NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

EVALUATION OF NON-CONVECTIVE WIND FORECASTING METHODS IN THE 15TH OPERATIONAL WEATHER SQUADRON AREA OF RESPONSIBILITY

by

Christopher S. Wireman

March 2012

Thesis Advisor: Wendell A. Nuss
Second Reader: Joshua P. Hacker

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE		*Form Approved OMB No. 0704–0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, 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 this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704–0188) Washington DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2012	**3. REPORT TYPE AND DATES COVERED** Master's Thesis
4. TITLE AND SUBTITLE Evaluation of Non-convective Wind Forecasting Methods in the 15th Operational Weather Squadron Area of Responsibility		**5. FUNDING NUMBERS**
6. AUTHOR(S) AND ADDRESS(ES) Christopher S. Wireman		**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A		**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number _____N/A_____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited	**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

Predicting critical wind thresholds for non-convective wind events is a challenge for today's operational forecasters. This study evaluates two different methods to forecasting non-convective wind gusts of ≥ 35 knots at five locations within the 15th Operational Weather Squadron's area of responsibility. In 2001, Olivier Brasseur developed the Wind Gust Estimate (WGE) as a physically based representation of the boundary layer parameters required to produce gusts at the surface. Previous research compared the WGE to the Air Force Weather Agency's non-convective wind gust algorithm. In this research, the WGE is statistically compared to the Rapid Update Cycle's (RUC) wind gust algorithm that is empirically derived to produce wind gusts forecasts in the RUC model. Utilizing a WRF ensemble data set, the statistical results show the RUC performed better overall at three of the five locations when evaluated with the ≥ 35 knot threshold. Case study analysis revealed that the WGE performed best on seven of the ten case studies. A best fit linear regression is applied to both algorithms and the performance is evaluated on ten independent case studies to analyze accuracy improvements and the potential use of such tuning to the algorithms for future applications. The results of this research suggest that integration of both non-convective wind gust forecast methods into operational forecasts at the 15th Operational Weather Squadron could prove valuable with further testing and evaluation against established rules of thumb and other accepted techniques.

14. SUBJECT TERMS Non-convective Wind Forecasting Methods, Wind gusts, Forecasting, Meteorology, Wind Gust Estimate, Rapid Update Cycle, Boundary Layer, Turbulence, Turbulent Kinetic Energy, Weather			**15. NUMBER OF PAGES** 129
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01–280–5500

Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

EVALUATION OF NON-CONVECTIVE WIND FORECASTING METHODS IN THE 15TH OPERATIONAL WEATHER SQUADRON AREA OF RESPONSIBILITY

Christopher S. Wireman
Captain, United States Air Force
B.S., North Carolina State University, 2003

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

**NAVAL POSTGRADUATE SCHOOL
March 2012**

Author: Christopher S. Wireman

Approved by: Wendell A. Nuss
                        Thesis Advisor

Joshua P. Hacker
Second Reader

Wendell A. Nuss
Chair, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Predicting critical wind thresholds for non-convective wind events is a challenge for today's operational forecasters. This study evaluates two different methods to forecasting non-convective wind gusts of ≥35 knots at five locations within the 15th Operational Weather Squadron's area of responsibility. In 2001, Olivier Brasseur developed the Wind Gust Estimate (WGE) as a physically based representation of the boundary layer parameters required to produce gusts at the surface. Previous research compared the WGE to the Air Force Weather Agency's non-convective wind gust algorithm. In this research, the WGE is statistically compared to the Rapid Update Cycle's (RUC) wind gust algorithm that is empirically derived to produce wind gusts forecasts in the RUC model. Utilizing a WRF ensemble data set, the statistical results show the RUC performed better overall at three of the five locations when evaluated with the ≥35 knot threshold. Case study analysis revealed that the WGE performed best on seven of the ten case studies. A best fit linear regression is applied to both algorithms and the performance is evaluated on ten independent case studies to analyze accuracy improvements and the potential use of such tuning to the algorithms for future applications. The results of this research suggest that integration of both non-convective wind gust forecast methods into operational forecasts at the 15th Operational Weather Squadron could prove valuable with further testing and evaluation against established rules of thumb and other accepted techniques.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

AFB—Air Force Base

AFWA—Air Force Weather Agency

AGL—Above Ground Level

AOR—Area of Responsibility

ARB—Air Reserve Base

ARW—Advanced Research WRF

CONUS—Continental United States

CRCM—Canadian Regional Climate Model

CSI—Critical Skill Index

DoD—Department of Defense

ESRL—Earth System Research Laboratory

ETS—Equitable Threat Skill Score

FAR—False Alarm Rate

HR—Hit Rate

IR—Infrared

KADW—Andrews AFB

KBLV—Scott AFB

KCEF—Westover ARB

KLFI—Langley AFB

KOFF—Offutt AFB

KSS—Kuiper Skill Score

MM5—Fifth Generation Mesoscale Model

MRF—Medium Range Forecast Model

MYJ—Mellor-Yamada-Janjić

NCAR—The National Center for Atmospheric Research

NCEP—National Centers for Environmental Prediction

NOAA—National Oceanographic and Atmospheric Administration

NWS—National Weather Service

OWS—Operational Weather Squadron

PBL—Planetary Boundary Layer

POD—Probability of Detection

RCA—Rossby Centre Regional Atmospheric Model

RMSE—Root Mean Square Error

RUC—Rapid Update Cycle

SWW—Strong Wind Warning

TKE—Turbulent Kinetic Energy

WGE—Wind Gust Estimate

WRF—Weather Research and Forecasting Model

YSU—Yonsei University

ACKNOWLEDGMENTS

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OBJECTIVES AND MILITARY SIGNIFICANCE

Non-convective winds are defined as high winds that occur in the absence of thunderstorms, tornadoes or tropical cyclones (Knox et al. 2011a). Non-convective wind events encompass many different types of weather phenomenon. For example, downslope winds, gap winds, dust storms, and other winds associated with extra-tropical cyclones are all events that are classified as non-convective. Winds caused by strong pressure gradients are also non-convective in nature (Ashley and Black 2008). The majority of focus in previous studies of non-convective wind events is placed on extra-tropical cyclones as the leading cause of most high wind events (Niziol and Paone 2000; Knox 2004; Lacke et al. 2007; Ashley and Black 2008; Knox et al. 2011a).

Non-convective wind forecasting has long since plagued even the most experienced forecasters. Unlike the better-known counterpart, convective winds, non-convective winds can occur in rather seemingly good weather days with clear skies and warm temperatures but can produce extensive damage and even loss of life (Kapela et al. 1995; Knox 2004; Lacke et al. 2007; Ashley and Black 2008; Knox et al. 2011a,b). The United States military has a vested interested in receiving timely and accurate watches and warnings of strong winds both associated and not associated with thunderstorms in order to properly protect people and secure valuable assets. High wind gusts can have a wide range of impacts on a military installation from unsecured maintenance equipment and materials on a flight line to personnel performing construction work on rooftops of high buildings such as hangars. Winds also play a dominant role in aviation operations. For example, details such as current runway usage at most airfields across the globe are typically determined by the prevailing wind direction and speed. Additionally, large aircraft may require notice of gusty surface winds before performing low-level missions where aircraft control is the highest priority (LaCroix 2002).

The 15th Operational Weather Squadron at Scott Air Force Base (AFB), Illinois, provides weather support to Air Force, Army, National Guard and other Department of

Defense (DoD) installations in the United States. The 15th OWS area of responsibility (AOR) consists of 150 locations across six distinct regions throughout the eastern half of the United States. From the Northern and Central Plains through the Great Lakes and Kentucky/Tennessee regions into New England and the Mid-Atlantic, the vast AOR provides many different types of weather phenomena and forecast challenges to the weather professionals at the 15th OWS providing timely, accurate and relevant weather products to the installations in the region (see Figure 1).

Figure 1.   Outline of the 15th OWS AOR.

In 2006, the 15th OWS requested research to evaluate the latest forecasting techniques of winds associated with convective activity with the focus placed on locations within their AOR. The result was a detailed analysis (Kuhlman 2006) on the value and accuracy of modeled derived indices such as T1, T2 and WINDEX methods. In 2011, a similar request was drafted by the 15th OWS to evaluate the forecasting of non-convective winds. Table 1 shows the winds in the 15th OWS AOR; this area encompasses 50% of the nation's top ten windiest cities from November through April (Niziol and Paone 2000).

Although various accepted techniques for forecasting non-convective winds are in use, the primary focus of this research is to statistically evaluate the performance of a physically based wind forecast model (the Wind Gust Estimate) at five different locations

within the 15th OWS AOR. The Wind Gust Estimate (WGE) is a relatively new physical approach developed by Brasseur (2001) to estimate wind gusts that incorporates known physical processes in the atmosphere, and in particular, the response of air parcels in the boundary layer to turbulent eddies. Previous studies of this technique indicate positive results (LaCroix 2002, Nordström 2005). An updated analysis using the latest weather prediction models and greater focus on location-based forecasting to identify possible localized performance differences will ideally produce meaningful results valuable to operational forecasters. More background on this will be presented in Chapter II.

1. **BLUE HILL, MA**	**17.0 MPH**
2. CASPER, WY	14.7
3. CHEYENNE, WY	14.5
4. DODGE CITY, KS	14.3
5. GREAT FALLS, MT	14.2
6. **ROCHESTER, MN**	**14.1**
7. AMARILLO, TX	13.8
8. **BOSTON, MA**	**13.5**
9. **NEW YORK, NY (LAGUARDIA)**	**13.5**
10. **BUFFALO, NY**	**13.2**

Table 1.   Top ten windiest U.S. cities during November through April. Bolded cities indicate those within 15th OWS AOR (After Niziol and Paone 2000).

The long-term goal of this work is to drive future research to improve the prediction of non-convective winds by operational forecasters to mitigate the unfavorable impacts to military operations. The short-term goal and focus of this thesis is to provide the 15th OWS a recommendation for the best methods and algorithms to utilize when forecasting non-convective wind events and particularly for winds greater than or equal to 35 kts. Supplementing this main goal are four secondary research goals listed below in order to ensure a thorough and valuable examination of different non-convective wind forecast models.

- Compare the forecast skill and performance of the WGE to an empirical forecast model.

- Examine the performance of each model at each individual location to identify trends.

- Determine if results can be used to tune the forecast algorithms to improve performance.

- Examine the diurnal performance of these models.

B. NON-CONVECTIVE WINDS DEFINITIONS AND THRESHOLDS

For the purposes of this research, non-convective wind events will be defined as wind events not associated with convection (i.e., thunderstorms, outflows, tornadoes, tropical systems). High wind events such as those due to pre and/or post-frontal winds associated with extra-tropical systems and winds due to strong pressure gradients will be included. Due to the inherent complexity of topographic effects, outside of the model's capability to resolve possible flow enhancement due to local terrain, phenomena such as gap winds or other types of events usually found in the Intermountain West will not be examined. Along with selection of the types of events to use in this study, it is equally important to examine the types of thresholds typically used in literature to analyze non-convective wind events.

This research focuses on Air Force Weather warning criteria and specifically the threshold for a Strong Wind Warning (SWW). SWWs are issued for winds not associated with thunderstorms greater than or equal to 35 kts (sustained or gust) (AFMAN 15–129V1 2011). This criterion differs slightly from National Weather Service (NWS) thresholds for high wind events often found as the criteria for which researchers choose to use. Some research uses a threshold of sustained wind of 40 mph (35 kts) for one hour or greater or a peak gust of 58 mph (50 kts; Knox et al. 2011a). Other research utilizes the NWS Central Region's criteria for a high wind advisory of sustained winds at least 30 mph (26 kts) for one hour or greater or gusts of 45 mph (39 kts) or greater (Crupi 2004). While analyzing storm report data for fatality information associated with high wind events, Walker and Black (2008) organized events by terms

such as "high wind," "gusty wind," and other similar basic terms and then further determined if these were associated with convective or non-convective environments.

C. METEOROLOGY ASSOCIATED WITH NON-CONVECTIVE WIND EVENTS

One cause of non-convective wind gusts can be physically explained by higher wind speeds aloft that are transported down to the surface. Other explanations and hypotheses of the causes of strong non-convective winds include topographic effects, strong winds due to the pressure gradient, tropopause folding, and a relatively new study area of "sting jets" (Knox et al. 2011a). Additional hypotheses such as the bent-back warm front are thoroughly summarized in research conducted by Asuma (2010).

Kapela et al. (1995) developed an operational forecast checklist based on 11 key atmospheric ingredients that have been shown to indicate a strong post cold-frontal wind event associated with extra-tropical cyclones in the Northern Plains. The operational checklist consists of the following ingredients: a) pressure gradient diagnosis from modeled output, b) strength and position of the 500 mb vorticity center, c) 3-hourly pressure changes from isallobaric analysis, d) subsidence, e) cold-air advection, f) lapse rate, g) satellite imagery and comma cloud features, h) jet position and strength, i) directional wind shear in the vertical, j) geostrophic wind, k) snow cover and l) cessation of strong winds. If some, or specifically if all, of the ingredients come together then dangerous surface winds could occur (Kapela et al. 1995). An idealized high wind event is shown schematically in Figure 2 detailing the culmination of all of these individual ingredients. Although the research conducted by Kapela et al. (1995) is focused on the Northern Plains, the meteorology presented on winds associated with extra-tropical cyclones can be applied in many locations prone to weather impacts due to passing cyclones (such as the Northeast United States) and was the focal point of the in-depth meteorological analysis conducted by Knox et al. (2011a).
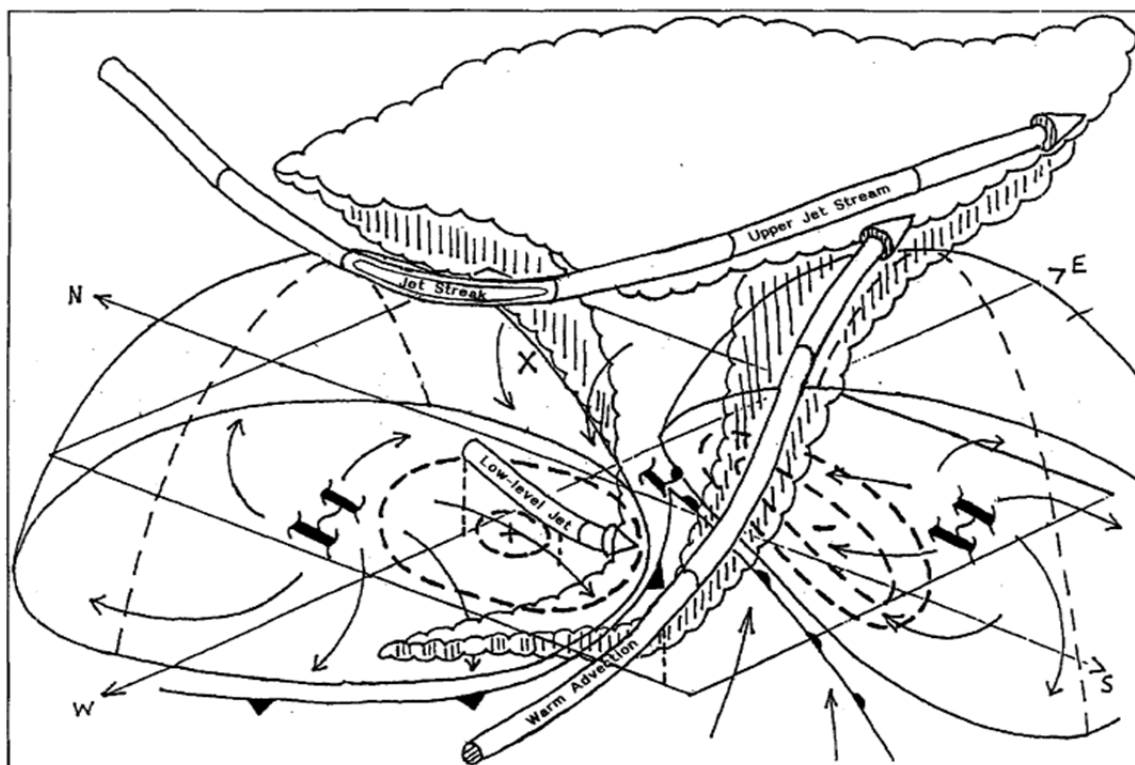
Figure 2. Idealized schematic of the 4 February 1984 strong wind episode in the northern plains. Surface features are moving southeast. Thick dashed lines are isallobars, with pressure rise-fall centers marked by +/- signs. Tubular arrows depict relative flow originating at low and high levels. The X represents a midlevel vorticity maximum. Surface anticyclones have an isentropic surface to represent the domelike structure of the air masses. Scalloped lines show associated clouds (From Kapela et al. 1995).

Kapela et al. (1995) found that locations just to the south or west of a passing vorticity maximum usually created the best scenarios for strong subsidence and a higher potential for momentum transfer to occur. Another subsequent feature associated with this subsidence is the atmospheric response as adiabatic warming occurs, effectively decreasing static stability and further supporting the downward transfer of momentum. It is important to note here that any inversion (including a subsidence inversion) will increase the static stability in the lower levels and decrease the momentum transfer into the boundary layer and likely result in lower surface wind speeds (Kapela et al. 1995).

An anomalous high-wind event over Upper Michigan occurred with the presence of a sharp inversion at 850 mb (Crupi 2004). The ability to predict or accurately

characterize the height of the boundary layer is critical in wind speed prediction (Niziol and Paone 2000). This is especially important when forecasting non-convective winds at night when the nocturnal boundary layer heights vary and are often difficult to predict combined with possible inversions.

Another identifiable feature associated with strong subsidence examined in various studies is that of subsidence associated with the "dry slot" noted in satellite images of mature cyclones often linked to tropopause folds (see Figure 3). This feature is commonly found south of the center of the low pressure system and indentified by its relatively cloud free region. High winds are often associated with this feature due to the large amount of subsidence occurring in this region as higher momentum air aloft is pulled down from the upper atmosphere into the lower atmosphere (Knox et al. 2011b). In weak or no static stability in the lower atmosphere, high speed winds can be mixed down creating gusty winds at the surface (Knox 2004). Utilizing isentropic charts is an integral way to analyze subsidence associated with the dry slot region. Identifying regions of higher momentum air aloft that has a greatest potential for downward transfer is possible when a strong gradient of post-frontal isobars on an isentropic surface is analyzed (Kapela et al. 1995).

Several previous studies identify the isallobaric wind, which occurs when the accelerating geostrophic wind is balanced with Coriolis force, as an important contributor to non-convective winds. The strongest winds will occur in the regions of the strongest pressure gradient (Knox et al. 2011a). The 3-hour pressure tendencies are a quick way to analyze the potential for strong gusty winds. The surface pressure tendency equation says that the pressure tendency is a function of the mass change in the vertical and therefore the vertically integrated temperature (represented as density) advection (Kapela et al. 1995). As such, strong cold air advection in the lower troposphere will be indicated at the surface by strong pressure rises (greater than 3 mb in three hours) which will alert forecasters to the possibility of gusty winds (Kapela et al. 1995 and Niziol and Paone 2000). Strong rise/fall pressure couplets are also a good indicator of high wind events (Niziol and Paone 2000). Niziol and Paone (2000), discuss the occurrences of strong 850 mb winds ahead of cold frontal passages that did not transport gusty winds to the surface

due to a lack of mixing as result of lower lapse rates or temperature inversions. Their analysis revealed that high wind events occurred when surface to 850 mb lapse rates reached 8° Celsius while non-events revealed lapse rates of 3.5° Celsius (Niziol and Paone 2000). Asuma (2010) concluded that strong wind speeds in the boundary layer as well as the near dry adiabatic lapse rates existed in the southwest quadrant of mid-latitude cyclones further developing an environment favorable for mixing gusty winds to the surface. Forecasters must key in on these areas of strong isallobaric pressure gradients and pressure rises as this usually implies a stronger measured wind (Kapela et al. 1995). Hourly pressure changes may be more important than 3-hourly tendencies (Niziol and Paone 2000). Furthermore, this type of analysis must be used with caution since the isallobaric wind may not be the dominant component of the ageostrophic wind and approximations for the isallobaric wind may be misleading (Knox et al. 2011a).
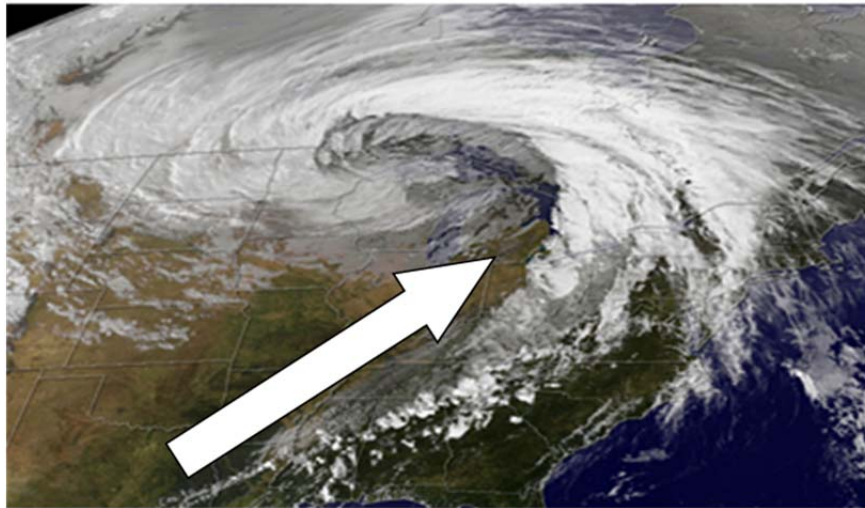
Figure 3. Satellite image (26 Oct 2010) of a mid-latitude extra-tropical cyclone with the dry slot highlighted by the white arrow (After Knox et al. 2011b).

D. CLIMATOLOGY OF NON-CONVECTIVE WIND EVENTS

The common goal among cataloging numerous similar weather events is that they likely reveal signatures that can be used in pattern recognition. This notion was detailed by Knight et al. (2005) as the authors explored common climatological predictors (anomalies) associated with various types of severe weather events. Although no single specific physical process can be identified as the cause of non-convective winds,

climatological analysis does reveal many consistencies of these events throughout the most prone regions of the United States where they occur (Knox et al. 2011a). Most climatology studies on non-convective winds focus on the Great Lakes region and the Northeast, with a limited number of studies conducted in the Midwest and Northern Plains. These events occur at much less frequency in the Southeast when compared to these other regions (Knox 2004).

One of the most cited studies in recent literature was conducted by a pair of meteorologists at the NWS Forecast Office in Buffalo, NY. Niziol and Paone (2000) conducted a climatological study of non-convective wind events in Western New York which has become a focal point for studies in the subsequent recent decade. Their research was developed to help operational forecasters predict high wind events as mid-latitude cyclones affected the Great Lakes region by analyzing synoptic scale weather patterns that typically produced these events. Non-thunderstorm related wind gusts greater than or equal to 50 kts at Buffalo, New York over a twenty-year period was chosen as the criteria for selecting events (Niziol and Paone 2000). This threshold is higher than most discussed previously.

The high wind events at Buffalo were climatologically analyzed and the results revealed most of the high wind events occurred during the cold season from October through April which correlates well to previous research of the frequency distribution of mid-latitude cyclones in the Great Lakes region. In terms of wind direction, the majority of the wind events occurred when the winds were from the southwest to west direction (Niziol and Paone 2000). This is consistent with typical "dry slot" winds as the orientation to the mid-latitude cyclone is such that upper level winds are from the south or southwest (Knox 2004). Surface track analysis of the cyclones was accomplished and revealed a direct correlation to the high wind events and Buffalo's location in the southwest to west quadrant of the low. For New York and much of the Northeastern United States mid-latitude storm tracks occur from southwest to northeast with storms crossing north and west of New York with vertical tilt north and west into the regions of coldest air. Strong east to west isallobaric gradients typically exist along cold frontal boundaries. Surface to 850 mb lapse rates are greater along with strong cold air

advection that exists in deep layers with wind events versus non-events (Niziol and Paone 2000). Results from research conducted in 2004 focused on the Midwest reveal similar conclusions. However, the southwest quadrant bias is dominant in high wind events, with few exceptions, and across a much more widespread region of analysis differing from previous conclusions that the bias was due to the Great Lakes influence and orientation (Knox 2004).

Expanding upon the research in the Great Lakes region, Lacke et al. (2007), created a 44-year climatology of non-convective wind events broken down by two different criteria: A) sustained winds of 40 mph (35 kts) or greater for at least one hour or B) any gust of 58 mph (50 kts) or greater. Thirty-eight observation stations from Minnesota through Ohio to New York were used for this study. Of the over six million observations analyzed, roughly 2,600 satisfied either criterion A or B. Nearly 30% of the observations that satisfied criterion A occurred during the month of March, however the peak (35%) for the wind gust criterion occurred in the month of January (Lacke et al. 2007).

For the majority of the wind events that met either criterion, the locations of the observation stations were primarily along the Great Lakes, with an additional peak in frequency of occurrence in the western part of the Midwest. Furthermore, when analyzing the cases by sea level pressures, the data revealed that non-convective wind events occurred with both high and low pressure systems although the frequency was greater for those associated with low pressure systems. Additionally, the authors examined wind direction preference for non-convective wind events to compare and contrast with results from previous research. The results were clear and strikingly similar to previous research done in both the Midwest and western New York State. Figure 4 shows an overwhelming high frequency (70% for criterion A and 76% for criterion B) of events occurred in the west through southwest cardinal direction (Lacke et al. 2007).

10
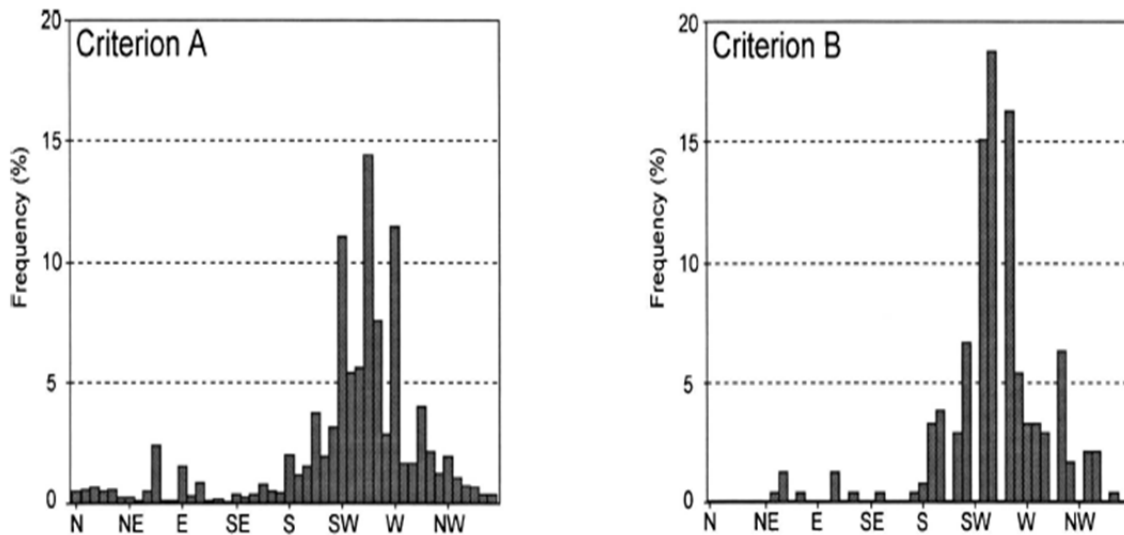
Figure 4.   Frequency of wind observations by wind direction satisfying criteria A) sustained winds of 40 mph (35 kts) or greater for at least one hour or criteria B) any gust of 58 mph (50 kts) or greater (After Lacke et al. 2007).

While results from the Great Lakes climatology study showed that non-convective wind events could occur associated with strong pressure gradients as high pressure builds into the region, the majority of the events occurred with mid-latitude cyclones supporting previous research results. This was an important conclusion when it appeared in 2007 since it was the first study conducted over a wide region to suggest the link between mid-latitude cyclones and non-convective wind events. Additionally, the result of the primary favored wind direction from the west through southwest was also an important discovery as it had only been noted in a small area in literature at that point. Although there were a few outliers in primary wind direction during these events, the authors conclude this is likely due to climatology of these wintertime cyclone tracks (Lacke et al. 2007). This conclusion was further supported by additional research that showed the primary wind direction for non-convective wind events in the Great Plains was from the northwest indicating the importance of the location of the observation station to the cyclones storm track (Knox et al. 2011b).

E. DAMAGE AND FATALITIES CAUSED BY NON-CONVECTIVE WINDS

The military community is not the only customer with a vested interest in accurate prediction of high wind events. The Internet is full of news articles that detail the impacts caused by non-convective wind events. Recent news stories range from relatively minor impacts at major sporting event festivities (Associated Press 2012) to the Notre Dame tragedy in 2010 (Knox et al. 2011b). It is clear that non-convective wind events have a wide range of impacts on the population within the United States.

History shows that many of the great natural disasters across the world were caused by non-convective winds. For example, "The Perfect Storm" of 1991, caused by a strong coastal cyclone that merged with the remnants of a hurricane caused five fatalities and $200 million in damage in eastern North America. In recent history, the October 2010 storm in the upper Midwest of the United States was one of the strongest ever recorded in the Continental United States (CONUS) with a low pressure of 955 mb and peak wind speeds of 78 mph (68 kts) recorded (Knox et al. 2011a). In terms of damage alone, in just a small four year survey from 2000–2004, high winds associated with non-convective events caused more property and crop related damages than did winds produced by thunderstorms or tornadoes (Lacke et al. 2007). Although various research reveals different numbers when it comes to comparing property and crop damage totals between convective and non-convective events, fatality trends associated with these storms remains consistent.

A study conducted by Ashley and Black (2008) provided the most in-depth analysis of fatalities associated with non-convective winds to date. The study analyzed fatalities from storm reports over a 26-year period from 1980–2005. Figure 5 shows that fatality numbers are similar to those associated with convective winds from thunderstorms and are greater than those winds associated with hurricanes and tropical storms (although flooding is usually the leading cause of fatalities in those events). The authors also analyzed fatalities among various regions within the United States. The Northeast represents a large percentage of overall fatalities likely due to its location with respect to strong low pressure systems crossing this region in the cold season months and higher population. 83% of fatalities analyzed during this time period were associated

with wind events caused by mid-latitude cyclones or post-frontal winds after passage of a cyclone. Gradient winds, likely associated with strong high pressure systems, accounted for another 10%. In summary, 93% of fatalities were caused by events that are less likely to draw public urgency from warnings when compared to severe thunderstorm or tornado warnings (Ashley and Black 2008). The lack of perceived danger likely results in people venturing into harm's way more often during these types of events as 91% of fatalities occur in vehicles, boats or outdoors (Knox 2004; Ashley and Black 2008; Knox et al. 2011b). The increased education of the general public and forecasting performance by meteorologists on non-convective wind events will work hand in hand to lower the dramatic impacts on society and provides for additional motivation for this thesis.
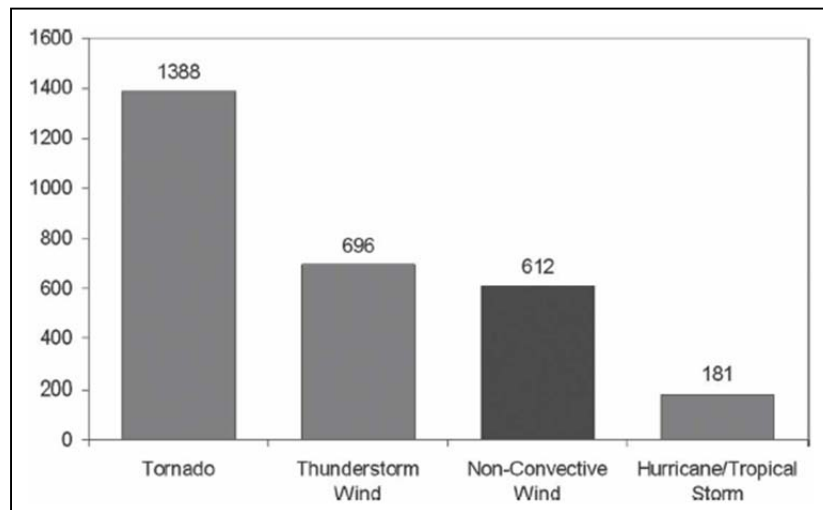
Figure 5.   Fatalities associated with various types of wind events from 1980–2005 (From Ashley and Black 2008).

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A. WGE METHOD

1. Overview

Through the mid-1990s, research on wind gusts and forecasting techniques largely centered around empirical or statistical approaches rather than methods designed around physical explanations of the causes of surface wind gusts. Empirical approaches can characterize the wind gusts with some accuracy, especially for weak to moderate wind gusts, but do not accurately predict the more severe wind gusts that occur (Brasseur 2001, Nordström 2005). Empirical methods are typically the preferred method at the operational forecasting level due to the lesser number of parameters required to run the model and their lower sensitivity to small errors. Empirical algorithms do not explain the physical processes in the atmosphere, which should be a requirement for the development of enhancements for increased model performance and reliability related to wind gusts (Brasseur 2001).

Brasseur (2001) developed an innovative technique for wind gust forecasting based solely on physical processes in the atmosphere (called the Wind Gust Estimate). This method was developed based on Brasseur's (2001) physical explanation of the intricate details of boundary layer interactions that caused wind gusts to occur at the surface. The goal was to not only develop a reliable wind gust forecast method, but in future research that examines the WGE methodology, increase the understanding of the physical processes in the atmosphere responsible for these gusts. In turn, this would enable enhancements to the forecasting method and increase reliability of the model. A key aspect of this method is not only the prediction of the wind gust at the surface, but also the prediction of the lowest and highest estimates of possible wind gusts speeds that encompasses the estimated wind gust with high probability, called the bounding interval (Brasseur 2001).

Before detailing the prediction model itself, it is important to understand the fundamental characterizations of the boundary layer processes of the WGE according to

Brasseur (2001). The most important of these physical explanations is that wind gusts occur due to the deflection of air parcels, flowing at higher speeds aloft, by large turbulent eddies in the boundary layer. It is clear from this explanation that these turbulent eddies must be strong enough to counterbalance the buoyancy forces in the atmosphere. Therefore, the stability of the boundary layer is one critical component of this method (Brasseur 2001). This supports previous research and explanations found in Chapter I that relate atmospheric stability to non-convective wind gust events.   In summary, the three major physical ingredients of the boundary layer that were utilized in the development of the WGE are the wind speeds within the layer, the turbulent eddies and the stability (Brasseur 2001). The accurate prediction of these components is important to the accuracy of the method.

Turbulent kinetic energy (TKE) plays a large role in the WGE equation and is parameterized within numerical models. TKE in the boundary layer is critical in the determination of which parcels will reach the surface. Several methods have been presented for the accurate prediction of TKE within the model. Brasseur (2001) details the standard prognostic TKE equation. This equation states that mean TKE is equal to shear production terms (includes both the x and y components) combined with the buoyancy, the vertical transport of turbulence as well as the dissipation (the latter two typically subtract from the mean TKE value). This equation is utilized in all turbulence parameterizations with 1.5 (or greater) order turbulence closure (Brasseur 2001). However, since TKE was not a standard variable output by the Air Force Weather Agency (AFWA) when analyzing the WGE performance with the Fifth-Generation Mesoscale Model (MM5), LaCroix (2002) developed an alternative method for calculating TKE. This calculation utilizes the Reynolds averaged perturbation velocities of the u, v, and w wind components. This alternative method was not tested in this research since the model data set used from the Advanced Research WRF Version 3.1.1 utilizes the 2.5 order turbulence closure model from the Mellor-Yamada-Janjić (MYJ) PBL scheme and calculates TKE, however this alternative method is an important aspect discussed in Chapter V (Skamarock et al. 2008).

## 2.	Calculation of the WGE

As mentioned previously, the fundamental characteristic of the WGE is that the air parcels are deflected from a specific height in the boundary layer by turbulent eddies that overcome the buoyancy within the layer (see Figure 6). The following equation summarizes these characteristics utilized in the WGE (Brasseur 2001):

$$\frac{1}{z_p}\int_0^{z_p} E(z)\,dz \geq \int_0^{z_p} g\,\frac{\Delta\theta_v(z)}{\Theta_v(z)}\,dz \tag{1}$$

The height of the parcel that is being deflected to the surface is represented as z_p. Large turbulent eddies that are described in Brasseur (2001) are represented in the numerical model by TKE. From the surface to height z_p the local TKE in the layer is represented by $E(z)$ which the integral then transforms into the average TKE in the layer. This integral on the left side of the inequality represents the energy associated with these large turbulent eddies. The right side of the inequality represents the energy associated with buoyancy. $\Theta_v(z)$ represents the virtual potential temperature (in degrees Kelvin) at the specific height, whereas $\Delta\theta_v$ represents the difference in virtual potential temperature in the specific layer (Brasseur 2001). For the purpose of this study and the model data utilized (examined in Chapter III), the starting height of 0 in the integral indicates the surface represented by the model in surface wind speed calculations as 10 meters above ground level (AGL). This is consistent with current automated observation systems currently utilized at most airfields across the country and is representative of the surface wind speeds.

The WGE is then represented by the maximum wind speed of all heights that satisfy Equation (1) and is given by Equation (2):

$$Wg_{estimate} = \max\left[\sqrt{U^2(z_p)+V^2(z_p)}\right] \tag{2}$$

U and V represent the x and y components of the wind at the particular height that air parcels are being deflected to the surface (Brasseur 2001).
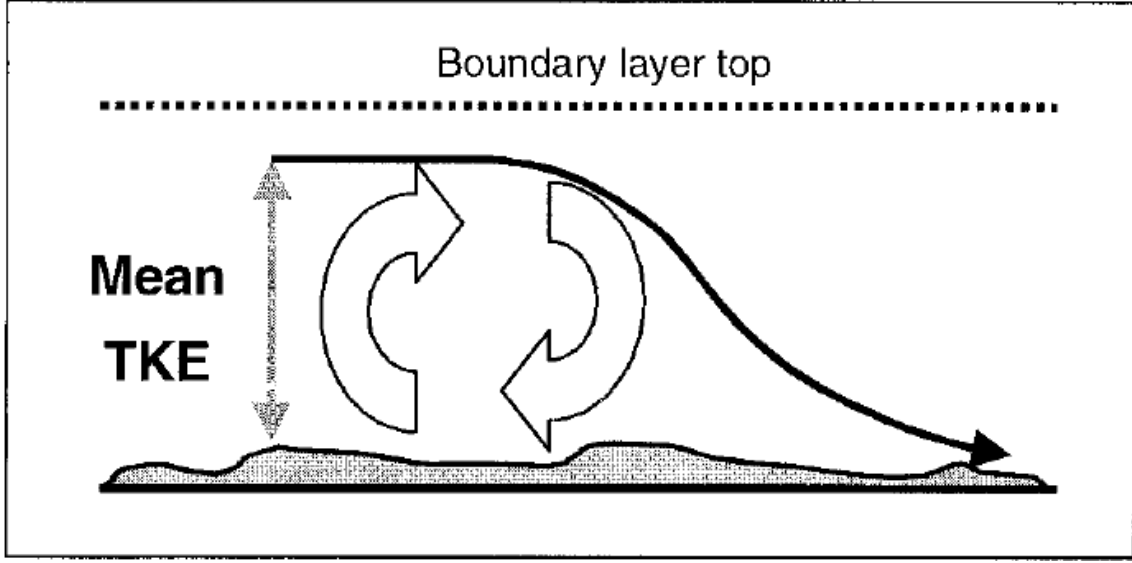
Figure 6.   Determination of the wind gust estimate based on turbulent kinetic energy averaged over a given depth (from the surface) in the boundary layer (From Brasseur 2001).

3.   Calculation of the Lower Bound

Though not analyzed in-depth in this research (brief results discussion in Chapter IV), the bounding interval calculation process from Brasseur (2001) is important to describe. Future research projects would benefit from analyzing this interval and the relation to probability of events occurring. The lower bound is similar to the WGE calculation except that it utilizes the local TKE at particular height versus the mean TKE through the layer as is the case for the WGE calculation. Since the lower bound utilizes the local TKE value versus the average, the value of the TKE variable will be smaller as height increases. This is the case for both stable and unstable layers, although the rate of change with respect to height will vary with different stabilities. The lower bound calculation is thus represented by the Equation (3):

$$\frac{\overline{w'w'}\left(z_p\right)}{2} \geq \int_0^{z_p} g \frac{\Delta\theta_v\left(z\right)}{\Theta_v\left(z\right)} \tag{3}$$

The right hand side of the inequality still represents energy associated with buoyancy in the atmosphere while the left hand side of the equation represents the vertical velocity variance associated with turbulence at the particular level. Brasseur (2001) further

explains that this vertical variance is not computed in the standard prognostic TKE calculation in a 1.5 order closure model but can be calculated using a simple ratio (2.5/11) multiplied by the local TKE, *E(z)* (Brasseur 2001). The lower bound wind gust speed is then represented by Equation (4):

$$Wg_{lower} = \max\left[\sqrt{U^2\left(z_p\right) + V^2\left(z_p\right)}\right] \tag{4}$$

4.   Calculation of the Upper Bound

The upper bound is simply the model's highest mean wind speed in the boundary layer. Given as Equation (5):

$$Wg_{upper} = \max\left[\sqrt{U^2\left(z_p\right) + V^2\left(z_p\right)}\right] \qquad \text{for } z_p \leq z_{top} \tag{5}$$

The variable z_{top} represents the boundary layer top (Brasseur 2001). Similar to the calculation of TKE, there are different methods for calculating the boundary layer top.

Numerical models represent the boundary layer top based on the planetary boundary layer (PBL) scheme chosen. For example, the Medium Range Forecast model (MRF) PBL scheme used in many models today, including variations of the WRF and MM5, calculates the PBL top utilizing the bulk Richardson number (LaCroix 2002, Skamarock et al. 2008). Additionally in the WRF model, the Yonsei University, MYJ, and Asymmetrical Convective Model Version 2 PBL schemes are available as choices to use. As mentioned previously, the MYJ scheme is utilized in the data set for this research and the main difference from this scheme and the MRF scheme utilized in previous research is that boundary layer top is calculated from TKE versus the Richardson number (Skamarock et al. 2008).

5.   Important Results and Conclusions from Brasseur (2001)

As part of the analysis section, Brasseur (2001) examined whether or not it appeared that inaccuracies in the prediction of the winds were due to the model or the WGE method. The first result from Brasseur's (2001) research is that of the dependency of the WGE to the model predicted meteorological fields and in particular the boundary layer. This seems intuitive, but the results of incorrect prediction of deepening low

pressure systems, for example, can cause large errors in the estimation of wind gusts. Moreover, strong vertical mixing with the boundary was shown to cause an overestimation of the bounding interval as well as the wind gust. This likely results in a key overestimation of the winds especially near sunrise when the nocturnal boundary layer is at its greatest (Brasseur 2001).

A large part of the results section detailed the impact of model horizontal resolution to the prediction of wind gusts using the WGE method. It was found that the higher the resolution of the model (50 km versus 25 km) the more accurate the wind gust estimates were due the model's ability to more accurately resolve mesoscale features within the cyclone (Brasseur, 2001).

In this day and age of budget cuts, especially in the DoD, managers should perform a cost/benefit analysis of accurate predictions balanced with the computing cost of increased resolution. Ideally, there exists a compromise where inaccuracies are accepted for a lower resolution model. It is hoped that the results of our research will shed some light on the ability of the WGE to predict non-convective wind gusts with a lower resolution model within the CONUS.

The bounding interval also proved to be a source of valuable information. Results on the reliability of the bounding interval for a three month period during the cold season at nine stations in Belgium were analyzed. On average, when the observed wind gusts ranged from 20 kts to 39 kts, the reliability of the predicted bounding interval was 81%. When winds were greater than 39 kts the reliability of the bounding interval decreased to 73% but still a high reliability percentage further emphasizing the need of accurately predicted parameters especially during severe events. The lowest reliability occurred when the observed winds were below 20 kts likely caused by overestimation of mean winds in the boundary layer (Brasseur 2001). One possible example of this, mentioned previously, is the overestimation of the vertical mixing by the model in the nocturnal boundary layer.

One final important result from Brasseur (2001) was the comparison of the results from the WGE to two different but widely used empirical methods for forecasting wind

gusts today. The WGE outperformed one method and performed similarly to the second method. These results indicate the WGE is a viable wind-gust forecasting method that can perform as well or better than widely used popular methods in today's operational forecasting realm (Brasseur 2001). This conclusion provided the motivation to compare the WGE to another empirically based wind gust model explained in Section 2C.

B. RESULTS FROM PREVIOUS STUDIES ON THE WGE METHOD

Several studies have been conducted on the WGE method since its development in 2001. LaCroix (2002) analyzed results from the WGE method as well as the relevancy to operational forecasting, and in particular to the Air Force Weather operational forecasting community. The purpose of the study was to compare and contrast the WGE method and the AFWA wind gust method run on 53 different 06Z and 18Z MM5 model runs. The results were compared and analyzed at 23 airfields (both military and civilian) throughout the CONUS that represented different geography from coastal regions to mountainous terrain. The verification criterion chosen was 15 kts. If a wind gust of 15 kts or greater was forecast, this triggered a "yes" in standard 2x2 contingency table verification procedures. For verification purposes in the absence of a wind gust, an observed sustained wind speed of 15 kts would be considered a "yes" observation in the contingency table verifications procedures (LaCroix 2002). Although, there are a few limitations identified in this study that were mitigated in our research (explained further in Chapter III), the results are important and relevant and were used to help formulate the goals and hypothesis as we began our research.

Overall RMSE showed that the WGE was more accurate than the AFWA algorithm although there were important variations that should be noted. Based on using standard meteorological verification scores (Hit Rate, Probability of Detection, etc.) the overall results showed that the WGE was better during the daytime hours at accurately predicting the 15 kt wind gust events than the AFWA wind gust algorithm. However, at night, the AFWA algorithm outperformed the WGE, although the AFWA algorithm still showed a bias of overforecasting at night (LaCroix 2002). This result suggests that the nocturnal boundary layer characteristics in the model are not accurately predicted in the WGE. Parameters such as TKE may be overestimated, especially during overnight hours

causing overforecasting of surface wind gusts at the surface. When analyzed for specific regions, the data from the coastal stations typically did not conform to the diurnal trend mentioned previously. Alternatively, when the results for the coastal plains locations were isolated, the data revealed that the AFWA algorithm performed better than the WGE in almost all categories at all times (LaCroix 2002).

Utilizing the Canadian Regional Climate Model (CRCM) at resolutions of 60, 20, 5 and 1 km, Goyette et al (2003) implemented the WGE method and analyzed two strong extra-tropical cyclones and the predictions of the WGE at various stations in Switzerland and Belgium. This research takes a slightly different avenue of approach from Brasseur (2001) and LaCroix (2002) as it explores the impacts of terrain, and ultimately the importance of the model's ability to resolve the terrain and associated meteorological parameters in the boundary layer on the WGE (Goyette et al. 2003).

One key result from this study was that not only do the atmospheric parameters in the boundary layer need to be accurately predicted by the model in order for the WGE method to be accurate, but the "atmospheric flow field" or impacts of terrain on the flow also needs to be resolved well by the model. Vertical resolution also played a factor in the results as it increased from 30 sigma levels at 20 km to 46 sigma levels at 1 km. Similarly, vertical and horizontal resolution had an impact on other small scale features being resolved as well. For example, the authors note a low-level jet detected by the 1 km (46 sigma level vertical resolution) grid, but not the 20 km (20 sigma level vertical resolution) grid. This low-level jet represented a statically stable layer that was otherwise undetected by the 20 km grid. The result of this feature being missed was a higher boundary layer height prediction (identified at the 14th sigma level) and therefore a higher maximum wind speed that could reach the surface at the critical level where TKE was greater than buoyancy. In the 1 km grid, the boundary layer height was noted at the 8th sigma level (roughly 1 km lower) and therefore predicted a lower maximum wind gust. Due to the increased vertical resolution of boundary layer wind speeds (with the low level jet) and TKE variations, the atmospheric conditions are much more representative (see Figure 7). Consequently, when compared to observations, the 1 km grid forecast was much more accurate (Goyette et al. 2003).

22

Finally, most results showed that the higher the spatial and vertical resolution of the model the more accurate the WGE prediction was. However, at several locations in Belgium, which has similar geography to locations like the Great Plains of the United States, the relatively flat surface provided only somewhat better gains in accuracy when resolution was increased. The greater gains, and ultimately if measured, the more cost-effective approach, were realized in the higher resolution output over large topographical changes such as the Alps (Goyette et al. 2003). This is a valuable conclusion during a time of budget constraints when calculating the cost-benefit analysis of running higher resolution models across the entire CONUS. Perhaps running this algorithm at lower resolutions over relatively smooth terrain is a better solution while running the algorithm in higher resolution nested grids for regions with high topographic changes not only in the CONUS but other regions of the world where assets could be employed and mountains are a factor. It was this conclusion that motivated our study and analysis of a 45 km model with higher vertical resolution.

(a) **(b)**
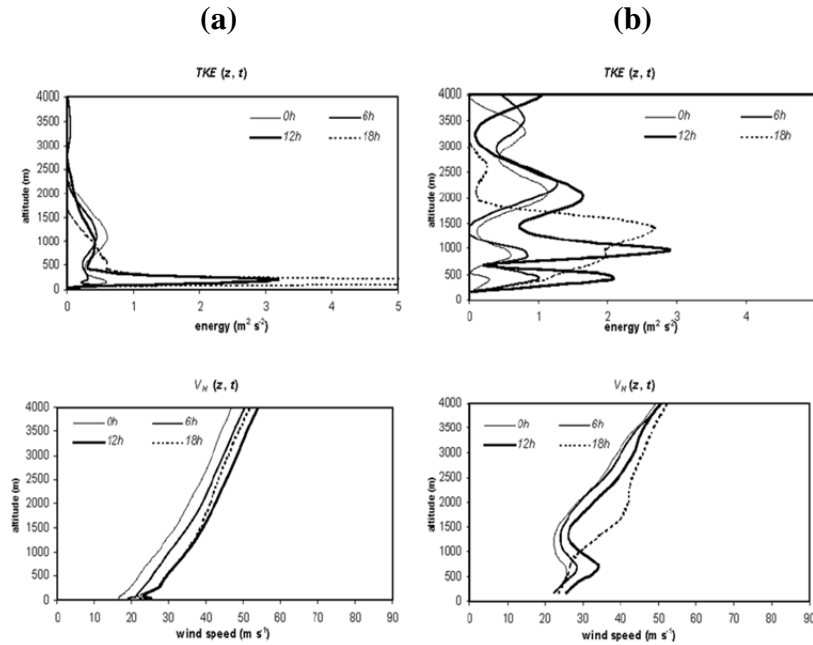
Figure 7. Simulated TKE and wind profiles, $V_H(z, t)$, at 0, 6, 12, and 18 hours UTC, at station Visp (Switzerland), elev. 2100 ft, during VIVIAN storm with CRCM at (a) 20 km grid spacing and (b) 1 km grid spacing. Vertical axes are wind speed difference in m s^{-1} and the height above the resolved surface in meters (After Goyette et al. 2003).

23

Nordström (2005) conducted research on the WGE method utilizing the Rossby Centre regional Atmospheric model (RCA) which was chosen due to the tendency of the model to underforecast wind gusts. The RCA is a regional climate model primarily used for climate studies across northern portions of Europe. With a higher resolution than a global model, the regional climate model downscales the global model data in order to predict phenomena at much smaller spatial scales than can be resolved at the global level. The model has 24 vertical levels and a horizontal resolution of 22 km and includes a 1.5 order closure which includes associated TKE values. The boundary layer height is determined by the bulk Richardson number (not TKE values such as the MYJ PBL scheme; Nordström 2005).

Nordström (2005) analyzed the results from storm systems over southern Scandinavia in 1999 and 2005 as well as a three month simulation from 2004 to 2005. The author notes that the WGE method showed a bias of overforecasting winds during the storms by as much as 25 ms^{-1} (49 kts). However, during the three month simulation it appears that the bounding interval captured the observed wind gusts at locations on land. In comparison, over the open water the WGE prediction of the magnitude of the winds were better represented by the model, however the bounding interval was less accurate overall in capturing the observed wind speeds. It was also apparent, as shown in other studies, the algorithm is highly dependent on the accurate prediction of main synoptic and mesoscale features and associated meteorological parameters that are used in the WGE (Nordström 2005).

Based on these results, Nordström (2005) developed a method to empirically tune the data to the results. The author concedes while not an explanation of the physical processes of the atmosphere, the tuning would dampen the wind gusts in order to produce more accurate results. The premise of this tuning is the assumption that the winds are slowed down through the surface layer after being deflected from above. The method produces an alpha variable to apply to the gust which was empirically chosen between 1.3 and 1.4. After applying the correction, the results improved over land, however over water where the original results were fairly accurate, the correction caused an overestimation of predicted wind speeds. Nordström (2005) concludes that although the

24

correction does not conform perfectly to the observed data, it does lower the predicted wind speeds over land which was the goal of the tuning (Nordström 2005).   This empirical tuning example, encouraged us to take a look at another possible way of tuning the algorithm in order to increase accuracy of the WGE method, particular for 35 kt to 49 kt wind warnings.

C.   RAPID UPDATE CYCLE (RUC) EMPIRICAL METHOD

Motivated by the conclusion from Brasseur (2001) that the WGE method performed as good as or better than two tested empirical relationships, this study also conducted a comparison to the RUC empirical method to forecast wind gusts. The RUC is an operational numerical weather model with 13 km grid spacing run every hour out to 18 hours by the National Oceanographic and Atmospheric Administration's National Centers for Environmental Predicton (NOAA/NCEP). For this research the actual RUC model was not used or analyzed, however, the algorithm for the post-processed variable of gust wind speed (hereafter referred to as the RUC method) was used with the WRF model data set described in Chapter III. The RUC method is a fairly simple empirical method that utilizes two model derived variables: boundary layer depth and wind speeds at each sigma level in the boundary layer to include the surface.   The RUC method process begins with first calculating the different in wind speed between the surface and each sigma level in the boundary layer. This difference is then multiplied by a coefficient that decreases from one at the surface to 0.5 at 1 km height and remains 0.5 for any height above 1 kilometer. The maximum "excess" value in the boundary layer computed from the wind speeds and associated coefficient is then added back to the surface wind speed for the peak gust. Equation (6) summarizes the RUC wind gust method:

$$Wg_{RUC} = WndSpd_{SFC} + \max\left[f(z) * \left(WndSpd(z) - WndSpd_{SFC} \right) \right] \quad (6)$$

where $f(z)$ denotes the coefficient at a particular sigma level and $WndSpd(z)$ represents the wind speed at that level (NOAA ESRL 2012).

THIS PAGE INTENTIONALLY LEFT BLANK

III.   DATA AND METHODS

A.   WRF-ARW MODEL

The Weather Research and Forecasting (WRF) model is the latest advanced state-of-the-art mesoscale weather model that was designed for education and research purposes as well as operational use (Skamarock et al. 2008). Utilized internationally, the WRF is a unique and collaborative effort. The WRF development and continuous improvement process spans across partners from the government sectors such as the Federal Aviation Administration and the National Oceanographic and Atmospheric Administration, to civilian institutions such as the National Center for Atmospheric Research (NCAR) and the University of Oklahoma, and importantly with the military organizations such as AFWA and the Naval Research Laboratory (Skamarock et al. 2008, Hacker et al. 2011 and WRF 2012). This collaboration across the spectrum of operational and research users in the Atmospheric Science community enables a wide range of enhancements and updates to the model while building closer bonds between these vital organizations with a similar goal: to improve weather forecasting. See Skamarock et al. (2008) for more details on the Advanced Research WRF (ARW) Version 3.

The WRF utilizes a hydrostatic pressure vertical coordinate (η). The calculation and equations are the same for the terrain-following coordinate σ used in many hydrostatic models. The following equations define η:

$$\eta = \frac{(p_h - p_{ht})}{\mu} \quad \text{where } \mu = p_{hs} - p_{ht} \tag{7}$$

The hydrostatic portion of the pressure is represented as p_h and the surface and top boundaries are represented as p_{hs} and p_{ht} respectively. Figure 8 depicts the terrain-following coordinate from the surface (shown as 1.0) to the top (shown as 0; Skamarock et al. 2008). Table 2 gives an example of how η level heights can vary with time based on surface pressure differences (only the lowest five levels are shown). Smaller differences (on order of <5 m) are noted at lower vertical levels, with larger differences (on order of 100 m) noted in the upper atmosphere (not shown). The model used for this

research contains 41 η vertical levels, with typically 10 to 15 levels in the boundary layer, which results in an increased vertical resolution over other models of similar horizontal resolution that were used in previous research on non-convective wind gust prediction.
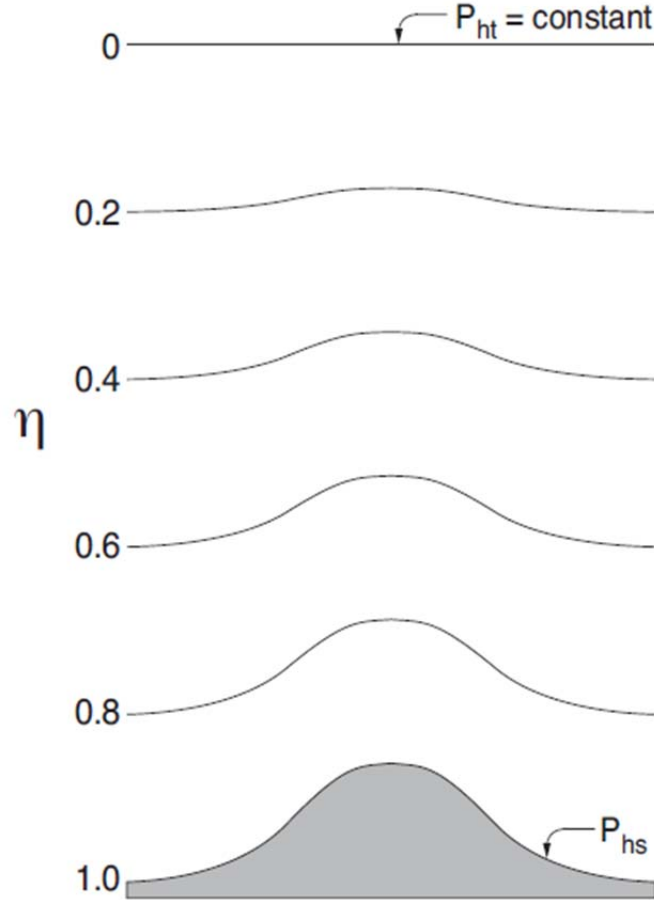
Figure 8.   ARW η coordinate (From Skamarock et al. 2008).

One main advantage of using the WRF in our research was the inclusion of TKE values throughout the boundary layer. The WRF includes a critical PBL scheme for our research; the Mellor-Yamada-Janjić (MYJ) which uses a 2.5 turbulence closure model. The top of the boundary layer is determined by TKE values as well as buoyancy and shear for the mean flow (Skamarock et al 2008). The inclusion of this key parameter is different than many of the previous research studies due to the need to solve for TKE values manually in their research.   The inclusion of TKE values from the model

minimizes errors associated with manual calculation, especially due to different methods noted in research to calculate this variable.

η level	Day A	Day B
1	19.34	18.61
2	58.11	55.90
3	97.05	93.32
4	144.07	138.41
5	207.32	198.83

Table 2.   Example of variations of η level heights (in meters) in a hydrostatic model due to differences in surface pressure between days in the lowest five model levels.

B.   WRF MESOSCALE ENSEMBLE DATA SET

The model utilized for this research was a 10-member WRF ensemble with 45 km horizontal grid and 41 vertical levels.   Hacker et al. (2011) tested nine different ensembles with a variety of model and initial condition perturbations in order to evaluate performance to identify best combinations for use in a mesoscale ensemble operational setting at AFWA.   The primary targeted audience: aviation customers.   As such, one main focus of the model evaluation was on lower atmospheric winds, which in turn is highly applicable to non-convective wind gusts and provided the confidence in the ensembles ability to accurately predict these winds (Hacker et al. 2011).   It should be noted that this study is not a continuation of their research, but instead uses the multiphysics ensemble to evaluate the performance of WGE and RUC algorithms.

Hacker et al. (2011) created an ensemble and incorporated 10 members with different sets of physics packages (see Table 3). Descriptions of these physics packages not explained previously can be found in Skamarock (2008). Each of the 10 members chosen for the ensemble represents a combination of different physics packages that run stably together and produced reasonable results. This multiphysics ensemble was shown to have better reliability for 10 m wind speeds than the control ensemble although the authors note the poorer performance for high wind events (Hacker et al. 2011). It is our

29

theory that applying non-convective wind gust algorithms to this data set will show increased performance for higher wind events. Furthermore, the choice of utilizing forecasts from this ensemble was further solidified by the results given by Hacker et al. (2011) on the multiphysics ensembles performance in the PBL. While highly accurate PBL predictions stand out as likely the most difficult to produce, it is shown that encompassing a broader spectrum of physics packages produces the best skill of predictions in the PBL compared to other individually tested ensemble methods (Hacker et al. 2011).

Overall, the data set consisted of 119 model runs with a mixture of 00Z and 12Z runs available. Each model run was initialized via cold-start interpolation from GEFS. The available model runs were from the months of June, November and December of 2008 and January and February of 2009. Our focus time period, given by research results in Chapter I, was November through February. Not all days throughout the available months included a model run. In fact, it was not uncommon to have two model runs (00Z and 12Z) for one day, with a day or two in between before the next available model run.

The data set included a maximum of 81 total variables (not including nine variables associated with character strings, model name, number of levels, etc.) in each ensemble run, however not all variables were available in each ensemble member. Due to the nature of the multiphysics ensemble, four of the 10 members utilized the Yonsei University (YSU) PBL scheme which does not include a prognostic TKE variable. Moreover, the YSU scheme uses the buoyancy profile to calculate the top of the PBL instead of TKE as in the MYJ scheme (Skamarock et al. 2008). The differences in the approaches between the WGE (physical) and the RUC (empirical) meant different variables were used for each. Table 4 highlights the main variables used to process each algorithm. Section E (Methodology) further examines the processes used by each algorithm to produce a forecast.

Member #	Land Surface	PBL	Microphysics	Cumulus	Long-wave	Short-wave
1	Thermal	YSU	Kessler	KF	RRTM	Dudhia
5	Thermal	MYJ	WSM6	KF	RRTM	CAM
7	Noah	MYJ	Kessler	BM	CAM	Dudhia
8	Noah	MYJ	Lin	Grell	CAM	CAM
10	Noah	YSU	WSM5	KF	RRTM	Dudhia
11	Noah	MYJ	WSM5	Grell	RRTM	Dudhia
15	RUC	YSU	Lin	BM	CAM	Dudhia
16	RUC	MYJ	Eta	KF	RRTM	Dudhia
17	RUC	YSU	Eta	BM	RRTM	CAM
19	RUC	MYJ	Thompson	Grell	CAM	CAM

Table 3.   Configuration of multiphysics ensemble adjusted to reflect our data sets member number. Member 10 uses the same physics suite as the operational configuration at AFWA (After Hacker et al. 2011).

Model variables used in WGE	Model variables used in RUC
10 m U wind component, 10 m V wind component, Sustained wind speed, Potential temperature, Ice mixing ratio, Rain mixing ratio, Cloud mixing ratio, Vapor mixing ratio, Geopotential height, Terrain height, PBL height, TKE	10 m U wind component, 10 m V wind component, Sustained wind speed, Geopotential height, Terrain height, PBL height

Table 4.   Variables used by each algorithm for this research. The TKE variable was available in six of the ten ensemble members.

C.   SELECTION OF LOCATIONS FOR EVALUATION

Five bases were chosen as test sites to evaluate the performance of the two algorithms. These sites were selected based on the criteria of minimal impacts of terrain

to wind gust enhancement as well as to sample and represent the geographical differences within the 15th OWS AOR. The five locations chosen were: Westover ARB, MA; Andrews AFB, VA; Langley AFB, VA; Scott AFB, IL; and Offutt AFB, NE. Figure 9 shows a map of terrain height and locations for the five bases used in this study.

D.   ALGORITHM PROCEDURES

Algorithms to calculate the WGE and RUC model wind gust forecasts from the WRF output were written and compiled using the program MATLAB. Model data netcdf files were converted to .mat files for use in MATLAB. For each forecast compiled, output was manually recorded into Microsoft Excel spreadsheets in order to perform various calculations and statistics described in Section E of this chapter. Surface sustained wind (10 m wind speed) and wind direction were not part of the algorithms analysis, however the variables were recorded in order to provide additional information on the model's performance of individual weather events. Prior to running the algorithms on the raw .mat file data, two interpolations were made.
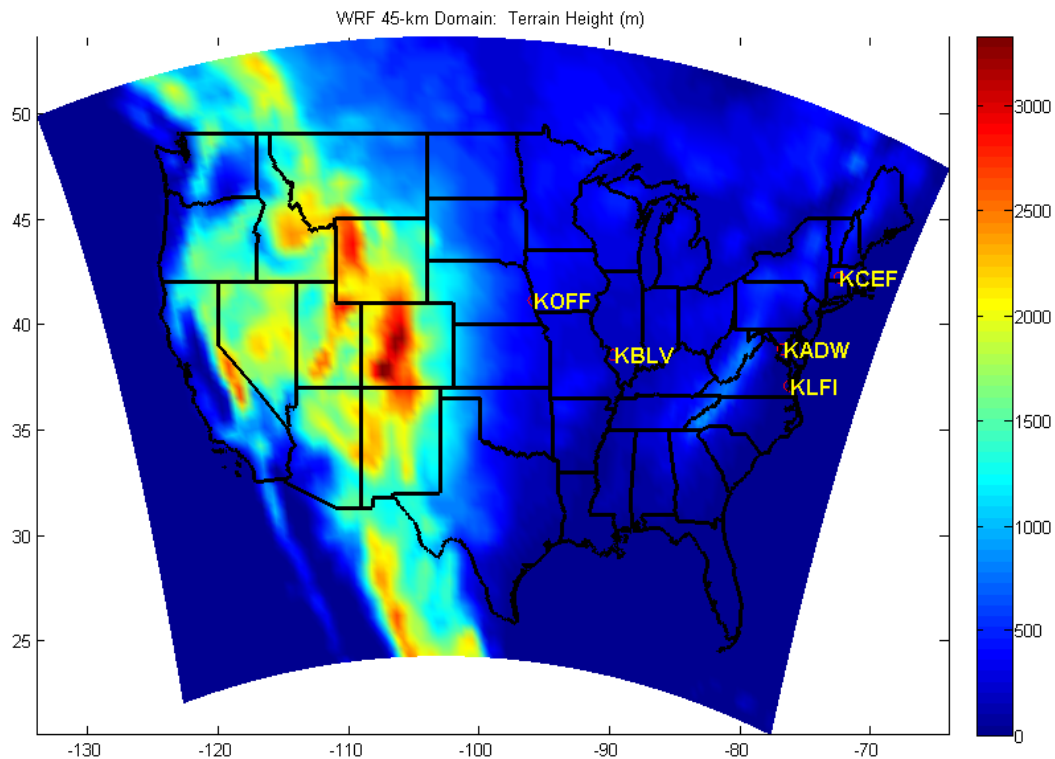
Figure 9.   Terrain map of 45 km WRF domain with locations used for this research plotted in MATLAB.

1.   Destaggering Variables in the Vertical and Bilinear Interpolation

Before processing the data with the non-convective wind gust algorithms, a subroutine (or function) was created to interpolate the data in two ways. First, the WRF model data set includes primary variables (U, V wind components, temperature, pressure) on 40 half η levels, however geopotential (φ) and vertical velocity (w) are available on 41 full η levels. Model data was unstaggered in the vertical which reduces, or interpolates, all variables to the whole η level. This process allows computations to be made more simply while retaining the integrity and resolution of the vertical grid. Due to the nature of the hydrostatic model, η above ground heights can vary as shown in Table 2. The max η level representing the PBL height at each location fell below the 20th level.

Another interpolation made to the model data was in the horizontal spatial realm of the grid. Since 45 km represents a large area between grid points, it was determined that an interpolation should be made at each location. The purpose of this was to ensure that more accurate representation of the variables at that location could be made versus using just the nearest grid point. Although terrain may not be a major factor at Westover ARB, MA, using a grid point 45 km to the west could mean a higher elevation point. Similarly, using a point 45 km to the east of Langley AFB, VA would result in an ocean data point. By taking the nearest four points, and interpolating that data to the actual location, a more accurate representation of the variable field can be made. This process is called bilinear interpolation.

Bilinear interpolation is a distance-weighted interpolation of gridded variables by using the nearest four grid points (Nett 2012). This interpolation includes all variables in an x, y, z grid. This means that a grid point relatively close to the actual location will get more weight than a neighboring grid point farther away. The interpolation is first made in the x-direction and then in the y-direction to the actual location. See Nett (2012) for further explanation on the process of bilinear interpolation.

After the data was bilinear interpolated, the actual and model terrain heights were compared to note any differences that might impact the accuracy of the data and forecasts. Table 5 shows the comparison between actual elevation and the model terrain

height. Westover ARB is the only outlier where the elevation difference was much larger than its actual elevation. Careful investigation reveals the western nearest grid points are located in higher terrain while the eastern grid points are similar to the actual station elevation. This difference will be accounted for in the data results chapter.

Instead of running the data across the grid to produce calculations, a column vector of data was extracted from the surface to the top of model layer for the point location. The two algorithms were then applied using this column vector of data. For each model run, data for all five locations were extracted. The data was placed in matrices of columns (one column for each of the locations) for each model variable.

	Actual elev (m)	Bilinear Interpolated elev (m)
Westover ARB	73	201.15
Andrews AFB	85	49.33
Langley AFB	3	5.04
Scott AFB	140	146.16
Offutt AFB	319	342.77

Table 5.   Comparison of actual and model estimated elevation heights in meters.

2.   WGE Algorithm

This section describes the procedures used in the WGE algorithm. The WGE algorithm used for this research is a modified version of the algorithm currently used to produce output for CONUS OWS modeled Skew-Ts. Due to the difference in available parameters (TKE, modeled PBL height, etc.), the adapted version differs in the process to compute values used in the algorithm. Therefore, the results produced here should not be considered verification of the OWS algorithm due to the differences in model data variables available as well as the modifications made to the algorithm.

The first step of algorithm is to set the initial gust estimate to the modeled 10 m gust variable. If the modeled PBL winds are less than that of the surface gust parameter,

then the 10 m gust variable represents the WGE prediction for that time period. In some cases, the 10 m gust parameter was missing, and therefore an artificial, but representative value was created by combining the u and v wind components at 10 m. The wind direction was also extracted from the model data field to provide additional information on the model's timing error of frontal passages.

Next, in order to calculate the surface parameters, variables are assigned values based on η levels. The surface potential temperature value is set to value at the lowest η level (1). As described above, this is a good approximation since the lowest η value is around 18–20 m AGL. The mixing ratios for rain, cloud, and ice are combined for each level with the value at the first level recorded as the surface combined mixing ratio. This same process is accomplished for the surface vapor mixing ratio. Next, a variable is created to represent part of the virtual potential temperature equation. The following equation converts potential temperature to virtual potential temperature (LaCroix (2002):

$$\theta_v = \theta\left(1 + .61r - r_L\right)$$

(8)

The vapor mixing ratio is represented by r and the combination of rain, cloud and ice mixing ratios is represented by r_L. After computation of the value inside the bracket is accomplished, the algorithm multiplies this result by the potential temperature at the surface to produce the virtual potential temperature.

The next step in the algorithm is to determine the η level that represents the top of the boundary layer. First the AGL level for each η level is computed. Then a one column matrix is created to determine how many levels are less than the MYJ produced PBL height variable. The maximum level in this matrix is indentified as the η level PBL top and the algorithm will not exceed this level when computing the WGE.

A "for" loop is then created to run from η level 1 to the η level identified as the top of the PBL. The values are calculated from the surface to the top of the PBL which is a departure from the top-down method used by LaCroix (2002). The algorithm steps up from the surface to each η level and calculates the variables in the layer required for both the TKE and buoyancy terms in Equation (1) shown in Chapter II. For each layer, both sides of Equation (1) are calculated and added to the running total of each respective

35

integral. After a layer is computed and added to the total integral, the TKE component is compared to the buoyancy component. If the TKE component is greater than or equal to the buoyancy component, then the wind speed at the highest η level used in the layer is compared to the previously recorded value. If the current level's winds are greater than the previous level's winds, the max gust is set as the new wind speed overriding the previous value. If the value is less than the previously recorded max gust value, it is discarded and the loop continues. As soon as the algorithm finds the point where the buoyancy term is greater than the TKE term, the loop is exited and the max gust value is recorded as the WGE. Calculations for the lower and upper bound of the WGE are also embedded in the algorithm and follow Equations (3), (4), and (5) from Chapter II. Due to the limited amount of analysis conducted on the bounding interval, detailed steps of algorithm will not be provided.

3. RUC Algorithm

The RUC algorithm is an empirically based estimate of non-convective wind gusts. As such, this makes the algorithm much shorter compared to the WGE. The boundaries of the RUC algorithm are similar to the WGE in that the values computed must stay within the PBL. Similar to the WGE, a determination of the top of the PBL, represented by the nearest η level, is calculated. This alerts the algorithm when to stop running the associated "for" loop and end the program.

The empirical portion comes from the coefficient that is applied at each level explained in Chapter II. The description of the code from NOAA ESRL (2012) notes that the coefficient decreases from 1 to .5 at 1 km AGL. Therefore, a column matrix was created with values of these coefficients. An average η level for 1 km was determined to be around the 12th level. Although this can vary as shown previously, the small fluction (+/- 50–100m) was determined not to vary enough to drastically alter the value of the coefficient applied. Therefore, coefficients linearly decreased from 1 at η level 1, to .4995 at η level 12 and remained .4995 at any level above 12 per the description given in Chapter II.

For each level the coefficient is multiplied by the wind speed at that level and added back to the surface wind speed value and stored in a matrix. When the algorithm reaches the top of the PBL, the matrix of adjusted wind speeds is analyzed and the maximum wind speed recorded as the RUC non-convective wind gust value.

E.   METHODOLOGY

The methodology and design of this project was tailored towards performance evaluation of the non-convective wind gust algorithms to forecast a specific wind warning threshold used in today's operational weather squadron setting. Therefore, the methodology uses slightly different verification processes and procedures compared to other studies. As such, the methodology and statistics presented in this research are designed to give a summary of the performance of the algorithms using a proven reliable model system (in this case the WRF model with the multiphysics ensemble) and less designed to evaluate the performance of the ensemble model itself. Therefore, statistics and results presented in this research will reflect the algorithms performance based on this methodology. Furthermore, the benefits of using the ensemble outweighed the benefits of using a single deterministic model due to the increased skill over using an equal resolution model run (Brennan 2012). For this reason, and due to the future of ensembles in the Air Force Weather community (Hacker et al. 2011), it was important to analyze the non-convective algorithm's performance using this ensemble.

1.   Selection of Events and Forecast Distribution

Beginning with selection of events for analysis, the first step was to gather all observations for the time period of 21 November 2008 through 21 February 2009 (coinciding with available model data sets) for the locations chosen and described in Section D. Therefore, observations for 92 days at each location were compiled from the 14th Weather Squadron's, the Air Force's Climatology Center, online database. For the five locations used, this yielded roughly 11,000 observations to analyze. To focus on the algorithms forecast of a specific threshold, we chose an observed value that would likely cause a forecaster to increase their situational awareness and to thoroughly analyze all possible tools for determination of wind warning issuance (in this care for winds $\geq$35 kts)

37

in an operational setting. This requires a key assumption: model skill under the threshold chosen is relatively good. Using the results of Hacker et al. (2011) and preliminary tests on a few cases, we felt confident that this assumption was valid especially for days where significant weather pattern changes (strong cyclones and associated frontal boundaries) did not occur.

The threshold was set to analyze forecasts on time periods when an observed value 30 kts or greater occurred. If a 30 kt observation was noted, it was grouped with surrounding observations closest to the nearest 3-hour forecast valid time. For example, a 0655Z 30 kt observation would be grouped with the 0455Z, 0555Z and 0755Z observations (-1 hr, valid time, +1 hr, +2 hr), hereafter referred to as an observation group, for error analysis. This narrowed the database from 11,000 observations down to nearly 630 and yielded 157 observation groups. These observation groups were then matched to the applicable 3-hour forecast panel from the closest available model run (explained further in Section E3). Figure 10 is a graphical representation of the 3-hourly forecasts with associated observation groups used for verifying maximum wind speeds.
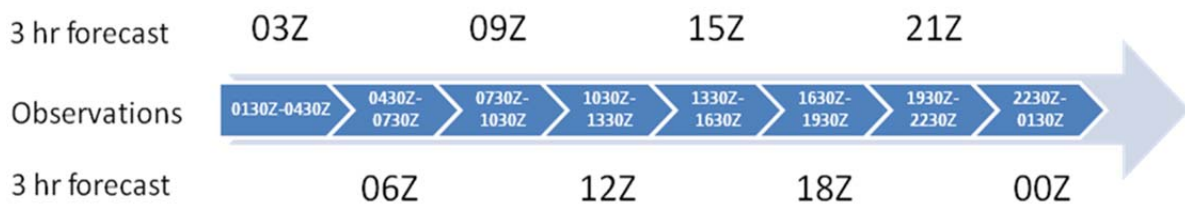
Figure 10. Three-hour forecast timeline with associated observation group times.
Max wind speed recorded in each observation time frame was used for verification of
accompanying 3-hour forecast.

Figures 11 and 12 show the forecast distribution by location as well as overall distribution partitioned by 6-hour forecast groups, respectively. Although the majority of the analysis was done with combined results, some results and conclusions are made based on the accuracy associated with specific forecast lead-times. However, due to the reliability of the model during the first 48-hour period (as shown by Hacker et al. 2011),

the combination of results from different forecast times should not greatly alter the overall statistics and results. 40% of the forecasts verified with within 12 hours of the model run and 56% within 24 hours. 26% of the forecasts verified were between 27 and 30 hours from the model run time indicative of the gap between days of model runs.
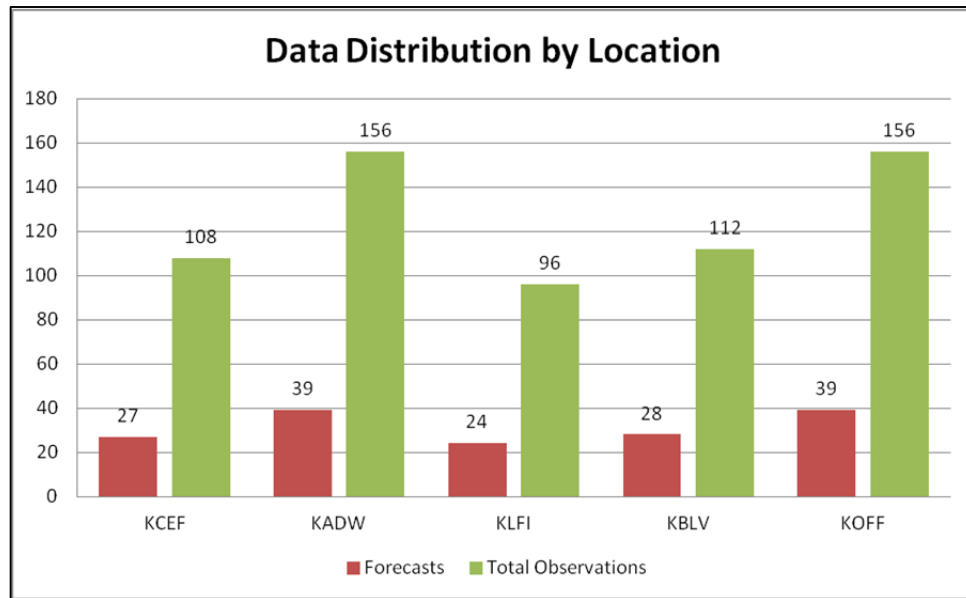
Figure 11.   Distribution of overall forecasts and accompanying observations by location. Red bars represent the number of forecasts analyzed at each location. Green bars represent the total number of observations analyzed at each location.
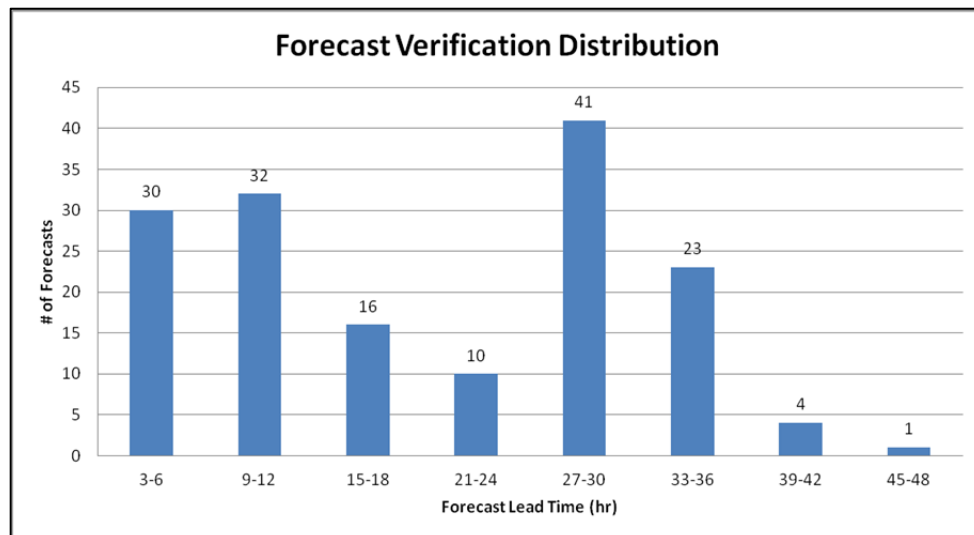
Figure 12.   Distribution of overall forecasts by lead time from closest model run binned in six hour forecast groups.

2.   Utilizing Ensemble Mean for Deterministic Forecast and Verification

Six of the 10 members utilized the MYJ PBL scheme, which produces a prognostic TKE value for each vertical level and computes boundary layer heights based on the TKE profile within the boundary layer. The other four members use the YSU scheme which does not produce a TKE value and calculates the boundary layer height from the buoyancy profile (Skamarock et al. 2008). Since TKE is a requirement for the WGE algorithm and to not incorporate error with differences in calculations of TKE, the ensemble mean forecast was taken from the six available members utilizing the MYJ PBL scheme (see Table 3). However, the RUC algorithm utilizes variables available in all members and therefore all 10 members forecast were utilized to calculate the ensemble mean of the RUC algorithm. The decision was made to incorporate all 10 members into the mean calculation for the RUC to produce a more reliable mean. This, in turn, could give the RUC method a statistical advantage; however when each members prediction were analyzed, the spread was such that it produced a minimal impact on mean value. Therefore the 10-member mean was used to analyze the RUC predictions.

A simple application of ensembles is to utilize the ensemble mean as a single forecast since this represents the most likely forecast from the initial atmospheric conditions. In fact, one major goal of ensemble forecasting is to improve the forecast by ensemble averaging (Kalnay 2003). Intuitively, using the ensemble mean then tends to average out large errors and enhance the similarities in member forecasts (Wilks, 2006). Using the mean as deterministic forecast gives users some information on the likelihood of an event occurring since the mean is comprised of its associated member's forecasts (Jolliffe and Stephenson 2003).   The ensemble mean is a non-probabilistic (or deterministic) forecast.

Verification of the ensemble mean yields pros and cons as explained in Ebert (2012b). Pros include filtering out smaller and unpredictable scales which, in turn, represents the ensemble's skill.   Furthermore general operational forecasters and other users of ensembles typically use the ensemble mean. Therefore, methods used to verify the ensemble mean can include deterministic (non-probabilistic) verification scores (Ebert 2012a,b). The purpose of this study is to show how well the algorithms performed

given a reliable data set. With this in mind, it was decided to utilize the mean as the single deterministic forecast and to compare verification statistics based on this forecast using documented statistics from Wilks (2006) rather than to use any individual member of the ensemble.

3. Method of Calculating Overall Forecast Error

The method for calculating the forecast error was determined by analyzing the initial results from the output of the algorithms. Error results were plotted on histograms to reveal the distribution of errors associated with the algorithms. Ideally, the errors would all be near zero, or have a Gaussian (or normal) distribution from the error distribution. For the determination of method to compute error statistics, two histograms were plotted. The bins associated with these histograms were in 2 kt intervals centered around a 4 kt central axis. The first histogram computed the wind gust error using the maximum wind speed recorded in the observation at the valid time of the forecast. For example, a 06Z forecast would be verified with at 0555Z observation using the highest wind speed recorded in the observation. The results show a fairly normal distribution of errors of the RUC algorithm, however, less of a normal distribution for the WGE. Moreover, the number of errors that exceeded 10 kts is also high for this verification (see Figure 13) indicating an overestimation of wind gusts.

Wind Gust Error (kts) All Hours/All Stations Obs at Forecast Valid Time

	<-10	-10 to -8	-8 to -6	-6 to -4	-4 to -2	-2 to 2	2 to 4	4 to 6	6 to 8	8 - 10	>10
WGE	9	10	2	11	11	28	18	20	10	10	28
RUC	5	6	8	18	20	46	16	3	11	7	17
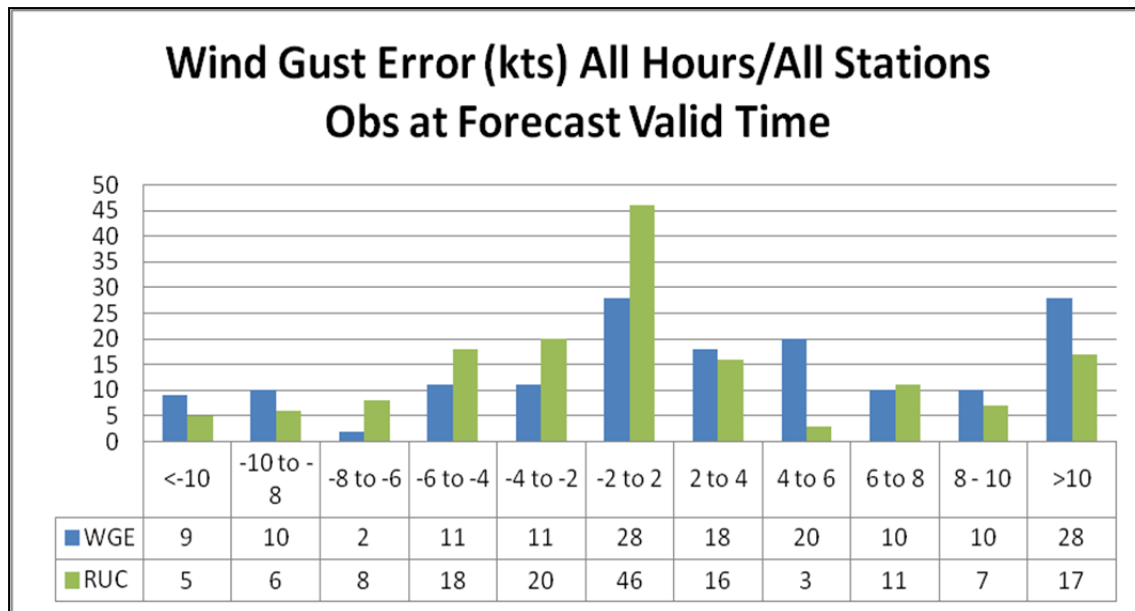
Figure 13.   Wind gust error distribution using maximum wind speed from observations at forecast valid time. Blue bars represent the number of WGE errors for each bin, green bars represent the number of RUC errors for each bin. Errors are binned by 2 kt intervals centered on a 4 kt central axis.

Due to inherent timing errors common in weather models, frontal passages and other mesoscale features that effect maximum wind forecasts could occur within the three hour gap not represented by the forecasts. Therefore, we also analyzed errors based on observation groups previously explained using the maximum wind +/- 1.5 hours from the forecast valid time. This method ensured that all gaps in between forecast valid times are covered by an observation for verification. For example, a peak wind remark recorded at 0135Z would be compared to the 03Z forecasts, while a peak wind remark recorded at 0125Z would be compared to the 00Z forecast. Increasing the verification window is a similar approach to a new form of verification called neighborhood (or "fuzzy") verification (see Ebert 2009). Since neighborhood verification is typically used for event verification, we chose this similar approach to increase the temporal window for wind speed verification. Intuitively, by increasing this window, only a higher observed wind speed would change the error value. Therefore, it is assumed this will minimize the overestimation and shift the errors to the left. WGE errors were more normally

distributed, while the RUC shifted to a slight underestimation of winds with 35% of errors falling between -2 kts and -6 kts (see Figure 14).

Wind Gust Error (kts) All Hours/All Stations Obs +/- 1.5 hr to Forecast Valid Time

	<-10	-10 to -8	-8 to -6	-6 to -4	-4 to -2	-2 to 2	2 to 4	4 to 6	6 to 8	8 - 10	>10
WGE	24	9	7	10	10	39	20	12	8	8	10
RUC	15	16	11	28	27	32	11	4	5	6	2
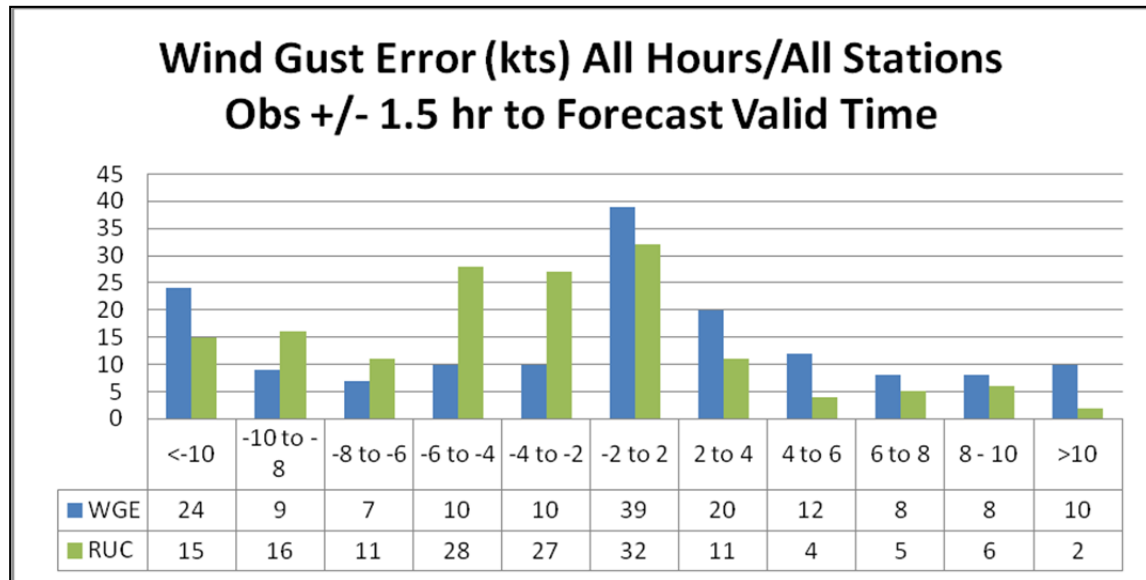
Figure 14.   Wind gust error distribution using maximum wind speed from observations +/- 1.5 hours from forecast valid time. Blue bars represent the number of WGE errors for each bin, green bars represent the number of RUC errors for each bin. Errors are binned by 2 kt intervals centered on a 4 kt central axis.

Mean error and RMSE also provide an indication of algorithm performance when compared with these different methods. According to Wilks (2006), mean error of greater than zero will reflect, on average, forecasts that are too high while the opposite is also true. Furthermore, the bias indicated here does not give information related magnitude and cannot be considered an accuracy measurement, but can give an indication of performance. Alternatively, RMSE does give an indication to the magnitude of the errors since it represents the error using similar units as the variable for wind. More on the results from RMSE calculations will be discussed in Chapter IV, but it is important to note the increased accuracy of the algorithms when compared to the observation groups (see Figure 15).
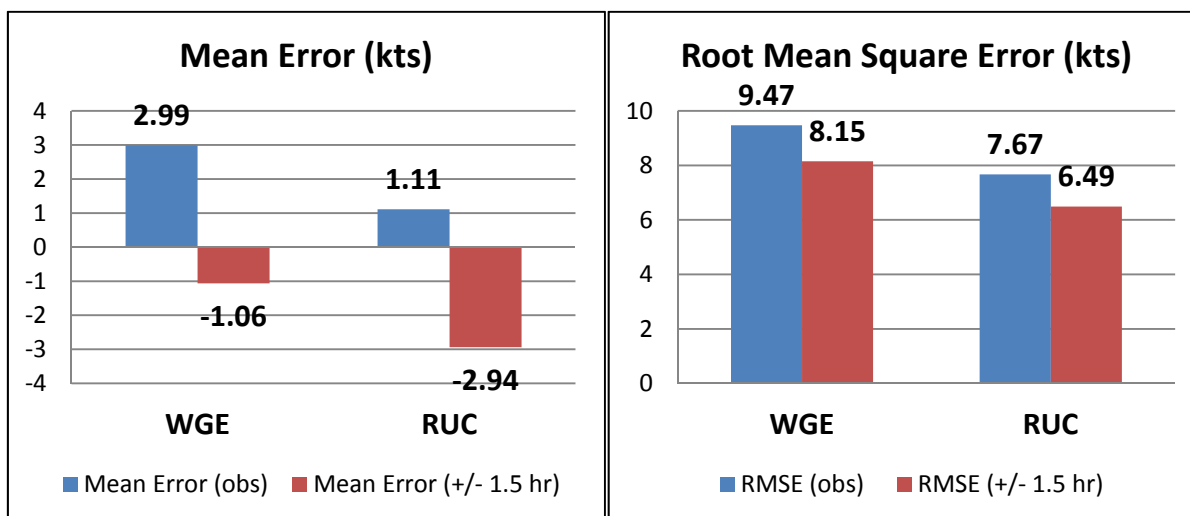
Figure 15.   Mean error and root mean square error comparisons. Blue bars indicate errors calculated using maximum wind speed from observations at forecast valid time. Red bars indicate errors calculated using maximum wind speed from observations +/- 1.5 hours from forecast valid time.

4.   Method of Calculating Hit/Miss/False Alarms

Choosing an appropriate method for calculating overall error was an important step in determining the methodology for event verification. Consistency is vital when analyzing data and producing results.   Using a strategy similar to neighborhood verification (Ebert 2009) allows us to verify the algorithms ability to predict warning level winds in an environment where winds are close to the threshold. Our method for calculating hits, misses, and false alarms makes use of standard practices for warning verification with an increased temporal window.

Using the observation groups determined from analysis of 30 kt observations, a threshold of 35 kts was set for event verification. In standard "yes/no" contingency table verification methods, if the algorithm produced an ensemble mean forecast greater than or equal to 35 kts (ensemble mean values rounded to the nearest knot) or an observed valued of 35 kts or greater occurred, then this constituted a "yes" in the contingency table.   Likewise, if either did not produce a results of 35 kts or greater, then this constituted a "no." The use of a 35 kt threshold is a rather strict but straightforward way

44

of verifying the ability of the algorithms to predict a specific warning level threshold. Other research has focused on other thresholds such as wind speeds of 15 kts (LaCroix 2002).

5. Statistics Computed

Based on Wilks (2006) and in computing statistics comparable to LaCroix (2002), standard two by two (2x2) contingency tables were developed (Figure 16). Air Force standard verification scores included Critical Skill Index (CSI), Probability of Detection (POD), and False Alarm Rate (FAR). Hit Rate (HR) measures the number of correct forecasts (yes/yes, no/no) compared to the total number of forecasts (n). Therefore, a perfect forecast would have a HR of 1. This score is important in assessing the performance for the two wind gust algorithms for our study. The ability of the algorithms to predict non-events in environments when observed winds were 30 kts or greater was critical to increasing confidence in the performance of the algorithms. This makes HR a useful statistic. Using the 2x2 table, this relationship is given by (LaCroix 2002):

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

The threat score, also referred to as the CSI, is often a more useful tool when verifying events that happen much less frequently than non-events (Wilks 2006). This is particularly the case when discussing the frequency of 35 kt events but may not be as important as HR. CSI is calculated as the proportion of correct forecasts to non correct forecasts for the threshold after removing non-occurrences (Wilks 2006). Therefore, using Figure 16, this is represented as:

$$CSI = a/(a+b+c)$$

Similar to HR, POD is also a useful calculation of algorithm performance as it measures the actual detection of verified "yes" events compared to the number of times the event actually occurred (Wilks 2006). This is represented as:

45

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

FAR is a measure of the ratio of the false alarms to the total number of events that did not occur (Wilks 2006). False alarms and misses (events not forecast, but those that occur) are critically important to DoD assets. A miss can cause damage or injury and associated monetary loss while a false alarm can cause wasted resources and cancelled operations also resulting in monetary loss. FAR is represented as:

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

Other scores associated with contingency table verifications were also computed outside of the typical warning verification statistics described in this section. The additional scores include bias, and two unbiased skill scores, the Kuiper (KSS) and Equitable Threat Skill Scores (ETS).

Similar to mean error previously mentioned, bias is not a measure of accuracy, however it is a measure of the algorithms performance in terms of over or underforecasting. Instead of positive and negative values representing this result as in the mean error, bias scores greater than or less than 1 indicate over or underforecasting, respectively, with a perfect score of exactly 1 (LaCroix 2002). Bias is represented as:

$$BIAS = (a+b)/(a+c)$$

KSS is an unbiased skill score that rewards forecasts of rare events and is applicable to our strict event threshold. ETS, also an unbiased skill score, utilizes CSI and accounts for random forecast events (LaCroix 2002). These skill scores are calculated by:

$$KSS = (ad-bc)/(a+c)(b+d)$$
$$ETS = (ad-bc)/(ab+ac+ad+b^2+bc+c^2+cd)$$

		Event Observed		
		Yes	No	
Event Forecast	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	n

Figure 16.   The 2x2 forecast verification matrix (From LaCroix 2002).

47

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DATA ANALYSIS

A. OVERVIEW

This chapter is broken into several sections to examine the variety of ways used to analyze the data set using the methodology described in Chapter III. The first analysis was done on the combined statistics for each algorithm at each location for sustained wind forecasts and the RMSE for wind gust predictions. Then location-based results were calculated using several techniques to form conclusions on the performance of the two algorithms at each site. After analyzing the algorithms performance at each location, a linear regression analysis was performed in order to assess accuracy differences. Finally, 10 cases studies (two at each location) were analyzed as independent data sets utilizing the results and conclusions from the location-based performance metrics as well as the linear regression tuning of the algorithms. A summary of the results from this analysis is presented in the last section of this chapter.

B. COMBINED RESULTS

The ability to accurately predict sustained wind forecasts can be an indicator of how well the model is handling the current weather features that are affecting a particular location. Since sustained winds do not vary as much as wind gusts, a time window surrounding the forecast was not used. The errors were combined to produce an overall sustained wind forecast error histogram similar to Figure 13 for wind gust errors. Figure 17 shows that the errors closely resemble a Gaussian distribution with a slight underforecast bias between two and six knots. This result, combined with the mean error and RMSE of the sustained wind forecasts shown in Figure 18, gave increased confidence in the model's ability to handle boundary layer features that effect surface winds forecasts.
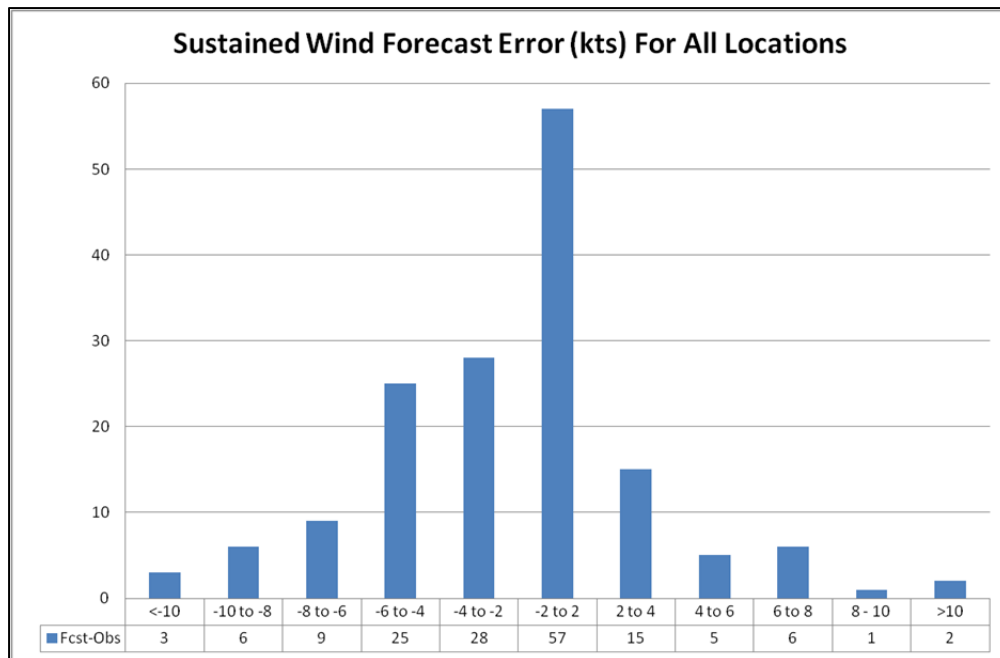
Figure 17.   Sustained wind forecast error histogram for all locations. Errors calculated using sustained wind speed from observations at forecast valid time. Blue bars represent number of errors binned by 2 kt intervals centered on a 4 kt central axis.
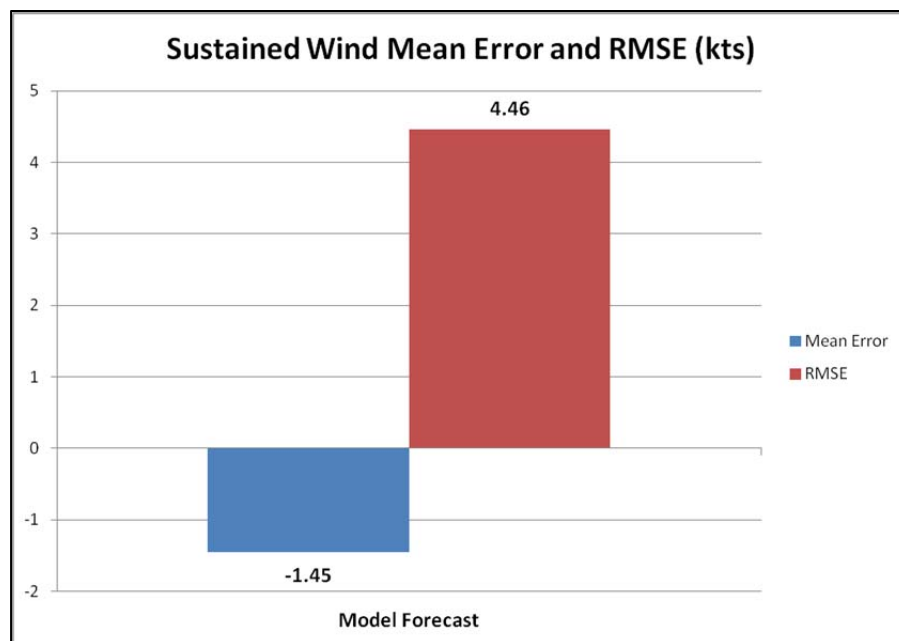
Figure 18.   Sustained wind forecast mean error and RMSE calculations for all locations. Errors calculated using sustained wind speed from observations at forecast valid time. Blue bar indicates mean error, red bar indicates RMSE.

Figure 15 reflects increased accuracy for both algorithms, represented by the RMSE, when forecasts were verified using wind speeds recorded +/- 1.5 hours from the forecast valid time. This analysis was also conducted on a location by location breakout to determine whether the algorithms performance changed for specific locations when evaluated using the RMSE (see Figure 19). For the WGE, three of the five bases showed improvement in RMSE when forecasts were verified against the +/- 1.5 hour observation groups. These RMSE improvements were seen at Westover (-3.56 kts), Andrews (-2 kts) and Scott (-3.29 kts). The increased RMSE values at Langley (+1.26 kts) and Offutt (+.23 kts) were smaller, comparatively. The improvements at the three locations outweighed the small losses at Langley and Offutt. For the RUC algorithm Andrews (-1.56 kts), Langley (-2.68 kts), and Scott (-2.25 kts) all revealed improvements in RMSE. Westover (+.06 kts) and Offutt (+.42 kts) both showed a small loss in accuracy. These results are comparable to the RMSE results from Brasseur (2001) which ranged from roughly 4 kts to 9 kts during a three month simulation in Belgium. Similarly, LaCroix (2002) showed RMSE values of 7 kts to 11 kts. While these two prior studies utilized different verification procedures and data methodology, the similar results indicate the validity of our data methodology and analysis approach.
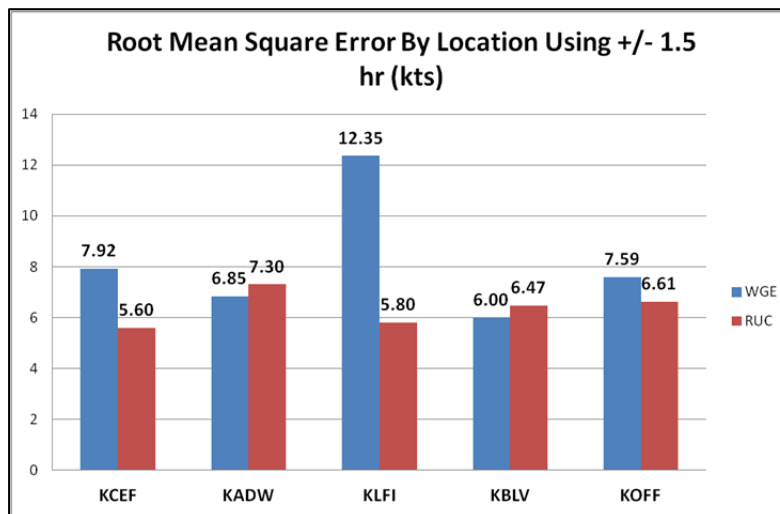
Figure 19.   RMSE error using maximum wind speed from observations +/- 1.5 hours from forecast valid time at each location. Blue bars represent the WGE RMSE in knots, red bars represent the RUC RMSE in knots.

C. LOCATION-BASED RESULTS

A location-based analysis was conducted and the algorithm performance varied at the five locations. This section analyzes the location-based results for a variety of different analyses conducted. The analysis conducted here will be summarized and used as a template for the case studies in Section E of this chapter.

1. Westover ARB

Westover ARB is approximately 45 miles to the east of the Massachusetts/New York border near the south-central portion of the state. Sixty-five miles to the east of the Hudson River that flows through New York, a spine of the Appalachian Mountains is situated between the Hudson River and Westover ARB (Google Earth 2012). As alluded to in Chapter III and shown in Table 5, the proximity of these mountains played a role in the modeled bilinear interpolated grid point elevation for this location.

A difference of 128 m was expected to produce differing results of predicted versus observed wind gusts at this location. Figure 19 shows the WGE RMSE values for Westover to be second largest of the five locations, and the RUC RMSE values are the lowest of the five locations. When the forecasts were compared to the nearest observation only, the WGE RMSE value was the highest of the five locations while the RUC RMSE remained the lowest (not shown).

The sustained and wind gust error histograms were plotted for Westover (see Figure 20). The RUC shows a relatively normal distribution of errors, where 56% of the forecasts errors were within +/- 4 kts. The WGE shows an overestimation of wind gusts (22% of forecast errors >10 kts), similar to what was expected based on higher modeled terrain. The sustained wind forecast errors also represent a relative Gaussian distribution with 70% of the errors within +/- 4 kts. It appears as though the RUC was able to overcome the model grid point resolution deficiency, while the WGE appears to be affected by this elevation difference.
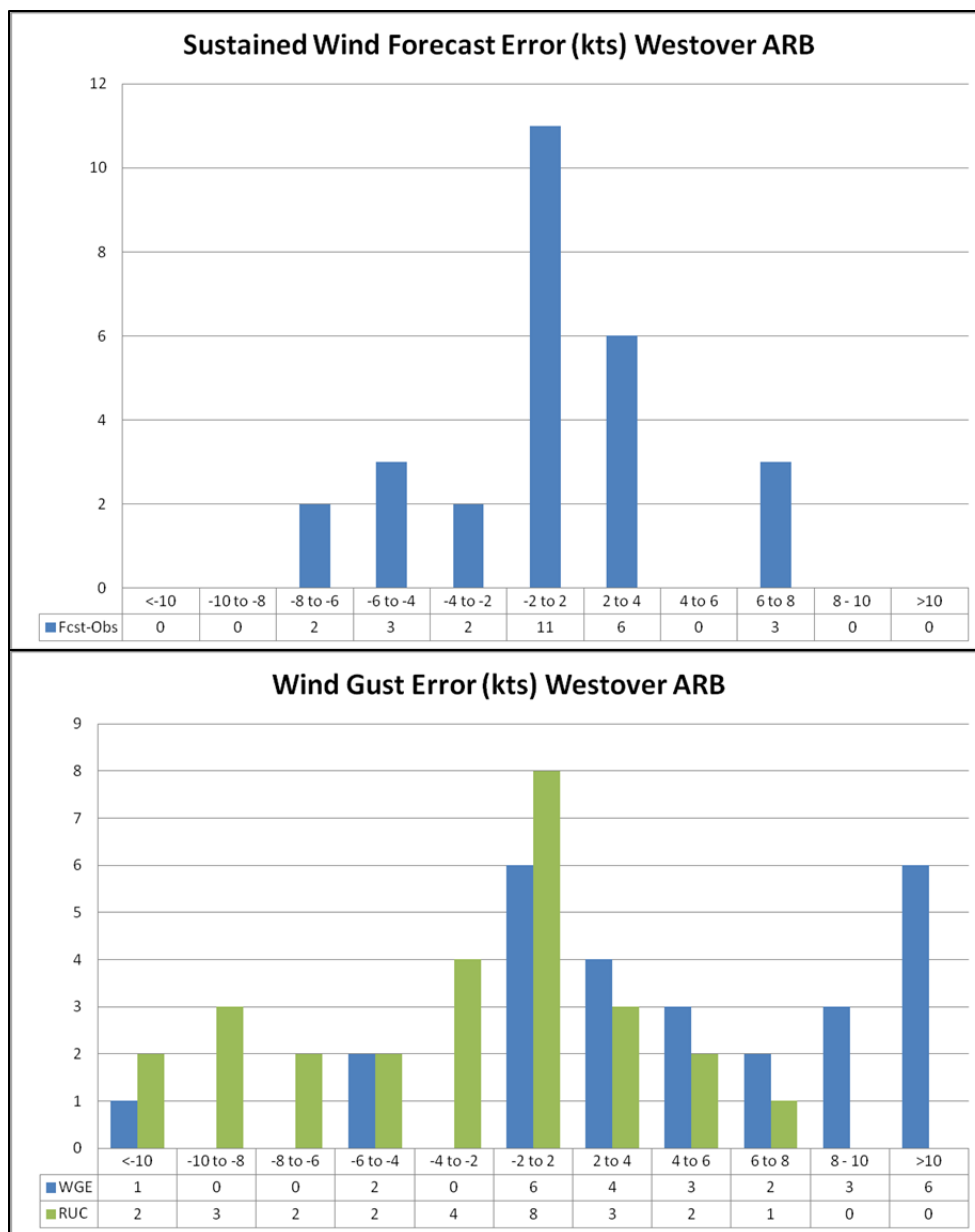
Figure 20.   Sustained (top) and wind gust error (bottom) histograms for Westover ARB. Sustained wind errors (blue bars, top graph) calculated using sustained wind speed from observations at forecast valid time. Wind gust error distribution using maximum wind speed from observations +/- 1.5 hours from forecast valid time. Blue bars represent the number of WGE errors for each bin, green bars represent the number of RUC errors for each bin. Both histogram errors are binned by 2 kt intervals centered on a 4 kt central axis.

There were 27 3-hour forecasts analyzed for Westover using 2x2 contingency tables constructed from the ensemble mean to evaluate the performance of forecasted winds greater than or equal to 35 kts. Overall, the RUC outperformed the WGE in all categories except CSI (-2%) and POD (-57%). The latter result indicates the WGE's ability to predict 35 kt events, however based on the results from the wind gust error histogram, it is likely this result is a product of an overestimation of winds rather than an increase in skill as the skill score comparisons shown in Figure 21 indicate a Kuiper Skill Score of 0.35 for the WGE versus 0.66 for the RUC. Near term statistics (3–24 hour period) reveals the WGE outperforms the RUC in all categories except HR (-9%).
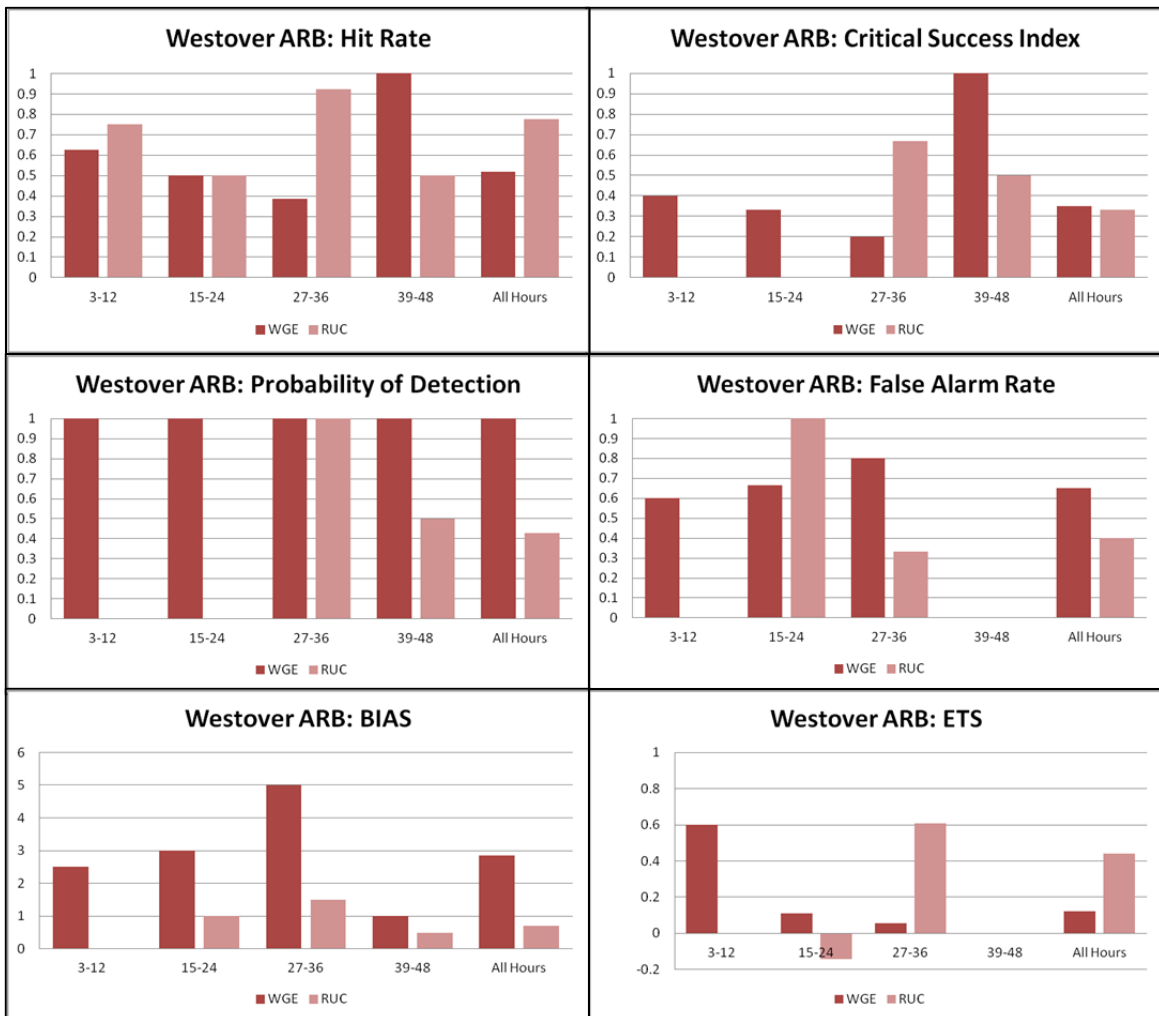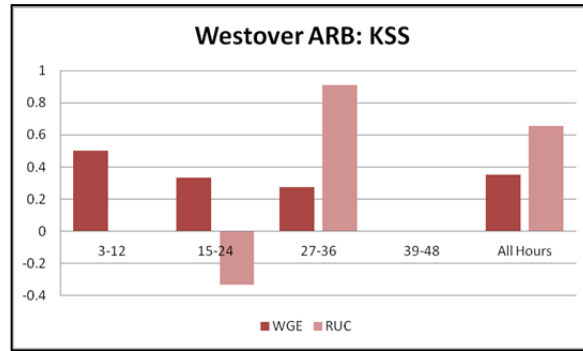
Figure 21.   2x2 contingency table statistics for Westover ARB (27 forecasts). Statistical performance is grouped by forecast lead-time in 12-hour groups. For example, the 3-12 hour bin in each chart represents statistics calculated on forecasts verified between three and 12 hours from model run time. The final group represents statistics calculated on all hours combined. Dark red bars indicate WGE statistics, light red bars indicate RUC statistics.

Another approach in assessing algorithm performance is to look at it from the standpoint of an operational forecaster. That is to say, how well did the algorithms verify forecasts when the prediction was for winds greater than or equal to 35 kts? Even though the forecast may not have verified in the +/- 1.5-hour window, extending the window indicates that timing errors were possible, which could reveal the algorithms accurate prediction of the threshold with just a simple timing error. Since, it is possible that a single observation verified more than one 3-hour forecast when the window was expanded, the total number of verified forecasts combined with the number of misses are not the same for each algorithm.

Figure 22 indicates this analysis for the WGE forecasts at Westover. Of the 27 3-hour forecasts analyzed, 20 forecasts were for winds that met or exceeded the 35 kt threshold. Of those 20 forecasts, seven verified within the +/- 1.5-hour window. One more verified when that window was extended to +/- 3-hours and two more when it was extended to +/- 6-hours. In all, 10 of the 20 forecasts verified within a 12-hour period while another 10 did not verify at all and were counted as false alarms. Furthermore, there were no missed forecasts for this location. This overforecast tendency appears to be a direct result from the modeled elevation difference and not an accurate evaluation of the algorithms as similar results at other locations were not replicated. Figure 23 is the same

analysis but conducted on RUC forecasts at Westover. The number of actual forecasts (five) for the threshold is much lower than WGE reinforcing that the empirical model was not affected as much by the elevation difference. All five verified within the 12-hour forecast window, and there were no false alarms. However, there were four missed forecasts using this algorithm.

This particular analysis was developed to judge the performance of the algorithms along the fine line between acceptable false alarms and misses. Ideally, both would be zero. However, is 10 false alarms and zero misses worse than four misses and zero false alarms? The results at Westover represent the unique and often challenging risk management decision process operational forecasters endure when making critical threshold decisions.
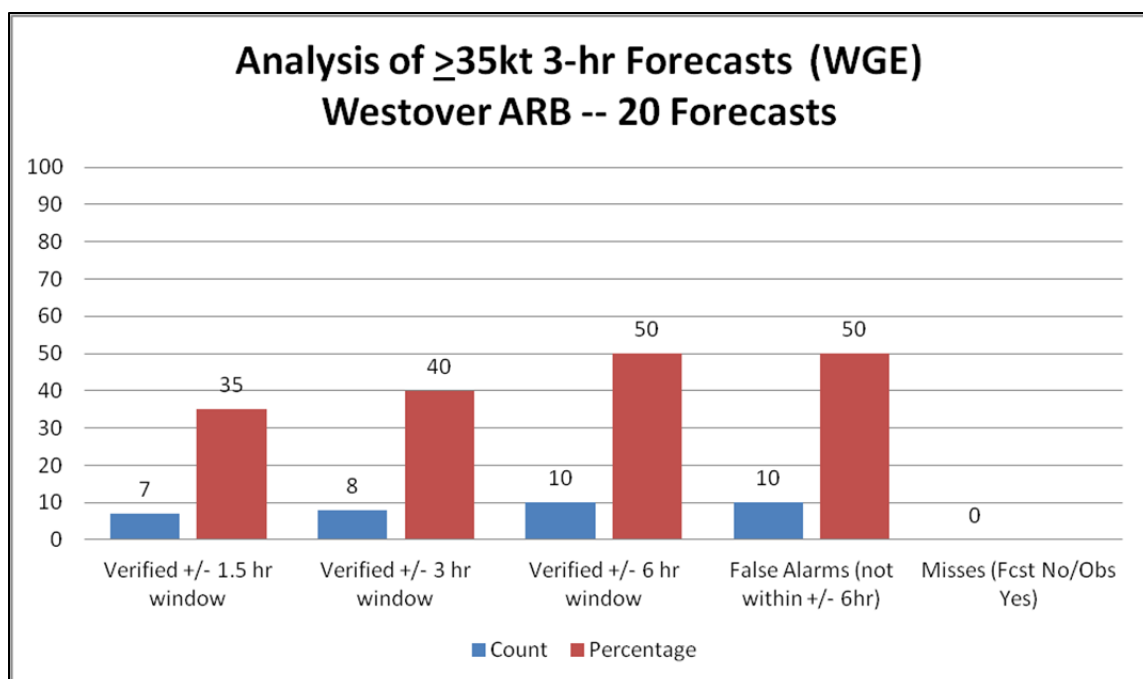
Figure 22. Threshold analysis using the WGE algorithm at Westover ARB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.

56
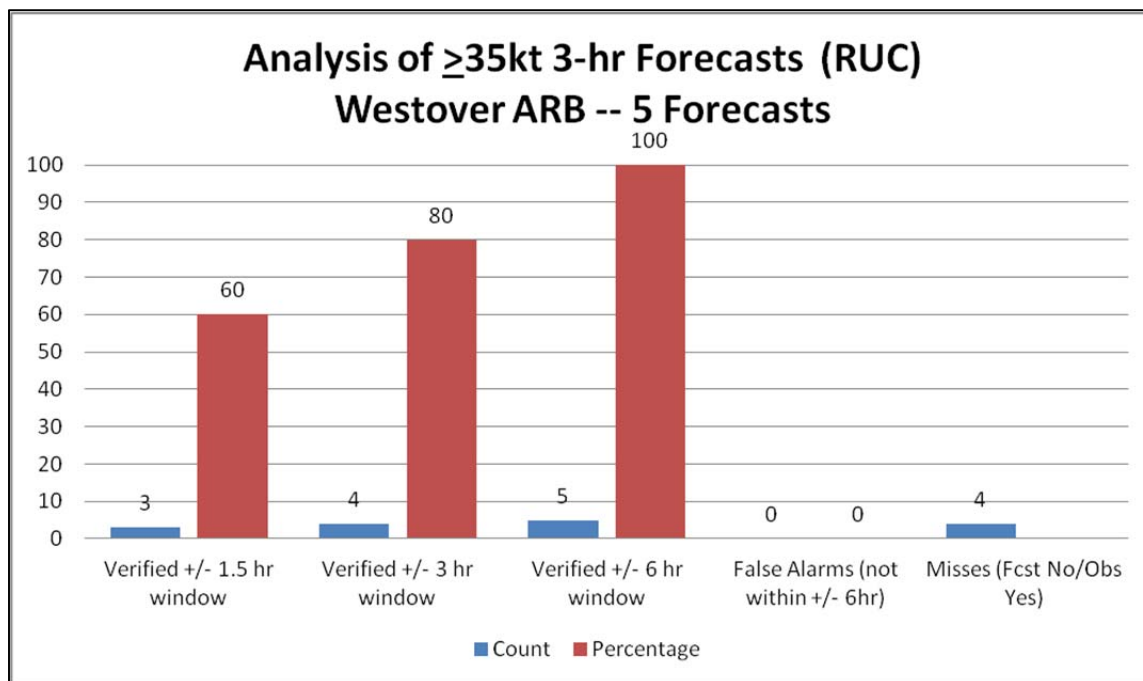
Figure 23.   Threshold analysis using the RUC algorithm at Westover ARB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.

2.   Andrews AFB

Andrews AFB is approximately 10 miles to the southeast of Washington D.C. In between the Potomac River to the west and the Chesapeake Bay to the east, the terrain around the base provides little in the way of wind enhancement with no major terrain differences in the surrounding region (Google Earth 2012). Modeled elevation for this location is approximately 36 m lower than actual elevation. With this difference in elevation, it was expected that wind gust predictions would be slightly underforecast for this base, an opposite effect from what was shown at Westover ARB.

There were 39 forecasts verified at Andrews. WGE RMSE was slightly lower than the RUC by less than .5 kt. Error histograms revealed that 54% of the forecasts were within +/- 4 kts for the WGE compared to only 36% for the RUC. Analysis of both algorithms revealed an underforecast bias, however the RUC was much more underforecast with 77% of the errors resulting in a prediction of >2 kts lower than the observed value (see Figure 24). This bias was also evident in the sustained wind error histogram where 54% of the errors were between +/- 4 kts with 28% between -4 kts and -6 kts (not shown).

Wind Gust Error (kts) Andrews AFB

	<-10	-10 to -8	-8 to -6	-6 to -4	-4 to -2	-2 to 2	2 to 4	4 to 6	6 to 8	8 - 10	>10
WGE	5	4	2	4	3	10	8	0	1	1	1
RUC	6	6	2	8	8	5	1	0	1	2	0
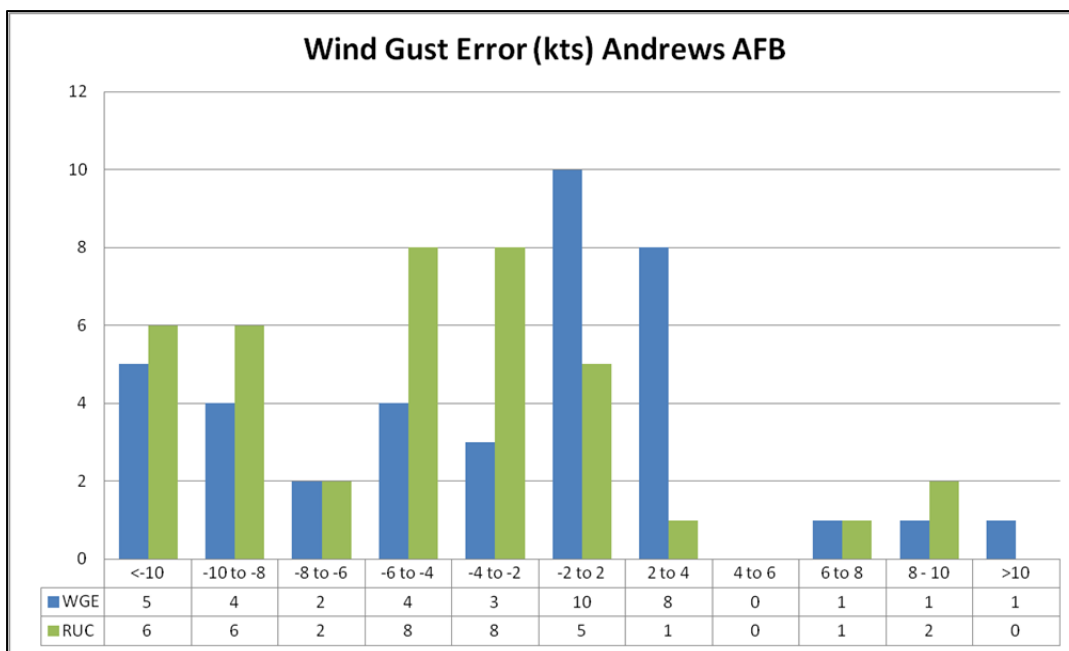
Figure 24.   Wind gust error distribution using maximum wind speed from observations +/- 1.5 hours from forecast valid time. Blue bars represent the number of WGE errors for each bin, green bars represent the number of RUC errors for each bin. Errors are binned by 2 kt intervals centered on a 4 kt central axis.

When all forecast hours were combined, the WGE outperformed the RUC in six of seven of the statistics and skill scores that were evaluated. However, in the short term window of less than 24 hours, the RUC outperformed the WGE in all categories except POD (-33%) but had one more missed forecast compared to the WGE (see Figure 25).

Figure 25. 2x2 contingency table statistics for Andrews AFB (39 forecasts). Statistical performance is grouped by forecast lead-time in 12-hour groups. For example, the 3-12 hour bin in each chart represents statistics calculated on forecasts verified between three and 12 hours from model run time. The final group represents statistics calculated on all hours combined. Dark red bars indicate WGE statistics, light red bars indicate RUC statistics.

Of the 39 3-hour forecasts analyzed, 13 WGE forecasts were for winds that met or exceeded the 35 kt threshold (see Figure 26). Of those 13 forecasts, six verified within the +/- 1.5-hour window. Nine verified when that window was extended to +/- 3-hours and one more when it was extended to +/- 6-hours. In all 10 of the 13 forecasts verified within a 12-hour period while another three were counted as false alarms. Additionally, there were four missed forecasts. The underforecast bias revealed previously was a direct contributor to missing these four forecasts. Eight threshold forecasts were produced using the RUC algorithm (see Figure 27). Four verified within the +/- 1.5-hour forecast window, with another two within the +/- 3-hour window. However, there were six missed forecasts and two false alarms using the RUC algorithm. Analyzing the results solely based on this analysis indicated the RUCs underforecast bias for this location yielded an increase in misses when compared to the WGE.
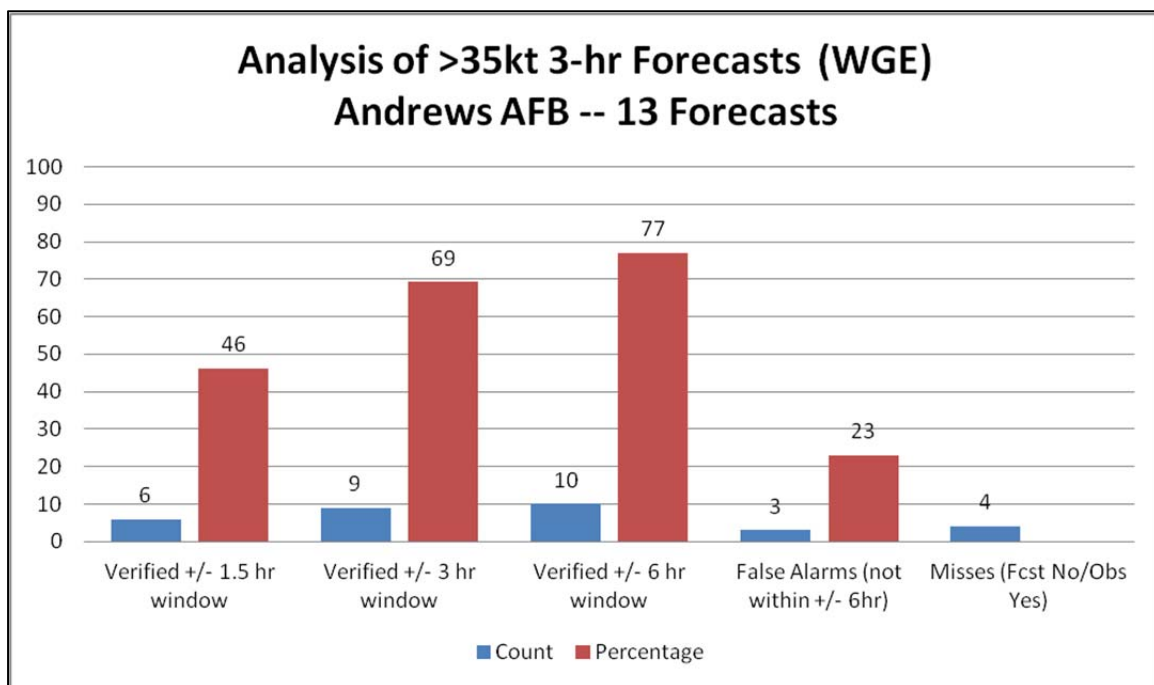
Figure 26.   Threshold analysis using the WGE algorithm at Andrews AFB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.
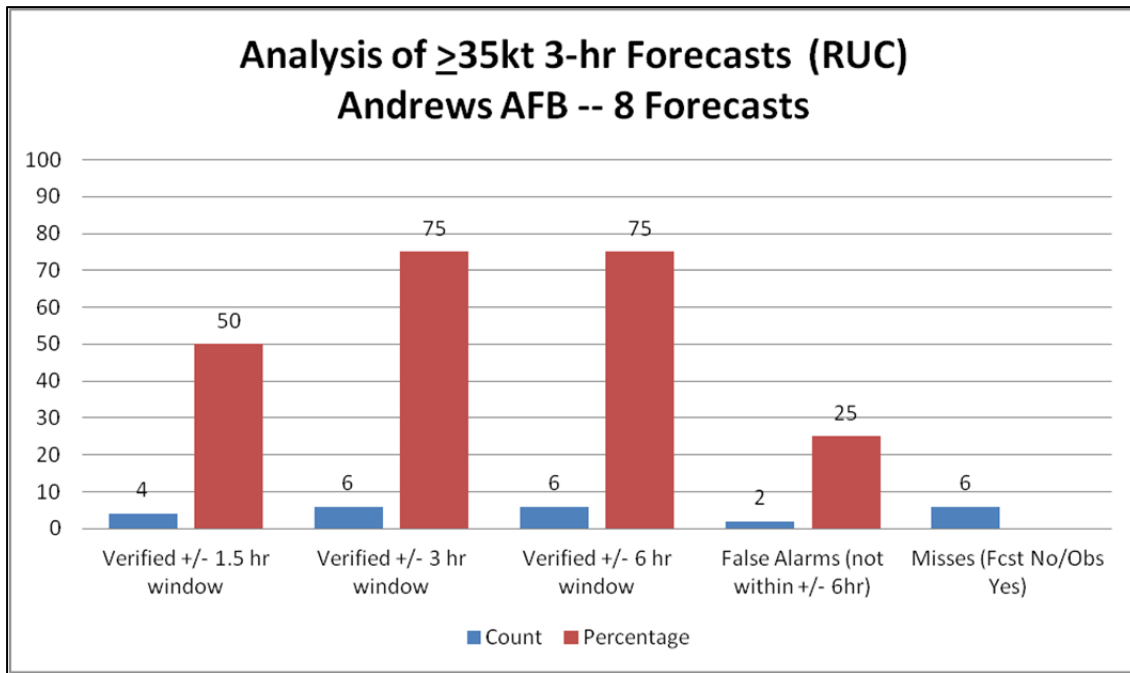
Figure 27.   Threshold analysis using the RUC algorithm at Andrews AFB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.

3.   Langley AFB

Langley AFB is located in the southeastern region of a coastal plain peninsula that extends betweens the James River and Chesapeake Bay in southeastern Virginia. Approximately five miles to the west of the Chesapeake Bay and eight miles to the east of the James River, the weather is influenced by the bodies of water nearby including the Atlantic Ocean which is 25 miles to the east (Google Earth 2012). Modeled elevation for this location is approximately 2 m higher than the actual elevation. Therefore, model predictions of wind gusts are not affected by the interpolated terrain height.

There were 24 forecasts verified at Langley AFB. This number reflected the lowest number of non-convective wind events during the evaluation period for the locations analyzed and is indicative of the lower number of events that occur in the southern portions of the CONUS. RUC RMSE revealed lower error than the WGE by 6.55 kts. Error histograms reiterated the WGE error for this location with a large underforecast bias that included 46% of the forecasts underforecast by greater than 10 kts. 54% of the forecasts were within +/- 4 kts for the RUC compared to only 25% for the WGE (see Figure 28). The sustained wind error histogram depicted accurate predictions with 71% of the errors were between +/- 4 kts (not shown).

Wind Gust Error (kts) Langley AFB

	<-10	-10 to -8	-8 to -6	-6 to -4	-4 to -2	-2 to 2	2 to 4	4 to 6	6 to 8	8 - 10	>10
WGE	11	1	0	2	1	4	1	2	0	1	1
RUC	2	0	1	5	3	7	3	0	1	2	0
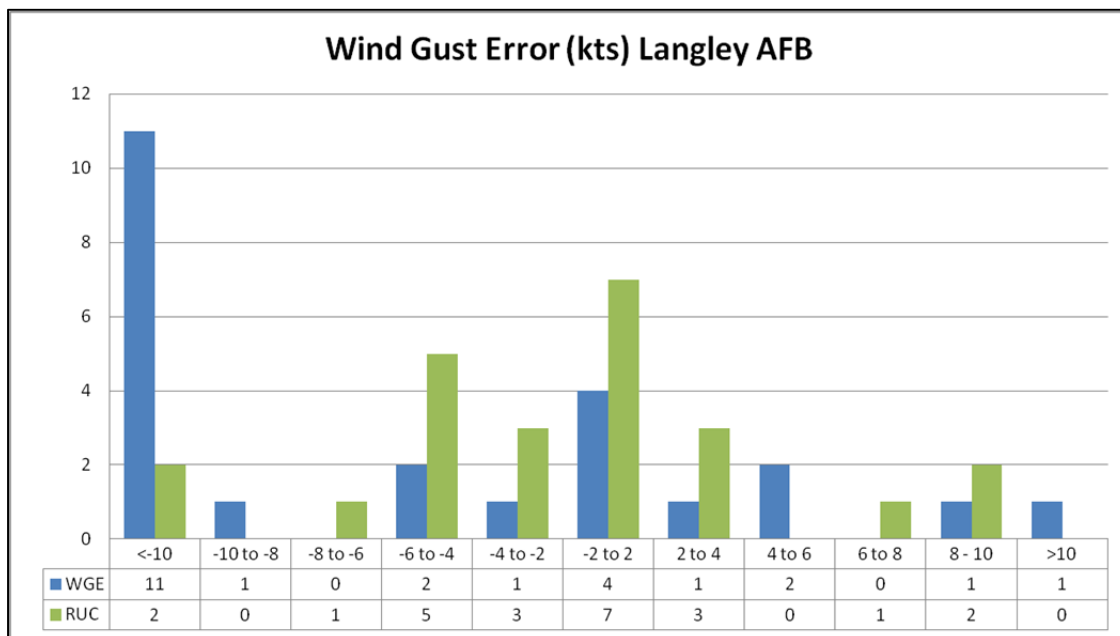
Figure 28.   Wind gust error distribution using maximum wind speed from observations +/- 1.5 hours from forecast valid time. Blue bars represent the number of WGE errors for each bin, green bars represent the number of RUC errors for each bin. Errors are binned by 2 kt intervals centered on a 4 kt central axis.

When all forecast hours were combined, contingency table statistics showed the RUC outperformed the WGE in all seven evaluated statistics and skill scores. This performance also occurred in the short term forecast window of less than 24 hours (see Figure 29).
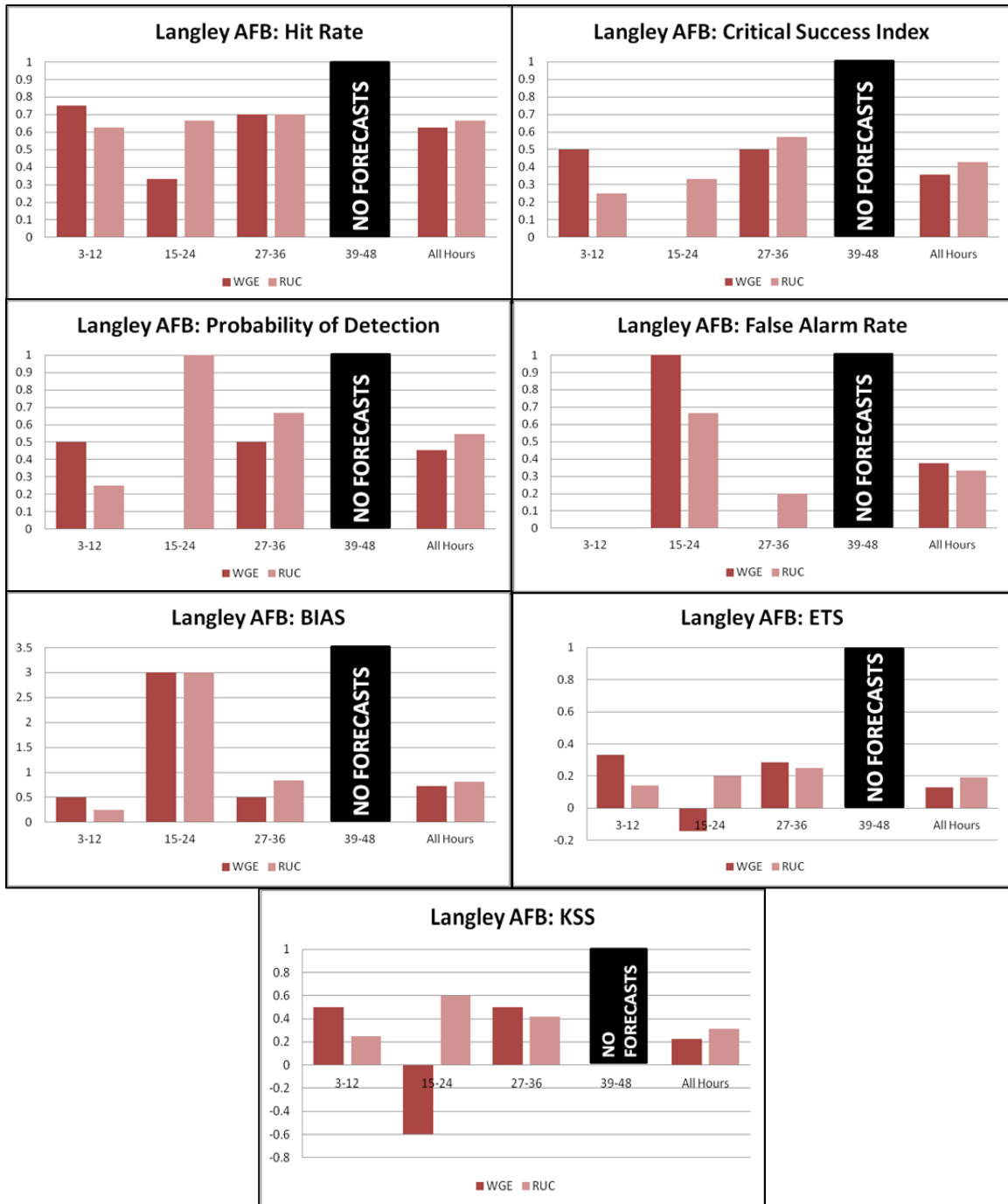
Figure 29.   2x2 contingency table statistics for Langley AFB (24 forecasts).
Statistical performance is grouped by forecast lead-time in 12-hour groups. For example,
the 3-12 hour bin in each chart represents statistics calculated on forecasts verified
between three and 12 hours from model run time. The final group represents statistics
calculated on all hours combined. Dark red bars indicate WGE statistics, light red bars
indicate RUC statistics.

Of the 24 3-hour forecasts analyzed, eight WGE forecasts were for winds that met or exceeded the 35 kt threshold (see Figure 30). Of those eight forecasts, five verified within the +/- 1.5-hour window and a total of six verified when that window was extended to +/- 3-hours. Six of the eight forecasts verified within a 12-hour period while two were counted as false alarms. Furthermore, there were six missed forecasts. The underforecast bias discussed previously was a direct contributor to missing these six forecasts. 10 threshold forecasts were produced using the RUC algorithm (see Figure 31). Seven verified within the +/- 1.5-hour forecast window, with another two within the +/- 3-hour window and all 10 verified within the +/- 6-hour window. Although there were four missed forecasts, the RUC algorithm was the best method of prediction for this location.
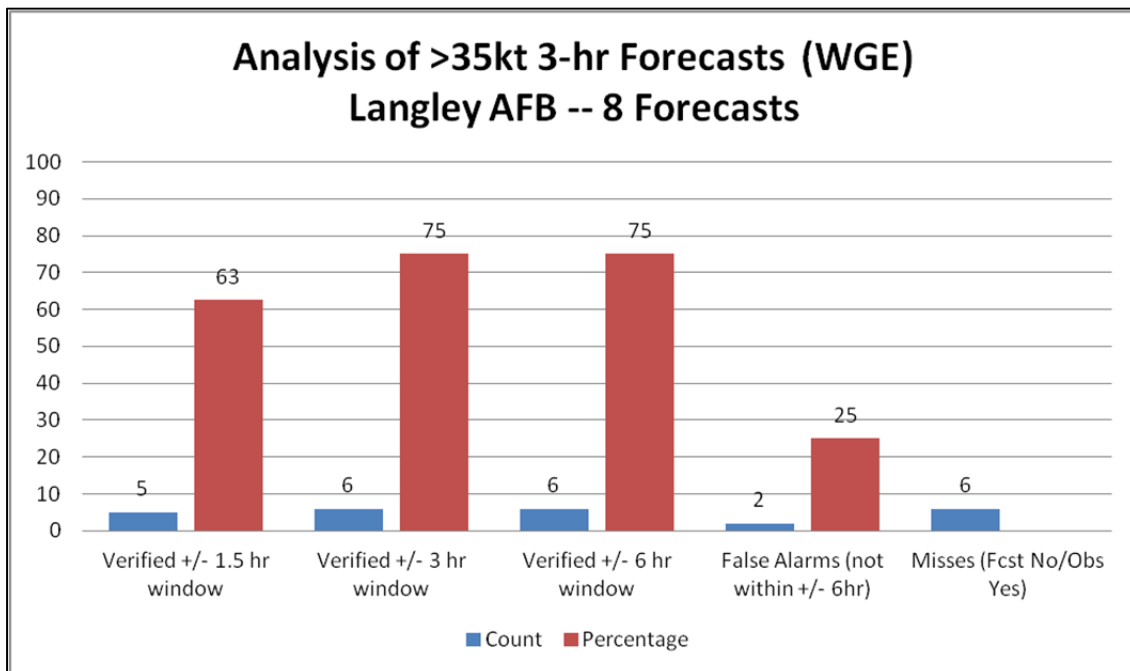
Figure 30.   Threshold analysis using the WGE algorithm at Langley AFB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.
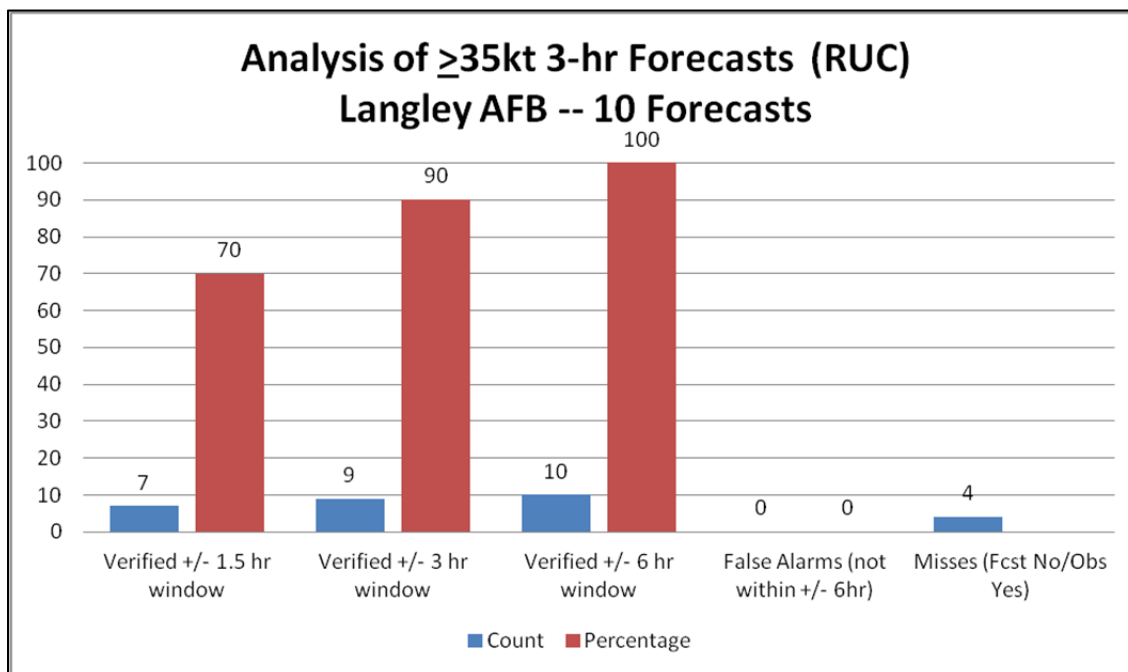
Figure 31.   Threshold analysis using the RUC algorithm at Langley AFB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.

4.   Scott AFB

Scott AFB is located in southwestern Illinois, approximately 20 miles to the east of St. Louis, Missouri and the Mississippi River (Google Earth 2012). Relatively flat terrain exists throughout much of the nearby region with little in way of terrain enhancement to wind speeds. Modeled elevation for this location was approximately 6 m higher than actual elevation. Therefore, model predictions of wind gusts were not influenced by errors due to the small difference in interpolated terrain height.

There were 28 forecasts verified at Scott. RMSE calculations for both algorithms were almost identical with less than .5 kt separating the WGE from the RUC. Error histograms revealed a relative normal distribution for the WGE error for this location with a slight overforecast bias. The RUC distribution was less normally distributed with a clear underforecast bias. 50% of the forecasts were within +/- 4 kts with 43% overforecast by greater than 2 kts for the WGE as compared to 39% within +/- 4 kts and 54% underforecast by greater than 2 kts for the RUC (see Figure 32). The sustained wind error histogram revealed accurate predictions of wind gusts as 64% of the errors were between +/- 4 kts with a definitive underforecast bias. Only two of the 28 sustained wind forecasts verified with an error greater than 2 kts. (not shown).

Wind Gust Error (kts) Scott AFB

	<-10	-10 to -8	-8 to -6	-6 to -4	-4 to -2	-2 to 2	2 to 4	4 to 6	6 to 8	8 - 10	>10
WGE	0	1	3	1	3	8	3	4	2	1	2
RUC	0	2	4	6	3	5	3	0	2	1	2
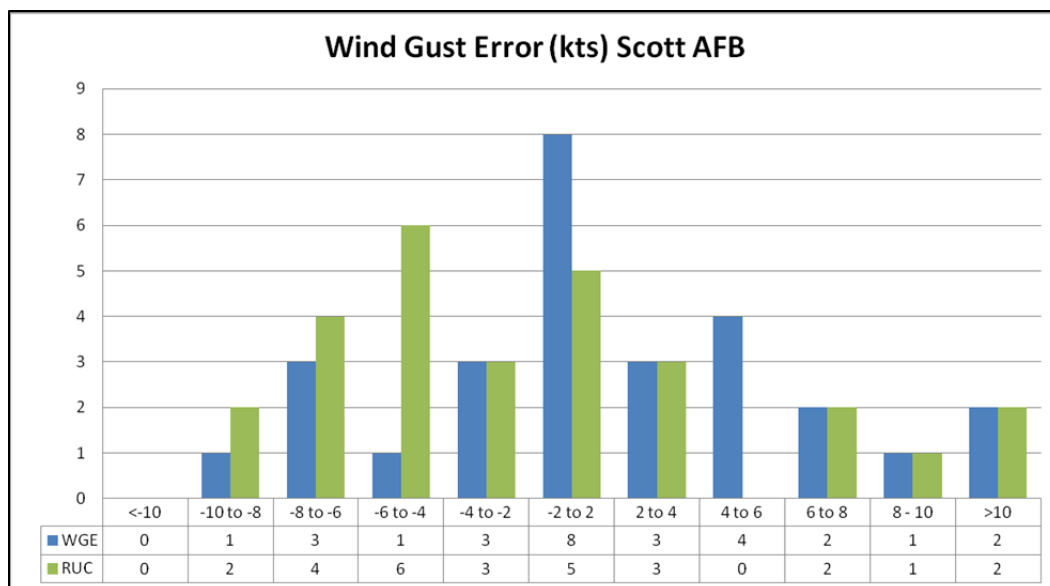
Figure 32. Wind gust error distribution using maximum wind speed from observations +/- 1.5 hours from forecast valid time. Blue bars represent the number of WGE errors for each bin, green bars represent the number of RUC errors for each bin. Errors are binned by 2 kt intervals centered on a 4 kt central axis.

When all forecast hours were combined, contingency table statistics showed the WGE outperformed the RUC in all categories except bias (-.88). This performance also occurred in the short term forecast window of less than 24 hours (see Figure 33). When all statistic and error calculations were examined at Scott, the best predictions were produced by the WGE, which was also seen in the analysis of the 35 kt wind forecasts.
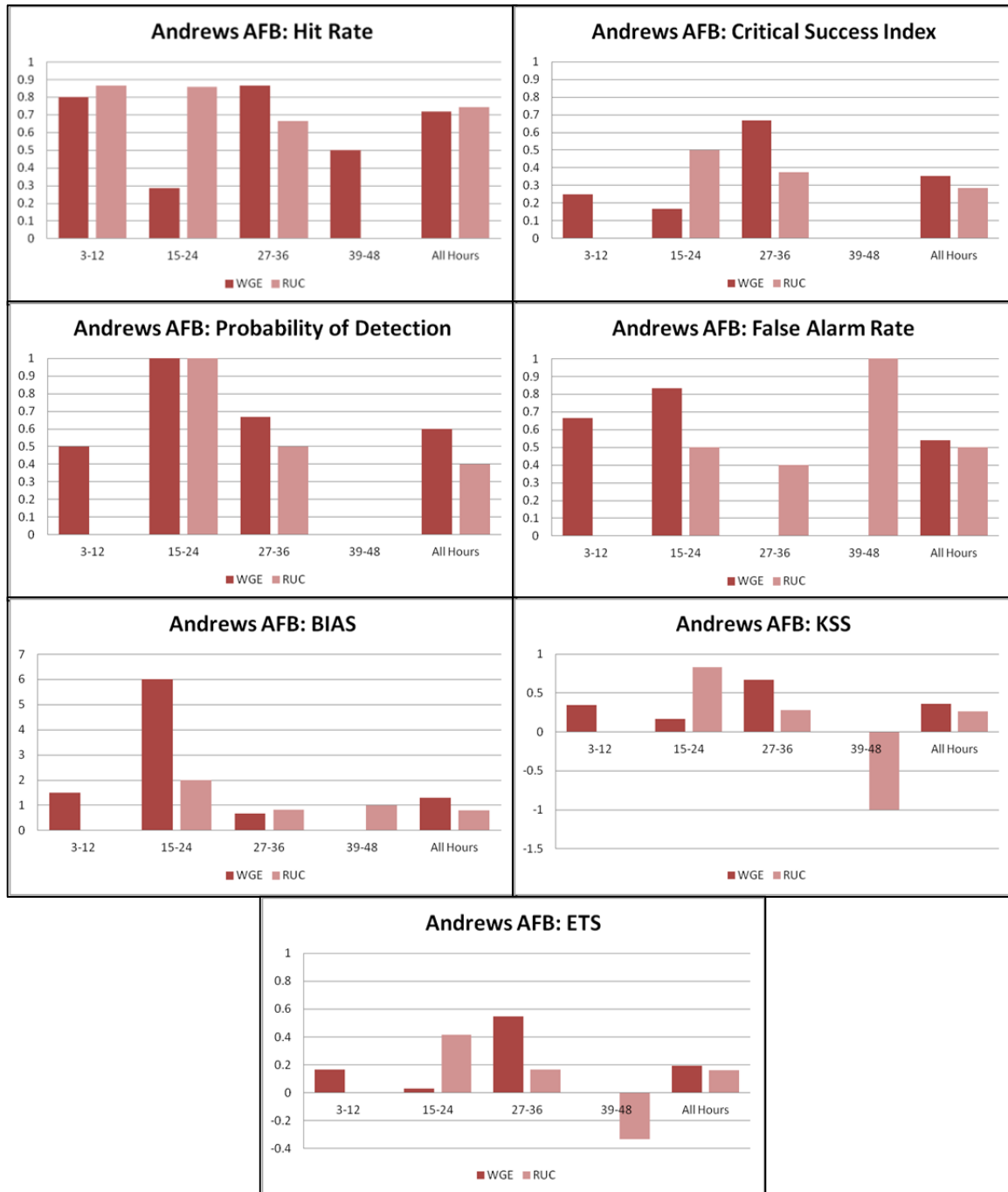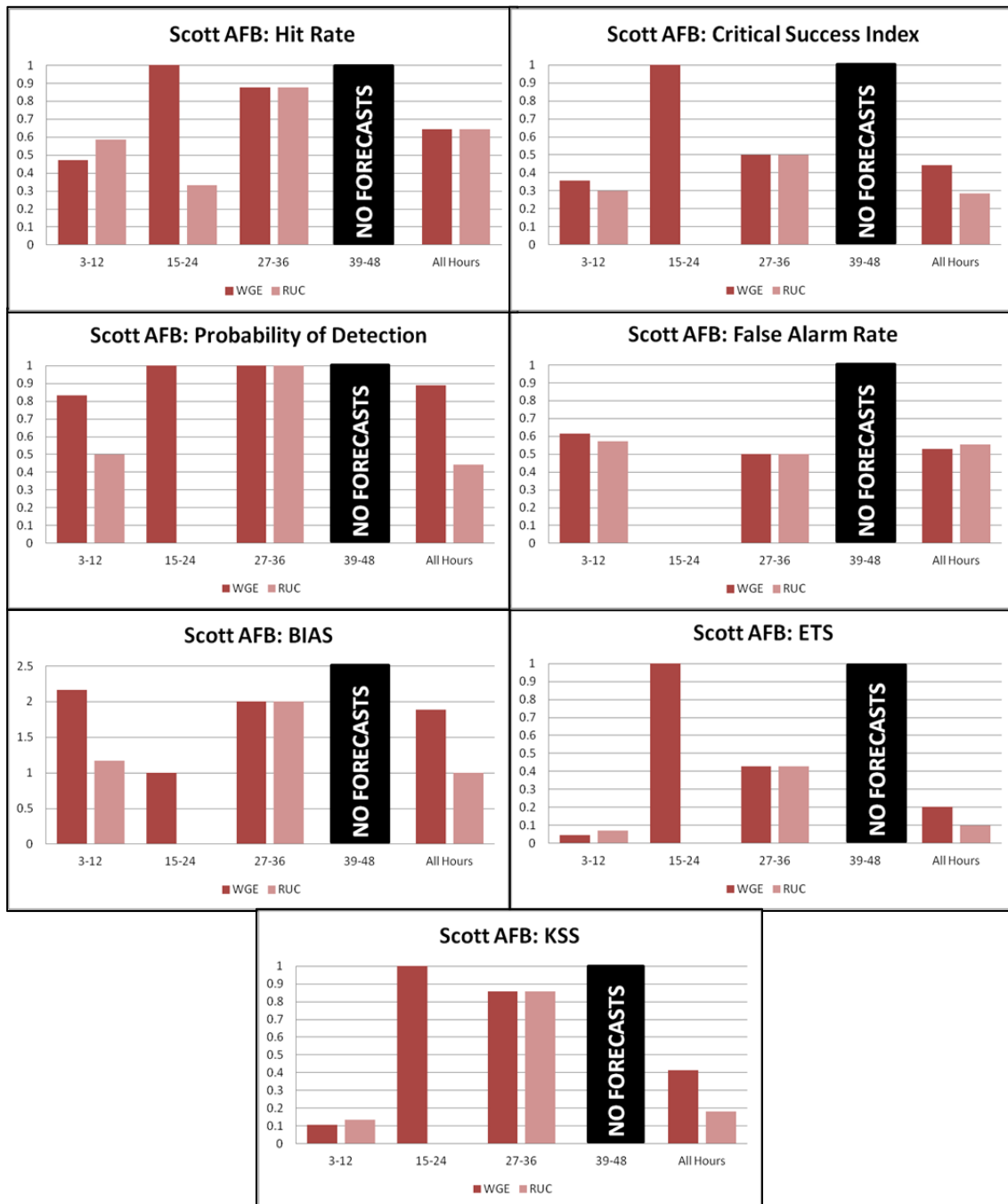
Figure 33.   2x2 contingency table statistics for Scott AFB (28 forecasts). Statistical performance is grouped by forecast lead-time in 12-hour groups. For example, the 3-12 hour bin in each chart represents statistics calculated on forecasts verified between three and 12 hours from model run time. The final group represents statistics calculated on all hours combined. Dark red bars indicate WGE statistics, light red bars indicate RUC statistics.

Of the 28 3-hour forecasts analyzed, 17 WGE forecasts met or exceeded the 35 kt threshold (see Figure 34). Of those 17 forecasts, eight verified within the +/- 1.5-hour window and a total of 10 verified when that window was extended to +/- 3-hours. The remaining seven forecasts were attributed to false alarms and there was one missed forecast. Nine threshold forecasts were produced using the RUC algorithm (see Figure 35). Four verified within the +/- 1.5-hour forecast window, with another two within the +/- 3-hour window and a total of six verified within the +/- 6-hour window. This resulted in three false alarm forecasts and five missed forecasts. The five missed forecasts were a result of the underforecast bias. Although there were seven false alarm and one missed forecasts, the statistics prove the WGE algorithm was the best method of prediction for this location.

Figure 34.   Threshold analysis using WGE at Scott AFB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.
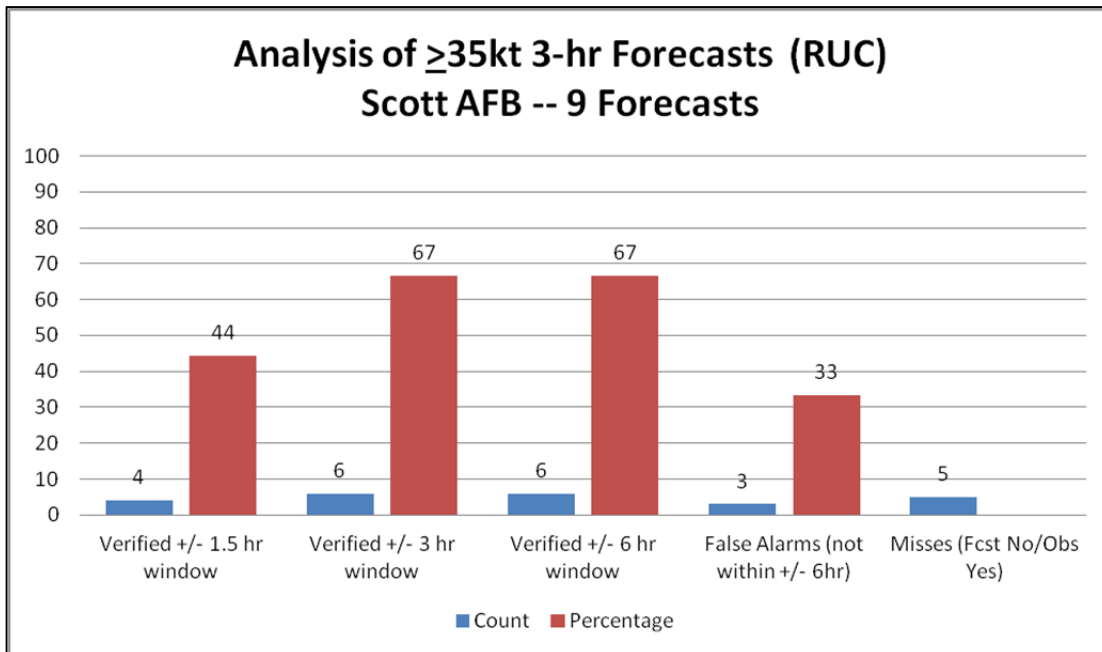
Figure 35. Threshold analysis using RUC at Scott AFB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.

5.   **Offutt AFB**

Offutt AFB is located in southeastern Nebraska, approximately 10 miles to the southeast of Omaha and approximately three miles from the Nebraska/Iowa border (Google Earth 2012). Eastern Nebraska into Iowa is dominated by flat terrain with little in the way of terrain changes and is often identified geographically as the High Plains. Modeled elevation for this location was approximately 24 m higher than actual elevation. Only small overforecast errors were expected due to this interpolation difference. The error analysis below reveals a primarily underforecast bias despite the expected outcome.

There were 39 forecasts verified at Offutt which tied with Andrews for the most forecasts verified during the evaluation period. This correlates well with previous research findings that show the High Plains to be a region with a high concentration of non-convective wind events. RMSE calculations for both algorithms were fairly similar with less than 1 kt separating the WGE from the RUC.   Error histograms revealed a relative Gaussian distribution for both algorithms. The WGE error distribution was more

normally distributed about the center of the axis for errors between -2 kts and 2 kts. The RUC errors were more distributed about the axis centered about errors from -2 kts to -4 kts indicating an underforecast bias for this location. 46% of the forecasts were within +/- 4 kts with 41% overforecast by greater than 2 kts for the WGE as compared to 44% within +/- 4 kts and 72% underforecast by greater than 2 kts for the RUC (see Figure 36). The sustained wind error histogram depicted accurate predictions with 64% of the errors between +/- 4 kts with an underforecast bias as only eight of the 39 sustained wind forecasts verified with an error greater than 2 kts. (not shown).

Wind Gust Error (kts) Offutt AFB

	<-10	-10 to -8	-8 to -6	-6 to -4	-4 to -2	-2 to 2	2 to 4	4 to 6	6 to 8	8 - 10	>10
WGE	7	3	2	1	3	11	4	3	3	2	0
RUC	5	5	2	7	9	7	1	2	0	1	0
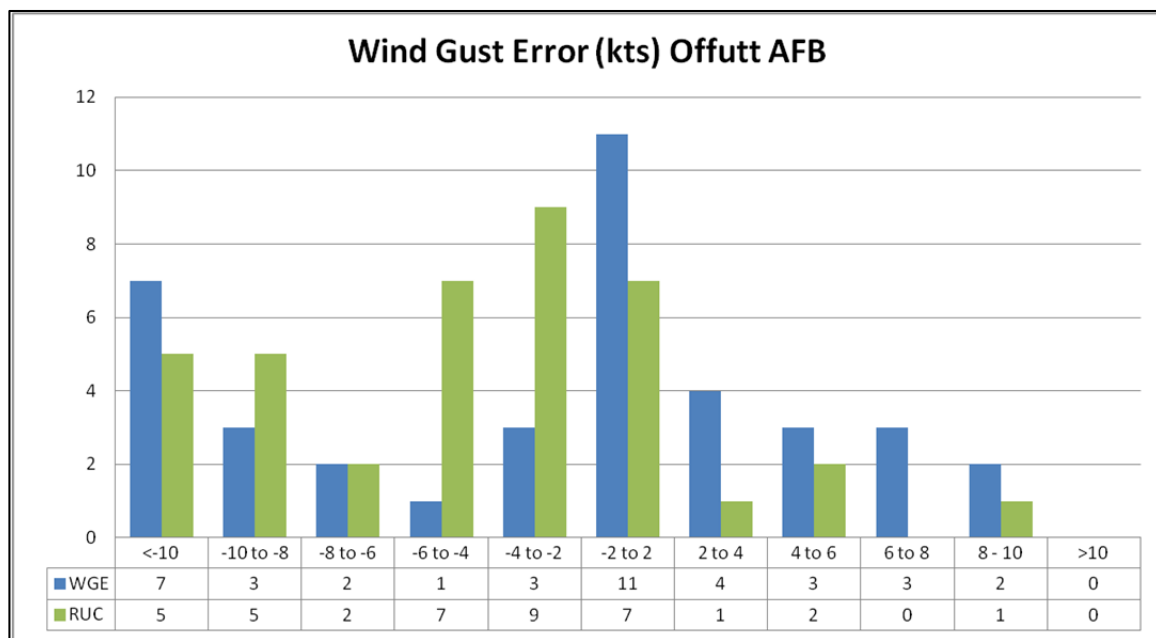
Figure 36.   Wind gust error distribution using maximum wind speed from observations +/- 1.5 hours from forecast valid time. Blue bars represent the number of WGE errors for each bin, green bars represent the number of RUC errors for each bin. Errors are binned by 2 kt intervals centered on a 4 kt central axis.

When all forecast hours were combined, contingency table statistics showed the RUC outperformed the WGE in all categories except bias (-.42). In the short term forecast window of less than 24 hours, the RUC outperformed the WGE in all categories (see Figure 37).   When all statistic and error calculations were examined, the best algorithm predictions were produced by the RUC.
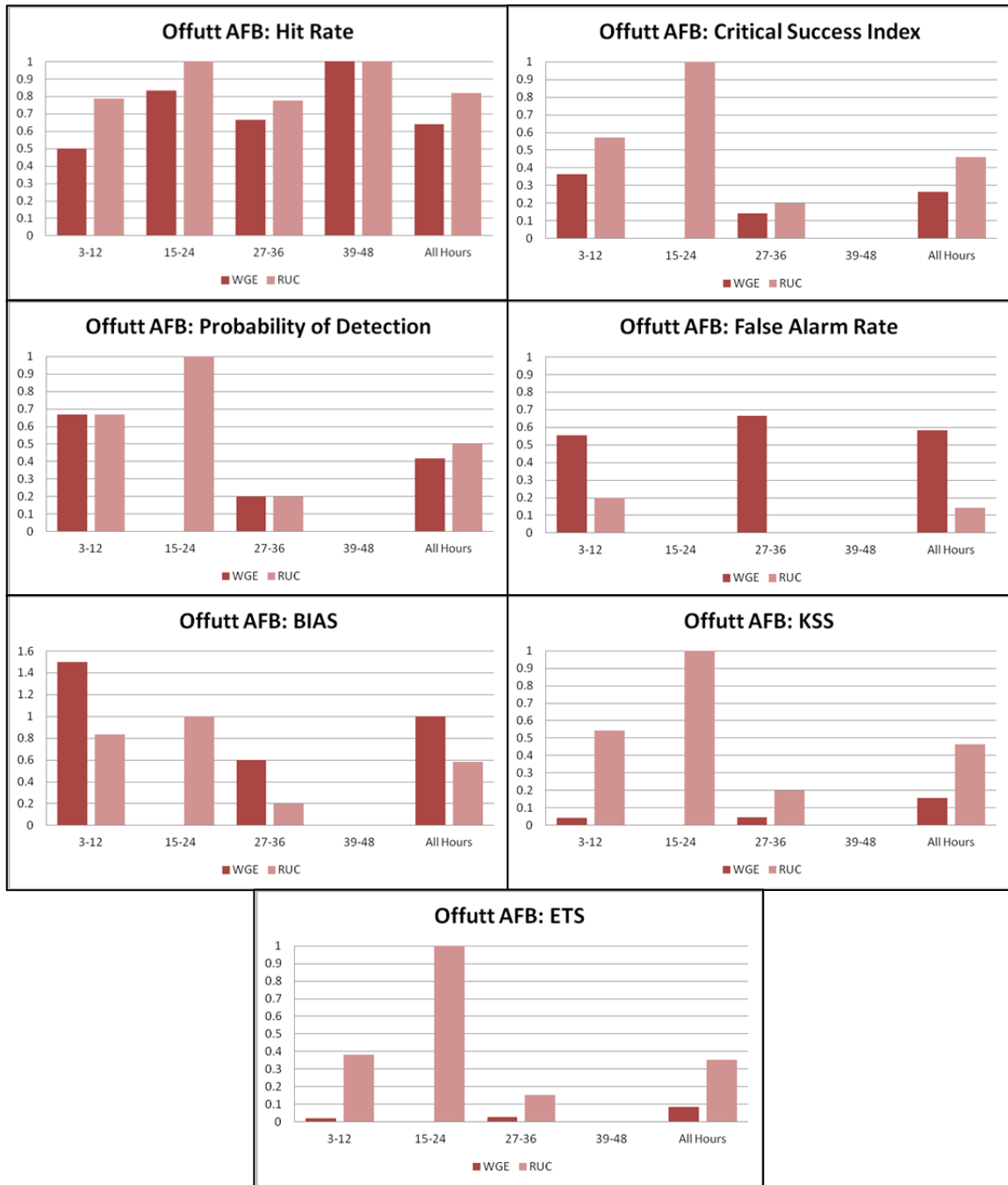
Figure 37.   2x2 contingency table statistics for Offutt AFB (39 forecasts). Statistical performance is grouped by forecast lead-time in 12-hour groups. For example, the 3-12 hour bin in each chart represents statistics calculated on forecasts verified between three and 12 hours from model run time. The final group represents statistics calculated on all hours combined. Dark red bars indicate WGE statistics, light red bars indicate RUC statistics.

Of the 39 3-hour forecasts analyzed, 12 WGE forecasts met or exceeded the 35 kt threshold (see Figure 38). Of those 12 forecasts, five verified within the +/- 1.5-hour window, two more when the window was extended to +/- 3-hours and a total of eight verified when that window was extended to +/- 6-hours. The remaining four forecasts were attributed to false alarms and there were seven missed forecasts. Seven threshold forecasts were produced using the RUC algorithm (see Figure 39). Six verified within the +/- 1.5-hour forecast window, with the seventh verifying within the +/- 6-hour window. This resulted in no false alarm forecasts but six forecasts were missed. This performance metric is somewhat concerning considering the results from the 2x2 contingency table statistics. While the decrease of false alarms to zero is a performance advantage, the decrease in missed forecasts by one forecast using the RUC algorithm does not produce strong confidence in either algorithms ability to detect or properly alert to the given threshold.
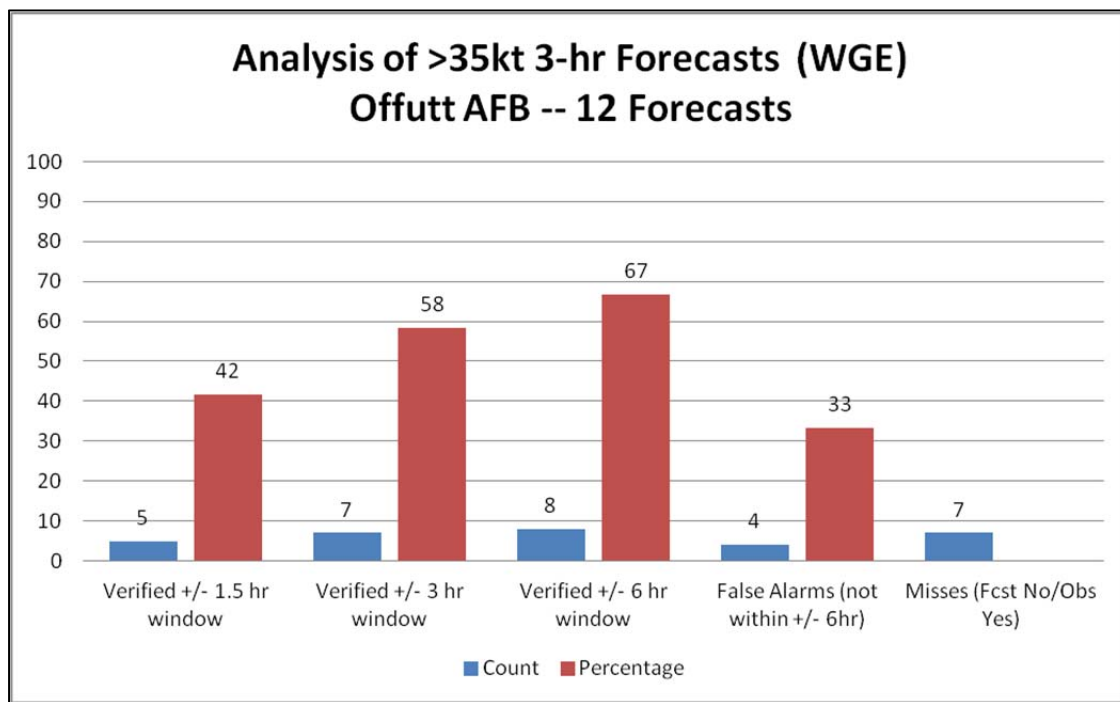
Figure 38. Threshold analysis using WGE at Offutt AFB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.
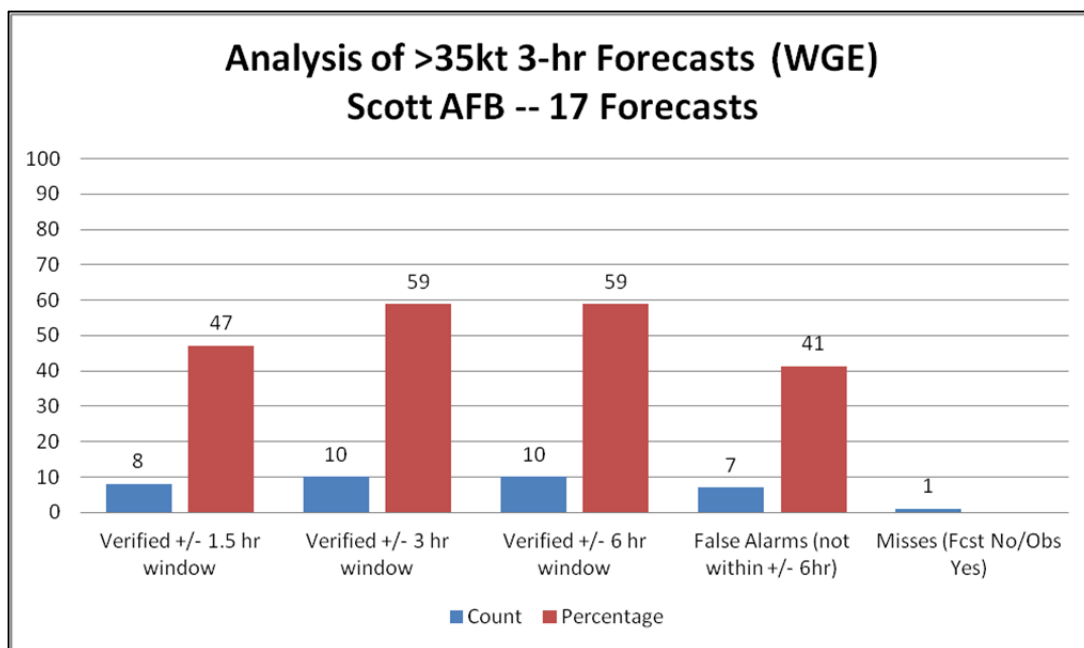
72
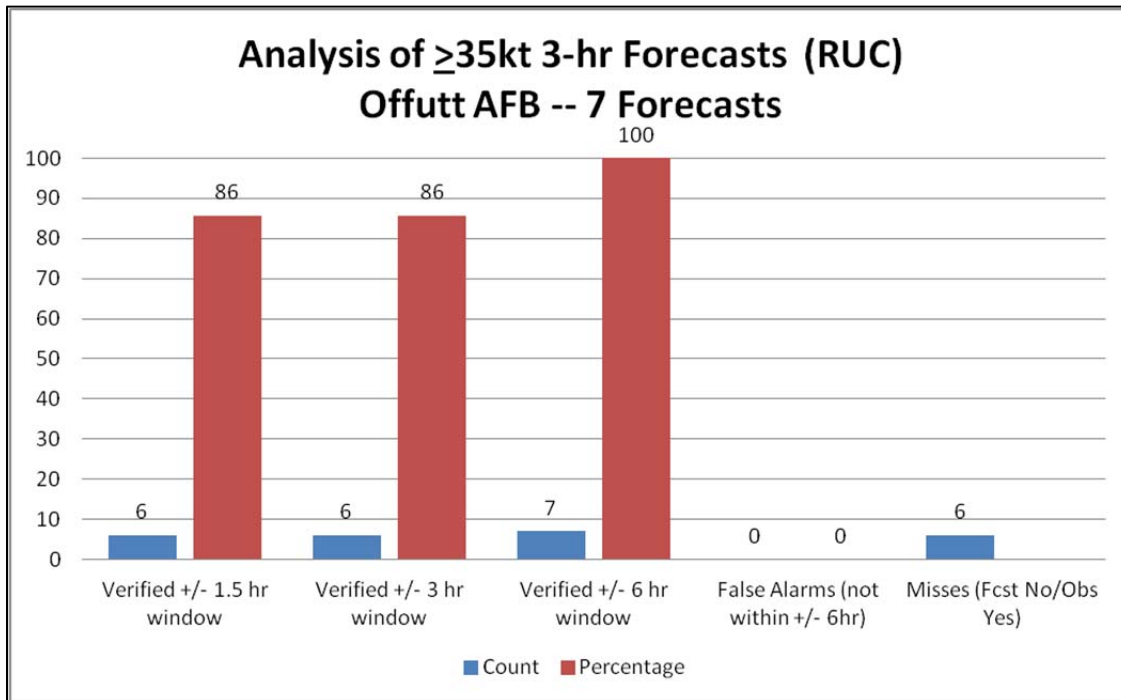
Figure 39.   Threshold analysis using RUC at Offutt AFB. Blue bars represent number of forecasts verified in each window (3, 6, or 12 hours), false alarms or missed forecasts. Red bars represent associated percentages of overall forecasts in each window or false alarm columns.

D.   LINEAR REGRESSION ANALYSIS

1.   Method

A simple best fit linear regression was applied to each algorithms results for each location. Similar to the explanation given by Nordström (2005) for tuning the algorithm to their data set, this method was designed to improve results of the analyzed data in order to determine whether this method is viable for use in the field. However, this does not provide an explanation of the effects of model tuning on the physical parameters forecast by the model.

The linear regression was accomplished by applying the best fit line to the data using the same results (i.e., 3-hour forecasts compared to observation groups) used in this chapter to produce a new prediction of maximum wind gust. RMSE and new 2x2 contingency table statistics were produced for each algorithm at each location to determine the increase in accuracy provided by the linear regression. For calculations of

RMSE, the "Leave One Out Cross Validation Method" was used to help ensure over fitting did not occur. See Moore (2012) for more information on this and other methods to detect and prevent overfitting of data. Each location summary also details the magnitude of the changed forecasts that became worse than the original forecast. The purpose of this analysis is to show that most forecasts changed slightly (≤3 kts) using the best fit, although this could be the difference in crossing the 35 kt threshold. These tools, along with the linear regression equations, were applied to independent case studies to assess the validity of using this method to tune the algorithms in order to produce more accurate results.

2. Results

a. Westover ARB

Table 6 shows the resulting linear regression equations for the two algorithms. The regression equations were applied to the original 27 forecasts. The WGE regression resulted in 21 improved forecasts and six worse forecasts. Of the six worse forecasts, three changed by 3 kts or less and the other three between 8 kts and 12 kts, a fairly large change that attempted to reduce the outliers to the mean. Misses also increased from zero from the original analysis to four when the regression analysis was applied as four original 35 kt forecasts were adjusted below the threshold. When new contingency table statistics were computed, four of the seven categories (HR, FAR, Bias, ETS) improved, while the remaining three categories worsened (see Table 7). The RUC regression resulted in 13 improved forecasts, seven worse forecasts (all changed ≤ 3 kts) and seven forecasts that did not change. No change was apparent in any of the contingency table statistics or in the number of misses from original analysis for the RUC. Figure 40 reflects the RMSE comparison for each algorithm.

WGE	**RUC**
y=.2456x + 24.071	**y=.3091x + 23.881**

Table 6. Linear regression equations for Westover ARB computed by applying best fit line to observed wind gust versus forecast data pairs.

	WGE	Adj WGE	RUC	Adj RUC
HR	0.52	0.70	0.78	0.78
CSI	0.35	0.27	0.33	0.33
POD	1.00	0.43	0.43	0.43
FAR	0.65	0.57	0.40	0.40
BIAS	2.86	1.00	0.71	0.71
KSS	0.35	0.23	0.66	0.66
ETS	0.12	0.13	0.44	0.44

Table 7.   Post-regression statistics for Westover ARB. The seven statistical categories analyzed were Hit Rate (HR), Critical Success Index (CSI), Probability of Detection (POD), False Alarm Rate (FAR), Bias, Kuiper Skill Score (KSS) and Equitable Threat Score (ETS). Green cells represent an increase in statistical accuracy when compared to the original value, red cells represent a decrease in statistical accuracy.

Figure 40.   RMSE comparison for Westover ARB. Blue bars represent the original RMSE value, red bars represent the RMSE values after linear regression was applied and green bars represent the cross-validated RMSE values.

b.   *Andrews AFB*

Table 8 details the linear regression equations for Andrews AFB. The regression equations were applied to the original 39 forecasts. The WGE regression resulted in 24 better wind gust predictions, nine worse predictions and six forecasts that were unchanged. The nine worse forecasts were a result of six changing by ≤3 kts and

three changing between 4 kts and 7 kts, however the number of misses remained the same at four. When new contingency table statistics were computed, they remain unchanged from the original statistics (see Table 9). The RUC regression resulted in 29 improved predictions, six worse predictions (five changed by ≤3 kts and one changed by 4 kts) and four forecasts that did not change. Five of the seven categories (CSI, POD, Bias, KSS, ETS) improved, while HR and FAR worsened slightly by less than 5%. Figure 41 reflects the RMSE comparison for each algorithm.

WGE	**RUC**
y=.2481x + 25.8181	**y=.2527x + 26.245**

Table 8.   Linear regression equations for Andrews AFB computed by applying best fit line to observed wind gust versus forecast data pairs.

	WGE	**Adj WGE**	**RUC**	**Adj RUC**
HR	0.72	0.72	0.74	0.72
CSI	0.35	0.35	0.29	0.31
POD	0.60	0.60	0.40	0.50
FAR	0.54	0.54	0.50	0.55
BIAS	1.30	1.30	0.80	1.10
KSS	0.36	0.36	0.26	0.29
ETS	0.20	0.20	0.16	0.17

Table 9.   Post-regression statistics for Andrews AFB. The seven statistical categories analyzed were Hit Rate (HR), Critical Success Index (CSI), Probability of Detection (POD), False Alarm Rate (FAR), Bias, Kuiper Skill Score (KSS) and Equitable Threat Score (ETS). Green cells represent an increase in statistical accuracy when compared to the original value, red cells represent a decrease in statistical accuracy.

Figure 41.   RMSE comparison for Andrews AFB. Blue bars represent the original RMSE value, red bars represent the RMSE values after linear regression was applied and green bars represent the cross-validated RMSE values.

c.   *Langley AFB*

Table 10 shows the linear regression equations for Langley AFB applied to the 24 forecasts. The extremely flat slope of the WGE regression indicates a poor fitting of the data (Wilks 2006). The WGE regression resulted in 17 better forecasts, five worse predictions and two unchanged forecasts. The five changed forecasts ranged from ≤3 kts (three) to 5 kts (two).   The main advantage became apparent when missed threshold event forecasts were assessed. Overall misses decreased from six to one using the WGE at this location even though the fit is shown to be poor. When new contingency table statistics were computed, all but CSI and POD worsened (see Table 11). However, POD doubled from 45% to 90% after the regression was applied. The RUC regression resulted in 13 improved predictions, eight worse forecasts (six by ≤3 kts and two by 4 kts) and three predictions that did not change. Misses also decreased from five to three. Similar to the WGE statistics, after regression was applied all statistical categories worsened except CSI (+4%) and POD (+18%). Figure 42 reflects the RMSE comparison for each algorithm. Of particular note, when WGE regression was applied, RMSE was

cut by 60% which is likely resulted in the increased accuracy of detection of threshold events as indicated by the new POD.

WGE	RUC
y=.0708x + 33.352	y=.417x + 21.259

Table 10.   Linear regression equations for Langley AFB computed by applying best fit line to observed wind gust versus forecast data pairs.

	WGE	Adj WGE	RUC	Adj RUC
HR	0.63	0.46	0.67	0.63
CSI	0.36	0.43	0.43	0.47
POD	0.45	0.91	0.55	0.73
FAR	0.38	0.55	0.33	0.43
BIAS	0.73	2.00	0.82	1.27
KSS	0.22	-0.01	0.31	0.27
ETS	0.13	-0.01	0.19	0.15

Table 11.   Post-regression statistics for Langley AFB. The seven statistical categories analyzed were Hit Rate (HR), Critical Success Index (CSI), Probability of Detection (POD), False Alarm Rate (FAR), Bias, Kuiper Skill Score (KSS) and Equitable Threat Score (ETS). Green cells represent an increase in statistical accuracy when compared to the original value, red cells represent a decrease in statistical accuracy.
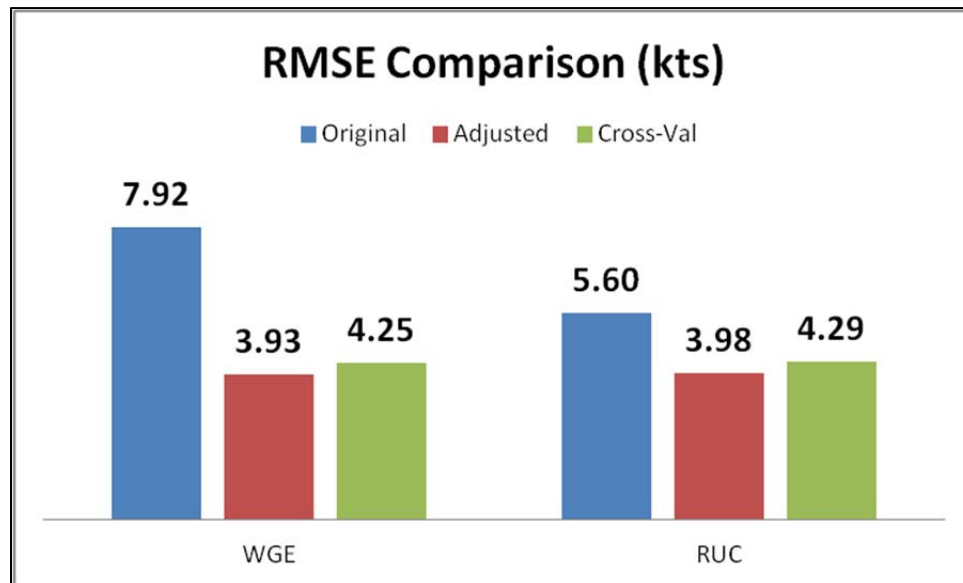
Figure 42.   RMSE comparison for Langley AFB. Blue bars represent the original RMSE value, red bars represent the RMSE values after linear regression was applied and green bars represent the cross-validated RMSE values.

d.   Scott AFB

Table 12 reflects the linear regression equations for Scott AFB applied to the 28 forecasts.   The WGE regression resulted in 17 better forecasts, nine worse predictions and two unchanged forecasts. Five of those nine forecasts changed by ≤ 3 kts while the other four changed by 4 kts to 8 kts. Misses also increased from one to two. Table 13 shows the updated statistics that were computed.   Three of seven categories were improved (HR, FAR, Bias).   The RUC regression resulted in 18 improved predictions, five worse forecasts and five predictions that did not change.   Of the five forecasts, three changed by ≤ 3 kts and two changed by 4 kts to 11 kts. Analyzing the post-regression RUC statistics showed that all categories improved except a slight decline in the bias statistic and the ETS remain unchanged.   Figure 43 details the RMSE comparison for each algorithm.

	WGE	RUC
y=.3742x + 20.722	**y=.3236x + 23.208**	

Table 12.   Linear regression equations for Scott AFB computed by applying best fit line to observed wind gust versus forecast data pairs.

	WGE	Adj WGE	RUC	Adj RUC
HR	0.64	0.68	0.64	0.68
CSI	0.44	0.44	0.29	0.36
POD	0.89	0.78	0.44	0.50
FAR	0.53	0.50	0.56	0.44
BIAS	1.89	1.56	1.00	0.90
KSS	0.42	0.41	0.18	0.28
ETS	0.20	0.15	0.10	0.10

Table 13.   Post-regression statistics for Scott AFB. The seven statistical categories analyzed were Hit Rate (HR), Critical Success Index (CSI), Probability of Detection (POD), False Alarm Rate (FAR), Bias, Kuiper Skill Score (KSS) and Equitable Threat Score (ETS). Green cells represent an increase in statistical accuracy when compared to the original value, red cells represent a decrease in statistical accuracy.
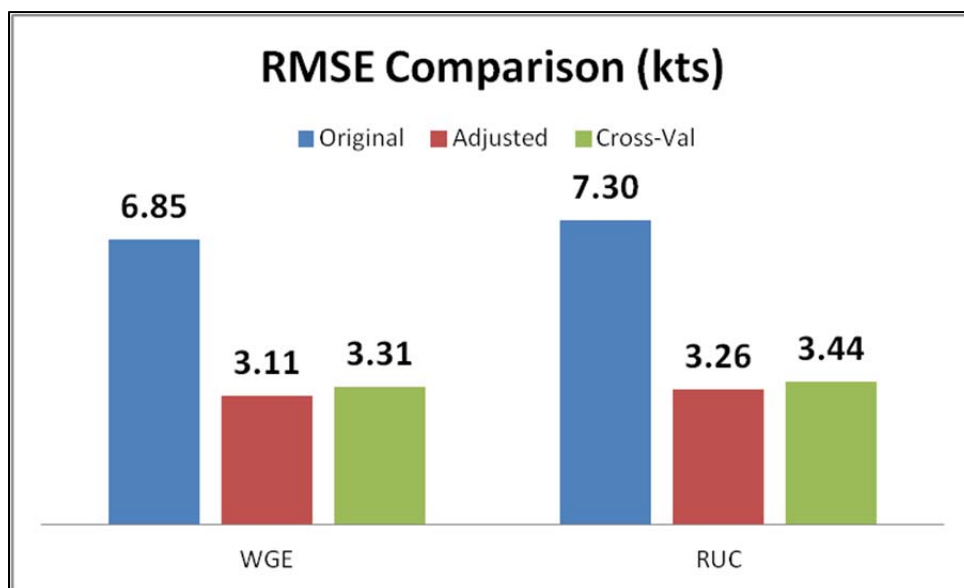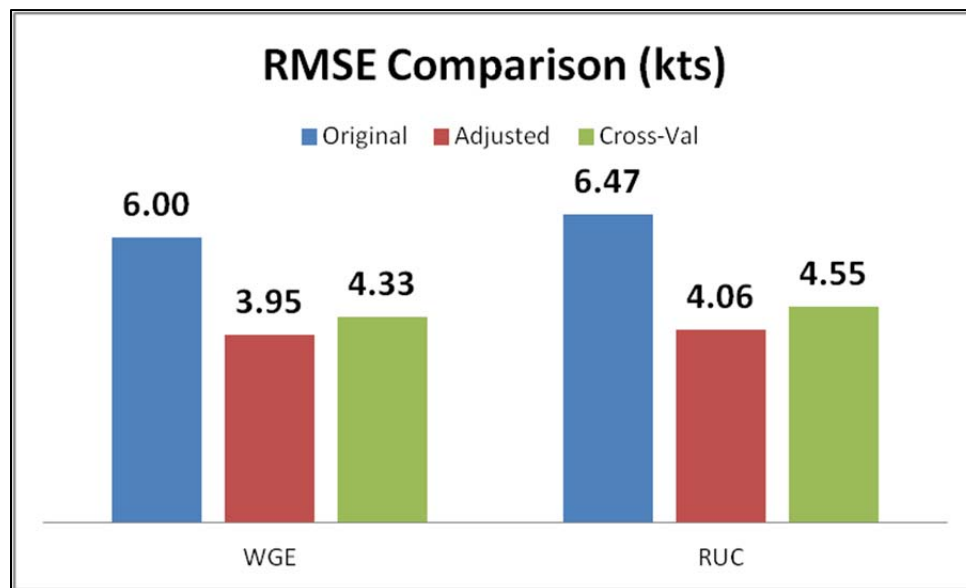

Figure 43.   RMSE comparison for Scott AFB. Blue bars represent the original RMSE value, red bars represent the RMSE values after linear regression was applied and green bars represent the cross-validated RMSE values.

e. *Offutt AFB*

The final linear regression analysis was accomplished on the data at Offutt AFB. Table 14 shows the linear regression equations applied to the 39 forecasts. The WGE regression resulted in 25 better forecasts, nine worse predictions (six changed by ≤3 kts and three changed by 4 kts to 5 kts) and five unchanged forecasts. Missed forecasts increased from seven to eight. When new contingency table statistics were computed, all but CSI, POD, and Bias improved (see Table 15). The RUC regression resulted in 25 improved predictions, 10 worse forecasts (seven changed by ≤3 kts and three changed by 4 kts to 5 kts) and four predictions that did not change. Missed forecasts decreased from six to four. All statistical categories improved except for a small decline in FAR (6%). Figure 44 reflects the RMSE comparison for each algorithm.

WGE	**RUC**
y=.1919x + 27.42	**y=.3304x + 23.741**

Table 14.   Linear regression equations for Offutt AFB computed by applying best fit line to observed wind gust versus forecast data pairs..

	WGE	**Adj WGE**	**RUC**	**Adj RUC**
HR	0.64	0.69	0.82	0.85
CSI	0.26	0.25	0.46	0.57
POD	0.42	0.33	0.50	0.67
FAR	0.58	0.50	0.14	0.20
BIAS	1.00	0.67	0.58	0.83
KSS	0.16	0.19	0.46	0.59
ETS	0.09	0.11	0.35	0.45

Table 15.   Post-regression statistics for Offutt AFB. The seven statistical categories analyzed were Hit Rate (HR), Critical Success Index (CSI), Probability of Detection (POD), False Alarm Rate (FAR), Bias, Kuiper Skill Score (KSS) and Equitable Threat Score (ETS). Green cells represent an increase in statistical accuracy when compared to the original value, red cells represent a decrease in statistical accuracy.
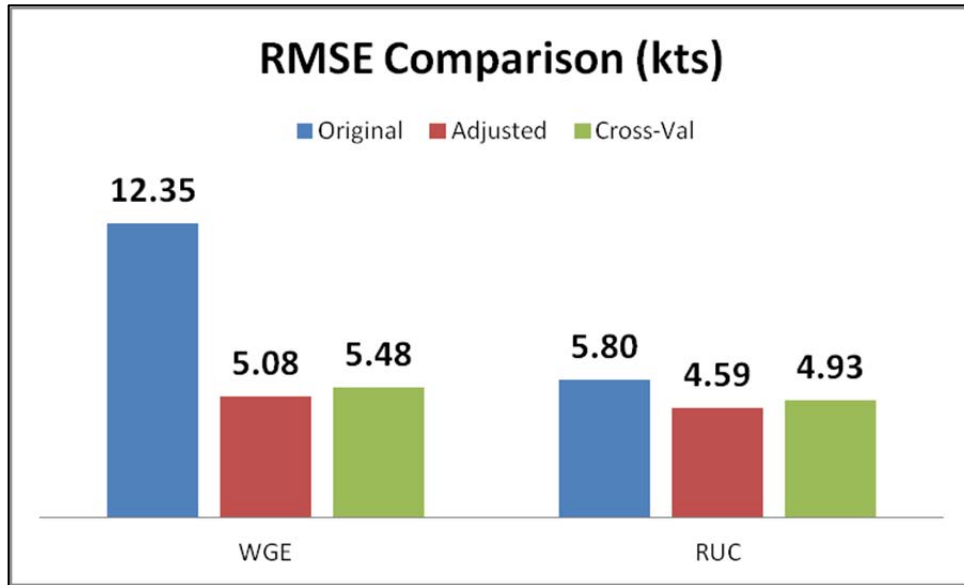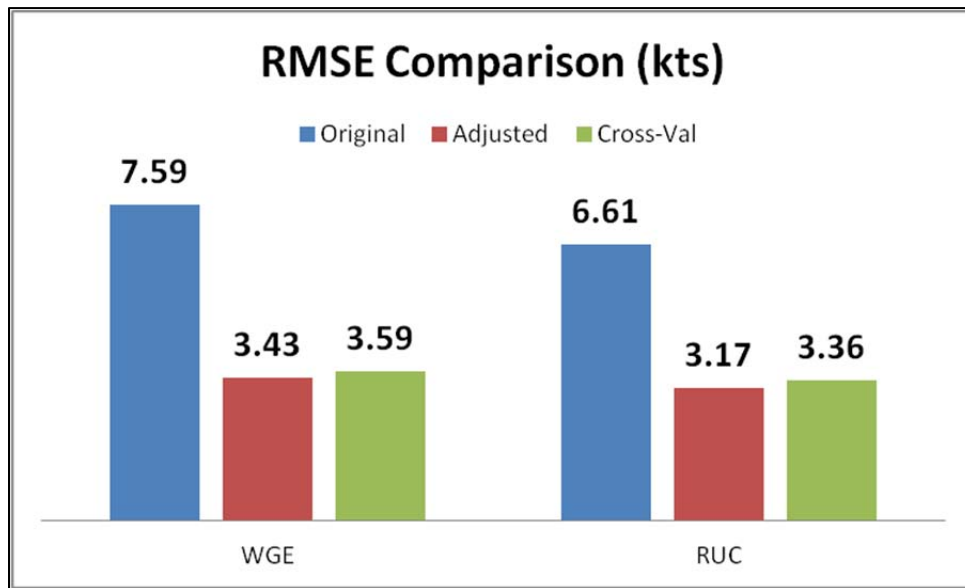
Figure 44.   RMSE comparison for Offutt AFB. Blue bars represent the original RMSE value, red bars represent the RMSE values after linear regression was applied and green bars represent the cross-validated RMSE values.

3.   Summary

Although best fit regression analysis was a simplistic way of fitting seemingly non-linear data to a line, it is our theory that this approach could be useful when applied to a collection of data over many years. Our data design restricted the results to center around the 30 kt to 35 kt threshold, however if all forecasts and all observations were used to produce a regression analysis, this could potentially be very useful in tuning the algorithms to the model. For the purposes of our research, the results discovered above were used during the case study analysis to provide additional information when preparing to forecast an independent case near the 35 kt threshold.

This regression must be used with caution, however. It is evident from the linear regression equations that low wind gust forecasts would result in large adjustment due to our data design. For example, a 5 kt wind forecast would likely be adjusted to near 30 kts due to our analysis parameters. After carefully analysis and consideration, a minimum forecast wind speed was set in order for the regression to be applied because the errors associated with the regression were non-Gaussian. During case analysis, wind forecasts

below 25 kts were not used to apply the regression and that forecast remained unadjusted. Section E describes the analyzed results of the data with and without linear regression applied.

E. CASE STUDY ANALYSIS

1. Overview

Two 24-hour periods where winds reached 30 kts or greater in at least one observation were withheld from the original data set at the beginning of the selection process. By masking this data very early on in the research project, we ensured objectivity would be accomplished during the analysis of this data. The only piece of information known was that at some point, winds reached at least 30 kts in order for the observations to be culled and saved from the original batch of observations. Each forecast was approached from an operational forecaster point of view, however with limited preliminary information. Only the surface analysis and infrared (IR) imagery three hours prior to the beginning of the forecast period were accessed. Furthermore, no model fields were analyzed due to the focus of this case study analysis on the algorithm performance with information and results found for each location.

The goal of the case study analysis was to analyze each 3-hour forecast panel for a pre-determined 24-hour period. During this 24-hour timeframe a decision would be made whether or not to issue a 35 kt wind warning for this location and, if so, the valid times of the warning to assess timing error. The observed data was then compared to the forecast and the results are summarized below. This method of analysis is not meant to encompass a "perfect world" scenario where all data is available; however the method was designed to provide an evaluation of the algorithms based on all of the information and results analyzed to this point to provide a thorough analysis of the algorithms performance with and without adjustments detailed previously. The following section provides examples of how the case study analysis was conducted as well as some of the results that were identified. An example of a hit, miss, false alarm as well as evidence of nocturnal boundary layer and timing errors from a selection of these case studies are presented with a summary of all results detailed at the end.

83

2. Results

a. False Alarm Example

As previously explained in this research, at Westover ARB, the WGE outperformed the RUC statistically in the near term and the threshold verification statistics further emphasized the WGE's ability to predict wind gusts using the 35 kt wind threshold. For these reasons, the WGE was chosen as the preferred algorithm for this case study at Westover. This does not mean the RUC information was discarded, only that it was used to supplement the forecast provided by the WGE. Additionally, climatological wind direction information for 35 kt wind events were analyzed. A peak in 35 kt wind events occurred when wind directions were from the west through northwest (270–330). This provided additional guidance likely not taken into account by the algorithms forecasts.

This case resulted from a storm system approaching Westover on 25 December 2008. A low pressure system located over the Great Lakes region was lifting to the northeast with an approaching cold frontal boundary located in extreme western New York by 00Z on the 25th. The forecast period is from 03Z (22L) on the 25th to 03Z on the 26th.

Both algorithms and both adjusted algorithms indicated 35 kts during the forecast period. The WGE increased the wind gusts to near 50 kts by 09Z which corresponded well with expected frontal passage time based on placement at 00Z. The gradual decrease in winds indicated by both algorithms was indicative of the predicted weakening pressure gradient behind the cold front. The initial difference in wind gust estimation and temporal evolution was likely attributed to the differences in physical processes (increase in TKE during frontal passage) versus the empirical relationship of the wind speeds in the boundary layer.

Based on the given results the warning would have been issued valid at 09Z until 12Z. Waiting until 09Z was based on the model predicted wind direction (not shown) not favorable for 35 kt wind events based on climatology until that time period.

Furthermore, due to the tendency for WGE to overforecast at Westover based on previous results, the expected wind speeds were reduced to between 35 kts and 39 kts.

After the observations were analyzed for this case, the warning would have been a <u>false alarm</u> (by only 1 kt). This gives confidence that the adjustments made to the forecast and the warning itself were certainly valid. Applying the linear regression increased the performance of the WGE algorithm and produced the best algorithm for this case. For the original WGE algorithm, of the nine forecasts (03Z through 03Z), five predicted winds greater than the 35 kt threshold, however once the linear regression was applied, this number was reduced in only one 3-hour forecast. Futhermore, the RMSE decreased from 11.88 kts to 8.67 kts. The RUC's performance was less changed by the linear regression adjustment with only an RMSE change from 14.67 kts to 10.87 kts. Figure 45 shows the comparison of the original and adjusted by linear regression predictions compared to the observations. The black lines indicate nighttime hours for this location to show potential overforecasting of the algorithms at night as inferred in previous research.

Figure 45.   25 Dec 08 original (top) and adjusted forecasts by linear regression (bottom) compared to observations at Westover ARB. Blue lines represent original (top) and adjusted (bottom) WGE forecasts. Red lines represent original (top) and adjusted (bottom) RUC forecasts. Green line represents maximum observed wind speed (by +/- 1.5-hour observations group) for each accompanying 3-hour forecast time. Space between vertical black lines represents hours of darkness.

b.   Hit Example with Nocturnal Boundary Layer Error

Both algorithms performed very similar at Andrews AFB, especially after the linear regression adjustments were made. However, the RUC's performance in the near term window became the basis of choosing this algorithm as the preferred solution for this location. Additionally, a peak in 35 kt wind events occurred when wind directions were from the west through northwest (270–330) based on climatological information.

86

This case resulted from a storm system approaching the location on 31 December 2008. A low pressure system located over northern Indiana was moving to the east towards Pennsylvania with an approaching boundary system approaching the location. The warm front extended southeast into western Virginia while the cold front extended southwest into southeastern Missouri at 00Z. It was expected that this forecast was going to be influenced by both frontal passages. The forecast period is from 03Z (22L) on the 31st to 03Z on the 1st.

Initial results indicated the model had a handle of the synoptic situation with winds gradually increasing with a peak wind occurring at 18Z which corresponded well with expected cold frontal passage in both algorithms. The WGE algorithm was more aggressive with wind speed changes with roughly 15 kts at 12Z to near 50 kts at 15Z, while the RUC is more gradual and peaks near 40 kts at 18Z. When the linear regression was applied, the WGE remained unchanged until 15Z due to predetermined minimum 25 kt threshold. The RUC crossed this threshold at 09Z. While the temporal evolution is relatively the same, the difference between the two predictions was much smaller beginning at 15Z after the regression was applied. Both adjusted algorithms predicted winds greater than 35 kts during the same time frame.

Based on the given results the warning would have been issued valid at 15Z until 03Z. Both algorithms temporal pattern reflected the expected wind speed variations that would accompany the associated weather pattern. Climatological wind analysis revealed favorable wind directions to occur at that time as well. Furthermore, due to the tendency for the RUC to underforecast this location based on previous results, the expected wind speeds would have been increased to between 44 kts and 48 kts.

After the observations were applied to the case study, the warning would have <u>verified</u> with a 1 hour 12 minute timing error. The maximum observed wind speed during this time period was 47 kts. This adjustments made to the forecast and the warning itself were certainly valid. Additionally, the temporal evolution of the observed winds was almost identical to the WGE and adjusted WGE pattern. Applying the linear regression increased the performance of the WGE algorithm and produced the best algorithm for this case. For the original WGE algorithm, of the nine forecasts, five were

predicting winds greater than the 35 kt threshold and four of those verified as hits. Once the linear regression was applied, these numbers remain unchanged but RMSE decreased from 5.84 kts to 4.97 kts. The RUC's performance was less changed as five of five "yes" forecasts verified. However, the RMSE for both the unadjusted and adjusted predictions were extremely high (12.58 kts and 15.17 kts, respectively) indicative of the nocturnal boundary layer overforecast bias. Figure 46 shows the comparison of the adjusted and unadjusted predictions compared to the observations.
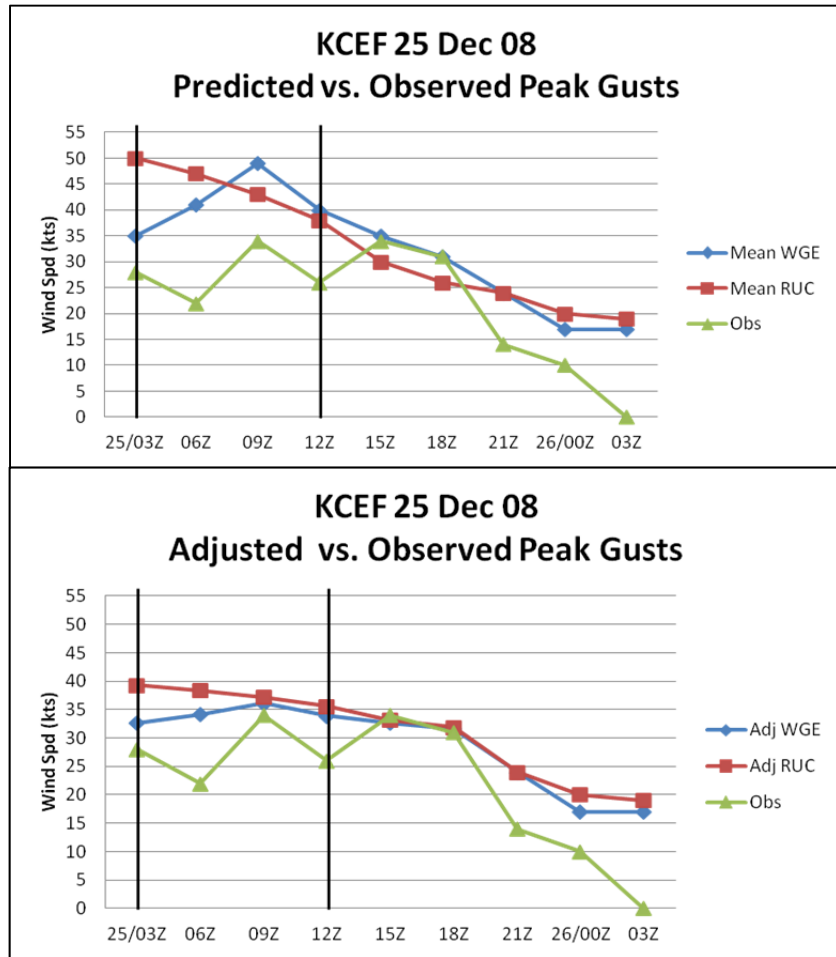

Figure 46.   31 Dec 08 original (top) and adjusted forecasts by linear regression (bottom) compared to observations at Andrews AFB. Blue lines represent original (top) and adjusted (bottom) WGE forecasts. Red lines represent original (top) and adjusted (bottom) RUC forecasts. Green line represents maximum observed wind speed (by +/- 1.5-hour observations group) for each accompanying 3-hour forecast time. Space between vertical black lines represents hours of darkness.

### d.	*Miss Example*

As shown in the statistical results, the WGE algorithm performed better than the RUC algorithm when statistical categories were analyzed at Scott AFB. Therefore, the WGE was used as the preferred forecast solution for this location during case study analysis. Adjustments were made based on the slight overforecast bias shown in the error histograms. Additionally, climatological preferred wind directions were from the west through northwest directions (270–300). Also of note, for research conducted on data during the 2008–2009 season, wind gusts also exceeded the 35 kt threshold at wind directions from the southeast through west (150–270).

This case resulted from a storm system very close to Scott AFB on 14 January 2009. A low pressure system was located along the northern Missouri and central Illinois border and moving to the east. The approaching warm front was in proximity to the base by 12Z while the cold frontal boundary extended southwest from the low into northern Missouri and back into Kansas. The forecast period is from 15Z (09L) on the 14th to 15Z on the 15th.

Both algorithms revealed a very similar evolution in wind gusts with the RUC prediction of 33 kts initially and decreased steadily through the end of the forecast period. The WGE predicted winds at 23 kts initially and then peaked at 35 kts at 18Z before decreasing through 21Z. The algorithms maintained an approximate 5 kt difference through 12Z where both algorithms were essentially identical. The adjusted algorithms revealed similar maximum wind speed forecasts below 35 kts for both algorithms through the period. The spike from the initial 15Z forecast to the 18Z forecast reflected in the WGE is more of a representative signature of a cold frontal passage.

Based on the given results the warning would not have been issued for this case. Both algorithms' predictions did not provide high confidence in meeting the threshold as only the unadjusted WGE forecast peaked at 35 kts, then rapidly dropped wind speeds afterwards. Favorable wind directions do occur for 35 kt winds and the

nocturnal boundary layer is not a factor for this case, however the tendency of the WGE to overforecast this location resulted in a lowered forecast maximum wind gust forecast of 30 kts to 32 kts.

After the observations were applied to the case study, the non issued warning would have been a miss. Two observations of 35 kts and 36 kts were recorded during the 18Z to 21Z forecast windows. The sustained wind speed was also analyzed and revealed a very low RMSE for this forecast window at 2.75 kts. This indicated the accuracy of the model's handling of the boundary layer pattern. Furthermore, the original 35 kt forecast by the WGE verified which indicated the ability of the WGE to accurately forecast the wind pattern given model's handling of the atmospheric and boundary layer parameters.

For the original WGE algorithm, of the nine forecasts, seven verified as "no/no" forecasts, one as a "yes/yes" and one as a miss with an RMSE of 6.05 kts. The unadjusted WGE prediction turned out to be the most accurate for this case. Both adjusted algorithms did not indicate warning level winds and RMSE increased for the WGE to 6.68 kts. While the RMSE decreased from 5.34 kts to 4.70 kts for the RUC, both adjusted and unadjusted failed to predict the warning level winds throughout the forecast period. Figure 47 shows the comparison of the adjusted and unadjusted predictions compared to the observations.
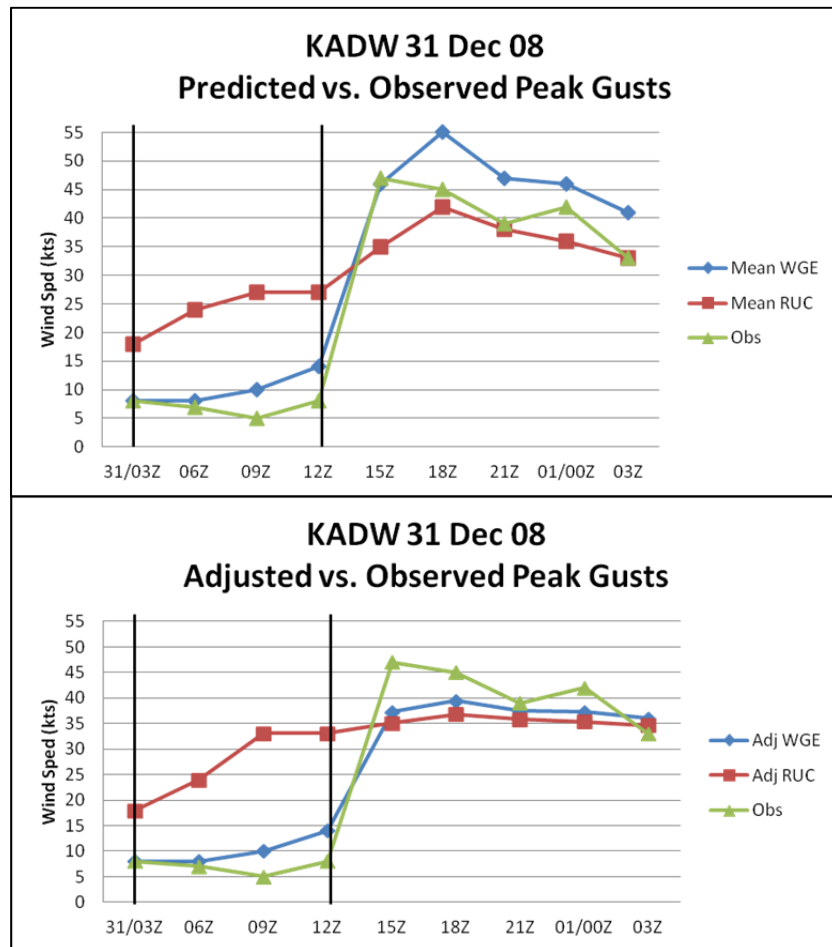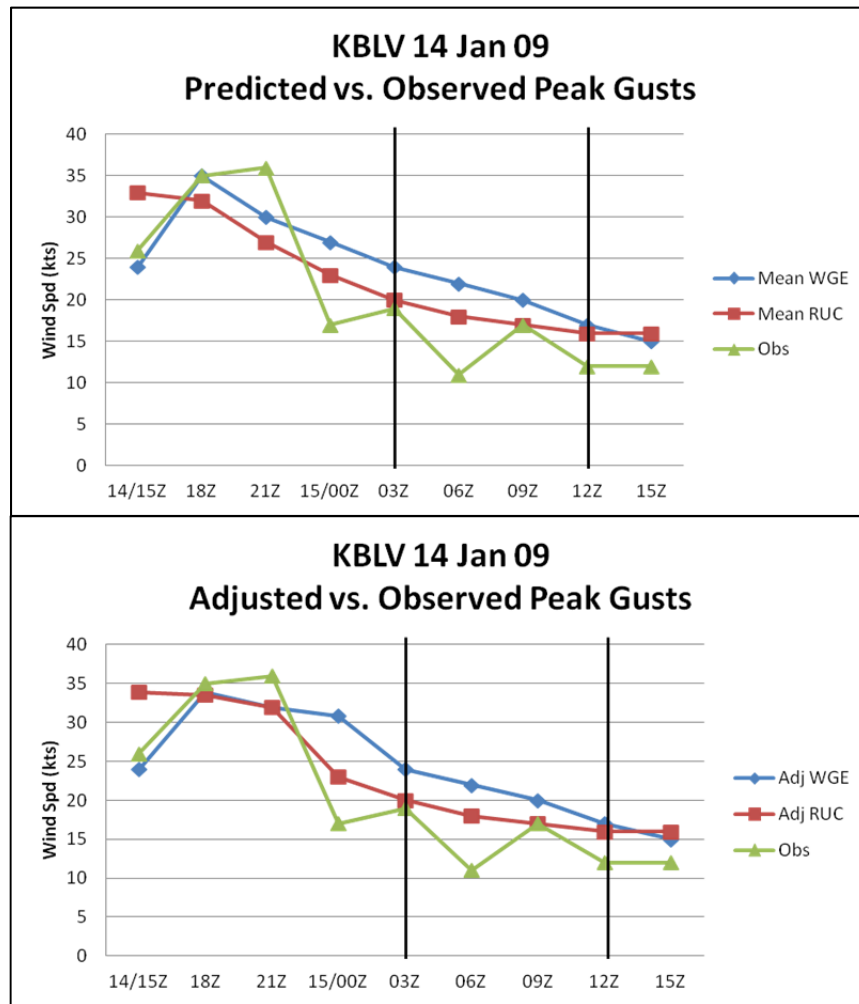
Figure 47.   14 Jan 09 original (top) and adjusted forecasts by linear regression (bottom) compared to observations at Scott AFB. Blue lines represent original (top) and adjusted (bottom) WGE forecasts. Red lines represent original (top) and adjusted (bottom) RUC forecasts. Green line represents maximum observed wind speed (by +/- 1.5-hour observations group) for each accompanying 3-hour forecast time. Space between vertical black lines represents hours of darkness.

e.   *Hit Example with Model Timing Error*

The results from the statistical analysis section revealed the WGE's better performance in the near term window of less than 24 hours at Offutt AFB. After linear regression was applied, the RUC also performed well statistically for this location. Therefore, the WGE was used as the preferred forecast solution for this location during case study analysis with consideration given to the adjusted RUC forecast.   The tendency

for the WGE to have a slight overforecast bias for this location was also taken into consideration when producing a forecast during the case study analysis process although this bias was minimal. Additionally, climatological preferred wind directions for 35 kt wind events occurred throughout almost all directions on the compass rose. The peak of events occurred when wind directions were from the northwest through north directions (300–360). However, secondary and tertiary maximums included wind gusts from the north through northeast, and southeast through west (000–030 and 150–270) indicative of the variable nature in which wind events occur at this location.

This case resulted from a storm system north of Offutt AFB on 12 January 2009. A low pressure system was located near the South Dakota, North Dakota, Minnesota border triple point. The approaching warm front was in proximity to the base by 12Z while the cold frontal boundary extended southwest from the low into central South Dakota. The forecast period is from 15Z (09L) on the 12th to 15Z on the 13th.

Both algorithms revealed identical temporal evolutions of wind gusts with winds decreasing through the first six hours then increasing rapidly over the next six hours with a peak wind at 00Z. The WGE max wind of 44 kts exceeded the 35 kt threshold where the RUC prediction of 34 kts did not and presented the only major difference between the two algorithms. The adjusted algorithms revealed almost the same exact temporal evolution of wind gusts and maximum wind speed forecasts except for 06Z. Additionally, the linear regressed WGE maximum wind speed at 00Z was reduced to the 35 kt threshold.

Based on the given results the warning would have been issued for this case valid from 21Z to 06Z. The approaching cold front was forecast to move through within six to nine hours of the analysis time with peak winds during frontal passage at approximately 00Z as indicated by both algorithms. Wind directions were favorable during this time (320–350) and with maximum wind speeds expected to occur during the afternoon to early evening hours the issuance of the warning would have been justified. With the WGE chosen as the algorithm of choice, wind speeds were predicted to reach 44 kts, although the RUC predicted max winds of 35 kts during the same period. This resulted in a lowered wind speed forecast of 40 kts to 45 kts.

After the observations were applied to the case study, the warning would have <u>verified</u> with a 9 minute timing error based on the adjusted warning valid time. Based only on the both algorithm's predictions, the timing error would have been around three hours as both algorithms predicted 35 kts between 21Z and 00Z. Maximum winds reached 36 kts near 21Z. The temporal evolution of the winds verified well with both algorithms although there appeared to be a timing error of frontal passage by approximately three hours (see Figure 48). Regardless, the spike in peak wind and rapid decrease was very well indicted by both algorithms. The sustained wind speed was also analyzed and revealed a very low RMSE for this forecast window at 2.31 kts with no timing error problems noted.

When RMSE, false alarm and missed forecasts were analyzed, the statistics were somewhat misleading due to the timing error. Both show large RMSE errors and missed 35 kt forecasts, but these are attributed to the timing errors as noted. For the original WGE algorithm, of the nine forecasts, six verified as "no/no" forecasts, one as a "yes/yes," one as a miss, and one was a false alarm with an RMSE of 10.27 kts. The unadjusted WGE prediction turned out to be the most accurate for this case although, as expected, the adjusted RUC was very close. Both adjusted algorithms indicated warning level winds, 00Z-03Z for the WGE and 00Z for the RUC, but neither verified due to the timing error. RMSE increased for the WGE to 10.39 kts, while the RMSE decreased from 9.13 kts to 8.87 kts for the RUC. Figure 48 shows the comparison of the adjusted and unadjusted predictions compared to the observations.
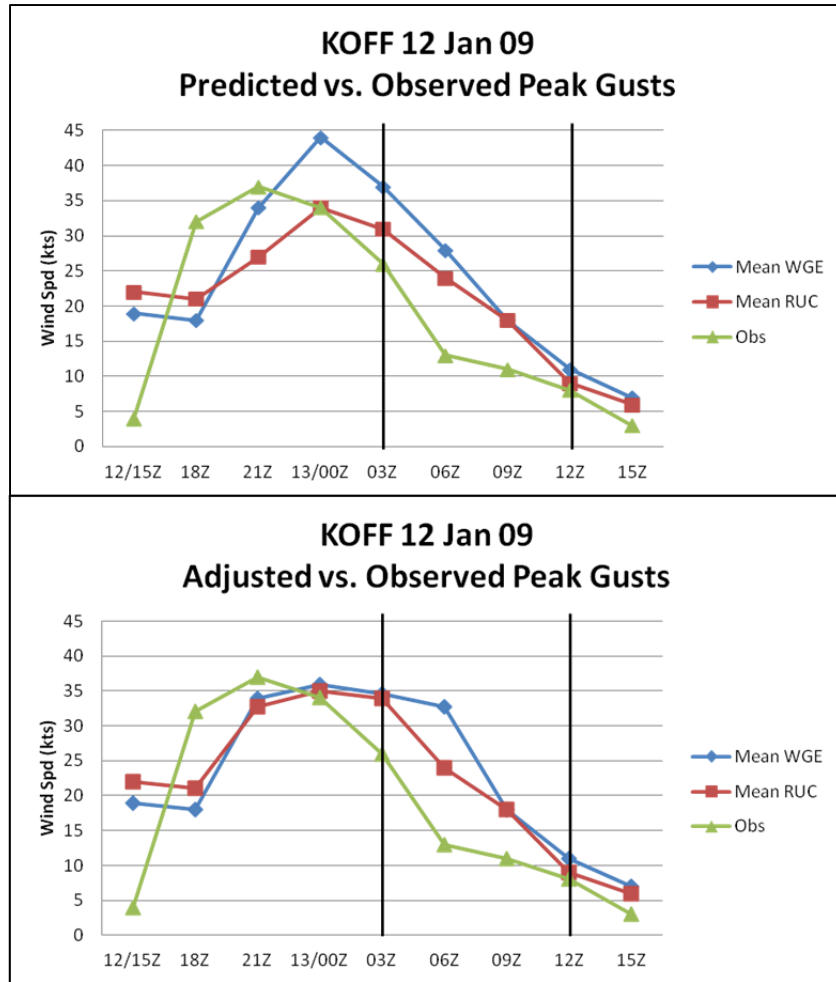
Figure 48.   12 Jan 09 original (top) and adjusted forecasts by linear regression (bottom) compared to observations at Offutt AFB. Blue lines represent original (top) and adjusted (bottom) WGE forecasts. Red lines represent original (top) and adjusted (bottom) RUC forecasts. Green line represents maximum observed wind speed (by +/- 1.5-hour observations group) for each accompanying 3-hour forecast time. Space between vertical black lines represents hours of darkness.

3.   Summary

The case studies revealed valuable results when analyzed together. Of the 10 cases, six 35 kt wind warnings would have been issued based on the analysis conducted. Four of those would have verified and two would have been false alarms (both by only 1 kt) which indicates accurate forecasts were made. Of the four non-issued warnings, two of those verified as non-events and another two were missed. In one missed case, one of the two algorithms forecasted 35 kt winds with a slight timing error from the observed 35

kt wind speed, but low confidence in the predictions based on analysis resulted in a forecast of lower wind speeds. The other miss resulted from a stronger than expected frontal passage at night that caused winds in excess of the threshold not predicted by either algorithm. There typically exists an inherent pressure of making a decision to issue a warning in an operational forecasting environment, especially when the models indicate above or near threshold values. However, for this research, objective analysis and reasoning without this pressure provided an opportunity to produce forecasts and assess the algorithms capability without the added pressure of costs associated by issuing or not issuing the warning.

When the WGE was compared to the RUC for overall performance and best method, the WGE was the better algorithm for seven of the 10 cases. For both cases at Scott AFB, the unadjusted WGE performed the best, confirming the statistical analysis conducted for this location. Three of the seven times the linear regressed WGE performed better than the original forecasts including both cases at Andrews AFB. Furthermore, the RUC algorithm performed better during three of the 10 test cases analyzed and only one of those three times did the linear regression improve the forecast. This validates key research goals presented in this thesis. First, different methods worked at different locations. Furthermore, the linear regression, while only a small statistical sample, showed promise when tuning the physically based wind algorithm. Less of an improvement is evident when applied to the empirically based wind gust method. Nonetheless, similar to the improved forecasting ability with the creation of ensembles, using these different wind algorithms together to produce a forecast proved valuable with the added benefit of additional information provided from another model.

The black lines indicated on each of the Figures 45 through 48 indicated hours of darkness to briefly investigate the tendency for the algorithms to overforecast nocturnal boundary layer wind gusts at the surface. There are indications that this is an issue with both algorithms, however the RUC algorithm seemed to be effected by this more in the ten case studies analyzed. With accurate PBL representation and especially of the lower TKE values in a nocturnal boundary layer, it was noted that the WGE handled this

phenomenon slightly better whereas the empirical method relies on accurate modeled wind speed and PBL height variables which could be overforecast.

The bounding interval described in Chapter II was not explicitly examined in this research, however preliminary results were collected for future analysis. Table 16 shows the results from this data period collected at each location along with the total results. In summary, approximately two-thirds of the observed wind gusts were within the bounded interval predicted by the WGE. Approximately 20% of the observations were observed below the lower bound while nearly 10% were observed above the upper bound. These results support Brasseur's (2001) conclusions that the bounded interval is a reliable source of information and a reasonable way to assess confidence in the algorithms prediction.

	Westover	Andrews	Langley	Scott	Offutt	TOTAL	%
Obs Low	12	6	2	7	8	35	22.29
Within Range	14	24	21	21	27	107	68.15
Obs High	1	9	1	0	4	15	9.55
Accuracy %	51.85	61.54	87.5	75	69.23		

Table 16.   WGE bounded interval analysis. Observed low column indicates the number of forecasts where the observed value was below the lower bound prediction. Observed high column indicates the number of forecasts where the observed value was higher than the upper bound prediction. Accuracy percentages based on the ratio of number of values within range to the total number of forecasts for each location.

V. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

Based on the results of this research, using the WGE and RUC algorithms to predict 35 kt wind events is beneficial. The analysis design for our research was rather strict and some statistical reliability was lost due to the fact that observations and forecasts below 30 kts were not analyzed. However, the inclusion of low wind events would only act to increase the performance statistics detailed in Chapter IV. This increase in performance would come from the amount of days with calm or light winds and the increased number of correct rejections and lower number of false alarms that would occur. Outside of the some predictions in the nocturnal boundary layer, the algorithms were shown to handle low wind situations well during a variety of sample cases (not shown), and at the end of several analyzed case studies such as the 12 Jan 09 case shown in Figure 48. Due to our strict threshold design of 35 kts, it is anticipated that statistics such as false alarm may be a pessimistic representation. This is a promising conclusion since false alarm rates were near 50% overall.

As mentioned in several studies conducted before this, these algorithms are highly sensitive and dependent upon the model's performance and accurate representation of the atmospheric parameters, especially in the boundary layer. Brasseur (2001) suggested that higher-resolution model forecasts would lead to more accurate WGE predictions. Our research showed that using a lower resolution model (45 km) still produced reliable results that provided reasonable confidence in the ability of both algorithms to produce accurate non-convective wind gust predictions. Certainly we expect that a higher resolution model should produce better results, but it is not clear whether skill gains would justify increases in computational costs.

There is still a requirement to quality check and adjust the algorithm's output. Examples of these types of adjustments include analysis of climatological factors, local rules of thumb and known model errors or tendencies (i.e., terrain interpolations). In many ways, this can be considered an example of tuning the algorithms as well. For

example, by applying location-based climatological wind direction analysis for 35 kt wind events, adjustments to the timing of 35 kt winds based on predicted wind direction can be made. In several of the analyzed case studies, this reduced the false alarm rate and timing error of our predicted wind warnings. Furthermore, the nocturnal boundary layer issues were only briefly examined in case studies where an entire 24-hour period was analyzed. Results support previous research conclusions of the tendency of the WGE to overforecast winds in the nocturnal boundary layer (LaCroix 2002) when compared to daytime wind speeds forecasts and was also noted in predictions from the RUC algorithm.

The linear regression analysis is a simple way to tune the algorithms to the data. Initial results show there were some improvements made to some independent case studies analyzed, but outliers had a large effect and could not be applied to lower wind speed predictions. While not perfect, it is suggested that a large database of prediction versus observation results would produce a more accurate regression analysis and could be applied to tune the algorithms. A collection of results for each location is ideal. Additionally, incorporating other parameters such as the wind direction for these events would make the tuning more reliable.

The results from the case studies show the positive advantages of introducing another forecast method into the forecast process. Although only three of the 10 cases resulted in the RUC as the best algorithm forecast, analyzing the differences between the algorithms is just as important as the similarities. The data and results presented suggest integration of the two algorithms would help to increase the accuracy of critical wind speed thresholds. The benefit of these algorithms could be fully realized if applied to each grid point in the model and a spatial representation of wind gusts were presented. The location-based results presented in this study show the benefits of applying and evaluating these algorithms a particular location, which then could be applied to individual grid points for further analysis.

An accurate representation the TKE field in the boundary layer is vital in WGE calculations. LaCroix (2002) presented a method using to produce this calculation using model perturbation variables when TKE values are not available due to the model's PBL

98

scheme. Due to the availability of TKE and boundary layer height parameters from the MYJ PBL scheme for our research, LaCroix's (2002) method was not utilized. However, it is our recommendation that a model data set incorporating a parameterization scheme, such as the MYJ PBL scheme, be used to ensure the accuracy of the TKE fields and other boundary layer variables. This ensures that calculation or boundary layer height errors are not a factor in the algorithms prediction of maximum wind gusts. A study of the cost/benefit analysis of calculating TKE versus utilizing a model with the MYJ PBL scheme should be accomplished. This is one of several recommendations of from our research.

B. RECOMMENDATIONS

There is still work needed in this area to benefit forecasting for other locations outside of the 15th OWS AOR. The following recommendations are suggestions for areas of future research. First, using the ensemble mean as the single deterministic model caused a loss of valuable ensemble information. This was an unintended consequence of our research, but necessary to focus on the performance of the algorithms versus the performance of the model itself. Future research should explore the evaluation of the value added by applying these algorithms to an ensemble forecast. Brier Skill Scores and reliability operating characteristics would be valuable analyses utilizing such ensemble information.

One such example of value added by the ensemble is assessing the risk of using the ensemble member's prediction of a threshold forecast. This is a hybrid approach to conversion of a probabilistic forecast to a nonprobabilistic forecast by choosing an appropriate threshold presented in Wilks (2006). A brief example of this is presented in Figures 49 and 50. Member 11 was chosen as the control forecast and used to produce a "yes/no" forecast compared to the percentage of the members forecasting a "yes" forecast. This analysis revealed that the lowest increase in miss rate compared to the lowest increase in false alarm rate existed at the 67% threshold (4 of 6 members) for the WGE and 60% for the RUC (6 of 10 members). This is shown in the second graph of each figure as miss rate is compared to false alarm rate. Based on these thresholds, waiting to issue a warning for the 35 kt threshold until these thresholds were met

produced the ideal miss/false alarm rates when compared to other thresholds. The black lines represent the 67% and 60% thresholds, respectively.

Finally, implementation of these algorithms, as well as the location-based results presented in this research, could prove useful in operational forecasts; however further testing is needed to measure performance against local rules of thumb and other established methods currently in place. Further research areas should utilize real-time information. This requires access to current model data sets that incorporate the MYJ PBL scheme and associated required parameters. As future models come online (such as the Rapid Refresh), this is a possible avenue to acquire real time information for similar analysis conducted in this research. This would also allow research to be conducted on high resolution models with similar PBL characteristics to evaluate performance differences such as was accomplished by Brasseur (2001).

Figure 49.   WGE ensemble prediction threshold using Member 11 as the control forecast. Each plot represents the statistics if the warning were issued based on the percentage of members predicting a "yes" forecast. Black line represents the best decision threshold based on analysis. Hit Rate (HR), Critical Skill Index (CSI), Probability of Detection (POD), False Alarm Rate (FAR) and Miss Rate plotted on left. FAR versus Miss Rate plotted on the right graph.
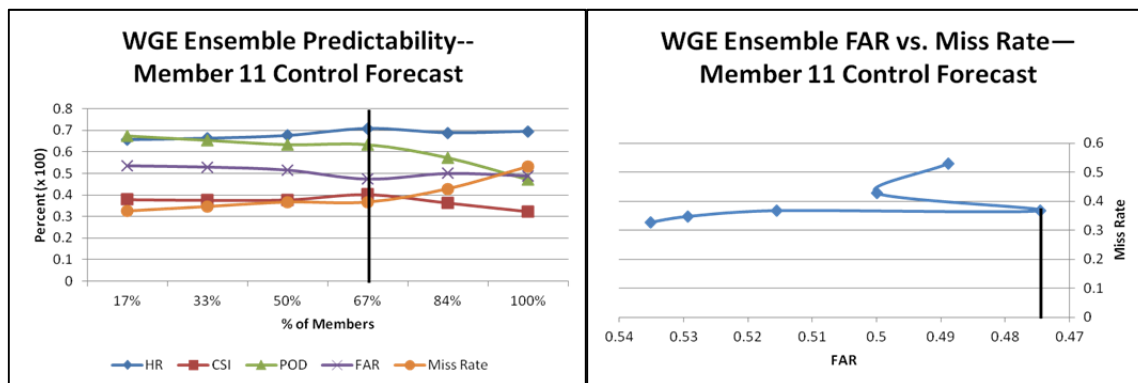
Figure 50. RUC ensemble prediction threshold using Member 11 as the control forecast. Each plot represents the statistics if the warning were issued based on the percentage of members predicting a "yes" forecast. Black line represents the best decision threshold based on analysis. Hit Rate (HR), Critical Skill Index (CSI), Probability of Detection (POD), False Alarm Rate (FAR) and Miss Rate plotted on left. FAR versus Miss Rate plotted on the right graph.
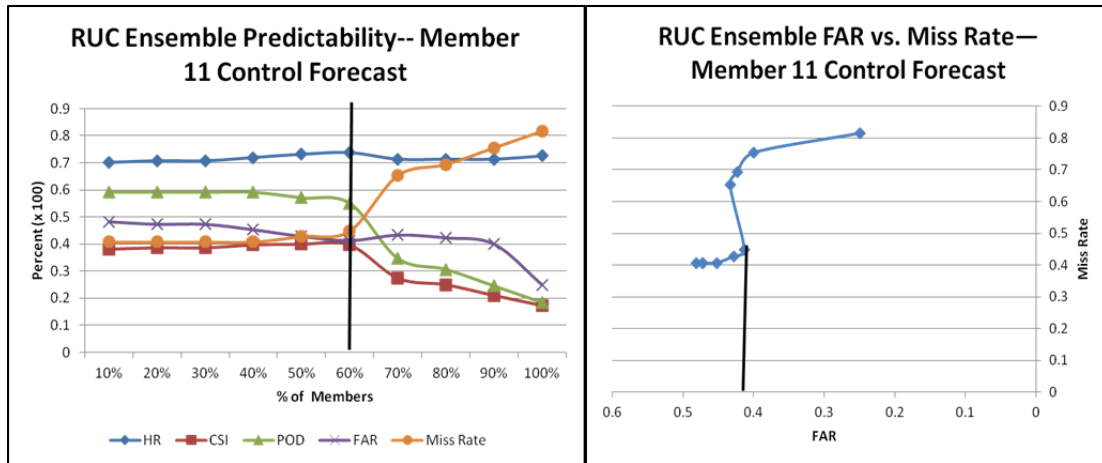
THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

14th Weather Squadron, Asheville, NC, cited February 2012: Climate Services, 14th Weather Squadron. [Available online at https://notus2.afccc.af.mil/SCIS/. Login required.]

15th Operational Weather Squadron, Scott AFB, IL, cited February 2012: 15th OWS AOR. [Available online at https://ows-public.scott.af.mil. Login required.]

AFMAN 15–129V1, 2011: Air and Space Weather Operations – Characterization, Air Force Manual 15–129 Volume 1, AF/AF3O-W. [Available online at http://www.e-publishing.af.mil/shared/media/epubs/AFMAN15–129V1.pdf]

Ashley, W.S., and A. W. Black, 2008: Fatalities associated with nonconvective high-wind events in the United States. *J. Appl. Meteor. Climatol.*, **47**, 717–725.

Associated Press, cited February 2012: High winds force shutdown of tents at NFL Experience. [Avalilable online at http://www.foxnews.com/sports/2012/01/29/high-winds-force-shutdown-tents-at-nfl-experience/?test=latestnews]

Asuma, J. V., 2010: Cool season high-wind events in the Northeast U.S. M.S. thesis, Dept. of Atmospheric and Environmental Sciences, University at Albany, State University of New York, 117 pp.

Brasseur, O., 2001: Development and application of a physical approach to estimating wind gusts. *Mon. Wea. Rev.*, **129**, 5–25.

Brennan, M. J., cited February 2012: Ensemble Prediction Systems. [Available online at http://www.aoml.noaa.gov/general/lib/lib1/nhclib/Powerpoints/WMO-NHC20110miami/Michael%20Brenan-Ensemble%20Prediction%20Systems-WMO2011/WMO%20Ensembles%202011.pptx]

Crupi, K. M., 2004: An anomalous non-convective high wind episode over upper Michigan. *Natl. Wea. Dig.,* **28**, 3–12.

Dudhia, J., 1993: A nonhydrostatic version of the Penn State-NCAR Mesoscale Model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, **121**, 1493–1513.

Ebert, E. E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting,* **24(6)**, 1498–1510.

——, cited February 2012a: Methods and scores used for verifying ensemble forecasts. [Available online at http://www.cawcr.gov.au/projects/EPSverif/scores/scores.html]

——, cited February 2012b: Verification of Ensembles. [Available online at: http://www.ecmwf.int/newsevents/meetings/workshops/2005/TIGGE/Ebert.pdf]

Google Earth version 6.1.0.5001, computer software, Google, Inc., Mountain View, CA.

Goyette, S., O. Brasseur, and M. Beniston, 2003: Application of a new wind gust parameterization: multiscale case studies performed with the Canadian regional climate model. *J. Geophys. Res.* **108**(D13): 4374, doi:10.1029/2002JD002646, 1–16.

Hacker, J.P., and Coauthors, 2011: The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. *Tellus*, **63A**, 625–641

Jolliffe, I.T., and D. B. Stephenson 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 240 pp.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimiliation and Predictability.* Cambridge University Press, 341 pp.

Kapela, A. F., P. W. Leftwich, and R. Van Ess, 1995: Forecasting the impacts of strong wintertime post-cold front winds in the Northern Plains. *Wea. Forecasting,* **10**, 229–244.

Kuhlman, C. J., 2006: Evaluation of convective wind forecasting methods during high-wind events. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 63 pp.

Knight, P., J. Ross, B. Root, G. Young, and R. Grumm, 2005: Fingerprinting Significant Weather Events. Preprints, *Fourth Conf. on Artificial Intelligence Applications to Environmental Science*, San Diego, CA, Amer. Meteor. Soc., 10 pp. [Available online at http://ams.confex.com/ams/pdfpapers/87741.pdf]

Knox, J. A., 2004: Non-convective windstorms in the Midwest United States: Surface and Satellite Climatologies. *Extended Abstracts*, *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., P5.3. [Available online at http://ams.confex.com/ams/pdfpapers/81844.pdf]

——, J. D. Frye, J. D. Durkee, and C. M. Fuhrmann, 2011a: Non-convective high winds associated with extratropical cyclones. *Geo. Compass*, **5:2**, 63–89.

——, ——, ——, and M. C. Lacke, 2011b: Death from a clear blue sky: extreme non-convective high winds. *Earthzine*, June 2011. [Available online at http://www.earthzine.org/2011/06/04/death-from-a-clear-blue-sky-extreme-non-convective-high-winds/]

Lacke, M. C., J. A. Knox, J. D. Frye, A. E. Stewart, J. D. Durkee, C. M. Fuhrmann, and S. M. Dillingham, 2007: A climatology of cold-season nonconvective wind events in the Great Lakes Region. *J. Climate*, **20**, 6012–6022.

LaCroix, K.W., 2002: Application of the wind gust estimate and comparison to the AFWA MM5 wind gust algorithm. M.S. thesis, Department of Engineering Physics, Air Force Institute of Technology, 88 pp.

MATLAB version 7.10.0.499 (R2010a), computer software, The MathWorks, Inc., Natick, Massachusetts.

Moore, A., cited 2012: Cross-validation for detecting and preventing overfitting. [Available online at http://www.cs.cmu.edu/~cga/ai-course/overfit.pdf]

National Oceanic and Atmospheric Administration, Earth System Research Laboratory (NOAA ESRL), cited 2012: RUC Post-Processing Diagnosed Variables. [Available online at http://ruc.noaa.gov/vartxt.html#gust]

Nett, J. cited February 2012: Bilinear Interpolation. [Available online at http://web.pdx.edu/~jduh/courses/geog493f09/Students/W6_Bilinear%20Interpolation.pdf]

Niziol, T. A., and T. J. Paone, 2000: A climatology of non-convective high wind events in western New York state. NOAA Tech. Memo. NWS ER-91, 36 pp.

Nordström, M., 2005: Estimation of gusty winds in RCA. M.S. thesis, Department of Earth Sciences, Uppsala University, 42 pp.

Skamarock, W.C., and Coauthors, 2008: A description of the advanced research WRF version 3. NCAR Tech. Note. NCAR/TN-475+STR, 113pp.

The Weather Research & Forecasting (WRF) Model, cited 2012: About the Weather Research & Forecasting Model. [Available online at http://www.wrf-model.org/index.php]

Wilks, D. S., 2006: *Statistics Methods in the Atmospheric Sciences, Second Edition*. Academic Press, 627 pp.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
    Ft. Belvoir, Virginia

2. Dudley Knox Library
    Naval Postgraduate School
    Monterey, California

3. Dr. Wendell A. Nuss
    Naval Postgraduate School
    Monterey, California

4. Dr. Joshua P. Hacker
    Naval Postgraduate School
    Monterey, California

5. 15th Operational Weather Squadron
    Scott Air Force Base, Illinois

6. 16th Weather Squadron
    Offutt Air Force Base, Nebraska

7. 26th Operational Weather Squadron
    Barksdale Air Force Base, Louisiana

8. Air Force Weather Technical Library
    14th Weather Squadron
    Asheville, North Carolina