ABSTRACT

CRAFT, JEFFREY AARON. Wind Ramp Events: Forecast Verification and Climatology. (Under the direction of Sukanta Basu.)

Implementing wind power into the global arena takes more than merely erecting wind turbines where wind is climatologically abundant: profitability is heavily dependent on accurate short-term wind forecasts. The wind forecasting sector of the wind power industry has seen recent advancements in technology due to the introduction of mesoscale numerical weather prediction (NWP) models. However, short-term wind forecasting has been somewhat deficient in its overall ability to produce extremely accurate wind forecasts, even with the recent industry advancements. Arguably the most problematic attribute in short-term wind forecasting is generated by fluctuations in the wind itself. Sharp changes in wind farm power production on small temporal scales are referred to as wind ramp events.

To fully understand the nature of wind ramp events, and their role in the wind industry, a study is needed to provide evidence as to where wind ramp events are most probable and the sources in which they are generated in various areas of the United States. Locations of strong air-sea temperature gradients, locations of high-frequency thunderstorm activity and mountainous regions are all found to be hot spots for wind ramp events in the study conducted.

Currently, numerous studies are being conducted in attempt to increase wind forecast accuracy in the scientific community. Modern forecast verification metrics generally evaluate point-by-point amplitude-based errors that are produced by model forecasts when compared to observed fields. However, the metrics used to calculate these errors do not always generate representative values of the inaccuracies presented. Therefore, two metrics will be introduced that have the ability to quantify inconsistencies between two fields that modern metrics cannot. The Partial Hausdorff Distance will be shown to calculate spatial displacements between two fields, while the Kantorovich Distance will quantify both spatial displacements and amplitude-based errors. To apply these metrics an idealized case study along with a real case study will be presented. The idealized case study will use ellipsoids to demonstrate the ability of these metrics. In the real case study the Weather Research and Forecasting Model will be used to generate forecast fields to be evaluated against Real-Time Mesoscale Analysis wind data and NEXRAD reflectivity

data. The proposed metrics will be evaluated against traditional forecast verification metrics used in the wind industry.

Wind Ramp Events: Forecast Verification and Climatology

by
Jeffrey Aaron Craft

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Marine, Earth, and Atmospheric Sciences

Raleigh, North Carolina

2013

APPROVED BY:

Ryan Boyles	Brian Reich

Sukanta Basu
Chair of Advisory Committee

DEDICATION

This thesis is dedicated to my amazing bride-to-be, Jessica, who stood by me during this wild ride for several years. Also, my parents for their ongoing support and love.

BIOGRAPHY

Jeff Craft was born in Phoenix, AZ where he can thank the North American Monsoon for first generating his passion for severe weather. After moving to Atlanta, GA at age 5 he truly fell in love with severe weather after watching a lightning storm from the upstairs loft in his house. At age 13 he moved to Ann Arbor, MI where he was introduced to yet another passion: University of Michigan football. At age 18 he and his family moved back to Phoenix where he attended Arizona State University and obtained his Bachelor of Arts degree in Human Communication. Shortly after graduation he landed a position in marketing at a local healthcare firm in Scottsdale, AZ, and realized that he wanted to finally pursue his fascination for weather.

In 2008 he re-enrolled at Arizona State University in pursuit of his Bachelor of Science degree in Meteorology/ Physical Geography and a Minor in Mathematics. He knew that he had only received a small taste of the world of atmospheric sciences, so he began the application process for graduate school. After discussing research topics/ interests with Dr. Sukanta Basu at North Carolina State University, they began a journey together to study wind ramp events caused by severe thunderstorms, and how forecast verification metrics can be used to help quantify the ability of numerical weather prediction models to anticipate these events.

After only 16 months in the program at North Carolina State University, Jeff received a job offer from a global leader in renewable energy consulting; GL-Garrad Hassan in San Diego. Currently he is working as a Meteorologist in the forecasting department where he is in charge of maintaining their 23-machine numerical weather prediction model cluster, along with assisting in short and medium-term wind forecasts to various clients around North America.

In his free time he enjoys spending time with his bride-to-be Jessica, and their amazing and rambunctious dog Dexter. Now that they live in San Diego, they can go to the beach and enjoy the mild climate that Southern California has to offer.

ACKNOWLEDGEMENTS

I would like to acknowledge my Graduate Advisor, Dr. Sukanta Basu, who single-handedly forced me to learn more in 18 months than I ever thought possible. Your guidance has been invaluable! This work was partially supported by the National Science Foundation (AGS-1122315) grant.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Humans have been heavily reliant on fossil fuels for over a century and have already begun to feel the negative impacts of their use. Current scientific research is showing that emissions from fossil fuel burning are a leading contributor to global climate change and to air and water pollution in densely populated cities around the world [Crowley, 2000, Hoel and Kverndokk, 1996]. With the global oil supply projected to run out by the end of the century, and a need for cleaner energy sources, a shift is being made towards renewable energy sources, like wind. In fact, it has been projected by the U.S. Department of Energy (DOE) that by the year 2030, the United States will generate 20% of its electricity from wind energy.

Implementing wind power into the global arena takes more than merely erecting wind turbines where wind is climatologically abundant: profitability is heavily dependent on accurate short-term wind forecasts [Alexiadis et al., 1998, Milligan et al., 1995]. The wind forecasting portion of the wind power industry has seen recent advancements in technology due to the introduction of mesoscale numerical weather prediction (NWP) models [Costa et al., 2008, Jones et al., 2005, Larson and Westrick, 2006]. However, short-term wind forecasting has been somewhat deficient in its overall ability to produce extremely accurate wind forecasts, even with the recent industry advancements [Bossavy et al., 2012].

One major issue that arises in wind forecasting is highly attributed to the fact that wind is intermittent by nature. The breezes we feel when we step outside are very rarely at a constant velocity for longer than a few seconds, which presents issues to forecasters. If wind forecasts are inaccurate by overestimating or underestimating wind speeds, the

repercussions can be costly. It was shown that an decrease of 10% in forecasting error across the U.S. could decrease annual operating costs by over $100M [Lew et al., 2011].

There are, however, two regions of the wind turbine power curve have posed most problematic for wind forecasters: one on the lower end where wind power is approximately the cube of wind speed, and one beyond the cut-out range [Tan and Islam, 2004]. Within the cubic power region, small errors in wind speed forecasts equate to extremely large errors in wind power estimations [Celik, 2003]. For instance, a change in wind speed from 12 to 14ms^{-1} may not have an impact in wind power production if these velocities are within the rated capacity region of the power curve. However, a change in wind speed from 6 to 8ms^{-1} could have a drastic impact on power production if within the cubic power region of the power curve. Since changes in wind speed can have such a great impact on wind power forecasts, a need for predicting these fluctuations in wind in our forecasts is of great importance.

Another area of the power curve that causes concern for wind forecasters is found in the transition from rated power to cut-out. In this region winds exceeding the cut-out range, roughly 25ms^{-1} in industry-leading turbines, will force the turbine to automatically shut down for an extended period of time and seize to generate power [Tan and Islam, 2004]. In the cut-out region of the power curve a change in wind speed of 1-2ms^{-1} can cause a rapid change in power production.

An additional challenge created for forecasters is generated when wind intermittency varies on several time-scales, ranging from a few seconds (microscale), several hours (mesoscale) or days ahead (synoptic-scale). Since the time-scales evaluated can vary so drastically, different approaches have been invoked in an attempt to maximize forecasting accuracy [Taylor et al., 2009]. One such approach is using time-series data to generate a forecast for short-term wind changes [Brown et al., 1984, Huang and Chalabi, 1995]. This approach can detect and forecast small-scale fluctuations in wind by using diverse statistical approaches (e.g, artificial neural networks, autoregressive moving average models) on very small temporal scales (ie. a few seconds). However, time-series approaches to forecasting lose much of their accuracy extending out to larger temporal scales [Giebel et al., 2011].

Due to modern advancements in technology, a different approach has recently become widely used in wind forecasting for time-scales on the order of a few hours to days ahead: NWP models. By using NWP models to forecast wind fluctuations, a larger temporal

span can be evaluated over much greater spatial areas. Nevertheless, running high-resolution simulations using NWP models can be computationally expensive, and thus model forecast may only be delivered every few hours. However, many of the modern NWP models used (North American Mesoscale, Global Forecast System, etc.) use very coarse horizontal resolution which means wind speed values are averaged over large spatial areas. When spatial averaging over large areas occurs, wind fluctuations are smoothed and make it difficult for wind/power ramps to be forecast.

Even so, regardless of what approach is used, the wind energy industry will be more profitable when using a forecasting technique, as opposed to using climatology [Lew et al., 2011, Orwig et al., 2012]. Nonetheless, there is a need for vast improvement in short-term wind forecasts to optimize efficiency in the wind industry. Although government agencies deliver short-term wind forecasts to the wind energy industry, it has been stated that these forecasts are relatively inadequate.

1.1 Wind and Power Ramp Events

Arguably the most problematic attribute in short-term wind forecasting is generated by fluctuations in the wind itself. Sharp changes in wind farm power production on small temporal scales are referred to as power ramps. Commonly, short-term wind power forecasts are given 6-48 hours in advance. Therefore, not anticipating these fluctuations can generate losses in revenue for wind farm operators and electrical emergencies if not foreseen in power forecasts.

However, wind ramps and power ramps must be distinguished from one another. Power ramps are more or less a function of wind ramps; in that power ramps are caused by wind ramps in certain areas of the power curve. As stated, a power ramp does not always occur with a wind ramp. Power ramps are dependent on the region of the power curve in which the wind ramp takes place. Therefore, it is important to state that these two are not necessarily synonymous with one another. However both will be used in this thesis in different experiments and thus must be loosely defined.

Ramp events create major issues in short-term power forecasting, as they can cause massive jumps or drops in power production that may not necessarily be anticipated [Greaves et al., 2009]. Ramp-up events cause excess power production which may then be wasted if it was not predicted in earlier forecasts, while ramp-down events can create

an overestimated power forecast that can force grid operators to use other means of energy to account for the lack of wind power production.

An example of the importance of wind ramp forecasting was displayed on February 26, 2008 in Texas [Ela and Kirby, 2008]. On this day, the Electric Reliability Council of Texas (ERCOT) was forced to utilize the Emergency Electric Curtailment Plan (EECP) and call on reserve capacity in response to a synoptic-scale wind ramp event that took place. The meteorological conditions of that day were not rare by any means, and even though the NWP forecast did predict a ramp-down event to occur, the power production was much lower than expected and emergency reserve energy was needed to fuel the grid.

These types of situations are not uncommon by any means. Forecasters have the daily challenge of using downscaled NWP data to generate forecasts at a single point within a wind farm. Usually an ensemble of NWP data will be used in conjunction with specified weights on each NWP based on evaluated statistical analysis of the performance of each NWP. For instance, a certain NWP may be extremely accurate at farm X, and therefore will be heavily weighted in the ensemble to produce the most accurate results. However, seasonal atmospheric features that cause ramp events can greatly affect the quality of various NWP and therefore these weighting parameters need to be adjusted regularly by using hind-casting methods.

1.1.1 Industry Definitions

Another issue regarding ramp events stems from a lack of an industry-wide definition in the term. Without a true definition to characterize a ramp event, the industry cannot fully study their long-term effects on wind farm profitability and if improvements are being made in forecasting such events.

Although there is not an industry-standard definition for a ramp event, several working rules have been generated in the past decades. For instance, according to AWS Truepower, a power ramp can be defined as a change in 20% of a wind farm's rated capacity in 30 minutes [Freedman et al., 2008]. Alternatively, Greaves et al. [Greaves et al., 2009] define a power ramp as a 50% change in wind farm power production on a temporal scale of four hours or less. Other, less quantitative approaches, merely examine the graphical slope of the power production to identify sharp gradients in installed nameplate capacity [Focken and Lange, 2008]. Having such a wide range in temporal characteristics associated with the defined wind ramp events can make forecasting these events even

more difficult [Greaves et al., 2009].

1.1.2 Wind Ramp Forecasting Errors

When analyzing errors produced by ramp events, several considerations are needed [Bossavy et al., 2010]. One such error that can be generated during a wind ramp forecast is a spatial displacement error. These errors occur when the spatial location of the ramp event is shifted in the forecast. Somewhat linked to a spatial displacement error is a temporal displacement error: A timing error can be considered a spatial displacement error of a wind ramp event, since an event's location can be contingent on velocity which considers time and space. Further, the forecast amplitude may be incorrect, and therefore could under-predict or over-predict the magnitude of the ramp event in what is referred to as a amplitude error. All three errors are problematic, but when using a single point within a wind farm for wind and power predictions, amplitude errors become the most prevalent errors.

However, it has been posed that as the length of forecast time increases, wind ramp forecasts become increasingly difficult to accurately predict. In fact, studies have stated that it may not be possible to generate highly accurate ramp forecasts greater than 48 hours into the future [Cutler et al., 2007], while others have stated that NWP model forecasts cannot be reliable for short-term forecasts shorter than 6 hours ahead [Negnevitsky and Johnson, 2008]. Therefore, there might only be a small time-frame that wind ramp events can be predicted using NWP models. Contrary to these statements, some NWP (GFS, ECMWF, etc.) generate forecast horizons of 168 hours ahead, which argues the validity of the previously stated studies. However, regardless of forecast length, ramp forecast errors can be generated on different time-scales and by various atmospheric phenomena.

1.2 Causes of Ramp Events

Ramp events can be caused by many different atmospheric phenomena. One such cause, and likely the most common, are synoptic-scale fronts associated with mid-latitude cyclones. These fronts can be on the order of thousands of kilometers in length. Since mid-latitude cyclones tend to be long lived and travel at a relatively predictable pace, their propagation speed tends to be forecast with relatively high accuracy. However,

small-scale features embedded within these synoptic-scale fronts can generate great error in short-term wind forecasts. As pressure and temperature gradients change due to approaching and passing fronts, strong changes in wind speed become more common. These smaller-scale changes in wind speed can cause ramp events across large spatial scales.

Mesoscale outflow boundaries associated with thunderstorms are another cause of wind ramp events. Outflow boundaries can be produced by the downdraft region of a thunderstorm. As the descending air of a thunderstorm reaches the surface, the air diverges radially and can cause local wind speeds to change rapidly. Since outflow boundaries are on smaller temporal and spatial scales than synoptic-scale features, they are found to be much more difficult to predict when compared to synoptic-scale fronts. Further, since outflow boundaries lose energy with time and distance, the strongest impacts are found closest to the source which also happens to be the most difficult location to forecast thunderstorm outflows.

Another ramp-causing phenomenon are downbursts. Downbursts are formed when a thunderstorm's updraft is no longer able to suspend the amount of liquid and ice mass that is held within the updraft. If the updraft is weakened, due to evaporative cooling, then the liquid water content in the updraft can begin to descend towards the surface. In fact, downbursts can create surface winds found to travel 60ms^{-1}, that diverge radially from the point of impact with the surface [Hjelmfelt, 1988]. Rear-flank downdrafts can form on the upwind-side of a supercell, and can generate intense surface winds which can surpass wind turbine cut-out speeds [Hirth et al., 2008].

Downbursts and outflow boundaries, unlike synoptic-scale fronts, are extremely difficult to forecast, especially the timing and exact location of their formations. Midlatitude cyclones and synoptic-scale fronts tend to be relatively easy to predict, however the embedded features are usually the most difficult part to forecast. Even though downburst generation is relatively rare, the United States Great Plains (USGP) and Midwestern United States are found to be the most common regions for downburst development on earth according to the National Climatic Data Center (NCDC). This region also happens to be the location of the bulk of the U.S. wind farms.

A major source of wind ramp events in the USGP in the warmer months are nocturnal low level jets (LLJ). Nocturnal LLJ are found to be located between 50m and 400m above ground level (AGL). Coincidentally, this elevation happens to be located in the

range of many of today's turbine rotor planes [Storm et al., 2009]. These LLJ, caused by temperature gradients, topography and atmospheric stability, can generate damaging turbine-height vertical wind shear, turbulence and wind ramp events [Cosack et al., 2007]. While LLJ climatology is well-studied, forecasting the amplitude and timing of these jets are still difficult, even with advanced NWP.

Wind ramp events are found at different amplitudes with differing spatial and temporal extents. Regardless of size or intensity, ramp events tend to produce similar results: inaccurate wind forecasts and decreased profitability of power production. Due to the challenges posed by ramp event forecasting, independent system operators (ISO), owner/operators and energy traders have been unable to fully maximize their profitability based on wind power production estimates. These challenges in forecasting can be attributed to physical and numerical errors associated with NWP models.

This thesis will focus on the NWP modeling approach to short-term wind forecasting, which can cover a broad spectrum of temporal scales. Further reference to short-term wind forecasts will hereafter refer to time-scales of hour-ahead to day-ahead forecasts (6-24 hours). The main objective of this thesis is to introduce new forecast verification metrics to be used in the wind industry, and to answer the following scientific questions: 1. Can the newly proposed forecast verification metrics generate more representative values associated with spatial displacement and amplitude errors in forecasts when compared to observed fields? 2. Are there localized hot-spots for wind ramp events across the United States, and how do the frequency of wind ramp events change during the year?

It is important to disclose that the study presented will focus on an approach to address forecast verification metrics based on spatiotemporal displacement of image characteristics between observed and forecast fields. This study, however, will not attempt to improve modeling techniques, model parameterization or model performance; the NWP model will merely be used to generate various forecasts which will be compared to the observational field to quantitatively show displacement and amplitude errors in forecasts. Future work may include using these metrics to analyze model sensitivity based on varying configuration.

The structure of this thesis is as follows: Section 2 will include a 32-year wind ramp climatology. In section 3, the deficiency of modern forecast verification methods will be addressed, along with the introduction of two proposed metrics. An idealized case study will be presented in section 4. Section 5 will address the shortcomings of data availability

and introduce meteorological covariates. Section 6 will include a real case study, followed by the conclusions and future perspectives in section 7.

Chapter 2

U.S. Wind Ramp Climatology

Since wind ramp events have only become a researched topic among the short-term wind forecasting community in recent years, the amount of literature on the spatiotemporal climatology of these events is sparse. To fully understand the nature of wind ramp events, and their role in the wind industry, a study is needed to provide evidence as to where wind ramp events are most probable and the sources in which they are generated in various areas of the United States. Without identifying the frequency of the wind ramp events throughout the country, there is a gap in knowledge of wind ramp events as an atmospheric phenomena. However, presented will be a non-comprehensive analysis since the main focus of this thesis is forecast verification metrics. Nonetheless, this analysis should be sufficient in addressing localized hot-spots for wind ramp events.

To conduct an analysis of the spatiotemporal distribution of wind ramp events, relatively long-term datasets were needed. However, since wind ramp events tend to be identified on a time-scale of a few hours, the temporal resolution of the data used did not need to be extremely high. Therefore, for this analysis, 32 years of North American Regional Reanalysis (NARR) data were used to locate not only general hot-spots for wind ramp events, but also their proximity to modern wind farms and the atmospheric conditions that create high-frequency events in certain locations.

NARR data are generated and monitored by several government agencies, including but not limited to National Oceanic and Atmospheric Administration (NOAA), NCDC and National Center for Environmental Prediction (NCEP). The reanalysis data are created by use of the Eta Data Assimilation System (EDAS) which generates 3 hour temporal resolution datasets for North America. Further, data from radiosondes, aircraft

data, dropsondes and geostationary satellites are assimilated into the NARR datasets.

Since the NARR datasets have a temporal resolution of 3 hours, which is consistent with some industry-used wind ramp definitions, no temporal scaling was needed in the analysis. One drawback may be that the horizontal resolution of the data is somewhat coarse at 32km. However, for this study the regional wind ramp climatology is desired, the horizontal resolution of the datasets should not largely affect the validity of the results throughout the domain(s) used. But, again, spatially-coarse NWP data are a major shortcoming of wind ramp forecasting since winds are averaged over a large area and thus lose much of the natural fluctuation found at high resolutions.

The first step in the post-processing of the data was to generate an algorithm to extract and format the data. Using this algorithm, the wind data are translated from wind speeds to wind ramp events by binning the grid cells that experience a change in wind speed (increase or decrease) greater than or equal to 4ms^{-1} during one time step (three hours). If a change in wind speed exceeding the threshold does occur, the grid cell in which it occurred will count one event. This will be administered during the entire 32 year time-frame. When these values are mapped, areas that experience higher frequency of wind ramp events were revealed. These regions will be evaluated in attempt to explain the atmospheric phenomena that occur to cause the high frequency of wind ramp events during the 32 year span (as seen in Figure 2.1).

Although some of the phenomena that generate wind ramp events do not necessarily produce only a single event within a three hour time-frame. For instance, a squall line moving through the Midwest will likely invoke wind ramps at each location during its life-cycle. Therefore, even though it is one atmospheric event, it can cause ramp events for hundreds or thousands of miles. However, for this evaluation, we will assume that one event takes place during each time-step.

Over the entire 32 years, there does seem to be a few hot-spots that are generated by using this algorithm. Over most of the Western CONUS, there are pockets of high frequency wind ramp events over the Rocky Mountains, and the lower deserts. This generates evidence that these features are not merely terrain-induced ramp events since there also tends to be high frequencies where lower terrain is present. Further, on the eastern portion of the CONUS there does seem to be topographically enhanced wind ramp events associated with the Appalachian Mountains. Furthermore, areas off the coast of the Atlantic Seaboard and Gulf of Mexico appeared to experience high frequencies of wind

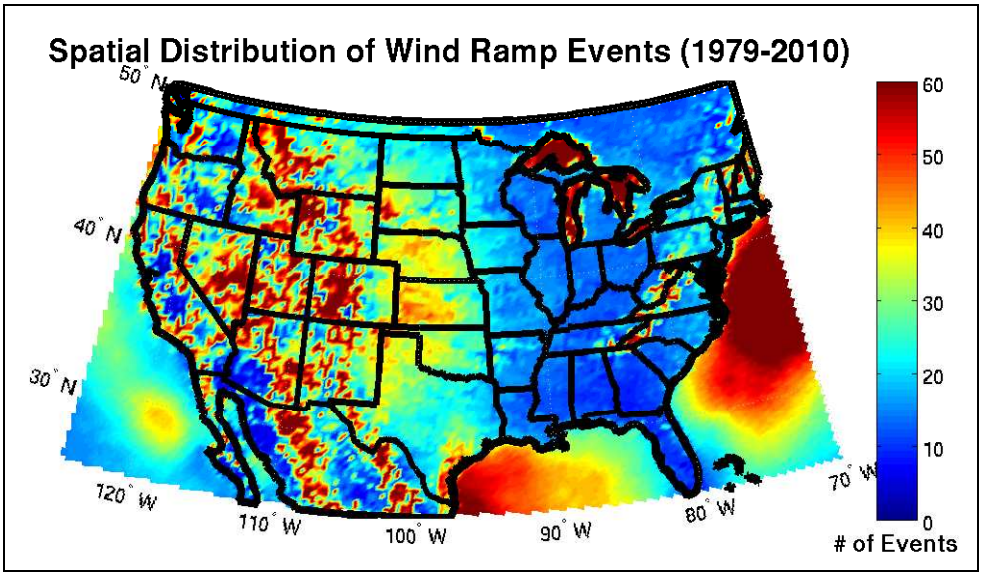

Figure 2.1: 32 year annual average wind ramp climatology using NARR data. Wind ramp events were defined as changes in wind velocity greater than or equal to 4ms^{-1} in three hours.

ramp events as well. Possible reasons for these areas will be provided in this section.

When evaluating the entire 32 year span, only long-term conclusions can be yielded from the evaluation. To see the year-to-year and seasonal distribution, a more in-depth study was conducted. When individual years were analyzed qualitatively, some drastic differences in wind ramp frequency between years transpired. In Figure 2.2, two years are shown: 1997 and 2008. It is quite easy to see that the spatial distribution and frequency of wind ramp events in these domains are drastically different between the two years. Because there did seem to be an annual change in distribution of wind ramp events, an evaluation of annual climate impacts was needed to see possible causes of such large annual fluctuations.

2.1 Impact of ENSO

Since there did seem to be a strong variability from year-to-year, an investigation of the influence of the El Niño Southern Oscillation (ENSO) was conducted in attempt to present evidence of how teleconnections might change the distribution of wind ramp

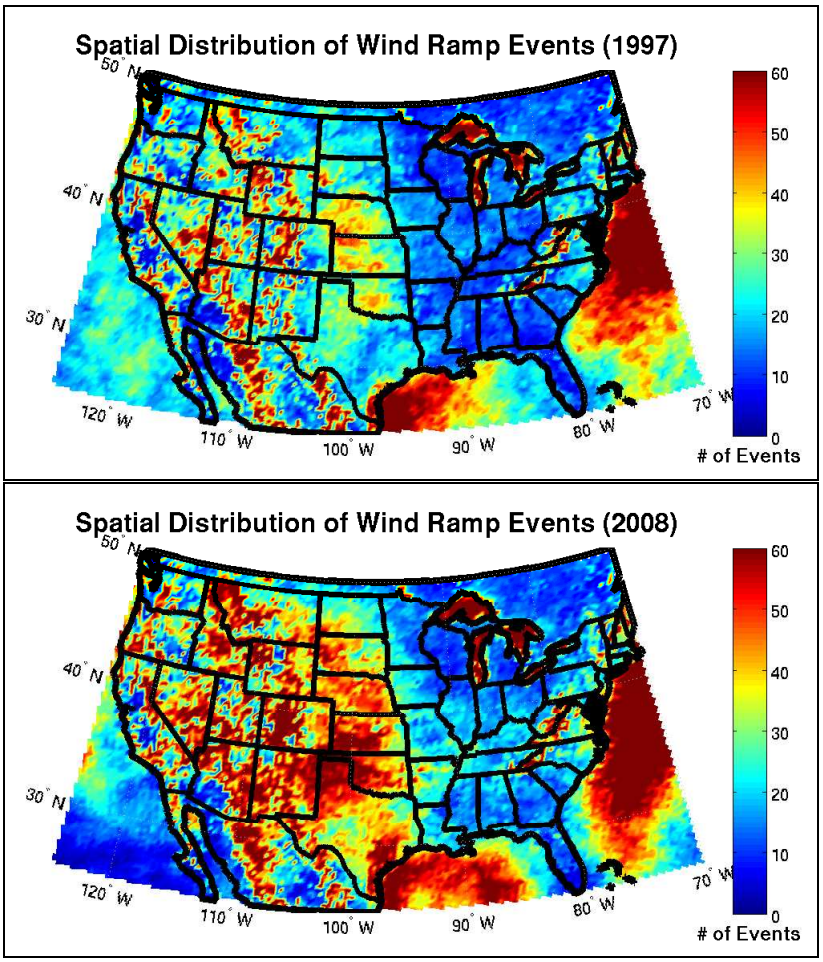

Figure 2.2: The annual average of two years of wind ramp distributions showing exceptionally different results annually over the Continental United States. The distribution of wind ramps in 1997 (top) was much less than in 2008 (bottom).

events. El Niño (La Niña) years are caused by warm (cold) sea-surface temperature (SST) anomalies in the Tropical Pacific Ocean. These SST fluctuations can cause the tropical and polar jet streams to vary, which greatly impacts weather patterns across North America Trenberth [1997]. Since El Niño years are found to generate stronger Pacific Jet Streams across the southern portion of the United States, it was hypothesized that El Niño years would increase wind ramp events, especially across the Southern United States.

The El Niño and La Niña years used for this study are defined as strong ENSO years, according to the Oceanic Niño Index (ONI) created by the NCEP. From a qualitative, standpoint there was no real large impacts of the ENSO compared to the neutral phase years. As seen in Figure 2.3, when evaluating La Niña, El Niño and neutral phase years, there does not seem to be an overwhelming trend presented. However, the frequency of wind ramp events over the Central Great Plains, Atlantic Coast and Gulf Coast did increase during El Niño years compared to La Niña and neutral phase years. Furthermore, a small area of increased frequency off the coast of Baja Mexico in the Pacific was presented during El Niño years. However, this is purely a qualitative analysis based on the maps of the spatial distribution of wind ramp events. A statistical analysis would likely indicate whether there is actually an increase over the entire domain.

As seen in Figure 2.4 (bottom), the mean value of wind ramp events for the domain during La Niña years was shown to be slightly higher than neutral phase and El Niño years. However, El Niño years were found to have a lower mean value than neutral phase years as well, possibly signifying that El Niño years may decrease overall spatial distribution of wind ramp events. A second plot,2.4 (top),shows the maximum value of wind ramps within the domain for each year. This plot shows that El Niño and La Niña years produced higher maximum values of wind ramp events as compared to neutral phase years, and that La Niña years, had a higher average value across the domain. Therefore, one could assume that over the CONUS, that during a La Niña year more wind ramp events will be likely when compared to neutral phase years. The same can not necessarily be said about El Niño years, since the mean value was found to be less than neutral phase years.

However, based on the analysis shown in Figure 2.4, there did not seem to be a large increase in the Southern CONUS associated with the increased jet stream over the region. In fact, spatially, there does not appear to be much difference between El Niño and La

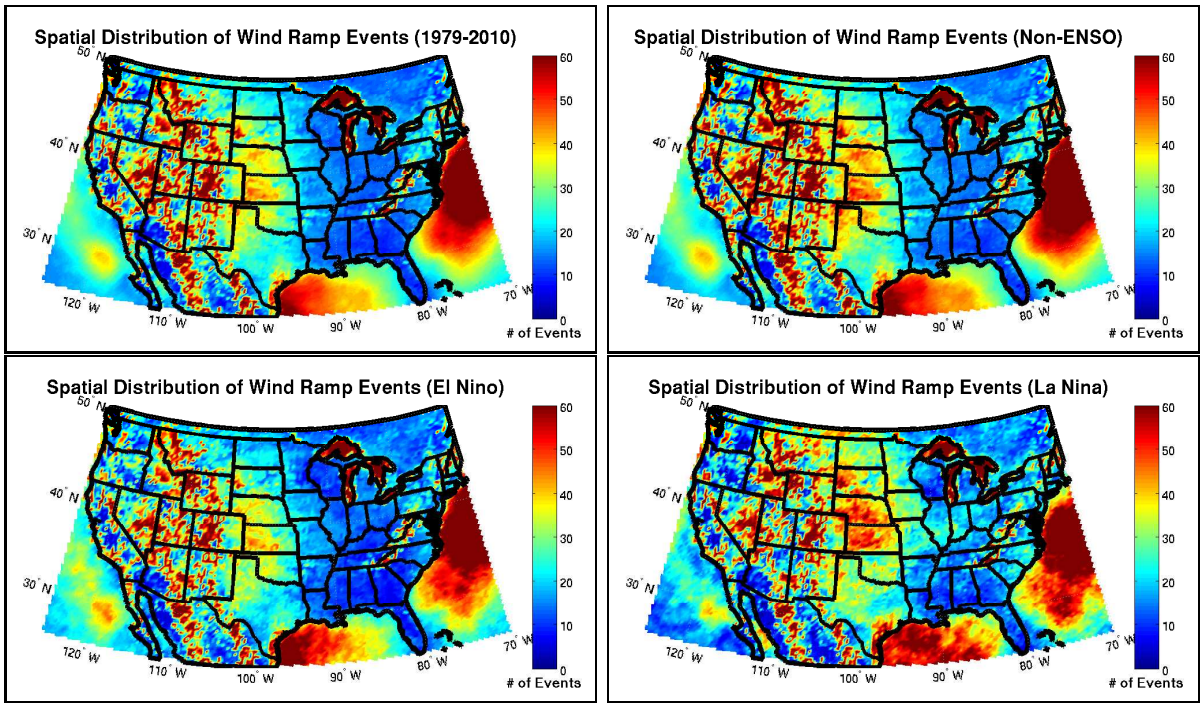

Figure 2.3: Wind ramp spatial distribution for 32 years (top left), during neutral phase years (top right), El Niño years (bottom left) and La Niña years (bottom right). El Niño and La Niña years used are considered strong El Niños and La Niñas, defined by the NCDC. Color bars are set to scales based on the annual average of wind ramp events.

Niña distributions at all. This may be partially due to the coarse horizontal resolution of the dataset used. Regardless, no significant results came from this preliminary ENSO analysis.

2.2 Inland Wind Ramp Events

When evaluating the 32 year analysis of wind ramp events over the CONUS, several high frequency locations appear. The Western United States and Rocky Mountain regions appear to experience a large area of high wind ramp event frequency during the examination period. A further investigation is needed to evaluate if these events are merely topographically induced, or if atmospheric roles are being played in the formation of wind ramp events. This will be explained later in this section.

Another region in which wind ramp events seem to be predominant is the Central Great Plains region, referred to as Tornado Alley. Tornado Alley is the geographic expanse from the eastern side of the Rocky Mountains to the Mississippi River Basin, from Mexico to Canada. In this region of the U.S. many geographic features aid in wind ramp event-generation year-round. Proximity to the Gulf of Mexico ensures copious amounts of moisture in this region. Warm, moist Gulf air is fuel for thunderstorm development. During the Spring and Summer months, severe weather is extremely active in this region, attributed to greater convective available potential energy (CAPE) generated from increased dew points and warmer temperatures.

Furthermore, sharp moisture gradients form in this area of the country, as warm moist air from the Gulf of Mexico clashes with warm dry air from the desert southwest. The area where these two airmasses become collocated is called the dryline. The buoyancy gradients generated by these air masses aids in severe weather generation Lin [2007] cause horizontal convective rolls (HCR) to form, which have been found to play a role in supercell formation Peckham et al. [2004].

During the cooler months, synoptic-scale frontal passages are extremely common in this region, as the polar jet stream tends to dig south from Canada, aiding in cyclogenesis. The polar jet stream also contributes to generating strong vertical wind shear, which helps intensify storms. Therefore, this region of the United States, although rich in wind resource, also carries with it the tendency to generate weather events that can make wind forecasting a challenge.

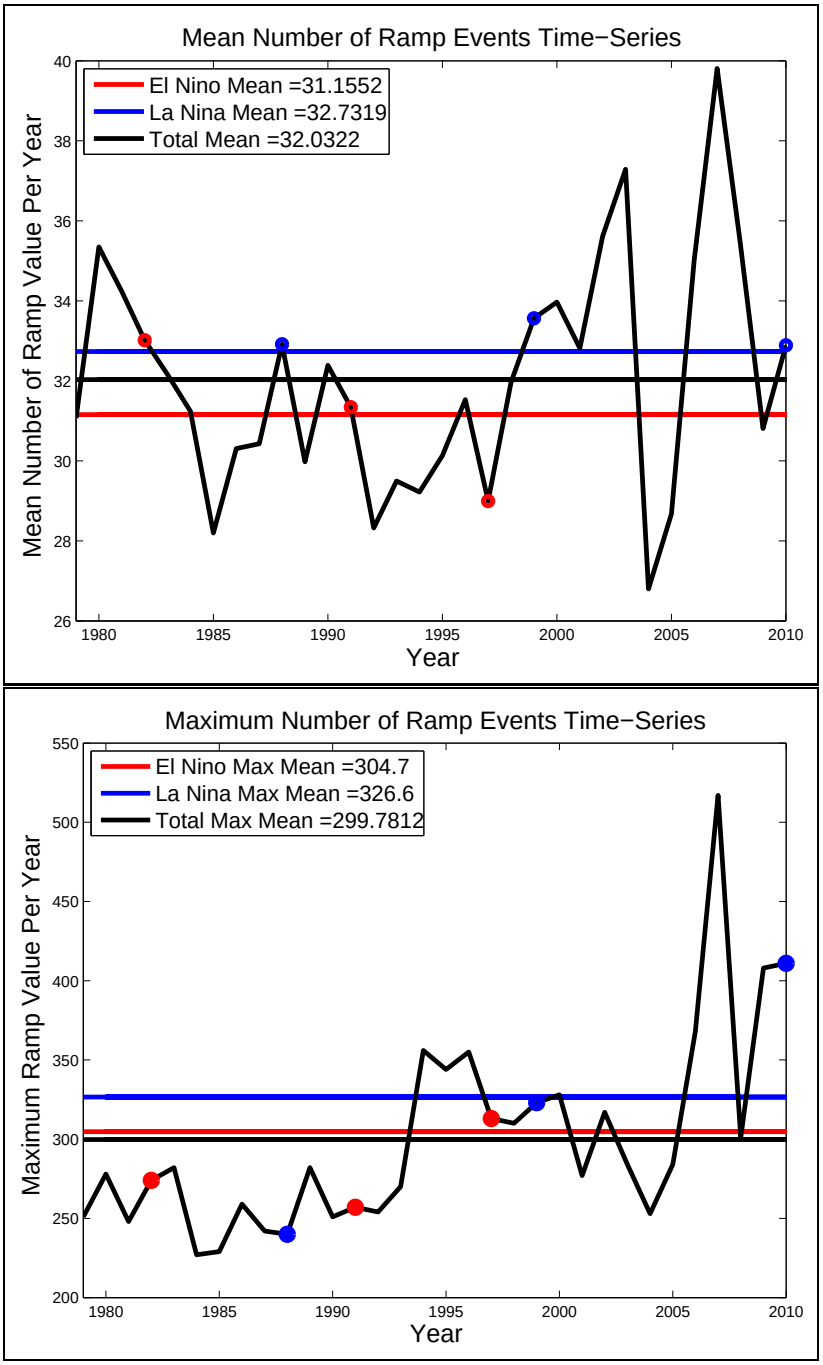

Figure 2.4: Plots of mean (top) and maximum (bottom) values of wind ramp events over the CONUS domain. La Niña years are shown to have higher domain-wide average values of wind ramp events, and higher maximum values. However, although El Niño has a higher maximum value than neutral phase years, the mean values for El Niño years are lower than neutral phase years. Large dots indicate strong ENSO years.

When using an algorithm to separate and confine the wind ramp events based on various seasons, the inland distribution changes drastically. It becomes apparent that during the spring months, as previously stated, wind ramp events in the USGP become an increasingly dominant feature. Further, the spatial expanse of high-frequency ramp events over much of the Rocky Mountains sheds light that the wind ramp events that take place over this region are not purely topographically forced: but instead rely on seasonality. Furthermore, the distribution of high frequencies in this region is also found in the lower desert region. Therefore, all of the high frequencies in this area cannot be assumed to be topographically forced.

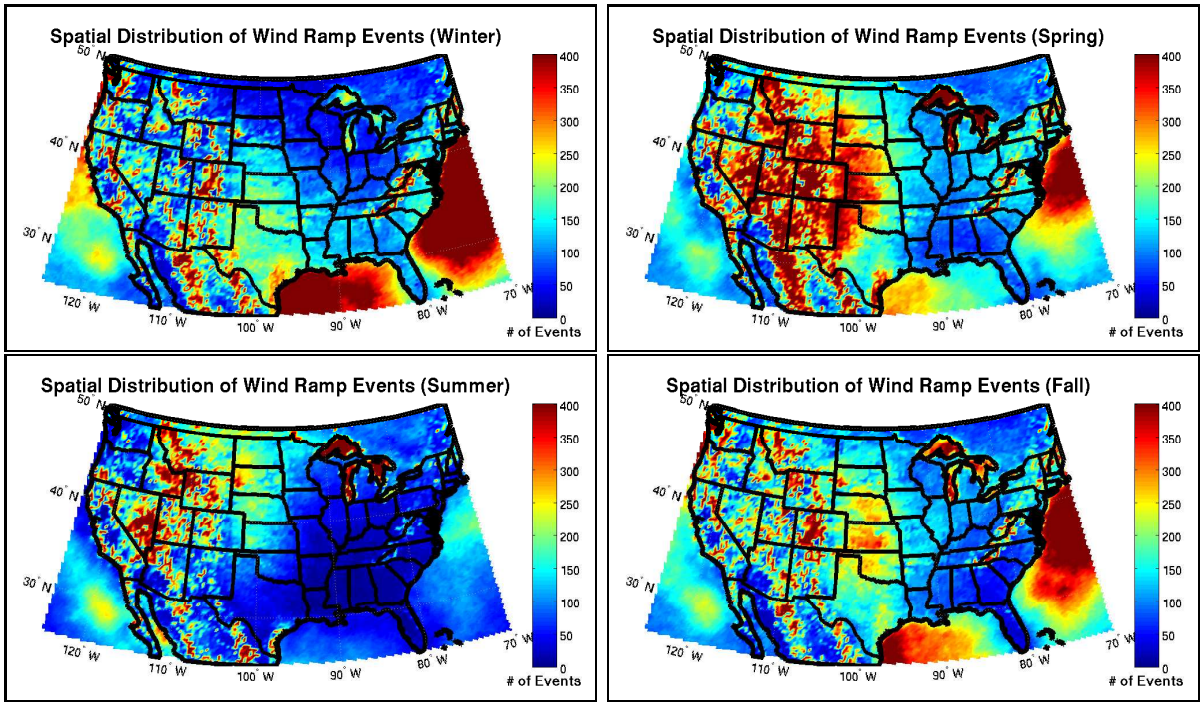

Figure 2.5: Total wind ramp distributions for the U.S. with varying seasons.

2.3 Coastal Wind Ramp Events

Areas of high frequency wind ramp events are also found in regions along the Mid and Northern Atlantic Coast and the Western Gulf of Mexico. For brevity, a discussion of the Atlantic Coast will be used to discuss some causes of coastal wind ramp events. A major contributor to the increased frequency of coastal wind ramp events in these regions is due to the sea-surface temperature (SST) gradients generated by the Gulf Stream off the eastern coast of the United States. Wind speed increases have been found to be very closely related to SST gradients. In fact, Oneill, et al. O'Neill et al. [2010] presented that a Gulf Stream SST gradient of 2-4^{0}C could increase wind speeds 2ms^{-1}. Further, Arritt Arritt [1989] found that warm SST's had a large affect on increasing wind speeds and could be strong enough to induce convective mixing within the marine boundary layer and found that when sea-surface water was warmer than the air above it, strong winds tended to develop within the boundary layer. However, these phenomena are not unique to this geographic area: a study conducted in Central America showed that strong near-surface winds were generated in response to SST gradients Xie et al. [2005].

Since there is a strong connection between SST gradients and wind speed accelerations, inland and coastal domains should yield different wind ramp event distributions. However, it can be seen that coastal wind ramp event frequency is seasonally dependent. It has been shown that during the winter months the Gulf Stream can generate strong near-surface wind convergence contributed to pressure adjustments to SST gradients Minobe et al. [2008]. These results are consistent with this study, in that there appears to be a much higher frequency of wind ramp events in the coastal regions during the winter months. Further, as the surrounding ocean water warms during the summer months, the SST gradients would diminish, yielding a much lower frequency of coastal wind ramp events.

Therefore, the results presented in this section appear consistent with several studies that introduce the notion of SST gradients having an effect on near-surface wind speeds. The seasonal distribution in this study can be seen in Figures 2.5 and 2.6, where coastal regions have much higher frequencies during the winter months, and virtually no frequency during the summer.

The climatology of wind ramp events over the United States is highly variable on both spatial and temporal scales. Although ENSO years were not found to have a strong impact on the frequency and distribution of wind ramp events, seasonal variations

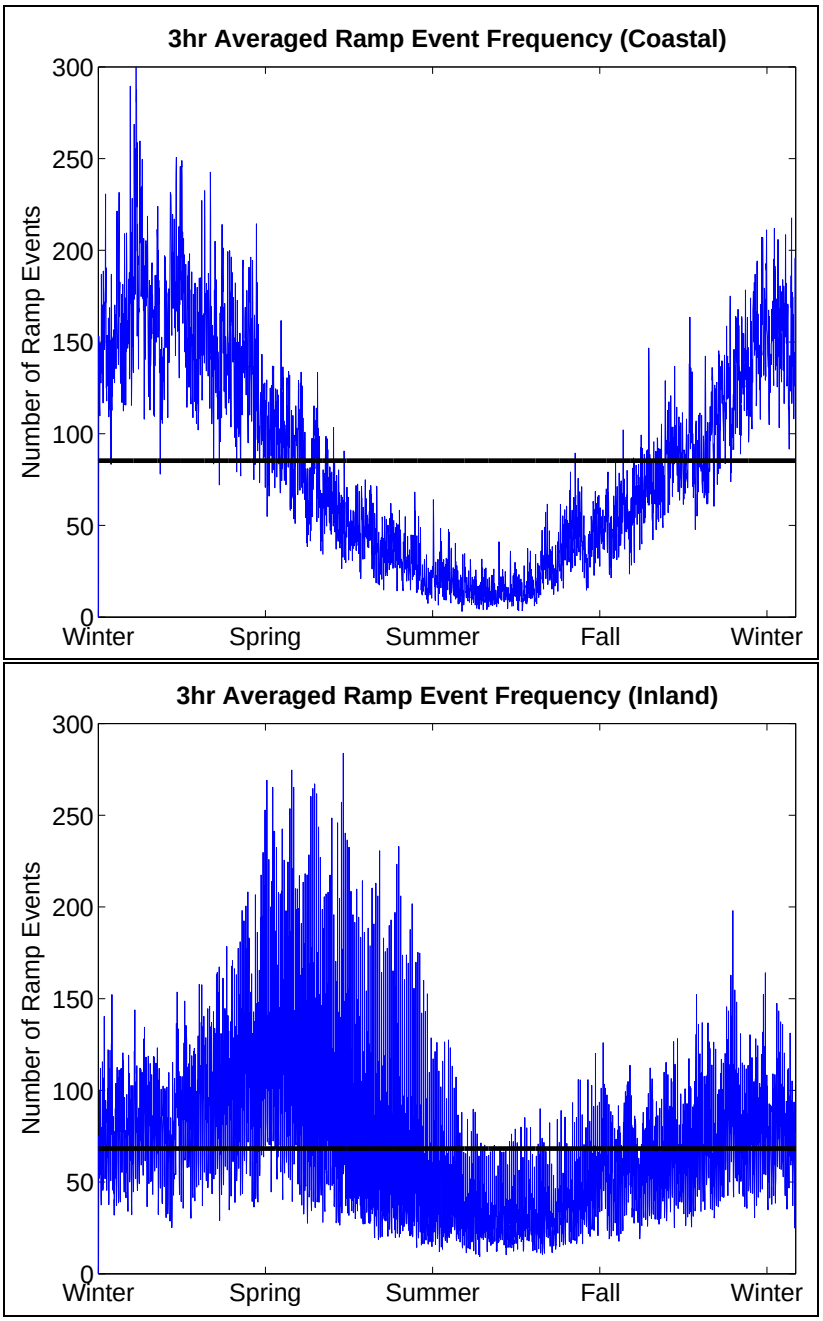

Figure 2.6: Wind ramp occurrences of coastal (top) and inland (bottom) domains. The coastal domain shows a clear maximum in frequency in Winter and minimum in Summer, while the inland domain illustrates the higher frequency during the Spring and early Summer convective season. The black line indicates the mean value of wind ramp events at each point over that domain.

were very apparent in both inland and coastal regions of the domain. Winter months were found to generate high-frequency wind ramp events in coastal regions due to SST gradients and air-sea interactions of cold near-sea surface air. Late Spring and early Summer was shown to decrease coastal activity, but drastically increase inland frequency of wind ramp events, especially over the USGP and Rocky Mountains due to increased convective activity. Further, there tends to be a northward shift in wind ramp events during the Summer and a southward shift during the Winter months, highly attributed to the migration of the Polar Jet Stream.

Chapter 3

Traditional and Proposed Forecast Verification Metrics

Currently, numerous studies are being conducted in attempt to increase wind forecast accuracy in the scientific community. However, the means by which the accuracy of the wind forecasts in these studies are evaluated are not necessarily the most representative when it comes to quantifying how well forecasts are improving with time. Modern forecast verification metrics generally evaluate point-by-point amplitude-based errors that are produced by model forecasts when compared to observed fields. However, the metrics used to calculate these errors do not always generate representative values of the inaccuracies presented.

There are attempts being made in the wind industry to increase short-term wind forecast accuracy. For instance, the Wind Forecast Improvement Project (WFIP) is a study attempting to provide added insight into the importance of increased observational wind data and enhanced modeling approaches in wind forecasting Freedman et al. [2008]. The WFIP is a two-year project, beginning in 2011, funded by the Department of Energy (DOE) and NOAA, that spans two areas of the United States (i.e., the Southern and Northern Great Plains). By using a slew of observational data from remote sensing systems, upper air soundings, meteorological towers and turbine anemometers, WFIP attempted to show that localized wind forecasts will gain accuracy with increased observational data. However, it will be shown in this thesis that the widely-used forecast verification metrics currently used in the wind industry are deficient in addressing the forecast errors addressed.

An example of the inefficiency of using some of the modern forecast verification metrics can be seen in the study by Ahijevych, et al.Ahijevych et al. [2009], in the National Center for Atmospheric Research (NCAR) Spatial Forecast Verification Methods Inter-Comparison Project (ICP). In this study, geometric shapes representing a rain forecast and observed rain fields are depicted by spatially displaced ellipsoids. During this study, various statistical metrics were used to quantify the error associated with each of the forecasts, and shown that the spatial displacement of these forecasts is not being depicted with commonly used forecast verification metrics.

The overall motive to this thesis is to illustrate not only that traditional forecast verification metrics used in the wind industry do not fully quantify the spatiotemporal and amplitude errors in forecasts, but that outside-of-the-box thinking can produce new metrics that can achieve what traditional metrics lack. Two new metrics will be addressed and tested against traditional metrics in hopes of illustrating that the traditional metrics used do not generate representative values for the aforementioned errors.

A recently created toolbox used to increase forecast verification accuracy was generated called Method for Object-based Diagnostic Evaluation (MODE) which is part of the Model Evaluation Toolkit (MET) in the Developmental Testbed Center (DTC). This toolbox does not use the aforementioned traditional metrics when delivering a measure of forecast accuracy. However, this toolkit does use around 50 metrics to deliver an accuracy value based on a weighting function of the metrics being used. Such metrics include, but are not limited to, centroid distance, area ratio, boundary distance, etc. These values are all calculated and then are weighted to generate a single value called the interest level. If the interest level exceeds a certain threshold value, the two fields are considered a match. However, a shortcoming of this approach is that the weighting values can drastically change the interest values and can be arbitrarily chosen. Further, the idea of generating 50 values for verification purposes is something that is not necessarily desirable in the wind industry. The wind industry usually attempts to represent forecast accuracy by delivering two or three statistical metric values. Therefore, although the MET may be powerful in some cases, this may not be the best answer for forecast verification in the wind industry.

3.1 Traditional Forecast Verification Metrics

As stated, the NCAR ICP study used several traditional metrics in forecast verification of ellipsoids. One such metric used in the Ahijevych et al. Ahijevych et al. [2009] study also happens to be one the most widely used forecast verification metric in the wind industry today: the root mean squared error (RMSE).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Fcst_i - Obs_i)^2}{n}} \tag{3.1}$$

Using the RMSE for evaluating the similarity of the amplitude between two objects can be useful, but will only account for amplitude-based errors in overlapping fields. For instance, when an area of the forecast field is collocated with a portion of the observed field, the RMSE can be calculated to quantify the absolute value of the mean difference between forecasts and observations. However, using the RMSE can yield a double-penalty by generating error for the pixel-to-pixel difference in the amplitude fields Gilleland et al. [2009].

An example of the inefficiency of using RMSE for forecast verification can be addressed when using the images from the NCAR-ICP. The RMSE calculated for each of the images does not depict the overall displacement of the amplitude cores in the forecast and observation fields. Furthermore, the RMSE values did not vary depending on the objects' shape, size or displacement differences. Using the RMSE yielded identical values for for several cases, even though one forecast was clearly much more representative of the observational field, as compared to the other forecasts.

Another metric used to illustrate the similarity between two fields is the correlation coefficient (r).

$$r = \frac{\sum_{i=1}^{n}(Fcst_i - \overline{Fcst})(Obs_i - \overline{Obs})}{\sqrt{\sum_{i=1}^{n}(Fcst_i - \overline{Fcst})^2 \sum_{i=1}^{n}(Obs_i - \overline{Obs})^2}} \tag{3.2}$$

The value of r can be used to describe the similarity between the data within the forecast and observed fields. However, like the RMSE, without an overlapping field, the metric will not accurately quantify the dissimilarities between the two fields. Furthermore, using a linear regression, which is merely r^2, will only exacerbate the error between the two fields when they are not overlapping.

Another commonly used metric in forecast verification is the mean absolute error

(MAE). The MAE is generated by averaging the absolute difference between the forecast and observation.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Fcst_i - Obs_i| \tag{3.3}$$

This metric, too, is deficient when comparing and quantifying the differences between two non-overlapping fields.

Frequency Bias or Bias (FB or B) can be calculated by the dividing the total forecast yes by the total observed at each grid point.

Table 3.1: Contingency Table

	Observed Yes	Observed No	Marginal Total
Forecast Yes	a (Hits)	b (False Alarms)	a + b
Forecast No	c (Misses)	d (True Negative)	c + d
Marginal Total	a + c	b + d	a + b + c + d

$$B = \frac{(a+b)}{(a+c)} \tag{3.4}$$

Again, if there is a displacement in amplitude of forecast and observed fields, the bias will not illustrate the difference between two differently displaced forecasts.

The equitable threat score (ETS) is a metric that calculates the ratio of accurate forecasts to total forecast yes and misses, with a correction expected by chance (a_r).

$$ETS = (a - a_r)/(a + b + c - a_r) \tag{3.5}$$

where,

$$a_r = (a+b)(a+c)/n \tag{3.6}$$

The ETS will encounter similar problems as the aforementioned metrics.

As shown, the forecast verification metrics currently being used in the wind industry do not quantitatively represent the accuracy of spatially displaced forecasts. This is problematic in that this does not enable forecasters the ability to evaluate the forecast improvements with time.

3.2 Proposed Metrics

Since there appears to be shortcomings in using the most common forecast verification metrics, new metrics are needed to quantify the displacement and amplitude errors that are generated in forecasts. The hypothesized outcome of this study will be to illustrate the proposed metrics can be used to better differentiate the spatial displacements and amplitude errors in forecast fields when compared to the observed fields.

3.2.1 Partial Hausdorff Distance

The Hausdorff distance (HD) has been used in image comparison studies for decades.

$$H(A,B) = \boxed{max(h(A,B), h(B,A))} \tag{3.7}$$

where

$$h(A,B) = \boxed{max_{a \epsilon A} min_{b \epsilon B} ||a-b||} \tag{3.8}$$

The HD is merely the distance each point within the observed field lies near a point in the simulation field, and vice versa [Huttenlocher et al., 1993]. To calculate h(A,B), the distance from a ϵ A to any point in B is calculated. Next, the shortest distance between each individual point a ϵ A is determined. This process is completed for the entire A and B field, and is ranked according to distance. The largest value is h(A,B), called the directed Hausdorff distance. This basically means that every a-value will be within h(A,B) of any point in the field B. The same process is completed for h(B,A), and the largest of the two values (h(A,B) and h(B,A)) is the HD.

For this study a computer algorithm is used for the calculation of the HD. This metric has been shown to use minimal computational expense as compared to other metrics in image comparison [Di Gesú and Starovoitov, 1999]. Since HD is a spatial displacement metric, and does not take into consideration the amplitude of the values being compared, a binary field is needed [Venugopal et al., 2005]. Therefore, the fields being used need to be threshold at a certain value.

After applying the threshold value to the fields, binary fields are generated, and images of the observed data and forecast data were compared (shown in Chapter 7). Since small features can be found in some of the images that may not have an impact in the overall wind field, a 75 percentile of similarity is calculated to remove outliers, thus the yielded

value is deemed the Partial Hausdorff Distance (HD_p). This process is conducted by ranking the distance values for each field, and omitting the top 25 percent which would be declared values in the far-field that may be of less importance.

As both images are in binary form, the overall spatial displacement of the threshold values can be compared using the HD_p. A HD_p value of zero means that the images are exact matches, while larger values are generated by greater spatial displacement. However, the HD_p can only be used to quantify spatial displacements between binary images. This can be very useful, however a threshold value can be arbitrarily chosen, and therefore the amplitude cannot be evaluated. For this reason, another metric will be used to evaluate the spatial displacements and amplitude-based errors in the forecast and observed fields.

3.2.2 Kantorovich Distance

It has been stated that a spatial displacement metric can be a very valuable tool when evaluating various data fields. However, spatial displacement quantification is only half of the achievable verification needed to fully understand if forecast fields are well-representative of the observed fields. For this reason, a metric is needed that will generate a single value for both the spatial displacements, along with the amplitude-based errors. The metric used for this evaluation is the Kantorovich Distance (KD).

In 1942, Kantorovich created a metric that stemmed from Monge's 1781 work on optimization of mass transfer Benamou and Brenier [2000]. The KD can be thought of the least expensive way to transport one field to another. The cost of this transportation is defined by a distance-function, which measures the distance between the grid cells of the image Kaijser [1998]. The KD can be applied to forecast verification in a similar way the HD_p is being applied: using two fields that are being compared for similarity, the metric will yield a smaller value to transport one field to another if the fields are more similar. In fact, the KD will quantify the spatial displacements and amplitude-based errors on a cell-by-cell comparison, to generate a single value.

Although a threshold value does not need to be applied to the fields, one can be applied to decrease the computational expense of the calculations. Applying a threshold to the fields using the KD will not generate a binary field as seen in the use of the HD_p. In fact, the threshold will be applied, and only values exceeding the threshold will be held in place during the calculations. All values below the threshold will be set to zero.

This can be helpful in thunderstorm fields, since low values of reflectivity tend to not be associated with thunderstorms.

As stated, the KD can be thought of as a transportation problem, where supply (source) and demand (destination) nodes are used as the observed and simulated fields, respectively. The least cost to transport the values of the observed field to the simulated field is calculated based on the distance from one node to another, and based on the supply and demand at these nodes Hillier and Lieberman [1977].

The calculation of the transportation problem using linear programming is merely the minimum of the summation of the cost (c_{ij}) and the distance (x_{ij}). The KD can be written as:

$$KD_{raw} = \boxed{min(\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij})} \tag{3.9}$$

assuming $\sum_{i=1}^{m} x_{ij}$ = demand$_i$, for i=1,2...m, $\sum_{j=1}^{n} x_{ij}$ = supply$_j$, for j=1,2...n, while c_{ij} and $x_{ij} \geq 0$.

However, the calculations associated with determining the KD can be computationally expensive, compared to the HD$_p$.

Bias Correction Coefficient:

Transportation problems in general are constructed to calculate the cost distance to transfer values from one field to another field. However, this is based on the idea that the supply will equal the demand. In other words, the summed amount of output from the supply nodes will equal the summed amount of the values at the demand nodes. Overall there is no loss or gain of values. However, in the case of comparing observed reflectivity or wind fields with simulated fields, there might not be a equal value from the two fields being used. Therefore, a bias correction coefficient (BCC) is needed to ensure that an apples-to-apples comparison is being conducted.

In order to compare two fields with equal summed values, a bias correction algorithm was needed when the two fields do not have an equal summed value (ie. supply $\neq$ demand). To do this, the threshold is first applied to the fields, such that only the values exceeding the threshold are left. Next, all of these values are summed within the fields. The next step is to calculate the ratio, or BCC, of the sum of the observed field and the simulated field. Finally, the BCC is multiplied to the forecast field. In doing this, the

ratio with a simulated field with a higher summed value than the observed field will yield a value less than 1. This multiplied to the simulated field, the field will decrease in value to equal that of the observed field. The same holds true for a summed observed field with a higher value than the simulated field. This bias correction generates two fields that are equal in summation, and can thus be compared as equal supply and demand nodes using the transportation problem. The KD algorithm is then administered to the equal fields. This value will be called KD_{raw}. After the KD_{raw} is determined, the BCC is re-applied to the value to generate the KD, as seen in equations 3.10 and 3.11. Shortcomings of bias correction will be discussed in the conclusions section.

$$if BCC > 1, KD = \boxed{KD_{raw} \times BCC} \tag{3.10}$$

and

$$if BCC < 1, KD = \boxed{KD_{raw}/BCC} \tag{3.11}$$

When analyzing the BCC, the farther the value is from 1, the less representative the forecast field is in amplitude to the observed field. For instance, if the sum of the observed field is extremely large, and the forecast drastically underrepresented the magnitude of the field, then the BCC would be very large. Likewise, if the forecast wildly overestimated the observed field, then the value of the BCC would be extremely small. The BCC can be thought of as a tie-breaker for KD values that are similar. An example would be if there were two forecasts, F1 and F2 and an observed field, Ob. If F1 is closer in magnitude to Ob than F2 is, and F1 and F2 are equally displaced from Ob, then it can be said that F1 is a 'better' forecast. However, after the bias correction is applied to the forecast fields, there is a chance that F1 and F2 could yield very similar KD values. Therefore, since the BCC associated with F1 would be closer to 1 than the BCC associated with F2, F1 would be considered the better forecast.

Chapter 4

Idealized Case Study

4.1 Idealized Cases

In the aforementioned (NCAR ICP) study conducted by Ahijevych, et al., several geometric shapes were used to show how spatial displacements in forecasts are not well-represented by most common forecast verification metrics Ahijevych et al. [2009]. In fact, the most widely used metrics in the wind industry were unable to distinguish between three relatively different forecasts: all three forecasts received the same value for RMSE, MAE, ETS, correlation coefficient, etc. Therefore, three similar idealized case studies will be used for validation of the HD_p and KD as forecast verification metrics that can generate values that are consistent with the forecast error. The values generated by the proposed metrics will be compared to the aforementioned traditional metrics widely used in the wind industry for forecast verification.

Case 1 will illustrate how both the HD_p and KD can be used to quantify spatial displacements in forecast fields with magnitudes identical to the observed field. This is important to show since both metrics will be used to calculate spatial displacement errors. Case 2 will be identical to Case 1, however the amplitudes of the forecast fields will be increased, while the observed field will stay constant. Case 2 will be used to show that when there is an amplitude error in the forecast field, the values of KD increase due to the amplitude error; Case 1 should yield lower KD values than Case 2 due to no amplitude error on Case 1, but identical HD_p values since the displacement has not changed. Case 3 will be used to illustrate that forecast fields with varying positive or negative amplitude error will result in higher KD values, when compared to no amplitude

error, and that increased amplitude error will yield increased KD values. Further, since all three forecasts are displaced the same distance, the HD_p value should be identical for all three forecasts. Furthermore, since all three cases will be using the same observed field, each idealized case study can be intercompared.

4.1.1 Idealized Case 1

In reference to Figure 4.1, the observed field is the figure in the upper left, while the other three figures show the forecasts (to the right of the observed field) along with the representative observed field (at 200 on the x-axis). In all three forecasts, the forecast field is spatially displaced, and no area is overlapping. The values in the observed and forecast fields are 100 in Case 1.

Forecast 1 is displaced slightly, while Forecast 2 is displaced relatively far from the observed field. Forecast 3 in this case has an inverted aspect ratio (compared to Forecast 1 and 2). These three forecasts are the three that yield the exact same values using the traditional forecast verification metrics in the Ahijevych et al. work. It is relatively easy to see in Figure 4.1 that Forecast 1 is superior to the other forecasts, with only small spatial displacement. Further, Forecast 3 should yield a smaller score for displacement than Forecast 2, but a larger value than Forecast 1 (assumed HD_p and KD values: Forecast 1 <Forecast 3 <Forecast 2).

It is shown in Table 4.1 that the HD_p and KD were able to quantify the displacements in these forecasts, and accurately generate the lowest scores for Forecast 1, while Forecast 2 yielded the highest scores. Further, Forecast 1, 2 and 3 were each given unique HD_p and KD values, which appear to be consistent with each of their respective spatial displacements. Finally, it should be noted that Forecast 1 and 2 are constructed to have an identical aspect ratio to the observed field, with only a spatial displacement to the 'east'. Therefore, it is important that the metrics were able to illustrate that Forecast 1 is a better forecast than Forecast 2, and by an amount that is attributed merely to the displacement. This was shown to be true in both (HD_p and KD) metrics in Table 4.1.

Further, it can be seen that the BCC values for the forecasts are all 1, meaning the summation of the amplitudes in their fields exactly match that of the observed field. Therefore, it can be assumed that the forecast and observed fields are identical in magnitude over the spatial domain, since their areas are the same.

In this case the RMSE, MAE, ETS and r^2 are all identical for each forecast, respec-

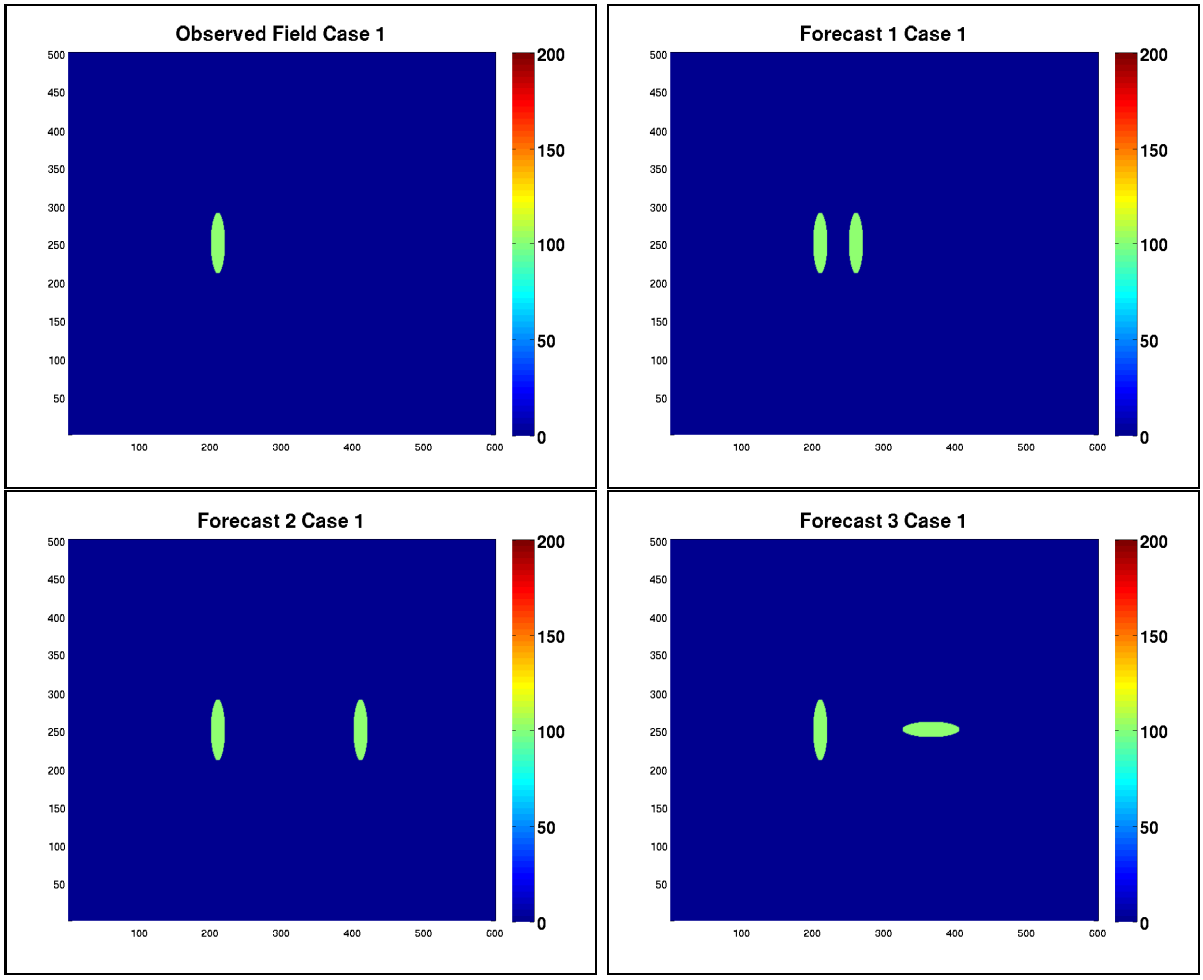

Figure 4.1: Idealized Case 1. Ellipsoids on the left are the observed field, while ellipsoids on the right are forecasts. This is consistent throughout all idealized case study images. All non-zero values are 100 with differing displacements.

Table 4.1: Partial Hausdorff (left), Kantorovich Distances (middle) and bias correction coefficient (right) values yielded from idealized Case 1.

Distance Values	HD_p	KD	BCC
Forecast 1	47	$6.19x10^6$	1
Forecast 2	197	$2.47x10^7$	1
Forecast 3	162	$1.93x10^7$	1

tively. This illustrates, as shown in the NCAR ICP [Ahijevych et al., 2009], that the metrics used cannot decipher between the forecasts since they do not have areas with overlapping amplitudes.

However, the observed and forecast fields shown in Case 1 above are of equal amplitude, therefore somewhat redundant results are generated, especially when using a metric that calculates amplitude-based errors (KD). For this reason, a second and third case study are conducted to illustrate how different amplitudes can change the values of KD and HD_p.

Table 4.2: RMSE, MAE, ETS and r^2 values for idealized case 1. All three forecasts generated identical values and therefore did not decipher spatial displacements in the fields.

Error Values	RMSE	MAE	ETS	r^2
Forecast 1, 2 and 3	141.42	200	-0.002	0

4.1.2 Idealized Case 2

Since the forecast fields and observed fields in Case 1 were all the same magnitude, they did not showcase the ability of the KD metric to generate values that can show amplitude errors as well as displacement errors. Therefore, a similar case study is repeated, only this time the forecast fields will have a higher magnitude than the observed field. The hopes of this study is to show that the KD and HD_p will still decipher between the

displacement of the fields in a representative manner, and allow for amplitude errors to determine the best forecast.

In Case 2 each of the forecast fields were multiplied by 2 (observed field values would be 100, forecast field values would be 200). This way, too, we can compare Case 1 to Case 2, since the observed fields are identical, with amplitude differences in the forecast fields. It would be assumed that since the amplitudes in Case 2 are more over-predicted, that the KD values for these fields would be greater than those in Case 1 (Table 4.1), since higher amplitude errors should generate in higher KD values. Furthermore, since the HD_p generates a binary field in association with the threshold values (Th $\geq$ 100), the HD_p value is anticipated to be the same, since the spatial displacements of the fields have not changed. This roots back to the shortcoming of the HD_p: the values for Case 1 and Case 2 will be identical if the threshold applied allows the entirety of the fields to be used even though the fields in Case 1 should be considered better forecasts.

It is shown that when the amplitudes were doubled in Case 2, the KD values increased when compared to Case 1. In fact, the KD values themselves are doubled. As stated, it should be assumed that since the spatial displacement did not change within these forecast fields, but the amplitudes increased in magnitude, then the KD would actually show that the forecasts in Case 1 were better than Case 2 by yielding a higher KD value in Case 2. Further, the BCC values in Case 1 were 1, meaning the sum of the amplitude of the observed and forecast fields were identical. However, in Case 2 the BCC values are 0.5, meaning the observed field is actually half of the forecast fields; the forecast fields are over-predicted. Therefore, the values of KD should deliver evidence that the Case 1 fields are more similar than Case 2.

Similar to Case 1, Case 2 generates identical values for all three forecasts using the traditional metrics.

Table 4.3: Partial Hausdorff (left), Kantorovich Distances (middle) and bias correction coefficient values (right) yielded from idealized Case 2.

Distance Values	HD_p	KD	BCC
Forecast 1	47	1.24×10^7	0.5
Forecast 2	197	4.95×10^7	0.5
Forecast 3	162	3.85×10^7	0.5

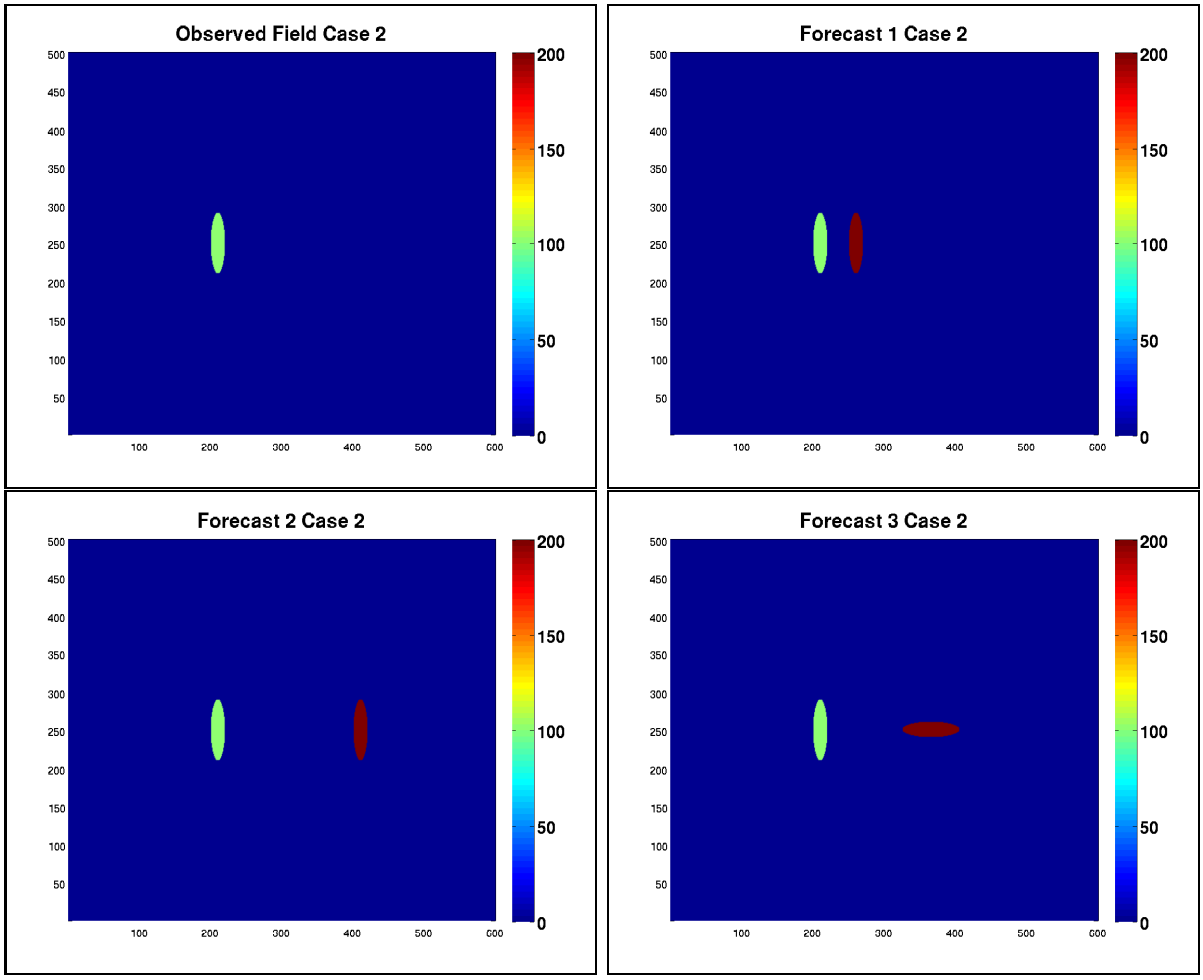

Figure 4.2: Idealized Case 2. All three fields are uniquely displaced, with amplitude error factor of 2. The observed field is 100 (green ellipsoids), Forecast 1, Forecast 2 and Forecast 3 are all 200 (red ellipsoids).

Table 4.4: RMSE, MAE, ETS and r^2 values for idealized case 2. All three forecasts generated identical values and therefore do not decipher spatial displacements in the fields.

Error Values	RMSE	MAE	ETS	r^2
Forecast 1, 2 and 3	223.61	300	-0.002	0

4.1.3 Idealized Case 3

One more idealized case study was conducted with fields of varying amplitudes with identical spatial displacements. This case uses only the observed field and the displacement illustrated in Forecast 1, but varies the amplitude of the field in Forecast 1. In Case 3, the following assumptions are made: F1 = (0.75*Forecast 1), F2 = (1.25*Forecast 1), F3 = (2*Forecast 1) seen in Figure 4.3. The premise behind this case is to show that the KD value of the three equally displaced fields will be greatly effected by their amplitude error when using the KD, but will remain constant with the HD_p; when the threshold applied allows for the entire field to be used.

Table 4.5: Partial Hausdorff (left), Kantorovich Distances (middle) and bias correction coefficient (right) values yielded from idealized Case 3.

Distance Values	HD_p	KD	BCC
F1	47	$8.24x10^6$	1.33
F2	47	$7.73x10^6$	0.8
F3	47	$1.24x10^7$	0.5

The results of Case 3 indicate that if three fields are all spatially displaced the same distance, but have differing amplitude errors, the KD will generate values that are indicative of this amplitude error. For instance, F2 actually has a smaller amplitude error than F1 and F3, and thus generates the lowest KD. Further, since the field was doubled in F3, the KD is higher than F1 and F2. These values are also consistent when analyzing

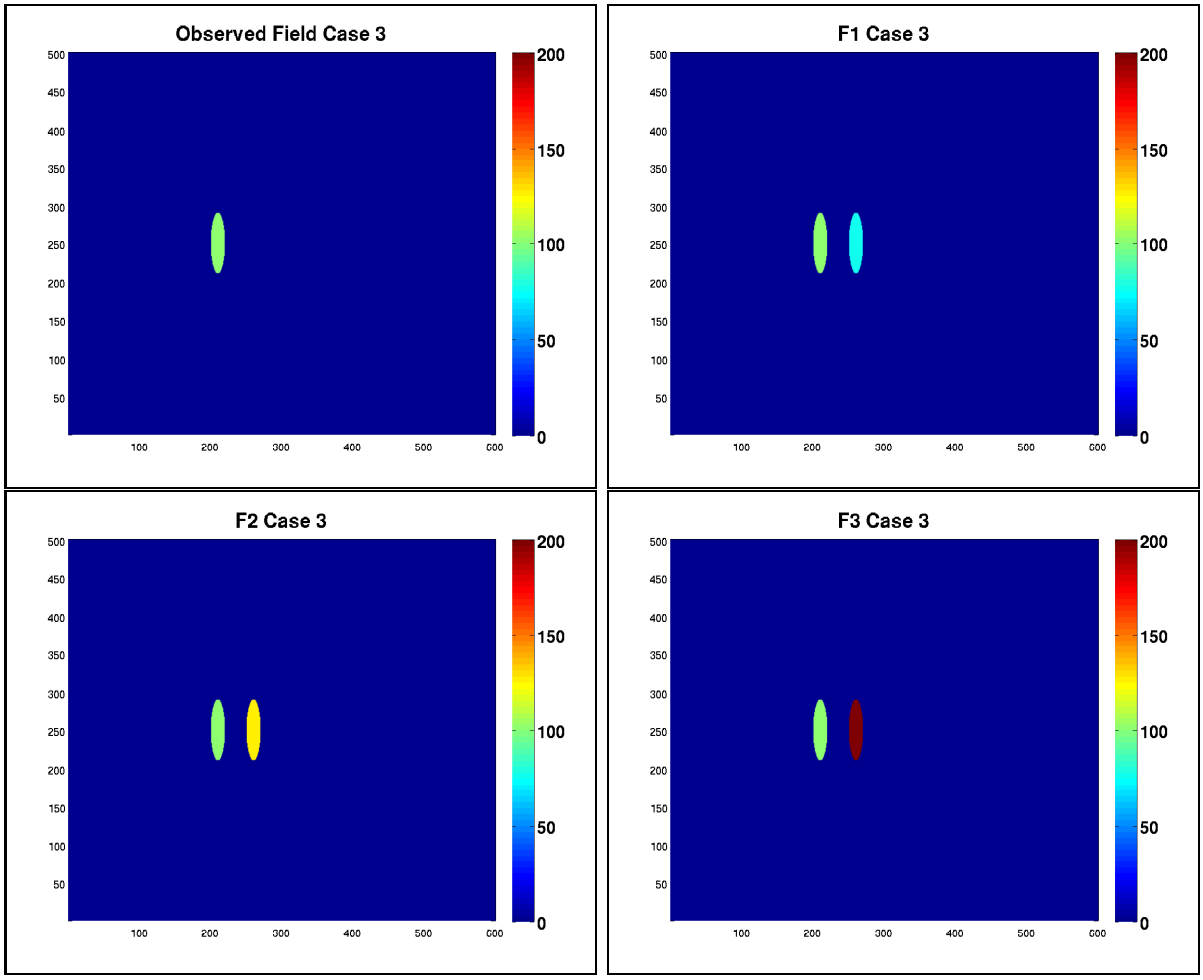

Figure 4.3: Idealized Case 3. All three fields are equally displaced, with amplitude error. The observed field is 100, F1 is 75 (top right), F2 is 125 (bottom left) and F3 is 200 (bottom right).

the BCC values for each forecast. The BCC for F2 is the closest to the value of 1, which would mean the most similar amplitude of the field. The BCC in F3 is the farthest from 1 out of the three, and thus has the greatest departure in amplitude. Therefore, both the KD and the BCC can be used in conjunction to analyze how similar the forecast fields are spatially and in magnitude when compared to the observed fields.

In Case 3, the traditional metrics deliver unique values since there is no spatial displacement difference but an amplitude difference. In this case, these metrics can be powerful, since they are able to differentiate between the three forecasts as the proposed metrics do.

Table 4.6: RMSE, MAE, ETS and r^2 values for idealized case 3. All three forecasts generated differing values since there were no spatial displacements in the fields, but only amplitude errors.

Error Values	RMSE	MAE	ETS	r^2
F1	125	175	-0.002	0
F2	160.08	225	-0.002	0
F3	223.61	300	-0.002	0

Results of the Idealized Cases:

In Case 1 it was shown that three spatially displaced forecast fields would generate different HD_p and KD values, with smaller values generated by less displacement. However, the amplitudes of the forecast and observed fields were all identical and therefore no bias correction was given to the three forecasts (BCC of 1). Further, the traditional metrics were unable to decipher the three forecasts as unique.

In Case 2 it was shown that when the forecast field is double the observed field, the HD_p values, based on threshold values used, generated the exact values as Case 1, showing the spatial displacement of the fields were identical. However, the KD values for the three forecasts were doubled due to a amplitude error of a factor of 2. The KD values in Case 2 when compared to Case 1 illustrate that more error was found in the forecast fields in Case 2, as expected. Further, the BCC values in Case 2 were indicative of a

greater amplitude error to the forecast field, and thus less-representative amplitude in the fields. In addition, the traditional metrics, as seen in Case 1, were unable to differentiate between the three forecasts.

Case 3 was used to show that the spatial displacement of all three fields were identical (exact same HD_p values), but with varying amplitudes on each field (both over and under-predicted). The KD values associated with each forecast field show that the forecast with the smallest magnitude of amplitude error has the smallest KD, while the field with the greatest amplitude error has the greatest KD. Results from all three idealized case studies were consistent with the hypothesized assumptions.

However, there were some questionable results with respect to orientation of the ellipsoids. It can be argued from a meteorological standpoint that Forecast 3 in Cases 1 and 2, is a less accurate forecast due to the orientation of the ellipsoid when compared to Forecast 2. The metrics are designed to calculate spatial displacements, and thus, since Forecast 3 is less displaced, it yields a lower value. For instance, if the ellipsoids were actually cold fronts, Forecast 3 would not actually appear to be more accurate than Forecast 2. This is one of the shortcomings of using the spatial displacement metrics on forecasts without including a parameter for the aspect ratio.

Although the traditional metrics were in this case able to generate unique values for each of the forecasts, the three cases shown illustrate how powerful the proposed metrics are compared to the traditional metrics. However, to show that these metrics can be applied to real-world cases, they will be used to evaluate the Buffalo Ridge Wind Farm case study.

Chapter 5

Issues With Observed Data

5.1 Data Availability

Likely the most accurate way to evaluate forecast verification of wind ramp events would be to analyze power production and economic data from the individual operators. This data, referred to as supervisory control and data acquisition or SCADA data, is wind speed and/or power data sent from the wind farm operators for scientific analysis. Unfortunately, these data are not always given for analysis. If analysis of projected wind power forecasts could be compared with actual power generation, a more accurate depiction of overall forecast quality could be generated. However, data pertaining to wind farm power production and energy pricing is proprietary, and is generally difficult for the scientific community to gain access to such data.

Although an economic analysis of forecast productivity may be considered the most desirable method of forecast verification, an analysis on the progress of the accuracy of the actual wind forecasts themselves could also shed light on if wind forecasts are improving. If wind speed forecasts are found to be improving with time, then the power production forecasts should be increasing in accuracy as well. However, high-resolution wind data are not readily available to the public, therefore it cannot always be used in wind forecast validation.

While base-state atmospheric variables are generally found to be relatively constant with only small fluctuations over time and space (temperature, pressure, etc.), horizontal wind velocity is an ever-changing phenomena and fluctuations can be an order of magnitude higher or lower in some cases. For this reason, weather forecasters usually have much

more accuracy in forecasting the short-term temperature profile than they do forecasting fluctuations in wind at the same location. Therefore, increased spatial coverage of wind data is needed in order to accurately predict wind intermittency Freedman et al. [2008]. However, a lack of spatially-abundant wind data has become an obstacle for short-term wind forecasters.

A major issue in wind ramp forecasting is attributed to the lack of accurate, reliable high-resolution wind data in close proximity to wind farms. Without proper wind datasets, the industry is forced to use less sufficient data for wind forecasts, which yields higher inaccuracies. Furthermore, high-resolution wind data at turbine hub-height is not available to wind forecasters. Sounding data are taken at various pressure levels in the atmosphere, and may not be useful at hub height depending on the size of the turbines. Therefore, vertical interpolation methods, such as implementing the log wind profile, are needed to estimate wind speeds at varying heights. These estimates yield wind values relatively close to the actual winds, but they still can be relatively inaccurate, since atmospheric stability plays a role in vertical interpolation of surface winds.

5.2 Real-Time Mesoscale Analysis Data

Although high-resolution wind datasets are generally sparse, the National Centers for Environmental Prediction (NCEP) in collaboration with the Earth System Research Laboratory (ESRL) generates a dataset that is available to the general public in the form of Real-Time Mesoscale Analysis (RTMA) data. RTMA data are 2.5km horizontal resolution with 1hr temporal resolution reanalysis data that span North America. The variables generated by RTMA are 2m temperature, 2m specific humidity, 2m dew points, surface pressure and 10m U and V wind components. Without looking into the source of this data, one may feel that RTMA could be the answer for the short-term wind forecasting community's lack of quality wind data. However, the way in which RTMA data are generated illustrates how deficient readily available wind data are in accuracy De Pondeca et al. [2011].

The rapid update cycle (RUC) model is used as the first guess for the RTMA data. However, the resolution of RUC is 13km and therefore needs to be downscaled onto the National Digital Forecast Database (NDFD) grid of 2.5km. Further, mesonet station data account for 84.5% of the RTMA input data, and of these 14,000 wind mesonet stations,

over 80% of them fail the quality control criteria Pondeca et al. [2007]. Therefore, the bulk of the input data are rejected before they can be used. Furthermore, according to DePondca et al. Pondeca et al. [2007], the wind components of the data from these sources are found to have the greatest inaccuracies of all variables. Therefore, using this dataset for short-term wind forecasting may be problematic. Therefore, a different approach may be implemented as an alternative.

5.3 Meteorological Covariates

Abundant radar reflectivity data are available to the general public. In addition, the resolution of radar reflectivity data are increasing both spatially and temporally as enhanced technology is used. One such radar network is used operationally by the National Weather Service (NWS), called the next-generation radar (NEXRAD) network.

The NEXRAD network is comprised of 159 S-band radars that span the United States. The radars have the capability to scan between 5 and 14 vertical levels, which corresponds to clear-air and precipitation modes, respectively. Each scan can take between 4.5 and 10 minutes, again, depending on scan mode. These radars have a horizontal resolution that has recently been increased from 1km to 250m at close ranges. This increased resolution allows for much greater detail in datasets produced by the network. Additionally, the enhanced range of each of the NEXRAD radars, from 230km to 300km, has allowed for greater scanning ability. The question then is raised: can radar reflectivity fields be used to estimate the occurrence of near-surface wind ramp events?

If near-surface wind assumptions can be made by using radar reflectivity data, then better wind ramp forecast validations may be in the near future. How would this bridging of wind and reflectivity data be used? A proxy, or meteorological covariate, for wind ramp events can be made when radar reflectivity values exceed certain thresholds. Brown and Murphy introduced the idea of meteorological covariates by using a highly observable variable in exchange for a lesser observable variable, if they are both measured on similar spatial and temporal scales Brown and Murphy [1996].

For this study, when reflectivity values are high it will be assumed that the near-surface wind increment in close proximity will be high as well. This notion can be seen in Figure 5.1, where the high reflectivity values are collocated with strong near-surface winds. Further, if sharp gradients in radar reflectivity are displayed (approaching squall

line, cold front, etc.), then the likelihood of a wind ramp event occurring at or near the leading edge of the gradient is extremely probable as well. Therefore, a proposed radar reflectivity/ near-surface wind proxy is being introduced to show that using reflectivity fields to estimate the location and timing of wind ramp events is useful when wind data are not available for forecast verification purposes.

In addition, a study conducted by Mackeen, et al., displayed evidence about longevity of storms based on reflectivity maximum MacKeen et al. [1999]. In this study, they found that storms producing radar reflectivity values exceeding 55dBZ have a much greater chance of surviving for longer periods of time, as compared to storms producing lower values. With this being said, if strong reflectivity values are seen upstream, chances of a wind ramp event taking place downstream are greater.

Using upstream data to predict wind ramp events is a method used in the wind forecasting industry. For instance, a study conducted at GL-Garrad Hassan indicated that the use of upstream wind data can be beneficial for downstream wind ramp forecasts. Additionally, the concept of using upstream data for modeling is the backbone of FDDA, which nudges the model forecast with ingested observational data. This technique can be a beneficial attribute in short-term wind ramp forecasts.

Using this near-surface/ radar reflectivity proxy, the spatial displacements of wind ramp events can be evaluated quantitatively by using the radar reflectivity field. To test this notion, a case study in which a ramp event took place will be evaluated.

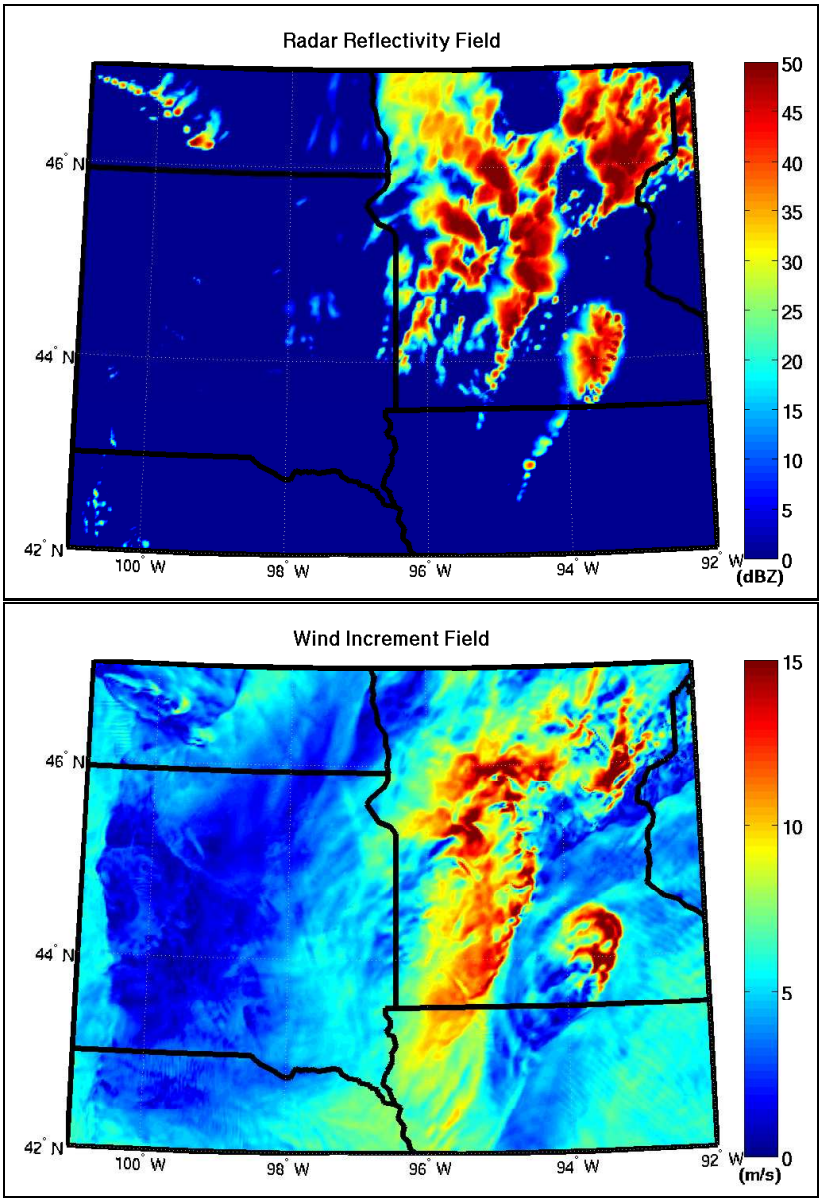

Figure 5.1: Using high values associated with radar reflectivity fields (top) as proxy for wind ramp events occurring at the near-surface (bottom). As shown in the plots created from WRF simulated fields, large values in wind increment at 10m over 60 minutes are co-located with high reflectivity values.

Chapter 6

Real Case Study

6.1 Case Study Introduction

On 2011 July 1, on the border of Minnesota and South Dakota, a wind ramp event occurred at a wind farm due to a severe weather event. The associated wind farm is referred to as Buffalo Ridge Wind Farm (BRWF). This wind ramp event makes for an intriguing study because it was also an extreme wind event. The severe thunderstorms associated with this event produced tornadoes, a downburst and many severe wind reports within the region, which in turn damaged several of the farm's wind turbines. And since high radar reflectivity values were generated during this event, the introduced proxy can be used for evaluation.

The National Weather Service (NWS) Storm Prediction Center (SPC) did, however, issue a slight risk for severe weather in the study area on the afternoon of July 1. Additionally, the SPC issued a significant wind event warning in a relatively large area of the region, including the area where the BRWF is located. This forecast was based on the synoptic conditions that were evolving over the region. However, the risk for high winds issued by the SPC covered a 16 hour window, and would not be extremely beneficial to the short-term wind forecasting community for this region.

6.1.1 Synoptic Conditions

The 2011 July 1 0UTC upper air charts provided by the NWS in Aberdeen, SD depicted a strong temperature inversion had developed in a layer from 700-850hPa. This capping

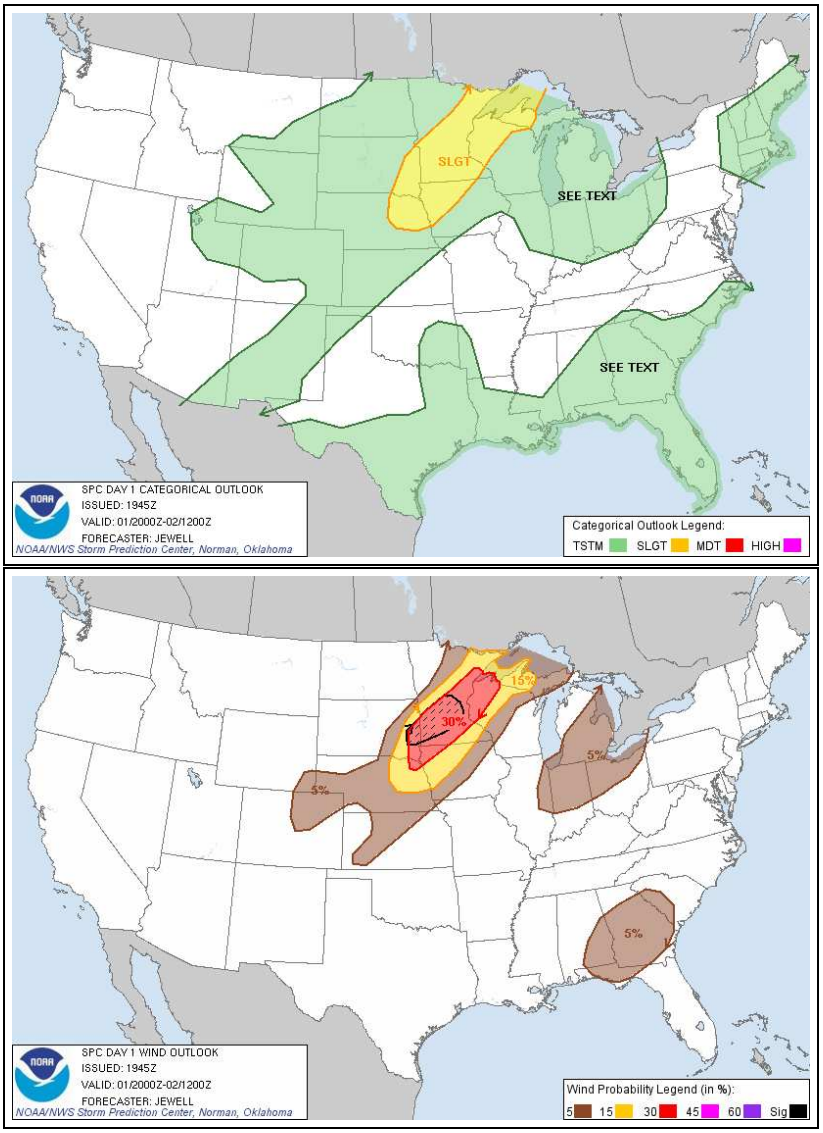

Figure 6.1: NWS Storm Prediction Center's 20UTC severe weather outlook (top) and severe wind prediction (bottom) maps. The dashed area where severe wind probabilities are extremely high is collocated with the Buffalo Ridge Wind Farm.

inversion would allow the boundary layer to become extremely unstable by enabling the surface moisture to heat during the day without mixing to the free atmosphere above.

With afternoon surface temperatures near 35°C and dew points nearing 25°C, mixed layer CAPE values exceeded 5000J/kg. A very prevalent shortwave trough entered the region in the afternoon, accompanied by a jet max of 25ms^{-1} at 500hPa and strong upper-level divergence at 300hPa. The winds created a veering wind profile and vertically sheared environment favorable for strong updraft rotation with 0-6km wind shear values of 60ms^{-1} and 0-3km storm-relative helicity values of 400m^{2}s^{-2}.

By late afternoon, a surface cold front entered the region and triggered a line of thunderstorms to the southwest of BRWF, in central South Dakota. As this line of storms propagated to the east-northeast and moved into an area of strong instability, storms began to intensify rather quickly and embedded supercells began to form. At the time this line of thunderstorms arrived at the location of BRWF, a downburst and a tornado had been reported in close proximity to the wind farm. The SPC reported wind gusts in close proximity to BRWF of 29-32ms^{-1} during this event. Uprooted trees, downed power lines and structural damage was also reported in the area around BRWF.

6.2 Mesoscale Numerical Weather Prediction Model

A series of simulations were conducted using the Weather Research and Forecasting (WRF) model to simulate the environment described in the previous section. The WRF model is a mesoscale NWP model used widely for weather forecasting and atmospheric research. It is equipped with many parameterization schemes, including but not limited to planetary boundary layer, micro-physics, longwave and shortwave radiation and cumulus schemes. Further, the WRF model is able to generate atmospheric simulations on the synoptic-scale, mesoscale and down to the microscale.

In these simulations, the aforementioned BRWF case study will be the foci of this evaluation. In addition, a series of varying model configurations will be used to simulate different outcomes that can be evaluated against the observed field. The parameterizations that differ for this study will be the planetary boundary layer (PBL) and surface layer schemes. These schemes are used to parameterize the turbulent fluxes of momentum, heat and other atmospheric variables found within this region of the atmosphere [Hu et al., 2010]. PBL schemes are needed since turbulent energy is found on the mi-

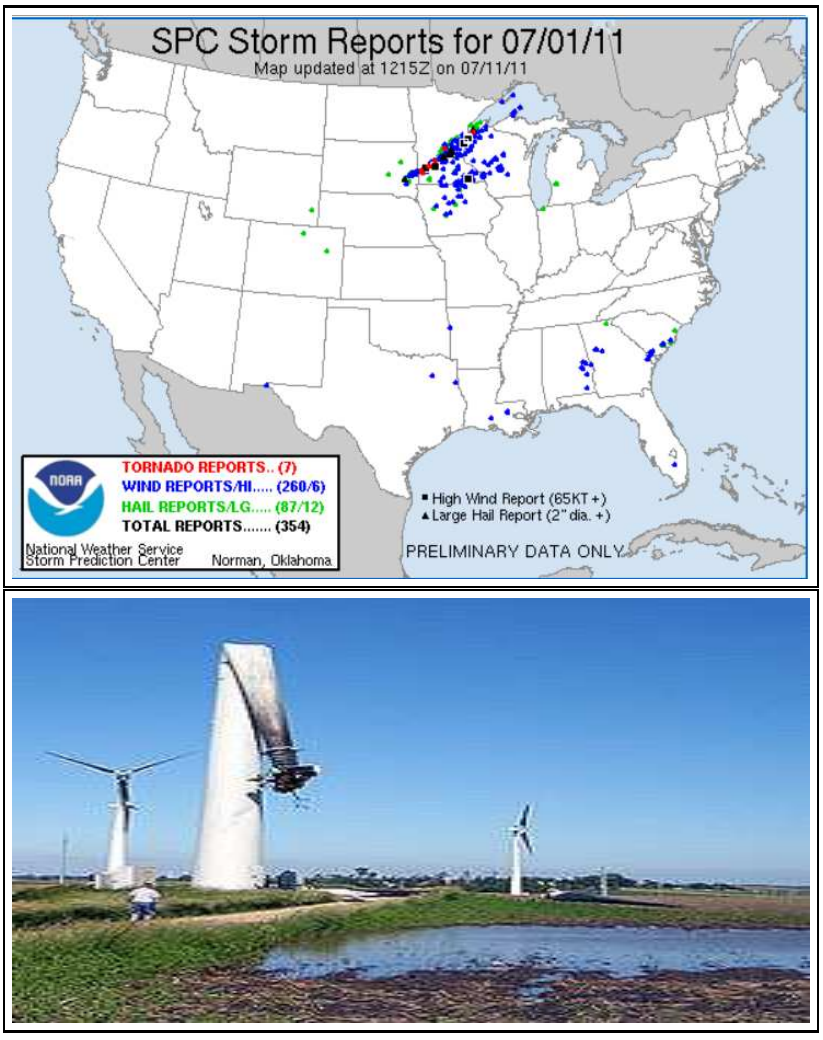

Figure 6.2: Storm reports from the National Weather Service Storm Prediction Center for July 1, 2011 (top). Two tornadoes were reported in close proximity to the Buffalo Ridge Wind Farm, along with a report of a downburst that damaged several of the wind farm's turbines (bottom). Bottom image courtesy of Lincoln County Emergency Management Office.

croscale, and needs to be parameterized since the resolution of the model is too coarse to simulate down to this level.

The three PBL schemes used in the corresponding model configurations are the Mellor Yamada Janjic (MYJ), the asymmetric convective model version 2 (ACM2) and Yonsei University (YSU); conf1, conf2 and conf3, respectively. Again, this thesis will not be addressing the model performance, but instead will show that the newly introduced metrics for forecast verification will yield a quantitative value of spatial displacement and amplitude error between varying forecasts.

In this case study, the WRF model version 3.3 was used to simulate the thunderstorm environment that generated a wind ramp event at BRWF on July 1, 2011. For these three simulations, initial conditions were used in the form of the North American Mesoscale (NAM) model, beginning 2011 July 1 at 12Z and running 18 hours. The domain used was 750km x 750km, with 2km horizontal resolution along with 51 vertical levels.

6.3 Buffalo Ridge Case Study

For the BRWF case study, both RTMA wind data and NEXRAD radar reflectivity data will be used. The RTMA data will merely be used to show that the wind fields in the forecasts can be compared using these metrics, and the NEXRAD data will be used to show the power of using a meteorological covariate when evaluating the performance of a numerical weather prediction model.

6.3.1 Real-Time Mesoscale Analysis Data

Wind ramp events tend to be defined in terms of wind power, as opposed to wind speeds, therefore the wind speed data needed to be converted to wind power based on a power curve. For this study, a wind turbine power curve was chosen arbitrarily; the Vestas 1.5MW. An algorithm was created to convert the RTMA and WRF wind speed data to wind power, based on this power curve. Further, since a change in wind power is what constitutes a power ramp event, the temporal difference between power values at each point is needed. For the 18UTC (22UTC) analysis, the change in wind power was calculated by taking the difference between 18UTC and 19UTC (22UTC and 23UTC).

When analyzing the results of the HD_p and KD for the wind data, only ramp-up events are considered. It is possible to generate a field that calculates the absolute value

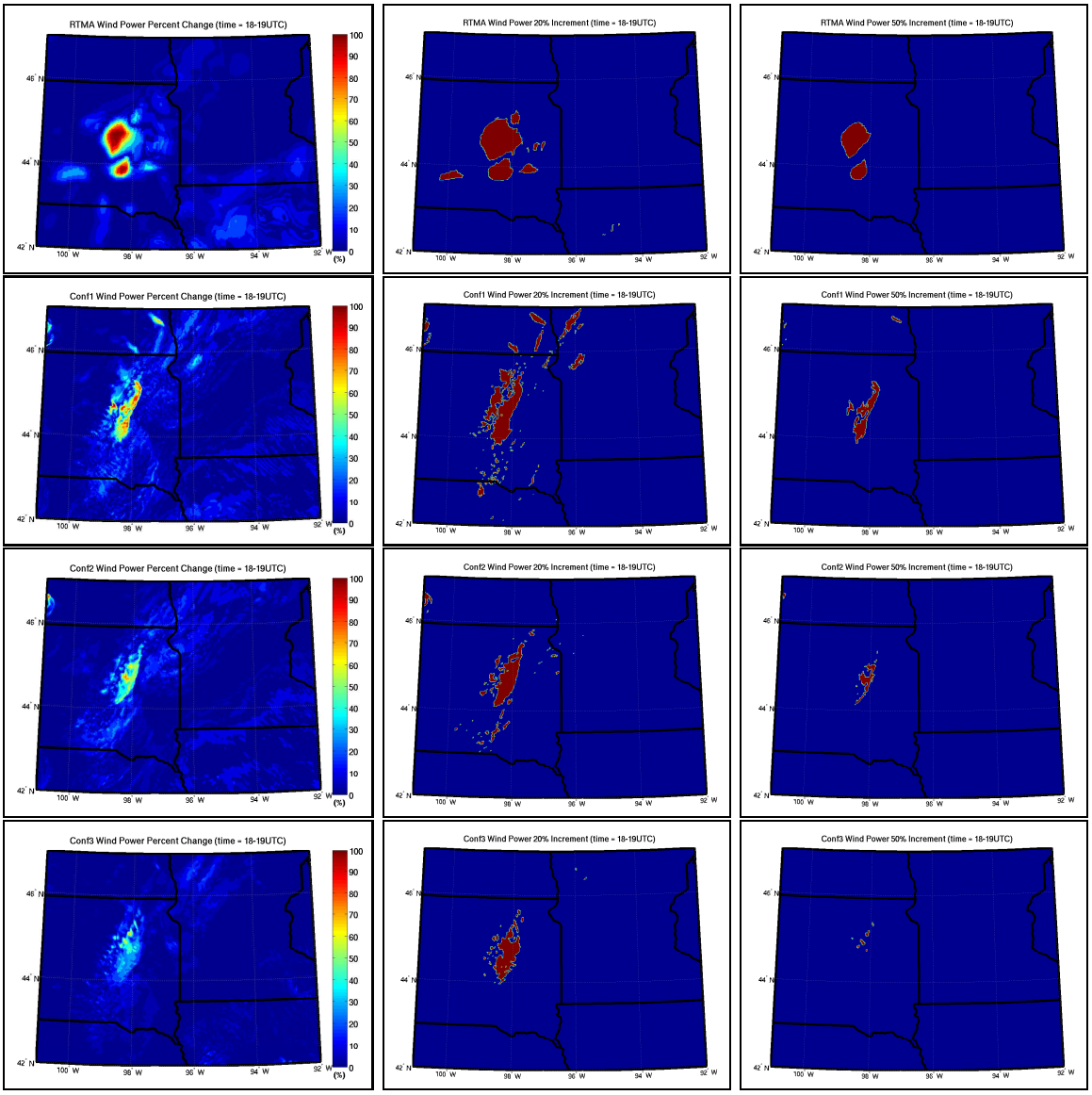

Figure 6.3: 10m wind speed plots using observed RTMA data (top row) and conf1, conf2 and conf3 (left), along with 20% wind power increment (center) and 50% wind power increment (right) thresholds at 18UTC. Threshold fields will be used to calculate the Partial Hausdorff Distance.

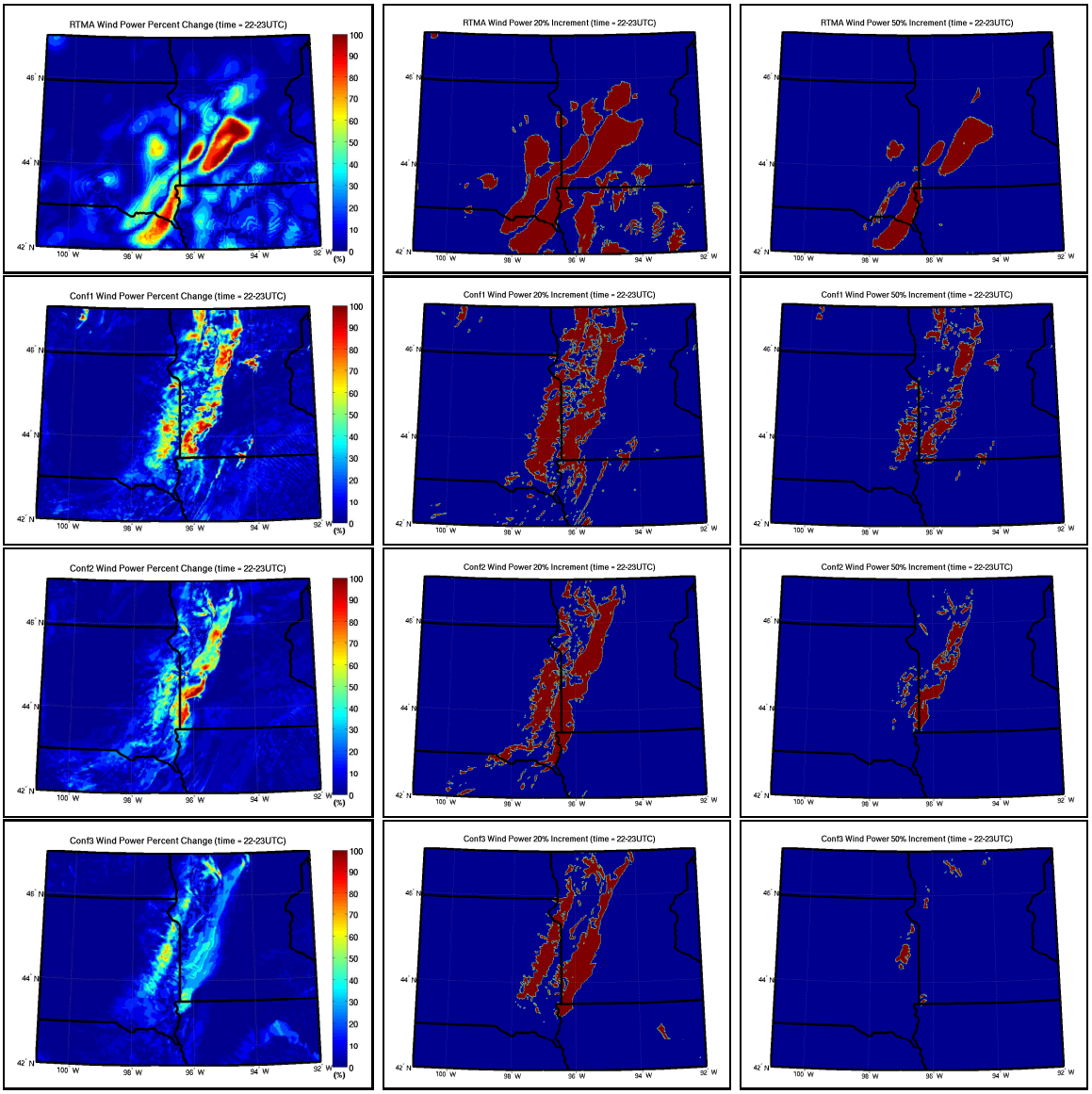

Figure 6.4: Similar to 6.3 for 22UTC.

of the wind increment values at each point, and consider both ramp-up and ramp-down events (as seen in the climatology study in Chapter 2). However, since the notion of meteorological covariates is based on the idea that high reflectivity values are thought to increase wind speeds at the near-surface, only positive ramp events (ramp-up) will be used, since the premise of this study is based on the notion that high reflectivity radar fields can be used as a proxy for wind ramp events. Further, in this case only 1 hour wind power increment is evaluated. Using a 20% and 50% threshold for wind power increment over the temporal span of 1 hour does fall within wind ramp definitions commonly used in the industry. However, a 50% increase in wind power in 1 hour could likely be constituted as an extreme wind event as well. Nonetheless, this case study did in fact involve an extreme wind event. Finally, to use the WRF simulations for comparison with the observed fields, the observed data were interpolated onto the WRF grid for spatial continuity. Without this step, the comparison would not generate accurate values since the images being evaluated would be at different resolutions. Resolution size of the images also impacted the HD_p value given for any given comparison. Higher horizontal resolution generates more pixels within the image. Therefore, in images with high resolution, and increased pixels, the displacement values of the HD_p will be larger than a more coarse resolution even though the accuracy might be increased. For instance, a 2km resolution image from a simulation might be more similar to the control image, but could result in a higher HD_p value than a 10km resolution simulation. However, if all of the compared images are of the same resolution, then the comparison with the smaller value is the most comparable image to the control image.

Partial Hausdorff Distance:

When looking at Figure 6.3, conf2 is shown to have values of 20% wind power percent change (wind increment) that look spatially similar to the RTMA dataset at 18UTC. Conf1 has values off to the north that are not seen in the observed field, while conf3 has a relatively small binary field. Further, when evaluating the 50% wind power change plots, conf1 can be seen to have a spatially consistent field with the observed field, while conf2 and conf3 have a much smaller overall field. This qualitative analysis is consistent with Table 6.1. Similar attributes can be seen when evaluating Figure 6.4 and 6.2.

In Tables 6.1 and 6.2, conf1 yields the lowest values (best representation) of HD_p for all runs, except when a threshold of 20% wind power increment is applied during 18UTC.

Interestingly, conf2 generates the lowest value for 18UTC with 20% threshold, but then generates the largest HD_p value for the same time with a threshold of 50% wind power increment. One could conclude from this that conf2 simulated the storm structure fairly well, but did not produce the higher values of reflectivity as well as the other simulations. Further, the frequency histogram of conf2 shows a lower frequency of values at 22UTC.

Table 6.1: Table of Partial Hausdorff Distance values for wind power increment data at 18UTC.

(18-19UTC)	HD_p @ Th=20%	HD_p @ Th=50%
Conf1	158	166
Conf2	144	358
Conf3	175	187

Table 6.2: Table of Partial Hausdorff Distance values for wind power increment data at 22UTC.

(22-23UTC)	HD_p @ Th=20%	HD_p @ Th=50%
Conf1	43	53
Conf2	49	55
Conf3	70	83

Kantorovich Distance:

Similar tables seen in the previous section are generated using the KD. However, there did not appear to be a clear best performing configuration when using the KD. One could argue that conf1 and conf3 performed roughly the same, while conf2 was again, generating relatively high values when high thresholds were applied at 18UTC. As seen in the previous section, conf2 generated a relatively large value at 18UTC (as in Table 6.1) with the high threshold which was reflected in the mean values. However, using the

KD, this anomaly is no longer present. In fact, for 18-19UTC, conf2 appears to perform better than conf3.

However, the KD yielded some conflicting values when compared to the values generated using the HD_p and the frequency histograms. For instance, in the HD_p analysis, conf1 was shown to generate the lowest overall values for the thresholds and times evaluated for the wind power portion and conf2 was likely the second best. However, the KD created evidence that conf1 and conf3 were actually both creating the lowest overall and conf2 was the least accurate of the three configurations. Furthermore, as stated in the previous chapter, the BCC can be thought of as a tie breaker in the instance that the KD values are similar, but the BCC are largely different. When evaluating the KD values for 18-19UTC for both 20% and 50% increment of wind power, all three configurations appear to perform similarly. However, the BCC values tell a different story. Conf3 wildly under-predicted the values of the RTMA wind field and thus has a much larger BCC than conf1 and conf2. Therefore, one can conclude that conf3, since it needed such a large bias correction, would be the least accurate of the three forecasts.

Table 6.3: Table of Kantorovich distance and bias correction coefficient values for wind power increment data at 18UTC.

(18-19UTC)	KD @ Th=20%	BCC (Th=20%)	KD @ Th=50%	BCC (Th=50%)
Conf1	7.22×10^6	3.86	4.13×10^6	2.29
Conf2	6.41×10^6	6.84	4.19×10^6	17.13
Conf3	6.90×10^6	36.98	3.97×10^6	33.33

The BCC values, however, appear to be consistent with the frequency histograms for the wind speed distributions. For instance, when looking at the histograms for the 18UTC wind speed distributions, conf3 appeared to under-predict the wind speeds when compared to the RTMA data. This is reflected in the BCC for conf3, which exceeds 33 for both thresholds at 18-19UTC: As stated earlier, larger BCC values are indicative of under-predicted forecasts since BCC = Sum of Observed Field / Sum of Forecast Field. However, the distribution of conf1 and conf2 appear to over-predict the wind fields, and thus should generate a smaller BCC. This can be seen in both 20% and 50% thresholds in Table 6.3.

Finally, using the KD, BCC and HD_p in conjunction with one another, it can be seen that since all three configurations generated similar KD values, but the BCC value was closest to 1 and HD_p value was lowest for conf1, that conf1 generated the best spatial and amplitude forecast when compared to the RTMA data.

Table 6.4: Table of Kantorovich distance and bias correction coefficient values for wind power increment data at 22UTC.

(22-23UTC)	KD @ Th=20%	BCC (Th=20%)	KD @ Th=50%	BCC (Th=50%)
Conf1	$9.52x10^6$	1.92	$1.79x10^7$	2.64
Conf2	$1.71x10^7$	3.94	$1.41x10^7$	12.32
Conf3	$1.98x10^7$	4.15	$1.59x10^7$	10.73

Relatively similar results are found in Table 6.4 for 22-23UTC. At this time, conf1 was shown to generate the lowest overall BCC values for the 20% and 50% thresholds. Also, conf1 generated the lowest HD_p value for both threshold values. Using the KD, BCC and HD_p values together can help to show the spatial displacement error along with the amplitude based error. In this case, conf1 can be seen to be the best forecast for 18-19UTC.

6.3.2 Meteorological Covariates

The first item that needs to be addressed about the reflectivity fields is that the NEXRAD radars do not give a value when there is no power returned to the radar. If there is nothing to bounce the radar feed back to the sensor, the radar will merely not have a value of reflectivity at that location for that time. However, WRF derives reflectivity differently, and assigns a minimum value of -30dBZ to a point in the domain that does not have a reflectivity return. Therefore, since a large majority of the domain does not have a return during the simulation, there was an over abundance of -30 values. As a result, there was an extreme amplitude error to the left of the distribution in the three simulations.

Secondly, and maybe most importantly, in this study we are interested in representing thunderstorm environments that will induce strong surface winds and wind increments. Therefore, the lower end of the distribution is essentially useless for examination, espe-

cially since it is filled with -30 values in the datasets. Since we are mainly interested in the thunderstorm or higher values of reflectivity to begin with, the decision was made to only evaluate the reflectivity values that are equal to or exceeded 30dBZ. This option would eliminate the amplitude error, and allow for the evaluation of only those higher reflectivity values.

When evaluating at the distributions of the 18UTC Figure 6.5, there appears to be much more values between 30 and 35dBZ in the NEXRAD dataset when compared to the model datasets. This feature creates a lower mean for the NEXRAD data, and a slightly higher mean value for the model datasets. However, it can be seen that the models tend to have maximum reflectivity values around 50 to 55dBZ, while the NEXRAD data shows the maximum values well into the 60dBZ range, albeit not very many.

Similarly, when looking at the 22UTC Figure 6.6 the distribution of values between 30 and 35dBZ in the NEXRAD dataset is much larger than the model datasets. However, at 22UTC NEXRAD data accounts for much larger values of reflectivity on the tail of the distribution, which may help to shift the mean back towards 40dBZ. A more substantial amount of 50+ dBZ values are found in the observed field as compared to the model fields, which do not exceed 60dBZ in any simulation. Regardless, the mean in conf1 does seem to be most consistent with the NEXRAD mean for reflectivity at 22UTC.

Partial Hausdorff Distance:

Observed NEXRAD radar reflectivity data were used to validate the model simulations. Preliminary images were generated to give a qualitative perspective on how well the reflectivity field was simulated using the WRF model. Before using the proposed and traditional metrics to evaluate the spatial displacement of the simulated and observed fields during this ramp event, data analysis in the form of time-series plots of wind and radar reflectivity fields were generated. Observed data used in the time-series plot was taken at Brookings Regional Airport (BRA) in South Dakota, about 30km due west of the wind farm. Twenty-minute observed data was generated using an Automated Weather Observing System (AWOS) at the BRA, for the same time as the NAM data (12UTC on July 1 to 06UTC on July 2, 2011). The wind speed data from the three simulations was extracted at roughly the latitude and longitude of the BRA.

As seen in the wind time-series plot, the model did manage to simulate a possible wind ramp event: however, with temporal displacement of about 2-3 hours. Further,

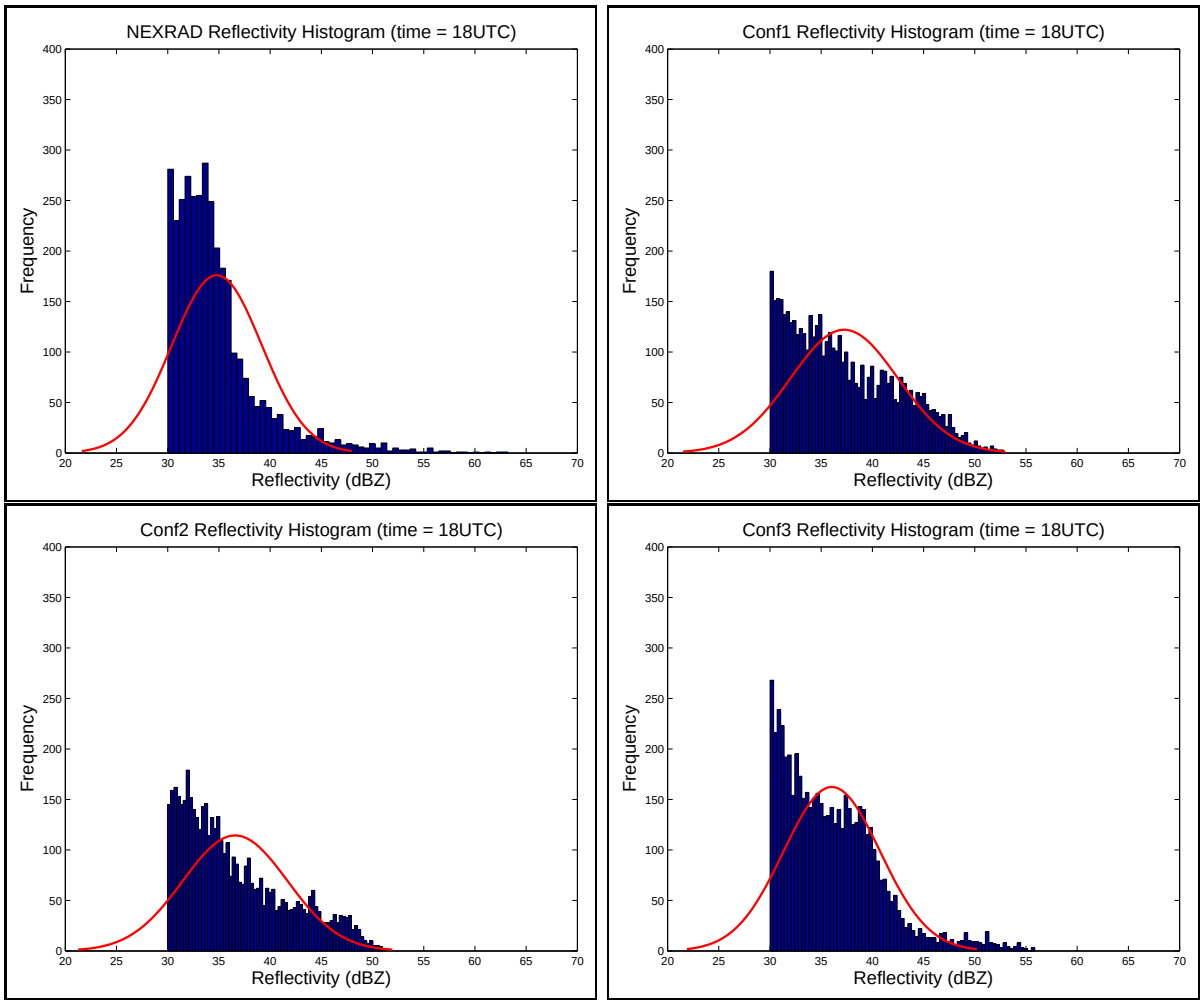

Figure 6.5: Radar reflectivity frequency histograms of NEXRAD data and model data for 18UTC.

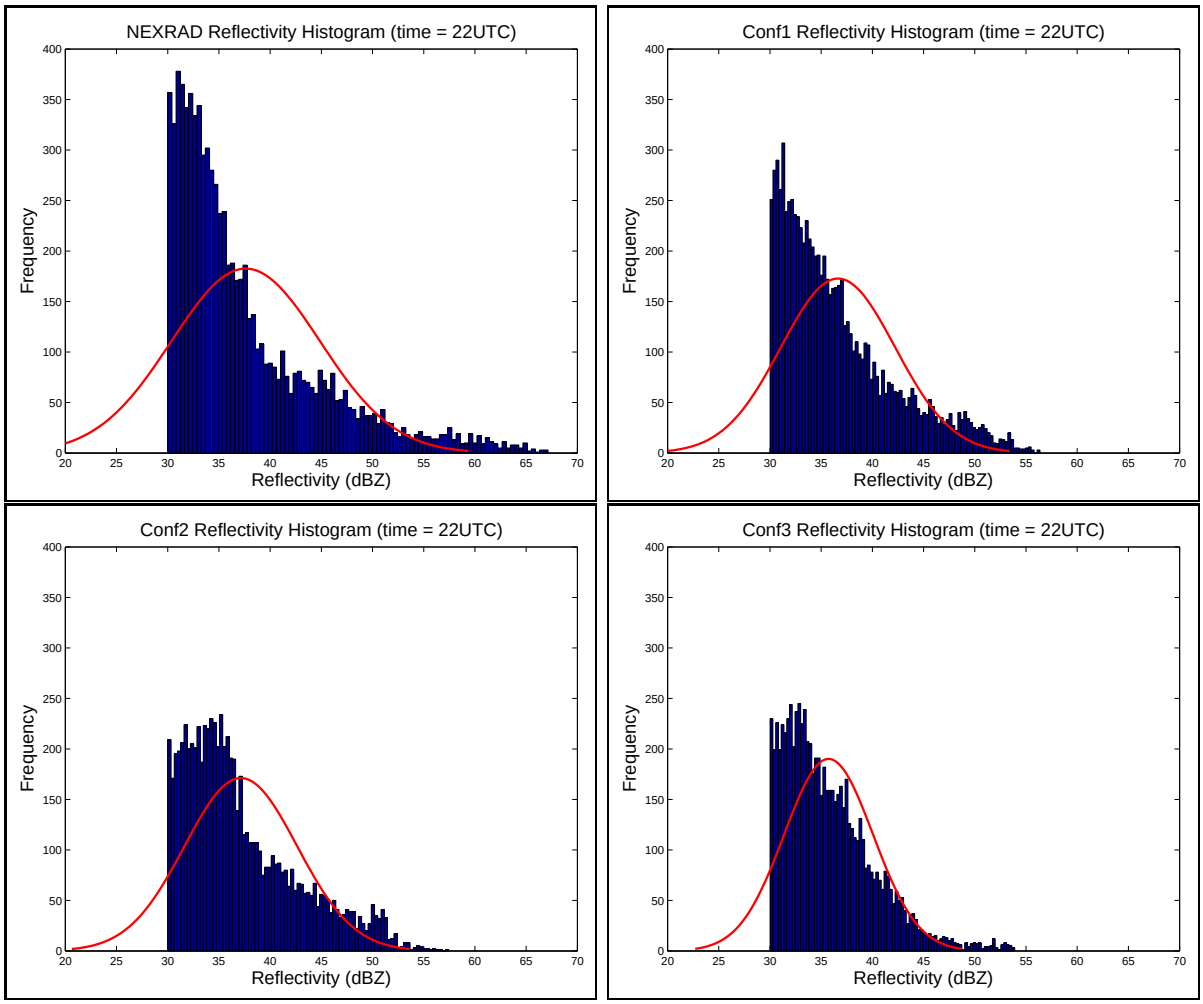

Figure 6.6: Radar reflectivity frequency histograms of NEXRAD data and model data for 22UTC.

the magnitude of the wind speeds for the given times was under-representative of the observed wind data by almost 5m/s.

The results from the reflectivity time-series plot are rather consistent with the results from the wind time-series, with respect to the lag of 2-3 hours and an under-represented amplitude. Further, there is a clear spike in reflectivity values in the NEXRAD data that is seen around 17-20UTC. This appears to be the time that the wind ramp event took place at the BRA. Therefore, it can be addressed that the wind ramp event was found to be initiated as the reflectivity values in the field began to rapidly increase.

A correlation of the meteorological covariates indicated a fairly strong relationship between the maximum values of radar reflectivity within the field and the associated maximum values of 60-minute wind increment within the field throughout the entire simulation time, as seen in Figure 6.8. All three simulations appeared to indicate that there was a general, positive relationship between the maximum reflectivity values and the maximum wind increment values within the domain. As the radar reflectivity values increased, the wind increment values increased as well, providing added validity to the use of meteorological covariates.

When evaluating the reflectivity fields there was some spatiotemporal discontinuity, however there were some similarities in storm structure between the simulations and the observed data. The simulations appeared to exacerbate the spatial extent of the reflectivity cores, which would have great impact on the HD_p values yielded.

Table 6.5: Table of Partial Hausdorff Distance values for radar reflectivity data at 18UTC.

(18UTC)	HD_p @ Th=30dBZ	HD_p @ Th=50dBZ
Conf1	204	202
Conf2	234	370
Conf3	210	197

For this evaluation, two reflectivity thresholds were used: 30 and 50dBZ. The value of 30dBZ may show overall storm structure, while 50dBZ may show convective cell structure and size. Evaluation using the HD_p of 18UTC with a threshold of 30dBZ generated results indicating that conf1 generated the most representative reflectivity field according to

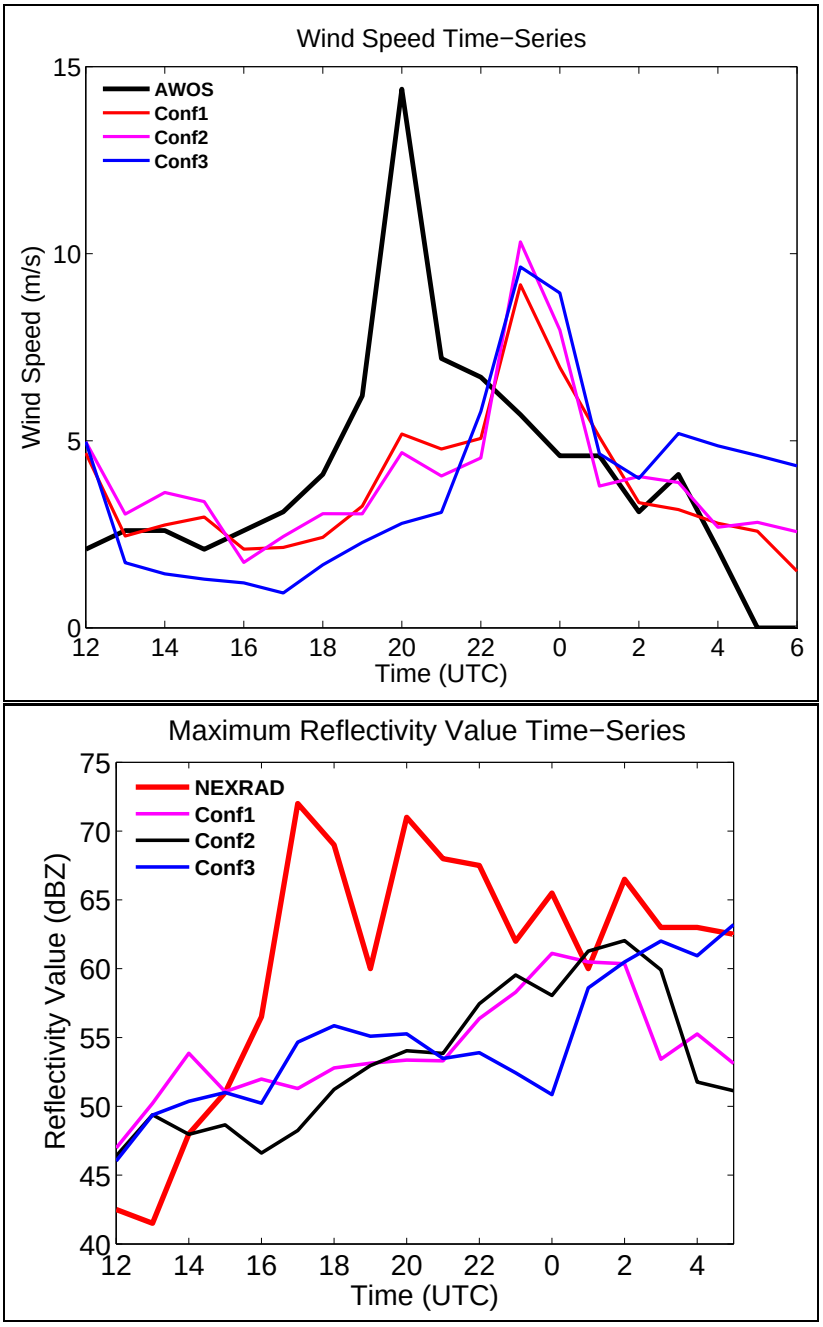

Figure 6.7: Time-series plot of wind speed at Brookings Regional Airport (top), along with the wind speed data from the three WRF simulations. The top plot depicts both an amplitude under-prediction and a temporal displacement of roughly three hours. Additionally, the WRF simulations under-predicted the maximum reflectivity fields in all three runs, as compared to NEXRAD observed data (bottom).

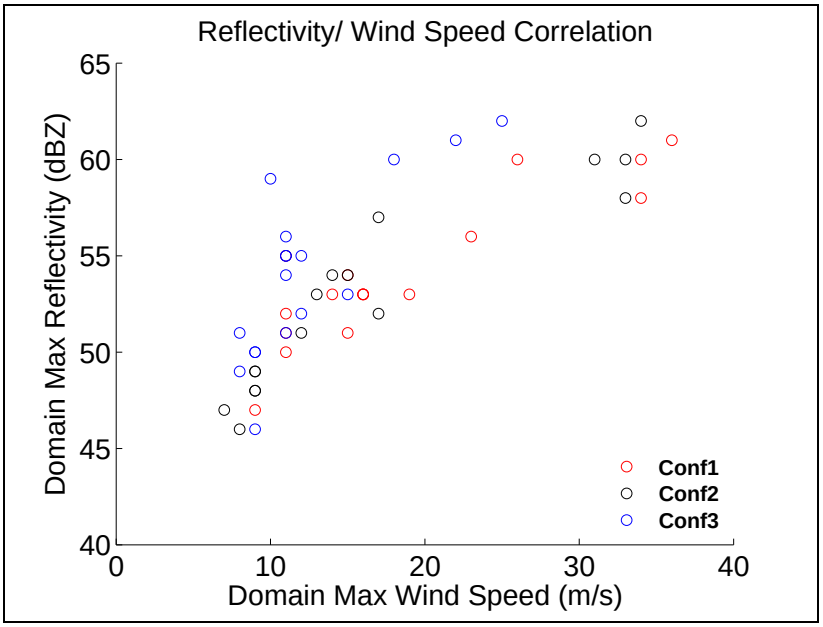

Figure 6.8: Scatter plot illustrating relationship of domain maximum reflectivity values and wind velocity values. As the maximum values in the reflectivity field increases, the maximum wind increment values increased as well.

Figure 6.9. However, when applying a threshold of 50dBZ to the reflectivity field conf3 generated the lowest value of the HD_p, while the HD_P for conf1 and conf2 for the threshold of 50dBZ yielded somewhat higher values. This would indicate that conf2 did a relatively poor job simulating high values of reflectivity over the domain at 18UTC, as also seen in the wind power increment data (6.1).

At 22UTC, there appeared to be much larger areas of stronger reflectivity values in all four of the fields as seen in Figure 6.10. However, by this time, the simulated reflectivity fields appeared to propagate at a slower velocity then the field generated by the NEXRAD mosaic: At 22UTC the strongest core of reflectivity produced in the NEXRAD field is much farther to the east than the simulated fields. It can be argued that the highest reflectivity cores within the simulated fields are farther to the north than the NEXRAD field, indicating a northward shift in reflectivity instead of a spatiotemporal lag. Regardless, there does seem to be a spatial displacement of roughly 200-250km between these cores.

Since the spatial coverage of the reflectivity values exceeding the two thresholds were

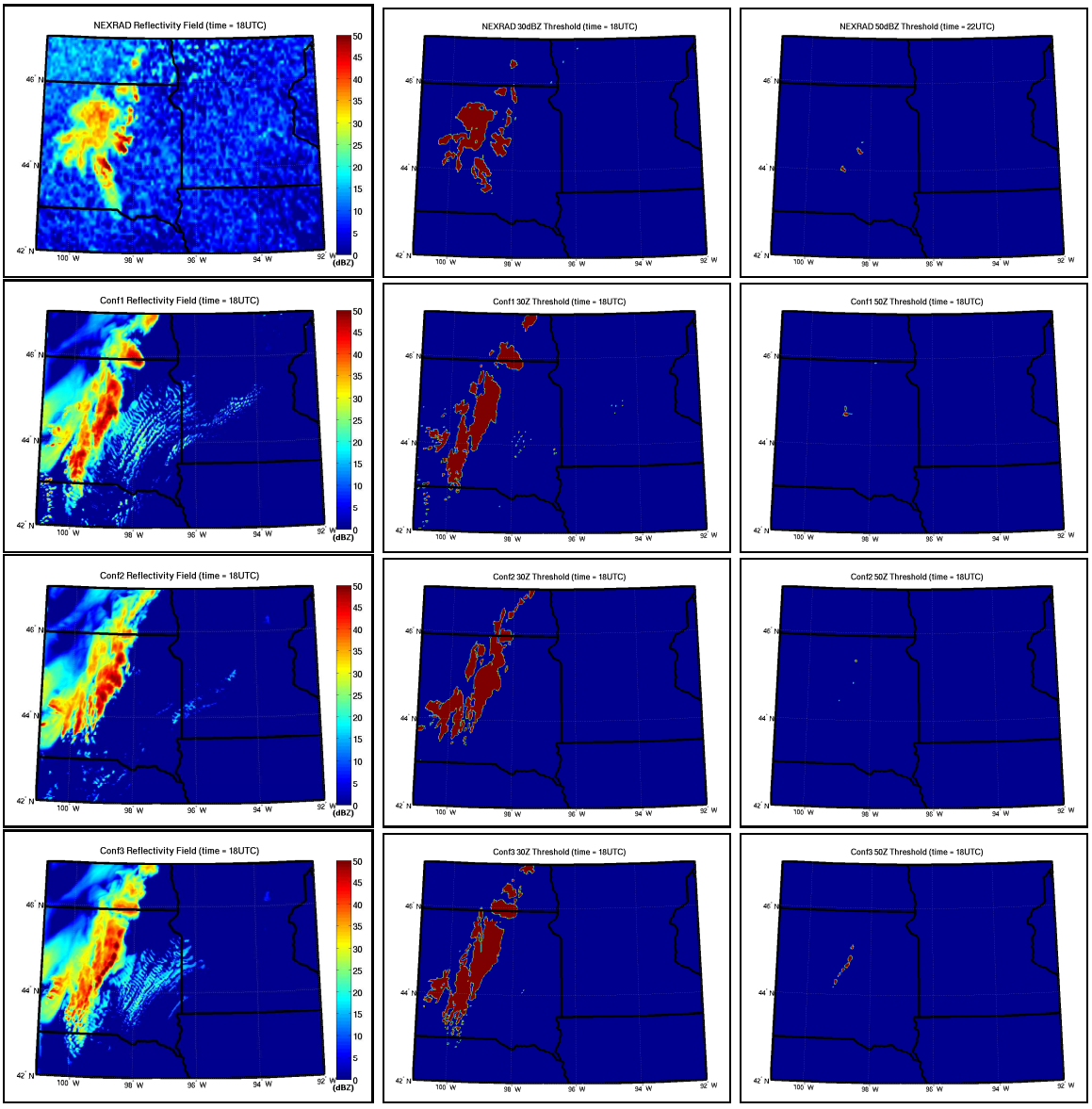

Figure 6.9: Radar reflectivity plots using observed NEXRAD data (top row) and conf1, conf2 and conf3 (left), along with 30dBZ (center) and 50dBZ (right) thresholds at 18UTC. Threshold fields will be used to calculate the Partial Hausdorff Distance.

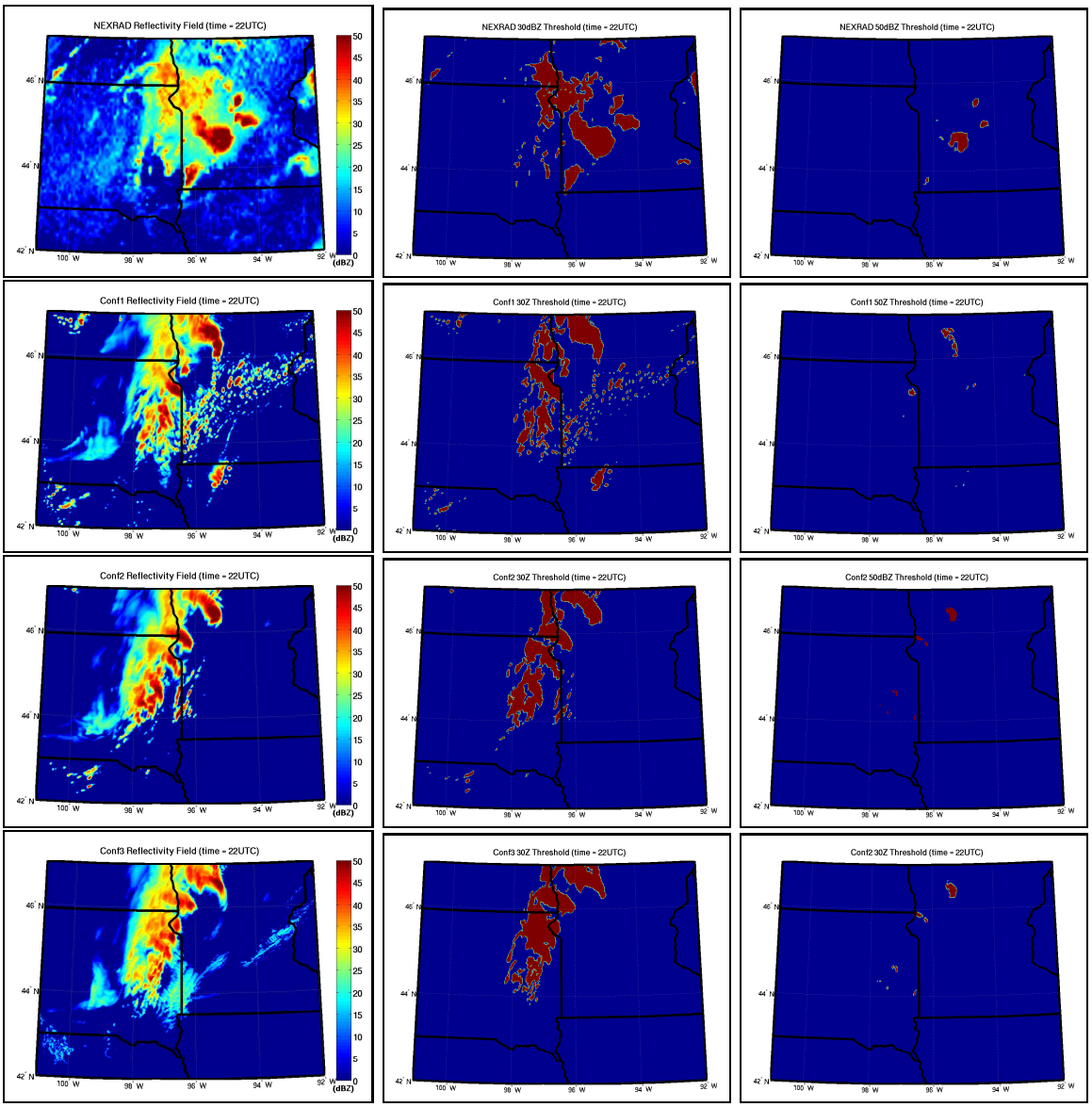

Figure 6.10: Similar to 6.9 for 22UTC

much greater at 22UTC when compared to 18UTC, the HD_p values are much lower (Table 6.5). When a threshold value of 30dBZ is applied to the field, conf1 yielded a very low HD_p value, while conf2 and conf3 generated higher values. This was the only time and threshold that conf2 generated a lower HD_p value than conf3. Further, the value yielded by conf1 at 22UTC is the lowest for the two times and threshold values.

Furthermore, at 22UTC all three of the configurations generated extremely similar HD_p values with a threshold of 50dBZ was applied (Table 6.6). Conf1 again generated the lowest HD_P value.

Table 6.6: Table of Partial Hausdorff Distance values for radar reflectivity data at 22UTC.

(22UTC)	HD_p @ Th=30dBZ	HD_p @ Th=50dBZ
Conf1	43	60
Conf2	83	65
Conf3	120	61

The quantitative results appeared to be consistent with the preliminary, qualitative image comparison and the previous section which evaluated wind power increment. After calculating the HD_p for all three configurations, conf1 was shown to generate the overall lowest mean HD_p value during 18UTC and 22UTC. However, conf3 generated the lowest mean value of HD_p for both times when a threshold of 50dBZ was applied to the field. In addition, all the three configurations generated the lowest values of HD_p during 22 UTC; when reflectivity values were shown to be the higher than at 18UTC. The HD_p did quantitatively represent the similarity between the reflectivity fields in the simulations and observed data. As shown, different times during the simulation created different outcomes of the HD_p values, as did the applied threshold. Regardless, it can be concluded from this analysis that during the two times of this study, and the two thresholds applied, that conf1 generated the most similarity in reflectivity field, when compared to the NEXRAD reflectivity mosaic. This information could be used in conjunction with the use of meteorological covariates to assess that conf1 was superior to conf2 and conf3 at simulating the spatial location of wind ramp events, as seen in the wind power increment section.

Kantorovich Distance:

Interestingly, the KD was not affected by the aforementioned issue with conf2 in the early portion of the simulation with high thresholds, as seen in Table 6.7. In this case, conf2 actually generated the lowest value for KD for a 50dBZ threshold at 18UTC and 22UTC. Regardless, conf1 was again shown to perform the best with the two thresholds, during the times analyzed. However, conf2 appeared to perform better than conf3 according to KD. Further, conf3 tended to generate the highest BCC values for the 22UTC time, but had relatively low values during 18UTC.

Table 6.7: Table of Kantorovich Distance values for radar reflectivity data at 18UTC.

(18UTC)	KD @ Th=30dBZ	BCC (Th=30dBZ)	KD @ Th=50dBZ	BCC (Th=50dBZ)
Conf1	2.19×10^7	0.49	4.35×10^5	0.54
Conf2	2.28×10^7	0.57	4.10×10^5	0.54
Conf3	2.81×10^7	0.52	4.16×10^5	0.4

Table 6.8: Table of Kantorovich Distance values for radar reflectivity data at 22UTC.

(22UTC)	KD @ Th=30dBZ	BCC (Th=30dBZ)	KD @ Th=50dBZ	BCC (Th=50dBZ)
Conf1	7.18×10^6	0.76	1.67×10^6	1.55
Conf2	9.52×10^6	0.93	1.33×10^6	1.69
Conf3	1.35×10^7	1.26	3.54×10^6	12.4

When using the BCC, KD and HD_p in conjunction, conf1 could again be considered the best forecast based on these values when using the reflectivity fields and the wind fields. Using just the KD and BCC can give evidence that conf1 was out-performing the other two configurations, but adding the HD_p analysis can help indicate that the spatial and amplitude errors associated with conf1 were the lowest overall of the three forecasts.

6.3.3 Traditional Metrics

In this final portion of the real case study, it is important to compare the traditional and proposed metrics' generated values to build on the notion that the proposed metrics are more powerful metrics than the traditional metrics. For this portion, only 22UTC is being used, because again, there are very few values in the 18UTC fields that exceed the largest thresholds. Using only 22UTC should be efficient enough to further illustrate the power of the proposed metrics.

Table 6.9: Values of RMSE, MAE, ETS and r^2 values for real case study, with configuration 1 at 22UTC.

Error Values	RMSE	MAE	ETS	r^2
dBZ50	64.78	79.89	-2.53×10^{-4}	1.3910^{-5}
dBZ30	45.56	55.82	-0.016	1.38×10^{-5}
Wind50	2.10	4.63	-4.57×10^{-4}	0.0012
Wind20	0.5938	0.55	-0.002	0.0052

When evaluating the outcomes for 22UTC using traditional metrics, there are some consistent results, and some inconsistent results when comparing the error values with the proposed metrics. First, the values of r^2 and ETS are extremely similar for all three configurations, for each of the respective thresholds. Further, the RMSE values are generally the same for the three configurations with only slight differences. The only traditional metric that appeared to be able to differentiate between the accuracy of the forecasts was the MAE.

Table 6.10: Similar to Table 6.9 with conf2 field.

Error Values	RMSE	MAE	ETS	r^2
dBZ50	63.88	78.41	-2.55×10^{-4}	0.003
dBZ30	48.34	58.15	-0.016	0.002
Wind50	12.59	158.63	-1.42×10^{-5}	0.0093
Wind20	1.13	1.48	-0.001	0.0119

It can be seen that the error recorded for the wind power increment field at 50% in conf1 was much less than the value generated for conf2 and conf3. This result is consistent with what was seen using the proposed metrics. It is important to note that since the RMSE, ETS and r^2 all generate similar values for rather dissimilar fields, it can be further concluded that the proposed metrics are more powerful as forecast verification metrics to quantify spatial displacement and amplitude-based errors than the traditional metrics.

Table 6.11: Similar to Table 6.9 with conf3 field.

Error Values	RMSE	MAE	ETS	r^2
dBZ50	61.99	72.98	-2.78x10^{-4}	1.7710^{-6}
dBZ30	44.71	54.93	-0.016	1.7610^{-6}
Wind50	12.58	158.56	-1.42x10^{-4}	0.0033
Wind20	1.97	4.11	-4.7x10^{-4}	0.0058

Chapter 7

Concluding Remarks and Future Perspectives

A short, non-comprehensive study on wind ramp climatology indicated that seasonality plays a larger role in wind ramp generation than the ENSO. As seen in chapter 2, the presence of either the El Nino or La Nina had little impact on annual wind ramp generation. However, future work might include a study that addresses the possible seasonal influence that these teleconnection phenomena generate. For instance, does the inclusion of the ENSO increase or decrease wind ramp frequency in various areas of the study domain during various seasons?

Further, the outcome of this analysis does suggest that spring and fall months may have a higher frequency of wind ramp events, when compared to summer and winter months. This is thought to mainly be attributed to the severe weather season in the Central CONUS during the spring and fall, the northern shift of the Polar Jet Stream in the summer and the lack of strong temperature gradients over most of the CONUS during the winter.

Additionally, the influence of air-sea temperature gradients may play a large role in coastal wind ramp generation. It was shown that there appeared to be a higher frequency of wind ramp events over the Gulf Stream and Atlantic Coast, due to the underlying warm ocean surface and cold continental air interactions. This occurrence appears to decrease drastically during the summer months when the air temperature and sea surface temperature gradient is much smaller.

It was also addressed in this thesis that modern forecast verification metrics have

shown flaws in quantifying spatial displacements and amplitude-based errors in model forecasts of various meteorological fields. Without the ability to quantify these forecast errors and how they are improving with time, forecast accuracy will increase at a slower pace. Lack of high-resolution wind data have been shown to be problematic in short-term wind forecasting.

However, using radar reflectivity fields as a meteorological covariate for near-surface wind speeds can be useful in the wind industry to estimate the temporal and spatial location of wind ramp events associated with thunderstorms. Further, radar reflectivity has been loosely linked to storm longevity; thus, using upstream reflectivity data could assist in short-term wind ramp forecasting by means of meteorological covariates and/or data assimilation. Using the aforementioned covariates, the spatial displacement errors of forecasts using high-resolution NWP model data can be quantified using the HD_p, when a threshold reflectivity value is administered.

It was shown in the previous section that using the HD_p to analyze the spatial displacement error of forecast fields when compared to observed data can quantify the displacement with a single value. As the HD_p value increased, the spatial displacements in the fields were shown to be increasing as well, indicating less similarity between the two fields. Although a threshold must be implemented to generate a binary field, there is still a valid representation of the displacement error when using the HD_p. Further, it was shown that as the reflectivity values increased in the fields, indicating stronger storms, the HD_p values began to decrease. This may shed light on the model's ability to simulate convection better than overall storm structure.

There are, however, some shortcomings of a metric that only produces values for spatial displacement, such as the one addressed in this thesis. First, applying a threshold value to a radar reflectivity field assumes that the wind at the near-surface is consistent with all reflectivity values with similar values; there is not a direct linear relationship between the radar reflectivity values and the near-surface wind values. However, since ramp events in this study are considered dichotomous, this metric can be used to indicate when a wind ramp is expected to occur due to the location of high-reflectivity values in the field.

Second, reflectivity fields are only being used as a meteorological covariate for the wind field at the near-surface. We have shown, through visualization, that there can be a relationship between the location of high radar reflectivity values and both high near-

surface winds, and strong wind increment values in the model simulations. However, the precise location of using a collocated high-reflectivity value and wind ramp event was not evaluated.

In addition, in this study 2km horizontal resolution simulations were conducted using the WRF model. However, the value HD_p can only be interpreted based on the size of the grid spaces used. Therefore, even when a small value for the HD_p is given, there can still be a displacement error of 2km within the grid cell itself, due to the resolution of the simulations. Further, a simulation using a 2km resolution may produce a much different HD_p value when a different resolution size is used. It is safe to say that lower resolutions for the same domain size, which would generate more grid cells within the domain, would generate a higher HD_p value since the spatial displacement is based on the amount of grid cells that are displaced. However, models have shown to generate more accurate results when a higher resolution simulation is conducted.

Furthermore, the parameter set for the HD_p can be changed. For this study, we used a 75% threshold for the values examined, which would decrease the effect of outliers. However, a change in this value would yield different results. A larger value, say 95%, would most likely generate a higher value of HD_p since more outlier values would be used. It should be noted that the 75% value was chosen relatively arbitrarily, but appeared to quantify the displacements between the models accordingly.

With that being said, a metric that will calculate the amplitude-based and spatial displacement errors of forecasts is needed. Without having to analyze a binary field, there is no ambiguity involved in selecting a reflectivity threshold value that is associated with thunderstorm activity. Therefore, a forecast verification metric that will quantify displacement and amplitude-based errors has been developed. This metric is used in a similar fashion as the HD_p; calculating a point-by-point similarity between observed and forecast fields. However, this metric will hold the values at each grid point and use these values to show the displacement of the amplitude field. This metric is the KD. The KD is a solution to a transportation problem that quantifies the cost of transporting one field onto another field. It was shown to quantify spatial, along with amplitude errors in forecast fields.

However, there are some issues that are generated when using the bias correction on the fields, since the sum of the observed field must be equal to the sum of the forecast field. For example, assume there are three fields: the observed field (Ob), and two forecasts (F1

and F2). In this example F1 is very similar in both spatial distribution and amplitude to Ob. Additionally, F2 has the same similar spatial distribution as F1, but the amplitude field is drastically overestimated. In this example, from a qualitative standpoint, F1 is clearly the better forecast due to the similar spatial distribution and amplitude that mirrors the Ob field. However, after applying the bias correction to the fields, F1 and F2 appear to be almost identical quantitatively. Therefore, the very algorithm that enables the fields to be compared to one another could actually play a role in generating some issues with the verification of which forecast is better.

Finally, the traditional metrics were compared to the proposed metrics using the same two case studies. The traditional metrics were shown to be unable to differentiate between many of the forecast field errors in both studies. However, the proposed metrics were indeed able to not only differentiate between the fields, but were also able to accurately quantify the errors. This notion is much more easily seen in the idealized case study, but was also identified in the real case study as well.

The main research question which asked if the traditional metrics used today in the wind industry were accurate in calculating spatiotemporal and amplitude errors was answered by using the two case studies. The modern metrics have a difficult time quantifying these errors, and new metrics are needed to increase the ability to assess forecast accuracy. The proposed metrics were able to more accurately quantify the errors found in the spatial displacements and amplitude errors of the forecast fields.

Future work may include a study with the bias correction of the fields to generate a result that may not be free of amplitude errors. Further, application of reflectivity gradient to these metrics may also be used, which may produce results consistent with gust front detection where sharp gradients in reflectivity can be used as opposed to merely a high reflectivity field for wind ramp events. In addition, a more comprehensive climatology study is needed to fully understand the spatial and temporal distributions of wind ramp events in the United States.

REFERENCES

D. Ahijevych, E. Gilleland, B. G. Brown, and E. E. Ebert. Application of spatial verification methods to idealized and nwp-gridded precipitation forecasts. Weather and Forecasting, 24(6):1485–1497, 2009.

M. Alexiadis, P. Dokopoulos, H. Sahsamanoglou, and I. Manousaridis. Short-term forecasting of wind speed and related electrical power. Solar Energy, 63(1):61–68, 1998.

R. W. Arritt. Numerical modelling of the offshore extent of sea breezes. Quarterly Journal of the Royal Meteorological Society, 115(487):547–570, 1989.

J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. Numerische Mathematik, 84(3):375–393, 2000.

A. Bossavy, R. Girard, G. Kariniotakis, et al. Forecasting uncertainty related to ramps of wind power production. In European Wind Energy Conference and Exhibition 2010, EWEC 2010, volume 2, 2010.

A. Bossavy, R. Girard, and G. Kariniotakis. Forecasting ramps of wind power production with numerical weather prediction ensembles. Wind Energy, 2012.

B. G. Brown and A. H. Murphy. Verification of aircraft icing forecasts: The use of standard measures and meteorological covariates. Preprints, 13th Conf. on Probability and Statistics in the Atmospheric Sciences, San Francisco, CA, Amer. Meteor. Soc., page 251252, 1996.

B. G. Brown, R. W. Katz, and A. H. Murphy. Time series models to simulate and forecast wind speed and wind power. Journal of climate and applied meteorology, 23: 1184–1195, 1984.

A. N. Celik. Energy output estimation for small-scale wind power generators using weibull-representative wind data. Journal of Wind Engineering and Industrial Aerodynamics, 91(5):693–707, 2003.

N. Cosack, S. Emeis, and M. Kühn. On the influence of low-level jets on energy production and loading of wind turbines. In Wind Energy, pages 325–328. Springer, 2007.

A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. Renewable and Sustainable Energy Reviews, 12(6):1725–1744, 2008.

T. J. Crowley. Causes of climate change over the past 1000 years. Science, 289(5477): 270–277, 2000.

N. Cutler, M. Kay, K. Jacka, and T. S. Nielsen. Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and wppt. Wind Energy, 10(5):453–470, 2007.

M. S. De Pondeca, G. S. Manikin, G. DiMego, S. G. Benjamin, D. F. Parrish, R. J. Purser, W.-S. Wu, J. D. Horel, D. T. Myrick, Y. Lin, et al. The real-time mesoscale analysis at noaa's national centers for environmental prediction: Current status and development. Weather and Forecasting, 26(5):593–612, 2011.

V. Di Gesú and V. Starovoitov. Distance-based functions for image comparison. Pattern Recognition Letters, 20(2):207–214, 1999.

E. Ela and B. Kirby. ERCOT event on February 26, 2008: Lessons learned. National Renewable Energy Laboratory, 2008.

U. Focken and M. Lange. Final report–wind power forecasting pilot project in alberta, canada. Oldenburg, May, 26, 2008.

J. Freedman, M. Markus, and R. Penc. Analysis of west texas wind plant ramp-up and ramp-down events. AMS Truewind Report, pages 250–278, 2008.

G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl. The state-of-the-art in short-term prediction of wind power: A literature overview. Technical report, ANEMOS. plus, 2011.

E. Gilleland, D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert. Intercomparison of spatial forecast verification methods. Weather and forecasting, 24(5):1416–1430, 2009.

B. Greaves, J. Collins, J. Parkes, and A. Tindal. Temporal forecast uncertainty for ramp events. Wind Engineering, 33(4):309–319, 2009.

F. Hillier and G. Lieberman. Introduction to mathematical programming. 1977.

B. D. Hirth, J. L. Schroeder, and C. C. Weiss. Surface analysis of the rear-flank downdraft outflow in two tornadic supercells. Monthly Weather Review, 136(7):2344–2363, 2008.

M. R. Hjelmfelt. Structure and life cycle of microburst outflows observed in colorado. Journal of Applied Meteorology, 27:900–927, 1988.

M. Hoel and S. Kverndokk. Depletion of fossil fuels and the impacts of global warming. Resource and Energy Economics, 18(2):115–136, 1996.

X.-M. Hu, J. W. Nielsen-Gammon, and F. Zhang. Evaluation of three planetary boundary layer schemes in the wrf model. Journal of Applied Meteorology and Climatology, 49 (9):1831–1844, 2010.

Z. Huang and Z. Chalabi. Use of time-series analysis to model and forecast wind speed. Journal of Wind Engineering and Industrial Aerodynamics, 56(2):311–322, 1995.

D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the hausdorff distance. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 15(9):850–863, 1993.

L. Jones, R. Zavadil, W. Grant, et al. The future of wind forecasting and utility operations. Power and Energy Magazine, IEEE, 3(6):57–64, 2005.

T. Kaijser. Computing the kantorovich distance for images. Journal of Mathematical Imaging and Vision, 9(2):173–191, 1998.

K. A. Larson and K. Westrick. Short-term wind forecasting using off-site observations. Wind energy, 9(1-2):55–62, 2006.

D. Lew, M. Milligan, G. Jordan, and R. Piwko. The value of wind power forecasting. NREL Conference Paper: 91st American Meteorological Society Annual Meeting, the Second Conference on Weather, Climate and the New Energy Economy, Washington D.C., January 26, 2011, 2011.

Y.-L. Lin. Mesoscale dynamics. Cambridge University Press, 2007.

P. MacKeen, H. Brooks, and K. Elmore. Radar reflectivity-derived thunderstorm parameters applied to storm longevity forecasting. Weather and forecasting, 14(2):289–295, 1999.

M. R. Milligan, A. H. Miller, and F. Chapman. Estimating the economic value of wind forecasting to utilities. National Renewable Energy Laboratory Golden, CO, 1995.

S. Minobe, A. Kuwano-Yoshida, N. Komori, S.-P. Xie, and R. J. Small. Influence of the gulf stream on the troposphere. Nature, 452(7184):206–209, 2008.

M. Negnevitsky and P. Johnson. Very short term wind power prediction: A data mining approach. In Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE, pages 1–3. IEEE, 2008.

L. W. O'Neill, D. B. Chelton, and S. K. Esbensen. The effects of sst-induced surface wind speed and direction gradients on midlatitude surface vorticity and divergence. Journal of Climate, 23(2):255–281, 2010.

K. Orwig, B. Hodge, G. Brinkman, E. Ela, M. Milligan, V. Banunarayanan, S. Nasir, and J. Freedman. Economic evaluation of short-term wind power forecasts in ERCOT: Preliminary results. National Renewable Energy Laboratory, 2012.

S. E. Peckham, R. B. Wilhelmson, L. J. Wicker, and C. L. Ziegler. Numerical simulation of the interaction between the dryline and horizontal convective rolls. Monthly weather review, 132(7):1792–1812, 2004.

D. Pondeca, G. MSFV, D. Parrish, R. Purser, W. Wu, G. DiMego, J. Derber, S. Benjamin, J. Horel, S. Lazarus, et al. The status of the real time mesoscale analysis system at ncep. 2007.

B. Storm, J. Dudhia, S. Basu, A. Swift, and I. Giammanco. Evaluation of the weather research and forecasting model on forecasting low-level jets: implications for wind energy. Wind Energy, 12(1):81–90, 2009.

K. Tan and S. Islam. Optimum control strategies in energy conversion of pmsg wind turbine system without mechanical sensors. Energy Conversion, IEEE Transactions on, 19(2):392–399, 2004.

J. W. Taylor, P. E. McSharry, and R. Buizza. Wind power density forecasting using ensemble predictions and time series models. Energy Conversion, IEEE Transactions on, 24(3):775–782, 2009.

K. E. Trenberth. The definition of el nino. Bulletin of the American Meteorological Society, 78(12):2771–2777, 1997.

V. Venugopal, S. Basu, and E. Foufoula-Georgiou. A new metric for comparing precipitation patterns with an application to ensemble forecasts. Journal of geophysical research, 110(D8):D08111, 2005.

S.-P. Xie, H. Xu, W. S. Kessler, and M. Nonaka. Air-sea interaction over the eastern pacific warm pool: Gap winds, thermocline dome, and atmospheric convection*. Journal of Climate, 18(1):5–20, 2005.