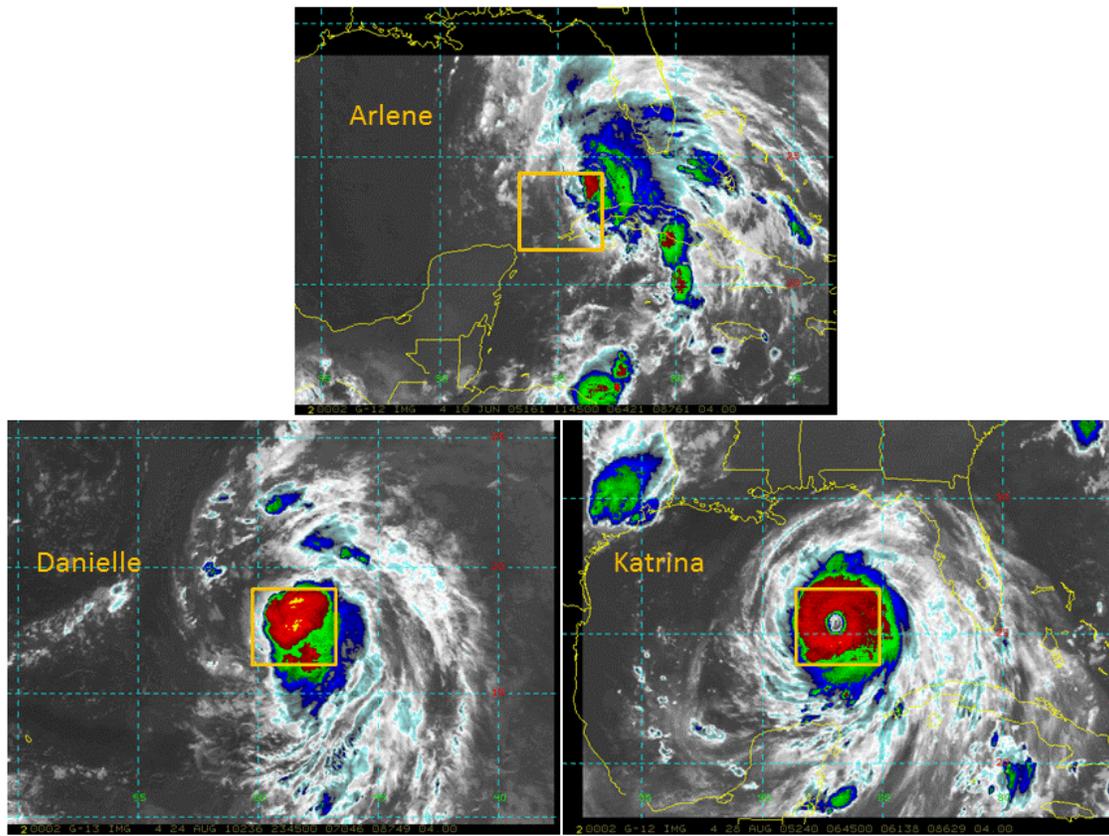


FIGURE 4.13. Wide view of Hurricane Katrina, Hurricane Danielle, and Tropical Storm Arlene. Orange box indicates pixels selected for use with the eye detection algorithm.

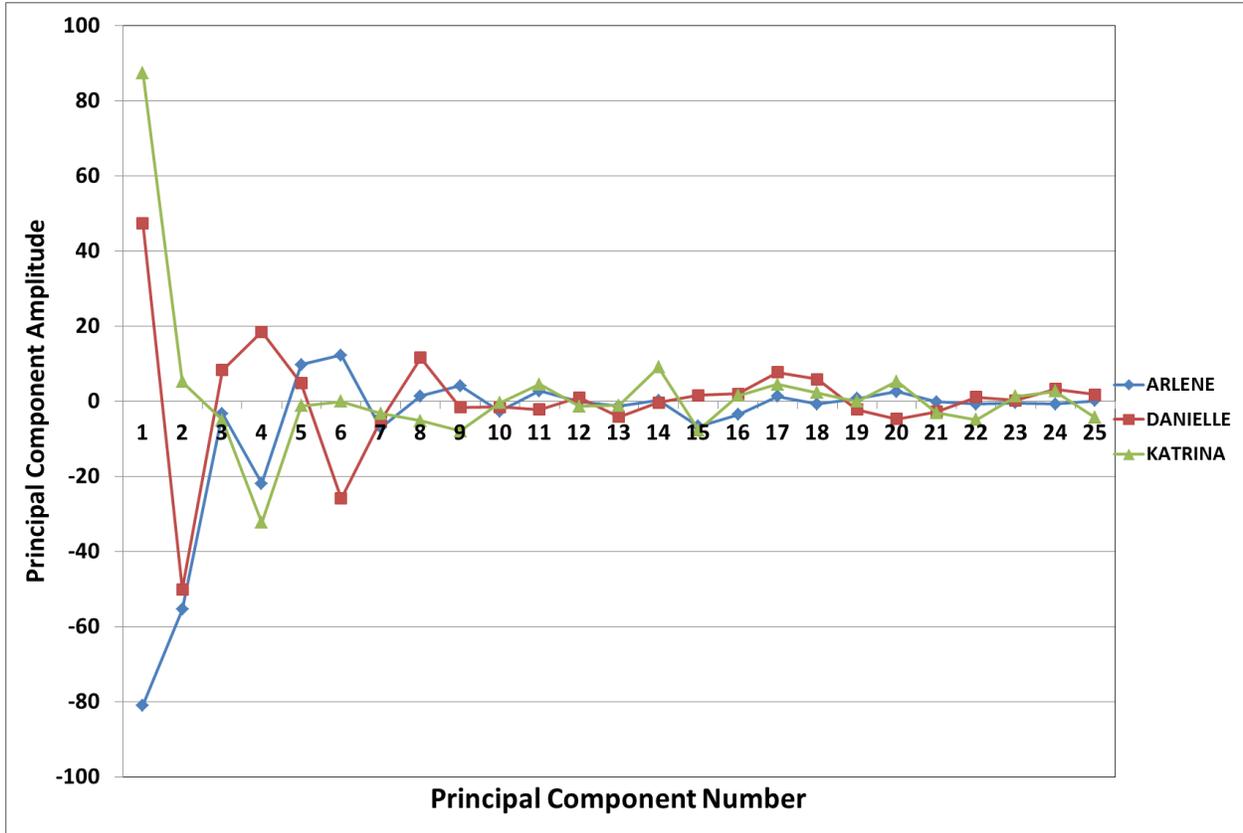


FIGURE 4.14. PC value amplitudes for Hurricane Katrina, Hurricane Danielle, and Tropical Storm Arlene.

4.3. SATELLITE DATA BASED ALGORITHM

The next step in the algorithm development is to determine the performance where the only input is from the IR satellite data. To accomplish this, the algorithm followed the steps outlined in Table 4.2. The key difference between the steps used in the baseline algorithm and the satellite data based algorithm is the addition of PCA for dimension reduction.

There are many possible choices with regard to how many and which PCs to use as input for the satellite algorithm. The starting point was a version with 25 EOFs with the highest explained variance. As described in Section 4.2 the first 25 eigenvectors explain 95% of the variance. Although 25 is much smaller than the original 6400 predictors, further reduction of the number of predictors used might be desirable because the covariance matrix will have 625

TABLE 4.2. Steps used for the satellite data based algorithm.

- (1) A matrix of N samples and D dimensions is prepared with a companion vector of N corresponding “true” classifications.
- (2) The matrix of samples and companion classifications are randomly shuffled and split into a “training” set and a “testing” set.
- (3) The mean value and standard deviation are found for each predictor using the training data.
- (4) Each predictor in both the training data and testing data is normalized by subtracting its mean and dividing by its standard deviation.
- (5) A set of eigenvectors is obtained by performing PCA on the training data.
- (6) The dimension of the training data and the testing data are reduced by projecting the datasets on to these eigenvectors.
- (7) Either LDA or QDA is trained against the reduced dimension training data and corresponding truth values.
- (8) Classification is performed on the reduced dimension testing data.
- (9) The classifications produced by the trained LDA or QDA implementation are compared to the “true” classifications.

elements. In the QDA method separate covariance matrices are calculated for the eye and no eye cases, but there are only 991 eye cases in the total sample. This may be inadequate to accurately estimate the covariance method for the eye sample. Therefore, versions with smaller number of EOFs were also developed.

To avoid arbitrarily selecting a subset of eigenvectors to use, the PSS of the algorithm was examined using a range from 1 to 25 EOF’s. The order of the inclusion of the predictors was determined by the corresponding magnitude of its significance. So the first eigenvector included had the largest magnitude in the significance vector, the second included eigenvector had the second largest magnitude in the significance vector, and so on. The results of that run can be seen in Figure 4.17. Examining Figure 4.17, it can be seen that the magnitude of the PSS levels off after about 10 EOF’S. For this reason, a total of 10 EOF’s with the highest magnitude in the significance vector were chosen to use in another set of runs. This was compared to a 10 EOF run where the included EOFs were simply the 10 EOFs with the highest explained variance.

It should be noted that although the significance vector was generated using the entire dataset, it should not bias the results towards improved performance. The precise magnitudes of the significance vector are unimportant for the purpose of sorting the inputs. Only the relative order of the vector components was used. The order of the significance vector remained the same regardless of whether the entire dataset was used or when significance vectors generated from random shufflings of training data were compared. To simplify the implementation of the algorithm, the order produced using the entire dataset was used.

Figure 4.15 shows the components of the sensitivity vector for the 25 EOF model, sorted by the magnitude of the components. Comparing this figure with Figure 4.7 shows that the majority of the eigenvectors with the highest sensitivity vector component magnitudes have a symmetric structure. Also there are sometimes large differences between the ordering by sensitivity vector magnitude compared with the ordering by variance explained.

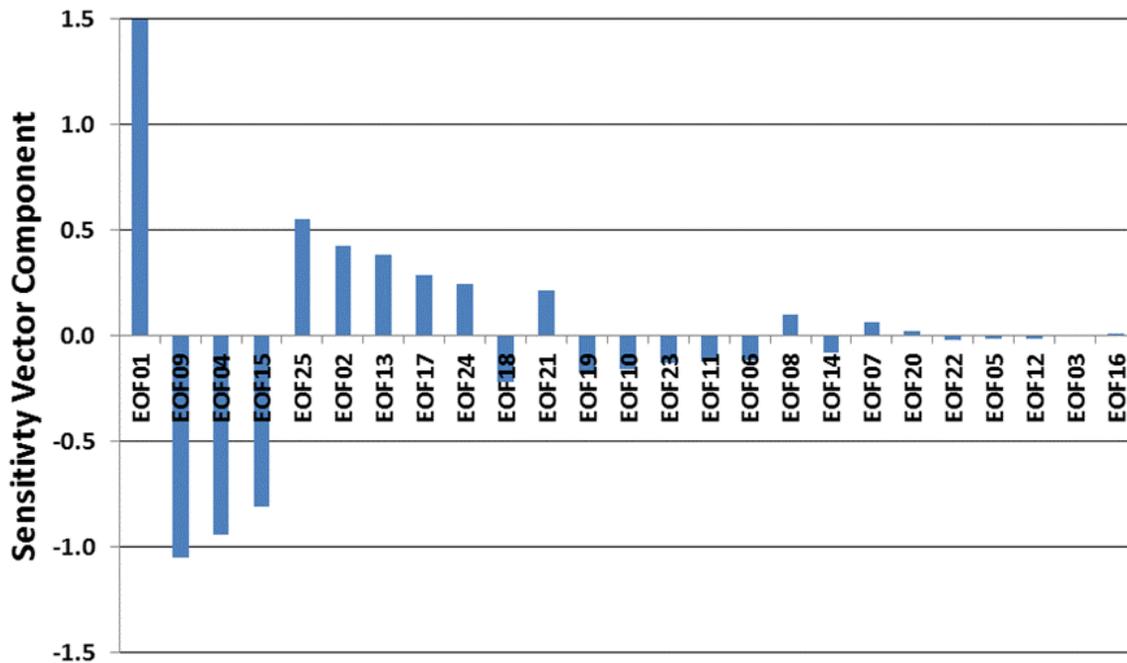


FIGURE 4.15. 25 EOFs with the highest variance explained sorted from largest to smallest sensitivity.

In summary, three versions of the satellite algorithm were developed, with LDA and QDA tested for each. The first includes 25 EOFs that explain 95% of the variance of the IR data. The second version includes the 10 EOFs in order of the magnitude of the sensitivity vector, and the third version includes 10 EOFs in order of the variance explained in the IR dataset. Figure 4.16 shows the verification metrics for the three versions of the satellite algorithm, and reveals a number of interesting results. All of the runs showed skill as measured by the BSS and PSS. The BSS values are all higher for LDA compared to QDA. The BSS uses the probabilities, which depends more on the details of the covariance matrices than the classifications. This may be a limitation of the sample size. Interestingly, QDA tended to have a higher true positive fraction and a smaller true negative fraction than LDA. The PSS, which depends only on the classification table (confusion matrix) is better for QDA. The 10 EOF version sorted by the sensitivity matrix has better BSS and PSS values for LDA and QDA than for the 10 version sorted by EOF variance explained. Also, the BSS and PSS for the 10 EOF version sorted by the sensitivity matrix are almost as good as or better than the 25 EOF version.

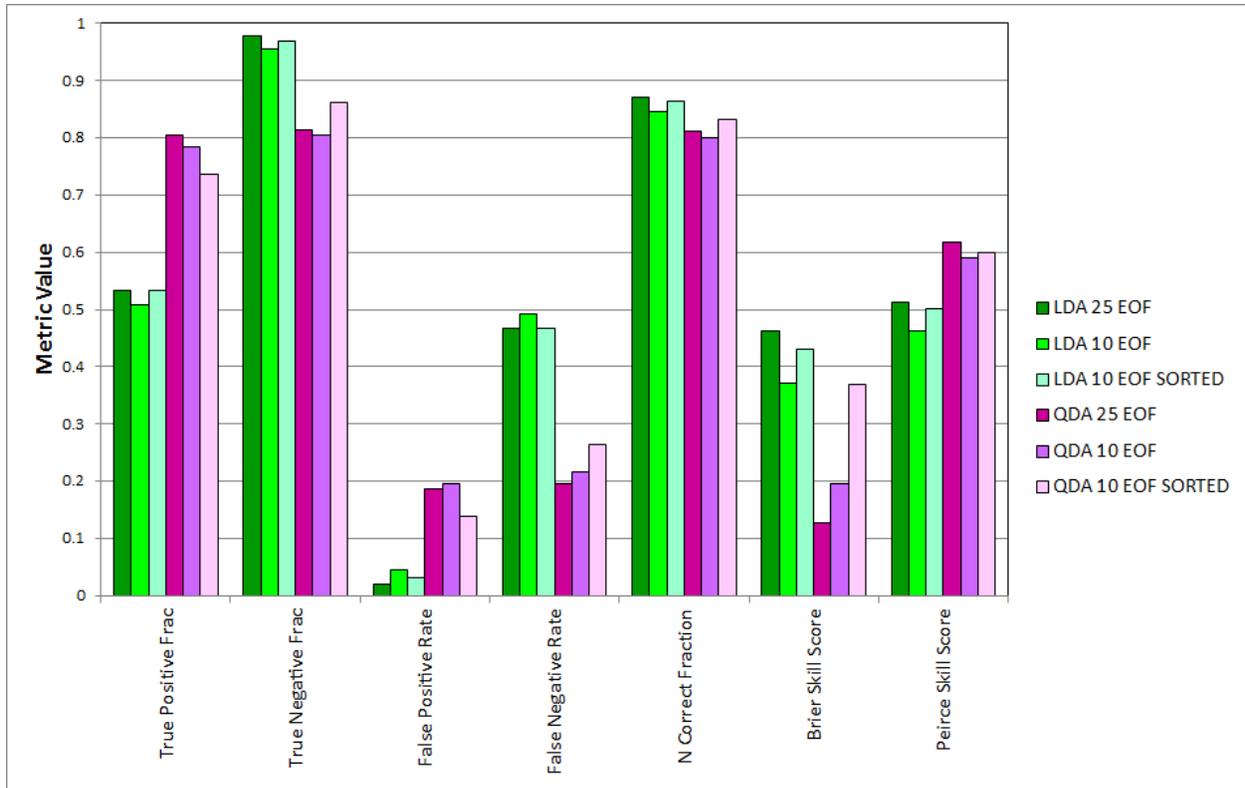


FIGURE 4.16. Comparison of satellite data algorithm runs using 25 EOF's sorted by variance explained, 10 EOF's sorted by variance explained, and 10 EOF's sorted by significance. All metrics are averages from 1000 runs using different shufflings of training and testing data.

Even though the satellite data based algorithm produced skillful classifications, it was unable to beat the baseline algorithm. Although the temperature values contained in the satellite images provided valuable information for separating the classes, the maximum wind speed value used in the baseline model appears to be more effective. For this reason, a hybrid algorithm was generated which combines the 10 sensitivity vector-sorted EOF version of the satellite model with the input from the baseline model.

4.4. COMBINED SATELLITE DATA AND ANCILLARY DATA ALGORITHM

The above results show that including the first 10 EOFs of the IR satellite data in order of the significance vector component magnitude provide results that are about as good as

the 25 EOF version. Figure 4.17 shows that the 10th sorted EOF sensitivity component is about 15% of the magnitude of the first one. Using this as a rough guide, the first four input variables from the baseline model were also included. From Figure 4.3 these include the maximum wind, latitude, and the two components of the storm motion vector.

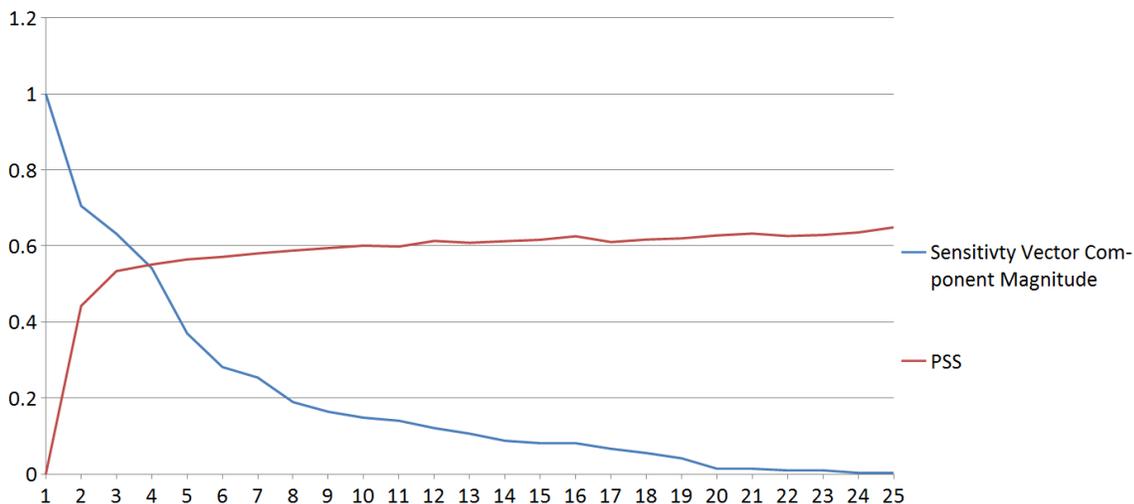


FIGURE 4.17. Performance of the satellite data based algorithm run with increasing number of EOF's where the order of included EOF's was determined by the magnitude of its significance.

The combined algorithm included 14 input variables, 10 from the IR data and four from the baseline model input. Again LDA and QDA were used. Two samples were used for the combined model. For evaluation of performance, 1000 partitions of testing and training sets were used, and the average verification statistics were calculated. For later model analysis, a version with the full sample was also run. This ensures that the full life cycle of each storm is included for the case studies that will be presented in Chapter 5.

Figure 4.18 shows the verification metrics for the combined model. Based on the BSS and PSS, all versions of the combined model are more skillful than their corresponding versions with the satellite data only and baseline models. The percent correct is near 90% for both LDA and QDA. Thus, combining the IR data with the baseline inputs is the best

approach for the eye classification algorithm. As expected, for both QDA and LDA the verification metrics are a little better when all the data are used together, compared to the use of training/testing partitions. However, the values are generally comparable, so the total sample version will be used in the further model analysis described below and in Chapter 5.

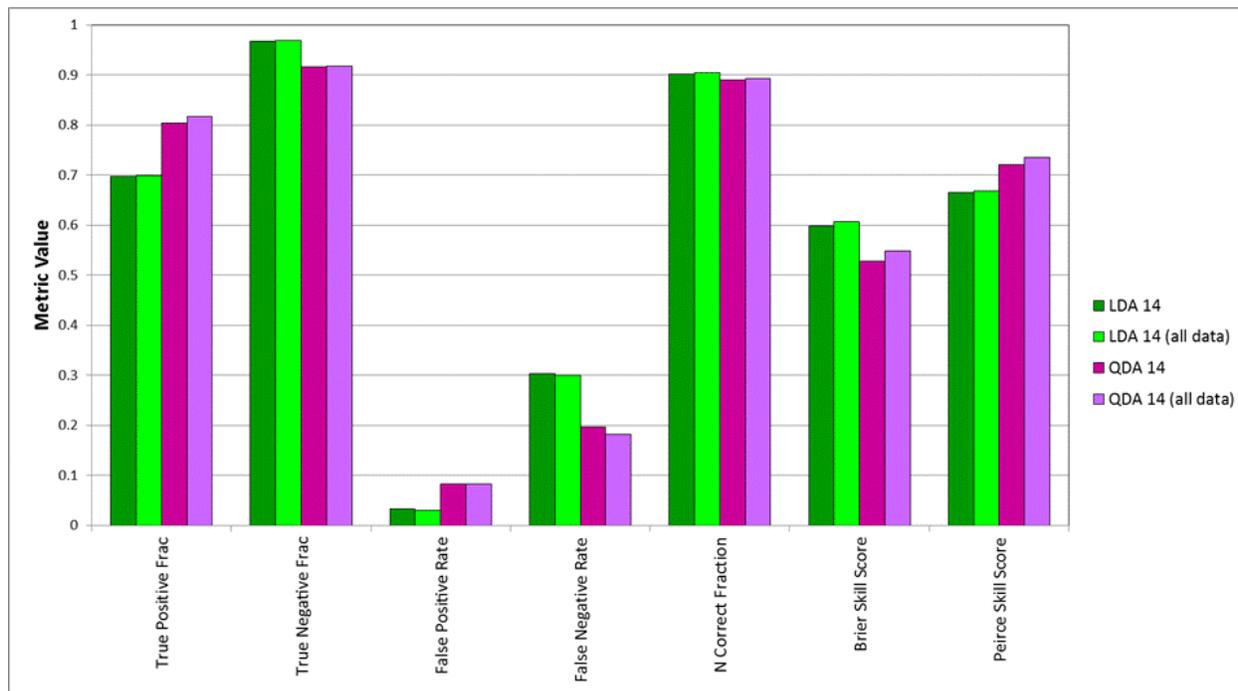


FIGURE 4.18. Comparison of combined satellite data and ancillary data algorithm runs. LDA and QDA were run using a randomly selected shuffling of the training and testing set as well as using the entire dataset for training and testing. All metrics are averages from 1000 runs using different shufflings of training and testing data.

Figure 4.19 shows the sensitivity for the combined model, sorted by the magnitude of the components. The maximum wind has the largest impact followed by EOFs 9, 4, 15 and 13. As illustrated in Figure 4.7, most of these EOFs have a symmetric structure. The components for the baseline variables are similar to those in Figure 4.3 for the baseline model by itself, although some have smaller magnitudes. This similarity suggests that these provide fairly independent information. The component for EOF01 is relatively small, even though it was the most important EOF when sorted by variance explained or by the sensitivity vector of

the satellite data model. It is likely that EOF01 was providing some of the same information as the maximum wind, so its importance dropped significantly when the maximum wind was included directly.

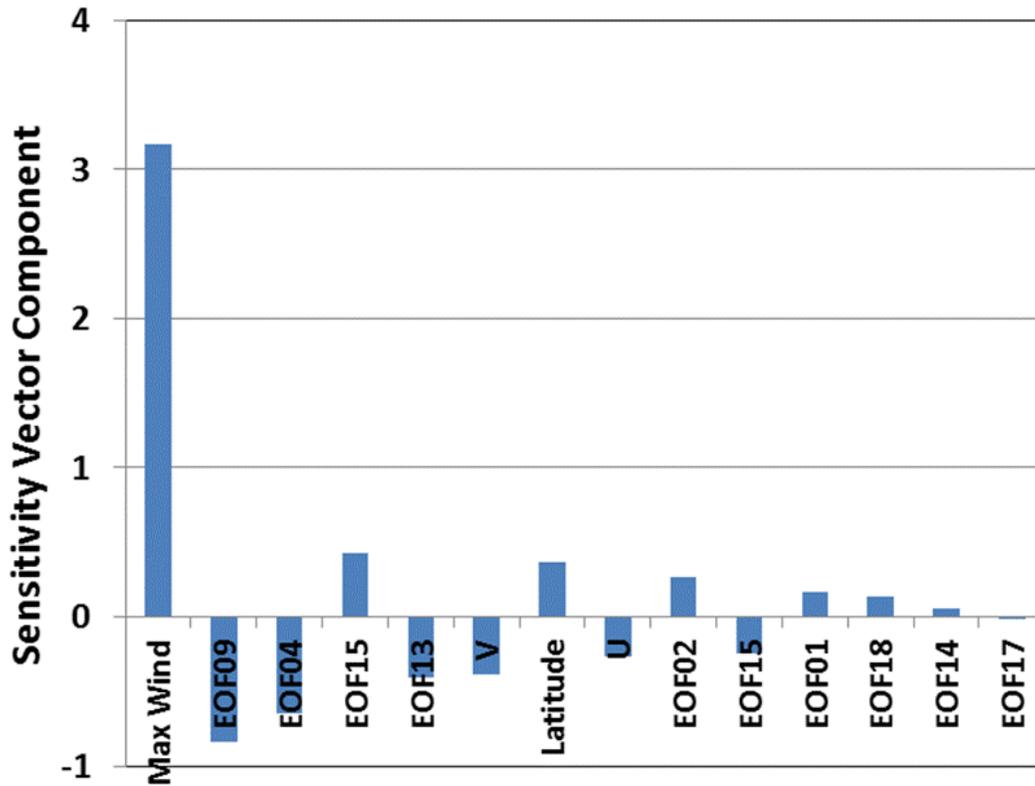


FIGURE 4.19. Sensitivity vector generated from the 14 predictor combined model.

To gain some insight into the differences between LDA and QDA, a simple two parameter version of the model was developed using the maximum wind and EOF 9 (the second most important input based on the sensitivity vector of the combined model). This version was developed with all of the input data and had verification statistics only a few percent worse than the 14 parameter model, depending on the metric. Examining Figure 4.20 it can be seen that QDA seems to divide the space along a curve which, in comparison to LDA, adds

a few false positives while adding several true positives. While this version only uses two predictors, based on the error statistics in Figure 4.18, it is likely that QDA and LDA are dividing space similarly using 14 predictors. Additionally, both the EOF and the maximum wind provide valuable information for separating the two classes. There are still many misclassified samples in the center of Figure 4.18. Given that the two predictor version of the model had slightly worse error statistics than the 14 predictor model, it appears that the 14 predictor model is able to use the extra dimensions of the data to separate some of the difficult to classify samples in the center of Figure 4.18.

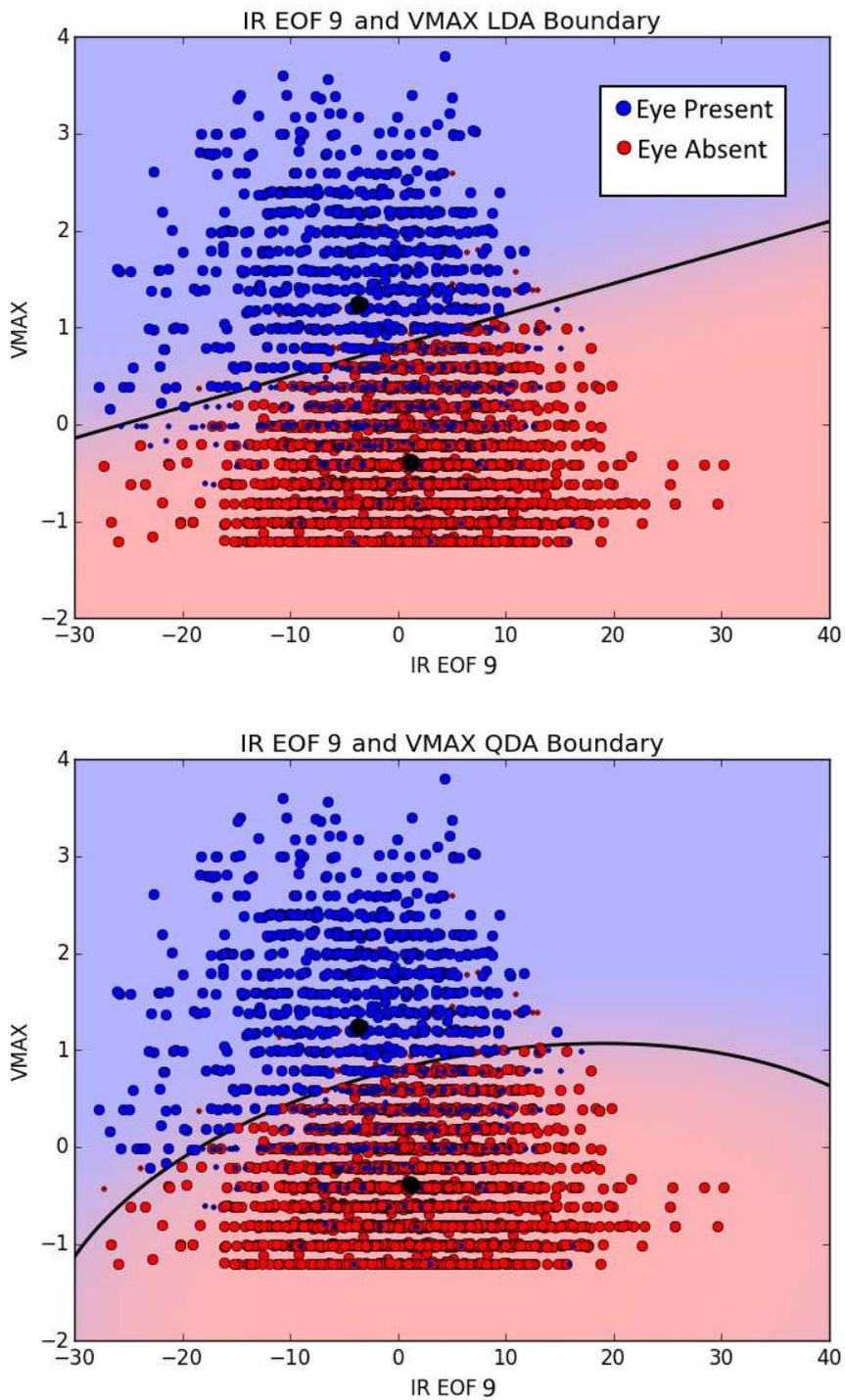


FIGURE 4.20. LDA and QDA boundaries using EOF 9 and VMAX. Generated using entire dataset for both training and testing.

CHAPTER 5

DISCUSSION

5.1. CASE STUDIES

The results in Chapter 4 showed that the 14 parameter combined model was able to correctly classify the eye/no-eye cases about 90% of the time. To get a feel for how this algorithm works on individual storms, the probability of an eye and the classification were determined for the full life cycles of the three test storms (Arlene, Danielle, and Katrina). Individual cases from these TCs where the algorithm incorrectly assigned the class will be examined in more detail to help provide suggestions for future improvements to the method.

Tropical Storm Arlene formed in the western Caribbean on June 8, 2005 and moved north into the Gulf of Mexico (www.nhc.noaa.gov). It never became a hurricane, but got pretty close when it was west of Tampa, FL with maximum winds of 60 kt. It hit land near the Florida-Alabama border late on June 11. According to the TAFB classifications, Arlene never had an eye. The LDA and QDA versions of the classification algorithm were run every 6 hours from 00 UTC on June 9 through 18 UTC on June 11. Both LDA and QDA correctly classified all 12 of these cases as no-eye. The maximum eye probability was only 2.7% from LDA and 0.1% from QDA on 06 UTC on June 11, when Arlene reached its peak intensity of 60 kt. Thus, the classification algorithm worked very well for Arlene.

Hurricane Danielle formed about 1000 km west of Africa and 1000 km north of the equator on August 22, 2010 (www.nhc.noaa.gov). It generally moved to the northwest for the next few days and became a hurricane at 18 UTC on August 23rd. Danielle reached a peak intensity of 115 kt at 18 UTC on August 27th when it was southeast of Bermuda. After that, Danielle turned towards the north and northeast, gradually weakening over the following few

days. Figure 5.1 shows the maximum wind, and LDA and QDA eye probabilities for the 32 Danielle cases. Also shown is the observed eye probability, which is 100% for eye cases and 0% for no-eye cases. Eleven of the 33 cases were observed to have an eye and 22 cases did not have an eye. For LDA and QDA, if the probability for a case is greater than 50% then the algorithm will classify the case as “eye-present”. This figure shows that all 22 of the observed no-eye cases were correctly classified by LDA. One no-eye case (number 9, which was at 5:45 UTC on August 24, 2010) was incorrectly classified by QDA. Case 18 (11:45 UTC on August 26th) was incorrectly classified as a no-eye case by LDA and QDA. Case 26 (17:45 UTC on August 28th) was incorrectly classified as a no-eye case by LDA. Thus, QDA correctly classified all but one no-eye and one eye case and LDA correctly classified all but two eye cases, which is about 94%. This is a little better than the sample average of 90%.

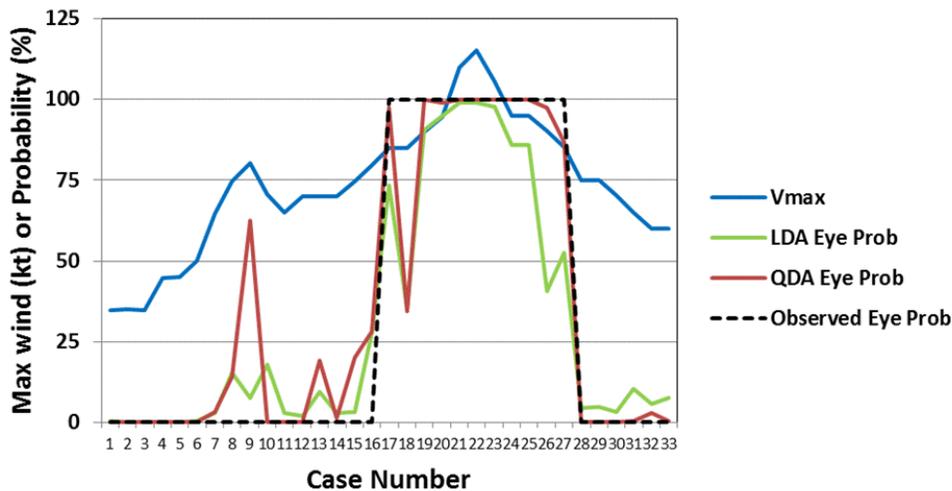


FIGURE 5.1. Algorithm probabilities/classifications vs truth for the lifetime of Hurricane Danielle.

Hurricane Katrina formed in the Central Bahamas on August 23, 2005 and struck South Florida a few days later as a category 1 hurricane (www.nhc.noaa.gov). It weakened a little as it crossed Florida but became a larger and powerful category 5 hurricane in the Central

Gulf of Mexico. It weakened to a category 3 storm before it made landfall near the Louisiana-Mississippi border on August 29th. Katrina was one of the top five deadliest hurricanes ever to hit the U.S. Figure 5.2 shows the maximum wind, and LDA and QDA probabilities for the 20 Katrina cases. This figure shows that QDA correctly classified all except for case 7, which was at 23:45 UTC on August 25th. That case was observed as any eye case, but was classified by QDA as a no-eye case. This case was just as Katrina was striking South Florida. LDA also mis-classified case 7 in addition to case 10, which was at 23:45 UTC on August 26th. However, the LDA probability was 44% so it did not miss this case by very much.

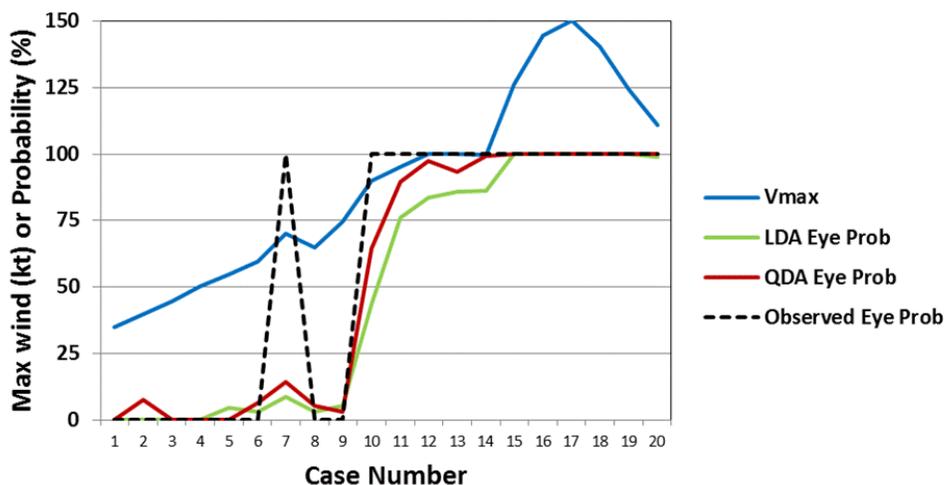


FIGURE 5.2. Algorithm probabilities/classifications vs truth for the lifetime of Hurricane Katrina.

Figure 5.3 shows the IR images for the three Danielle cases that were mis-classified by either LDA or QDA. Case 9 was mis-classified by QDA as an eye case. The QDA probability was only about 54%, so the algorithm was close to the boundary between the classes. The cold cloud area was very symmetric in this case, which may have contributed to it being classified as an eye case. Also, there is a small warm "wedge" to the northeast of the center, which may have contributed to the eye classification. Case 18 was mis-classified by both

LDA and QDA as a no-eye case. The IR image has a warm area near the center, but more of the area is to the northeast. It is possible that some improvement could be made to the algorithm if the IR data was used to re-center the image when there is a warm area near, but not exactly at, the image center. The cold area is also very asymmetric in this case, with more of the cold area to the south of the center. This may also have contributed to the mis-classification. Case 26 was mis-classified by LDA as a no-eye case. The warm area near the center is again shifted somewhat to the northeast, and the cold area is asymmetric, with more cold area to the south and west of the center. [16] showed that the thunderstorm activity near the center of tropical cyclones becomes asymmetric when they are surrounded by wind fields that vary in speed or direction with height. This is called environmental vertical wind shear, and causes the strongest thunderstorms (and coldest cloud tops in the IR imagery) to occur in the direction and to the left of the shear vector, defined as the upper level wind minus the lower level wind. The environmental vertical shear vector can be estimated in most cases, so the algorithm might be improved by including that information in the LDA or QDA.

Figure 5.4 shows the IR images for the two Katrina cases that were mis-classified by either LDA or QDA. Case 7 was mis-classified by both LDA and QDA as a no-eye case, and Case 10 was mis-classified by LDA as a no-eye case. Similar to the Danielle cases, the cold area is very asymmetric. Also, for case 10, there is a warm area but it is displaced to the north of the center of the image. Thus, re-centering the image and using the vertical shear vector to help account for asymmetries could be used to improve the algorithm.

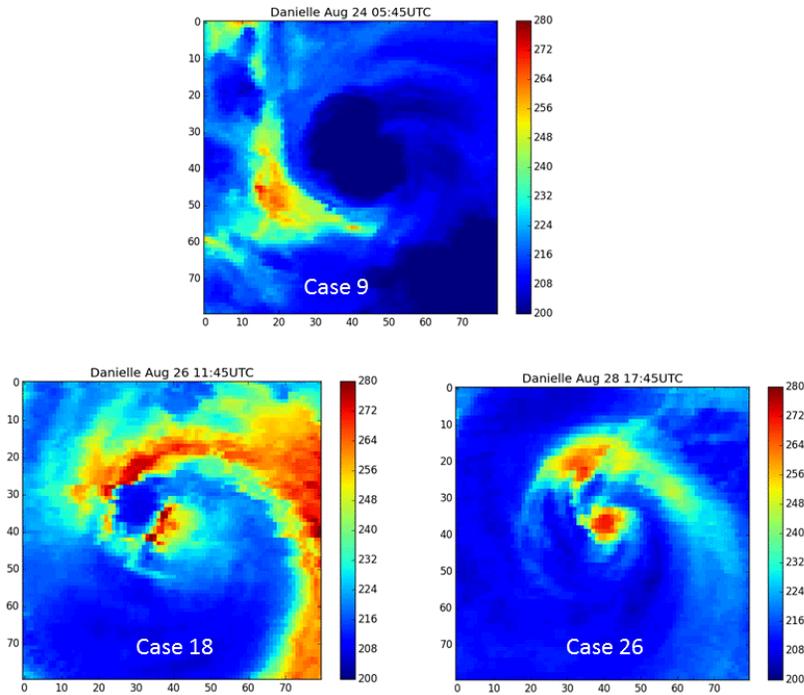


FIGURE 5.3. IR images from Hurricane Danielle used as input to algorithm. The algorithm incorrectly classified cases 18 and 26 as “eye-absent” cases while case 9 was mis-classified as “eye-present”.

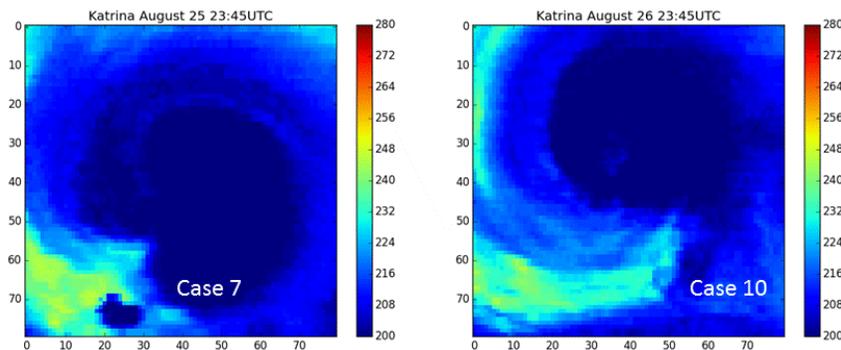


FIGURE 5.4. IR images from Hurricane Katrina used as input to algorithm. The algorithm incorrectly classified the images as “eye-absent” cases.

5.2. INDEPENDENT TESTING DATA

After development completed, it became clear that the training and testing datasets may not be completely independent. Since the entire set of available images was randomly shuffled and split into training and testing sets, it was likely that an image may appear in

the training set and a very similar image separated by 6 hours would appear in the testing set. This may bias the results of the algorithm towards improved performance. For this reason, the “combined” version of the model was trained using all of the available imagery from 1996-2012 and then tested on independent imagery from 2014. A total of 131 images from the 2014 hurricane season in the Atlantic basin were used for this test. The 2014 data was used for the independent test because the 2013 Atlantic hurricane season was extremely inactive, with almost no eye cases. Figure 5.5 illustrates the comparison between the tests using the 2014 data and the performance of the “combined” algorithm discussed in Section 4.4. It can be seen that the majority of the metrics were about the same between the 2014 tests and the 1996-2012 tests. The 1996-2012 runs of LDA saw slightly better performance over the 2014 runs. However, a clear bias improving the results can be seen for the QDA version of the algorithm. This bias improved the fraction of correct classifications by about 3%. The largest difference was the BSS which received improved performance from the bias by about 0.2. This performance still indicates that the algorithm is a useful tool with LDA appearing to be better overall. To get a more accurate measure of performance, further work should be performed to use an improved method of generating the training and testing sets. The imagery should be divided so that all the imagery for a particular storm will be placed exclusively in either the training or testing set.

5.3. ALTERNATE SENSITIVITY VECTOR

The sensitivity vector as defined in Section 3.1 multiplies the vector by the standard deviations of each predictor. This step was formulated to view how much each predictor within the sensitivity vector changes with a one standard deviation change. Strictly speaking, this is not an essential step and may obscure the information produced from examining the

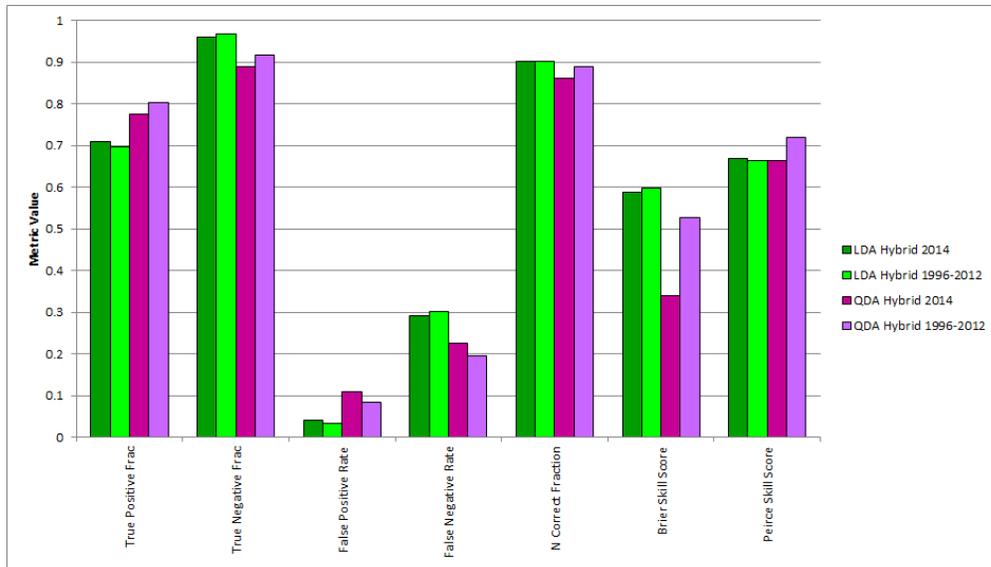


FIGURE 5.5. Comparison between average LDA/QDA performance among 1000 runs of the “combined” model using randomly shuffled training/testing sets and LDA/QDA trained using all available data and tested against 131 images from the 2014 hurricane season in the Atlantic basin.

difference between the discriminant functions. Using the sensitivity vector, a subset of 10 EOFs were selected from the set of 25 EOFs to produce the “combined” model. It is possible that an alternate set of EOFs may have been selected if the standard deviation term was excluded from the formulation of the sensitivity vector. For that reason, the 25 element sensitivity vector was generated without multiplying by the standard deviation. The comparison of these vectors can be seen in Figure 5.6. It can clearly be seen that the two different formulations produce different arrangements of the predictors. However, 6 of the first 10 elements appear in both vectors. This indicates that most of the same predictors would be included in the “combined” model regardless of which of the two sensitivity vectors was used. More work needs to be performed to examine how different selections of predictors impact the performance of the “combined” model.

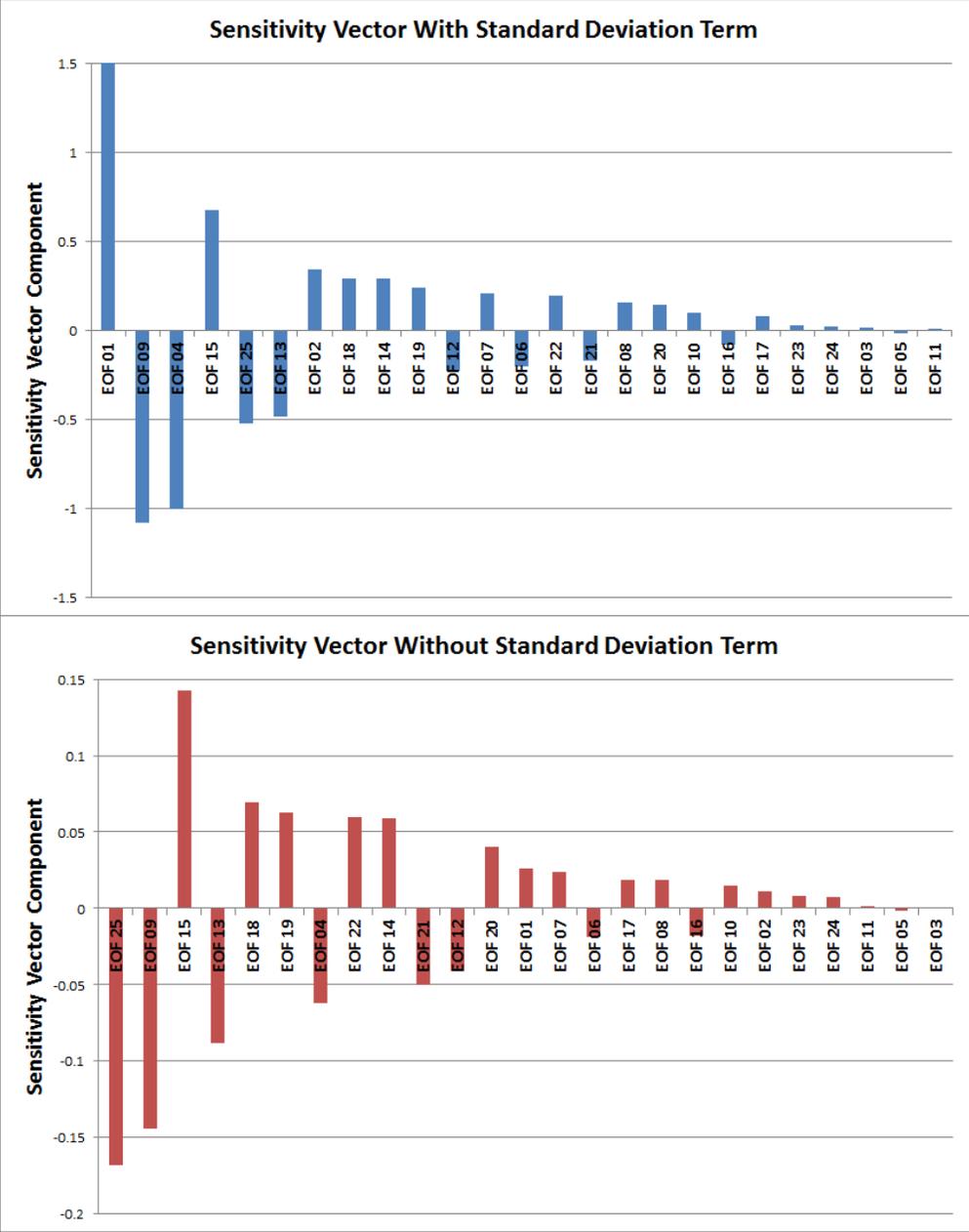


FIGURE 5.6. Comparison between the 25 element sensitivity vector generated using the standard deviation term and the vector generated without the standard deviation term. Elements are sorted by their magnitude within the sensitivity vector.

CHAPTER 6

CONCLUSIONS

Since eye formation is often associated with the rapid intensification of tropical cyclones, this information is very valuable to hurricane forecasters. Currently, eye analysis is performed subjectively by forecasters through analysis of satellite and aircraft data. As a result, much of the data available for prediction are unused. An automated procedure for detecting whether satellite imagery contains an eye would be a useful tool for forecasters. This project represents a novel effort to create an algorithm that can perform automated eye detection. Since infrared (IR) imagery from geostationary satellites is available virtually continuously, the algorithm was developed to use this imagery. Basic storm information that is also routinely available, including maximum wind speed, center latitude, center longitude, and storm motion was also used. 4109 IR images and their associated basic storm information were used for algorithm development. Each satellite image contains an Atlantic basin tropical cyclone with imagery spaced 6 hours apart. The selected images span from 1996 through 2012. Roughly 25% of the images were considered to have eyes as determined subjectively by the NHC Tropical Analysis and Forecast Branch (TAFB). The subjective estimates were used as the “truth” for algorithm development and evaluation.

The dimension of a single IR image around each tropical cyclone contained 6400 pixels, which is larger than the number of images available for development. Thus, Principal Component Analysis was employed to reduce the dimension of the satellite imagery. A total of 25 EOFs were sufficient to explain 95% of the variance of the IR dataset. LDA and QDA were used to classify each case into “eye-present” and “eye-absent” cases. Additionally LDA and QDA were used to estimate the probability that a case had an eye. The Peirce Skill

Score (PSS) was employed to evaluate the accuracy of the algorithm. The PSS is a combined measure of probability of detection and false positive rate. The PSS also measures skill above or below that of a “no skill model” to randomly assign classifications based on the distributions of the classes in the data. A variety of combinations of the basic storm parameters and satellite imagery were used as input to the algorithm in order to determine the best use of the information. Using only the basic storm information, LDA and QDA had a PSS of about 67%. Using only the IR satellite data, LDA and QDA had a PSS of about 60%. This indicates that the exclusive use of the IR images was not as useful as the exclusive use of basic storm information at performing classification. Using a combination of the 10 best IR and 4 best baseline predictors, LDA and QDA had a PSS of about 72%. This combined use of input data was able to correctly classify about 90% of the testing data. This result indicates that the algorithm would be very useful as objective guidance on eye formation for use by forecasters.

The performance of QDA was compared against LDA. QDA produced slightly better classifications than LDA. However when the Brier Skill Score, which evaluates the skill of probabilistic predictions, was compared, LDA performed better than QDA. A possible application of the eye detection algorithm is to provide input to statistical forecast methods for predicting rapid changes in storm intensity. A rapid intensification prediction scheme would likely find the probability of an image containing an eye more useful than a classification. For this purpose, the use of LDA is recommended over QDA.

After testing the algorithm on an independent dataset from 2014, it appears that the overall performance is slightly biased towards improved results. The Brier Skill Score for QDA saw the largest improvement due to bias. Further work will be performed to more accurately measure performance using improved generation of the training and testing sets.

Examination of the classification results for some individual storms suggested that the method might be further improved by re-centering the IR images in some cases, and using the vertical shear vector to account for asymmetries in the thunderstorms near the storm center. The vertical shear vector is also routinely available to forecasters from analysis of meteorological observations around tropical cyclones. These are topics for future research.

BIBLIOGRAPHY

- [1] R. Anthes, *Tropical Cyclones: Their Evolution, Structure and Effects*. Meteorological Monographs, American Meteorological Society, 1982.
- [2] J. Vigh, J. Knaff, and W. H. Schubert, “A climatology of hurricane eye formation,” *Monthly Weather Review*, vol. 140, pp. 1405–1426, 2012.
- [3] C. Velden, B. Harper, F. Wells, J. L. Beven II, R. M. Zehr, T. Olander, M. Mayfield, C. Guard, M. Lander, R. Edson, L. Avila, A. Burton, M. Turk, A. Kikuchi, A. Christian, P. Caroff, and P. McCrone, “The dvorak tropical cyclone intensity estimation technique,” *Bulletin of the American Meteorological Society*, pp. 1195–1210, 2006.
- [4] J. Kaplan, C. M. Rozoff, M. DeMaria, C. R. Sampson, J. P. Kossin, C. S. Velden, J. J. Cione, J. P. Dunion, J. A. Knaff, J. A. Zhang, J. F. Dostalek, J. D. Hawkins, T. F. Lee, and J. E. Solbrig, “Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models,” *Weather and Forecasting*, vol. in press.
- [5] S. Q. Kidder and T. H. Vonder Haar, *Satellite Meteorology an Introduction*. 525 B Street, Suite 1900, San Diego, California 92101-4495, USA: Academic Press, 1995.
- [6] A. B. Schumacher, M. DeMaria, and J. A. Knaff, “Objective estimation of the 24-h probability of tropical cyclone formation,” *Weather and Forecasting*, vol. 24, pp. 456–471, April 2009.
- [7] J. Knaff, S. Longmore, and D. Molenaar, “An objective satellite-based tropical cyclone size climatology,” *Journal of Climate*, vol. 27, pp. 455–476, 2014.
- [8] C. Landsea, D. Glenn, W. Bredemeyer, M. Chenoweth, R. Ellis, J. Gamache, L. Huffstetler, C. Mock, R. Perez, R. Prieto, J. Sanchez-Sesma, D. Thomas, and L. Woolcock,

- “A reanalysis of the 1911-20 atlantic hurricane database,” *Journal of Climate*, vol. 21, pp. 2138–2168, 2008.
- [9] E. N. Rappaport, J. L. Franklin, L. A. Avila, S. R. Baig, J. L. Beven II, E. S. Blake, C. A. Burr, J.-G. Jiing, C. A. Juckins, R. D. Knabb, C. W. Landsea, M. Mainelli, M. Mayfield, C. J. Mcadie, R. j. Pasch, C. Sisko, S. R. Stewart, and A. N. Tribble, “Advances and challenges at the national hurricane center,” *Weather and Forecasting*, vol. 24, p. 398, 2009.
- [10] T. Zito, N. Wilbert, L. Wiskott, and P. Berkes, “Modular toolkit for data processing (mdp): a python data processing framework,” *Frontiers in Neuroinformatics*, vol. 2, no. 8, 2009.
- [11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [12] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences Second Edition*. 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA: Elsevier Inc., 2006.
- [13] S. Aberson, “Five-day tropical cyclone track forecasts in the north atlantic basin,” *Weather and Forecasting*, vol. 13, pp. 1005–1015, 1998.
- [14] B. Jarvinen and C. Neumann, “Statistical forecasts of tropical cyclone intensity,” tech. rep., NOAA/NWS/NHC, 1998.
- [15] J. Cangialosi and J. Franklin, “National hurricane center forecast verification report,” tech. rep., NOAA/NWS/NHC, 2015.

- [16] K. L. Corbosiero and J. M. Olinari, “The relationship between storm motion, vertical wind shear, and convective asymmetries in tropical cyclones,” *Journal of the Atmospheric Sciences*, vol. 60, pp. 366–376, 2003.