A NEW ANALYSIS OF THE GÁLVEZ-DAVISON INDEX FOR CONVECTIVE FORECASTS IN NORTHERN AFRICA

THESIS

Gabriel D. Donndelinger, Captain, USAF

AFIT-ENP-MS-18-M-078

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENP-MS-18-M-078

A NEW ANALYSIS OF THE GÁLVEZ-DAVISON INDEX FOR CONVECTIVE FORECASTS IN NORTHERN AFRICA

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Atmospheric Science

Gabriel D. Donndelinger, BS

Captain, USAF

March 2018

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENP-MS-18-M-078

A NEW ANALYSIS OF THE GÁLVEZ-DAVISON INDEX FOR CONVECTIVE FORECASTS IN NORTHERN AFRICA

Gabriel D. Donndelinger, BS

Captain, USAF

Committee Membership:

Maj H. R. Tseng, PhD
Chair

Maj O. A. Nava, PhD
Member

Dr. R. A. Stenger
Member

AFIT-ENP-MS-18-M-078

Abstract

Severe wind gusts and thunderstorms have been difficult to forecast in Africa. Traditional convective forecast tools (e.g., Total Totals Index, Lifted Index, K Index (KI) and Convective Available Potential Energy) do not accurately portray potential for thunderstorms in Africa. To increase forecast accuracy for thunderstorms in northern Africa, this research effort used the Gálvez-Davison Index (GDI), a convective index created for the tropics, and assessed its applicability to northern Africa. GDI was produced for the Caribbean and Central America, and utilizes temperature, moisture, mid-level stability, dry air entrainment and an elevation factor to calculate convective potential. As such, these characteristics make GDI especially useful for forecasting thunderstorms in the tropics. In this research, GDI and KI were calculated using National Center for Environmental Prediction (NCEP) Global Forecast System (GFS) reanalysis data. K-means clustering was used to conduct an error analysis on both indices and the resulting location and area error values. These error values were then bootstrapped and confidence intervals were calculated using the bias-corrected and accelerated method.

Results indicate GDI and KI have similar location error in both the intra-annual and intra-seasonal studies. In comparison with KI, GDI had lower area error values in the intra-annual study and in most convective synoptic cases with 95% confidence.

Acknowledgments

This thesis was by no means the work of one individual, but an effort of many who I offer my heartfelt thanks to:

Major Rose Tseng (USAF) for taking me on as her thesis student, even when a lot of interaction was through distance correspondence. She led me, encouraged me and cared about my well-being through this journey. I could not have done it without her.

Major Omar Nava (USAF) for fostering me as his thesis student and for helping me along the way with various aspects of the project. He offered ideas on possible modifications to make to the index codes and guided me through the statistical methods.

Dr. Robert Stenger (USAF Ret.) for mentoring and instructing me on tropical meteorology, serving on my committee and offering ideas on cluster analysis.

Dr. Jose Gálvez and Mr. Michel Davison (NOAA) for being very supportive of this project, not only allowing their research to be worked on, but also for being available along the way for correspondence.

14th Weather Squadron (USAF) for providing the lightning data in a form that was usable for this project in a timely manner.

Naval Research Laboratory (NRL) for providing the satellite data.

University Corporation for Atmospheric Research (UCAR) Research Data Archive (RDA) for granting access to their GFS reanalysis data.

Contact at the 21st Operational Weather Squadron (USAF) for keeping in correspondence on the nature of the research question

My classmates for being on this journey together, through thick and thin. I can honestly say I would not have made it through this program without them.

Last but certainly not least, thank you to my family: To **my parents** for their love and support, for teaching me the value of hard work and faith in God through the craziness of life. To **my siblings**, for keeping me sane and making me laugh. To **my nieces and nephews** for continually reminding me the value of the little things and letting me act like a kid again.

In honor of my **grandparents**: This is a part of your legacy.

Table of Contents

Abstract iv

Table of Contents vi

I. Introduction 1

- **General Issue** 1
- **Problem Statement** 2
- **Hypotheses** 3
- **Research Objectives, Focus and Questions** 3
- **Assumptions/Limitations** 4
- **Implications** 6
- **Preview** 6

II. Background on African Thunderstorms and Literature Review 7

- **Chapter Overview** 7
- **African Thunderstorms** 25
- **Relevant Research** 12
- **The K Index** 12
- **The Gálvez-Davison Index** 14

III. Methodology 21

- **Chapter Overview** 21
- **NCEP GFS Reanalysis Data and GDI Calculation** 21
- **K Index (KI) Calculation** 25
- **Plotting ATDNET Lightning Data** 26
- **NRL IR Satellite Images** 27

Methods for Comparing Index Forecasts 28

Summary 37

IV. Analysis and Results 39

Chapter Overview 39

Intra-Annual Study 39

Intra-Seasonal (An In-Depth Late Summer/Early Fall Study) 42

Potential GDI-As Study 48

V. Conclusions and Recommendations 58

Chapter Overview 58

Conclusions of Research 58

Significance of Research 60

Recommendations for Action 61

Recommendations for Future Research 61

Summary 63

Appendix A: Intra-Annual Error Values 64

Appendix B: Intra-Seasonal Error Values 67

Appendix C: Potential GDI-A Error Values 71

Bibliography 75

List of Figures

Figure 1.1: The blue rectangle outlines the region of interest, 0°-20°N latitude and 20°W-50°E longitude. 2

Figure 2.1: The blue arrow highlights a portion of the AEJ at 650 mb. 9

Figure 2.2: The yellow rectangle highlights airmass thunderstorms on an infrared satellite image from 27 September 2017 at 1500Z (NRL 2017). 10

Figure 2.3: Yellow rectangles highlight MCSs on an infrared satellite image from 25 August 2017 at 0000Z (NRL 2017).....11

Figure 2.4: The locations of the prominent wind features, the African easterly jet (AEJ) at approximately 650 mb and the tropical easterly jet (TEJ) at approximately 175 mb. 12

Figure 2.5: A model depiction of the layers used in the GDI algorithm to predict convective potential (Gálvez and Davison 2016). 14

Figure 2.6: Convective potential based on the GDI value (Gálvez and Davison 2016). 19

Figure 3.1: A comparison of NOAA's model GDI (top) and the reanalysis GDI (bottom) from 01 August 2017 at 00Z. 24

Figure 3.2: KI mapped out over Africa for 15 August 2016 at 06Z. 25

Figure 3.3: GDI values (colored contours) at 15 August 2016 at 06Z, with Sferics lighting data (cyan asterisks) from 15 August 2016 from 05-07Z. 26

Figure 3.4: A color IR satellite image from 06Z on 15 August 2016 with cloud top temperatures (°C) indicated by the color filter (NRL 2017). 27

Figure 3.5: An example K-pick plot, full view from 1-10 clusters (left) and a zoomed-in view, 2-10 clusters (right). 31

Figure 3.6: GDI values 35 and above (red dots) and lightning data (cyan asterisks) on 15 August 2016 at 06Z. 32

Figure 3.7: KI values 30 and above (red dots) and lightning data (cyan asterisks) on 15 August 2016 at 06Z. 32

Figure 3.8: Lightning data (left) and GDI at or greater than 35 (right) divided into clusters from 06Z on 15 August 2016. 33

Figure 3.9: An example confidence interval plot with the mean (circle) and 95% confidence interval (error bars) for the GDI location error from 19-21 August 2017. 37

Figure 4.1: Intra-annual location error values of GDI (blue) and KI (red) forecasts. Means are indicated with circles and the 95% confidence intervals are indicated with error bars. 40

Figure 4.2: Intra-annual location error values of GDI (blue) and KI (red) forecasts. Means are indicated with circles and the 95% confidence intervals are indicated with error bars. 41

Figure 4.3: IR satellite image from 17 September 2017 at 00Z. 42

Figure 4.4: IR satellite image from 16 August 2016 at 18Z. 43

Figure 4.5: IR satellite image from 20 August 2017 at 00Z. 43

Figure 4.6: IR satellite image from 25 August 2017 at 00Z. 43

Figure 4.7: IR satellite image from 27 September 2017 at 15Z. 44

Figure 4.8: Intra-seasonal study location error grouped by days with predominant convective synoptic situations indicated, MCS for Mesoscale Convective Systems and AT for airmass thunderstorms.....45

Figure 4.9: Intra-seasonal study area error grouped by days with predominant convective synoptic situations indicated, MCS for Mesoscale Convective Systems and AT for airmass thunderstorms.....46

Figure 4.10: Airmass thunderstorms area error values with 90% confidence intervals. 48

Figure 4.11: GDI-A_{aveVV} from 15 August 2016 at 06Z with lightning (cyan asterisks). 50

Figure 4.12: GDI-A_{RH850} from 15 August 2016 at 06Z with lightning (cyan asterisks). 51

Figure 4.13: GDI-A_{RH700} from 15 August 2016 at 06Z with lightning (cyan asterisks). 51

Figure 4.14: GDI-A_{RH300} from 15 August 2016 at 06Z with lightning (cyan asterisks). 51

Figure 4.15: GDI-$A_{EPTP900}$ from 15 August 2016 at 06Z with lightning (cyan asterisks). 52

Figure 4.16: GDI-$A_{EPTP850}$ from 15 August 2016 at 06Z with lightning (cyan asterisks). 53

Figure 4.17: GDI-$A_{EPTP800}$ from 15 August 2016 at 06Z with lightning (cyan asterisks). 53

Figure 4.18: GDI-$A_{EPTP700}$ from 15 August 2016 at 06Z with lightning (cyan asterisks). 53

Figure 4.19: GDI-$A_{EPTP600}$ from 15 August 2016 at 06Z with lightning (cyan asterisks). 54

Figure 4.20: Location error values from GDI, KI and GDI-As for August 2016. 55

Figure 4.21: Area error values from GDI, KI and GDI-As for August 2016. 56

List of Tables

Table 2.1: K values and their respective thunderstorm frequency estimates (George 1960). 13

Table 4.1: GDI and KI location error ranges by date range. 44

Table 4.2: GDI and KI area error ranges by date range. 46

A NEW ANALYSIS OF THE GÁLVEZ-DAVISON INDEX FOR CONVECTIVE FORECASTS IN NORTHERN AFRICA

I. Introduction

General Issue

Forecasting convective weather on the African continent remains one of the most difficult challenges in meteorology. Due to numerous limiting factors, forecasters cannot accurately predict the weather to the degree of specificity that can be achieved in other areas of the world. Africa is a sparse region when it comes to quality environmental data, which creates many challenges for weather forecasting. Additionally, the dynamics of some weather phenomena remains unclear. For these and other reasons, forecasting weather in Africa remains a formidable task. One major challenge when forecasting environmental conditions in Africa is predicting convective storms with high winds and precipitation.

Forecasters use convective indices to aid in predicting thunderstorms. These indices utilize various parameters from real-time or modeled atmospheric soundings to gauge the probability of convective weather occurring. Common convective indices include: Lifted Index (LI), Showalter Stability Index (SSI) and the Total Totals Index (TTI). The K index (KI), which will be investigated further in Chapter II, is regarded as a quality index for tropical regions (Gálvez and Davison 2016). Unlike North America where LI, SSI and TTI are widely accepted as forecast standards to predict thunderstorms, currently no such standard exists for Africa. Common indices can be used over northern Africa, but not with the same level of confidence.

The area of interest for this research is northern Africa, specifically 0°-20°N latitude and 20°W-50°E longitude (Figure 1.1).

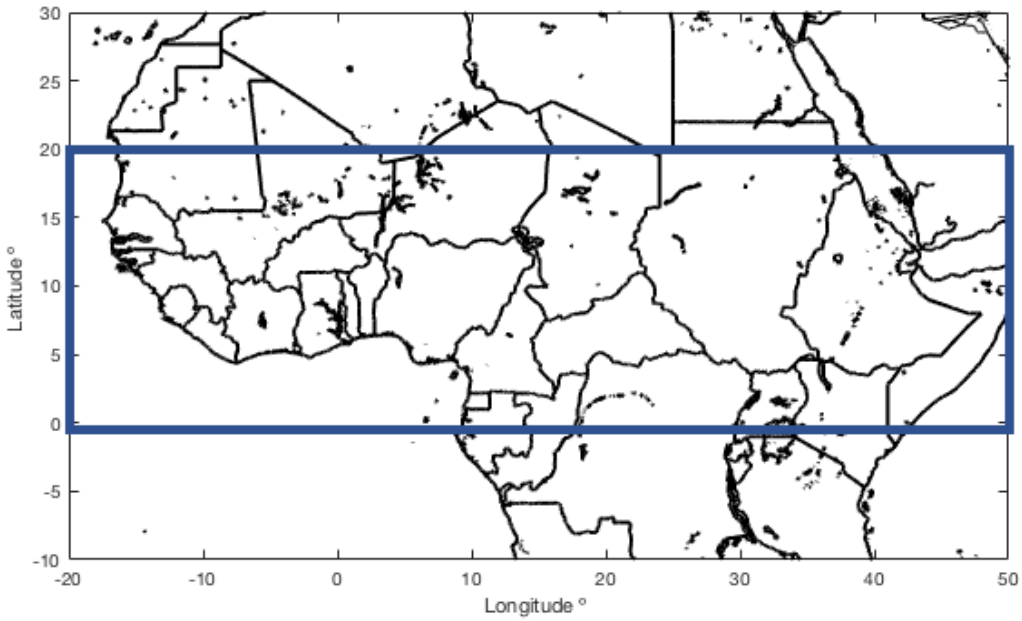

Figure 1.1: The blue rectangle outlines the region of interest, 0°-20°N latitude and 20°W-50°E longitude.

National Oceanic and Atmospheric Administration (NOAA) researchers, Gálvez and Davison (2016), have recently developed a new convective index tailored for the Caribbean and Central America. Validation studies have been conducted and variations have been made to tailor their Gálvez-Davison Index (GDI) for Costa Rica, South America and even South Korea (Omar Nava, Air Force Institute of Technology, written communication 13 July 2017). The purpose of this research is to assess and compare the forecasting skill of both GDI and KI for northern Africa.

Problem Statement

Current forecasting techniques for predicting convection in northern Africa are based on indices created with meteorological understanding and weather data from other

parts of the world. Thus far, there is no solid convective index tailored specifically for northern Africa. Consequently, forecasters cannot place sufficient confidence in predicting high winds, lightning or precipitation associated with convective activity.

Hypotheses

GDI will portray convection over northern Africa more accurately than KI. This is due to the nature of the GDI proven to work well for tropical regions, as it considers parameters important to tropical convection (Gálvez and Davison 2016). A new index should consider the parameters in GDI with some adjustments and possibly additional terms to account for the differences between Central America and the Caribbean and northern Africa.

Research Objectives, Focus and Questions

The research objectives are as follows:

1. Replicate GDI algorithm in Matlab, mapping out the index over northern Africa exactly as displayed on the NOAA website (http://www.wpc.ncep.noaa.gov/international/gdi/)
2. Replicate KI in Matlab, mapping out the index over northern Africa
3. Assess the skill of both indices for Africa by comparing forecasts with lightning data and satellite imagery, statistically analyzing their skill in predicting convection over northern Africa
4. Modify GDI into new GDIs for Africa (GDI-As) considering the differences in the regions they are targeting: the Caribbean and northern Africa

respectively; adjust the parameters within GDI and/or add new terms to modify the into GDI-As

5. Assess the skill of the GDI-As for Africa by comparing forecasts with lightning data and satellite imagery, statistically analyzing their skill in predicting convection over northern Africa

Research questions for investigation:

- How well does GDI predict convection over northern Africa, spatially, temporally and intensity-wise?
- What makes GDI work well for tropical regions? What makes it especially accurate over Central America and the Caribbean?
- Does GDI need improvement for forecasting over northern Africa, and if so, how? Are the same parameters in GDI applicable to GDI-A? What additional parameters, if any, need to be considered for northern Africa?
- How well does the new GDI-A work over northern Africa, spatially, temporally and intensity-wise? Why does it work better than GDI in this region?
- Does GDI-A improve confidence when forecasting convection over northern Africa?

Assumptions/Limitations

The data for this research is the Global Forecast System (GFS) reanalysis model data, as further explained in Chapter III. Although this data is not perfectly representative

of the true atmospheric conditions, it is the most realistic depiction of the atmosphere (UCAR 2014). Limitations of this data are the vertical and horizontal resolutions of the model. In the vertical, model data points are set at the surface, 1000 mb, 975 mb, 950 mb, 925 mb and 900 mb, and then every 50 mb above that until 100 mb (UCAR 2017). Increased emphasis is put on the lower levels, with several more layers represented than in the upper levels. Horizontal resolution is 1° by 1° of latitude and longitude, where 1° is approximately 111 km or 69 miles in northern Africa (UCAR 2017). Convection processes occur on smaller scales than the vertical and horizontal resolutions of the model data, as further explained in Chapter II.

Furthermore, since the forecast model data was mapped onto a 1° by 1° grid, each point is assessed to see if the forecast correctly identified lightning to occur. Two problems arise during this process. Of note, lightning rarely strikes at whole latitude and longitude degree values; instead the lightning strikes are scattered at various locations with fractional degree values of both latitude and longitude. This is problematic when comparing index values with truth, attempting to match model output data at whole degree values with sporadic lightning strikes. Moreover, interpretation of forecast index values indicating various levels of convective potential was subjective. Because of these two issues, a clustering method is used, not point-by-point analysis.

Another limitation is the independence of error in the samples. Model error carries over from one six-hour run to the next. In the field of meteorology, 24 hours is often the effective time between independent samples (Miller 1962). However, in this study, the samples are six hours apart. This is to maximize the use of available GFS reanalysis data and lightning observations.

Implications

With a new index for predicting convection over northern Africa, the confidence and specificity of forecasts in the region could increase greatly. Improving forecasting abilities will aid users in overcoming the challenges of predicting the extent, timing and intensity of convection over northern Africa. Increasing environmental situational awareness could contribute to a better understanding of the overall weather patterns and severe weather phenomenon in Africa and ultimately promote further knowledge of its role in the Earth's climate system and agriculture.

Preview

The organization of this thesis is as follows: Chapter II describes the past research on tropical convection and a background on African thunderstorms; Chapter III describes the background on data and methodology for this research; Chapter IV presents and analyses the results; Chapter V discusses how the results of this research impact the usefulness and influence of GDI and GDI-A when forecasting convection in Africa and concludes this research by presenting recommendations for future work in this field.

II. Background on African Thunderstorms and Literature Review

Chapter Overview

The purpose of this chapter is to lay out the foundation of current research and background on convection patterns over northern Africa. This information is critical to fully understanding the research findings and conclusions in later chapters.

African Thunderstorms

Research on thunderstorm activity focuses on convective initiation and life-cycle forcing in the mid-latitudes. Common convective indices from Chapter I forecast for mid-latitude convection because of the research focus in meteorology on these regions due to data availability, social and economic welfare. Similar research on convection in tropical regions could add to the knowledge base of global weather patterns and climatology.

Tropical convection varies in many ways from mid-latitude convection. First, latent heat release initiates and fuels convection in the tropics and available potential energy from strong temperature gradients drives convection in the mid-latitudes (Holton and Hakim 2013). Since most of the latent heat release in the tropics is from convective systems, thunderstorms upstream are an indicator of increased convective potential (Gálvez and Davison 2016). Overall, frontal movements, boundaries between airmasses, resulting from strong temperature gradients primarily cause mid-latitude convection, while large-scale circulations and latent heat release drive convection in the tropics. These large-scale circulations include the Hadley cell, the Intertropical Convergence Zone (ITCZ) and the Walker circulation.

Uneven heating on the earth's surface gives rise to the large-scale circulations in the atmosphere. The dominant circulation in tropical regions is the Hadley cell where trade winds in both hemispheres converge near the equator, causing air to rise (Holton and Hakim 2013). The Hadley cell is a main driver of heat transport from the equator poleward. Rising air transports heat from the surface to upper atmosphere through pseudoadiabatic ascent and formation of cumulonimbus clouds (Holton and Hakim 2013). These clouds form a discontinuous band of deep convection circling the globe along the meteorological equator called the Intertropical Convergence Zone or ITCZ (Galvin 2016). The location of the ITCZ moves north and south following the most direct solar radiation on the earth's surface. Vapor supplies required for this large scale convective circulation come from converging trade wind flow, providing the latent heat, and energizing persistent convection in the ITCZ (Holton and Hakim 2013).

Persistent convection in the ITCZ promotes another large-scale circulation in the atmosphere—the Walker circulation (Holton and Hakim 2013). This feature is the zonal movement of air along the equatorial region of the globe. This circulation is important to discussions and research on El Niño and Southern Oscillation in regards to disruptions in the prevailing patterns.

Tropical waves are features that promote convection in the tropics. When air rises in columns of convective clouds in the ITCZ upper level divergence occurs and, by mass continuity, low level convergence also occurs, forming tropical waves (Holton and Hakim 2013). Latent heat release from convective precipitation propels these weak disturbances in the ITCZ westward. (Holton and Hakim 2013). Perturbations in the easterly trade winds or 24-hour surface pressure changes identify the location of tropical

waves (Kirshnamurti et al. 2013). Waves that move off the African continent and into the Atlantic Ocean are important phenomena in the genesis of tropical cyclones (Kirshnamurti et al. 2013).

There are three main forms convection takes over northern Africa: 1) African easterly waves, 2) airmass thunderstorms and 3) mesoscale convective systems. Over the continent of Africa, unique processes occur in the atmosphere that generate specific types of tropical waves, African easterly waves (AEWs). During the summer in the Northern Hemisphere, intense surface heating in the Sahara Desert induces a strong positive temperature gradient between the equator and 25°N (Holton and Hakim 2013). This strong temperature gradient causes the African easterly jet (AEJ) to form at approximately 13-16°N with its core at 650 mb (Holton and Hakim 2013). Monsoonal flow and the lower Walker circulation induce westerly flow at approximately 10°N with its core at 950 mb. This set-up creates a cyclonic shear zone that promotes the initiation and propagation of synoptic-scale tropical waves (Holton and Hakim 2013). These AEWs depend more on the barotropic and baroclinic conversions of energy from the AEJ as opposed to latent heat release (Holton and Hakim 2013). High winds at 650 mb indicate the location of the AEJ (Figure 2.1).

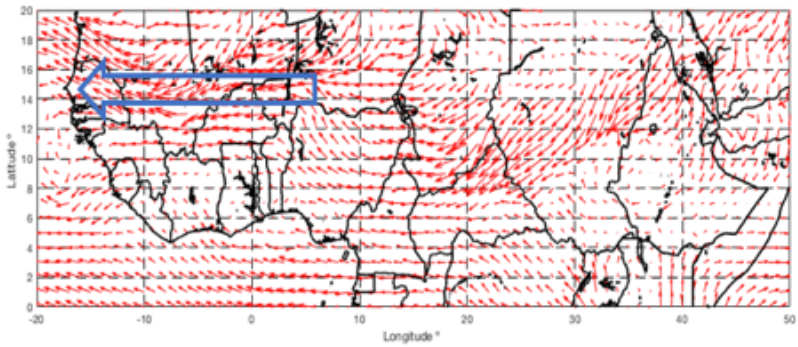

Figure 2.1: The blue arrow highlights a portion of the AEJ at 650 mb.

AEWs have distinct characteristics. They can range from 1,500-4,500 km, averaging around 2,500 km in length from north to south (Kirshnamurti et al. 2013). Lasting for 3-5 days, they travel at approximately 8 ms^{-1} or 5-7° longitude per day. Originating somewhere between 15-30°E, they reach their maximum amplitude anywhere from 10°E -20°W longitude, over West Africa or the coast. Ahead of the waves are northeasterly winds, low level divergence and sinking air while behind them are southeasterly winds, low level convergence and rising air (Kirshnamurti et al. 2013). Since this region is an easterly shear environment, convection is on the west side of the AEW axis. Although AEWs can generate convection, this is the least prominent form convection takes over northern Africa.

Convection can often be in the form of airmass thunderstorms in both the tropics and mid-latitudes. These are small columns of air that rise once daytime heating warms the surface past the convective temperature. Without the need for any mechanical forcing, air rises, moisture condenses and small, localized thunderstorms form. These systems range in size from a radius of 1° of latitude, near the equator approximately 111 km, and smaller. Most of these systems are smaller than 1° by 1° (Figure 2.2).

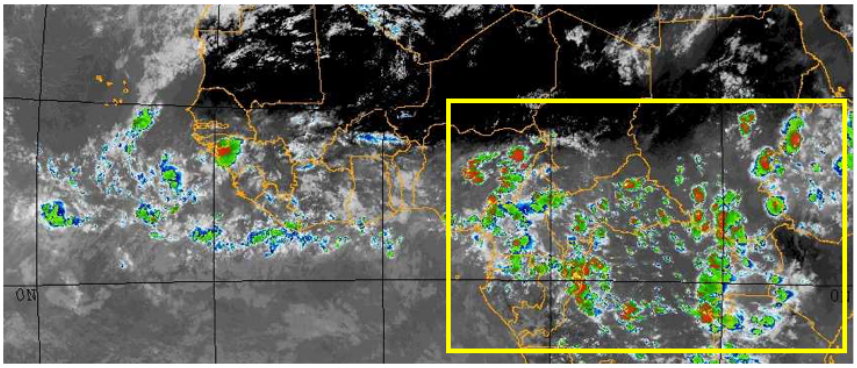

Figure 2.2: The yellow rectangle highlights airmass thunderstorms on an infrared satellite image from 27 September 2017 at 1500Z (NRL 2017).

Another source of convection in Africa is mesoscale convective systems (MCSs). MCSs are larger groupings of convective activity on a broader scale than independent smaller convective storms. MCSs are a cloud region at -52°C or colder covering a minimum of 30,000 km^2 (Jirak et al. 2003). They have strong vertical velocities, large amounts of precipitation and large areas of cold cloud tops (Figure 2.3; Kirshnamurti et al. 2013). Over northern Africa there is a prime environment especially conducive to the formation and maintenance of these MCSs with the tropical easterly jet (TEJ), at approximately 7°N and 175mb, and the AEJ, at approximately 13-16°N and 650mb (Figure 2.4). The anticyclonic shear side of the TEJ in the upper levels overlays the cyclonic shear side of the AEJ in the mid-levels, inducing convergence in the low levels and divergence in the upper levels. This environmental set-up is favorable for the development and maintenance of convection. Nearly all squall line systems in West Africa have been observed in this ideal environment between the TEJ and AEJ.

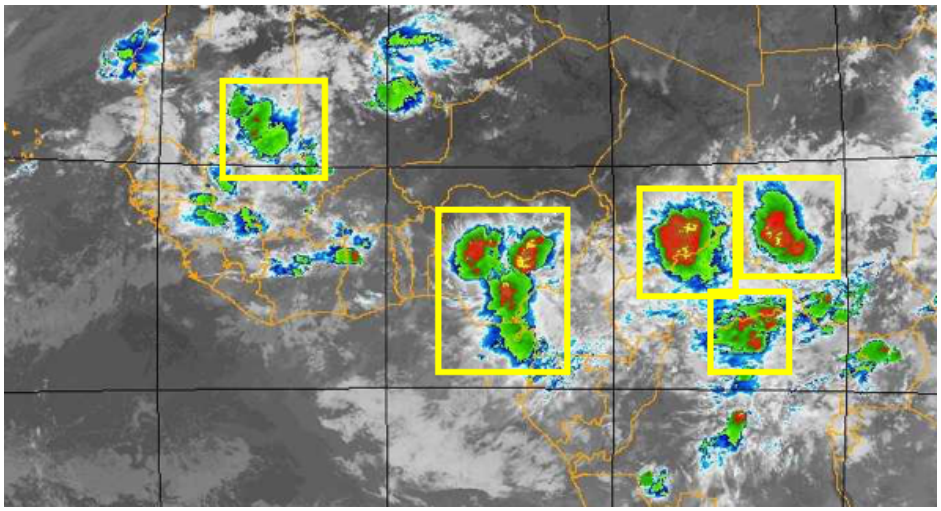

Figure 2.3: Yellow rectangles highlight MCSs on an infrared satellite image from 25 August 2017 at 0000Z (NRL 2017).

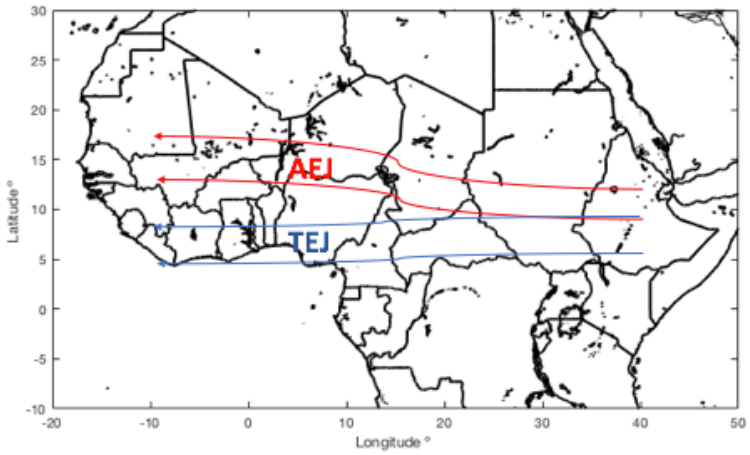

Figure 2.4: The locations of the prominent wind features, the African easterly jet (AEJ) at approximately 650 mb and the tropical easterly jet (TEJ) at approximately 175 mb.

There are other factors that can contribute to the ideal conditions in northern Africa for thunderstorms. During summer in the Northern Hemisphere there is warm, moist southwestern monsoonal flow in the low levels over northwestern Africa that is capped by dry easterly mid-level flow (Kirshnamurti et al. 2013). With a steady supply of heat and moisture at the surface, the shear environment is prime for thunderstorm development as the surface air rises into the mid-levels. Often convection on a smaller scale will dissipate once surface heating ceases. However, MCSs can form and enhance at night due to cooling cloud tops, promoting higher development in the atmosphere.

Relevant Research

The K Index

The K index is currently considered the index of choice for the tropics because it is targeted for forecasting airmass thunderstorms instead of convection from frontal systems or orographic lift (George 1960). It was created to forecast airmass

thunderstorms over the north central plains. A critical difference between K and the other common indices is the consideration of 700 mb moisture in the K index. Computing the K index is done using Equation 2.1:

$$KI = (850mb\ T - 500mb\ T) + (850mb\ Td) - (700mb\ T - 700mb\ Td) \quad (2.1)$$

where T is the air temperature and Td is the dewpoint temperature. KI is unique in its inclusion of 700 mb dewpoint depression, which incorporates "buoyancy and dry air entrainment in the tropical mid-troposphere" (Gálvez and Davison 2016). The values of the KI determine convective potential (Table 2.1). Since the KI is strictly used to forecast airmass thunderstorms, any lower atmosphere (below 700 mb) convergence (divergence) will increase (decrease) the frequency of thunderstorms (George 1960). The KI calculation does not include convergence or divergence; therefore, a forecaster must take these processes into consideration (George 1960).

Table 2.1: K values and their respective thunderstorm frequency estimates (George 1960).

K Value	Frequency Category
1. Less than 20	None
2. Above 20 but less than 25	Isolated thunderstorms
3. Above 25 but less than 30	Widely scattered thunderstorms
4. Above 30 but less than 35	Scattered thunderstorms
5. Above 35	Numerous thunderstorms

Considering the strength of the KI in forecasting airmass thunderstorms, its weaknesses in regards to tropical environments include "low variability in shallow convective regimes, […] and the disregard of thermodynamic properties below 850 mb" (Gálvez and Davison 2016). The K index neglects important factors for determining the

contribution of stability and moisture below 850 mb, which are important for tropical convection.

The Gálvez-Davison Index

The most imperative past research for study is the creation of the Gálvez-Davison Index (GDI) for tropical convection (Gálvez and Davison 2016). GDI is a new index that fills the gap in knowledge left by common convective indices lack of skill and accuracy in tropical regions.

GDI was built on knowledge of tropical convection and is specifically tailored for forecasting such events. It has four main components: equivalent potential temperature proxies core index (ECI), mid-level warming index (MWI), inversion index (II), and surface pressure correction for elevation (Co). GDI considers factors from three different layers in the atmosphere to calculate these components (Figure 2.5).

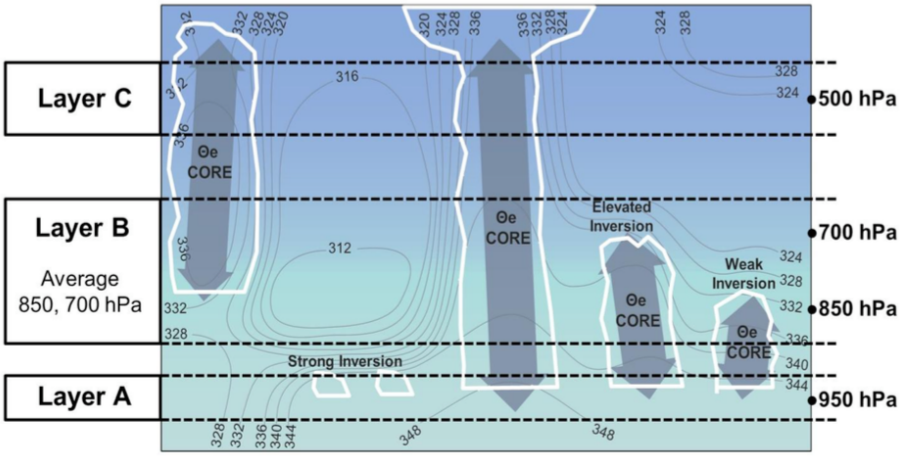

Figure 2.5: A model depiction of the layers in the GDI algorithm to predict convective potential (Gálvez and Davison 2016).

Equivalent Potential Temperature Proxies Core Index (ECI)

Equivalent potential temperature (EPT) is a quantity that encompasses both temperature and moisture in the air (Gálvez and Davison 2016). EPT can indicate "column moisture and potential release of latent heat" (Gálvez and Davison 2016). Higher EPT values are more favorable for convection, and a slow decrease with height is favorable for deep convection. Layer A was set at 950 mb to capture the characteristics of the boundary layer, since 925 mb proved be too high at times. The air temperature at the lifted condensation level (LCL) should be used when calculating the EPT. However, in order to simplify the calculations of EPT, GDI uses air temperatures at 850 mb, instead of the LCL air temperature.

Important factors to consider when forecasting tropical convection include moisture and trade wind inversions (TWIs). In the tropics, moisture is mainly a product of foregoing convection (Gálvez and Davison 2016). A feedback mechanism of moisture and convection indicates upstream convection could be a predictor for convection at the local site (Gálvez and Davison 2016). Another tropical phenomenon important for convection is the TWI, which is a minor decrease in lapse rate or a small increase in temperature with height. It is caused by subsidence from the descending air in the Hadley cell. Convective development depends on the strength and height of these inversions. Stronger and lower inversions will inhibit vertical development, while some growth can occur with weaker and higher inversions. Regions with ample moisture and limited inhibiting TWIs are prime for the development of convection.

Calculating the equivalent potential temperature proxy (EPTP) term for the GDI consists of incorporating ETPs from all three layers, A, B, and C.

$$\theta_A = \theta_{950} = T_{950}(1000/950)^{2/7} \tag{2.2}$$

$$\theta_B = 0.5(\theta_{850} + \theta_{700}) = 0.5\left[T_{850}(1000/850)^{2/7} + T_{700}(1000/700)^{2/7}\right] \tag{2.3}$$

$$\theta_C = \theta_{500} = T_{500}(1000/500)^{2/7} \tag{2.4}$$

Final EPTP values are calculated using the ETPs above in the following manner:

$$EPTP_A = \theta_A e^{\left(\frac{L_o r_{950}}{c_{pd} T_{850}}\right)} \tag{2.5}$$

$$EPTP_B = \theta_B e^{\left(\frac{L_o * (0.5(r_{850} + r_{700}))}{c_{pd} T_{850}}\right)} + \propto \tag{2.6}$$

$$EPTP_C = \theta_C e^{\left(\frac{L_o r_{500}}{c_{pd} T_{850}}\right)} + \propto \ . \tag{2.7}$$

In the equations above, the *r* values are all the mixing ratios at the specified levels, the empirical adjustment constant $\propto = -10\ [K]$, the latent heat constant $L_o = 2.69 * 10^6\ \left[\frac{J}{kg}\right]$, and the specific heat of dry air at constant pressure $c_{pd} = 1005.7\ \left[\frac{J}{kg\ K}\right]$.

Finally, a mid-level EPTP (ME) and a low-level EPTP (LE) determine the value of the ECI.

$$ME = EPTP_C - \beta \tag{2.8}$$

$$LE = EPTP_A - \beta \tag{2.9}$$

In the equations above, $\beta = 303\ [K]$ and is an empirical constant. The final ECI is calculated using Equation 2.10 below.

$$ECI = \begin{cases} \gamma * (EPTP_A - \beta) * (EPTP_C - \beta), & LE > 0 \\ 0, & LE \leq 0 \end{cases} \tag{2.10}$$

In Equation 2.10, $\gamma = 6.5 * 10^{-2}[K^{-1}]$ is an empirical scaling constant. Convective potential increases based on the difference between $EPTP_A$ and $EPTP_C$ and the β threshold. The GDI examines the 500 mb (Layer C) and 950 mb (Layer A) levels to

identify heat and moisture at the low and mid-levels. If both levels have significant amounts of heat and moisture, the whole column of air should be sufficiently warm and moist, favorable for convection.

Mid-Level Warming Index (MWI)

The MWI is a factor within the GDI that quantifies the stability change in the mid-levels based on temperatures at 500 mb. It identifies whether the presence of a warm ridge is increasing stability and inhibiting convection or a cool trough is decreasing stability and aiding convection growth. This index relies on the 500 mb air temperature departure from $\tau = 263.15$ [K] ($\sim -10°\mathrm{C}$). If the 500 mb temperature is warmer than τ, then the MWI is negative and decreases the magnitude of GDI. If the 500 mb temperature is cooler than τ, there is no contribution from MWI to the final GDI. MWI is only included when it indicates convection will be inhibited due to warmer mid-level air.

$$MWI = \begin{cases} \mu * (T_{500} - \tau), & T_{500} - \tau > 0 \\ 0, & T_{500} - \tau \leq 0 \end{cases} \tag{2.11}$$

In Equation 2.11, $\mu = -7\ [K^{-1}]$ is an empirical scaling constant. The MWI encapsulates the factor of mid-level stability by decreasing the final GDI value if warming is present at 500 mb.

Inversion Index (II)

The II is also an inhibiting portion of the GDI, but it considers the stability across the inversion and the presence of dry air entrainment above the inversion. These phenomena are captured in the stability factor (Equation 2.12) and drying factor (Equation 2.13).

$$S = \sigma * (T_{950} - T_{700}) \tag{2.12}$$

$$D = \sigma * (EPTP_B - EPTP_A) \tag{2.13}$$

In Equations 2.12 - 2.13, $\sigma = 1.5\ [K^{-1}]$ is an empirical scaling constant. Both the *S* and *D* factors are used to calculate the II.

$$II = \begin{cases} 0, & S + D > 0 \\ S + D, & S + D \leq 0 \end{cases} \tag{2.14}$$

GDI incorporates II only if II is negative. *S* is more negative when temperatures increase with height in the low levels. *D* is negative with when dry air entrainment is present. If there is not a sufficient inversion or dry air entrainment, II will not lower the final GDI value.

Correction for Elevation (Co)

GDI is optimal for quantifying convection potential for locations below 950 mb. In order to apply GDI to higher elevations, a terrain correction factor must be applied.

$$Co = 18 - \frac{9000mb}{P_{sfc} - 500mb} \tag{2.15}$$

This correction was proven to work well over the Mexican highlands in the original GDI study (Gálvez and Davison 2016). This correction is important for this study, especially in the Ethiopian highlands with heights up to 14,928 feet (ft). This height lies between 600-550 mb over northern Africa. As surface pressures reach 500 mb the denominator in the correction factor approaches zero, which is problematic. This is not a concern for this research.

The final calculation of GDI is the sum of the four factors (Equation 2.6). Values of GDI correspond to various convective potentials (Figure 2.6).

$$GDI = ECI + MWI + II + Co \tag{2.16}$$

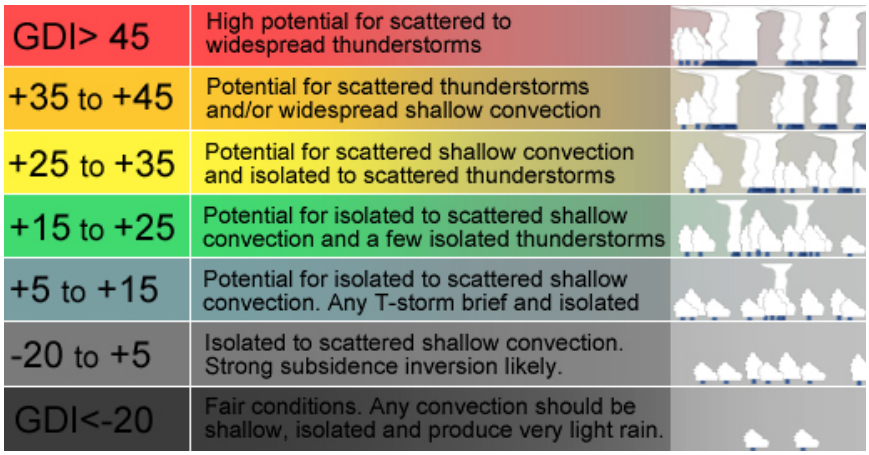

Figure 2.6: Convective potential based on the GDI value (Gálvez and Davison 2016).

Their findings concluded that GDI outperformed the TTI, LI, KI and the Convective Available Potential Energy (CAPE) for the region of their concern, Central America and the Caribbean (Gálvez and Davison 2016). They compared brightness temperatures with GDI values over the course of one rainy season in 2013. Out of the tested indices, GDI matched most closely with brightness temperatures for their area of concern. GDI excels in the 15°-25° latitudes "especially over oceans and eastern fringes of continents, where trade wind climates prevail" (Gálvez and Davison 2016). Accuracy in the GDI forecasts decreases near the intertropical convergence zone (ITCZ) because of less prevalent TWIs and fewer cool mid-level troughs in this region.

This research examines the applicability of the GDI to northern Africa, considering its strengths and weaknesses. One weakness is GDI's decreased skill in areas of "persistent ITCZ and deep-tropical convection environments" (Gálvez and Davison 2016). GDI's applicability to northern Africa is impacted by the lower number of mid-level troughs and TWIs compared to the Caribbean and the presence of the ITCZ. Another factor to consider is GDI's over-estimation of convective potential in thermal

low environments (Gálvez and Davison 2016). These limitations may hinder its applicability to northern Africa.

III. Methodology

Chapter Overview

The purpose of this chapter is to describe the data in this research as well as the methodology for analyzing the effectiveness of GDI and KI convective forecasts over northern Africa. This involves reanalysis weather data, lightning data which serves as truth for verification, GDI and KI parameters.

NCEP GFS Reanalysis Data and GDI Calculation

In order to analyze convective forecasts, the first step is to replicate the GDI calculation and confirm its accuracy with NOAA GDI forecasts. The National Center for Environmental Prediction (NCEP) GFS model reanalysis data is acquired from the University Corporation for Atmospheric Research (UCAR) Research Data Archive (RDA) where they host reanalysis data back to July of 1999 (UCAR 2014).

GFS reanalysis data is a rerun of the GFS model forecast with observations incorporated into the model to increase its accuracy. Reanalysis data incorporates 10% more observational data than the GFS model run (UCAR 2014). Some of this observation data includes radiosonde and satellite data. Since the reanalysis data incorporates more real-time observations, it is the best archived data describing the state of the atmosphere (UCAR 2014). The goal of creating these reanalysis files is to offer "the most realistic atmospheric analysis" or to archive data that is as close to the actual environmental conditions as possible (UCAR 2014).

GFS reanalysis data is the best option for this research. It is archived in gridded binary (GRIB2) format with a 1° by 1° horizontal resolution (UCAR 2017). The files

consist of numerous parameters characterizing the atmosphere at pressure levels from 1000 millibars (mb) up to 10 mb. Parameters required for the GDI calculation include temperature and relative humidity at 950 mb, 850 mb, 700 mb and 500 mb. For KI, the required parameters are temperature at 850 mb, 700 mb and 500 mb and relative humidity at 850 mb and 700 mb.

Equations 2.2-2.16 calculate the GDI and require mixing ratio values at certain levels. Since the GFS reanalysis files do not contain mixing ratio directly, it is calculated using relative humidity values. First, the dewpoint temperature (*Td*) is computed using relative humidity (*RH*) and air temperature (*T*) at the desired levels (Sensirion 2001).

$$H = \frac{log10(RH) - 2}{0.4343} + \frac{17.62 * T}{243.12 + T} \tag{3.1}$$

$$Td = \frac{243.12 * H}{17.62 - H} \tag{3.2}$$

Next, the saturation mixing ratio (es) and the mixing ratio (*r*) are computed using the dewpoint temperature (Davies-Jones 2009).

$$es = 6.112 * e^{\frac{17.67*(Td - 273.15)}{Td - 273.15 + 243.5}} \tag{3.3}$$

$$r = \frac{0.6220 * es}{X\ mb - es} \tag{3.4}$$

In Equation 3.4, *X mb* refers to the value of the pressure level in millibars (e.g. 950 mb, 850 mb, 700 mb or 500 mb). EPTP values are then calculated using the mixing ratios and Equations 2.5-2.7 for each layer: A, B and C. After ingesting the temperature and relative humidity data for each level and computing the mixing ratios, the components of GDI are calculable with the addition of certain empirical constants.

After calculating the GDI, the next step is to plot it over Africa. This requires a map of Africa with geographical and political boundaries. The data for the map is acquired from the CIA World Databank II website where they store files of geographical and political data by continent (Pape 2004). Political borders for these files are current as of 1990. Data for both the continents of Africa and Asia are used to map the continent of Africa along with a portion of the Middle East to match the NOAA Africa model GDI maps. Each file consists of sub text files that are broken down by continent and type of data. For each of the "national boundaries" and "coastline, islands and lakes" files are used in order to map out the geographical features and political boundaries.

To create a map similar to the NOAA Africa model GDI maps, the GDI values and Africa map data are all plotted on the same graphic. This data is bounded by 40°S-40°N latitude and 25°W-60°E longitude. The contour scale is set in increments of 10 GDI (5-15, 15-25, etc.) to match the NOAA Africa model GDI scales (Figure 2.2). For comparison colors are chosen to mimic NOAA's color scheme. White areas have insufficient reanalysis data to calculate GDI.

The area of interest for this research is between 0°-20°N latitude and 20°W-50°E longitude. Validation of these replicated GDI output maps was confirmed by the developer (Dr. Gálvez, NOAA) through visual inspection of a GDI forecast image (Gálvez, National Oceanic and Atmospheric Administration, written communication 26 July 2017).

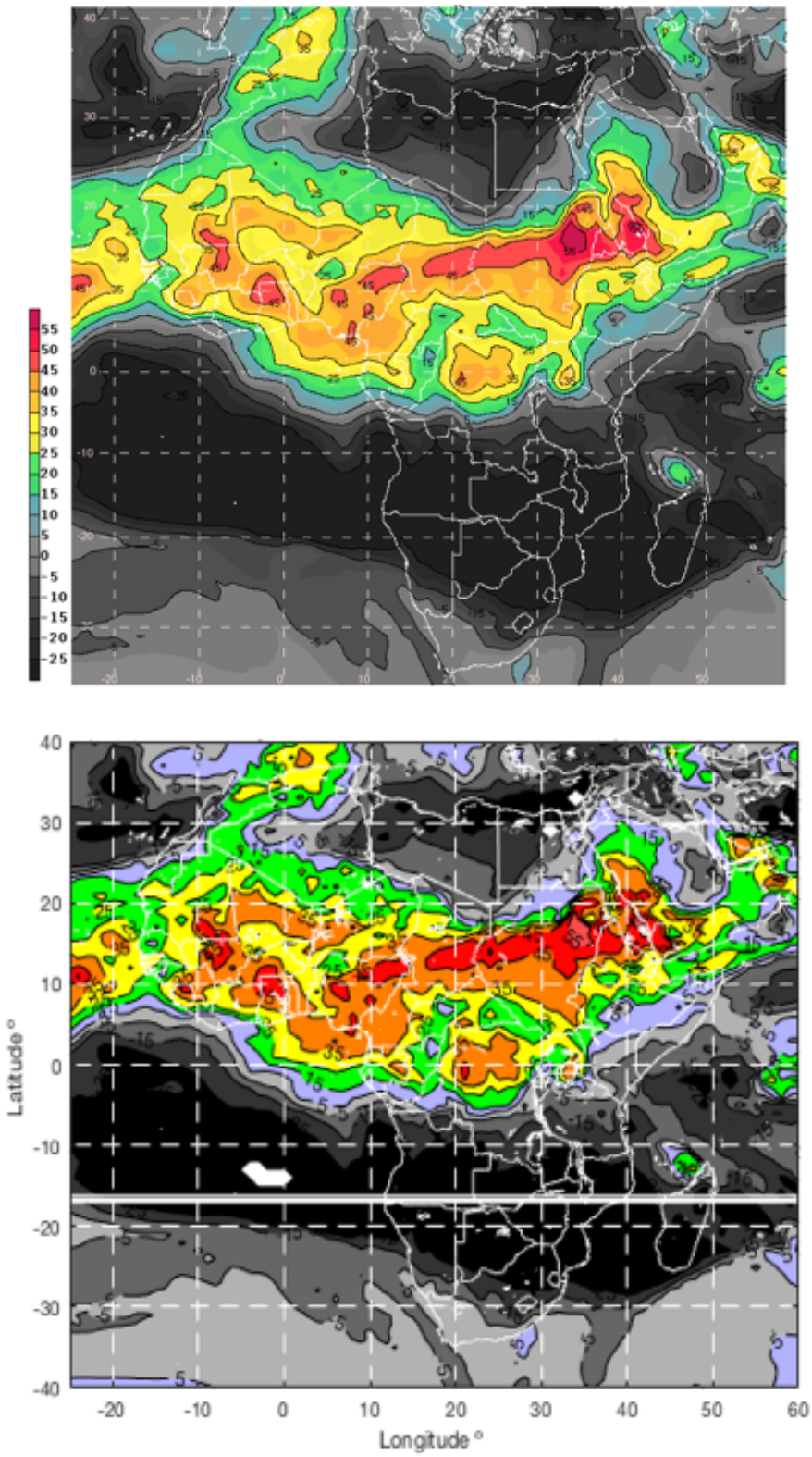

Figure 3.1: A comparison of NOAA's model GDI (top; NOAA 2017) and the reanalysis GDI (bottom) from 01 August 2017 at 00Z.

K Index (KI) Calculation

Plotting the KI is similar to plotting the GDI. An identical process plots the geographic and political boundaries in the region of interest, and calculates the dewpoint temperatures using Equations 3.1 and 3.2. When plotting the KI, minor changes in the contour scale are made to match the KI convective potential thresholds (Table 2.1). KI is in increments of five, while GDI is increments of ten and different values of KI indicate different convective potentials compared to GDI. The KI map is made to look similar to the GDI map with grey and black representing low KI values (low convective potential), while other colors represent various convective potentials from green to red (Figure 3.2).

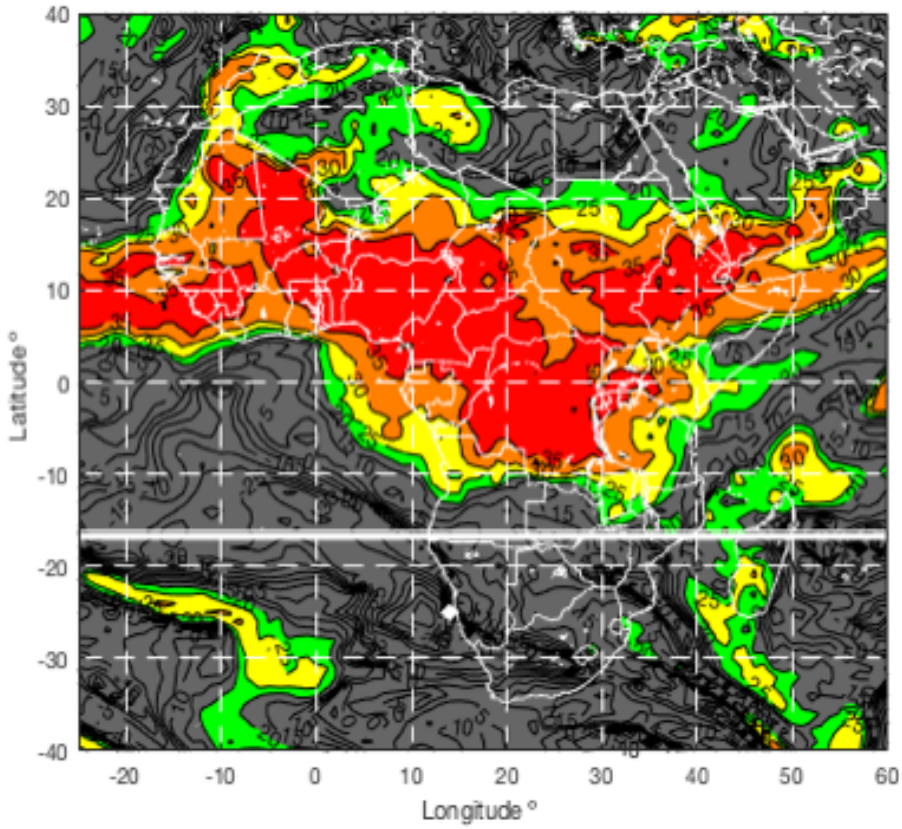

Figure 3.2: KI mapped out over Africa for 15 August 2016 at 06Z.

Plotting ATDNET Lightning Data

For verification purposes, lightning data is also plotted across northern Africa. This data is from the arrival time difference (ATD) thunderstorm detection system known as Sferics or ATDNET (AFWA 2012). Sferics is a system employed by the United Kingdom Meteorological Office and utilizes the arrival time differences of the signals from lightning strikes to identify their location.

The ATDNET is a network of sensors for lightning detection. Each New OutStation (NOS) sensor "listens" at a very low frequency (VLF) of about 13.7 kHz, a radio wave on the electromagnetic spectrum (AFWA 2012). This allows for the sensors to have a very long range for listening. Once four NOS sites detect a lightning strike, the system locates the flash based on the arrival time at all four stations. Strike location error ranges from 8-24 km across northern Africa (AFWA 2012).

Data extracted from these files includes the date, time, latitude and longitude of each strike in order to display them spatially. Lightning strikes are cyan asterisks on the index plots to indicate where lightning occurred (Figure 3.3).

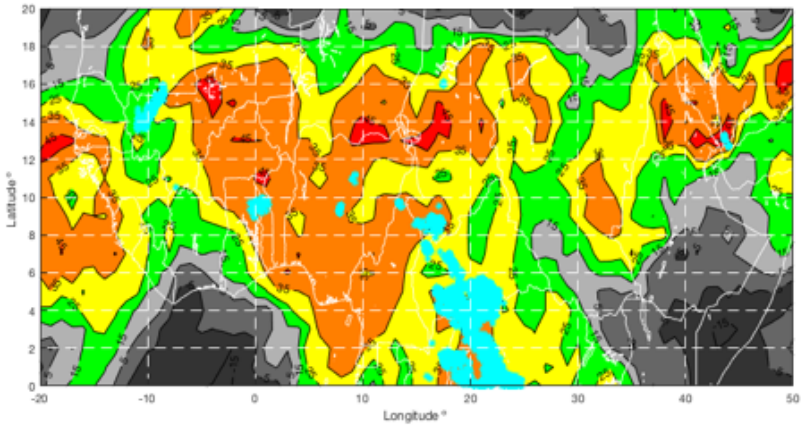

Figure 3.3: GDI values (colored contours) at 15 August 2016 at 06Z, with Sferics lighting data (cyan asterisks) from 15 August 2016 from 05-07Z.

NRL IR Satellite Images

Lightning data is compared with corresponding satellite imagery. Satellite images are acquired from the Naval Research Laboratory (NRL) Next Generation Weather Satellite Demonstration Project (NexSat), which is a partnership with the Cooperative Institute of Research in the Atmosphere (CIRA; NexSat 2011). For this project, Meteosat8 infrared (IR) images are selected, with a color filter to highlight the cloud top temperatures (-20°C). These images encompass a majority of the region of interest only missing a few degrees of longitude on the eastern side of the image (Figure 3.4).

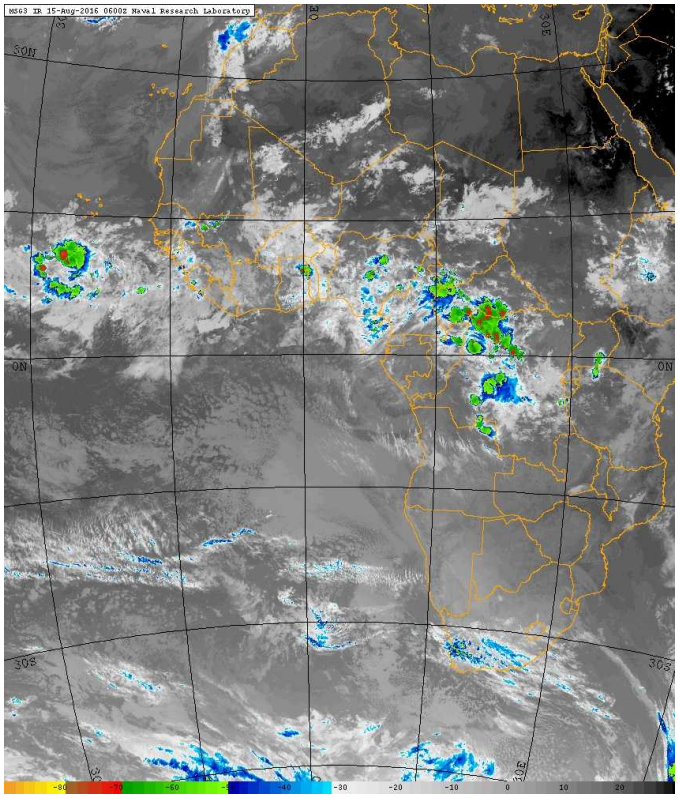

Figure 3.4: A color IR satellite image from 06Z on 15 August 2016 with cloud top temperatures (°C) indicated by the color filter (NRL 2017).

Once the lightning data is plotted, visual comparison determines if the lightning correlates to cold cloud tops. The freezing level over Africa ranges between 550-500 mb or 16,000-19,000 ft in height on forecast model skew-t profiles. Cloud electrification

requires the presence of frozen drops or graupel particles (Toracinta et al, 2001). Most lightning over Africa occurs with 40 decibel echoes reaching heights of 8 km, or approximately 26,000 ft, with cloud tops extending above that height (Toracinta et al, 2001). For this research, areas of cloud heights at or above the -20°C height, which is approximately 25,000 ft high on model skew-t vertical profiles in December, are sufficient for lightning to occur. True lightning strikes match cloud regions at or below -20°C on IR images.

Methods for Comparing Index Forecasts

Once the forecast index values and lightning data are plotted, the next step is to quantify the quality of the forecast for that moment in time. A point-by-point method is insufficient for this research. Since the forecast model data is mapped onto a 1° by 1° grid, this analysis requires assessing each point to see if the forecast correctly identifies lightning to occur. Two problems arise during this process. Of note, lightning rarely strikes at whole latitude and longitude degree values. Instead the lightning strikes, as expected, scatter out at various locations with fractional degree values of both latitude and longitude. This is problematic when comparing index values with truth, attempting to match model output data at whole degree values with sporadic lightning strikes. Moreover, interpretation of convective index forecast values indicating various levels of convective potential is subjective. Because of these two issues, a clustering method is appropriate, not point-by-point analysis.

Clustering methods are used in similar research, such as identifying storms, clouds and precipitation fields (Marzban and Sandgathe 2005; Singh and Gill 2013).

Overall, cluster analysis recognizes desired features in both forecast and observation fields for the purpose of comparing their characteristics (Singh and Gill 2013). This research divides the lighting and forecast data into the same number of clusters, or groupings of data points, then matches clusters to compare their location and spatial coverage differences. The differences in location are location error values and spatial or coverage differences are area error values. In the end, assigning each observation cluster to one forecast cluster is the best solution for error analysis in this case.

Hierarchical Clustering Method

A study on precipitation fields uses the agglomerative hierarchical clustering method (Marzban and Sandgathe 2005). This method separates data points into clusters starting with each point as its own cluster and then matching each cluster to the closest other cluster in each matching process (Marzban and Sandgathe 2005). Hierarchical clustering places each data point into clusters through this manner until there is one cluster. In this method, the number of clusters is treated as a variable changing over time and producing different error values (Marzban and Sandgathe 2005). Therefore, error fields are created for each data set. The user can decide where the cut off will be in the number of clusters based on the dendrogram graphic displaying how the clustering of the data points in sequential iterations (Marzban and Sandgathe 2005). Because the number of lightning data points is much greater than the forecast data points, the number of observed clusters would be significantly more than the forecast clusters. Creating the dendrograms and determining the ideal number of clusters is temporally and computationally expensive. Hierarchical clustering has been shown to be useful in other research, but would not work well for this research. This project requires the same

number of clusters to be specified between the observed and forecasted data to calculate location and area error. Since there was no solid manner to identify the ideal number of clusters from the dendrogram that works well for both forecast and observed data, hierarchical clustering is not the ideal method for analyzing data in this project.

K-Means Clustering Method

Another way to group data into clusters is the k-means clustering method. Unlike hierarchical clustering, k-means clustering focuses on idealizing the number of clusters by balancing the number of clusters with the total sum of the distances between data points and their centroids (Singh and Gill 2013). A centroid is the center of a cluster. K-means clustering groups data points by randomly placing k number of centroids in the data and assigning each data point to the closest centroid (Singh and Gill 2013). The distance from each data point to its centroid is summed up into the total point-to-centroid distance for that number of clusters. Then k number of centroids are placed throughout the data again and the data points are grouped into new clusters. This is done ten times for each k number of clusters and the smallest sum of total point-to-centroid distances is saved.

At first the number of clusters, k, is one and the lowest total point-to-centroid distance is saved. Then the minimum total point-to-centroid distance is determined for two clusters, then three and so on, up to ten clusters. Once all the minimum total point-to-centroid distances are plotted, the idealized number of clusters is identified using what is called a K-pick plot (Figure 3.5).

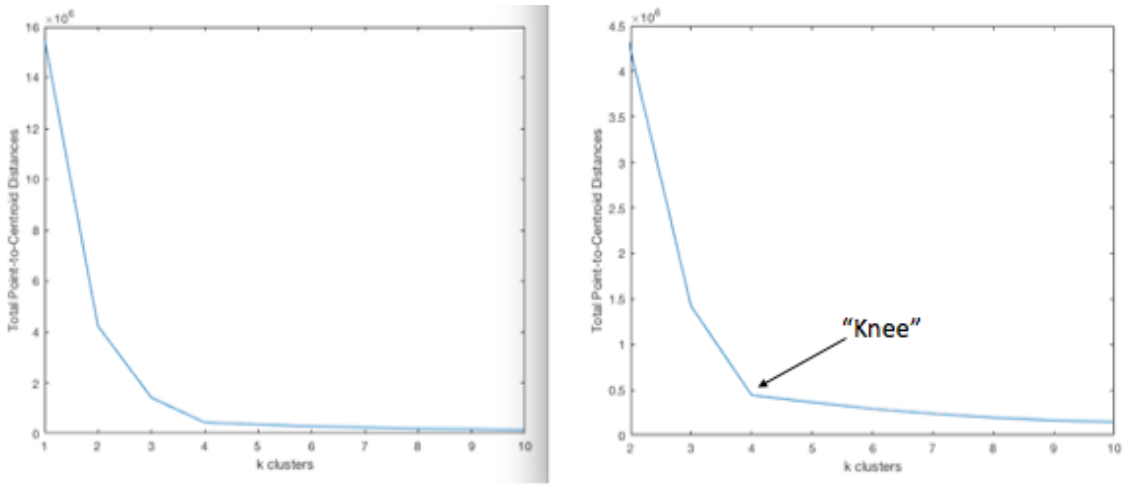

Figure 3.5: An example K-pick plot, full view from 1-10 clusters (left) and a zoomed-in view, 2-10 clusters (right).

The total point-to-centroid distances from the lightning data resulted in distances on the order of 10^8, as seen on the y-axis (Figure 3.5). It is difficult to properly choose an ideal number of clusters on such a plot, so a zoomed-in version of this plot was created to highlight the ideal number of clusters. The K-pick plots emphasize the decrease in total point-to-centroid distance with increasing cluster number, k (Singh and Gill 2013). In the example above, four was chosen for k because it is the last increase in cluster number associated with a large decrease in total point-to-centroid distance. The ideal k number of clusters is found at the bottom of the "knee" made by the curve in the K-pick plot. Beyond that point on the curve, the total sum of distances does not decrease significantly with each added cluster. Using this method, the number of clusters is chosen in an objective manner based on the K-pick plot of the lightning data at each time analyzed. The ideal number of clusters is then applied to the lightning data and each index analyzed for each particular time.

Error Analysis Method

Error analyses are conducted to assess the forecast quality of each applicable index at desired timeframes. The process begins by plotting the observed lightning data on top of the index being examined for that particular time (Figure 3.4). In order to conduct the analysis, only certain data from the index points are kept, the data points indicating scattered thunderstorms to occur. This scattered thunderstorm threshold is GDI values at or above 35 and KI values at or above 30 (Figure 2.6 and Table 2.1). Index values at or above those thresholds are selected and all others are omitted (Figures 3.6 and 3.7).

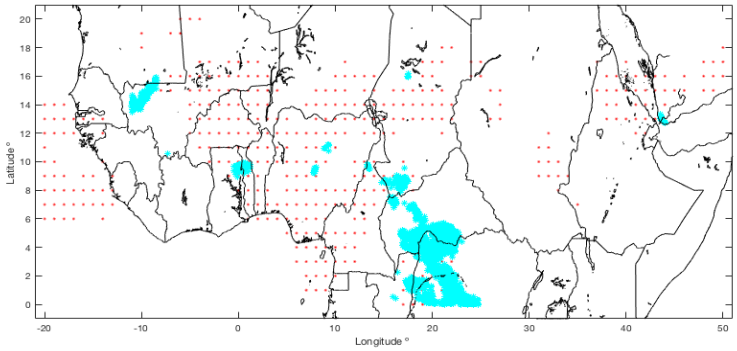

Figure 3.6: GDI values 35 and above (red dots) and lightning data (cyan asterisks) on 15 August 2016 at 06Z.

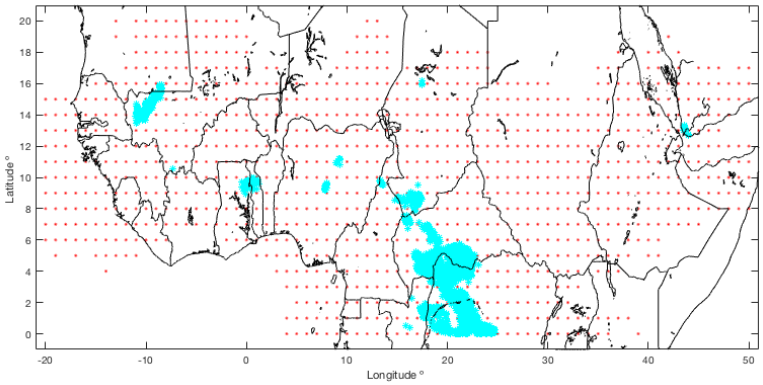

Figure 3.7: KI values 30 and above (red dots) and lightning data (cyan asterisks) on 15 August 2016 at 06Z.

Next, the lightning data is assessed for the ideal number of clusters to divide the data into using the K-pick plot process described above. Once the ideal number of clusters is acquired, both the lightning and index data at or above the scattered thunderstorm threshold are divided up into k clusters (Figure 3.8). Specified colors are assigned to clusters randomly. Cluster one is blue, cluster two was red, etcetera. Lightning cluster one, blue, does not always match index cluster one, also blue. Clusters are matched by the researcher examining and assigning clusters from east to west and north to south. Once the clusters are matched, the location and area error values are calculated.

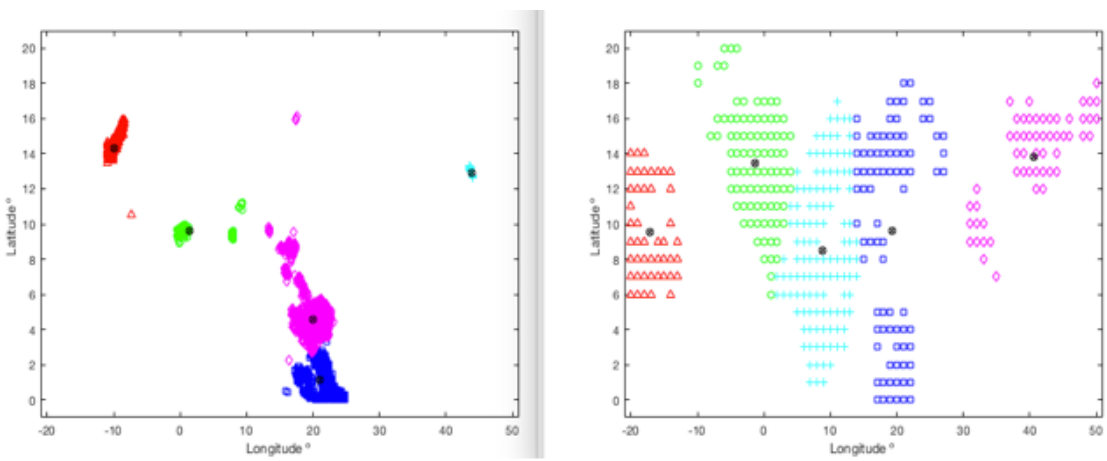

Figure 3.8: Lightning data (left) and GDI at or greater than 35 (right) divided into clusters from 15 August 2016 at 06Z.

Clustering is the appropriate method chosen for this research because of the ability to assess both location and area errors. These factors are calculated in the following manner:

$$Location\ Error = Distance\ between\ forecast\ cluster\ centroid\ and\ observed\ cluster\ centroid;$$

$$Area\ Error = Average\ distance\ between\ data\ points\ and\ forecasted\ cluster\ centroid - average\ distance\ between\ data\ points\ and\ observed\ cluster\ centroid.$$

Location error is the distance between an observed centroid and a forecast centroid. Area error is the difference of the average distance between a centroid and each of the data points in an observed and forecast cluster. Although more complex than the location error, area error is valuable in portraying the horizontal expanse of the cluster, as each point-to-centroid distance in the cluster is given the same weight. The number of observed and forecast cluster pairs ranges from four to six, depending on the lightning K-pick plot. Location and area error values from each cluster pair are averaged for these six-hour time increments.

This research examines the effectiveness of both GDI and KI forecasts using error values across all seasons in an intra-annual study with an in-depth analysis of their capabilities during the most active convective season: The Northern Hemisphere late summer and early fall. To obtain sufficient knowledge on the forecast ability of GDI and KI over northern Africa in all seasons, time periods evenly spaced throughout the year are chosen for analysis. Times include 00Z, 06Z, 12Z and 18Z on 15-17 February, 15-17 May, 15-17 August and 15-17 November of 2016. These times are selected to reflect even amount of time between them and analyzing data across all seasons, without regard for any particular synoptic situation. Since most thunderstorms occur in Africa during the Northern Hemisphere summer, an in-depth analysis of 15-17 August 2016, 19-21 August 2017, 25-27 August 2017, 15-17 September 2017 and 26-27 September 2017 is conducted to include various synoptic situations in late summer and early fall to garner further knowledge on the forecast ability of convective indices.

For a further analysis on the data collected, statistical methods are needed to assess the robustness of the data and the confidence levels that can be applied to

conclusions. Bootstrapping is the statistical technique used in this research, which utilizes the data to expand itself and determines confidence intervals based on synthetic values.

Bootstrapping Statistical Method

Often times parametric statistics are utilized to calculate confidence intervals or standard error of data sets (Ong 2014). However, these methods make the underlying assumption that a data set is normally distributed. Unless a data set is symmetric with a standard deviation of one, this assumption is not applicable and error results are unreliable (Ong 2014). On the other hand, non-parametric methods do not rely on such assumptions of the distribution of sample data sets. These methods resample the data with replacement and assume the sample distribution is representative of the population (Ong 2014). Since a normal distribution cannot be assumed for the data sets in this research, a resampling technique called bootstrapping is utilized.

Bootstrapping is a statistical method for expanding a data set by inflating it without changing its characteristics for statistical analysis. Because it is based on the law of large numbers, bootstrapping is a solid method for creating sufficient data so that the "empirical distribution will be a good approximation of the true distribution" (Orloff and Bloom 2014). This technique became well known in the late 1970s, but practical only later with high speed computational resources for implementation. Computations are conducted "on the data itself to estimate the variation of statistics that are computed from the same data" (Orloff and Bloom 2014). It is named after the metaphor of pulling oneself up by one's bootstraps. For this research, each set of error data, location and area from each two to three-day time period is bootstrapped and expanded to 10,000 points.

The end goal is to compare the confidence intervals of each error data set to the others in order to assess how well each index performed.

The bootstrapping statistical analysis begins by calculating artificial means of the data. Error data is sampled—GDI location error for example—with replacement and averaged. This creates an artificial mean value close to but not equal to the true mean. 10,000 artificial means are calculated from each data set to acquire a quality estimate of the 95% confidence interval (Orloff and Bloom 2014). A high level of confidence, or 95% confidence interval, is desired in this project. Bootstrapping allows researchers to estimate confidence intervals with high accuracy even with small data sets.

There are multiple ways to calculate confidence intervals from bootstrapped data. For a 95% confidence interval, the percentile method would use the 0.975 and 0.025 critical values, or 9,750th and 250th largest values in a 10,000-member data set as error bar end points above and below the actual mean, respectively (Orloff and Bloom 2014). A more accurate method is the bias-corrected and accelerated (BC_a) method (Efron and Tibshirani 1993). The BC_a method comes closest to fulfilling the standard of good confidence intervals, meaning they "closely match exact confidence intervals" and "give dependably accurate coverage probabilities in all situations" (Efron and Tibshirani 1993). Confidence intervals constructed using BC_a are more accurate overall and recommended especially for small sample sizes, like the 12-member data sets in this research (Wilks 2011). BC_a is more advanced through its incorporation of the cumulative distribution function (CDF) of the standard Gaussian distribution along with a bias correction parameter that "reflects median bias of the bootstrap distribution" to account for partiality (Wilks 2011). It also includes the acceleration parameter, which corrects for the skewness

of the data. By incorporating more parameters into its calculation that encapsulate the characteristics of the data, the BC_a method produces more accurate confidence intervals.

BC_a confidence intervals are calculated using the desired number of bootstrapped samples (10,000) the calculation desired (averaging) and the data for calculating the confidence intervals. Error bars are plotted using the mean of the data set and the lower and upper bounds of the confidence interval (Figure 3.9). A 95% confidence interval is shown, where 95% of all possible mean values fall into that range. Through BC_a, confidence intervals are created for location and area error data for each index across all the times analyzed.

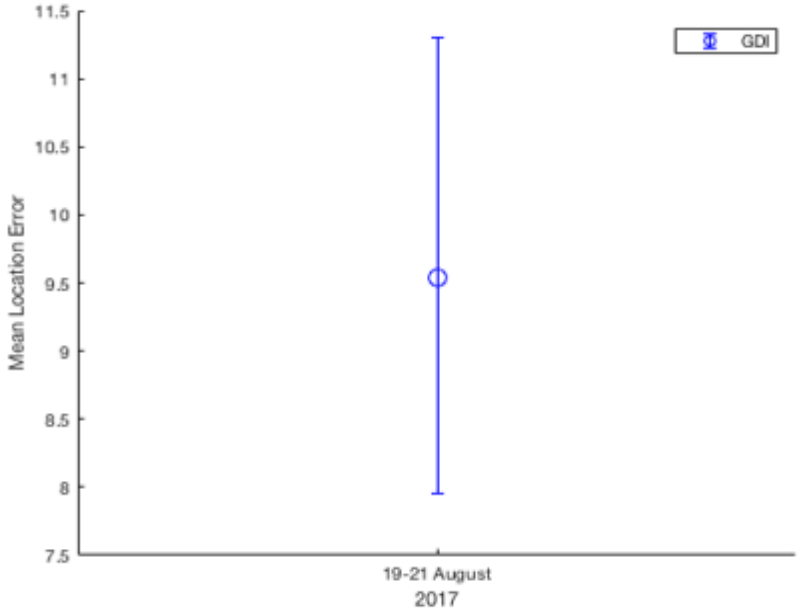

Figure 3.9: An example confidence interval plot with the mean (circle) and 95% confidence interval (error bars) for the GDI location error from 19-21 August 2017.

Summary

In order to conduct an error analysis of convective index forecasts in this project, the following methodology is implemented. First, convective indices are plotted using

GFS reanalysis data, with lightning data overlaid, and validated using real-time IR satellite imagery. Then, the lightning and index data are separated into the same number of clusters through the k-means clustering method. Paired clusters are examined to calculate both location and area error values. Finally, the error data is expanded using bootstrapping statistical methods and confidence intervals are calculated using the BC_a method.

IV. Analysis and Results

Chapter Overview

The purpose of this chapter is to evaluate and convey the results from the intra-annual and intra-seasonal (summer) studies. Also included are the results from an initial analysis on GDI modifications tailored specifically for Africa (GDI-As). The results include both the location and area error values.

Intra-Annual Study

A study examining the index forecast errors across all seasons is conducted to assess the effectiveness of each index's forecasts throughout the year. Error analysis method described in Chapter III is conducted for the following dates and times in 2016: 00Z, 06Z, 12Z and 18Z on 15-17 February, 15-17 May, 15-17 August and 15-17 November. These dates and times are chosen because they are evenly spaced throughout a year and encompass each season.

Intra-Annual Study: Location Error

Location error quantifies the distance between the observed and forecast lightning clusters. Mean location error values are plotted with 95% confidence intervals across all seasons (Figure 4.1). The closer these error values are to zero, the closer the forecast convection was to the actual lightning strikes. Results show mean error values of both GDI and KI are relatively close throughout the year. Both indices' location error values follow the same pattern, being greatest in February and least in August. More confidence

can be placed in both GDI and KI forecasts for location of convection in the summer than in the winter.

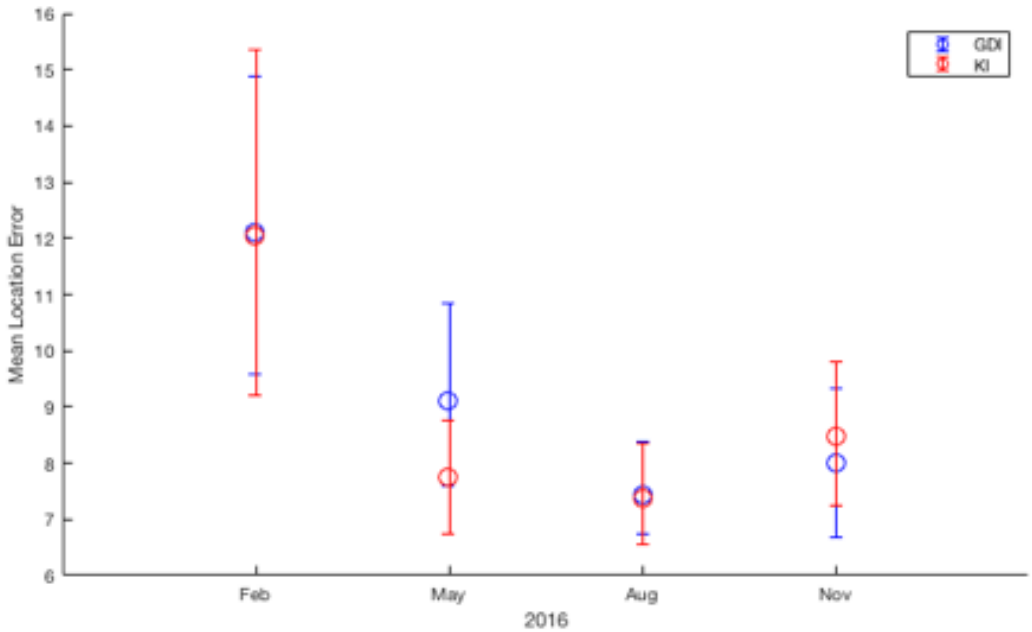

Figure 4.1: Intra-annual location error values of GDI (blue) and KI (red) forecasts. Means are indicated with circles and the 95% confidence intervals are indicated with error bars.

When comparing GDI and KI forecasts for location error intra-annually, little difference in locating convection can be concluded. For February, August and November there is little difference in the mean location error values and the 95% confidence intervals between the two indices. The largest difference can be seen in the location error values for May, where the mean values are the furthest apart. However, since the confidence intervals overlap, not much confidence can be placed in stating KI's forecast was more accurate than GDI's. Overall, GDI and KI location error values are similar when forecasting convection intra-annually.

Intra-Annual Study: Area Error

Area error quantifies the accuracy of the spatial coverage depicted in index forecasts when compared to observed lightning. GDI error values (blue) and KI values (red) with error bars attached displaying the 95% confidence intervals are plotted (Figure 4.2). Although the intra-seasonal location errors between GDI and KI are similar, a seasonal pattern can be identified with GDI area error values being lowest in February and highest in August, which is the opposite of the location error intra-annual pattern. KI area error values fluctuate less in the intra-annual case and are higher than any of the GDI area error values. Because these are all positive values, convective index forecasts almost always depict larger spatial coverage for lightning than what occurs. That difference in spatial coverage varies between indices intra-annually.

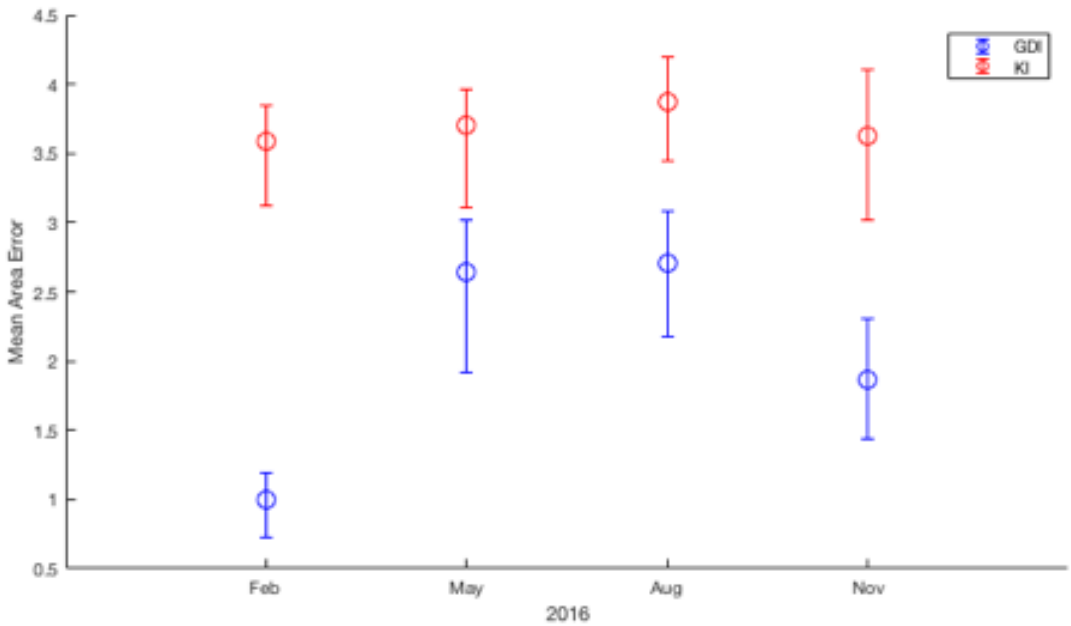

Figure 4.2: Intra-annual location error values of GDI (blue) and KI (red) forecasts. Means are indicated with circles and the 95% confidence intervals are indicated with error bars.

GDI consistently forecasts the areal coverage of lightning more accurately than KI throughout the year. This is shown by the GDI mean values being lower than KI mean

values in all seasons, as well as the confidence intervals never overlapping. The confidence intervals are corrected for biases within each data set through the bias-correction and acceleration method (described in Chapter III).

Intra-Seasonal (An In-Depth Late Summer/Early Fall Study)

An in-depth study is conducted to examine the forecast skill of GDI and KI during the most active convective season. In this study the dates are chosen specifically for the synoptic situation present over northern Africa, analyzing days with predominant MCSs (15-17 September 2017; Figure 4.3), a mix of MCSs and smaller airmass thunderstorms (15-17 August 2016, 19-21 August 2017 and 25-27 August 2017; Figures 4.4-4.6), and just airmass thunderstorms (26-27 September 2017; Figure 4.7).

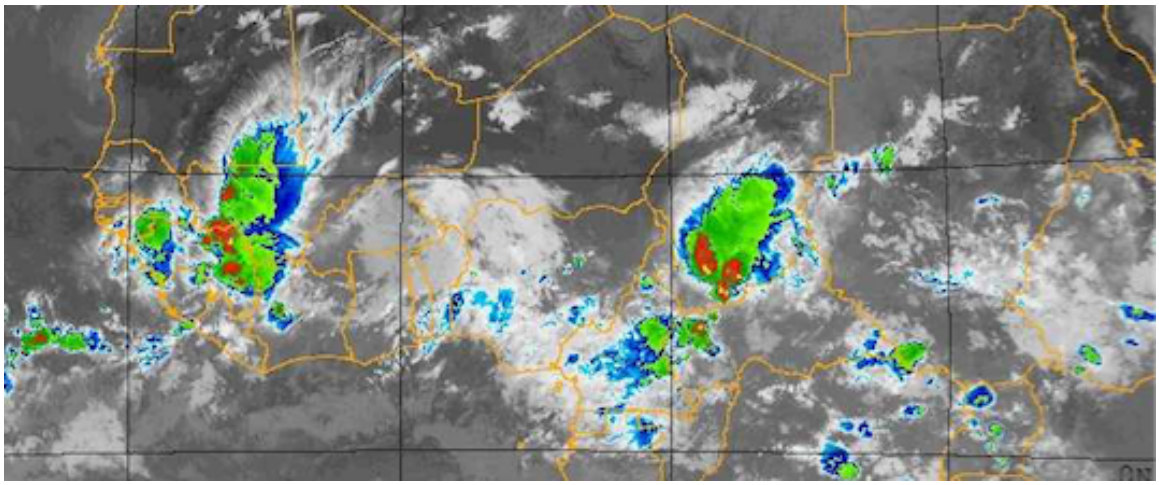

Figure 4.3: IR satellite image from 17 September 2017 at 00Z (NRL 2017).

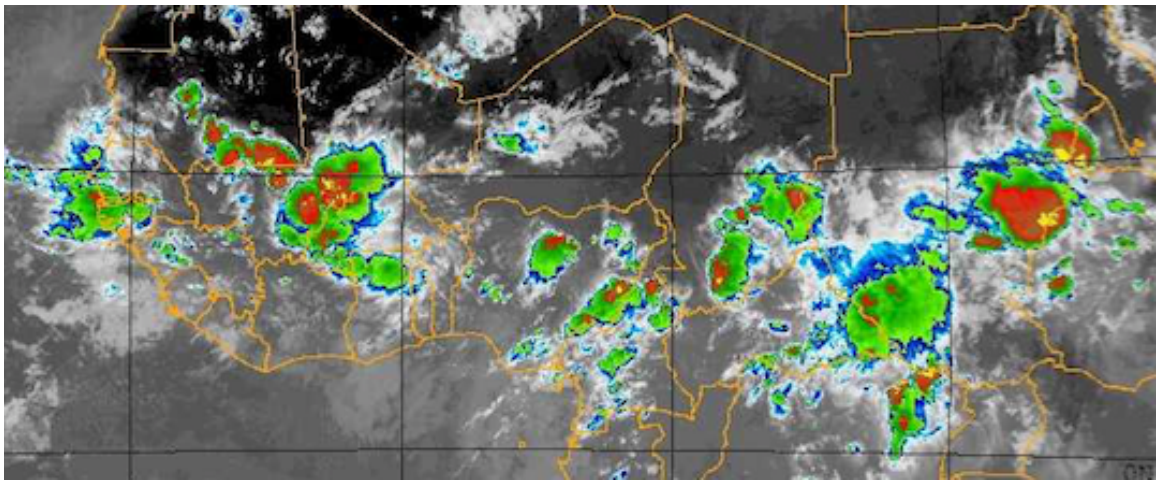

Figure 4.4: IR satellite image from 16 August 2016 at 18Z.

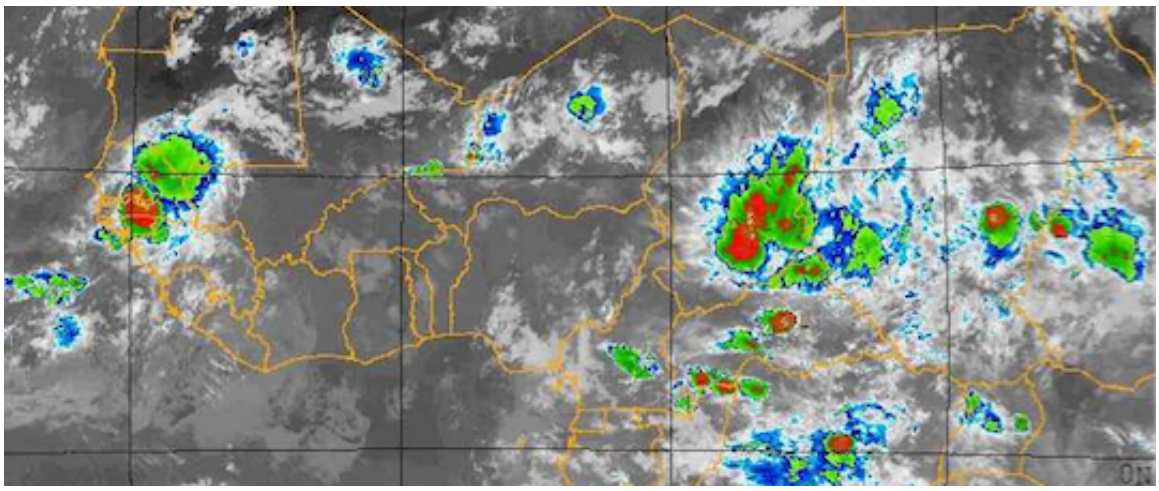

Figure 4.5: IR satellite image from 20 August 2017 at 00Z.

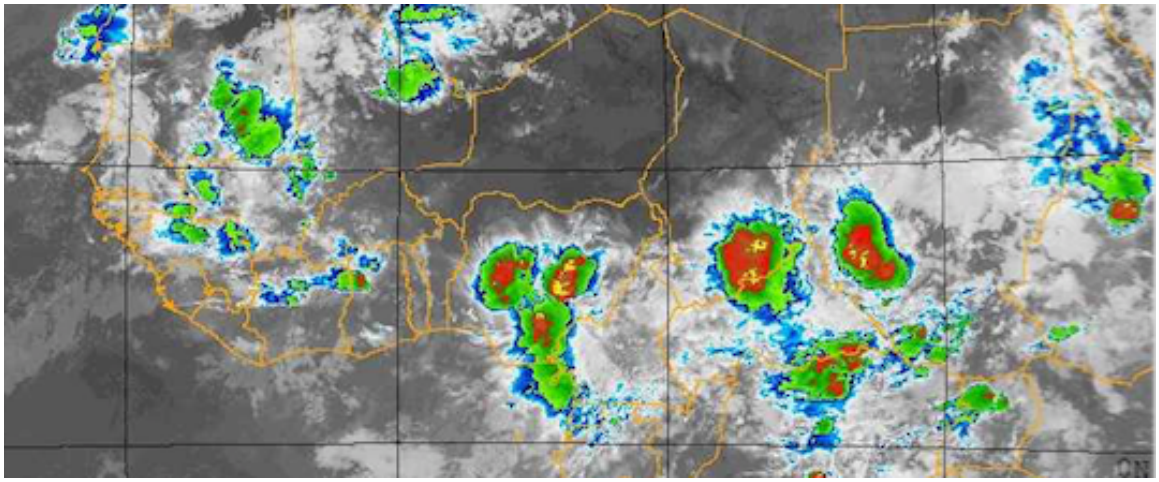

Figure 4.6: IR satellite image from 25 August 2017 at 00Z.

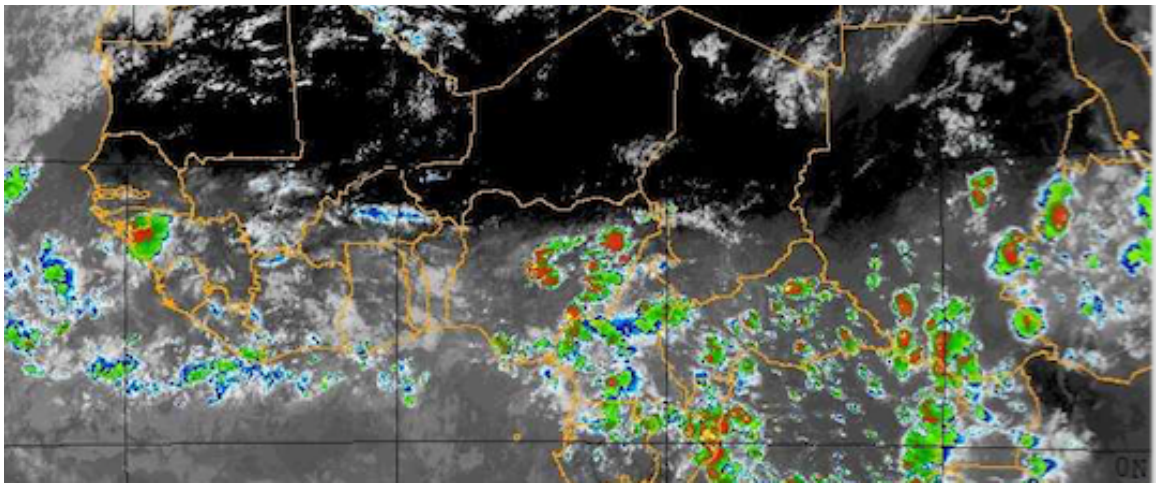

Figure 4.7: IR satellite image from 27 September 2017 at 15Z.

Intra-Seasonal Study: Location Error

For the intra-seasonal study location error values (Table 4.1 and Figure 4.8) each date range is labeled with the dates as well as the predominant convective synoptic situation for those days where MCS denotes convection is predominantly from mesoscale convective systems, and AT denotes convection is predominantly airmass thunderstorms. If a date range has both synoptic situations, the first one listed is the prevailing cause of convection. For 19-21 August, 2017 a majority of the convection is from MCSs, but airmass thunderstorms also exist. However, during 25-27 August 2017 the prevailing cause of convection is airmass thunderstorms with some MCSs present.

Table 4.1: GDI and KI location error ranges by date range.

	Convection Type	GDI Location Error Range in °	KI Location Error Range in °
15-17 Aug 2016	MCS/AT	6.7465 - 8.4119	6.5795 – 8.3739
19-21 Aug 2017	MCS/AT	7.8819 - 11.2938	6.6512 – 9.4888
25-27 Aug 2017	AT/MCS	7.4489 - 10.2761	6.8591 – 10.3150
15-17 Sep 2017	MCS	6.3466 - 9.8571	5.6765 – 9.1606
26-27 Sep 2017	AT	7.0432 – 13.0533	5.4714 – 10.4461

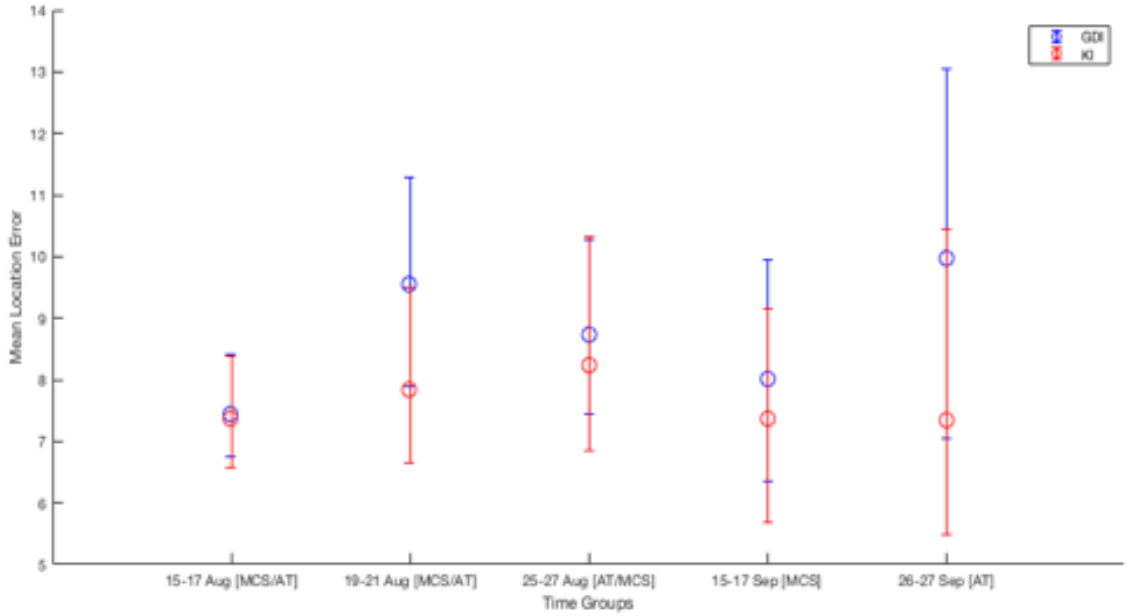

Figure 4.8: Intra-seasonal study location error with 95% confidence intervals grouped by days with predominant convective synoptic situations indicated, MCS for mesoscale convective systems and AT for airmass thunderstorms.

The values from August 2017 and mid-September of 2017 relate to the values from August of 2016, further supporting that GDI performs similar to KI in terms of location error. Late September exhibits a wider range of values where the predominant convective synoptic situation is airmass thunderstorms. These values are comparable to the November, May and February values from 2016. This indicates, once again, that both indices are more accurate at locating convection in the Northern Hemisphere during summer months and decrease in accuracy in transition and winter seasons.

Similar to the intra-annual study, the location errors between GDI and KI in the summer cases are not significantly different. Even in various convective synoptic situations, predominant MCSs or airmass thunderstorms, the 95% confidence intervals for GDI and KI overlap. Whether convection is present in large conglomerates or spread out in smaller clusters over a larger area, both indices perform at the same level.

Intra-Seasonal Study: Area Error

Area values from the intra-seasonal study are also analyzed (Table 4.2 and Figure 4.9). The same date ranges are shown, as well as the predominant convective synoptic situations. The intra-seasonal study further supports the intra-annual patterns in both GDI and KI area error values. The GDI is most inaccurate, indicated by higher area error values, in the summer. KI accuracy increases and error decreases in the September date ranges. However, KI has no significant change in area error values, which remain consistently in the same range as the intra-annual KI values.

Table 4.2: GDI and KI area error ranges by date range

	Convection Type	GDI Area Error Range in °	KI Area Error Range in °
15-17 Aug 2016	MCS/AT	2.3177 – 3.2114	3.535 – 4.2692
19-21 Aug 2017	MCS/AT	1.8735 – 2.6472	2.8435 – 3.8641
25-27 Aug 2017	AT/MCS	2.5772 – 3.7677	3.7325 – 4.8509
15-17 Sep 2017	MCS	1.5782 – 2.2652	3.1127 – 3.7854
26-27 Sep 2017	AT	1.3840 – 2.6961	2.1596 – 3.8264

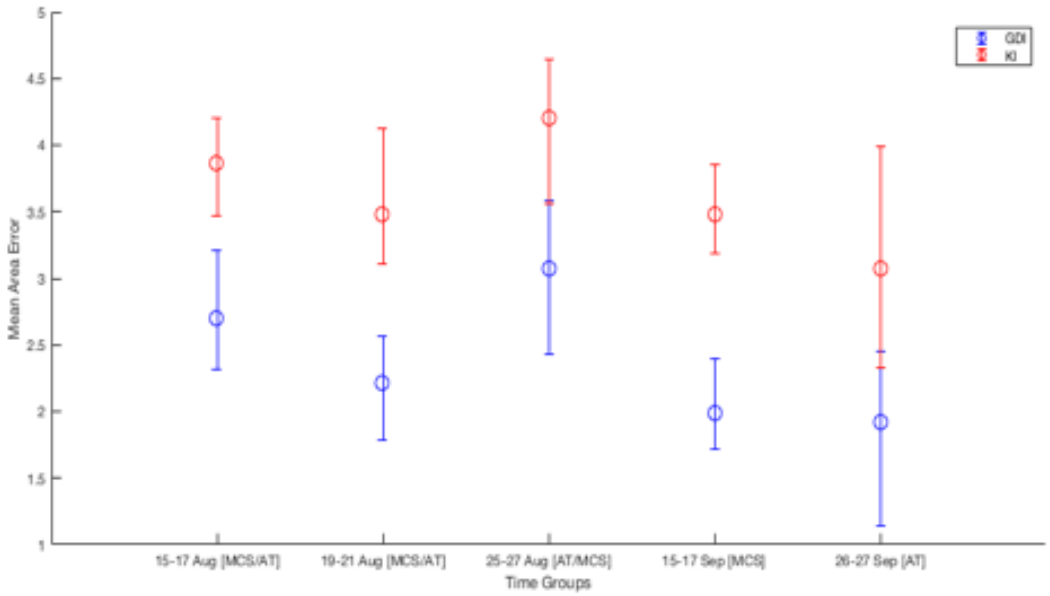

Figure 4.9: Intra-seasonal study area error grouped by days with predominant convective synoptic situations indicated, MCS for mesoscale convective systems and AT for airmass thunderstorms.

When comparing GDI and KI area error values from the intra-seasonal study, GDI is more accurate than KI at portraying the spatial coverage of lightning in each synoptic situation, but not with 95% confidence. For the date ranges where the predominant convective synoptic situation is MCSs, the GDI error values are significantly lower than KI values, with no overlap of the error bars. However, no such conclusion can be made with the same high level of confidence when convection is present predominantly as airmass thunderstorms. During these date ranges (25-27 August 2017 and 26-27 September of 2017) the error bars indicating the 95% confidence interval of the two indices overlap. With 90% confidence GDI has lower area error than KI when airmass thunderstorms are the prevailing convective situation (Figure 4.10). An interesting distinction is made between the two convective settings in regard to areal coverage of lightning. When MCSs are predominant, GDI more accurately portrays how spread out the areas of convection will be with 95% confidence. With mostly airmass thunderstorms present, GDI has lower area error than KI, however only with 90% confidence. KI forecasts are consistently more spread out and cover more areas with high potential for convection when compared to GDI forecasts. GDI's area error values are consistently less than KI's when MCSs are predominant. However, area error values are more similar when airmass thunderstorms are predominant.

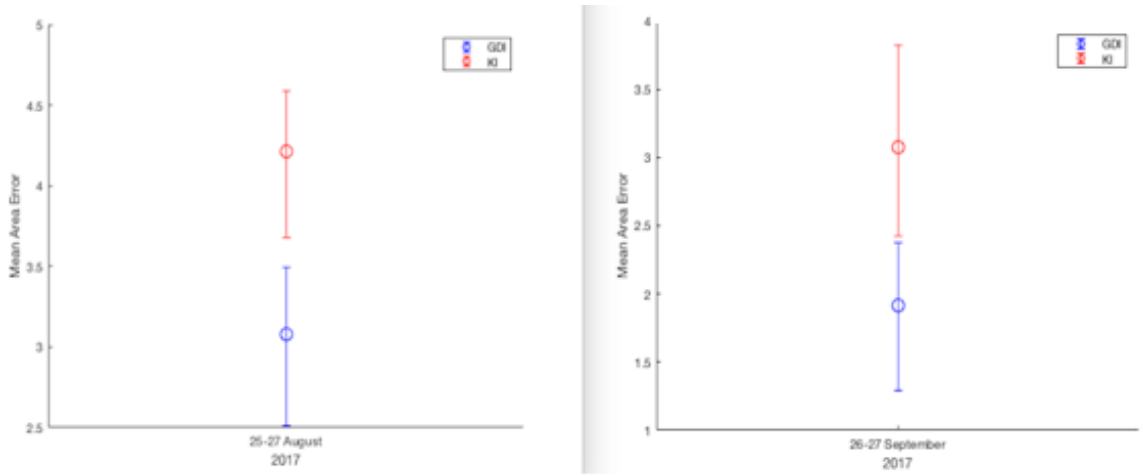

Figure 4.10: Airmass thunderstorms area error values with 90% confidence intervals.

Potential GDI-As Study

Since GDI and KI have comparable location error values across all seasons and convective synoptic situations, the next step is to find an index that would further reduce location error values. Several modifications are made to GDI (GDI-As) which are guided by past research. Attempts at improving location error through GDI-As fall into three categories: average vertical velocity (AveVV), relative humidity (RH) and equivalent potential temperature proxies (EPTP).

GDI-A: Average Vertical Velocity (AveVV)

One attempt to modify GDI into a new GDI-A is to include the average vertical velocity throughout the atmospheric column as another factor contributing to the resulting index value. Vertical velocity values are averaged through the column from 1000-200 mb every 50 mb (1000 mb, 950 mb, 900 mb, etc.). Rising air is essential for convection, because updrafts fuel thunderstorms. No singular vertical velocity value at a pressure level shows any promise for identifying updrafts, so a column-averaged vertical velocity value is used. To locate these updrafts and incorporate them into a convective

forecasting index, vertical velocities are calculated, averaged throughout the column and integrated into a GDI-A calculation.

Vertical velocities in pressure coordinates, ω $\left[\frac{Pa}{s}\right]$, are native to the GFS reanalysis data files. These values are converted into height coordinates, w $\left[\frac{m}{s}\right]$. The conversion requires pressure values and partial pressure of water vapor (Pv) at the various levels (Barani Design 2012). To calculate the Pv, the relative humidity is multiplied by the saturation vapor pressure (e_s; Equation 3.3). Once the Pv values are obtained, the air density (ρ) at each level is calculated:

$$\rho = \left(\frac{100 \ * X\ mb}{Rd \ * T}\right) * \left(1 - \frac{0.378 * Pv}{100 * X\ mb}\right). \tag{4.1}$$

In Equation 4.1, ρ is the air density, $X\ mb$ is the pressure in mb at various levels (1000 mb, 950 mb, etc.), $Rd = 287.05$ $\left[\frac{J}{kg*K}\right]$ the gas constant for dry air, and T is the temperature $[K]$. After calculating the air density, the final part of the conversion is conducted:

$$w = \frac{\omega}{-\rho * g}. \tag{4.2}$$

In Equation 4.2, w $\left[\frac{m}{s}\right]$ is the vertical velocity, ω $\left[\frac{Pa}{s}\right]$ is the vertical velocity, ρ is the air density and $g = -9.8$ $\left[\frac{m}{s^2}\right]$ (Holton and Hakim 2013).

Vertical velocities are averaged throughout the column ($AveVV$) and added to the GDI Equation 2.16 to create the new GDI-A equation (Figure 4.11):

$$GDIA_{aveVV} = ECI + MWI + II + Co + AveVV. \tag{4.3}$$

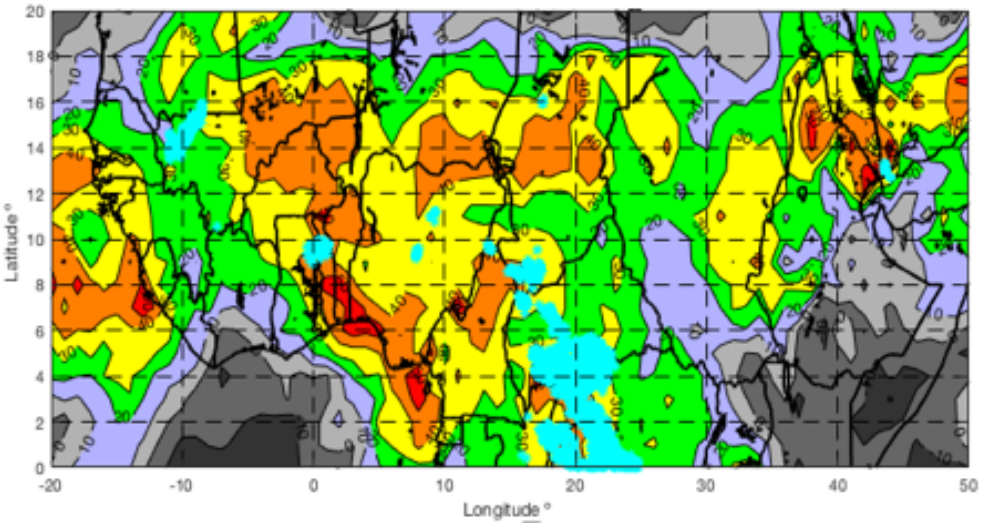

Figure 4.11: GDI-A_{aveVV} from 15 August 2016 at 06Z with lightning (cyan asterisks).

GDI-A: Relative Humidity (RH)

GDI is also modified to incorporate relative humidity values at certain pressure levels. This idea came from KI's calculation (Equation 2.1) that includes the 700 mb dewpoint depression value. Since this index is typically applied to tropical locations and the dewpoint depression value is its unique attribute, a similar concept is applied to modifying the GDI.

The levels chosen for these relative humidity modifications are not included individually in the EPTPs. 700 mb is chosen because it is used in the KI calculation. Two other layers are chosen, 850 mb and 300 mb, to test incorporating relative humidity values above and below 700 mb. Relative humidity values are added to the original GDI Equation 2.16 for these relative humidity GDI-As:

$$GDIA_{RH} = ECI + MWI + II + Co + RHXXX. \quad (4.4)$$

In Equation 4.4, $RHXXX$ is the relative humidity values at the desired levels, 850 mb, 700 mb or 300 mb (added separately into distinct GDI-As; Figures 4.12-4.14).

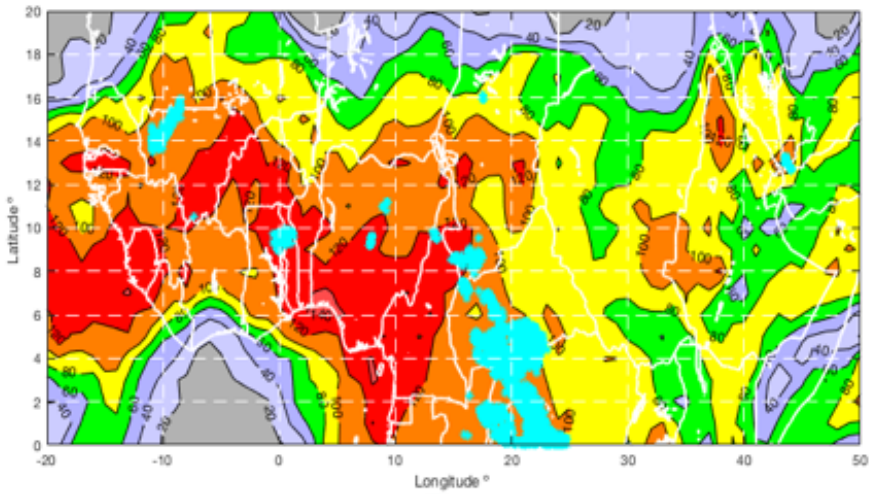

Figure 4.12: GDI-A_{RH850} from 15 August 2016 at 06Z with lightning (cyan asterisks).

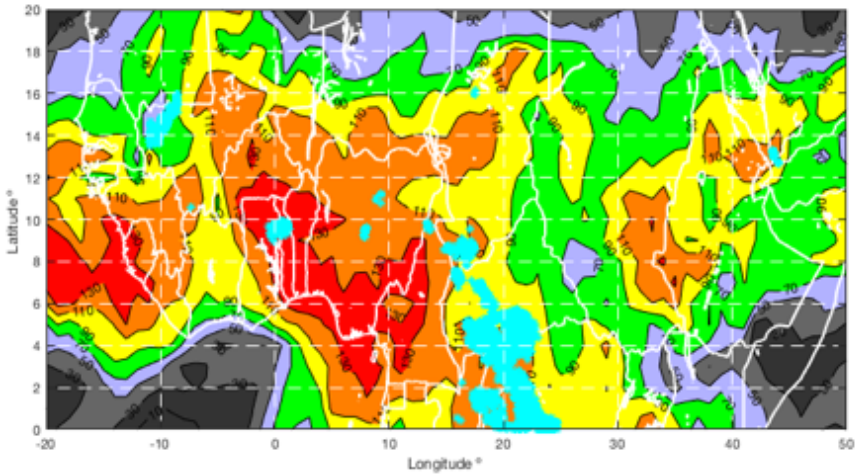

Figure 4.13: GDI-A_{RH700} from 15 August 2016 at 06Z with lightning (cyan asterisks).

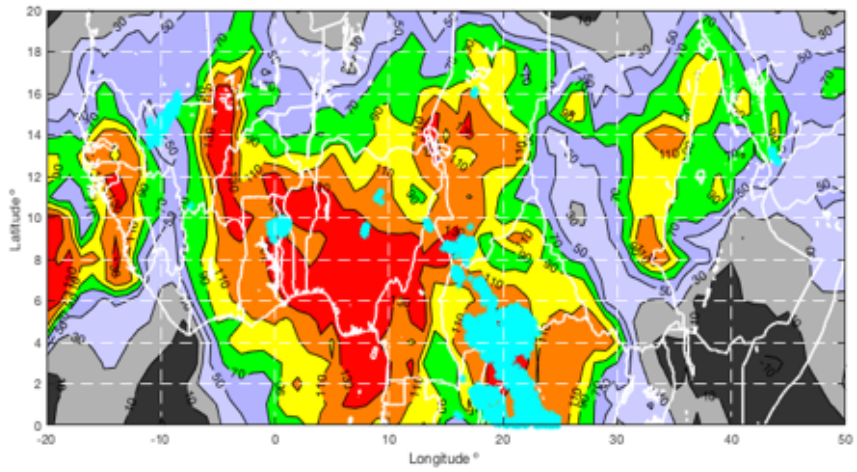

Figure 4.14: GDI-A_{RH300} from 15 August 2016 at 06Z with lightning (cyan asterisks).

GDI-A: Equivalent Potential Temperature Proxies (EPTPs)

The third category of GDI modifications involves changing the level at which the highest EPTP is calculated (Equations 2.5-2.7). For these GDI-A_{EPTP} modifications, the level at which the $EPTP_C$ is calculated is changed from 500 mb to 900 mb, 850 mb, 800 mb, 700 mb and 600 mb. These levels are chosen because the dynamics captured at 500 mb in the higher latitudes of the Caribbean Sea are captured at lower levels in the lower latitudes of northern Africa. No additional factors are added for these GDI-As, just the change in where the $EPTP_C$ is calculated:

$$GDIA_{EPTPXXX} = ECI\ (EPTPXXX) + \ MWI + II + Co. \tag{4.5}$$

Changing where the $EPTP_C$ is calculated alters the values of the resulting index (Figures 4.15-4.19).

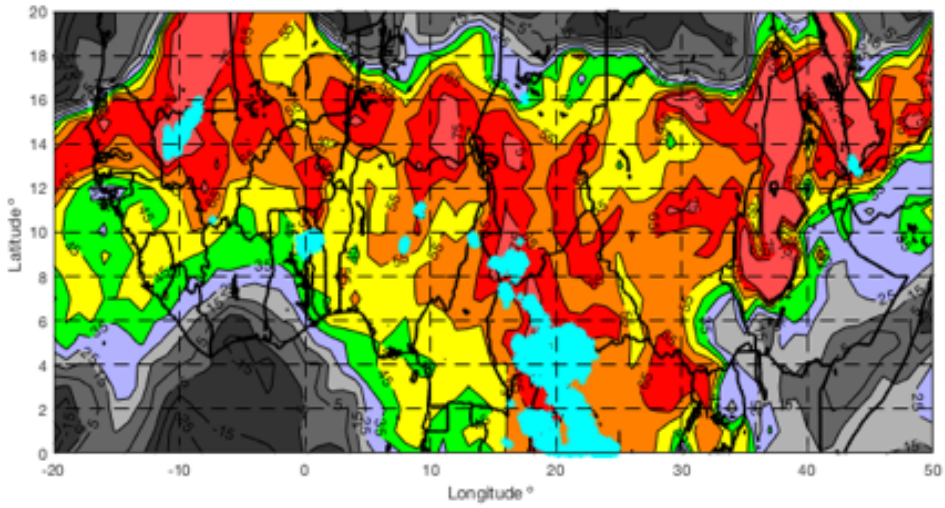

Figure 4.15: GDI-$A_{EPTP900}$ from 15 August 2016 at 06Z with lightning (cyan asterisks).

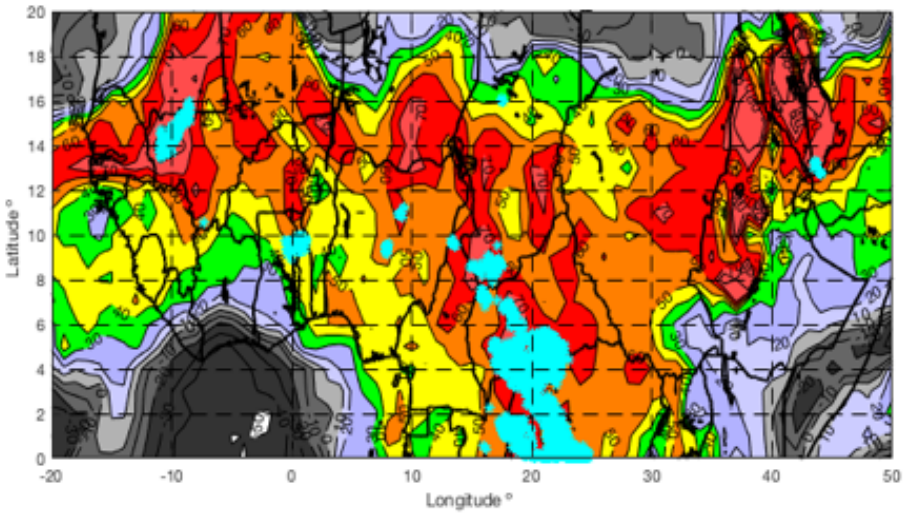

Figure 4.16: GDI-$A_{EPTP850}$ from 15 August 2016 at 06Z with lightning (cyan asterisks).

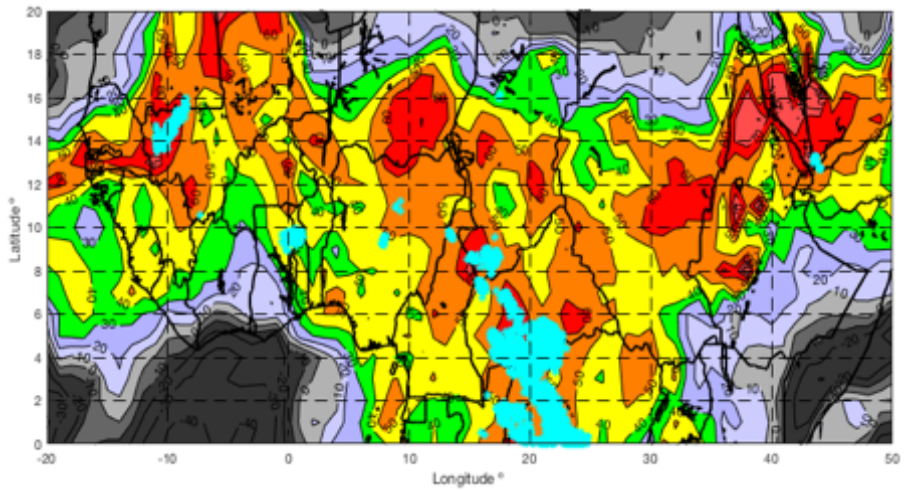

Figure 4.17: GDI-$A_{EPTP800}$ from 15 August 2016 at 06Z with lightning (cyan asterisks).

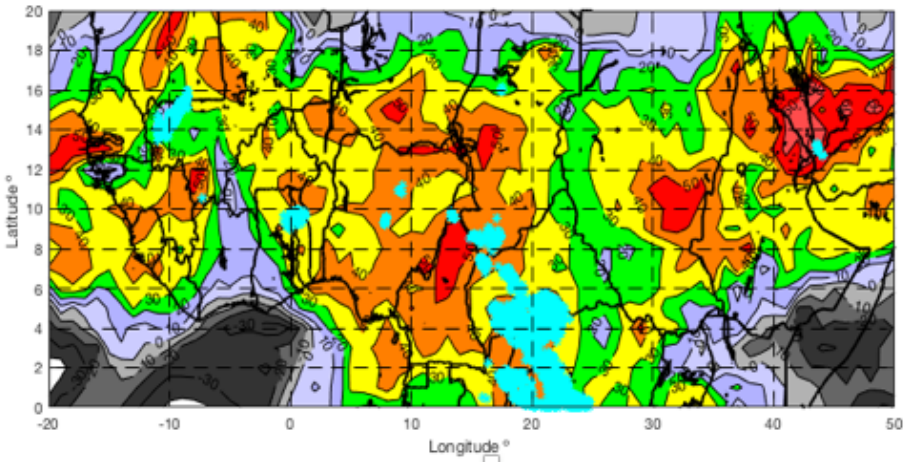

Figure 4.18: GDI-$A_{EPTP700}$ from 15 August 2016 at 06Z with lightning (cyan asterisks).

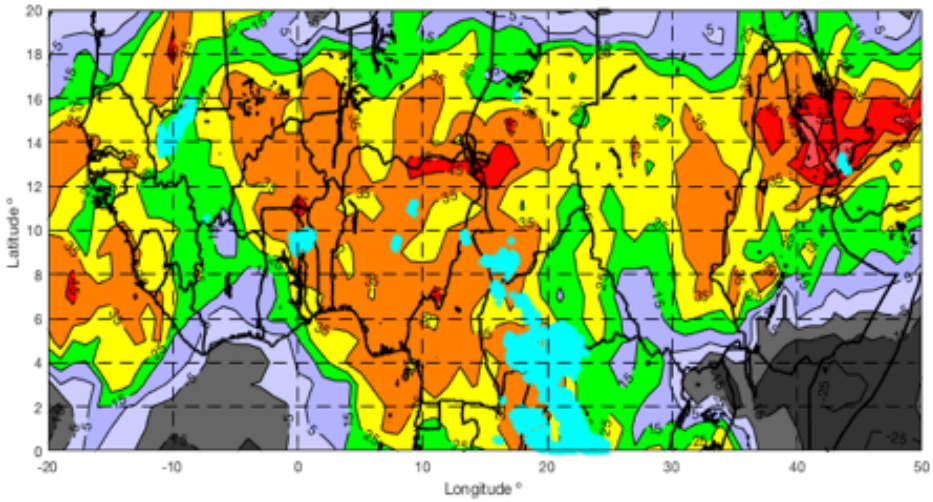

Figure 4.19: GDI-$A_{EPTP600}$ from 15 August 2016 at 06Z with lightning (cyan asterisks).

Potential GDI-As: Location Error

Error analysis is conducted on each GDI-A utilizing the same number of clusters based on the lightning data for k-means clustering. Then each error data set is bootstrapped and confidence intervals are calculated. This date range of 15-17 August of 2016 is chosen because it is the most active thunderstorm season for northern Africa. GDI values (blue) with black horizontal lines indicate the upper and lower bounds of the GDI 95% confidence interval for comparison (Figure 4.20). KI values (red) and GDI-A values (black) are indicated as well. The difference between the EPTP600 35 and EPTP600 40 is the scattered thunderstorm threshold. Since the index values are altered with each modification, 35 no longer means scattered thunderstorms for every GDI-A tested. EPTP600 35 considers 35 to be scattered thunderstorms, while EPTP600 40 considers 40 to be the threshold. The result is a slight change in location error between the two GDI-As, and EPTP600 40 indicates significantly high error values than GDI.

Changing the scattered thunderstorm threshold adjusts the location error values from similar GDI-As, and dramatically affect the GDI-As significance when compared to GDI and KI.

Overall, most of the GDI-A options tested have similar location error values as GDI and KI. Almost all the GDI-A confidence intervals fall into the GDI 95% confidence interval. Only the EPTP600 40 values are higher, meaning this GDI-A has significantly higher location error than GDI and KI. GDI-As that consider factors in the lower levels, 900-700 mb, seem to have lower mean location error than GDI, but remain within the bounds of the GDI 95% confidence interval. This proves most of these GDI-As would forecast the location of convection with the same skill as GDI and KI.

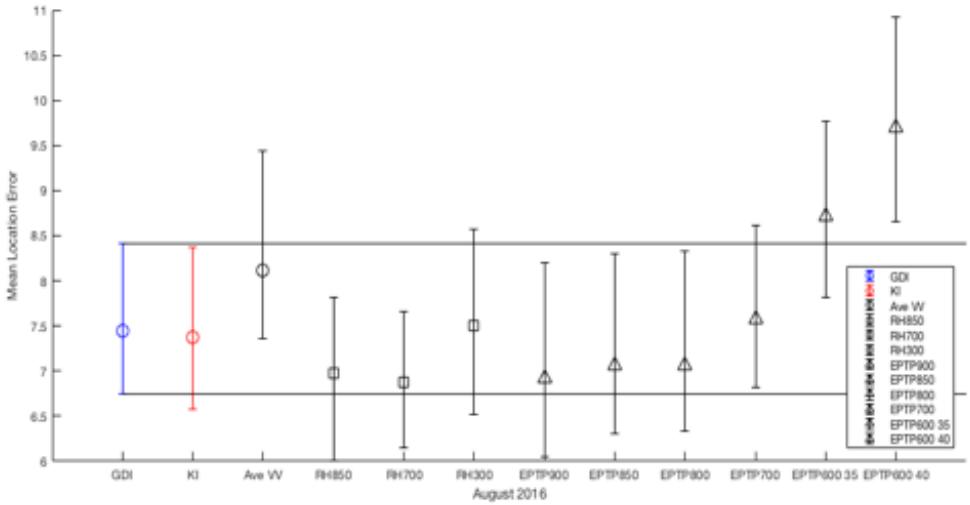

Figure 4.20: Location error values from GDI, KI and GDI-As for August 2016.

Potential GDI-As: Area Error

Area error values are also collected from the error analysis on all the GDI-As (Figure 4.21). Almost all the GDI-As have significantly lower area error values than KI, as with GDI. When considering the difference in scattered thunderstorm threshold, the

EPTP600 35 has higher area error values than the EPTP600 40, the opposite of location error. Higher scattered thunderstorm threshold values will decrease the spatial extent of the convective forecast. The entire EPTP600 40 confidence interval lies within the EPTP600 35 confidence interval, meaning increasing the threshold narrows the possible mean error values, but does not decrease error values in this case. Since increasing the threshold significantly increases the location error, this could mean the scattered thunderstorm threshold has more of an effect on location error than area error. This is an important conclusion when considering methods for reducing location error.

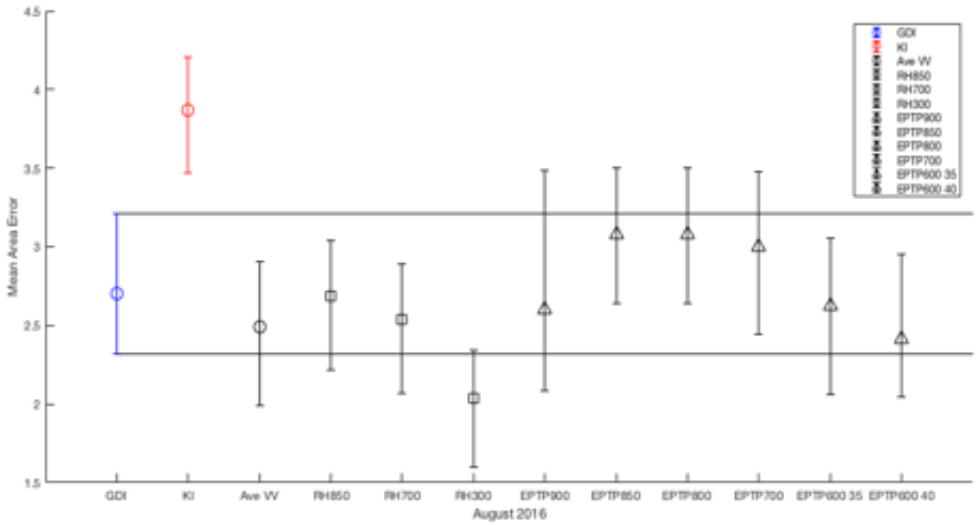

Figure 4.21: Area error values from GDI, KI and GDI-As for August 2016.

All the GDI-A confidence intervals overlap with the GDI 95% confidence interval. For area error, the GDI-As that include RH and EPTP factors from higher levels in the atmosphere prove to be the most promising. The EPTP600s confidence intervals indicate the lowest error values out of all the EPTP GDI-As. The most promising prospective GDI-A to significantly reduce area error is RH300 with its 95% confidence interval barely overlapping with GDI's. The results from RH300 indicate an increase in

vertical resolution of the data has potential to decrease GDI area error values significantly.

V. Conclusions and Recommendations

Chapter Overview

The purpose of this chapter is to state the conclusions of this research and recommend further research to improve forecasting of convection in northern Africa. Conclusions are made from the analysis and results in Chapter IV above.

Conclusions of Research

The guiding question of this research is how well GDI forecasts depict convection over northern Africa when compared to the only applicable existing convective index: KI. Two types of error are measured: (i) location error, which measures the distance between cluster centers of observed and forecasted convection, and (ii) area error, which measures the difference in spatial coverage between observed and forecasted convection. Overall, GDI and KI consistently has similar location error values. However, GDI has significantly lower area error values than KI in almost all cases, except when convection is mostly airmass thunderstorms in the intra-seasonal study.

Intra-Annual Study Conclusions

In the intra-annual study, the forecast skill of the indices is tested throughout each season in one year. GDI and KI have similar skill when forecasting the location of where lightning will occur. Location error values from both indices are lowest in the summer and highest in the winter. However, GDI shows significantly lower area error values than KI in the intra-annual cases. An opposite trend from location error is found in that GDI area error values are the lowest in the winter and highest in the summer. GDI consistently

depicts the spatial coverage of convection more accurately than KI. This difference in error values is most drastic in the winter and least in the spring.

Intra-Seasonal Study Conclusions

The intra-seasonal study examines different date ranges across the late summer and early fall to assess the indices' forecast skill in different convective synoptic situations. In regards to location error, both indices perform similarly when locating convection. Regardless of synoptic situation, neither index is significantly more accurate than the other. GDI outperforms KI in area error when the predominant convective synoptic situation is MCSs. When convection is mostly found in large systems, the GDI depicts the spatial coverage of lightning more accurately than KI. However, no such conclusion can be made from this study for days when convection is predominantly found in the form of airmass thunderstorms. This is possibly due to the inability of the coarse horizontal resolution of the GFS reanalysis data to resolve airmass thunderstorms.

Potential GDI-A Study Conclusions

Further conclusions on the impact of the scattered thunderstorm threshold can be made from the potential GDI-A study. An adjustment in the threshold changed both the location and area error values. Area error values experience minor changes with an adjustment in scattered thunderstorm threshold. However, altering the threshold has greater impacts on location error values. Since the change in error values was much greater for location error than area error, the scattered thunderstorm threshold has a greater impact on location error than area error. Scattered thunderstorm thresholds could be adjusted to find the minimum location error values for any particular index.

The primary goal of testing GDI-As is to identify modifications that could produce significantly lower on GDI error values. All GDI-As indicate similar location error values as GDI and KI (Figure 4.20). All the GDI-A confidence intervals exist within the GDI and KI confidence intervals. For area error, most of the GDI-A confidence intervals are within the bounds of the GDI confidence interval, if not higher in error values. The GDI-A that shows potential in lowering GDI's area error values is the RH300, which incorporates upper-level relative humidity values into its calculation. With improved vertical resolution, the small overlap in the confidence intervals may disappear, indicating a significant reduction of area error.

Significance of Research

GDI forecasts the location of lightning similar to KI, and GDI more accurately portrays the spatial expanse of convection than KI. Since KI often paints most of the north African region with high potential for convection in the summer months, GDI is a higher quality convective forecast than KI. GDI narrows down and highlights the areal coverage of convection and does not sacrifice the accuracy in locating lightning. When compared to KI forecasts, which depict high potential for convection across the continent, GDI is more accurate at portraying the spatial coverage of lightning. Instead of predicting much of the continent to be covered in thunderstorms all summer, like KI forecasts, GDI reduces the spatial coverage of high convective potential areas with the same accuracy in locating lightning occurrences. In the end, this greatly aids forecasters in identifying and forecasting the environmental conditions in northern Africa.

Recommendations for Action

Based on the results from this research, forecasters should utilize the GDI when predicting convection in northern Africa. By using the GDI, they can forecast the location and spatial coverage of convection and more accurately predict environmental conditions. This index should be used in combination with other forecasting tools—such as real-time satellite imagery, other model data and even the KI—to estimate the nature of convection in the future. Since GDI's skill decreases when airmass thunderstorms are the predominant convective synoptic situation, it should be used in conjunction with other forecasting tools. The GDI forecasts are currently depicted over Africa on the NOAA website link where the Gálvez and Davison article is found in the reference section below. Users may access the data, which forecasts 168 hours past the model run time.

Recommendations for Future Research

This project reveals that GDI is useful when forecasting convection over northern Africa, especially for areal coverage, but also can be investigated for improving location error. The focus of this future research section is on possible leads in the GDI-As and which modifications to the GDI are most promising.

One category of GDI-As that shows promise is the relative humidity modifications, specifically the inclusion of upper-level relative humidity. There is minimal overlap of the GDI and RH300 confidence intervals for area error. With enhanced vertical resolution, results could show inclusion of upper-level relative humidity significantly lowers GDI area error. The reduction in area error comes at no cost of location error; this modification leaves GDI's ability to locate lightning in-tact.

Further research is needed to pinpoint which level in the atmosphere gives an optimal GDI-A, since this project only considers 850 mb, 700 mb and 300 mb. Identifying an optimal relative humidity level could significantly reduce area error and perhaps the location error as well.

Another category of GDI-As with a promising lead was the EPTPs. Low and mid-levels are considered, 900 mb, 850 mb, 800 mb, 700 mb and 600 mb, but higher levels (e.g. less than 600 mb) should also be tested. A trend of decreasing area errors is indicated as EPTP modifications increase in height; however, these also result in worsening location error values. When considering EPTP changes to the GDI, balancing the location and area error values must be the guiding method to deciding whether this modification is beneficial.

A third avenue of further research is adjusting the scattered thunderstorm threshold. This threshold impacts the location error values while maintaining similar area error values. Different thresholds should be tested to determine whether a GDI-A could produce significantly lower location error. Another possible area of research is adjusting the scattered thunderstorm threshold on the existing indices, both GDI and KI, to find the optimal threshold for Africa. This could possibly influence the location error values for each index.

Moreover, as mentioned in Chapter I, the samples in the data sets in this research are six hours apart. In order to achieve greater independence amongst the sample error, analysis of data samples that are 24 hours or more apart could be used to future research (Ong 2014).

Finally, in testing the forecast skill of GDI against KI using a higher resolution model could be included in further research. With higher horizontal resolution, GDI could highlight airmass thunderstorms previously overlooked in GFS reanalysis data. Using a higher horizontal resolution model output could show increased promise in GDI's ability in lowering area error values.

Summary

The GDI, created for the Caribbean and Central America, has applicability to northern Africa. GDI forecasts depict convection in this region more accurately than KI forecasts. Forecasters can place confidence in the GDI forecasts when predicting thunderstorms in northern Africa. GDI can be utilized, along with other forecasting tools, to piece together an accurate depiction of environmental conditions, local weather patterns and, thus, the earth's climate system. Although challenges in forecasting African weather remain, employing the GDI will enhance environmental situational awareness. Further research on African thunderstorms is needed and an index specifically tailored for Africa may prove even more beneficial.

Appendix A: Intra-Annual Error Values

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
15-Feb				
00Z	16.68413	-1.74096	22.689166	-5.0627
06Z	11.398325	-0.722825	17.6698	-3.247675
12Z	10.681075	-1.077275	14.604225	-3.277175
18Z	10.03716	-0.98667	12.864166	-3.28753

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
16-Feb				
00Z	16.951925	-1.0761	17.311675	-3.2622
06Z	3.6673	-0.9466	13.3121	-3.079875
12Z	8.275575	-0.827	14.2167	-4.0301
18Z	10.2467667	-0.5418667	5.729533	-4.184966

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
17-Feb				
00Z	13.743975	-1.04755	8.2334	-3.092525
06Z	21.2962	-1.84403	4.891933	-3.14126
12Z	15.30304	-0.41014	8.12988	-3.19718
18Z	6.86363	-0.775266	5.111767	-4.1327

Monthly Ave	12.09575848	-0.999690225	12.06369542	-3.5829905

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
15-May				
00Z	8.834375	-3.451425	6.100425	-4.111025
06Z	10.0927	-2.27368	8.71378	-3.41764
12Z	9.8750667	-2.4033	10.78861667	-3.4467167
18Z	11.52615	-2.1844	9.106825	-3.217575

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error

16-May				
00Z	8.08885	-2.3141	6.9928	-3.520933
06Z	15.739833	-5.1007667	9.7792667	-5.669667
12Z	10.6567	-2.388	8.930316667	-3.64968333
18Z	6.09478	-2.3274	6.4844	-3.3495

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
17-May				
00Z	9.90905	-2.8715	5.16255	-3.7142
06Z	7.9955	-2.91375	5.05775	-3.843475
12Z	6.4434	-2.306525	9.187475	-3.61215
18Z	3.882575	-1.268375	6.720175	-2.8431

Monthly Ave	9.094914975	-2.650268475	7.75203167	-3.699638753

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
15-Aug				
00Z	10.89016	-2.72186	10.6562	-3.96676
06Z	7.46786	-3.49776	5.68908	-4.48582
12Z	7.22178	-2.27686	7.28274	-3.96508
18Z	6.32755	-1.60928	5.9715	-2.98031

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
16-Aug				
00Z	8.05258	-2.3992	7.5083	-3.459967
06Z	6.555	-4.1313	7.0744	-4.906875
12Z	6.78771	-2.12295	8.4034	-3.27648
18Z	8.64083	-3.05825	9.54555	-4.12745

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
17-Aug				
00Z	8.50773	-2.35025	7.7113	-3.54655
06Z	6.379925	-4.0407	4.99435	-5.114325

12Z	4.87143	-2.2058	5.82258	-3.32583
18Z	7.65712	-2.06632	7.87024	-3.27482

Monthly Ave	7.446639167	-2.706710833	7.37747	-3.869188917

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
15-Nov				
00Z	10.707933	-0.7083	12.12355	-3.10611
06Z	11.31132	-2.37938	9.92518	-3.1267
12Z	4.798825	-1.45645	6.01165	-2.447475
18Z	8.16555	-1.76235	6.720625	-3.344825

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
16-Nov				
00Z	11.24144	-2.0219	11.91074	-3.00874
06Z	7.14596	-1.23978	7.87892	-2.41312
12Z	9.712725	-2.931175	10.2542	-3.963975
18Z	7.4169333	-2.647667	7.343933	-5.4537667

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
17-Nov				
00Z	6.251675	-2.011925	6.990275	-4.2147
06Z	5.9892	-3.1132	6.5313	-5.2821667
12Z	4.1946	-1.3681	5.327675	-3.966325
18Z	9.231875	-0.694525	10.63465	-3.3114

Monthly Ave	8.014003025	-1.861229333	8.471058167	-3.636608617

Appendix B: Intra-Seasonal Error Values

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
15-Aug				
00Z	10.89016	-2.72186	10.6562	-3.96676
06Z	7.46786	-3.49776	5.68908	-4.48582
12Z	7.22178	-2.27686	7.28274	-3.96508
18Z	6.32755	-1.60928	5.9715	-2.98031

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
16-Aug				
00Z	8.05258	-2.3992	7.5083	-3.459967
06Z	6.555	-4.1313	7.0744	-4.906875
12Z	6.78771	-2.12295	8.4034	-3.27648
18Z	8.64083	-3.05825	9.54555	-4.12745

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
17-Aug				
00Z	8.50773	-2.35025	7.7113	-3.54655
06Z	6.379925	-4.0407	4.99435	-5.114325
12Z	4.87143	-2.2058	5.82258	-3.32583
18Z	7.65712	-2.06632	7.87024	-3.27482

Monthly Ave	7.446639167	-2.706710833	7.37747	-3.869188917

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
19-Aug				
00Z	6.12675	-3.23305	6.18105	-4.53445
06Z	10.41225	-2.0515	9.24575	-4.079025
12Z	14.12338	-2.52144	11.92456	-3.85724
18Z	5.8485	-1.8663	5.9282	-3.3032

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error

20-Aug				
00Z	7.7118	-2.593175	7.2431	-4.1324
06Z	11.66976	-2.18808	8.17654	-3.65398
12Z	14.236375	-1.642875	13.0058	-1.339525
18Z	12.1234	-2.20358	9.036	-3.958989

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
21-Aug				
00Z	8.31698	-1.8398	4.9601	-3.66856
06Z	8.440075	-3.62572	6.84955	-4.0122
12Z	5.0617	-1.779875	4.96075	-2.50495
18Z	10.3671	-1.061125	6.414325	-2.787675

Monthly Ave	9.536505833	-2.21721	7.82714375	-3.486016167

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
19-Aug				
00Z	8.4213667	-2.5625667	7.6030333	-3.5450667
06Z	8.8192333	-4.74530	9.457	-5.7624
12Z	7.39665	-2.824225	7.448725	-4.33495
18Z	11.436525	-1.963775	12.46775	-3.1333

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
20-Aug				
00Z	4.4409	-3.1285	4.8486	-4.208075
06Z	14.459667	-5.2213	15.4282	-6.2944667
12Z	5.82942	-3.5486	5.53806	-4.45826
18Z	9.188833	-1.63445	5.3298667	-2.98831667

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
21-Aug				
00Z	9.15572	-2.76912	6.02964	-3.70116
06Z	9.97908	-2.38362	9.52982	-3.6669

12Z	8.0648	-3.555433	7.3907667	-4.4449
18Z	7.41685	-2.5416	7.8181	-3.9465

Monthly Ave	8.717420417	-3.073207475	8.240796808	-4.207024589

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
15-Sep				
00Z	4.533475	-2.268725	6.566975	-3.3462
06Z	6.79712	-2.88980	10.20418	-4.37076
12Z	9.299675	-2.240825	11.705825	-3.862325
18Z	2.88585	-1.7103	3.571075	-3.55835

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
16-Sep				
00Z	11.41955	-2.4986	5.582975	-3.85445
06Z	13.377533	-0.6109	4.0234	-3.3807333
12Z	9.669225	-1.720825	12.09025	-3.42315
18Z	5.4143	-1.1867	3.52133	-2.420525

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
17-Sep				
00Z	6.580025	-2.462725	9.121325	-4.1335
06Z	5.9563	-2.293525	3.7619	-4.00465
12Z	12.82518	-2.05852	10.13548	-2.99032
18Z	7.30458	-1.92622	8.1644	-2.46308

Monthly Ave	8.005234417	-1.988972083	7.370759583	-3.484003608

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
26-Sep				
00Z	4.332825	-2.758425	3.3867	-3.2332

06Z	14.05305	-3.78008	14.91275	-4.979675
12Z	14.0741667	-1.059225	5.322133	-0.86675
18Z	5.45565	-1.7792	6.3457	-2.570025

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
27-Sep				
00Z	5.902075	-1.707925	7.6313	-3.8174
06Z	7.6262	-2.0818	8.3298	-3.670575
12Z	16.4456	-1.5360667	9.2063667	-3.6042
18Z	11.77142	-0.6653	3.76238	-1.8244

Monthly Ave	9.957623338	-1.921002088	7.362141213	-3.070778125

Appendix C: Potential GDI-A Error Values

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
15-Aug				
00Z	10.89016	-2.72186	10.6562	-3.96676
06Z	7.46786	-3.49776	5.68908	-4.48582
12Z	7.22178	-2.27686	7.28274	-3.96508
18Z	6.32755	-1.60928	5.9715	-2.98031

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
16-Aug				
00Z	8.05258	-2.3992	7.5083	-3.459967
06Z	6.555	-4.1313	7.0744	-4.906875
12Z	6.78771	-2.12295	8.4034	-3.27648
18Z	8.64083	-3.05825	9.54555	-4.12745

	GDI Location Error	GDI Area Error	KI Location Error	KI Area Error
17-Aug				
00Z	8.50773	-2.35025	7.7113	-3.54655
06Z	6.379925	-4.0407	4.99435	-5.114325
12Z	4.87143	-2.2058	5.82258	-3.32583
18Z	7.65712	-2.06632	7.87024	-3.27482

Monthly Ave	7.446639167	-2.706710833	7.37747	-3.869188917

EPTP 900	Location Error	Area Error	EPTP 850	Location Error	Area Error
	8.78908	-3.5812		8.64528	-3.46286
	5.3191	-3.636228		6.4042	-3.5437
	7.28018	-2.86374		7.17472	-2.98556
	5.8474	-1.989		6.3032	-2.08565
900	Location Error	Area Error	850	Location Error	Area Error
	5.63798	-2.52388		6.12185	-2.598767

	6.660325	-0.48815		5.969075	-4.126225
	5.977367	-2.42625		6.01711	-2.45648
	9.813775	-3.260875		9.8536	-3.336275

	6.915650875	-2.596165375		7.061129375	-3.074439625

EPTP 800	Location Error	Area Error	EPTP 700	Location Error	Area Error
	7.66822	-3.45798		10.09388	-3.58045
	6.65866	-3.26712		6.371	-3.56802
	7.6743	-2.86554		8.27412	-2.05014
	6.1989833	-1.991633		6.69455	-2.18201
800	Location Error	Area Error	700	Location Error	Area Error
	6.63681667	-2.3391667		7.52316	-2.417566
	5.786675	-3.918275		6.765925	-4.27785
	6.0244	-2.346		6.1287667	-2.5763
	9.464725	-3.1714775		8.73465	-3.332325
	7.014097496	-2.919649025		7.573256463	-2.998082625

EPTP 600_35	Location Error	Area Error	EPTP 600_40	Location Error	Area Error
	10.93578	-2.29252		11.29442	-2.65206
	6.90276	-3.43622		8.55274	-2.9851
	7.4095	-2.34388		10.02268	-2.70344
	8.79016667	-1.639266		9.2085	-1.1923
	Location Error	Area Error	600	Location Error	Area Error
	7.79565	-2.228667		8.1281667	-2.0182
	10.08668	-3.905025		12.662425	-3.122375
	7.6095	-2.10721		7.41095	-1.7656
	10.16225	-3.039175		10.36375	-2.867625

	8.711535834	-2.623995375		9.705453963	-2.4133375

RH 800	**Location error**	**Area Error**	RH 700	**Location error**	**Area Error**
	10.11674	-3.30856		9.92444	-1.367266
	4.26862	-2.84968		5.97208	-3.11846
	7.39792	-2.7555		7.213426	-2.63278
	6.93046	-1.77158		4.7723833	-2.0375333
800	**Location error**	**Area Error**	**700**	**Location error**	**Area Error**
	8.3069	-2.35285		7.675116667	-2.3519
	6.830175	-4.16435		6.24905	-3.54635
	7.6768	-2.133		7.69071667	-2.08151
	8.00645	-2.5817		7.03435	-2.8906
800	**Location error**	**Area Error**	**700**	**Location error**	**Area Error**
	3.93765	-2.3526		7.47215	-2.179716667
	5.7847	-3.835225		4.73625	-4.054875
	7.1702	-2.38105		6.6486	-2.1949
	7.12354	-1.73988		6.94794	-2.00568
	6.962512917	-2.685497917		6.86137522	-2.538464247

RH 300	**Location error**	**Area Error**
	8.62396	-3.2623
	7.15038	-2.47214
	6.2202	-1.97236
	4.76314	-1.39088
300	**Location error**	**Area Error**
	6.7942333	-1.584933
	6.298375	-2.541825
	10.26023	-1.50791667
	7.776375	-1.4611
300	**Location error**	**Area Error**
	8.48085	-1.81885

	5.0708	-3.177575
	10.8433	-1.5938833
	7.7226	-1.6409
	7.500370275	-2.035388581

Ave VV	Location Error	**Area Error**
	9.9494	-3.52732
	12.76124	-3.0932
	7.03582	-2.04088
	6.08304	-1.64176
Ave VV	Location Error	**Area Error**
	8.15811667	-1.818
	6.483875	-3.479225
	8.1085667	-1.56453
	8.293725	-2.8104
Ave VV	Location Error	**Area Error**
	8.5835333	-2.13625
	7.71885	-3.9123
	6.359833	-2.0238667
	7.80624	-1.81546
	8.111853306	-2.488599308

Bibliography

AFWA, 2012: Weather techniques and procedures lightning detection systems. 1-26pp.

Barani Design, 2012: Air Density Calculator. Accessed 16 October 2017, http://barani.biz/apps/air-density/

Climogram DRZA, 2017. *14 Weather Squadron USAF* https://climate.af.mil/product_locator/?userinput=&product=all&country_cd=NI&latitude=&longitude=&distance=&sortBy=name&ascendingOrDescending=ASC

Davies-Jones, Robert, 2009: On formulas for equivalent potential temperature. *Notes and Correspondence*, **137**, 3137-3148, DOI: 10.1175/2009MWR2774.1

Efron, Bradley, and R. Tibshirani, 1993: *An Intorcution to the Bootstrap.* Chapman and Hall/CRC, 437 pp.

Gálvez, J. M. and M. Davison, 2016: The Gálvez-Davison index for tropical convection. Accessed 15 January 2017, http://www.wpc.ncep.noaa.gov/international/gdi/

Galvin, J. F. P., 2016: *An Introduction to the Meteorology and Climate of the Tropics.* Wiley-Blackwell, 328 pp.

George, J. J., 1960: *Weather forecasting for aeronautics.* New York and London Press, 673 pp.

Holton J. R. and G. J. Hakim, 2013: *An Introduction to Dynamic Meteorology.* Academic Press, 532 pp.

Jirak I. L., W. R. Cotton, and R. L. McAnelly, 2003: Satellite and radar survey of mesoscale convective systems development. *Mon. Wea. Rev.,* **131**, 2428-2449, http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(2003)131%3C2428:SARSOM%3E2.0.CO%3B2

Kirshnamurti, T. N., L. Stefanova and V. Misra, 2013: *Tropical Meteorology: An Introduction.* Springer, 423 pp.

Lutzak, Philip: Hurricane Isabel September 2003. Accessed 28 September, 2017. http://www.philip-lutzak.com/weather/meteo%20241/PROJECT_3%20HURRICANE%20ISABEL%20PAGE%202%20EASTERLY%20WAVE.htm

Marzban, C. and S. Sandgathe, 2005: Cluster analysis for verification of precipitation fields. *Wea. Forecasting*, **21**, 824-830, http://journals.ametsoc.org/doi/abs/10.1175/WAF948.1

Miller, R. G., 1962: *Statistical prediction by discriminant analysis.* American Meteorological Society, 54 pp.

NexSat, 2011: About NexSat. Accessed 07 December 2017, www.nrlmry.navy.mil/nexdat/headliner/About%20the%20NexSat%20web%20page.pdf

NOAA, 2017: The Gálvez-Davison Index. Accessed 02 August 2017, http://www.wpc.ncep.noaa.gov/international/gdi/

NRL, 2017: NexSat NRL/JPSS Next-Generation Weather Satellite Demonstration Project. Accessed 17 August, 2017, https://www.nrlmry.navy.mil/htdocs_dyn_pregen_sat/PUBLIC/nexsat/pages/single/CONUS/focus_regions/AFRICA/Overview/ir_color/meteo8/Thumb/Latest.html

Ong, D. C., 2014: A primer to bootstrapping; and an overview of doBootstrap. *Stanford University Department of Psychology*, https://web.stanford.edu/class/psych252/tutorials/doBootstrapPrimer.pdf

Orloff J. and J. Bloom, 2014: Bootstrap confidence intervals. *MIT OpenCourseWare*, *https://ocw.mit.edu/courses/mathematics/18-05.../MIT18_05S14_Reading24.pdf*

Pape, Dave, 2004: CIA World DatabankII. Accessed 15 July 2017, https://www.evl.uic.edu/pape/data/WDB/

Piana, Mark E.: Hadley Cells. Harvard University, Accessed 28 September, 2017, https://www.seas.harvard.edu/climate/eli/research/equable/hadley.html

Sensirion, 2001: Dew-point calculation. Application note. 1-3pp. irtfweb.ifa.hawaii.edu/~tcs3/tcs3/Misc/Dewpoint_Calculation_Humidity_Sensor_E.pdf

Singh, S. and N. A. Gill, 2013: Analysis and study of k-means clustering algorithm. *International Journal of Engineering Research & Technology*, **2**, 2546-2551, *www.ijert.org/download/4586/analysis-and-study-of-k-means-clustering-algorithm*

Toracinta, E. R., D. J. Cecil, E. J. Zipser, and S. W. Nesbitt, 2001: Radar, passive microwave and lightning characteristics of precipitating systems in the tropics. *Mon. Wea. Rev.*, **130**, 802-824, http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(2002)130%3C0802%3ARPMALC%3E2.0.CO%3B2.

UCAR, 2014: NCEP FNLvGFS. Accessed 15 July 2017, http://rda.ucar.edu/datasets/ds083.2/docs/FNLvGFS.pdf

UCAR, 2017: NCEP FNL Operational Model Global Tropospheric Analysis, description. Accessed 15 July 2017, https://rda.ucar.edu/datasets/ds083.2/

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 676

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
22-03-2018	Master's Thesis	August 2016 – March 2018

TITLE AND SUBTITLE
A NEW ANALYSIS OF THE GÁLVEZ DAVISON INDEX FOR CONVECTIVE FORECASTS IN NORTHERN AFRICA

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Donndelinger, Gabriel D., Captain, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/ENY)
2950 Hobson Way, Building 640
WPAFB OH 45433-8865

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENY-MS-18-M-078

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
21 Operational Weather Squadron
Unit 3082, APO AE 09012
DSN 314-489-2100; Comm +49-0631532100
ATTN: POC

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES
This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
Severe wind gusts and thunderstorms have been difficult to forecast in Africa. Traditional convective forecast tools (e.g. Total Totals Index, Lifted Index, K Index (KI) and Convective Available Potential Energy) do not accurately portray potential for thunderstorms in Africa. This research effort used the Gálvez-Davison Index (GDI), a convective index created for the tropics, and assess its applicability to northern Africa. GDI was produced for the Caribbean and Central America, and utilized temperature, moisture, mid-level stability, dry air entrainment and an elevation factor to calculate convective potential. In this research, GDI and KI were calculated using Global Forecast System (GFS) reanalysis data. K-means clustering was used to conduct an error analysis. These error values were then bootstrapped and confidence intervals were calculated using the bias-corrected and accelerated method.

Results indicated GDI and KI had similar location error in both the intra-annual and intra-seasonal studies. In comparison with KI, GDI had lower area error values in the intra-annual study and in most convective synoptic cases with 95% confidence.

15. SUBJECT TERMS
African thunderstorms, tropical, convective index, Gálvez-Davison Index, K Index, K-means clustering

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Maj H. R. Tseng, AFIT/ENP
U	U	U	UU	90	19b. TELEPHONE NUMBER *(Include area code)* (937) 255-6565, ext 4520 (Hsien-Liang.Tseng@afit.edu)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18