

Final Technical Report Cover Page

Federal Agency to which Report is submitted: DOE EERE – Wind & Water Power Program

Recipient: AWS Truepower, LLC Award Number: DE-EE0004420

Project Title: The Wind Forecast Improvement Project (WFIP): A Public/Private Partnership for Improving Short Term Wind Energy Forecasts and Quantifying the Benefits of Utility Operations – the Southern Study Area.

Project Period: September 1, 2010 - January 31, 2014

Principle Investigator: Jeffrey M. Freedman, Lead Research Scientist, AWS Truepower,
jfreedman@awstruepower.com, 518-213-0044 (now at the Atmospheric Sciences Research Center, jfreedman@albany.edu, 518-437-8737)

Report Submitted by: Jeffrey M. Freedman, Lead Research Scientist

Date of Report: April 30, 2014

Covering Period: September 1, 2010 – January 31, 2014

Working Partners: National Oceanographic and Atmospheric Administration (James Wilczak, Team Lead, Boundary Layer Processes and Applications, james.m.wilczak@noaa.gov, 303-497-6245)

National Renewable Energy Laboratory (Erik Ela, Erik.Ela@nrel.gov, 303-384-7089)

Electric Reliability Council of Texas (Isabel Flores, Manager, Operations Analysis, IFlores@ercot.com, 512-248-6531)

MESO, Inc. (John W. Zack, Principal, john@meso.com, 518-283-5169)

Texas Tech University (John Schroeder, Director, National Wind Institute Professor of Atmospheric Science, john.schroeder@ttu.edu, 806-742-2813; Brian Ancell, Assistant Professor, Atmospheric Science Group, brian.ancell@ttu.edu, 806-742-3143)

University of Oklahoma Center for Analysis and Prediction of Storms (Keith Brewster, Senior Research Scientist and Associate Director, kbrewster@ou.edu, 405-325-6115)

North Carolina State University (Sukanta Basu, Associate Professor, Department of Marine, Earth, and Atmospheric Sciences, sukanta_basu@ncsu.edu, 919-513-7776)

ICF International (Venkat Banunarayanan, Principal, Venkat.Banunarayanan@icfi.com, 703-218-2516)

Cost-Sharing Partners: ERCOT, MESO, TTU, OU, NCSU, ICF International

DOE Project Team: DOE HQ Program Manager – Jose Zayas
DOE Field Contract Officer – Pamela Brodie

DOE Field Grants Management Specialist – Jane Sanders
DOE Field Project Officer – Brad Ring
DOE/CNJV Project Monitor – Yelena Onnen

Signature of Submitting Official: _____
(electronic signature is acceptable)

This report is based upon work supported by the U. S. Department of Energy under Award No. DE- EE0004420. Any findings, opinions, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the Department of Energy.

Table of Contents

Executive summary	1
1 Introduction.....	5
2 Background.....	7
2.1 Project objectives	7
2.2 Team members.....	8
2.3 User Group	10
3 Measurement campaign	10
3.1 Selection of study area.....	10
3.2 Phenomena affecting short-term wind forecasts	11
3.3 Field campaign – instrument deployment	15
3.3.1 Existing networks	16
3.3.2 Observation targeting.....	17
3.3.3 Proposed deployment.....	19
3.3.4 Deployment of NOAA, DOE, AWST, and partner equipment.....	20
3.3.5 LiDAR deployments	24
3.4 Results and discussion.....	25
3.4.1 Data acquired	25
3.4.2 Ramp frequency	28
3.5 Gap filling analysis and related work (NCSU)	28
3.5.1 Spectral Characterization.....	28
3.5.2 Gap-Filling and De-noising	30
3.5.3 Results and summary.....	31
4 Model systems and forecast campaign	33
4.1 Forecast systems.....	33
4.1.1 AWST/MESO	33
4.1.2 CAPS OU	42
4.1.3 TTU.....	44
5 Results and discussion – Forecasting.....	49
5.1 AWST/MESO	49
5.1.1 Deterministic forecast performance.....	50
5.1.2 Ensemble forecasts	56
5.1.3 Probabilistic Ramp Rate Forecasts	59
5.1.4 Data denial experiments.....	61
5.1.5 Ramp event analysis.....	65
5.1.6 Ramp Verification	66
5.1.7 Phenomena-based model performance.....	68
5.2 OU CAPS	73
5.3 TTU	77
5.3.1 Month-long period sensitivity study	77
5.3.2 Wind ramp forecasts	82
6 Accomplishments.....	86
7 Summary and Conclusions.....	87
7.1 AWST WFIP Forecast System	87
7.2 OU CAPS sensitivity studies.....	89

7.3 TTU data assimilation sensitivity studies	90
8 Recommendations	91
9 References	92
Appendices.....	100
9.1 Appendix A - Publications.....	100
9.2 Appendix B - Conferences, Workshops, and User Group Meetings	101
List of Tables.....	102
List of Figures	102
List of Acronyms.....	106

Executive summary

This Final Report presents a comprehensive description, findings, and conclusions for the Wind Forecast Improvement Project (WFIP)--Southern Study Area (SSA) work led by AWS Truepower (AWST). This multi-year effort, sponsored by the Department of Energy (DOE) and National Oceanographic and Atmospheric Administration (NOAA), focused on improving short-term (15-minute – 6 hour) wind power production forecasts through the deployment of an enhanced observation network of surface and remote sensing instrumentation and the use of a state-of-the-art forecast modeling system. WFIP encompassed two regions in the central U.S.: 1) the Northern Study Area (NSA), covering parts North and South Dakota, Minnesota, Nebraska, Iowa, and Wisconsin, and 2) the Southern Study Area (SSA), encompassing most of Texas. WindLogics led the NSA, effort while AWST directed the SSA campaign.

The key objectives of WFIP were to:

1. Assemble a top-notch team of private, government, and academic partners with collective experience and expertise in all facets required to ensure a successful completion of the proposed two-year project.
2. Select a region of interest (ROI), in this case geographically centered on the Electric Reliability Council Of Texas (ERCOT), that contains a sufficient number of wind farms and characteristic weather phenomena that enables transfer of the resulting work to other regions of the U.S.
3. Design and execute a targeted observation campaign, in consultation with DOE AND NOAA, to facilitate deployment of an enhanced network of existing and newly deployed surface and remote sensing observation platforms that will successfully capture the spectrum of representative spatial and temporal phenomena that principally influence the short-term operation of wind power production.
4. Using a diverse approach of individual model, statistical, ensemble-based approaches, and improved data assimilation systems, determine the value of additional atmospheric observations on wind power production forecasts leading to improvements in forecasting the timing, magnitude, and persistence of wind ramp events.
5. Develop improved economic metrics to demonstrate cost savings for utilities and operators from improved short-term (0 – 6 hour) wind power forecasts.
6. Based on improved short-term (0- to 6-hour) wind power forecasts, demonstrate how applied forecast performance metrics better reflect the manner in which operators and other wind forecast users are sensitive to wind power production forecast errors on electric power systems.
7. Disseminate project results to interested stakeholders, and define the necessary spatial and temporal characteristics of a nation-wide mesonet observation system devoted to wind energy applications.

There were three key attributes that formed the core foundation for WFIP: 1) the field deployment and data dissemination schemes; 2) the tailored forecast model systems; and 3) the economic evaluation. These qualities set WFIP apart from most classic field and modeling studies – an emphasis on determining the quantitative value of deploying additional instrumentation assimilated by an improved wind power production

forecast system and how those savings may be shared among the beneficiaries (that is, utilities, ISOs, wind generators, and rate payers).

The AWST SSA team included the Electric Reliability Council of Texas (ERCOT): the Balancing Authority (BA) responsible for managing much of the electrical grid in Texas, three private companies: AWST, MESO, and ICF International; three academic institutions: North Carolina State University, the University of Oklahoma (OU) Center for Analysis and Prediction of Storms (CAPS), and Texas Tech University (TTU); and the National Renewable Energy Laboratory (NREL).

Work on the SSA campaign began in November 2010 with a modeling sensitivity study performed by AWST to determine the best locations for the WFIP enhanced observation network (composed of surface meteorological, tall tower, and remote sensing platforms, assets contributed by DOE, NOAA, AWST, TTU, and NCSU). Instrument deployment occurred during July 2011, and all WFIP SSA instruments were operational by 17 July 2011. The official field and modeling campaign was conducted from 26 August 2011 through 13 September 2012.

The WFIP SSA modeling systems included

1. The AWST WFIP Forecast System (WFIPFS) composed of a 9-member ensemble of three different numerical weather prediction (NWP) models incorporating a variety of data assimilation and model physics schemes. Real-time forecasts were produced for the entire 1-year campaign, and a variety of sensitivity studies were performed;
2. The OU CAPS Advanced Regional Prediction System (ARPS) which produced real-time 3-km grid scale forecasts for the entire field and modeling campaign with enhanced vertical resolution covering the WFIP domain and adjacent portions of neighboring states; and
3. The TTU Weather Research and Forecast (WRF) system that performed several data assimilation sensitivity studies during select periods of the field and modeling campaign.

ICF International and NREL conducted a comprehensive economic analysis of the value of WFIPFS forecasts and assimilated observations. The study used the results of the pre-WFIP Baseline and WFIPFS power production forecasts as applied to operational ERCOT market rules to determine the system benefits savings, reduction in curtailment costs and load payments, and ancillary costs savings (the Non-Spinning Reserve Service, or NSRS).

DOE, however, has decided to undertake additional studies to explore the complex interactions between wind forecasting and power system operations prior to publication of the economic analysis results. The initial work performed by the WFIP teams provided important insight into the benefits and shortcomings of various power system assumptions, market designs, and modeling tools in identifying costs and savings. The desire to explore these important issues in more detail is the impetus for the new analysis. Over the next year, DOE plans to engage with industry experts, grid operators and economic modelers to accurately define methodologies that provide quantification of total financial savings and other ancillary benefits of improved short-term wind power production forecasts

For the SSA, a highly successful field deployment and data acquisition operations produced well over 90% data recovery during real-time operations. The yearlong field campaign captured an excellent set of observations

depicting the phenomena responsible for the characteristic weather and spectrum of ramp events typical of the ERCOT domain. This data set should prove invaluable for future post-WFIP analysis and atmospheric boundary layer studies.

Key findings from the SSA modeling and forecast effort include:

1. The AWST WFIP modeling system produced an overall 10 – 20% improvement in wind power production forecasts over the existing Baseline system, especially during the first three forecast hours;
2. Improvements in ramp forecast skill, particularly for larger up and down ramps;
3. The AWST WFIP data denial experiments showed mixed results in the forecasts incorporating the experimental network instrumentation; however, ramp forecasts showed significant benefit from the additional observations, indicating that the enhanced observations were key to the model systems' ability to capture phenomena responsible for producing large short-term excursions in power production;
4. The OU CAPS ARPS simulations showed that the additional WFIP instrument data had a small impact on their 3-km forecasts that lasted for the first 5-6 hours, and increasing the vertical model resolution in the boundary layer had a greater impact, also in the first 5 hours; and
5. The TTU simulations were inconclusive as to which assimilation scheme (3DVAR versus EnKF) provided better forecasts, and the additional observations resulted in some improvement to the forecasts in the first 1 – 3 hours.

The results summarized briefly here and presented more in-depth in the following chapters have been disseminated, in preliminary form, at several conferences and workshops (see Appendix B - Conferences, Workshops, and User Group Meetings) and at two special User Group meetings held in conjunction with the Utility Variable Integration Group meetings in Tucson AZ (9 February 2012) and at Salt Lake City UT (27 February 2013).

There is still opportunity for significant improvement, and better understanding (and therefore better parameterizations or better model physics) of the phenomena that drive the wind and large changes (i.e. ramp events) in the wind field. In particular, follow on efforts should include:

1. More comprehensive economic analysis inclusive of all forecasting time scales and indirect cost savings;
2. Identification of the ultimate recipients of economic benefits;
3. Additional, more extensive analysis of phenomena responsible for ramp events/outliers;
4. Further analysis of the forecast model performance, including
 - a. A more in-depth study of the enhanced observation network value, through data denial efforts focused on particular phenomena and the largest ramp events;
 - b. An inter-comparison of the data denial experiments for the Northern and Southern Study Areas that further quantifies the advantages of having a denser network of remote sensing instrumentation;
 - c. A regime-based analysis (i.e. cold season versus warm season);

5. Additional sensitivity experiments to determine how the different components of the WFIP system contributed to more accurate forecasts, in terms of overall improvement, regime-based performance, and selected ramp events.

1 Introduction

Wind energy is a fast-growing source of electrical power in many countries, and it currently supplies around 3% of the world's generating capacity.¹ Wind is also playing an important role in weaning the world's economies off of non-renewable energy sources that contribute to environmental degradation. With growing concerns regarding climate change, reductions in cost of energy, and state policy initiatives such as Renewable Portfolio Standards (RPS)², a rapidly expanding fleet of utility-scale wind power facilities is now being incorporated into the grid. As of 2012, nine states obtain more than 10% of their electricity from wind energy (AWEA 2013). The number of transmission-connected systems above 20 MW in size is expected to increase dramatically, which will affect electric power system planning and operations processes. Also, there is increasing interaction between the distribution and transmission systems with the advent of demand-side strategies, electric vehicles, and more affordable storage.

Unlike conventional energy generation, the ability to accept wind power into electric power systems is dependent upon calculations using atmospheric observations and wind forecasts. In the event that power generation from wind does not meet projections, electric power system Balancing Authorities (BAs) are required to have sufficient energy reserves. As a result, BAs are proposing new rules, charges, and penalties on wind energy suppliers for not meeting energy generation projections. Several industry investigations (e.g. Fabbri et al. 2005; Cardell and Anderson 2009) have estimated that substantial savings in annual system production costs can be achieved with improved wind forecasting accuracy, particularly if predicting the magnitude and timing of ramp³ events in the 0 to 6 hour range can be improved.

A natural response to the increasing amounts of wind energy being integrated into local and regional grids has been Independent System Operator (ISO) and utility requirements mandating forecast systems. However, with wind now poised to become a major source of power for the U.S. (DOE 2008; DOE 2013), its inherent variability combined with system loads drawing more power from wind requires further improvements in forecasting, especially on the short (0-6 hour) time scales critical for system reliability and economic dispatch. This is a priority for the utilities, balancing authorities, and other market regulators, as the uncertainty of variable generation forecasts is still a major obstacle for many users (Makarov et al. 2010).

The first computer models devoted to wind power forecasting were developed during the 1980s, an outgrowth of a Pacific Northwest National Laboratory (PNNL) working group (Wendell et al. 1978; Bossanyi 1985). Throughout the 1990s, a variety of statistical approaches were employed.⁴ In 1999, eWind™, the predecessor of the forecasting system used in the Wind Forecasting Improvement Project (WFIP) Southern Study Area (SSA), was developed by AWS Truepower (AWST). In the early 2000s, the California Independent System Operator (CAISO) developed a centralized wind power forecasting system (Makarov et al. 2002). Since

¹ See <http://www.gwec.net/global-figures/wind-in-numbers/>

² Thirty-seven states now have mandatory or voluntary RPS programs. See <http://www.cleanenergystates.org/projects/state-federal-rps-collaborative/state-rps-annual-reports-and-compliance-reports-beta/>

³ A ramp is generically defined as an excursion of power production of some magnitude in a specified time frame.

⁴ See Argonne (2009) for a thorough review of the history of wind and wind power forecasting.

then, a large number of ISOs, utilities, and balancing authorities have deployed wind power forecasting systems (WPFS). And with deeper penetration of wind energy, forecasting the wind is ever more critical for developing and managing the modern electrical grid.

In response to this need, in 2010 the U.S. Department of Energy (DOE) put forth a solicitation: Enhancing Short Term Wind Energy Forecasting For Improved Utility Operations. This is now known as WFIP, a multi-year, DOE/National Oceanographic and Atmospheric Administration (NOAA) sponsored study whose main purpose is to demonstrate the value of additional atmospheric observations and model enhancements for improving 0- to 6-hour wind energy production forecasts. AWST and its partners were selected as one of two teams to develop and carry out a comprehensive field deployment and modeling exercise to fulfill the goals of WFIP. The AWST Team, which performed work for the Southern Study Area (SSA) centered in Texas, is composed of 1) the Electric Reliability Council of Texas (ERCOT), the Balancing Authority (BA) partner; 2) MESO, Inc.; 3) the University of Oklahoma Center for Analysis and Prediction of Storms (CAPS); 4) Texas Tech University (TTU); 5) North Carolina State University (NCSU); 6) ICF International; and 7) the National Renewable Energy Laboratory (NREL).

The key objectives for the WFIP SSA team were to

1. Assemble a top-notch team of private, government, and academic partners with collective experience and expertise in all facets required to ensure a successful completion of the proposed multi-year effort.
2. Select a region of interest (ROI), in this case geographically encompassing ERCOT, that contains a sufficient number of wind farms and characteristic weather phenomena that facilitates transfer of the resulting work to other regions of the U.S.
3. In conjunction with NOAA and DOE, design and execute a targeted observation campaign using a network of existing and newly deployed surface and remote sensing observation platforms to successfully capture the spectrum of representative spatial and temporal phenomena that principally influence the short-term operation of wind power production.
4. Use a diverse approach of individual model, statistical, and ensemble-based approaches, and improved data assimilation systems, to determine the impact of additional atmospheric observations on wind power production forecasts that lead to improvements in forecasting the timing, magnitude, and persistence of wind ramp events.
5. Develop improved economic metrics to demonstrate cost savings for utilities and operators from improved short-term (0 – 6 hour) wind power forecasts.
6. Demonstrate how existing and improved economic and new forecast performance metrics better reflect the manner in which operators and other wind forecast users are sensitive to wind power production forecast errors on electric power systems.
7. Disseminate project findings to interested stakeholders, and define the necessary spatial and temporal characteristics of a nation-wide mesonet observation system devoted to wind energy applications.

WFIP is the first nationally sponsored project to specifically focus on improvements to short-term wind power production forecasts, leveraging state-of-the-art forecast models, high resolution remote sensing and surface observation observations, and sophisticated economic models to capture the hourly cost savings resulting from improved power production forecasts. It is a large-scale experiment designed to test various methods of

improving short-term forecasts, including rapid-update-cycle forecasts, ensembles of forecasts, and dedicated observational networks. Furthermore, forecasting systems until now have not aimed specifically at forecasting ramps on short time scales to improve system reliability, but instead have emphasized general forecast performance metrics such as mean absolute error (MAE), with focus on next-day plant scheduling. Here, tailored metrics were applied (such as the Critical Success Index and Ranked Probability Skill Score; see Section 5) to provide appropriate performance statistics on the short-term (0 – 6 hr) forecast windows.

Although the SSA forecasting and observational work focuses on a geographical area covering western and central Texas, the results, techniques developed, and lessons learned from this study should be of value to utilities, ISOs, and Transmission System Operators (TSOs) in other regions interested in investing in or improving short-term wind power forecasting. To disseminate the information gained and lessons learned from WFIP, a Users Group was formed consisting of selected utilities, TSOs, and ISOs who are already engaged in or contemplating wind forecasting operations and have a strong interest in reviewing the progress, findings and conclusions from this work. Before, during, and after the field deployment and modeling phases of the project was completed, team members presented an overview of WFIP and preliminary findings at conferences and workshops sponsored by industry, utility groups, and the scientific community (see Appendix B - Conferences, Workshops, and User Group Meetings). Frequent interaction with NOAA and DOE scientists and policy makers also enhanced the visibility and vetting of the WFIP work. Through this review and outreach process, other forecasting programs, as well as the advancement of the country's community of practice, have already directly benefited.

The work performed and results presented here for the WFIP SSA constitute a team effort. Although AWST led the endeavor and oversaw all components of the project (field deployment, model development and forecasting, and the economic analysis), each team member performed admirably in their respective roles, with AWST/MESO providing the foundational modeling system and real-time power production forecasts throughout the one-year experiment period; ERCOT providing critical real-time power production and meteorological data from participating wind farms (accounting for nearly 90% of total capacity in the ERCOT domain); OU CAPS running a version of the ARPS model that was incorporated into the WFIP forecast system and performed post-operational sensitivity studies; TTU performing case study and data denial simulations and providing valuable data from its 200-m instrumented tower and 915 MHz wind profiler; NCSU providing a SoDAR at the TTU site and developing a data gap-filling scheme for the SoDAR and profiler data; and ICF and NREL performing the tailored and detailed economic analysis.

2 Background

2.1 Project objectives

The DOE acknowledged the need for enhancements in modeling and observation platforms in its 2008 report: U.S. Department of Energy Workshop Report: Research Needs for Wind Resource Characterization (Schreck et al. 2008). Key goals in that report relevant to this work include

- 1) Acquisition of observations for model validation and forecast enhancement; and
- 2) Improvement of industry and atmospheric modeling in

- a) the near real time, from a few minutes up to one hour ahead, to provide alerts on weather events which could significantly change the output of the regional wind capacity and affect system reliability; and
- b) one to several hours ahead, to anticipate rapid changes in output from a wind plant to the regional electrical power grid and the need for scheduling adequate reserve capacity to accommodate up and down movements in the wind plant output.

Recognizing these needs, the AWST Team engaged in the following efforts to meet the principal project objectives:

1. Assembled a preeminent team of private, government, and academic partners with collective experience and expertise in all facets of wind energy forecasting and field observations required to ensure a successful completion of the project.
2. Selected a region of interest (ROI), in this case geographically centered on the Electric Reliability Council Of Texas (ERCOT), that contains a sufficient number of wind farms with a large penetration of wind energy and characteristic weather phenomena that facilitates transfer of the resulting work to other regions of the U.S.
3. In collaboration with NOAA and DOE, deployed surface and remote sensing observation platforms to capture the spectrum of representative spatial and temporal phenomena that influence short-term operations of wind power production.
4. Used a diverse approach of individual model, statistical, and ensemble-based methods and advanced data assimilation systems, to determine through data denial experiments and other sensitivity studies, the potential improvement on wind power production forecasts, including the incorporation of the NOAA High Resolution Rapid Refresh (HRRR) model in the forecast suite and performance analysis.
5. Developed a tailored economic analysis approach with appropriate metrics that show quantifiable cost savings for utilities and operators from improved short-term (0 – 6 hour) wind power forecasts.
6. Demonstrated how existing and improved forecast performance metrics better reflect the manner in which operators and other wind forecast users are sensitive to wind power production forecast errors on electric power systems.
7. Disseminated project results to interested stakeholders, including utility, industry, and government groups.

2.2 Team members

The AWST Team members have strong skills and experience in the relevant disciplines. The team includes ERCOT, which manages an electric system with over 10,000 MW of installed wind power, three private companies with decades of experience in the renewable energy industry, three academic institutions, each of which have made contributions to the advancement of forecasting model capabilities, and the National

Renewable Energy Laboratory (NREL), a Federally Funded Research and Development Laboratory (FFRDC) that is world renowned for its work in renewables.

Team roles were tailored to team members' resources and capabilities. The distribution of team member expertise is summarized in Table 2-1.

Table 2-1. WFIP Southern Study Region Team members: capabilities, qualifications, and expertise.

Capabilities/Expertise							
Organization Name	Industry	Weather Forecasting	NWP Model Expertise	Instrumentation	Economic Analysis	Transmission Analysis	
ERCOT	Private/RTO					x	
AWS Truepower	Private	x	x	x			
MESO	Private	x	x				
ICF International	Private				x		
North Carolina State University	Academic	x	x	x			
Oklahoma University	Academic	x	x	x			
Texas Tech University	Academic	x	x	x			
National Renewable Energy Laboratory	FFRDC		x	x	x	x	

The team member organizations and their respective roles include:

1. **AWST**: led and managed all aspects of the project. AWST Principal Investigator, Dr. Jeffrey M. Freedman, coordinated the technical and logistical aspects of the work for the team including the kick-off and subsequent technical review meetings with NOAA and DOE, surveys required before instrumentation deployment, field campaign, data and economic analyses, all SSA modeling activities aside from those conducted by NOAA, and the drafting of this Final Report. AWST personnel also identified features from the observed data responsible for ramp events and worked with other team members on data quality control, forecast validation, and economic analyses.
2. **ERCOT**: as the BA partner, ERCOT provided necessary data and supported the evaluation of systems operations benefits and provided logistical support for the economics team (ICF and NREL) in determining the economic savings attributable to the improved short-term wind energy forecasts. Thirty-four wind-powered generating resources (WGRs--the wind farms), representing 85% of the wind generating capacity in ERCOT's territory, participated in WFIP in accordance with formalized non-disclosure agreements signed by team members and NOAA.
3. **MESO, Inc.**: designed and operated the WFIP forecasting system, and performed a number of analyses and multi-model (ensemble) sensitivity studies in collaboration with NOAA and other team members.
4. **Texas Tech University (TTU)**: prepared, deployed, operated, and maintained various measurement systems at the Reese Technology Center in Lubbock, TX and contributed a subset of model runs as part of a sensitivity study using an ensemble Kalman filter (EnKF) in a version of the Rapid Refresh/High Resolution Rapid Refresh (RR/HRRR) model.

5. **North Carolina State University (NCSU):** contributed a SoDAR to the field campaign and developed automated algorithms based on statistical learning theory for quality control of SoDAR and profiler data; NCSU also investigated the mesoscale spectral characteristics of the atmospheric boundary layer.
6. **Oklahoma University (OU) Center for Analysis and Prediction of Storms (CAPS):** used a version of the Advanced Regional Prediction System (ARPS) at 3-km horizontal resolution with enhanced vertical resolution to assimilate all available data including all measurements collected during the field campaign.
7. **ICF International (ICF):** oversaw the economic analyses to assess benefits resulting from improved short-term wind energy forecasting.
8. **National Renewable Energy Laboratory (NREL):** developed metrics and assisted ICF on the economic analysis.

2.3 User Group

A main goal of WFIP was to create a manner in which the successful project components can be replicated and transferred to others. A non-federally funded portion of WFIP included a User Group composed of utilities and ISO's engaged in or contemplating using wind forecasting (operations and/or research) and acknowledged an interest in reviewing the progress, findings and 'lessons learned' from this project. Although the WFIP field campaign and forecasting work focused on a geographical area within the ERCOT domain, the results here have implications for all utilities and TSOs interested in short-term (and ramp) forecasting. Through this review and outreach process and the forecasting programs of User Group members, the advancement of the country's community of practice directly benefited.

Users Group members included the Hawaiian Electric Company, Southern California Edison, PJM Interconnection, the New York ISO, and the New England ISO. Other observers and participants in the workshops included representatives from NOAA, DOE, ERCOT, and other federal, state, and local authorities.

3 Measurement campaign

One of the main objectives of WFIP was to investigate how short-term (0 to 6 hours) wind forecasts may be improved by enhancing existing observation networks. A key component to meeting this goal was determining what kind of additional measurements would be most effective in contributing to model forecast improvement and where those instruments should be deployed most efficiently. The following subsections discuss why the ERCOT domain was selected, what phenomena affect the performance of short-term wind forecasts, how additional instrument sites were identified, and how they functioned during deployment.

3.1 Selection of study area

The WFIP Southern Study ROI was selected because of:

1. The ROI already contained high quality multiple surface-based and remote sensing networks (Figure 3-6);
2. An excellent road network and the number of favorable sites (i.e. local airports) facilitated easy access for instrument deployment, operations, and maintenance;
3. The ERCOT territory, wholly within the ROI, contained the largest number of operating wind farms in the US, with an aggregate installed wind generating capacity (at the time of the field campaign) exceeding 10,000 MW, representing about 15% of system-wide peak demand;
4. AWST had already developed the ERCOT Large Ramp Alert System (ELRAS), a ramp forecasting system; and
5. The ROI experiences a wide variety of atmospheric phenomena (LLJs, frontal systems, convective outflow boundaries; see section 3.2 below) relevant to short-term wind power forecasting and common to many wind-rich areas of the country. Thus, lessons learned here should be applicable in many regions experiencing similar phenomena.

3.2 Phenomena affecting short-term wind forecasts

Accurate wind energy forecasting requires identifying and predicting a spectrum of meteorological phenomena with varying temporal and spatial characteristics. On short-term time scales (0 - 6 hours), these phenomena can be difficult to forecast, resulting in significant errors because of inaccuracies in the timing, magnitude, and placement of wind speed/wind direction discontinuities. More significantly, they can produce undesirable effects on the power grid, especially those associated with ramp events. For the purposes of the WFIP SSA, a ramp event is defined as an aggregate wind power generation excursion (that is, up or down) defined by the thresholds set forth for the operational ELRAS (see Zack 2011) given in Table 3-1.

Table 3-1. Ramp (MW) thresholds for 15-, 60-, and 180-minute periods for WFIP wind farm regions and the WFIP aggregate.

Ramp period (minutes)	Ramp type	East Texas	McCa mey_	North Texas	San Angelo	Sweet water East	Sweet water North	Sweet water West	WFIP Agg.
15	High	200	320	80	60	840	160	280	900
	Medium	150	240	60	45	600	120	210	700
	Low	100	160	40	30	360	80	140	450
60	High	400	640	160	115	1440	320	560	1850
	Medium	350	480	120	85	1080	240	420	1400
	Low	200	320	80	60	720	160	280	900
180	High	560	720	180	130	2880	360	1120	3700
	Medium	480	640	150	110	2160	300	840	2800
	Low	320	480	100	75	1440	200	560	1850

Ramp events can occur at any time of year within and surrounding the WFIP region. Freedman et al. (2008) identified several meteorological features responsible for up or down ramp events occurring within and adjacent to the ERCOT domain. The instrument deployment scheme was designed to capture with sufficient resolution the horizontal and vertical characteristics of the ramp-producing features. A brief discussion of the relevant phenomena follows.

Frontal system/trough/dry line. These are density fronts or air mass discontinuities usually associated with synoptic scale (hundreds to thousands of km) features. The largest ramp event during the WFIP campaign resulted from a frontal system moving across the region on 7 – 8 September 2012. As depicted in Figure 3-1, this system caused a sudden increase in wind power production (from 17% to over 80% capacity factor in less than 95 minutes) from WFIP WGRs in the ERCOT domain

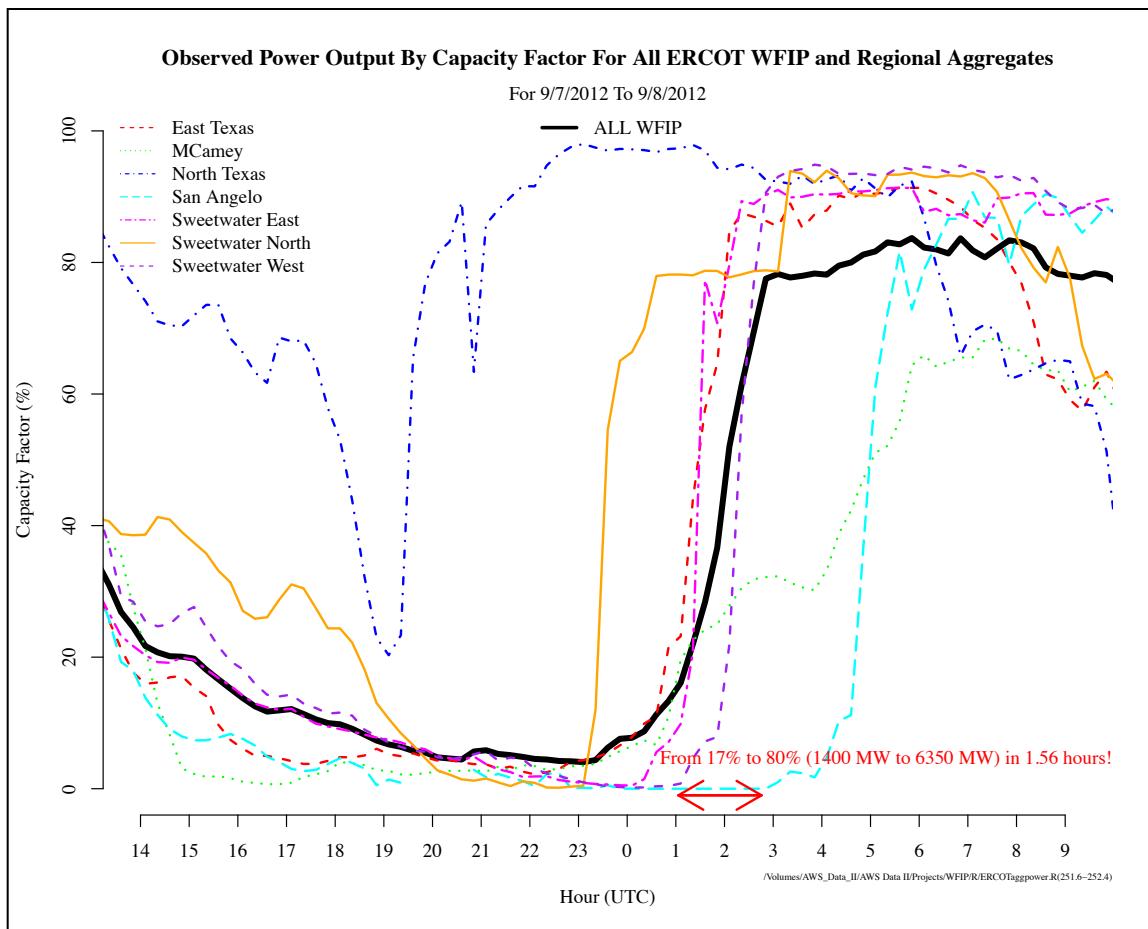


Figure 3-1. Observed wind power production (expressed by capacity factor in %) for regional wind farm aggregates (varying colors and line types) and the entire aggregate of WFIP wind farms (solid black line) during a large up-ramp event on 7 – 8 September 2012.

Thunderstorms and convection-induced outflow boundaries or gust fronts. These features occur on the mesoscale (tens to hundreds of square kilometers) and can move in any direction and at speeds in excess of 25 m s^{-1} (Figure 3-2). The frequency of these events varies considerably from year to year within the ERCOT domain (see Freedman et al. 2008). Outflow boundaries usually propagate radially outward from thunderstorm clusters (or other mesoscale convective systems). Although gust fronts often lose strength rather quickly, they can initiate additional convection and subsequent gust fronts. Since individual convective elements and clusters of storms are small-scale phenomena, short-term evolution of their temporal and

spatial characteristics is difficult to forecast. Thus, a sufficiently dense network of observations and incorporation of high frequency data is necessary for more accurate forecasts of these features, especially in the 0 to 3 hour time scale characteristic of outflow boundaries and gust fronts.

Low-level Jets (LLJs). This phenomenon occurs regularly throughout the year in the southern Great Plains. Southerly LLJs tend to be strongest (wind speeds can reach in excess of 25 m s^{-1}) but northerly LLJs do occur (Song et al. 2005). Different classification schemes for LLJs have evolved over time (e.g. Bonner 1968) but two types commonly occur over the ROI: (1) the nocturnal LLJ, caused by radiative cooling after sunset, and (2) a pre-frontal LLJ caused by an increasing pressure gradient ahead of a cold front. (Both types were frequently observed during WFIP.) The height of the LLJ varies between about 50 m and 400 m, but typically occurs at about 200 m (Banta et al. 2002). A special concern introduced by LLJs is the large vertical shears (upwards of 15 m s^{-1} per 100 m) that can occur across the turbine rotor plane (see Figure 3-4).

Critical observational and forecasting issues concerning LLJs are 1) their variation in height, 2) the magnitude of the vertical wind speed gradient, 3) their formation and persistence, 4) spatial characteristics such as width and depth, and 5) intermittent turbulence leading to propagation of strong winds towards the surface. One observational issue is that the LLJ structure can frequently extend below 100 m (Figure 3-3). To ensure sufficient vertical resolution of the full profile of the LLJ, the field measurement campaign (see section 3.3) included the deployment of several integrated observation sites (IOSs) featuring the co-location of a surface meteorological station, SoDAR, and wind profiling radar.

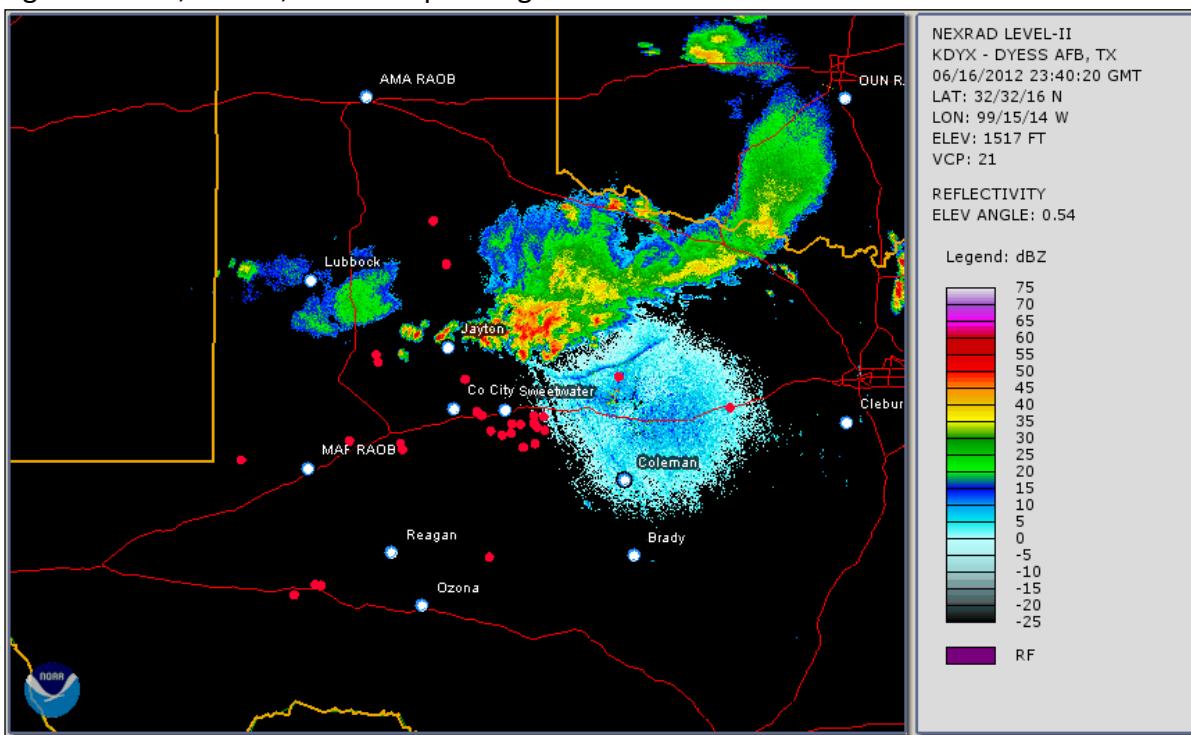


Figure 3-2. Radar imagery of a thunderstorm complex near Sweetwater aggregates on 17 June 2012 2340 UTC. Markers depicting points of interest include: instrument sites (light blue), and ERCOT wind farms (red). Note the line of convection-induced outflow to the northeast of Sweetwater.

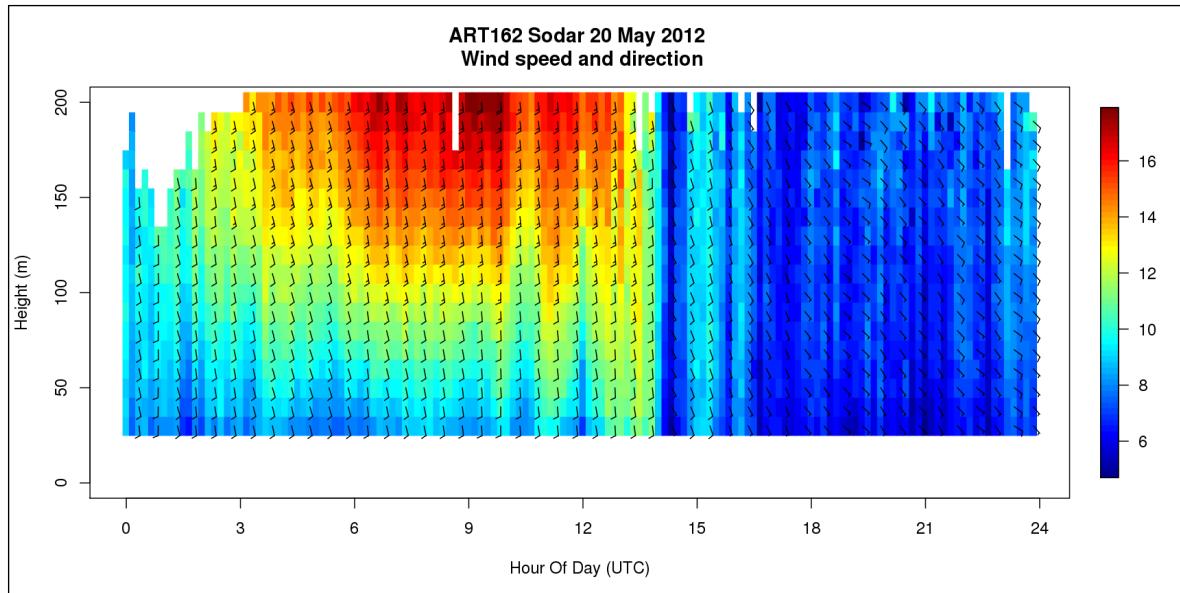


Figure 3-3. Time-height cross-section of 10-minute horizontal wind speed (ms^{-1}) from the AWST SoDAR unit at Reagan on 5 May 2012. Note the low-level jet feature (warmer colors--higher wind speeds) 03 UTC - 12 UTC (9PM - 6AM CST).

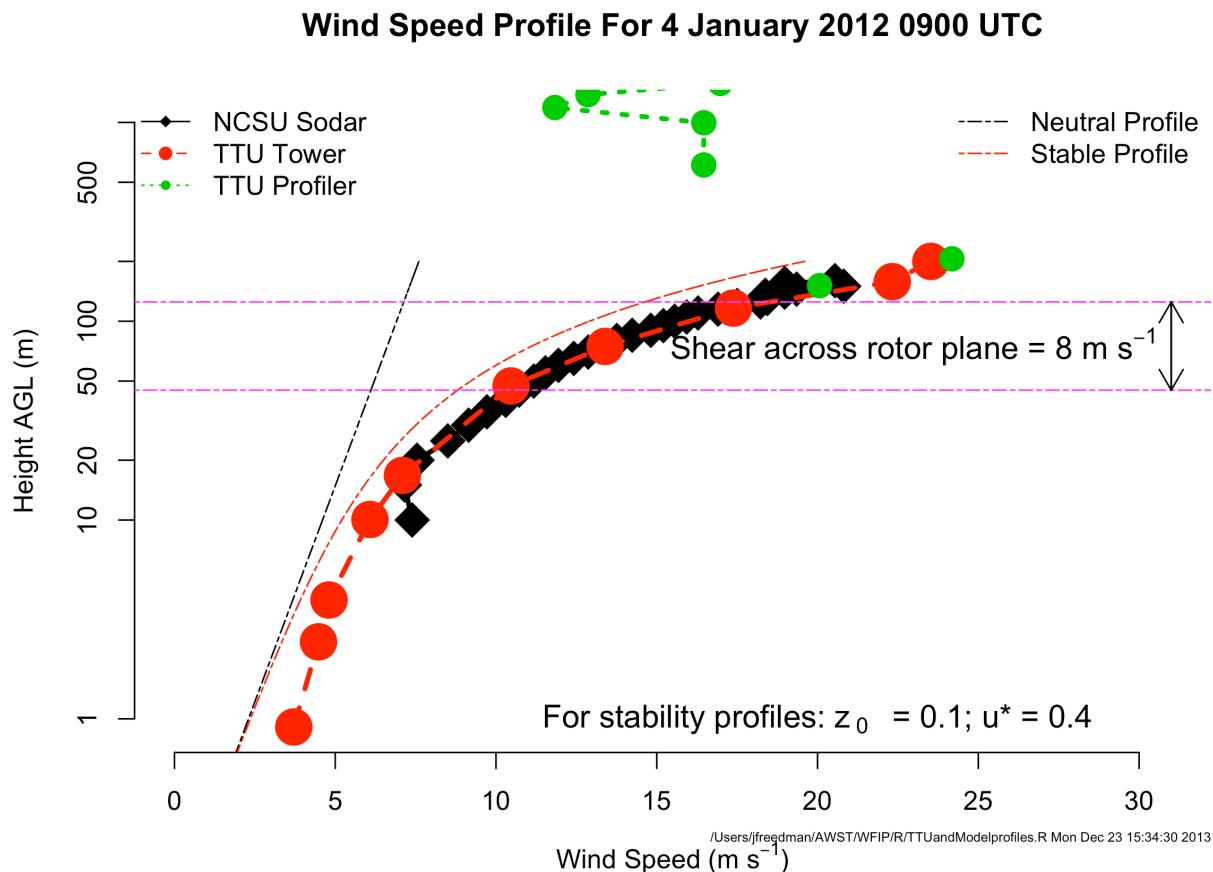


Figure 3-4. Wind speeds for the TTU 200-m tower (large red circles), the NCSU SoDAR (large black diamonds), and the TTU 915 MHz wind profiler (green circles) for 4 January 2012 0400 UTC. Dot-dash lines represent neutral (black) and stable (red) profiles for $z_0 = 0.1$ and $u^* = 0.4$. Y-axis is logarithmic.

Qualitative analysis indicates that the LLJ is a regular, periodic, and dominant feature that frequently drives capacity factors to over 60% (and therefore a large fraction of power production) during the nocturnal hours (see Figure 3-5). In essence, the plethora of wind farms throughout the Great Plains is a product of the ubiquity of the LLJ.

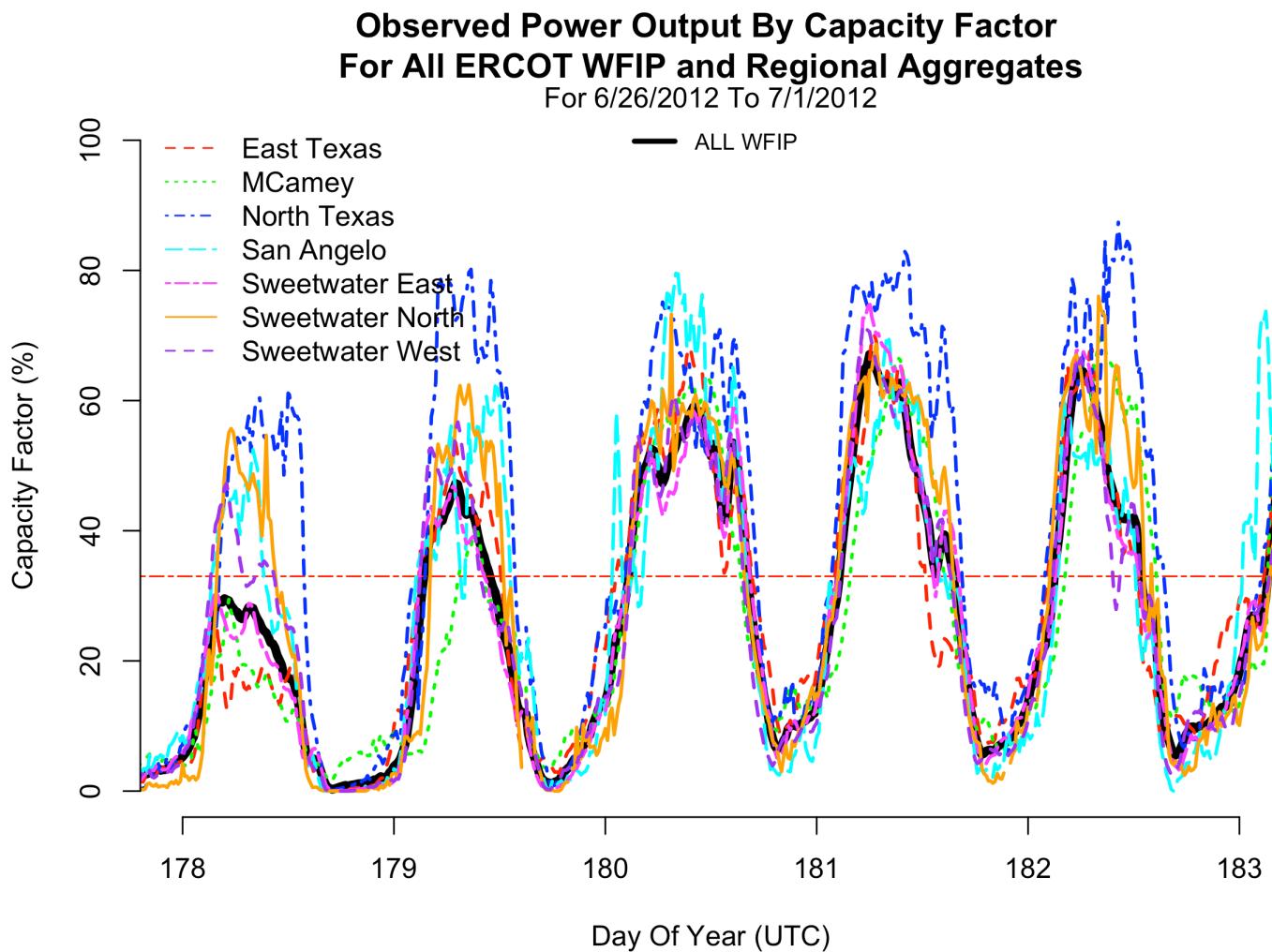


Figure 3-5. Observed power output by capacity factor (%) for regional wind farm aggregates and all WFIP for days 178 -183 (26 June - 1 July 2012. Horizontal red dash-dot line depicts average capacity factor (35%) for the period.

3.3 Field campaign – instrument deployment

A key goal in meeting WFIP objectives was to determine where additional measurements would be most effective in facilitating model forecast improvement. A subset of the WFIP program surface and remote sensing instrument suite was made available to the AWST team (see section 3.3.4) complementing those contributed by AWST, TTU, and NCSU.

3.3.1 Existing networks

As mentioned in section 3.1, a major reason for choosing the ROI is the existing high quality surface-based and remote sensing networks already in place. As depicted in Figure 3-6, surface and tower observation networks include the Automated Surface Observation System (ASOS) and Automated Weather Observation System (AWOS) stations operated by the National Weather Service (NWS) and Federal Aviation Administration (FAA), the West Texas A&M University Alternative Energy Institute data base of public towers, the West Texas Mesonet (Schroeder et al. 2005), the Oklahoma Mesonet, and Remote Automated Weather Stations (RAWS), part of NOAA's Real-time Observation and Analysis Network. Together, these networks include hundreds of stations throughout the region of interest reporting surface and near-surface wind and weather data on time scales of one minute to one hour. Much of, but not all, of this data are available through NOAA's Meteorological Assimilation Data Ingest System (MADIS) and were assimilated into the WFIP forecasting systems (see section 4). Most of the surface observation wind measurements were made at the standard 10 m observation height; tall tower wind observations ranged between 40 and 80 m. Thus, the additional observations made available through the DOE/ NOAA collaboration were necessary to capture the wind profile through typical rotor plane depths: 40 – 140 m AGL.

In addition to the surface networks discussed above, there are over a dozen operational wind profilers within and around the ROI (Figure 3-6). These include facilities operated by NOAA, and other entities such as the Texas Commission on Environmental Quality, and DOE's Atmospheric Radiation Measurement (ARM) Program. The NWS also operates several Next-Generation Radar (NEXRAD) 88D Doppler Radar (NEXRAD) sites throughout the ROI and launches twice-daily (or more often, depending upon

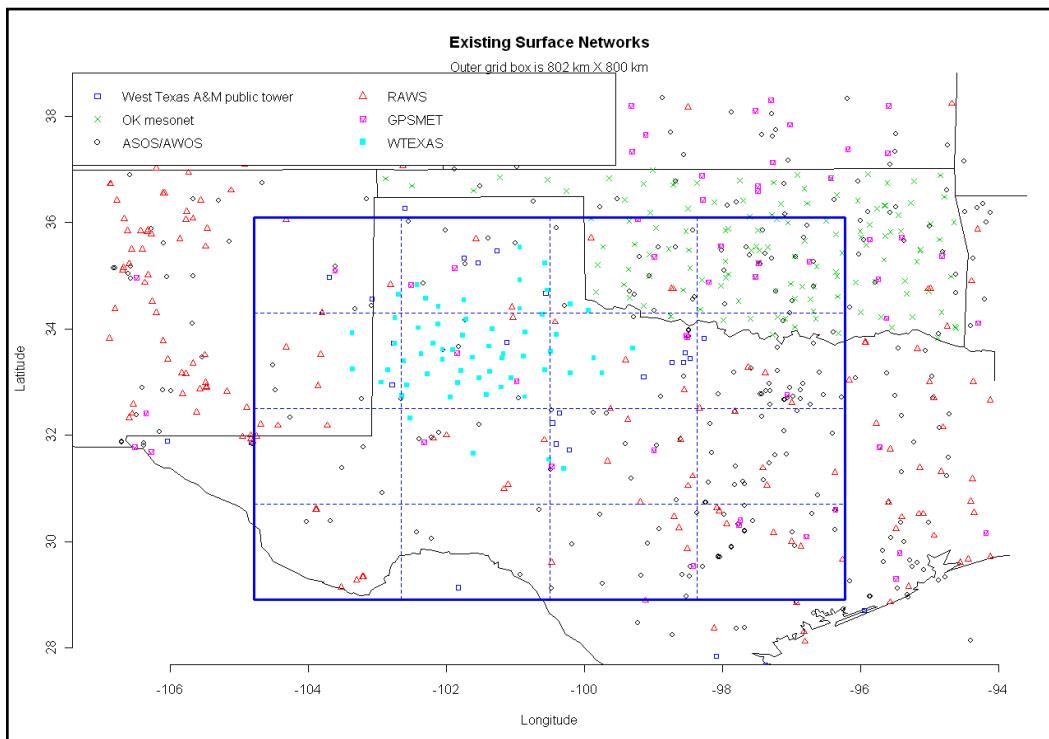


Figure 3-6. Existing surface observation networks within and adjacent to the region of interest (outlined in thick blue line). Dashed lines represent approximately 200 km grid squares.

weather conditions) rawinsondes from several sites usually co-located with the NEXRAD 88D or NWS Weather Forecast Offices within the ROI. Note that the NEXRAD radars and wind profilers only provide wind information at the lowest one or two range gates within the rotor plane.

Finally, TTU operates a 915-MHz wind profiler and 200-m tower at the Reese Technology Center, about 15 km west of Lubbock (see Figure 3-8 and Table 3-2). The tower has 10 levels including measurements of the u, v, and w wind components, humidity, and temperature. This facility, upgraded with funding from WFIP, was critical in providing detailed and continuous information regarding the wind and temperature profile within the lower part of the boundary layer (BL), particularly the stable BL, and served as a verification site for model validation and NCSU data gap filing efforts.

3.3.2 Observation targeting

AWST/MESO has developed two objective methods to predict error reduction in a wind speed forecast due to the deployment of a given sensor at a specific location. The first method employs an ensemble of NWP models to determine regions of high forecast sensitivity where observational data for selected variables are likely to improve the forecast of wind speed for the target location (Zack et al. 2010). A second, less computationally expensive method uses a time-lagged analysis from a single NWP model. The second approach was employed in this project owing to computing and project timeline constraints.

Output from a 13-month series of NWP simulations was used to compute the simulated time-lagged correlations between the prior change in a measurable variable at a potential observation site and the future change in wind speed at a forecast site. By repeating this calculation for every model grid point, correlation maps were constructed for different seasons, weather regimes, forecast look-ahead times, and observed variables. Only correlations that exceeded 0.25 with a confidence of greater than 99% were considered. This confidence was assessed by assuming the range of possible errors in the correlation followed a normal distribution with the standard deviation as a function of effective sample size. The effective sample size was assumed to be a function of the total sample size and the time-lagged autocorrelations of both the predictor and forecast variables.

The next step was to estimate the reduction in forecast root mean square error (RMSE) or sensitivity for a given location to each model grid point. This process required computing a weighted sum of the variance (correlation squared) for the observed variable with the highest time-lagged correlation. The weighted sum was computed over all sites, look-ahead times, seasons, and regimes. Error reduction was then determined for representative regime classification schemes (i.e. all times, four time-of-day bins, and four initial wind direction bins). Finally, the predicted forecast error reduction was mapped for individual forecast sites, 1- to 6-hour look-ahead times, and seasons. Error reduction was also computed for a combination of locations that could serve as prospective instrument locations.

Error reduction computed in this manner is effectively the predicted reduction in wind speed root-mean-square error (RMSE) assuming that none of the information from the new sensors is already available in the current state-of-the-art forecast. The method does not account for reductions in benefit when multiple sensors are installed due to correlations between the sensors. It also does not account for additional benefits from multiple variables measured by the same instrument – only the most highly correlated variable measured by the sensor is considered for a given site, season, look-ahead time or regime. Finally, it does not

account for or resolve small-scale, short-lived features that are not adequately represented by the NWP simulations. These may include convective elements such as thunderstorms, large eddies and circulations induced by local variability in cloudiness, soil moisture and other factors. This in turn would lead to a reduction in forecast benefit for deployed sensors (see e.g. section **Error! Reference source not found.**).

Recommended sensor locations were based on the overall predicted error reduction, predicted error reduction for specific look-ahead times and seasons, and location of existing sensors. Figure 3-7 shows a subset of the existing and recommended sensor locations. Background color shading depicts the mean predicted fractional wind speed RMSE reduction for four seasons, four forecast sites (two in the Sweetwater East region and one each in the Sweetwater West and Sweetwater North), and six look-ahead times (1- to 6-hours ahead) for a single proposed IOP site (co-located SoDAR, profiler, and surface station) sited at each model grid point location. The green and yellow shading shows potential sensor locations with the highest predicted error reductions. Note that both the recommended co-located instrument sites at 99.4W, 31.2N and 100.6W, 32.5N (orange and pink filled circles with black borders (Figure 3-7) are in regions where the predicted wind speed RMSE reduction is in the 12 to 16% range.

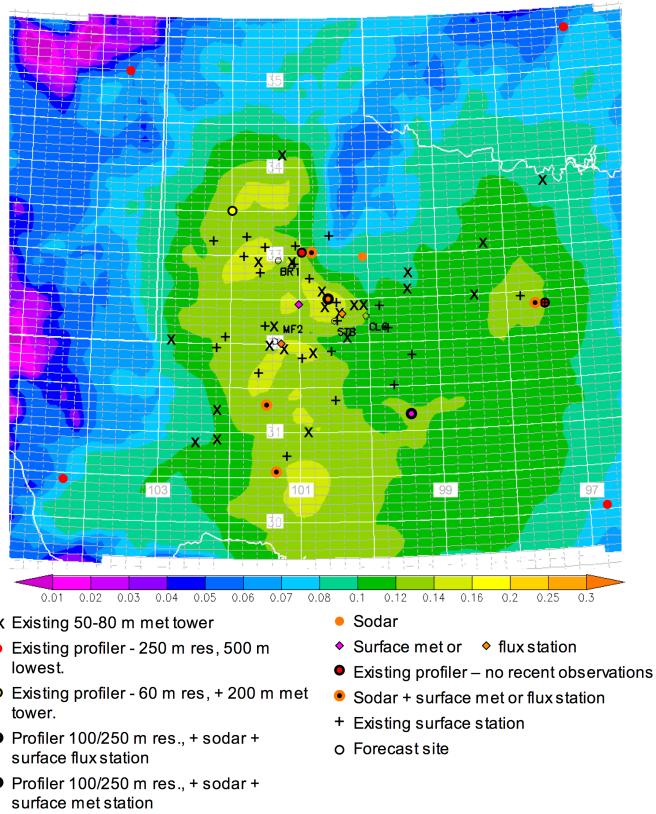


Figure 3-7. The mean predicted fractional wind speed RMSE reduction (color shading) for all seasons, forecast sites, and look-ahead times for a site with a collocated profiler, SoDAR, and flux station. Existing and recommended sensor locations area also shown. Recommended sensor sites were chosen based on the overall predicted error reduction, predicted error reduction for specific look-ahead times and seasons, and location of existing sensors. Only a representative sample of existing met towers and surface stations is shown. Background color contours depict the mean predicted fractional wind speed RMSE reduction for all seasons, forecast sites, and look-ahead times for a site with a collocated profiler, SoDAR, and flux station.

3.3.3 Proposed deployment

Most sensible weather and associated phenomena described in Section 3.2 move into and through the ROI from the south, west, and north (e.g. Bonner 1968; Chang et al. 2002; Song et al. 2005). Specifically, the LLJ tends to be oriented from the south or southwest (Song et al. 2005); convection (and hence most, but not all, outflow boundaries) tends to move from the northwest, west, or southwest, depending upon local and regional dynamics; and frontal systems tend to move from the northern, western, and less frequently, southern sectors, as guided by larger scale circulation patterns. The targeted observation results largely conformed to this climatology, and the proposed field deployment scheme (while not duplicating observations from existing networks) was weighted in these direction sectors, to afford model assimilation schemes the greatest probability of capturing these features as they develop and move through the ROI.

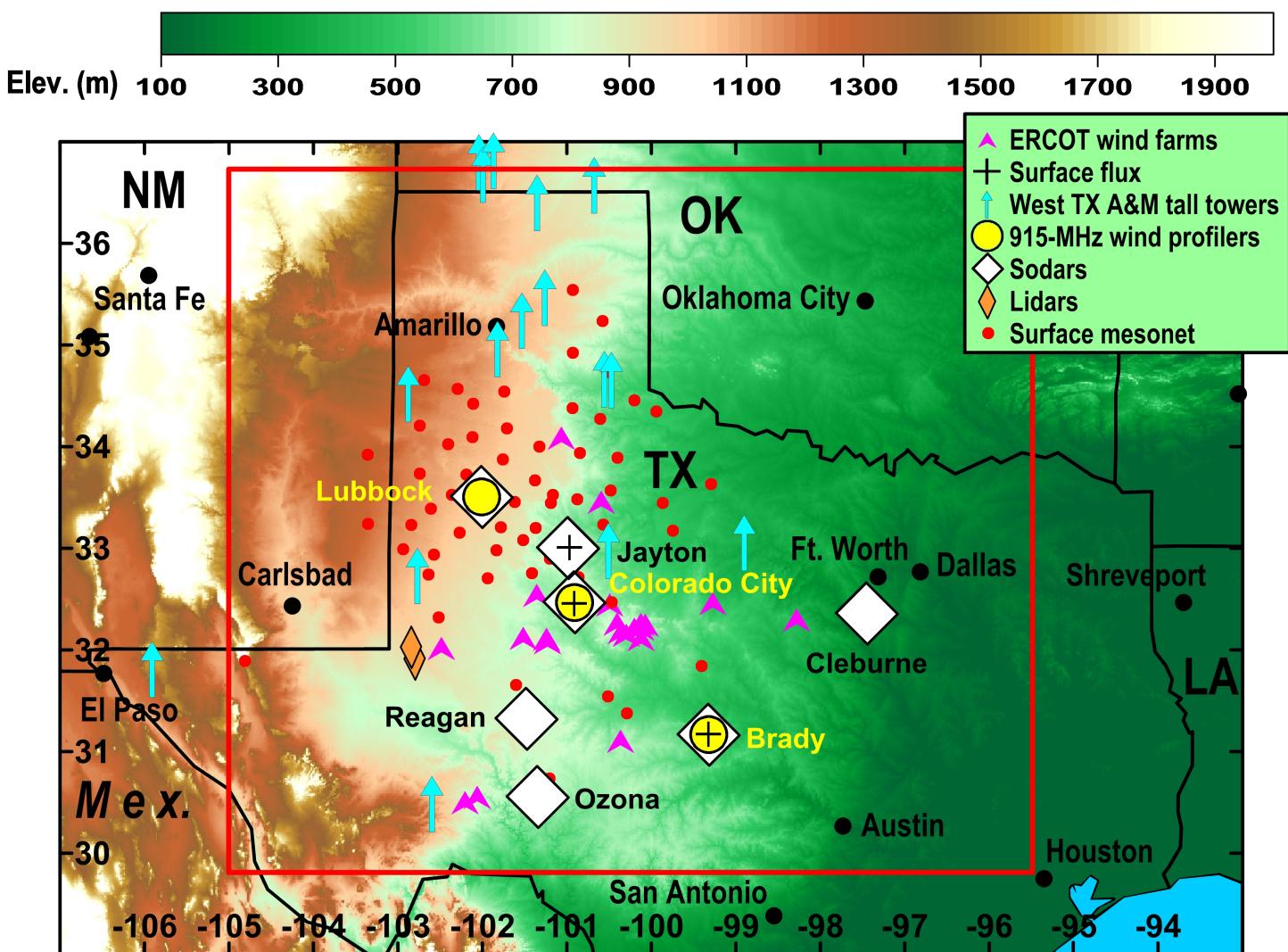


Figure 3-8. NOAA/PNNL/AWST WFIP deployment with existing profiler, NEXRAD, and upper air (rawinsonde) networks. IOSS suites are referenced in Table 3-2. Solid red line (outlined rectangle) shows the ERCOT ROI and center of the model domain. Figure courtesy of J. Wilczak, NOAA ESRL.

A key component of the field deployment was the combination of surface and remote sensing platforms as integrated observation sites (IOSs), to ensure the complete capture of the wind profile from the surface through and beyond the top of the atmospheric boundary layer (typically 1 – 3 km in this region). Nonetheless, limitations of the instrument suite available to the WFIP SSA necessitated a considerable scaling back of the eventual deployment scheme ([Figure 3-8](#)).

One attribute of the SSA was the geographically uneven distribution of wind farms in the ERCOT territory (see Figure 3-9). Approximately 40% of the WGR capacity (over 4000 MW) is clustered in Sweetwater region of north central Texas, making wind power production susceptible to large ramp events being initiated by relatively small-scale features (e.g., convective outflow). Although the targeted observation studies (see section 3.3.2) were designed to observe these events, the limits of their predictability, data assimilation constraints, and the scaled back instrument availability for the SSA likely produced the mixed results in the data denial experiments and the ability to successfully forecast the phase and amplitude of ramp events from these features (see sections **Error! Reference source not found.**, 5.2, and 5.3).

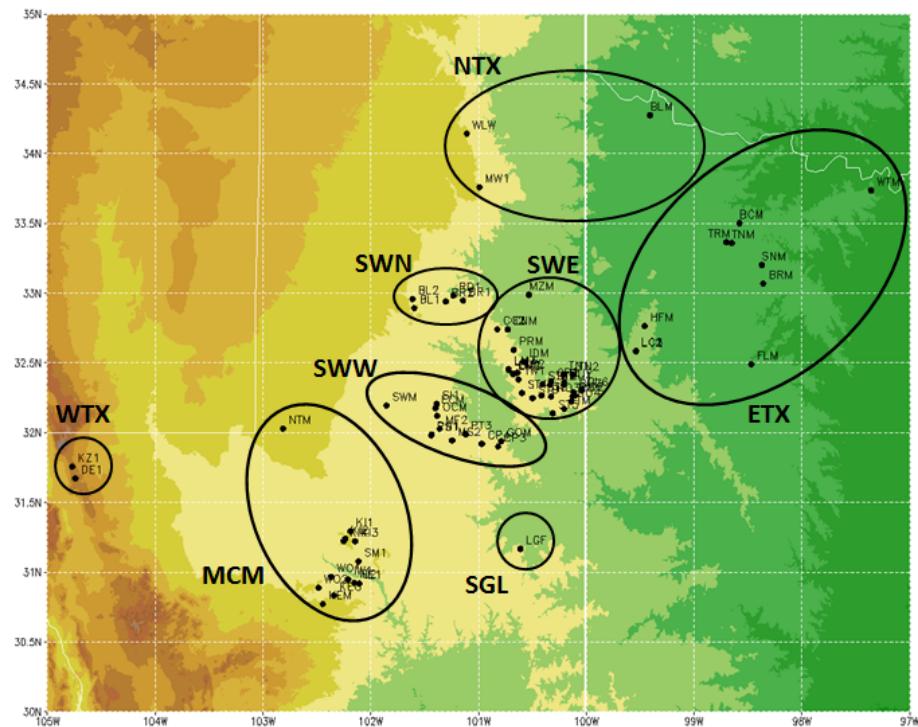


Figure 3-9. Regional aggregates of wind farms in the WFIP SSA (ERCOT domain). Aggregates are defined as NTX: North Texas; SWN: Sweetwater North; SWE: Sweetwater East; SWW: Sweetwater West; ETX: East Texas; WTX: West Texas; MCM: McAmey; SGL: San Angelo.

3.3.4 Deployment of NOAA, DOE, AWST, and partner equipment

The following paragraphs describe the instrument platforms and observation suites that were used in the field measurement campaign.

DOE/NOAA instrument suite. NOAA, DOE, and the Pacific Northwest National Laboratory (PNNL) made available for the WFIP South Team two 915-MHz Wind Profiling Radars (WPRs), three SoDARs, three surface flux stations, and three surface meteorological stations (see [Table 3-2](#)). The three surface stations provided by the NOAA/DOE deployment package measure standard meteorological variables such as wind speed and DE- EE0004420

direction, temperature, dew point, and surface pressure. The three flux stations use eddy covariance systems that measure the turbulence quantities u' , v' , w' , and T' , and hence calculated the surface heat and momentum fluxes.

TTU – Reese Technology Center. TTU contributed data collected from several measurement platforms during the field campaign phase of the project. These data were collected from a 200-m meteorological tower and a boundary layer radar profiler, both of which are located at the Reese Technology Center 15 km west of Lubbock, Texas.

The 200-m meteorological tower (Figure 3-10) includes 10 levels of instrumentation and provided wind speed, direction, temperature, relative humidity, and barometric pressure sampled at 50 Hz 24-hours per day. Boom arms containing the instrumentation are mounted at 1.0, 2.4, 3.9, 10.0, 16.8, 47.3, 74.7, 116.5, 158.2, and 200 meters.

For the WFIP experiment, the 50 Hz data was averaged to produce 30-minute summary files for the duration of the WFIP project.. Additional post-processed 30-minute temperature summaries were constructed by request for the period January 20, 2012 – September 25, 2012 for all tower levels.

Upgrades were made to the 200 m tower and processing applications to support real-time transmission of quality-controlled data, and additional web plots and statistics were developed to ease daily monitoring of the wind observations at each level. Outside of outages due to tower maintenance and upgrades in January 2012, data recovery from the tower was above 95%. On a few isolated occasions the microwave link between the field site and the campus network failed, but these cases were resolved quickly and nominally impacted real-time data transmission for less than a day. While this occurred, data was continually collected at the tower without interruption; files were processed and uploaded to the web once the microwave link was fixed.

A Vaisala LAP-3000 Boundary Layer Profiler (Figure 3-10) located at the Reese Technology Center was also made available for this project. The profiler provides vertical profiles of wind speed, direction, and virtual temperature data constructed over 20 minute intervals. Between 160 – 1600 m AGL, profile bin spacing is 60 m and between 1600 – 6000 m, profile bin spacing is 160 m. Funding from the WFIP project contributed to a major system upgrade including replacement of the power supply, upgrade of the modulator interface unit, repair of the amplifier, replacement of the main system server, and an update of the system software.

Table 3-2. WFIP instrument deployment (July 2011 – September 2012). Shaded rows indicate grouping by IOS.

Station	Station ID	Instrument	Operator	Latitude (N)	Longitude (W)	Elevation (m)
Brady	BDYTX	Profiler915	ARL	31.175	99.326	548
Brady	BDYFLUX	Flux	ARL	31.175	99.326	548
Brady	BD2TX	SoDAR	ANL	31.175	99.326	548
Brady	BDY	SfcMet	ARL	31.175	99.326	548
Colorado City	COCTX	Profiler915	ARL	32.472	100.921	673
Colorado City	COCFLUX	Flux	ARL	32.472	100.921	673
Colorado City	CC2TX	SoDAR	ARL	32.472	100.921	673
Colorado City	COC	SfcMet	PNNL	32.472	100.921	672
Jayton	JTNT2	Profiler404	NOAA	33.017	100.978	707
Jayton	JT2TX	SoDAR	ARL	33.017	100.978	707
Jayton	JTNFLUX	Flux	ARL	33.017	100.978	707
Jayton	JTN	SfcMet	PNNL	33.017	100.978	707
Cleburne	KCPT	AWOS	NA	32.354	97.435	258
Cleburne	CLETX	Profiler915	TCAQ	32.354	97.435	256
Cleburne	CPTTX	SoDAR	AWST	32.354	97.435	258
Cleburne	CPT	SfcMet	PNNL	32.354	97.435	259
Lubbock	RESTX	SoDAR	NCSU	33.611	102.051	1017
Lubbock	TTU Tower	200m tower	NA	33.611	102.051	1010
Lubbock	LBKTX	Profiler915	NA	33.611	102.051	1010
Reagan	RGNTX	SoDAR	AWST	31.199	101.473	793
Reagan	BIGL	SfcMet	TTU	31.199	101.473	823
Ozona	KOZA	AWOS	NA	30.736	101.203	726
Ozona	OZA	SfcMet	PNNL	30.736	101.203	721
Ozona	OZATX	SoDAR	AWST	30.736	101.203	720
Sweetwater	KSWW	AWOS	NA	32.469	100.467	725
Sweetwater	SWW	SfcMet	PNNL	32.469	100.467	717
Coleman	KCOM	AWOS	NA	31.841	99.406	517
Coleman	COM	SfcMet	PNNL	31.841	99.406	514

NCSU. The NCSU group installed and operated a Scintec Flat Array SoDAR (SFAS) at the Reese Technology Center as part of the field deployment. The SoDAR was collocated with the TTU 200 m tall meteorological tower and wind profiler. It was installed 2.9 m above ground level (AGL; see

Figure 3-11). The SFAS SoDAR has a flat array acoustic antenna with 64 piezoelectric transducers and emits beams in 9 directions in up to 10 frequencies operating in the frequency range of 2525 to 4850 Hz.

AWST deployed three ART model VT-1 SoDARs with a vertical range of 30 - 200 m with 10 m resolution. These SoDARs were located at local or municipal airports at Cleburne, Ozona, and Reagan TX (see Figure 3-12). The SoDARs were collocated with PNNL surface met stations at Reagan, Ozona (where an existing ASOS station was operating), and Cleburne (where a profiler and AWOS station were also located).



Figure 3-10 (left): Texas Tech 200-m tall tower at Reese Technology Center; (upper right): lowest three levels of instrumentation on the 200 m tower showing the various instrumentation on each boom arm; (lower right): TTU Vaisala LAP-3000 Boundary Layer Profiler



Figure 3-11 (left): the deployment locations of NCSU SoDAR, TTU 200 m meteorological tower, and TTU wind profiler; (right): an image of the NCSU SoDAR (red box). The 200 m tall tower is visible in the background.



Figure 3-12 (top left): AWST ART-1 SoDAR commissioning at Reagan TX; (top right) SoDAR commissioning at Ozona TX;(bottom left): PNNL surface met station at Cleburne TX with AWST ART-1 SoDAR in background; (bottom right): NOAA flux station at Colorado City TX.

3.3.5 LiDAR deployments

During the latter part of the WFIP observation and forecast campaign, an opportunity arose to deploy two Leosphere LiDARs for a two-month period at a participating wind farm in west Texas. The purpose of this deployment was to ascertain the value of on-site LiDARs to very short-term forecasting (0 – 1 hour) and to gather data for possible wake modeling studies. In late August 2012, a Windcube 7 and Windcube 8 were deployed downwind and upwind of a wind farm in west Texas (see Figure 3-13). However, because of a variety of technical issues (e.g. air temperature exceeding the operational constraints of the LiDARs and power supply interruptions), only a few weeks of viable data were collected. Data from simultaneous operation of both LiDARs was limited to just 2 – 3 weeks. Results did show distinct wakes, especially for the along row wind direction (i.e. ESE) and the presence of the nocturnal LLJ.

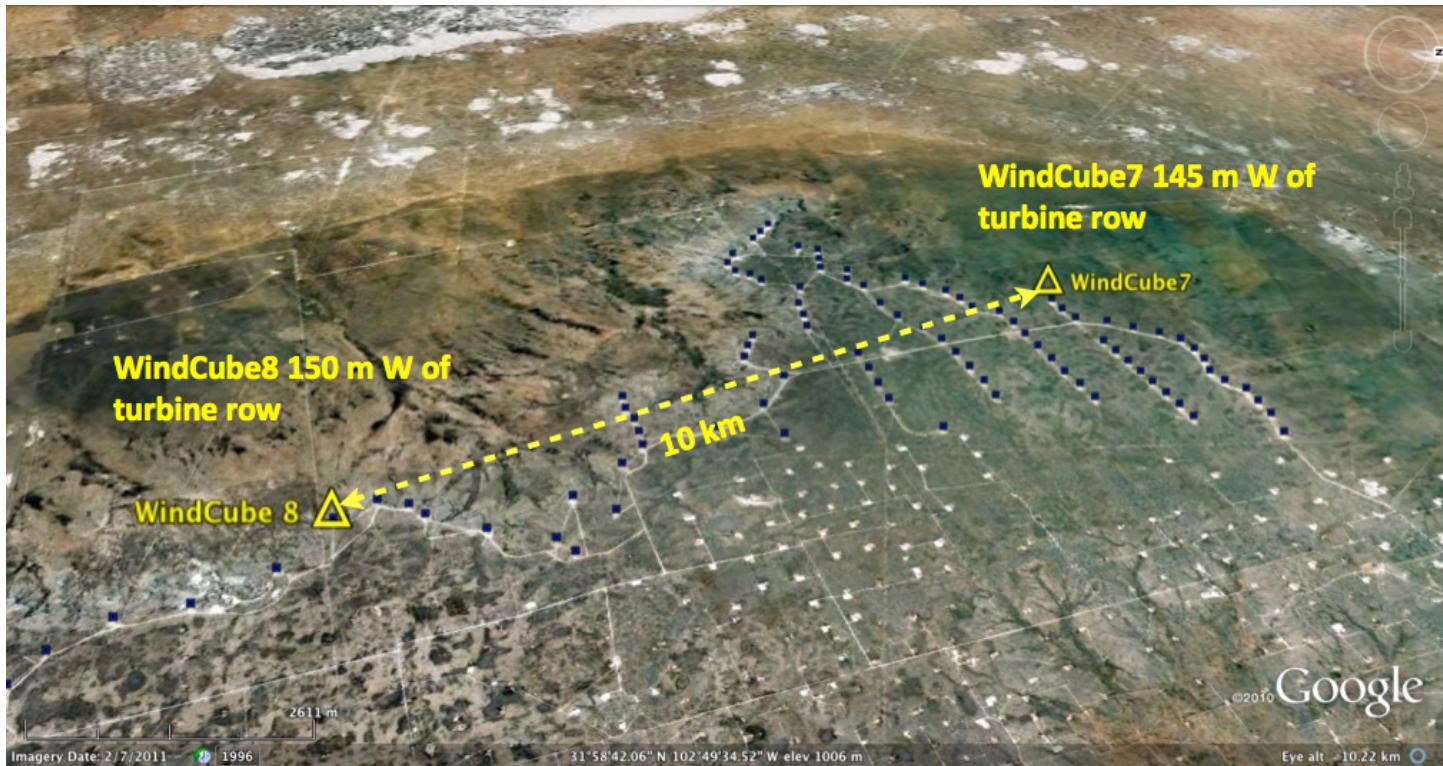


Figure 3-13. Deployment of Leosphere Windcube LiDARs at a wind farm in Texas.

3.4 Results and discussion

3.4.1 Data acquired

AWST SoDARs. The three AWST SoDARs (at Cleburne, Ozona, and Reagan TX; see Figure 3-8 and Table 3-2) operated, with the exception of regular maintenance (approximately every 3 months), without interruption for the entire field and modeling campaign. Commissioning of these SoDARs occurred in mid-July 2011, and continued operating through the end of the deployment period (13 September 2012). Data was processed on site and compiled into 10-minute averages of wind speed and direction, individual u, v, and w components, turbulence intensity (TI), turbulent kinetic energy (TKE), and inclined flow angle. All data was uploaded in real-time to MADIS.

The NCSU SoDAR data. The NCSU SoDAR was operational from 1 July 2011 to 12 September 2012. The SoDAR wind data were uploaded in real-time to an ftp server. All the Team members were given access to this secured server for data downloads. The data were also transferred to MADIS from the ftp server in an automated fashion. During WFIP, it provided 59 vertical layers of wind data (speed and direction) from 10 m up to a maximum of 300 m with a vertical resolution of 5 m and an averaging time of 10 minutes. These data were quality controlled in real-time by Scintec's APRun software.

TTU staff was responsible for the continuous operation of the NCSU SoDAR. With the exception of a few power failures and communication interruptions, there was an almost continuous stream of SoDAR observations, resulting in only a few temporal data gaps for the entire field campaign. The availability of SoDAR data is shown in Figure 3-14. The raw SoDAR data were compared with the corresponding TTU tower-based data as described in Figure 3-14. Beyond the elevation of 100 m AGL, the SoDAR data availability decreased quite rapidly (below 50% above 140 m). This behavior is common among commercial-grade

SoDARs, especially in a dry environment such as the panhandle of Texas. Due to these data gaps, the mean wind speed above 100 m AGL is quite noisy and unreliable in comparison to the tower measurements. At an elevation of 78 m AGL (close to a typical wind turbine hub-height), the SoDAR data underestimate the mean wind speed significantly, and both Weibull distribution parameters are underestimated. Based on the wind rose plots (Figure 3-14), there is also a bias in the wind direction distributions of the SoDAR as compared with the tower data.

Table 3-3 Comparison of NCSU SoDAR-based and TTU tower-based wind data for the period 1 July 2011 – 12 September 2012.

Level	Altitude (m)	R^2	RMSE (ms ⁻¹)	Weibull Parameters				Number of data points
				NCSU SoDAR C (ms ⁻¹)	k	TTU Tower C (ms ⁻¹)	k	
5	16.76	0.763	1.27	5.47	2.19	6.05	2.55	11611
6	47.24	0.722	1.77	6.40	2.01	7.37	2.57	18274
7	74.67	0.751	1.77	7.63	2.19	8.33	2.46	17770
8	116.43	0.722	2.09	8.93	2.15	8.95	2.24	15405
9	158.19	0.713	2.23	9.33	2.36	8.39	2.16	5698
All data Points*		0.74	1.82					68758
*(Levels 5-9 (16.7 – 158.2 m altitude) data are used for calculating the Statistics)								

Note: In order to collocated with 30-minutes tower data and calculated the statistics, 10-minutes SFAS (SoDAR) data are averaged to 30 minutes and interpolated to tower altitude. However, in Figure 3-14, all raw data from 10-minutes sodar at 78 m and 30-minutes tower at 75 m are used to calculate the statistics.

TTU tower and profiler data. During the WFIP project, the 200-m TTU tower data operated continuously from 16 July 2011 – 13 September 2012. The 915 MHz profiler was operational from 1 December 2011 – 1 April 2012; and from 26 June 2012 – 13 September 2012. The 20-minute wind and temperature data were assimilated into separate daily text files that were uploaded in near real-time to MADIS for sharing. Additional web plots and statistics were developed to ease daily monitoring for the profiler data stream.

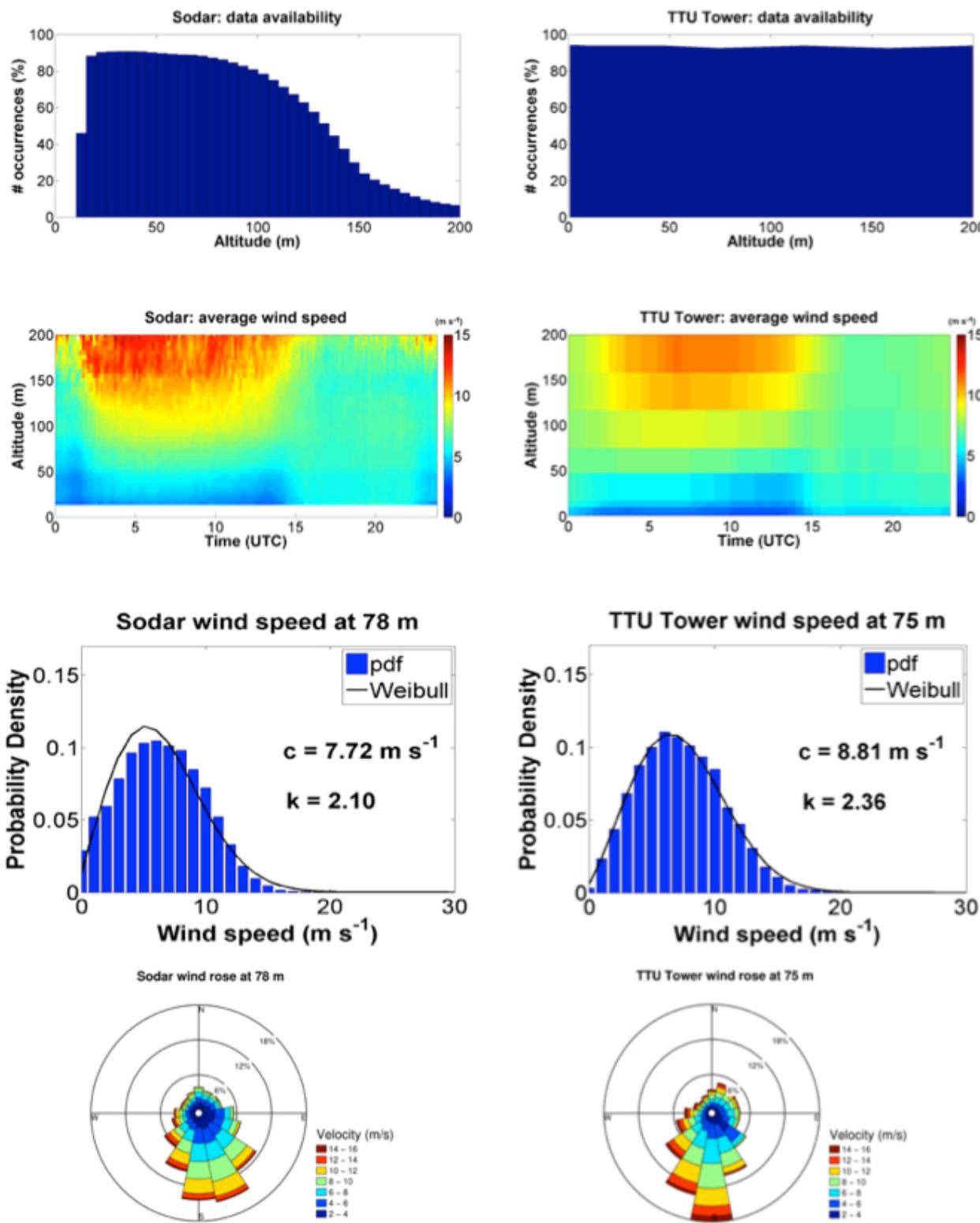


Figure 3-14. Comparison of NCSU SoDAR-based (left panel) and TTU tower-based (right panel) wind data. From top to bottom, the panels represent data availability, mean wind speed, histograms, and wind roses, respectively. Time frame: 1 July 2011 – 12 September 2012.

3.4.2 Ramp frequency

Ramp frequency was evaluated for events of varying duration, and the atmospheric phenomena responsible were identified through the seasonal pattern of ramp occurrence. For all regions of the ERCOT domain, longer-duration (60- and 180-minute; see Table 3-1) ramp events are more frequent during the cold months and into the spring transition (WFIP aggregate region shown in Figure 3-14). This pattern is seen in both up- and down-ramps.

These events are attributed to frontal passages, which are more frequent during the winter and spring because of the southerly location of the polar jet. There is a peak in the 15-minute ramp occurrence with the spring transition, most likely due to increased convective activity.

The atmospheric phenomena discussed above are associated with both up- and down-ramp events. A previous study (Freedman et al 2008) found that up ramps tend to be more common within

the ERCOT domain. This bias towards positive ramps arises because rapidly moving transient features such as frontal systems and convective outflow tend to produce abrupt increases in wind speed, with a more gradual slackening of the wind speed gradient after their passage (Freedman et al 2008). The present study found a roughly equal number of up- and down-ramp events for the WFIP aggregate region when considering all 15-, 60-, and 180-minute ramp periods (approximately 48% up-, 52% down-ramps). For all duration ramp events only of the medium and high threshold, approximately 61% were up-ramp and 39% were down-ramps. The annual proportion of threshold up- and down-ramp events observed in this study is consistent with Freedman et al (2008).

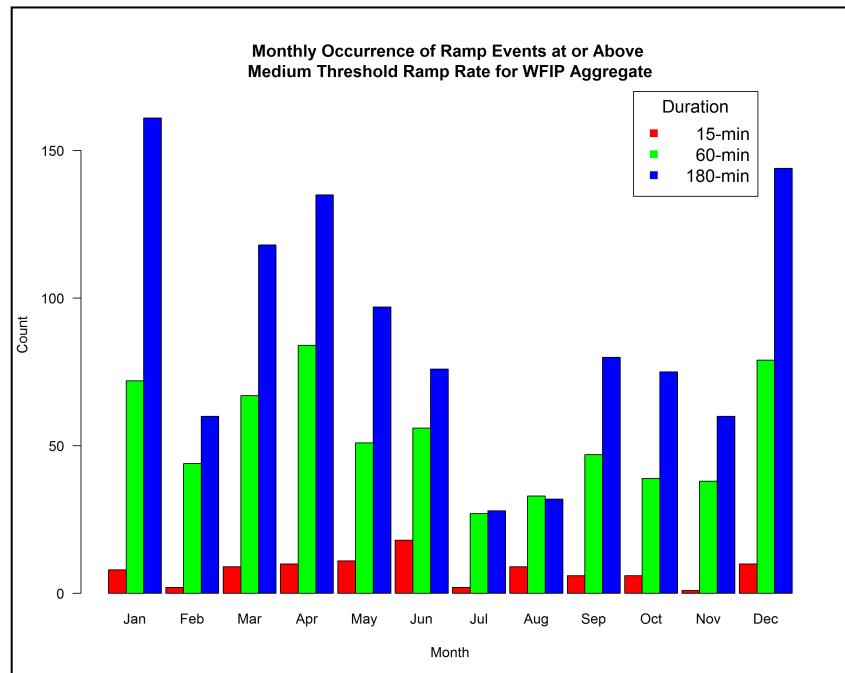


Figure 3-15. Monthly occurrence of ramp events in the WFIP aggregate region meeting or exceeding the medium ramp rate threshold, separated by duration of event: 180-minute (blue), 60-minute (green), and 15-minute (red).

3.5 Gap filling analysis and related work (NCSU)

3.5.1 Spectral Characterization

To further assess the quality of the SoDAR-based wind data, the NCSU team performed a rigorous spectral characterization.. Wind speed power spectra were calculated using three complementary estimation techniques: fast Fourier transform (FFT), Haar wavelet transform (WT), and auto-regressive moving average (ARMA) models. The FFT and WT techniques require continuous (i.e., without any missing observations) and uniformly spaced data series. Given the sporadic presence of missing observations in the SoDAR dataset, a (temporal) linear interpolation was applied to create a gap-filled series prior to the FFT-based and WT-based spectra calculations. However, this ad-hoc temporal gap filling was not very effective, as its adverse impact on

spectral characterization cannot be quantified in a reliable manner. Therefore, an alternative approach based on the ARMA models, which is insensitive to randomly missing data, was also utilized for wind spectra computations. This approach is available via an open-source code called “ARMAsel for Irregular or Missing Data”, or “ARMAsel-mis” (Broersen, 2006).

Qualitatively, all the estimation techniques lead to very similar wind spectral shapes (Figure 3-16). The FFT-based spectra show strong diurnal and semi diurnal variations, which is physically expected. However, quite surprisingly, the wind spectra from all the vertical levels (only two levels are shown in Figure 3-16) portray a $-5/3$ scaling regime in frequency range of $\sim 1.2 \times 10^{-5}$ Hz to $\sim 2.7 \times 10^{-4}$ Hz (called mesoscale range). In the time-domain, this scaling regime corresponds to ~ 1 - 24 h. Starting with the classic works by Naström and Gage (Naström and Gage 1985; Gage and Naström 1986), a $-5/3$ scaling in mesoscale wind spectra were reported several times based on observational data from the upper troposphere and lower stratosphere. However, the presence of a similar regime in boundary layer wind data has rarely been reported in the literature⁵. This highlights the uniqueness of the WFIP SoDAR datasets.

Note that at the high frequency end (beyond $\sim 3 \times 10^{-4}$ Hz) the spectra show a tendency to flatten. This tendency is more prominent in FFT- and ARMA-based spectra compared to the wavelet-based spectra. This plateau is likely related to ‘spectral gap’; however, the temporal resolution of the SoDAR data (10 min) is not fine enough to reach a definite conclusion.

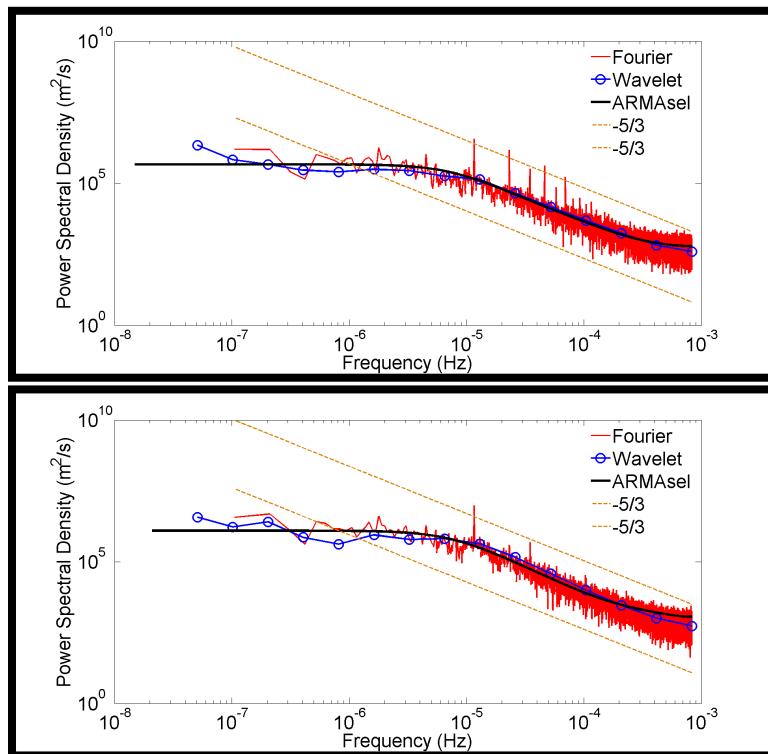


Figure 3-16. Power spectral density using FFT (red lines), WT (blue lines with circles), and ARMAsel (black lines) for SoDAR wind speed from 1 July 2011 to 12 September 2012. The top and bottom panels correspond to altitudes at 18m and 123m above ground level, respectively. The dashed orange lines represent a slope of $-5/3$.

⁵ Note that the mesoscale range represents lower frequencies (or wavenumbers) in comparison to the so-called inertial-range turbulence. The existence of a $-5/3$ scaling regime in the inertial-range is well documented in the boundary layer turbulence literature.

3.5.2 Gap-Filling and De-noising

For WFIP, a novel gap filling and de-noising algorithm was developed based on Vapnik-Chervonenkis Statistical Learning theory (Cherkassky et al. 1999; Cherkassky and Shao 2001; Cherkassky and Mulier 2007). The algorithm is fully automated and is schematically explained in Figure 3-17. The salient features of this algorithm are as follows:

1. The algorithm is applied to synthetically merge U and V wind components from the spatially and temporally collocated SoDAR and wind profiler;
2. Gap-filled and de-noised U and V wind components are calculated using Chebyshev polynomial based on the best polynomial order;
3. The best polynomial order is calculated using Vapnik-Chervonenkis (VC) theory (Cherkassky et al. 2007);
4. Unacceptable or erroneous gap-filled wind data are eliminated using the small median test (Lambert et al., 2003);
5. Log-linear polynomial (Oncley et al., 1996; Grachev et al., 2005) is applied further to gap-fill the U and V wind components;
6. The process of polynomial fitting and erroneous data elimination are repeated in the algorithm to calculate gap-filled and de-noised wind speed and wind direction;
7. Resulting gap-filled and de-noised wind speed data are validated against the observational data from the TTU 200 m meteorological tower.

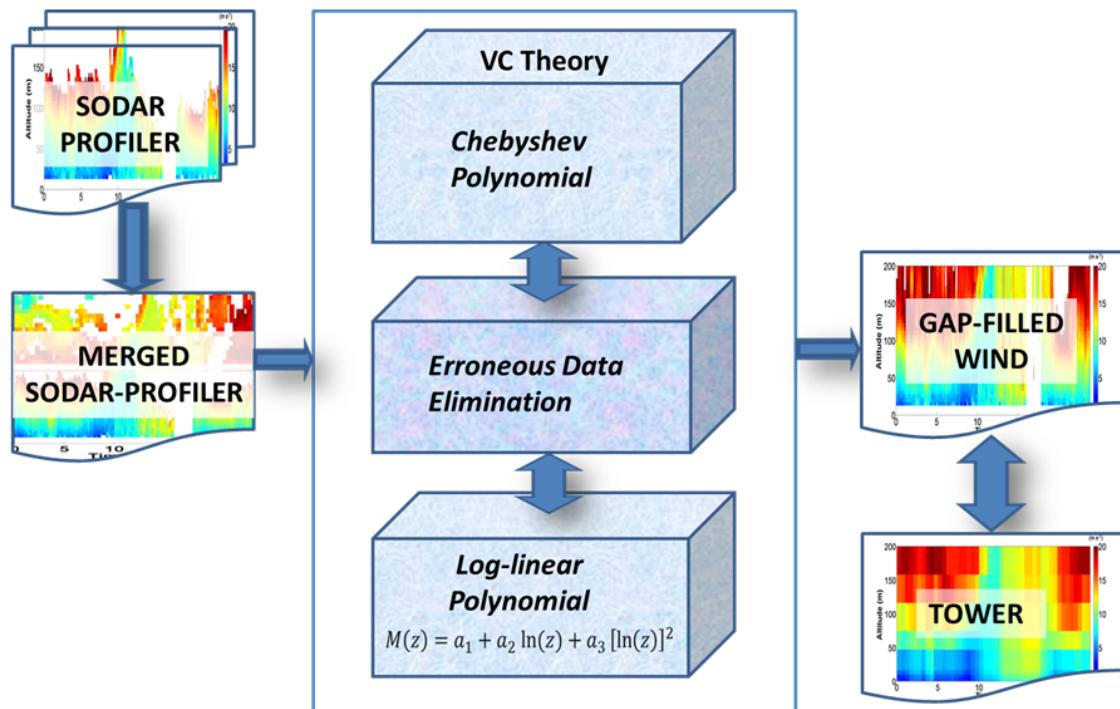


Figure 3-17. Proposed gap-filling and de-noising algorithm (see text for details).

In Figure 3-18, the performance of this algorithm is shown using wind data from 30 January 2012. Time-height plots of raw SoDAR and tower wind speeds are shown in the bottom panels of this figure. Visual examination of the spatial-temporal evolution of wind speed indicates that high wind speeds occurred in late evening to morning hours (associated with a low-level jet event). The SFAS SoDAR is able to capture the wind speeds up to an altitude of 125 m. However, the data availability above this altitude is limited. There is also a data gap during the early evening hours (SoDARs do not perform well during near-neutral conditions that occur during the transition from a convective to a stably stratified boundary layer). The top-left panel of Figure 3-18 shows the time-height plot of wind data from the TTU wind profiler located adjacent to the SoDAR and the meteorological tower. Note that the profiler does not measure any wind data below 160 m. The proposed gap-filling and de-noising algorithm is applied to the SoDAR and the wind profiler data. The wind profiler data are available at 20 minutes interval and are interpolated to match with the 10-minute resolution of the SoDAR data. The algorithm-generated wind speed data are shown in the top-right panel of Figure 3-18. After the application of the algorithm the data availability was increased reasonably, with improved data consistency, especially at higher altitudes.

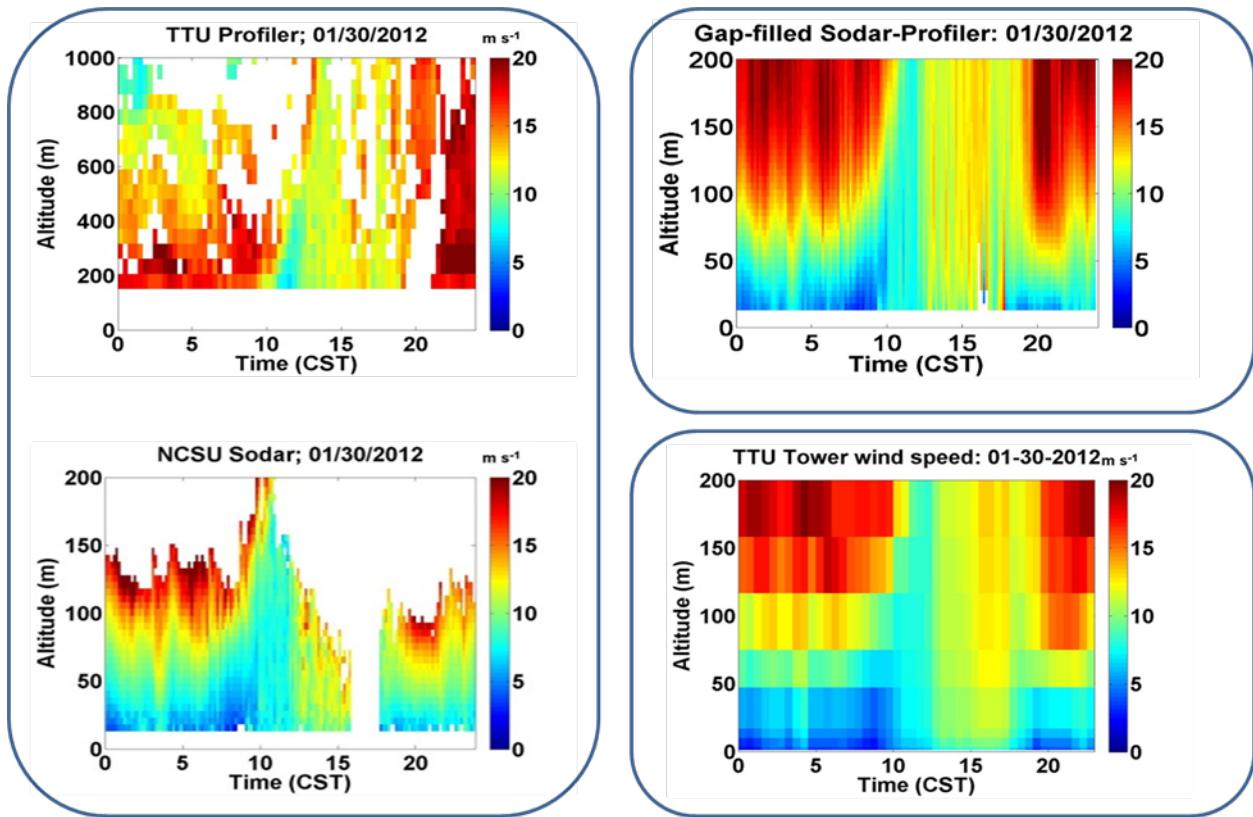


Figure 3-18. Time-height plots of average wind speed from a wind profiler (top-left panel), a SoDAR (bottom-left panel), and a meteorological tower (bottom-right panel). The gap-filled SoDAR-profiler wind data are shown in top-right panel. Time of interest: 30 January 2012.

3.5.3 Results and summary

The spectral analysis of the NCSU SoDAR data revealed the existence of a mesoscale scaling regime with a slope of $-5/3$. No significant variations in the slopes were found with respect to altitude. At present, it is not known if this result is specific to the southern Great Plains region or more universal. From a fundamental atmospheric science perspective, the spectral results are quite intriguing. From a practical wind energy

standpoint, they offer a rigorous framework for NWP model verification. Furthermore, the spectral information can be extremely useful for wind time-series forecasting and in mesoscale data assimilation (e.g., spectral nudging).

The proposed gap-filling and de-noising algorithm was applied to approximately four months (30th November, 2011 – 2nd April, 2012) of SoDAR and profiler data collected during the WFIP campaign. The results are shown in Figure 3-19. Clearly, the SoDAR data availability above 120 m AGL is low. Because of this lack of data, the average wind speed from SoDAR at higher elevation is noisy and unreliable. However, the application of the proposed algorithm significantly increases the data availability. More importantly, the gap-filled average wind speed values compare more favorably with the tower observations.

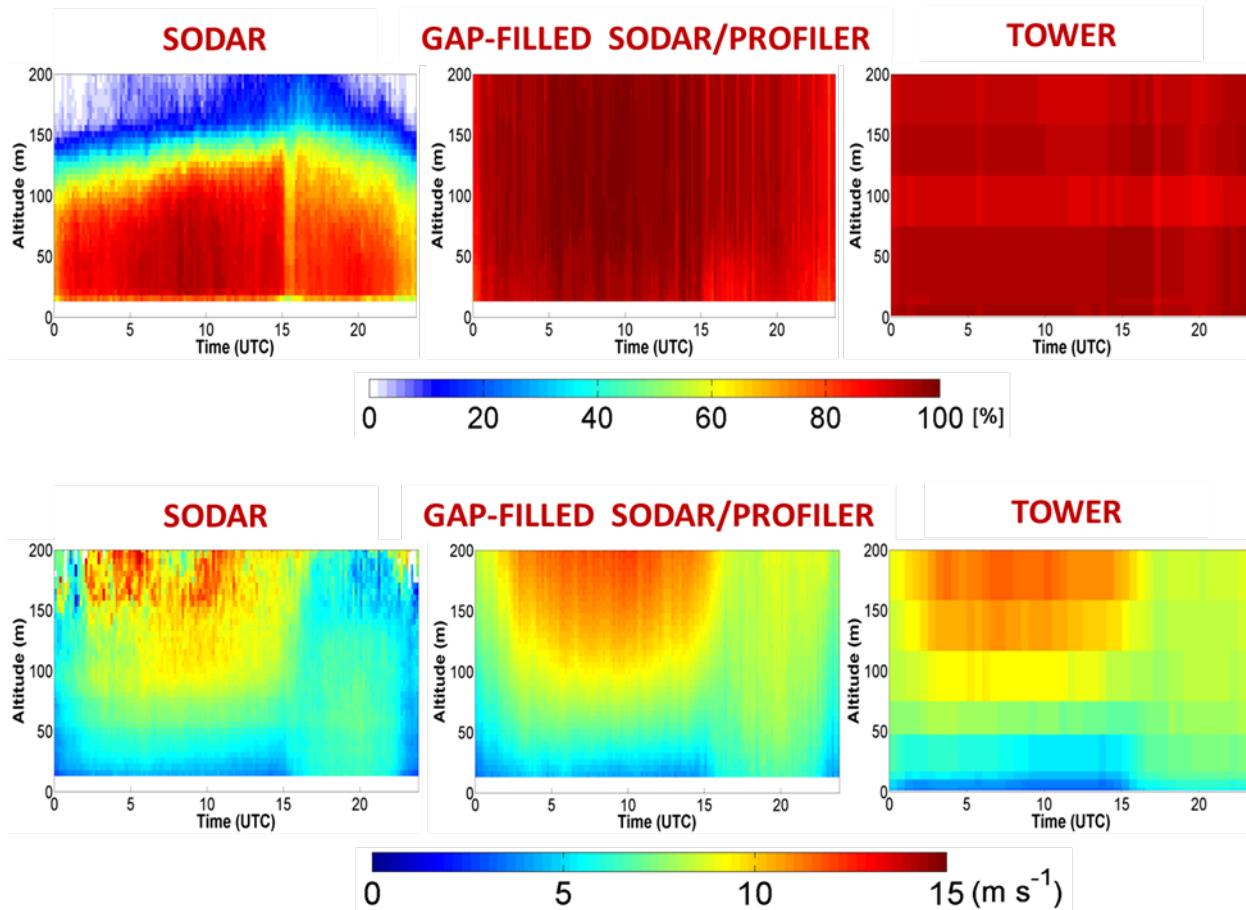


Figure 3-19. Time-height plots of data availability (top panels) and average wind speeds (bottom panel). Time frame: 30 November 2011 – 2 April 2012. See text for details.

4 Model systems and forecast campaign

4.1 Forecast systems

4.1.1 AWST/MESO

The WFIPFS is an enhanced and expanded version of AWST's operational eWind forecast system with five core components (Figure 4-1): (1) an ensemble of rapid-update short-term NWP forecasts; (2) a statistical adjustment procedure for each of the NWP forecasts; (3) a set of statistical time series prediction schemes; (4) an ensemble composite algorithm; and (5) a wind plant output model. Details of all five components are discussed in this section.

The NWP component of the forecast system is composed of three different modeling systems run by AWST and the High Resolution Rapid Refresh (HRRR) run by NOAA's Earth Systems Research Laboratory (ESRL). The high-resolution domains for each model are shown in Figure 4-2; most of the models have similar grid configurations (Table 4-1). However, the following attributes were varied among the ensemble members:

- NWP models used to generate the simulations
- Source of lateral boundary conditions (BC)
- Boundary layer physics schemes
- Convective cloud schemes, and
- Data assimilation schemes used to initialize the models.

The three models used by AWST were 1) the Weather Research and Forecasting (WRF) model (Skamarock et al. 2005), 2) the Advanced Regional Prediction System (ARPS) model (Xue et al. 2000, Xue et al. 2001), and 3) the Mesoscale Atmospheric Simulation System (MASS) model (Manobianco et al. 1996). All simulations had a horizontal grid resolution of 5 km and an update frequency of 2 hours with the exception of a low resolution Ensemble Kalman Filter (EnKF) member which was only used to produce initial conditions (ICs) and BCs for two of the high resolutions ensemble members. These models have numerous configuration options and considerable overlap in their routines (moist convection, turbulence, boundary layer, etc.). For WFIP, configurations were sufficiently varied but still consistent with the objective of generating the best possible low-level wind forecast (see Table 4-1).

The lateral BCs provided information about the evolution of the larger scale flow to the high resolution small domain of the rapid update cycle model. Two different sets of lateral BC data were used: (1) the Rapid Refresh (RR); and (2) the AWST EnKF forecasting system (discussed later in this section). The fundamental purpose for varying BCs was to incorporate variability in the evolution of the larger scale flow. By increasing the range of forecasted conditions, statistical techniques can identify and correct systematic observation and numerical modeling errors (biases) that can lead to an overall improved deterministic forecast.

All models were run in a cold-start mode using a new set of initial conditions from the large scale model (the Global Forecast System, or GFS, and the Rapid Refresh, or RAP), every cycle except for the MASS and AWST EnKF ensemble members which were run in a warm-start mode. The warm-start mode used the forecasted

state from a previous model run (every two hours for MASS and every 6 hours for the EnKF) as the starting point for the creation of the new initial state--therefore initialization data from a large-scale model was not needed.

A key component of forecasting hub-height (for WFIP, 80 m) wind speeds is predicting variations in turbulent mixing within and just above the planetary boundary layer (PBL), a historically difficult task for atmospheric models. For example, many parts of Texas experience a fairly strong diurnal cycle in low-level wind speeds associated with the LLJ. Transition periods, such as the morning rapid growth phase of the PBL (Freedman and Fitzjarrald 2001) can cause significant wind power ramps. In order to simulate the details of this cycle, NWP models need to accurately account for the vertical mixing of heat and momentum near the earth's surface. Therefore, several turbulent transport routines were used including the ARPS 1.5-TKE (turbulent kinetic energy; Sun and Chang 1986, Xue et al. 1996), WRF MYNN2 (Mellor-Yamada Nakanishi and Niino Level 2 PBL; Nakanishi and Niino 2006), WRF UW (University of Washington scheme; Bretherton and Park 2009) and MASS TKE (turbulent kinetic energy; Therry and Lacarrere 1983; Benoit et. al. 1989).

The cumulus parameterization in each model estimates the amount of convective development and precipitation on the sub-grid scale. Varying the cumulus parameterization in the ensemble, and therefore the evolution of precipitation-induced, low-level outflow boundary winds, allows the ensemble forecast to reflect uncertainty in the development of vertical cloud structures and precipitation. The convective parameterizations used for the real-time runs were the WRF KF (Kain-Fritsch scheme) for ARPS and WRF (Kain 2004), ETA KF for MASS (Kain and Fritsch 1990), and Grell (Grell and Devenyi 2002) for MASS.

Five different data assimilation schemes were employed including the: (1) ARPS 3D variational (3DVAR) scheme (ARPS 3DVAR; Gao et al. 2004); (2) Bratseth scheme using the ARPS data assimilation system (ADAS; Brewster, 1996); (3) Ensemble Kalman Filter (EnKF) scheme using WRF and WRF-DART (Hacker et al. 2005); (4) four-dimensional data assimilation (FDDA) scheme based on Newtonian relaxation (nudging; Stauffer and Seaman 1990); and (5) WRF variational (WRFVAR) hybrid scheme (Demirtas et al. 2009). Both the ADAS and ARPS 3DVAR schemes use isotropic covariance structures while the hybrid scheme uses anisotropic covariance structures estimated from a lower resolution EnKF ensemble forecast used as the initial conditions. The WRFVAR assimilation system was used only on the one of the three WRF ensemble members because the covariance structure could not be easily modified for mesoscale forecasting applications without the use of a hybrid technique. The objective in varying these schemes is to incorporate variance in the initial state of different ensemble members to better represent perturbation states in the subsequent model forecasts.

It was not possible to incorporate all permutations of these attributes into a 9-member ensemble. Computational resource constraints prevented a larger ensemble from being used. However, by producing a 13-hour forecast every two hours, a more recent set of ICs and BCs were available and updated meteorological observations were assimilated more often than most real-time high resolution ensembles. The varied model configuration for each ensemble member is summarized in Table 4-2.

The assimilation and forecast cycle for MASS, ARPS, and WRF are illustrated in Figure 4-3. The increased frequency of assimilation and model initializations allows each forecast to resolve specific meteorological phenomena better through the use of recent observations. The rapid update cycle also increases the number of model runs predicting specific wind-related meteorological events and provides up to five independent forecasts per member overlapping at a specific time. Given the size of the ensemble (13 members), 45 different overlapping NWP forecasts were produced for a given event. This ensemble generated a larger sample of raw model output for the statistical components of the forecast system and verification.

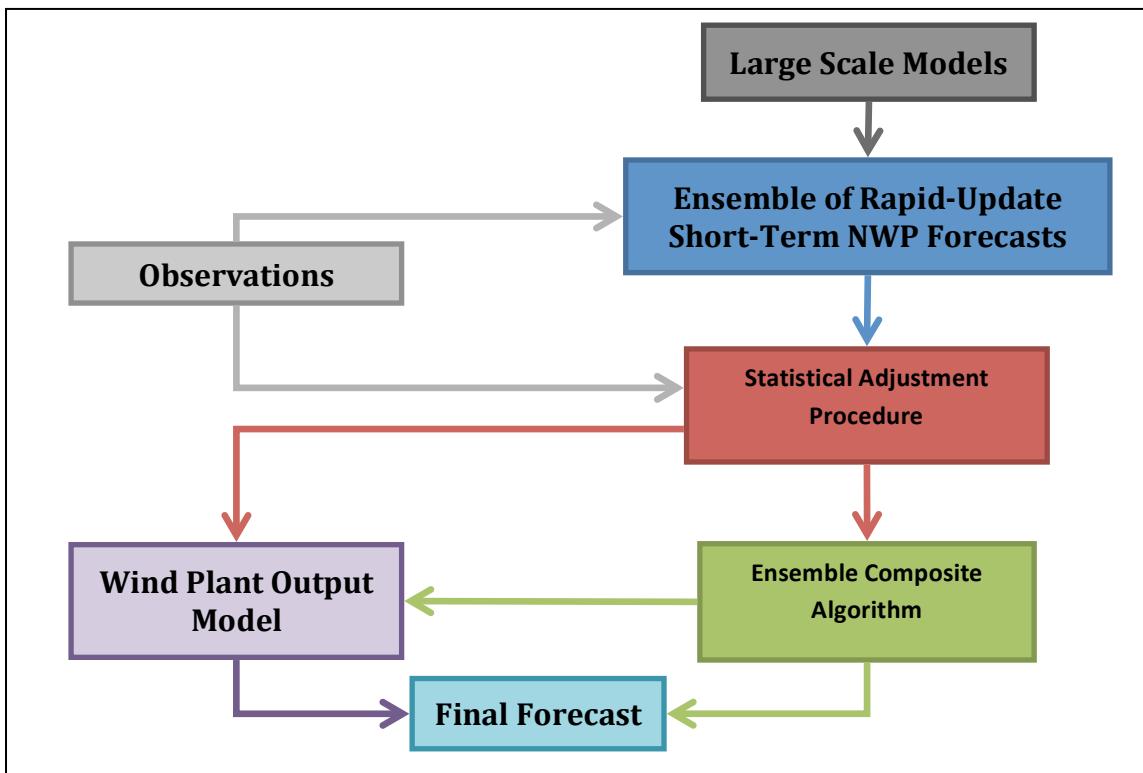


Figure 4-1: Overview of the WFIP forecast process.

The second major component of the forecast system is the statistical adjustment of forecasts produced by each NWP configuration. This approach is routinely referred to as Model Output Statistics (MOS) and its purpose is to remove systematic errors due to unresolved sub-grid processes, limitations in model physics, or data assimilation techniques. There are numerous ways in which the MOS concept can be applied. In this system, a screening multiple linear regression approach (Press et al. 1992) was used and applied separately to each wind plant site and each NWP model. A set of model state variables are used as input into the regression and is trained twice, once for observed power, and separately for observed hub-height wind speed. The model state variables that explain the greatest amount of variance in the observed data set are used. The underlying concept is that model error patterns are dependent upon the site location and so different adjustments are applied for each location.

The MOS training sample must be sufficiently large to avoid over fitting (Everett 2002). The training sample at the start of the project contained an insufficient number of forecasts so no MOS adjustment was performed until 60 days of forecasts were completed (30 days needed for training each model and 30 days to train ensemble forecasts). A separate MOS adjustment was formulated for each NWP configuration and updated daily using a rolling 30-day sample of forecasts and observed data.

The third component of the forecast system was a set of statistical time series prediction schemes. These schemes use information from the recent time series of meteorological and power production data at a forecast site to extrapolate future production. They are especially useful in very short-term (0- to 2-hour) look-ahead periods because the NWP models at the 5-km scale often do not have enough spatial resolution or sufficient data to initialize and forecast smaller-scale features (e.g. large eddies) that cause a substantial amount of wind variability on these time scales.

Table 4-1. AWST model members and static configuration

Parameter	ARPS	WRF	MASS
Number of Domains	1	2	2
Resolution	5 km	15 / 5 km	15 km / 5 km
Domain Size	153 x 135 x 50	150x135x50/154x139x50	150x135x40/150x135x40
Forecast Length	13 hours	13 hours	13-hour forecast 4 additional hours nudging
Initialization	Cold start	Cold start	MASS warm start IC /RR BC (Cold start at 0000 UTC)
Time Step	7.5 s large 2.5 s small	Adaptive 90 s /30 s	27 s/9 s
Advection	4 th Order	5 th order horizontal, 3 rd order vertical	MPDATA ⁶
Mixing	1.5 TKE / 4 th order mixing	MYNN 2 or UW	MASS TKE
Parameterization	Lin Ice ⁷	WSM6 ⁸	Mixed Phase ⁹
Microphysics	WRF KF	WRF KF	ETA KF or Grell
Cumulus			
Parameterization	Simplified surface radiation physics ¹⁰	RRTMG ¹¹	SRPH ¹²
Radiation	ARPS ¹³	Thermal diffusion ¹⁴	SRPH ¹⁵
Parameterization			
Surface			
Parameterization			

⁶ Multi-dimensional positive definite advection transport algorithm; Smolarkiewicz 1998⁷ Lin et al. 1983; Tao et al. 1989⁸ WRF single-moment 6 class microphysics; Hong and Lim 2006⁹ Lin et al. 1983; Rutledge and Hobbs 1983¹⁰ Xue et al. 2000¹¹ Rapid radiative transfer model G; Iacono et al. 2008¹² Surface energy budget-radiation-planetary boundary layer-hydrology scheme

Stephens and Webster 1979; Savijarvi 1990

¹³ Sun and Chang 1986, Noilhan and Planton 1989¹⁴ Dudhia 1996¹⁵ Surface energy budget-radiation-planetary boundary layer-hydrology scheme; Noilhan and Planton 1989

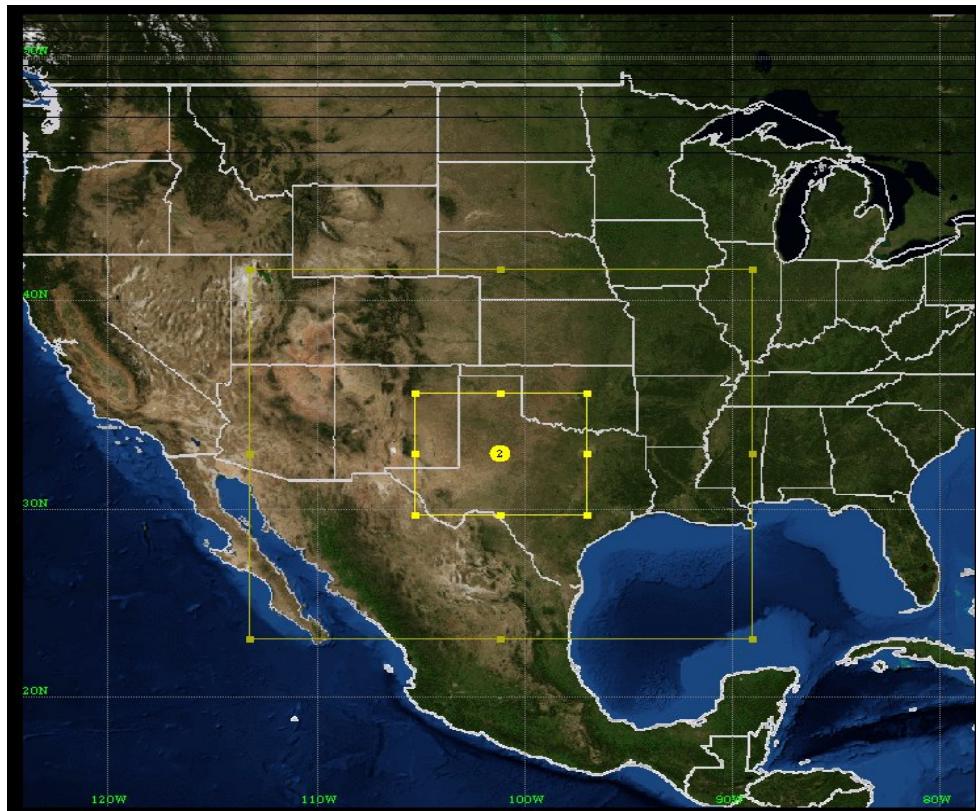


Figure 4-2: Domain of the WRF two-way nested and MASS one-way nested runs. The outer grid has 15-km grid space while the inner grid has 5-km grid spacing. For ARPS, only the inner grid shown here was run.

Table 4-2. Table of AWST model members and run configuration parameters, ARPS members IDs are denoted as DOEA1-3, MASS members are denoted, DOEM1-3, and WRF members are denoted as DOEW1-3, with the last integer representing each member of the ensemble and model.

Name	ID	Model	Assimilation Method	PERTURBATION	IC/BC	Grid Spacing
ARPS 3DVAR	DOEA1	ARPS	IAU-3DVAR	3DVAR	RR/RR	5km
ARPS ENKF	DOEA2	ARPS	IAU-3DVAR	EnKF	EnKF/EnKF	5km
ARPS ADAS	DOEA3	ARPS	IAU-ADAS	ADAS	RR/RR	5km
MASS NUDGING	DOEM1	MASS	Nudging	KF Convection	WARM/RR	15/5km
MASS OBS DENIAL	DOEM2	MASS	Nudging	No Project Observations	WARM/RR	15/5km
MASS GRELL	DOEM3	MASS	Nudging	Grell Convection	WARM/RR	15/5km
WRF ENKF	DOEW1	WRF	Hybrid-WRFVAR	MYNN2 PBL	EnKF/EnKF	15/5km
WRF MYNN	DOEW2	WRF	None-DFI	UW PBL	RR/RR	15/5km
WRF UW HRRR	DOEW3	WRF	None-DFI	MYNN2 PBL	RR/RR	15/5km
HRRR	HRRR	ERSL	None	-	RR/RR	3km
ENKF	EnKF	WRF	DART	-	WARM/GFS	45/15km

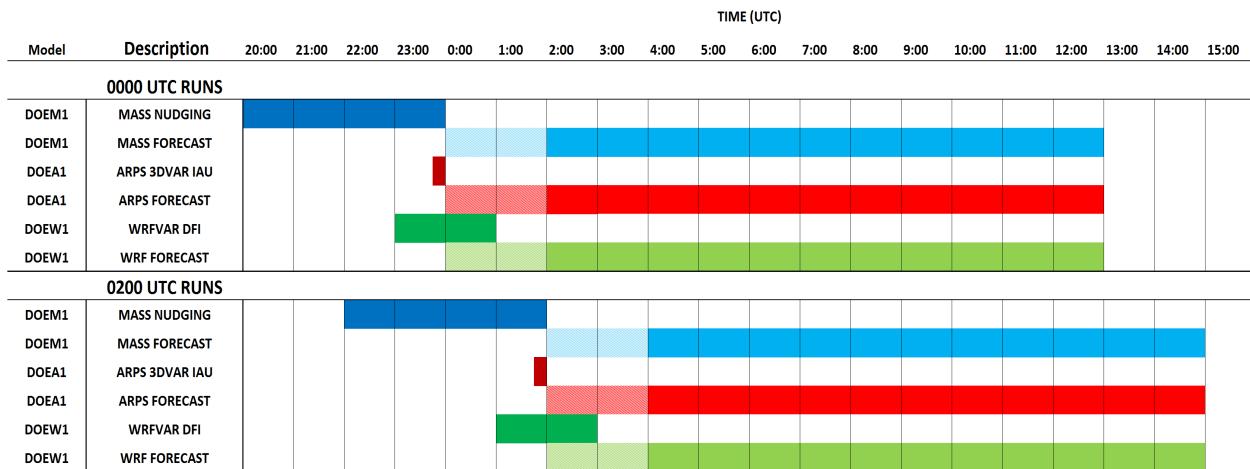


Figure 4-3: Assimilation and forecast timeline for one member of the 0000 and 0200 UTC ARPS, MASS and WRF WFIP cycle, the hatched area denotes the part of the forecast that was unavailable for real-time applications.

There is a lag before NWP forecast data are available due to the time required for data gathering, analysis, initialization, and model execution. Thus, there is typically more recent data available than what is used to initialize the latest available NWP run. There are, however, time series prediction schemes that take advantage of the newer data to improve the NWP forecast performance for the 0- to 2-hour look-ahead period. The use of the time series prediction system, known as the Persistence Adjust (PA) method, was incorporated into the WFIPFS NWP forecast and also included in the MOS-adjusted forecast.

The PA method determines the initial forecast bias at the time of forecast generation and then applies a static bias correction to the remainder of the 6-hour forecast. Similarly, the MOS method corrects the initial bias but applies a separate bias correction independently to each forecast interval. For both bias correction techniques under persistent conditions, the first available forecast interval (i.e. the two-hour forecast at 15-minute intervals) will be a persistence forecast using the latest observations translating to an extremely low forecast error.

The fourth component of the forecast system is the ensemble composite model. This model ingests forecasts from other system components and statistically generates the final deterministic or probabilistic forecast product. For a deterministic forecast, the role of the ensemble composite model is to weight each individual forecast according to its performance in a training sample from the previous 30 days. The fundamental concept is to assign greater weight to forecast members that are likely to perform better based on previous forecasts and observations in the training sample. This approach also permits the weighting of the individual methods to shift from one look-ahead period to another.

In a probabilistic mode, all model forecasts are used to generate the probability of exceedance for several ramp rate thresholds and a probability distribution of ramp rates based on quantile regression (Koenker 1994). For both the deterministic and probabilistic forecast, separate prediction equations were developed for each look-ahead period (i.e. 1 hour, 2 hour, etc.) and each wind plant. The deterministic ensemble forecast system was trained using the most recent 30 days of observed data and model forecasts, while the probabilistic ensemble forecast system was trained using the previous 60 days of observed data and model forecasts.

The experiments run for the deterministic ensemble forecast system are shown in Table 4-3. During the forecast campaign, only two ensemble forecasts were run; the optimized ensemble (OPTENS) used as the final WFIP real-time forecast and the OPTENS_STWPF or short-term wind power forecast system (used as a Baseline) to

compare with the OPTENS forecast. The OPTENS_STWPF used MASS model forecasts with ICs and BCs from the GFS and NAM. Both real-time ensemble methods were retrained daily, producing a new set of prediction equations used to produce the forecast.

Several WFIP probabilistic forecasts were also generated in real-time during the WFIP campaign. One for each individual WFIP model forecast and then an ensemble forecast which included forecasts from all WFIP model members. The forecasts produced a probability of exceedance value for several ramp rate thresholds. These forecasts were compared to a probabilistic ramp forecast produced using the ERCOT Large Ramp Alert System (ELRAS), an operational ramp forecasting system based on an earlier version of the ARPS model that does not include incremental analysis update (IAU; Bloom et al. 1996) observation assimilation.

The entire forecast process depicted in Figure 4-1 is summarized as follows: First, the NWP component is run and the forecast information is passed to a statistical adjustment package. The statistical adjustment package extracts an unadjusted 6-hour NWP model forecast at 15 minute intervals for input into the MOS, PA, and unadjusted methods in which an updated 6-hour forecast is produced. Then the 15-minute forecast methods are combined and used as input into the ensemble forecast system, which generates a 6-hour deterministic forecast every 15 minutes.

Due to the data volume and computing resources needed to process the HRRR forecasts in a real-time setting, the HRRR was not included in the real-time OPTENS forecast. To determine the sensitivity of including the HRRR as inputs for an ensemble forecast, several additional experiments were run historically with static weighting coefficients that were not updated (Table 4-3) with (ENS_WHRRR) and without (ENS_NHRRR) the HRRR model. The static weights provide an indication of HRRR member performance and are preferred over the time varying weighting due to the large computational resources needed to retrain each predictive equation on a daily basis.

For an additional sensitivity study, the ensemble was generated from the best three performing WFIP members (ARPS 3DVAR, WRF MYNN2 and MASS Grell) based on forecast performance discussed in Chapter 5. This new experiment was run to determine the performance of an ensemble forecast using only the best performing NWP members. For the second experiment the ensemble was run with (ENS_BWHRRR) and without (ENS_BNHRRR) the HRRR forecast to determine an additional sensitivity of the 3-member ensemble forecast performance to including the HRRR model (especially given the HRRR's 3 km resolution 1 hr update cycle versus the model members' 5 km resolution 2-hr update cycles).

Table 4-3: List of ensemble forecast methods for the WFIP project by type and included members. The real-time WFIP forecast method is optimized ensemble (OPTENS) and the baseline forecast method is OPTENS_STWPF (short term wind power forecast). To compare sensitivity to model members, ENS_WHRR, ENS_NHRR, ENS_BWHRR, and ENS_BNHRR were run but without time varying weighting as used in the OPTENS method.

Ensemble Name	Method Type	Members
OPTENS (WFIP Composite)	Time Varying, Site Varying Weighting	WFIP members without HRRR
OPTENS_STWPF (Baseline)	Time Varying, Site Varying Weighting	GFS/NAM Operational MASS members
ENS_WHRRR	Static Weighting	All WFIP with HRRR
ENS_NHRRR	Static Weighting	All WFIP members without HRRR
ENS_BWHRRR	Static Weighting	Best 3 WFIP members with HRRR
ENS_BNHRRR	Static Weighting	Best 3 WFIP members without HRRR

The fifth and final component of the forecasting system is the wind plant output model. This statistical model represents the relationship between meteorological variables and power production for each wind plant. This relationship is constructed from a sample of measured meteorological and power production data from each wind plant. The forecasted power production for regional or system-wide aggregates of wind plants is constructed by adding the predictions from the individual facilities. The plant output model is used to convert forecasts of meteorological variables to predictions of power production. An ensemble of power production predictions was generated for each forecast time using the meteorological forecasts from each ensemble member as well as the ensemble composite forecast.

The five components discussed previously were integrated to produce both real-time operational and historical forecasts for several sensitivity studies during the project. In the operational implementation, wind power production forecasts were generated every 15 minutes for the duration of the modeling campaign. The 15-minute forecast frequency was greater than the update frequency of all NWP cycles (Figure 4-3).

Each forecast timeline extended 6 hours after the time of forecast delivery. For each site, the forecasts included the (1) 15-minute averaged wind speed and direction, (2) 15-minute average power production, (3) probability of the 15-minute, 60-minute, and 180-minute ramp rates exceeding a set thresholds beginning at the start of each 15-minute interval in the forecast period, and (4) ramp events and attributes (start time, duration, amplitude, and maximum 15-minute embedded ramp rate) that were expected to begin within the 6-hour forecast period. All individual and ensemble composite forecast data for the target sites and subsets of the NWP output data were archived for the evaluation and analysis activities.

Numerous types of project and non-project data were incorporated into the assimilation system used in the physical NWP forecast and statistical models (Table 4-4). Each forecast system assimilated a particular suite of observations dependent upon what data were supported by the assimilation system. All members assimilated observations with the exception of the two WRF members initialized from the RR without any additional data assimilation (Table 4-2). The vast majority of routine observations were assimilated by the WRF, ARPS, and MASS including upper air (rawinsonde), land (Automated Surface Observation System, ASOS), and marine surface (buoy) observations.

The capability for all three models to assimilate SoDAR, profilers, tall tower, and mesonet observations was developed at the beginning of the project. In addition, several additional data types were supported for specific assimilation systems, including radar data for ARPS and satellite winds for WRF and the EnKF assimilation system.

All project and non-project data were subjected to quality control (QC) procedures. This QC included out-of-range checks and procedures to identify bad data based on the spatial gradient of analysis increment and large differences of the observation from the background state. The assimilation of project and non-project observations in the NWP models improved the analysis used to initialize each short-term forecast ensemble member (see results discussed in Chapter 5).

Assimilating observed project and non-project data may lead to dynamically unbalanced initial conditions that can degrade short-term forecasts. Therefore, each model used a technique to address this issue (Figure 4-3 and Table 4-2). For WRF, a digital filter (WRF DFI; Lynch and Huang 1992) was run forward and backward for one hour before starting each run. ARPS used an IAU scheme that nudged a set of previous initial conditions 15 minutes before forecast initialization. In MASS, a nudging scheme was used for assimilation of observations over a 4-hour period ending at the initialization time. In addition MASS was only cold started once per day from a larger-scale model run instead of a previous MASS forecast. This approach was used to limit the amount of imbalance from a new set of lower resolution RR initial conditions. The other models were cold

started each cycle with the exception of the EnKF--it was never cold-started except in the rare instance of hardware or software malfunction.

Several sensitivity studies were performed to determine the impact of observations from the sensors deployed in this project on forecast performance. For the real-time 9-member ensemble, one MASS member (MASS OBS DENIAL) omitted project observations to determine their impact on a daily basis in real-time. In addition to the real-time, data denial ensemble member, the ARPS 3DVAR member (DOEA4) was run historically for two 30-day periods in December 2011 and April 2012 with all observations or only non-project observations (DOEA5) producing a 13-hour forecast every 6 hours.

The two, 30-day historical observation denial runs provide insight into the impact of project observations on forecast performance given the large sample size and variability of weather conditions in the 30-day period. These simulations were run using the RR that assimilated project observations for ICs and BCs. Therefore DOEA4-DOEA5 forecasts could be influenced by project observations still present in the IC and BC. To address this issue, ESRL ran a 7-day period of forecasts without assimilating project observations from 1 - 7 December 2011. This period was selected due to a large ramp event that occurred on 3 December 2011. The ARPS forecasts (DOEA7-DOEA8, Table 4-5) were generated using ICs and BCs from the experimental data-denied RR forecasts to determine the impact of project observations over this 7-day period.

Table 4-4: Table of observations assimilated into the AWST ensemble members. Project data are highlighted in grey.

Data Type	Ingest Source for AWST Model Members			
	WRF	ARPS	MASS	EnKF
Surface	MADIS ¹⁶	NOAAPort	NOAAPort	MADIS
Rawinsonde	MADIS	NOAAPort	NOAAPort	MADIS
Profiler	MADIS			MADIS
Mesonet	MADIS			MADIS
WFIP SoDARs and profilers	MADIS			No decoder
SAT WINDS	MADIS	No decoder	No decoder	MADIS
ERCOT WGR met towers	ERCOT			No decoder
TTU tower	TTU			No decoder
Radar velocity azimuth display (VAD)	No decoder	88d2arps (ARPS radar data decoder).	88d2arps	No decoder
Radar reflectivity	No decoder	88d2arps	No decoder	No decoder
Radar radial velocity	No decoder	88d2arps	N/A	No decoder
Buoy	MADIS	NOAAPort	NOAAPort	MADIS
Rapid Refresh 3-D gridded data	WPS ¹⁷ ungrib	ext2arps (ARPS gridded data decoder).	WPS ungrib	WPS ungrib

¹⁶ Meteorolgocial assimilation data ingest system

¹⁷ WRF preprocessor system

The historical sensitivity study quantified the impact of project observations on raw, short-term power forecasts directly from the NWP model without using a statistical model to post process the NWP output data as is done in the full forecast system. The real-time sensitivity study helped determine the effect of project observations using MOS-corrected forecasts. Results from both studies focused on short-term power forecast errors over all sampled cases (see Chapter 5).

Table 4-5: List of model members for the data denial sensitivity analysis.

NAME	ID	Project Observations	Assimilation Method	Initial Conditions (IC) Boundary Conditions (BC)
ARPS All Obs	DOEA4	Yes	ARPS 3DVAR	RR
ARPS Obs Denial	DOEA5	No	ARPS 3DVAR	RR
ARPS All Obs	DOEA7	Yes	ARPS 3DVAR	Data Denial RR
ARPS Obs Denial	DOEA8	No	ARPS 3DVAR	Data Denial RR

Table 4-6: List of AWST and ESRL data denial periods.

Name	Period
AWST Winter	December 2011 (4 initializations per day)
AWST Spring	April 2012 (4 initializations per day)
ESRL Winter	1 - 7 December 2011 (12 initializations per day)

4.1.2 CAPS OU

The Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma (OU) produced real-time 3-km grid scale forecasts covering the WFIP domain and adjacent portions of neighboring states (Figure 4-4). The 12 hour-long forecasts were produced every 6 hours using the CAPS ARPS numerical model (Xue et al., 2000, Xue et al., 2003). The ARPS model has a TKE 1.5-order sub grid turbulence closure model used with a Sun and Chang (1986) planetary boundary layer (PBL) parameterization for the mixed boundary layer. Surface fluxes are specified by similarity theory, using a diagnosed PBL height. Single-moment 5-category ice-microphysics (Lin-Tao, Tao et al, 1993) was used.

The ARPS model was configured to provide high vertical resolution in the boundary layer to better resolve the difference between the surface wind (typically measured by anemometers at 10 m above ground level (AGL) and the wind turbine hub height (80 m AGL). The number of vertical grid levels was set at 58, and average grid spacing of 360 m was specified with stretching of the vertical distance between adjacent grid levels following a hyperbolic tangent profile rather than the usual cubic function. The result was the placement of 5 vertical grid levels in the lowest 200 m AGL rather than the more typical 3 vertical levels (Figure 4-5), with similar improvement in spacing elsewhere in the lowest 1-km AGL.

Data assimilation, including Doppler weather radar wind and reflectivity data from the NEXRAD radars in the region, was done using a 3DVAR method with cloud analysis (Gao et al. 2004, Brewster et al. 2005, Hu et al. 2006) and Incremental Analysis Updating (IAU, Bloom et al. 1996) following real-time strategy similar to Brewster et al. (2008). Observation data used included available standard surface data, mesonet stations (Oklahoma Mesonet and West Texas Mesonet), and surface observations, SoDAR, and profiler data from the

WFIP field deployment. For each forecast the 3DVAR analysis was run at 15 minutes before the hour, and 15-minutes of IAU was performed (Figure 4-6). The latest model forecasts available in real-time from the operational 12-km NAM were used as the background and BCs.

Post processing included the generation of model forecast soundings every 15 minutes at pre-determined verification and ensemble processing points relevant to wind production in the ERCOT area. These soundings were automatically transferred to AWST/MESO for use in the production of multi-model ensembles and forecast verification.

WFIP forecasts at OU ran from 26 July 2011 through 13 Sept 2012. Computing resources provided to the project by the OU Supercomputing Center for Education and Research (OSCER) included 120 cores on an Intel Harpertown Xeon cluster (Sooner) from August 2011 to March 2012 and then 64 cores on an upgraded Intel Sandy Bridge Xeon cluster (Boomer) from March to September 2012. Although the computer and forecast system are an academic systems without 24/7 support, about 92% percent of forecasts ran to completion. An additional 1.5 % of forecasts ran but not to completion. Lost forecasts were due to unscheduled computer outages or networking downtime that led to loss of needed background or boundary condition data. Verification of the OU production runs is presented as part of section 5.2.

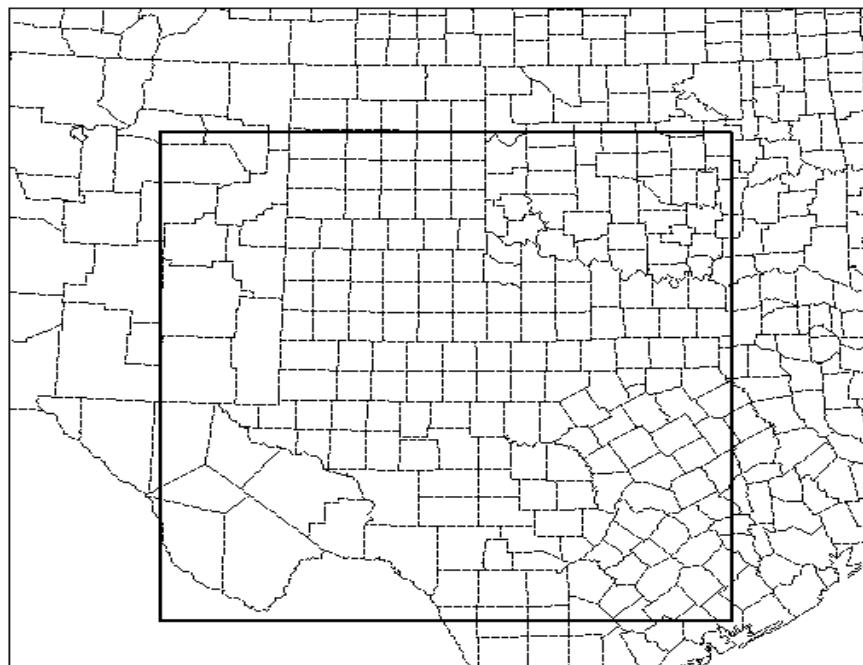


Figure 4-4, Map of the CAPS 3-km forecast domain (1200 x 1080 km), with the WFIP-South Study Area in the inner box.

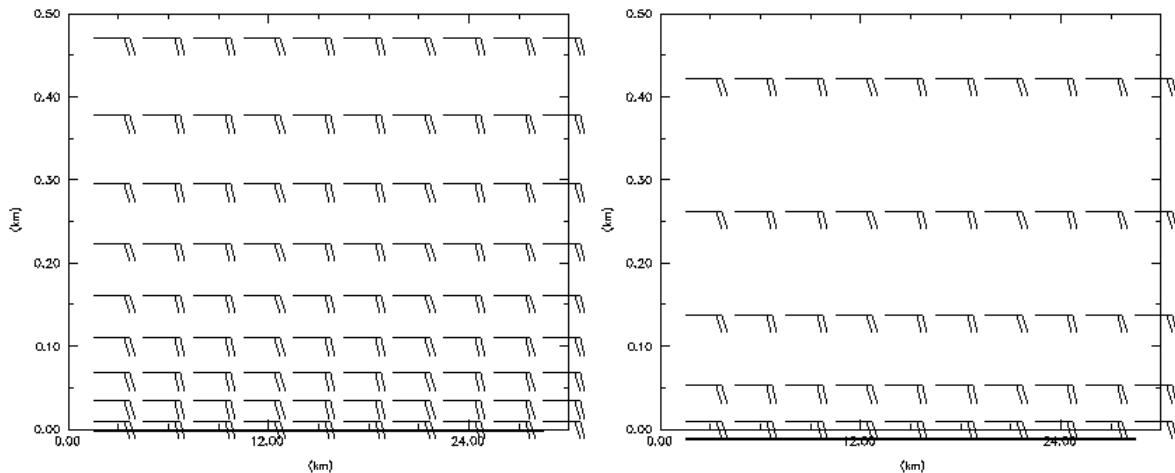


Figure 4-5. Vertical spacing of mass and horizontal wind levels in the lowest 500 m. Left: the ARPS production runs for WFIP, right: typical vertical spacing in real-time high-resolution models.

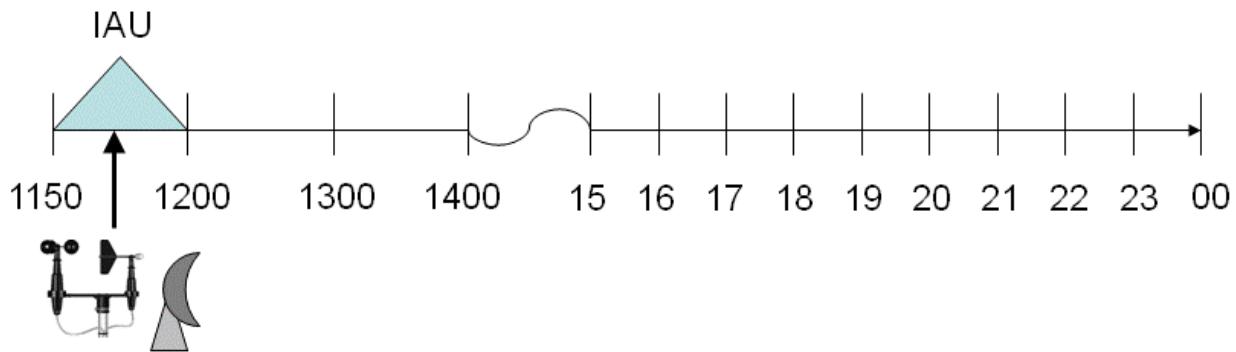


Figure 4-6: Schematic of data assimilation and forecast timeline for a sample run time (1200 UTC) showing IAU applied with a triangular time weighting scheme in the 15-min pre-forecast period and a 12-h forecast.

4.1.3 TTU

The TTU WFIP modeling system used for WFIP is the Advanced Research WRF model version 3.3 (Skamarock et al. 2008). The model physics used for all experiments are the Yonsei University (YSU) PBL scheme (Hong et al. 2006), the Kain-Fritsch cumulus parameterization (Kain and Fritsch 1990, Kain and Fritsch 1993), the Noah land surface model (Chen and Dudhia 2001), Thompson microphysics (Thompson et al. 2004), the Rapid Radiative Transfer Model (RRTM) longwave radiation scheme (Mlawer et al. 1997), and the Dudhia shortwave radiation scheme (Dudhia 1989). All model integrations use 38 vertical levels and are performed at 12-km grid spacing (with the exception of the runs performed on the 36-km domain that provide BCs for the 12-km EnKF experiments). Ensemble boundary conditions for the 36-km domain are produced through perturbations about the GFS global model described in Torn et al. (2006). For the GSI experiments, the boundary conditions are provided by the GFS model every 6 hr to the 12-km domain.

This TTU effort utilized both an EnKF and a 3DVAR data assimilation system. The EnKF employed is the Data Assimilation Research Testbed version (DART, Anderson et al. 2009). It uses the ensemble adjustment Kalman

filter (Anderson 2001) technique, which Anderson (2001) showed performs better than the traditional EnKF. To mitigate the effects of small sample size (Anderson and Anderson 1999), adaptive prior covariance inflation (Anderson 2009) in space and time are employed, as well as a Gaspari-Cohn localization radius (Gaspari and Cohn 1999) in both the horizontal and vertical direction. EnKF covariances are purely flow-dependent and calculated from the ensemble. The ensemble size is 50 members, which is the same ensemble size within the real-time Texas Tech ensemble prediction system (TTEPS) that is able to run operationally on a large computing cluster. Every assimilation experiment for WFIP ran on a 6-hr cycle. The ensemble is initialized through random perturbations drawn from climatological covariances within the WRFVAR system (Barker et al. 2004). Figure 4-7 shows the 12-km domain on which the EnKF system is run. In addition, there is a larger parent 36-km domain (not shown) that stretches across the majority of the Pacific Ocean used to provide flow-dependent boundary conditions in the form of a 1-way nest to the ensemble members on the 12-km grid.

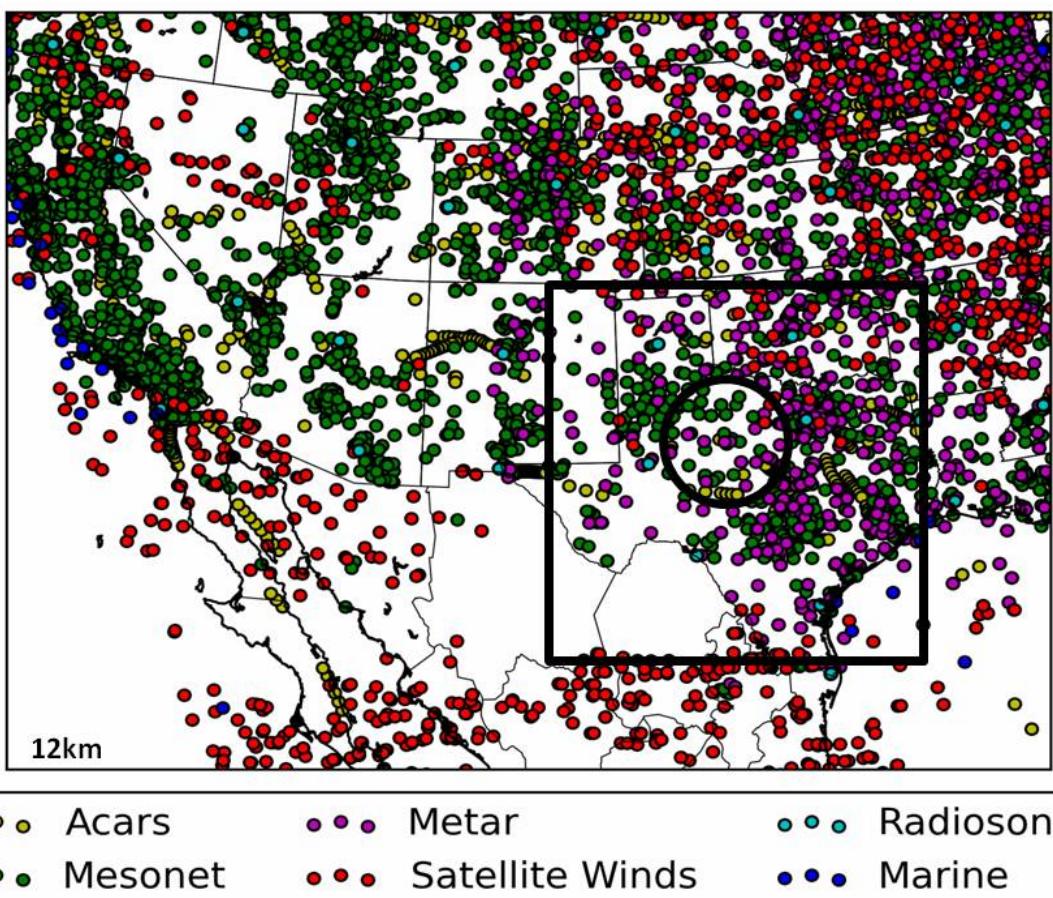


Figure 4-7. The 12km modeling domain used in this study. The black circle indicates the ramp verification area containing the 80-m meteorological towers, and the black box shows the verification area for the month-long experiments. A typical distribution of routine assimilated observations is also shown by the colored circles.

The 3DVAR system employed in WFIP is that of the WRF-based Global Statistical Interpolation system Version 2 (GSI, WRF Developmental 2 (GSI, WRF Developmental Testbed Center 2010). Analogous to the EnKF, the GSI possesses a number of specific parameters that affect the specific parameters that affect the horizontal and vertical extent of analysis increments (similar to EnKF localization), and a weight placed on the localization), and a weight placed on the first-guess variance (similar to EnKF inflation). GSI covariances are static and are generated using the static and are generated using the NMC method (Parrish and Derber 1992) with NAM 12km model forecasts. Both data assimilation and Both data assimilation and forecasts within the GSI 3DVAR system occur over the same domain shown in Figure 4-7. BCs to the GSI 3DVAR

forecasts are provided by the GFS. Both the EnKF and GSI data assimilation parameters are tuned since the quality of analyses and forecasts in each system depend on these variables and can vary with, for example, different domain sizes and terrain complexity. The tuning procedure was performed by cycling the assimilation system for a range of parameter values over a week of data assimilation cycles during the month-long verification period of December 2011 used in this study (28 assimilation cycles). Extended 24-hr forecasts were produced from each analysis, and the mean absolute errors measured against domain-wide surface-based and rawinsondes wind observations, averaged over all assimilation cycles were inspected to choose the optimal parameters. The tuned parameters were those that produced the smallest errors over the forecast period. These parameters were tuned specifically for optimal wind forecasts, and different parameters might exist that optimize other forecast aspects (although the same parameters found to optimize wind forecasts also optimized temperature forecasts). For the EnKF, the inflation parameters (6 total) and both the vertical and horizontal localization radii were tuned. For GSI, there are three parameters that weight the first-guess variance and control vertical and horizontal localization. Unlike the EnKF, for which a single parameter value collectively addresses all analysis variables, the GSI parameters exist independently for the following analysis variables: streamfunction, velocity potential, temperature, surface pressure, and specific humidity. In turn, a very large number of runs were required to tune these parameter values. Figure 4-8 shows an example of two such tuning runs, one associated with the vertical localization within the EnKF, and the other associated with the horizontal localization within the GSI. Three parameter values are shown in each example, with the optimal value depicted by the black line. The final tuned values for both the DART EnKF and GSI 3DVAR systems are shown in

Table 4-7.

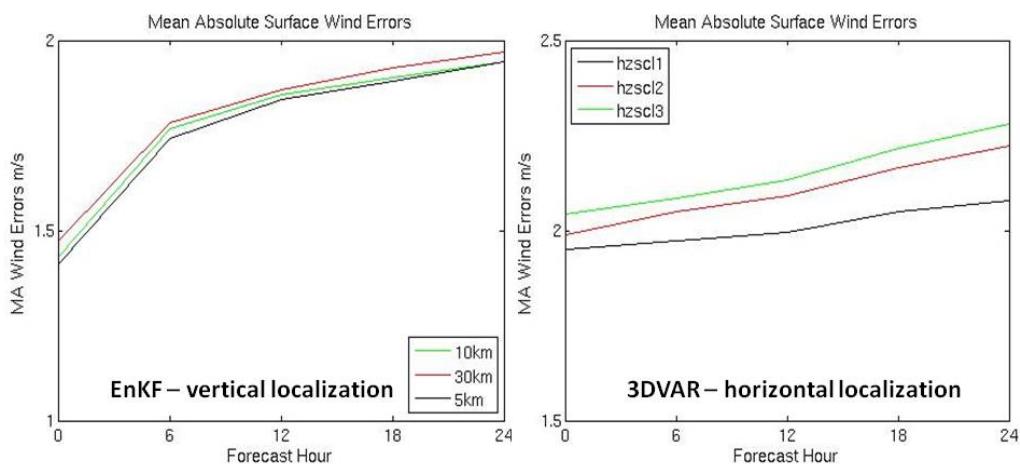


Figure 4-8. Two examples of tuning runs for DART EnKF vertical localization and GSI 3DVAR horizontal localization performed over the first week of December 2011.

A variety of experimental runs were performed in order to compare the performance of the DART EnKF and GSI 3DVAR, and to test the observational impacts of both mesonet surface data and profiler/SoDAR observations in each system. Two sets of experiments were performed: 1) a month-long continuously cycled integration during December 2011, and 2) individual periods of cycled assimilation runs that capture 10 individual wind ramp events (more specifically, 7 up ramps and 3 down ramps). High priority wind ramps for sensitivity studies were selected based on their internal ramp thresholds over the winter 2011/spring 2012 season. Each of these wind ramps were synoptically forced, either through frontal/dry line passages or synoptic-scale forcing of the surface pressure field, and do not include ramp cases from other phenomena such as convectively-driven outflow events or LLJs. A series of 24-hr forecasts were run from all analyses for both the month-long integration as well as each individual cycling period in order to evaluate any dependence

of performance on forecast lead time. Furthermore, for each cycled integration, a spin-up period of 2 days was run on a 6-hr assimilation cycle in both the EnKF and 3DVAR systems, mostly to allow flow dependence to build within the EnKF runs.

Table 4-7. Parameter values for the DART EnKF and GSI 3DVAR systems

DART EnKF	
Horizontal Localization	500 km
Vertical Localization	5 km
Inflation:inf_initial	1.02
Inflation:inf_sd_initial	0.9
Inflation:inf_damping	0.9
Inflation:inf_lower_bound	1.0
Inflation:inf_upper_bound	10000000
Inflation:inf_sd_lower_bound	0.9
GSI 3DVAR	
as_op	0.5, 0.5, 0.25, 0.35, 0.35
vs_op	1.5
hzscl	0.1, 0.2, 0.4

The month-long runs are verified through mean absolute errors against wind observations from surface METAR stations (scattered throughout the black box in Figure 4-7). The wind ramp events are verified against wind observations from twenty 80-meter meteorological towers located at WGRs (contained within the black circle in Figure 4-7). Four aspects of the wind ramps are used to measure the forecast skill of each experimental run: 1) ramp onset, 2) ramp duration, 3) ramp magnitude (that is, the difference between the minimum and maximum wind speed occurring during the ramp event), and 4) ramp maximum wind speed. This object-oriented verification approach was taken because large mean absolute errors were found with wind ramps that possessed relatively small timing errors (but showed otherwise well-forecast characteristics such as ramp duration and magnitude) compared with other runs that appeared to capture the overall ramp much less accurately. This verification strategy has the potential to allow wind plant operators to decide which forecasts are best for their purposes based on the skill associated with the different ramp aspects.

5 Results and discussion – Forecasting

5.1 AWST/MESO

Forecast performance can be validated using several different metrics. The choice of metric depends upon which aspects of the forecast are being validated. Deterministic forecasts are traditionally validated by computing the MAE, RMSE, and bias with respect to observed wind speed or power production. These error characteristics highlight the deviation of the forecast from the observed state. The results can be subdivided by delivery hour, forecast look-ahead hour, and model initialization time to emphasize the performance associated with a specific subset of forecasts or observations. In this study, observations and forecasts locations were aggregated to provide an overall regional or system-wide representation of forecast performance.

The MAE and RMSE metrics are most appropriate for large samples and longer-term trends in time series datasets. For ramp events or large changes in power production over short periods of time, these traditional metrics penalize phase errors and reward persistence forecasts. For WFIP forecasts, it was important to resolve these events given their disruptive nature and impact to system operations. Therefore, in addition to the standard deterministic power and wind speed forecasts, probabilistic and deterministic ramp forecasts were also produced.

The deterministic ramp forecasts were validated using a critical success index (CSI, Wilks 1995). The first step in computing CSI is to define a time window around an observed ramp. If the predicted ramp is forecasted within the window, either before or after the ramp event, then the forecast is considered a “hit”, otherwise it is considered a “miss.” When a predicted ramp event does not occur, it is considered a false alarm. The CSI is calculated as:

$$\text{csi} = \frac{\text{hit}}{(\text{hit}) + (\text{miss}) + (\text{false alarm})} \quad (1)$$

CSI is the ratio of the hits to the sum of the hits, misses, and false alarms. When the CSI=0, it means there are no hits and therefore there is no forecast skill. When the CSI=1, all the observed ramp events were predicted with no false alarms representing perfect forecast skill. For the CSI to be a representative forecast performance metric, a large enough window must be chosen to accommodate some degree of phase errors. A forecasted ramp event that occurs outside the determined window will be penalized twice; first from classifying the ramp event as a missed forecast, and second from a false alarm, since the ramp event was predicted outside of the ramp window. Therefore, for this study, several CSI hit windows were used, offering a more comprehensive set of results than can be provided by a single window.

An alternative to deterministic ramp event forecasts is probabilistic ramp rate forecasts. The probabilistic forecasts can communicate a range of possible forecasted outcomes by providing a probability distribution as a function of ramp rate and probability of exceedance for several ramp rate thresholds. This metric is more appropriate for determining ramp event forecast performance. For instance, convective weather can cause

both up ramps and down ramps. A deterministic ramp forecasting system may forecast only one of these events without the indication of a possible ramp of the opposite sign. On the other hand, a probabilistic forecast system can show that an up and down ramp are probable and the possibility of persistent conditions is low. This information cannot be communicated with a deterministic forecast.

To validate the ramp-based probabilistic forecasts, a Ranked Probability Skill Score (RPSS, Murphy 1969) is used to determine the forecast skill for each ramp rate bin using the Ranked Probability Score (RPS). The RPS (equation 2) is the sum of the squared deviation of the predicted (P_k) and observed (O_l) event cumulative probabilities summed over all bins:

$$RPS = \frac{1}{n-1} \sum_{m=1}^n \left[\left(\sum_{k=1}^m P_k \right) - \left(\sum_{l=1}^m O_l \right) \right]^2 \quad (2)$$

The RPSS (equation 3) contains the ratio of the RPS of the probabilistic forecast to that of the climatological probabilities (based upon the period December 1 2011 through Sept 1 2012) and used as a reference in this study:

$$RPSS = 1 - \frac{\overline{RPS}_{Forecast}}{\overline{RPS}_{climo}} \quad (3)$$

The RPSS represents the improvement of the ramp probability forecast over the climatological ramp probabilities. A RPSS value less than zero indicates the reference (climatological) forecast has more skill than the actual forecast while values greater than zero indicate greater forecast skill than the reference (climatological) forecast. For completeness, a RPSS contribution metric was also computed to determine the contribution of each ramp rate bin RPSS to that of the RPSS for the whole sample of all ramp rate bins.

In addition to RPS and RPSS, probability-based forecasts can also be verified using a reliability diagram, which compares forecasted probability to the observed frequency of an event. This metric can communicate information about biases in the probability forecast for a sample of events. A forecast exhibits perfect reliability when the forecast probability of an event equals the observed frequency. When the observed frequency is less (greater) than the forecast probability, the forecast shows a positive (negative) bias.

5.1.1 Deterministic forecast performance.

The nine AWST high-resolution models produced 13-hour forecasts¹⁸ every two hours for each forecast location for one year. During the project, most of the NWP runs had high availability. Both the ARPS 3DVAR

¹⁸ Note, as set forth earlier (see section 4.1.1), there were two forecast types: a bi-hourly model forecast and a 15-min delivery forecast. The WFIP models were run out 13 hours every 2 hours for each model member; the bi-hourly forecast were then used to generate a deliverable 6 hour forecast every 15 minutes using the 3 bias correction methods

and the WRF members (excluding the WRF EnKF) had the highest forecast availability around 94% (Figure 5-1) due to their previous use in an operational setting before the beginning of the project. The EnKF initialized members were missing BCs and ICs from the continued development of the EnKF forecasting system and resulted in both high-resolution members having lower forecast availability (60 - 75%). The HRRR forecast model states were generated sequentially by forecast hour at ESRL. The HRRR availability was much higher for forecast in the first 1-6 hours after model initialization and decreased sequentially for longer forecast hours due to missing model state data for longer forecast times.

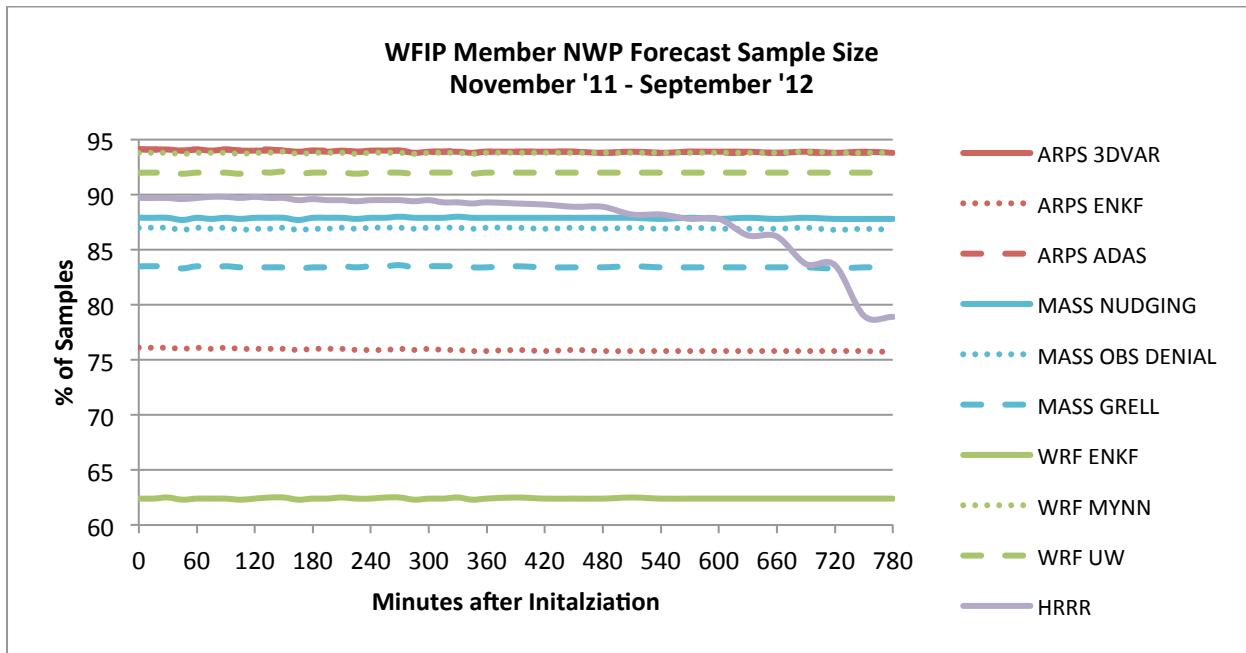


Figure 5-1. Forecast system sample size as a percent of maximum number of possible model runs for each of the 10 high-resolution WFIP NWP members.

The raw NWP model forecasts were used to characterize initialization error and performance since they do not include the impact of either recent observations or bias correction. The RMSE of the system-wide WFIP aggregate power production forecast from each raw NWP ensemble member and the HRRR is shown in Figure 5-2. At the beginning of the forecast period, the MASS members have the lowest short term forecast error. This result was most likely due to the four-hour nudging scheme and warm start model initialization that prevented the model from drifting out of dynamic balance after data assimilation. The HRRR, ARPS 3DVAR, and ARPS ADAS forecast members also had a lower forecast error compared with the WRF and EnKF initialized members at the beginning of the period.

There is a rapid decrease in error (~4% of capacity) within 30 minutes after observation assimilation for the ARPS members that may be attributed to the assimilation of radar data resolving convective features not present in the course resolution IC. ARPS, MASS and the HRRR also had the lowest error during the end of the forecast period. The EnKF ensemble members had the highest forecast error (18-22% of capacity) most likely due to the lower resolution of the EnKF forecasting system used as IC and BC (which could not resolve the mesoscale phenomena as well as the RR which was also used as IC and BC for the high resolution members).

The results for the system wide WFIP aggregate wind speed RMSE are similar to that of power (Figure 5-3). The wind speed verification only shows minor differences in performance when compared to power verification. All models showed a general increasing trend towards a positive bias during the forecast period (see Figure 5-4). The ARPS model had the greatest negative bias at the beginning of the forecast period (at -10% of capacity), most likely from biases in the assimilation system, while WRF and HRRR consistently showed a high positive bias (5% of capacity) in the aggregate power forecast (which has been documented in Mass 2011). The MASS model, that was warm started and included a nudging scheme, had the lowest model bias (0-4 % of capacity).

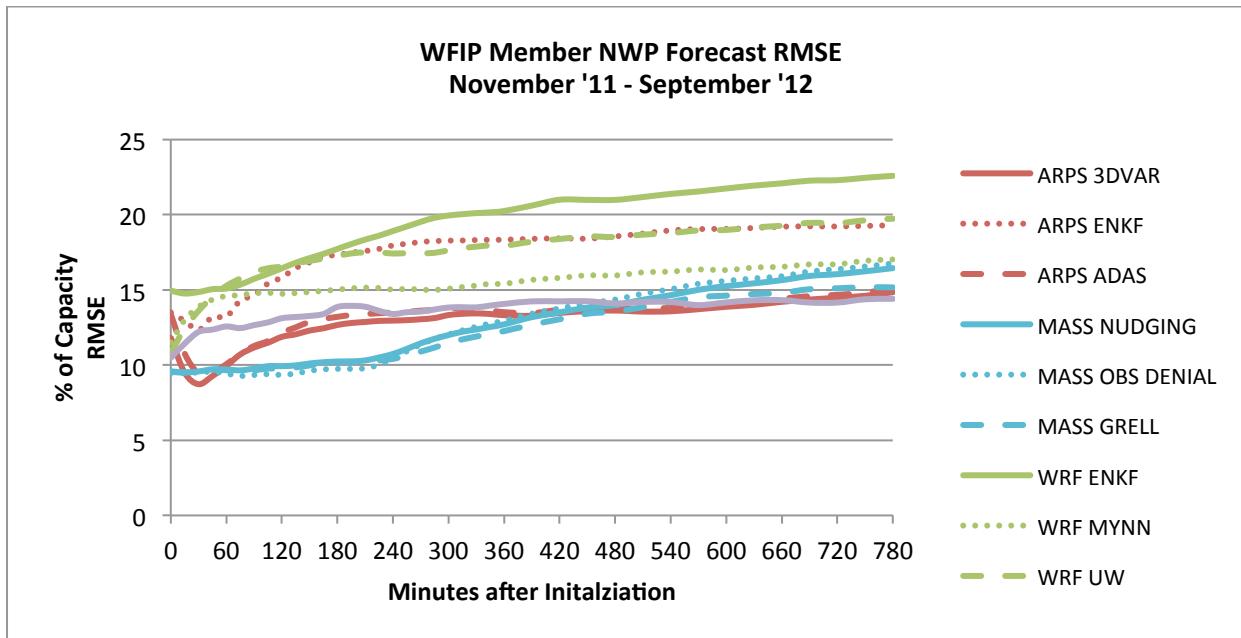


Figure 5-2. Root Mean Square Error (RMSE) of 0- to 13-hour power production forecasts for the WFIP system-wide aggregate as a percent of nameplate capacity for the each of the 10 high-resolution WFIP NWP members. The “zero” represents the time of model initialization,

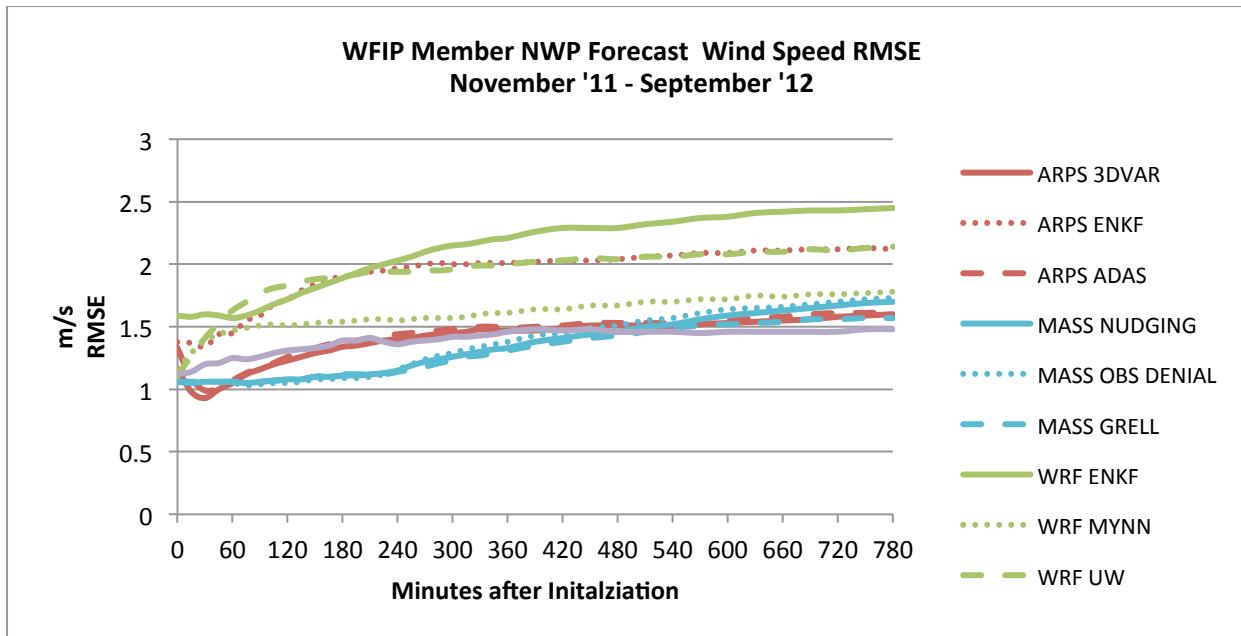


Figure 5-3. Root Mean Square Error (RMSE) in m/s of 0- to 13-hour capacity-weighted forecasts of average 80-m wind speed for the WFIP system-wide aggregate for the each of the 10 high-resolution WFIP NWP members. The “zero” time represents model initialization.

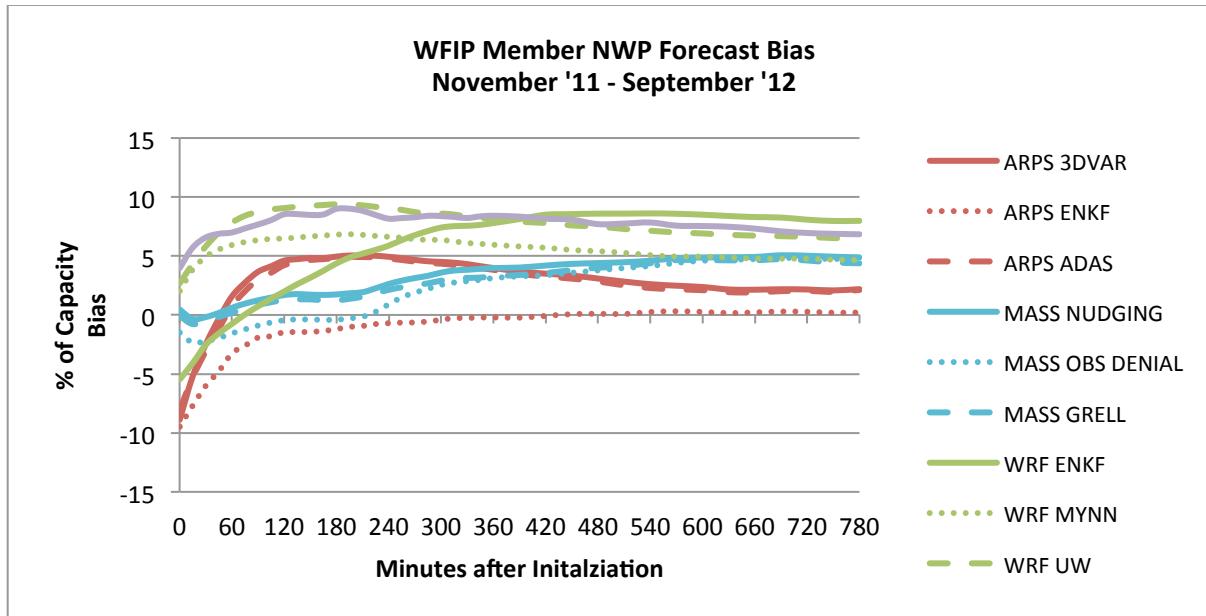


Figure 5-4: Bias of the 0- to 13-hour power production forecasts for the WFIP system-wide aggregate as a percent of nameplate capacity for each of the 10 high-resolution WFIP NWP members. The “zero” time represents model initialization,

The unadjusted forecast method verification is shown in Figure 5-5 (top) and differs from the forecast verification shown in Figure 5-2 as follows: (1) forecasts are issued every 15 minutes instead of every 120 minutes (2 hours), (2) the forecast extends 360 minutes (6 hours) being at the first 15 minute interval instead of 780 minutes (13 hours) beginning at the time of initialization, and (3) missing NWP model data are substituted using an older NWP run to generate the 15-minute forecast (up to 2 previous runs). ARPS (3DVAR, ADAS) and HRRR were the best performing unadjusted methods with the lowest RMSE (13 -14 % of capacity in Figure 5-5 (top) for a 360 minute (6-hour) forecast, followed by the MASS and WRF (MYNN, UW) models.

The unadjusted forecasts are produced from model runs that are only available for bias correction in real-time (see Figure 4-3). These forecasts exclude the first two hours of model integration generated while the model members are running. Unlike the unadjusted forecasts, the previous initialization results include these data from the latest model cycle and are not representative of the current real-time forecast system. Similar to the NWP initialization performance results (Figure 5-2) the EnKF initialized members (WRF and ARPS) had the highest forecast error of 18-22% of capacity (Figure 5-5 (top)).

The MOS forecasts were also created every 15 minutes using a time-weighted bias correction technique and recent power observations (Figure 5-5 (bottom)). The MOS technique can remove the systematic bias that was displayed in Figure 5-4. The improvement between the unadjusted and MOS forecasts is quite dramatic since recent power observations lower the RMSE significantly (removing up to 85% of the error in the first 15 minute forecast). The improvement is most notable in the 1- to 3-hour time frame when a persistence forecast is expected to be a good predictor. For the remaining 3 hours, ARPS (3DVAR, ADAS), WRF (MYNN, UW) and MASS members perform the best, with the HRRR and EnKF initialized forecasts (ARPS and WRF) having a substantial (2-5% of capacity) higher RMSE.

The forecast bias for the unadjusted and MOS methods is shown in Figure 5-6. A substantial amount of systematic bias has been removed by applying the MOS method, leaving the non-systematic counterpart. The

most dramatic bias reduction is for the HRRR forecast. The unadjusted HRRR has a very positive high bias while the HRRR MOS method has a slightly negative bias. Overall, the magnitude of the bias for all unadjusted methods is greatly reduced by MOS.

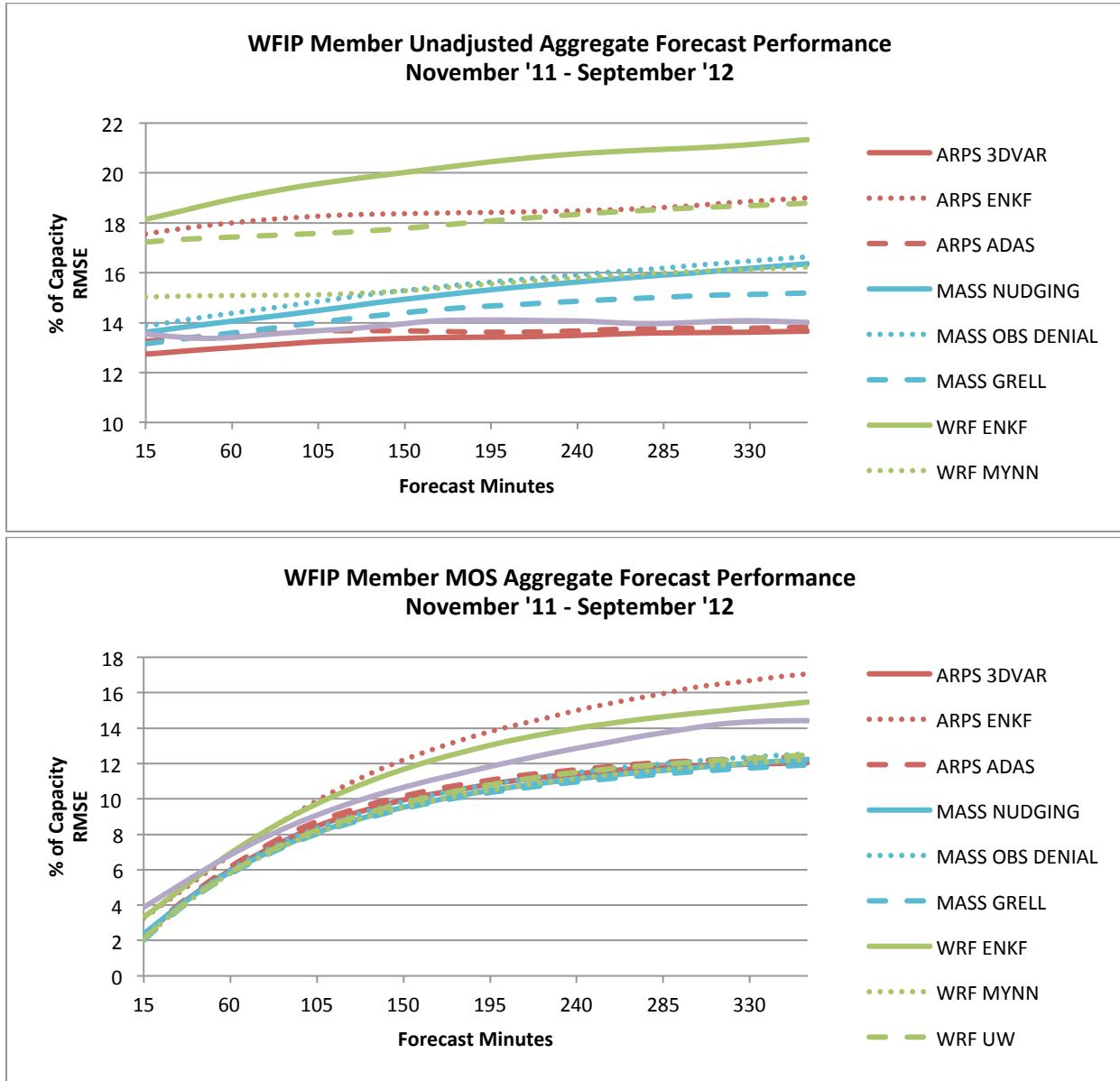


Figure 5-5: Forecast system aggregate power RMSE as a percent of nameplate capacity for unadjusted (top) and MOS (bottom) 15-minute forecast methods for each of the high-resolution WFIP NWP members.

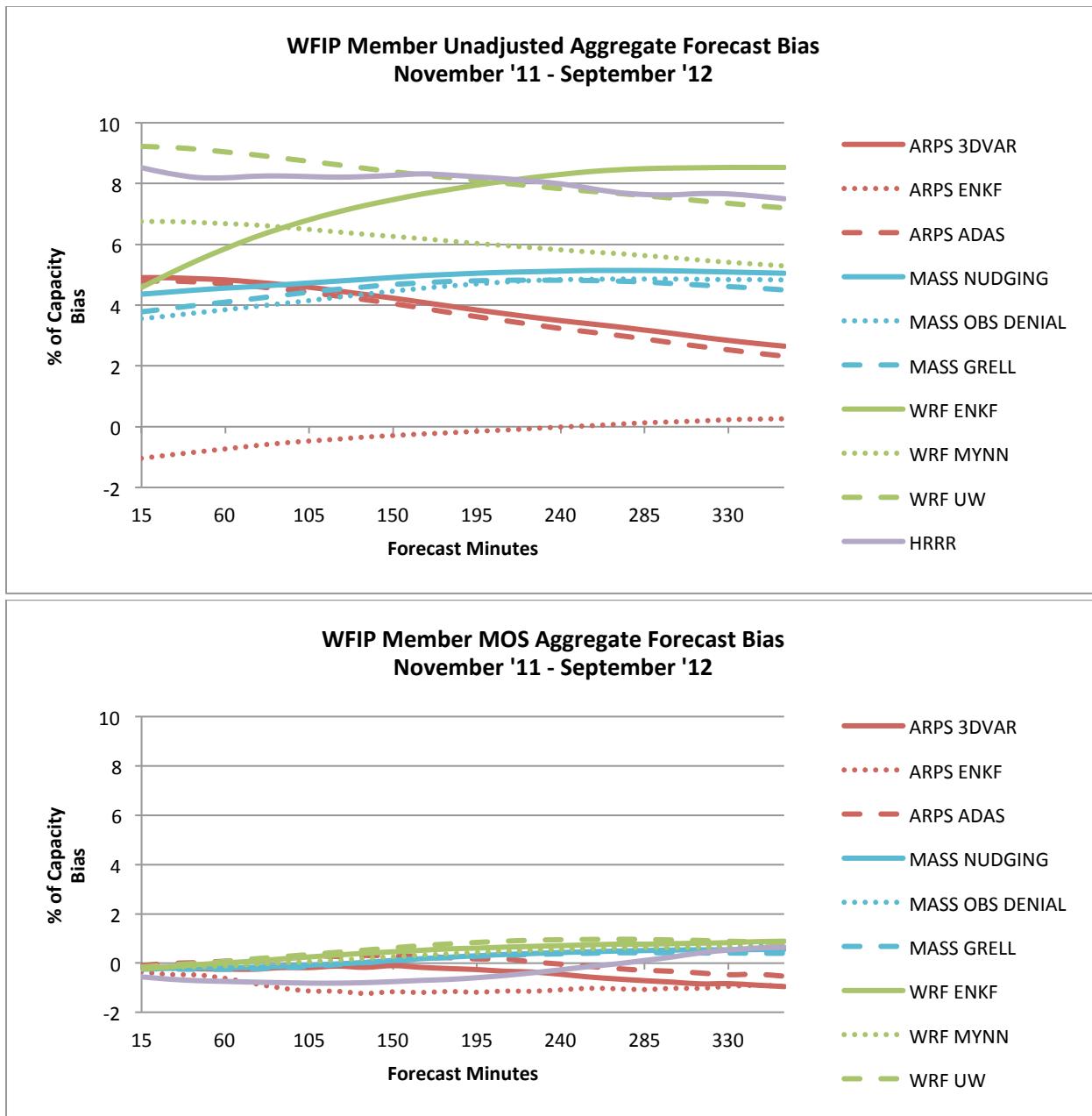


Figure 5-6: Forecast system aggregate power bias as a percent of nameplate capacity for unadjusted (top) and MOS (bottom) 15-minute forecast methods for each of the 10 high-resolution WFIP NWP members.

Table 5-1 summarizes the forecast error of the 6-hour forecast for each of the three 15-minute forecast methods. “Unadjusted” clearly has the highest forecast error as no bias correction is applied. The PA method, which applies a very basic bias correction, does not perform as well as the more dynamic, time varying bias correction method which is used in MOS. Overall, the HRRR member has the lowest forecast error when the PA method is applied while the MASS GRELL and WRF MYNN members have the lowest forecast errors when the MOS method is applied.

Table 5-1. Time averaged aggregate ensemble member performance for three forecast methods. Error is RMSE as a percent of nameplate capacity of the aggregate. Larger errors are denoted with darker red shading. The RF MYNN PA method was unavailable during the forecast project due to an error in the configuration software.

Time Averaged % capacity aggregate RMSE by model and method November '11 – September '12			
	Unadjusted	PA	MOS
ARPS 3DVAR	13.35	10.49	9.52
ARPS ENKF	18.39	11.86	12.31
ARPS ADAS	13.65	10.85	9.76
MASS NUDGING	15.14	10.58	9.34
MASS OBS DENIAL	15.45	10.82	9.61
MASS GRELL	14.45	10.48	9.17
WRF ENKF	20.14	unavailable	11.60
WRF MYNN	15.55	13.62	9.31
WRF UW	18.03	11.70	9.53
HRRR	13.87	10.07	10.87

5.1.2 Ensemble forecasts.

The individual forecasts were combined to generate an ensemble forecast for each wind plant, regional aggregate, and the system-wide aggregate. The ensemble forecast algorithm, known as OPTENS, weighted each of the three methods from each of the 10 different NWP models based on the relative performance over the previous month.

Figure 5-7 shows the improvement of the WFIP OPTENS versus the OPTENS_STWPF baseline forecast. The greatest error reduction occurs in the first 90 minutes of the forecast. Beyond 90 minutes, the forecast error reduction steadily decreases but improvement persists out to 6 hours. The improvement is thought to be due to the more accurate high resolution WFIP models (with little impact from the assimilation from project observations, discussed in section [Error! Reference source not found.](#)).

To determine the value in adding the HRRR to the ensemble, several historical ensemble forecasts were generated to include the HRRR (without dynamic weighting as discussed in section 4.1.1). The forecast performance for each ensemble forecasts is shown in Figure 5-8. The best forecast is obtained using a dynamically weighted ensemble of all WFIP members (OPTENS_AGG) with the other WFIP methods having a slightly higher error (~0.5% of capacity over the OPTENS_AGG method). The baseline (OPTENS_STWPF) has the highest error (8-12% of capacity) even though it is dynamically weighted. The exclusion of high resolution WFIP members degrades the forecast significantly. This result highlights the improvement in forecast performance by running a high resolution NWP ensemble. There was very little difference in the performance of the ensemble forecasts with and without the HRRR. This use of dynamic weighting with HRRR may have shown better performance in the ensemble.

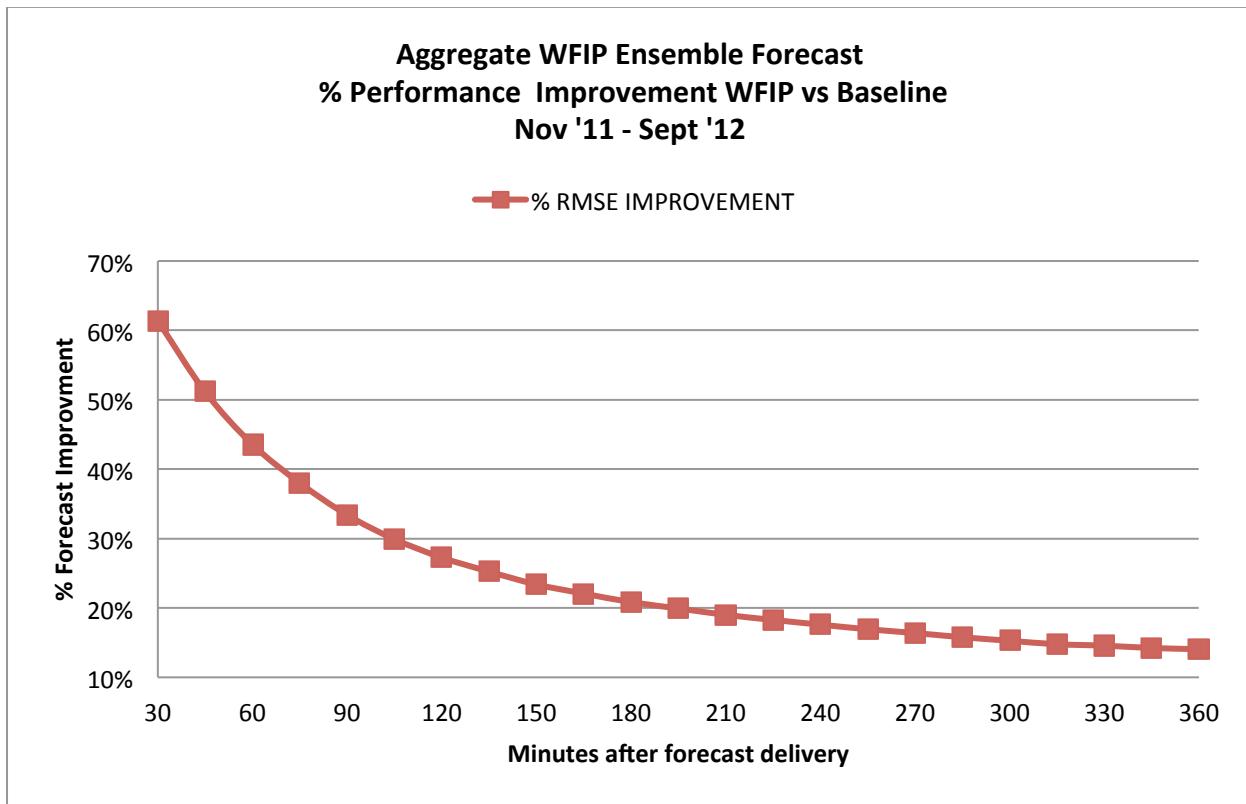


Figure 5-7: Comparison of the percent improvement in forecast performance as a function of forecast look-ahead time using the OPTENS versus OPTENS_STWPF (baseline) method. The values represent the percent reduction in RMSE of the OPTENS_WFIP over that of the OPTENS_STWPF method.

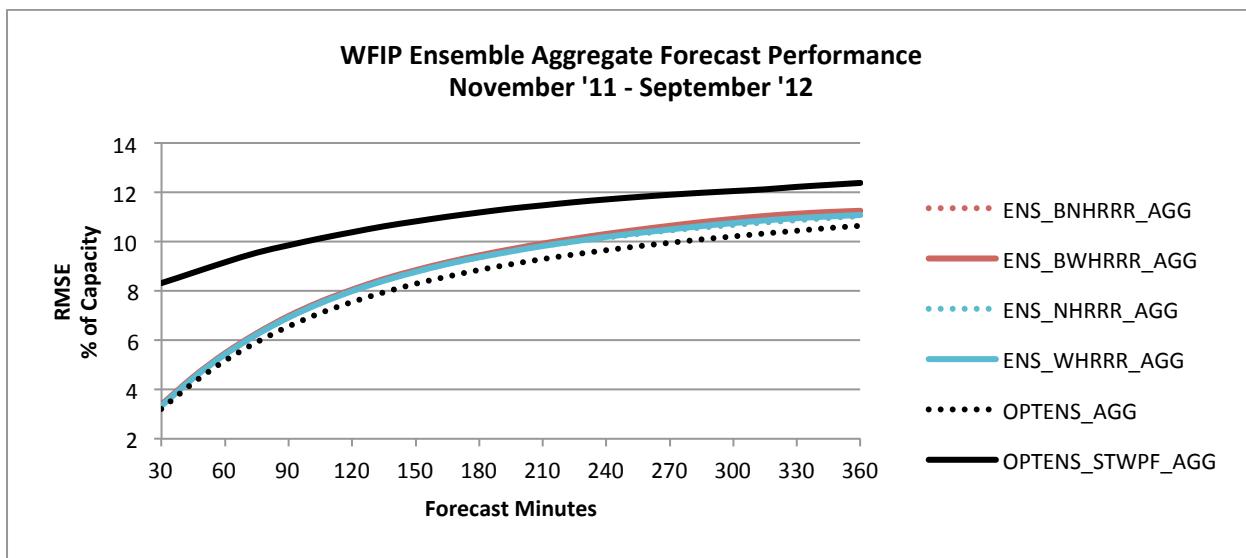


Figure 5-8: RMSE (% capacity) of six ensemble forecasts of the WFIP system-wide aggregate power production (discussed in Chapter 4.1.1 and Table 4-3).

Taylor diagrams. One way to graphically summarize how closely the ensemble performance (individual members and the optimized forecast) matches the observations is through a Taylor diagram (Taylor, 2001). The similarity between the ensemble performance and observations is quantified in terms of their correlation (or coefficient of determination— R^2), their centered root-mean-square difference, and the amplitude of their variations (represented by their standard deviations). Thus, Taylor diagrams can be especially useful in evaluating multiple aspects of model performance in a phase/amplitude space. In Figure 5-9, the black asterisk (observations) is the benchmark to which the model member/ensemble performance is measured. Note that the individual model members (un-optimized—no statistical post

WFIP (Entire Domain) 3-hr Forecasts (Cap. Fac. %) Versus Observations

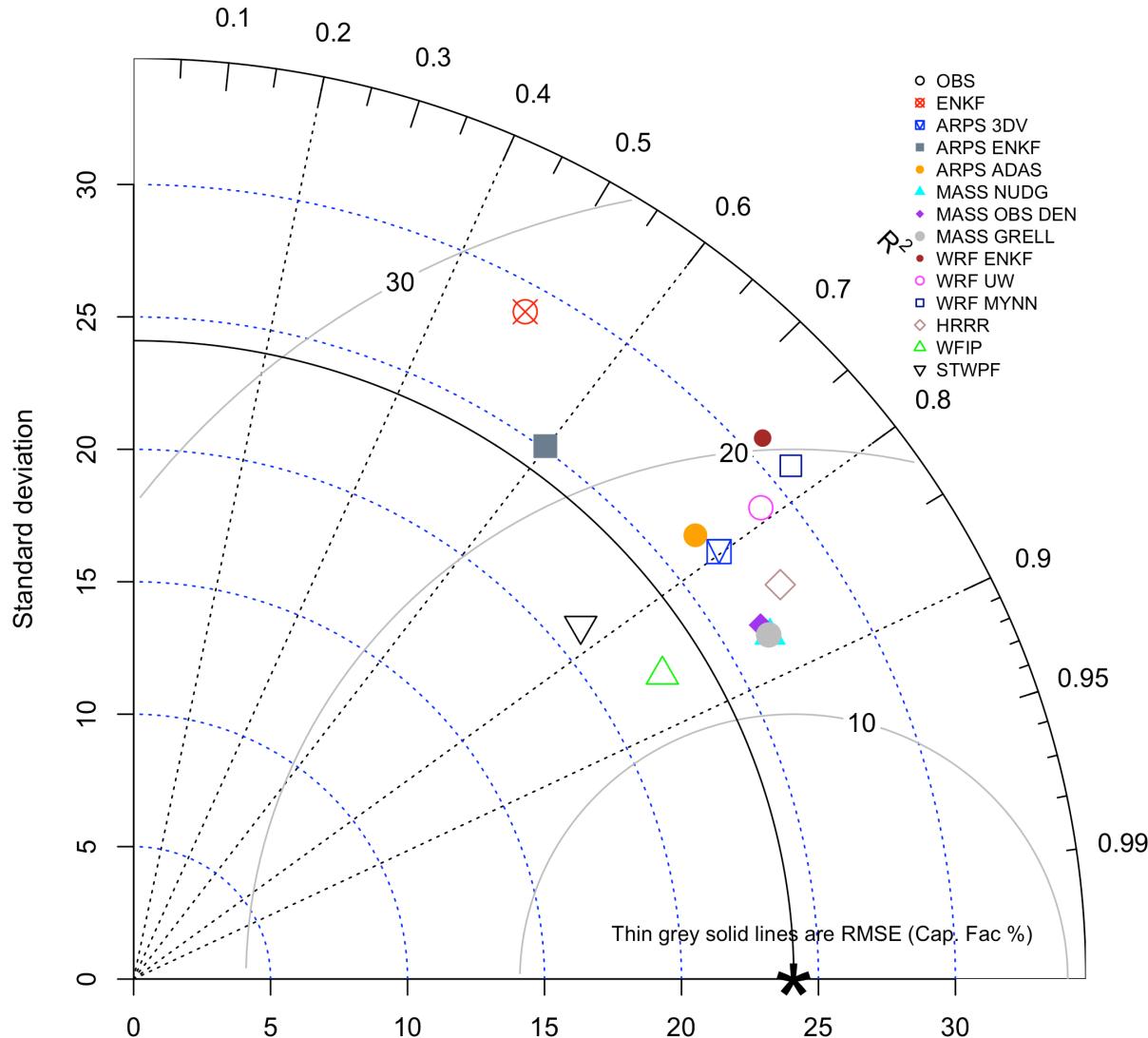


Figure 5-9. Taylor diagram showing individual ensemble member, STWPF (open upside down black triangle), and WFIP (optimized ensemble, depicted by the open green triangle) 3-hr forecast performance as compared with observations (black asterisk). Thin grey solid lines represent the centered RMSE, black dotted lines depict the coefficient of determination (R^2), and blue dotted lines show the standard deviation.

processing) show considerable scatter in the phase/amplitude space, with the MASS and HRRR members performing best. There is also significant increase in overall skill shown by the WFIP ensemble as compared with the baseline STWPF, with definitive movement towards minimizing RMSE, increasing R^2 , and capturing observational variability.

5.1.3 Probabilistic Ramp Rate Forecasts

The deterministic ramp forecasts discussed previously cannot communicate the likelihood of occurrence or likelihood of different ramp event scenarios. Therefore, 6-hour probabilistic ramp event forecasts were created every 15 minutes. The probabilistic ramp event forecasts contain the probability of exceedance for several ramp rate thresholds and a probability distribution of ramp rates. These probabilistic ramp forecasts were compared to an operational ELRAS ramp forecasting system based on the ARPS model.

A reliability diagram shown in Figure 5-10 compares the predicted ramp rate probabilities to the observed frequency of occurrence. For most of the distribution, the forecasted probabilities are similar to the expected frequency with two exceptions. First, in the 20-60% probability range the probability forecasts of most members show a high bias of about 10%, forecasting higher probabilities than the observed event frequency. Second, the ARPS members show a high bias in the 90-100% probability bin. These results indicate that for the most part, the probabilistic ramp event forecasting system generates reasonable probabilities within 10% of the actual observed frequency.

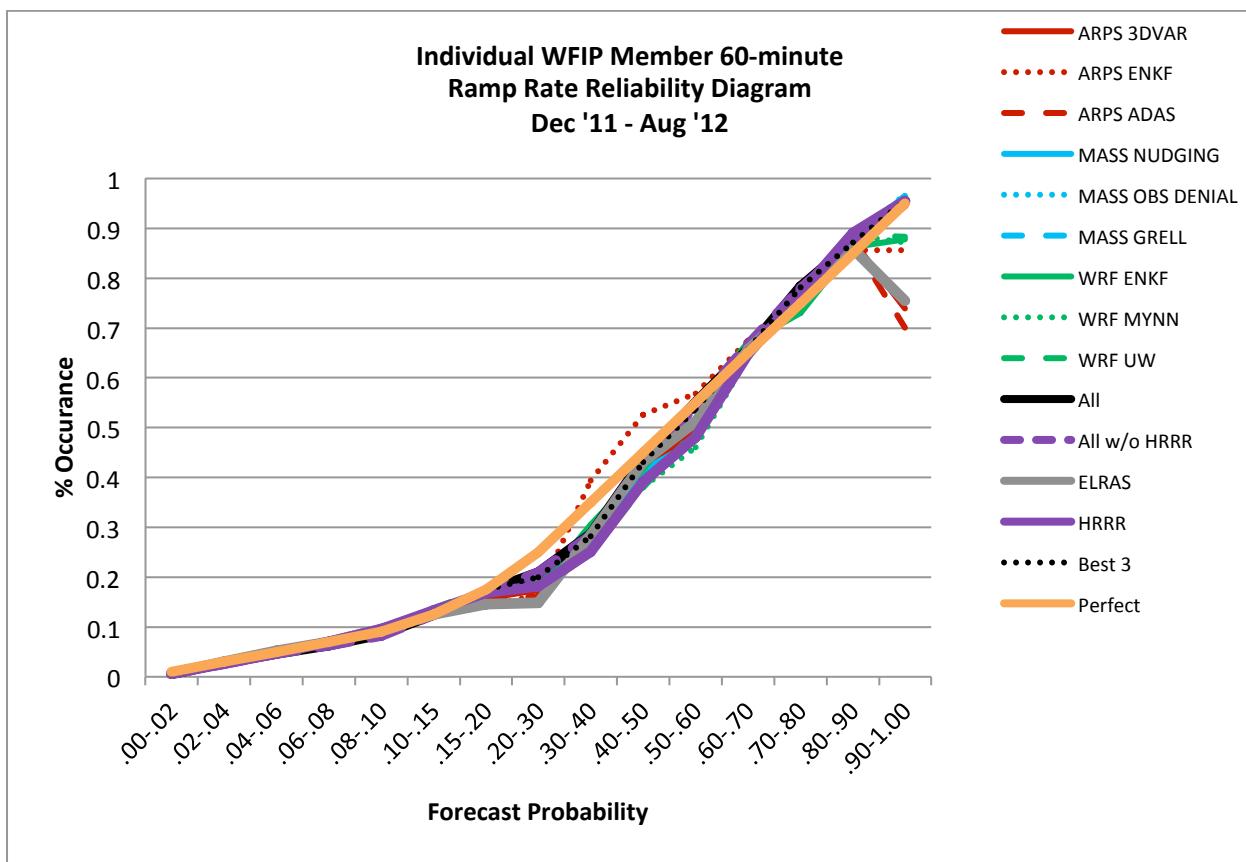


Figure 5-10: Reliability diagram for all individual WFIP member ramp forecasts, and ramp forecasts generated from a combination of WFIP forecast members (all, best 3, all w/o HRRR). For reference, a perfect forecast is shown in orange.

The probabilistic ramp forecasts were also verified using the RPSS to determine how well each of the probabilistic ramp forecasts performed compared to climatology. The results show that the ensemble forecast methods produced a more accurate probabilistic forecast of 60-minute ramp rates than any one of the single forecast member methods (Figure 5-11). There is on average a 20% improvement in RPSS (forecast skill) of the ensemble forecasts ('Best 3', 'All') over the best performing single member methods (i.e. HRRR, MASS). This result highlights the additional value from using an ensemble to generate a probabilistic ramp forecast. Of the single member methods, the HRRR probabilistic ramp forecast performed the best followed by the MASS and WRF (MYNN, UW) forecast members. The ARPS (3DVAR, ADAS) and EnKF members (ARPS, WRF) performed poorly, mostly due to the higher false alarm rate. The baseline ELRAS also performed poorly when compared to the other WFIP members (grey line in Figure 5-11).

The RPSS results shown in Figure 5-11 help determine the predictive skill of the ensemble forecast system based on its members for a particular ramp rate period and look-ahead time but do not differentiate forecast skill between varying ramp rate thresholds. The RPSS contribution (discussed in section 5.1) for 60-minute ramp rates at all look-ahead times is shown in Figure 5-12**Error! Reference source not found.**. For forecasted events without verifying ramps, the RPSS is negative indicating that climatological ramp probabilities are a more accurate forecast in persistent conditions. The largest contributions to RPSS occur for larger up and down ramp events (750-2000 MW/hour), indicating greater forecast skill in predicting ramp events of this magnitude as compared to ramp events of smaller and larger magnitudes. The contribution is still positive, but much smaller for the largest binned ramp rates (>2000 MW/hour), most likely due to the small sample size compared with other bins. The ensemble ramp forecasts tend to perform better than a ramp forecast generated using a single NWP model as noted by the higher RPSS contribution scores for the 'Best 3' and 'All' forecasts as compared to the other single model members.

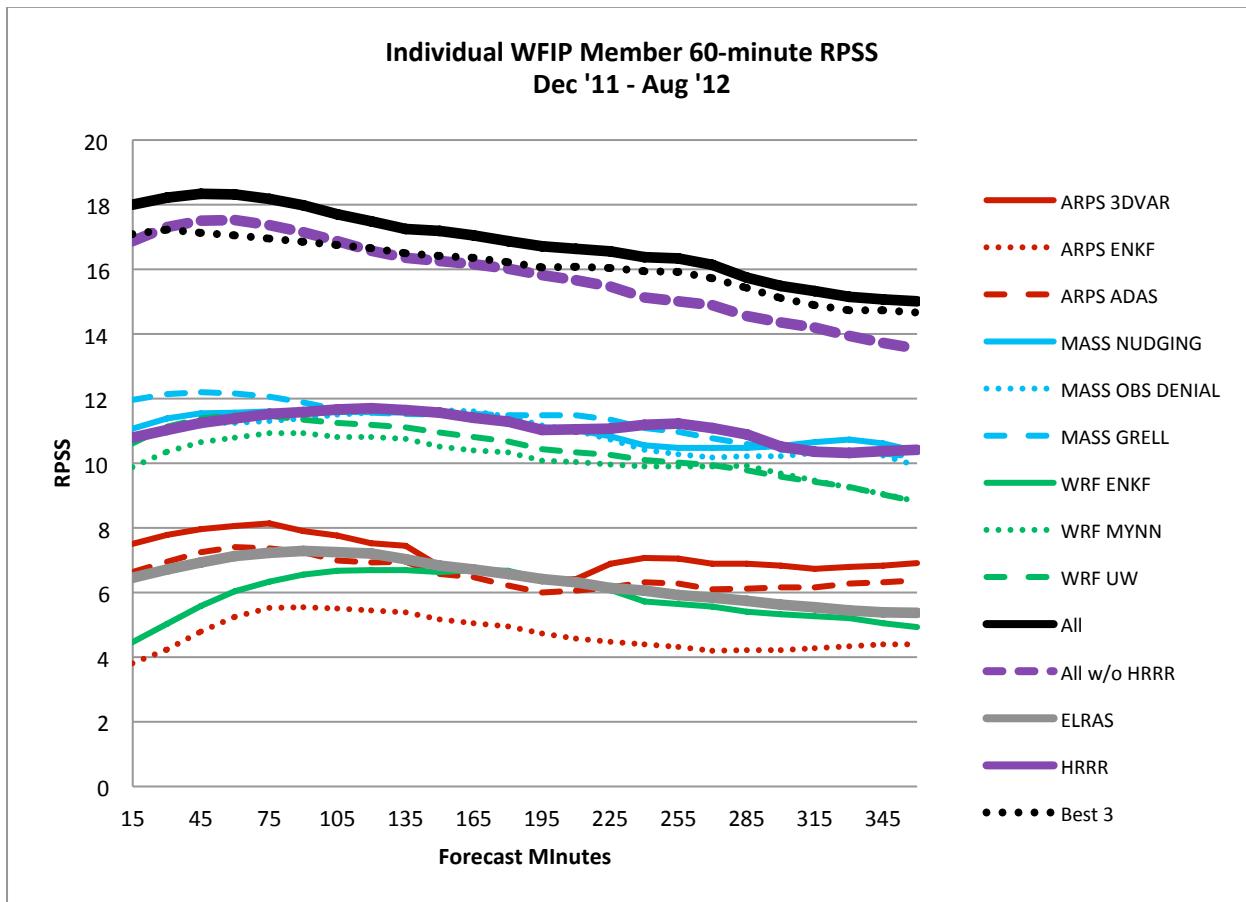


Figure 5-11: Ranked Probability Skill Score (RPSS) for all individual WFIP member ramp forecasts, and ramp forecasts generated from a combination of WFIP forecast members (all, best 3, all w/o HRRR).

5.1.4 Data denial experiments.

To determine the impact of project observations on forecast performance, several additional historical forecasts were performed for two 30-day periods: one in December 2011 and the other in April 2012. The forecast performance based on RMSE for WFIP system-wide aggregate power production forecasts is shown separately for December (Figure 5-13, top) and April (Figure 5-13, bottom). It is important to note that while the type and amount of data assimilated varied among the experiments, the RR dataset was used for ICs and BCs in all of the experiments and the RR runs assimilated all project data.

For these experiments, there is only a marginal impact from project observations. This result is most evident by comparing the run that included the assimilation of the project data (ARPS All Obs) with the run that did not include the assimilation of these data (ARPS Obs Denial). The improvement is most prominent in the first 30 minutes after initialization time (ARPS All Obs RMSE decreases 2% of capacity). These results also are also similar for aggregate wind speed verification (Figure 5-14).

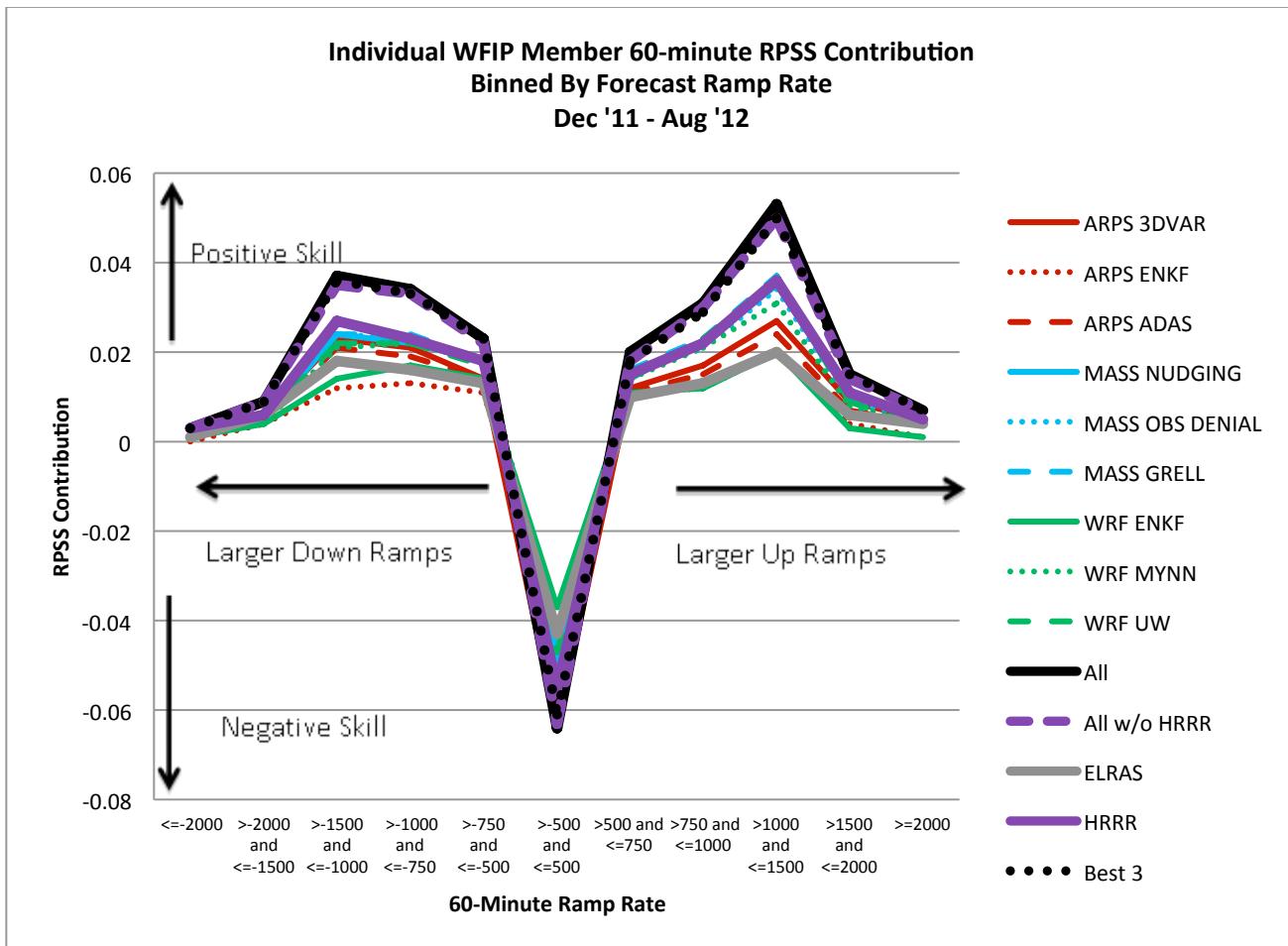


Figure 5-12: Ranked Probability Skill Score (RPSS) contribution averaged over all forecast times shown by 60-minute ramp rate. RPSS contribution is shown for all individual WFIP member ramp forecasts and ramp forecasts generated from a combination of WFIP forecast members (all, best 3, all w/o HRRR). Observed ramp magnitude values along the x-axis are in MW/hour units.

Data from instrumentation deployed for WFIP in the southern study area produced only marginal improvements in power production forecasts. This minimal impact may be caused by a number of factors including the weather regime, assimilation system, and configuration of the NWP model. The project observations would likely have more impact if assimilated using regime-dependent instead of isotropic structure functions in the NWP modeling systems. Such functions vary for specific weather events and more accurately represent how observations influence the analysis of model parameters at nearby grid points (discussed later in this section).

An additional ESRL historical reanalysis period was also run using a set of RR runs for BC and IC that did not assimilate project observations. The ARPS model was run in two different configurations: 'ARPS All Obs' assimilated all observed data and 'ARPS Obs Denial' assimilated only non-project observations. The key difference for these runs was the use of the RR model for ICs and BCs that did not assimilate project data. The forecast errors for the WFIP system-wide aggregate power production and wind speed forecasts are shown for the first week of December (Figure 5-15). The results are very similar to that of the previous December 2011 and April 2012 historical runs showing little to no impact of project observations on the forecast performance for the period.

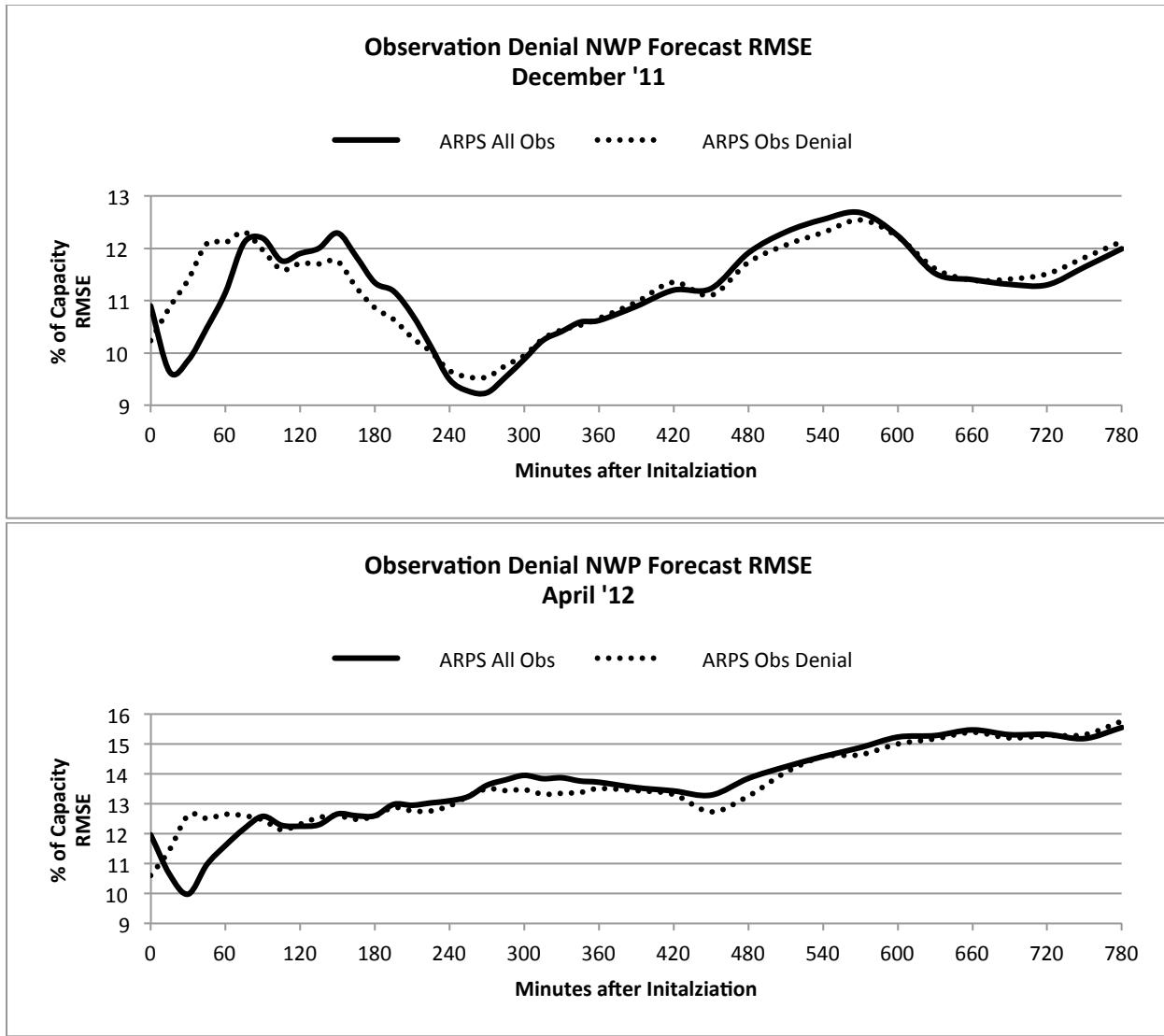


Figure 5-13: The RMSE of power production forecasts for the WFIP system-wide aggregate for the AWST data denial experiments for a 30-day period (top) December and (bottom) April.

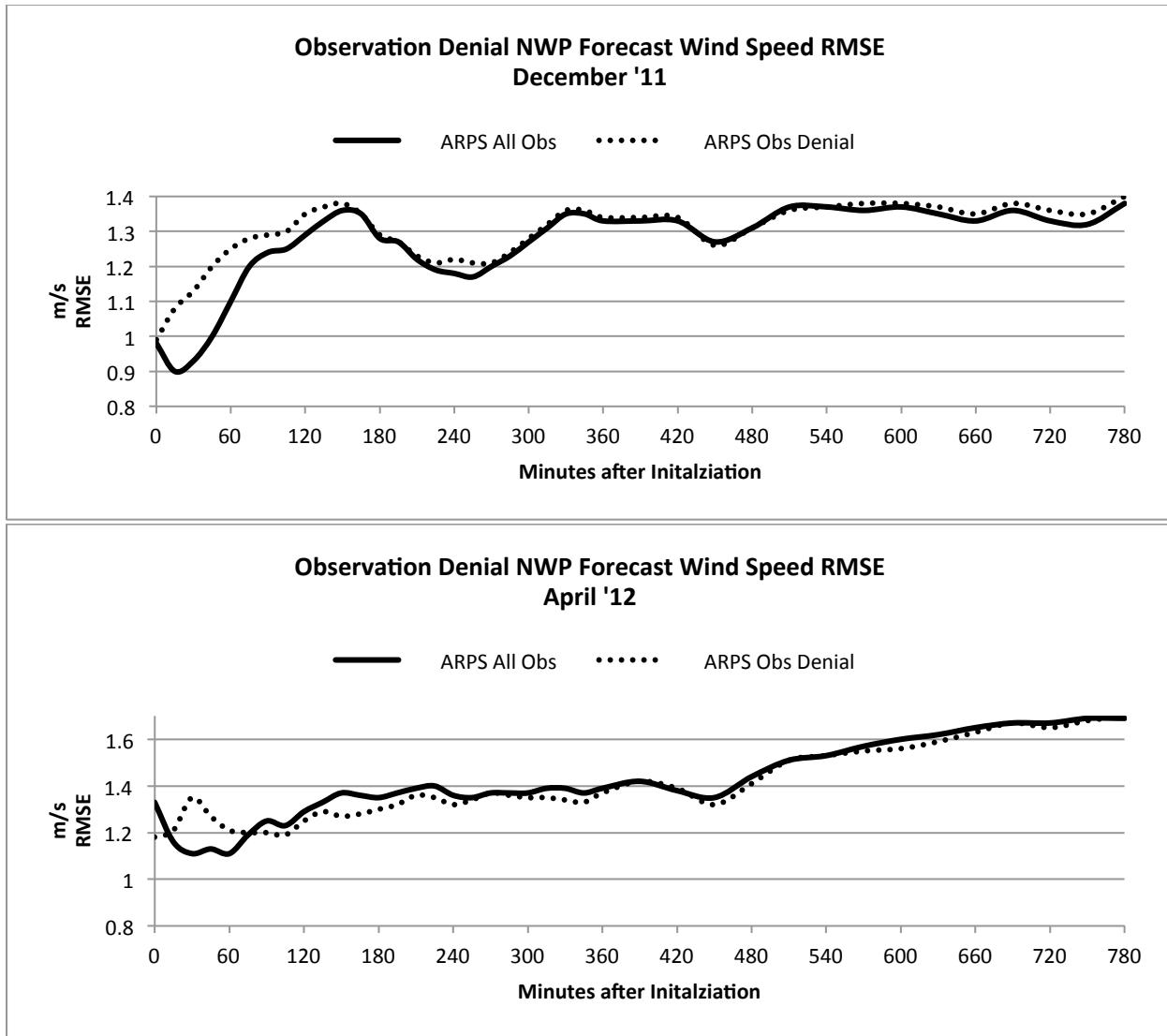


Figure 5-14: The RMSE of capacity weighted average wind speed forecasts for the WFIP system-wide aggregate for the AWST data denial experiments for a 30-day period (top) December and (bottom) April.

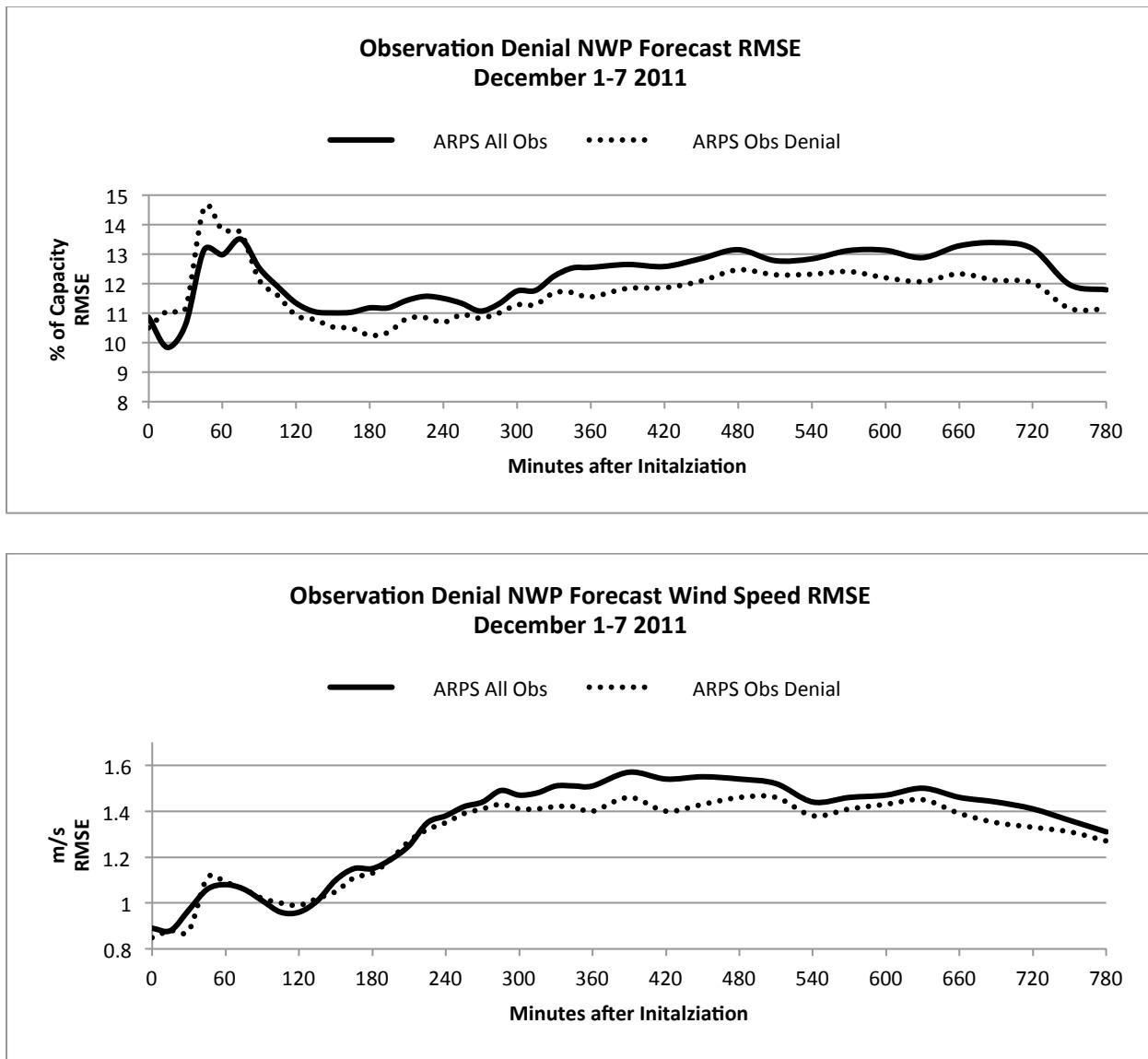


Figure 5-15: RMSE of forecasts of the WFIP system-wide aggregate (top) power generation (% of capacity) and (bottom) wind speed (m/s) for data denial experiments that included denial of project data to the ESRL RR data assimilation process for one week in December 2011.

5.1.5 Ramp event analysis

To determine the impact of project observations on specific ramp events, 20 ramp cases were compiled for a 6-month warm season period and a 6-month cold season period (Table 5-2). The ARPS model was run in rapid update mode with 13-hour forecasts generated every 2 hours beginning 6 to 8 hours before each ramp event. This strategy yielded several sets of overlapping runs for each ramp case.

For each ramp event, several experiments were performed: (1) a forecast in which ARPS assimilated all observations and (2) a forecast in which ARPS assimilated all observations except project observations. A 9-hour deterministic ramp forecast was generated at 15-minute intervals for the entire forecast period for each experiment. The CSI was computed to determine impact of observation assimilation on the performance of deterministic ramp forecasts.

Table 5-2: List of days covering each ramp event during the winter and summer periods. An asterisk (*) denotes that more than one ramp event occurred during the listed date.

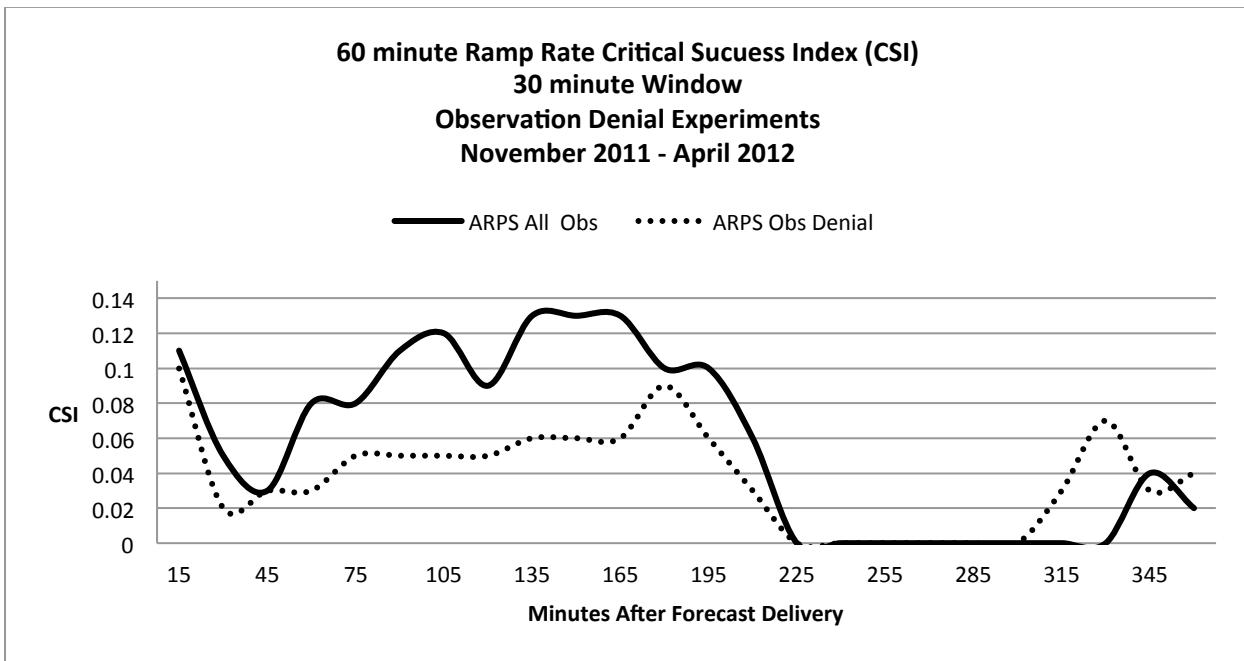
Cool Season Cases	Warm Season Cases
<i>2 November 2011*</i>	<i>31 March 2012</i>
<i>8 November 2011</i>	<i>7 April 2012</i>
<i>16 November 2011*</i>	<i>15 April 2012*</i>
<i>3 December 2011</i>	<i>17 April 2012</i>
<i>26 December 2011</i>	<i>25 April 2012</i>
<i>30 December 2011 – 1 January 2012*</i>	<i>6 May 2012</i>
<i>28 January 2012</i>	<i>20 May 2012</i>
<i>29 January 2012</i>	<i>16 June 2012</i>
<i>20 February 2012 – 21 February 2012*</i>	<i>15 July 2012</i>
<i>8 March 2012</i>	<i>7 September 2012</i>

5.1.6 Ramp Verification

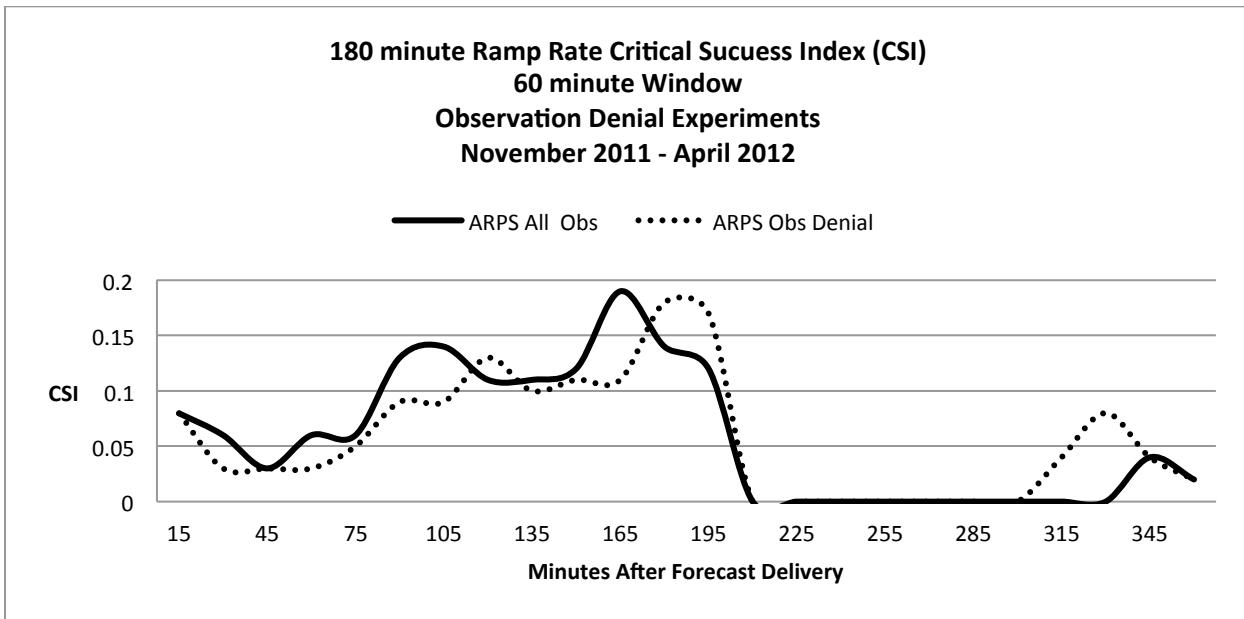
Although the data denial experiments indicated that the assimilation of project observations had very small impact on the wind speed and power production forecasts over the entire data denial experiment period, it was still possible that the assimilation of observations may improve the skill of ramp event forecasts. Therefore, deterministic ramp forecasts were created for several observed high magnitude ramp events occurring within two 5 month periods (Table 5-2). Then, CSI results were generated to determine if there was an improvement of the deterministic ramp event forecasts from the assimilation of project observations.

The results for the winter and summer ramp cases are shown in Figure 5-16. For both periods, there is an improvement in the CSI for the 1- to 3-hour forecast when project observations were assimilated (ARPS All Obs) compared with no assimilation of project observations (ARPS Obs Denial) for both 60-minute (Figure 5-16a) and 180-minute (Figure 5-16d) ramp rates. The improved CSI scores (40% for previous examples) for the ARPS All Obs forecast highlights the importance of project observations to ramp prediction in the 1- to 3-hour forecast time, although, other periods and ramp rates did not show conclusive results.

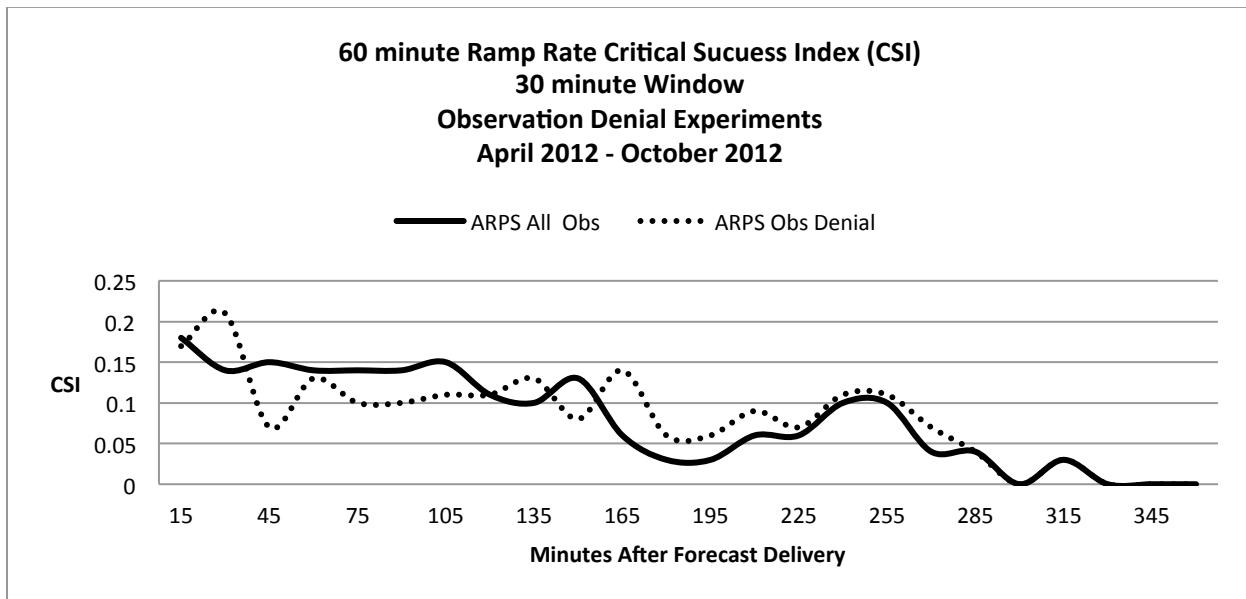
Overall, the improved CSI results in the 1- to 3-hour forecast time frame when the project data were assimilated are in contrast to the impact of the assimilation of project data on the wind speed and power time series forecasts shown in the previous subsection.



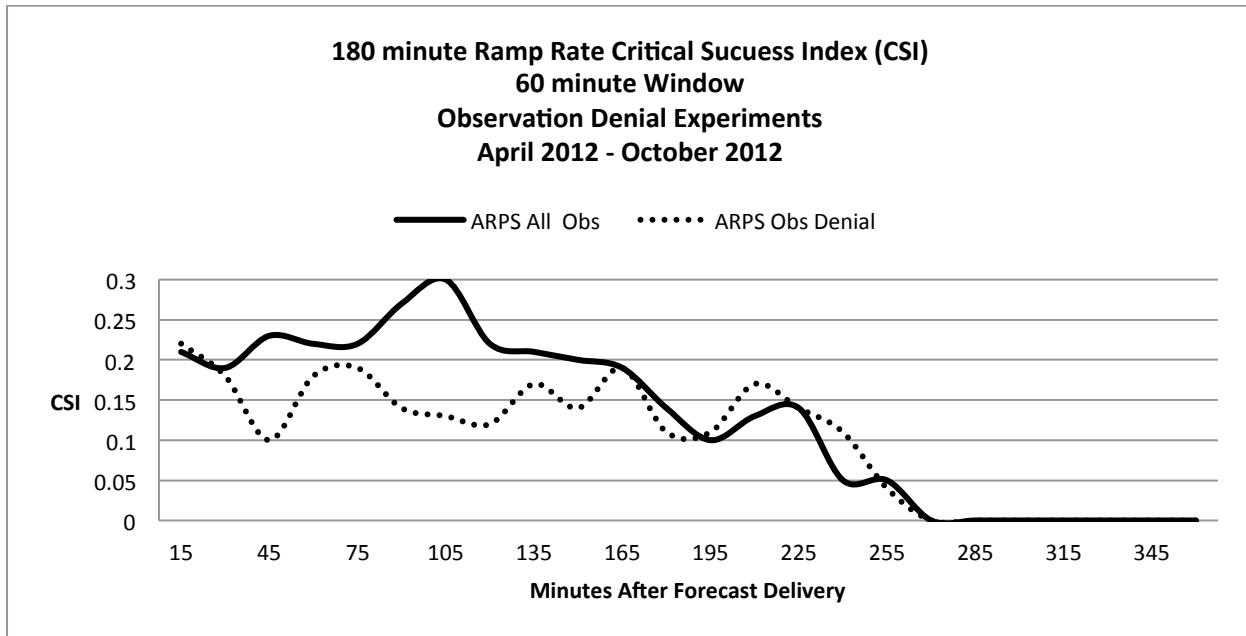
(a)



(b)



(c)



(d)

Figure 5-16: Critical Success Index (CSI) for ramp event forecasts for the November 2011 – April 2011 (panels a and b) and April 2012 – September 2012 (panels c and d) periods. Panels (a) and (c) are for 60 minute ramp rates using a 30-minute “hit” window and panels (b) and (d) are for 180 minute ramp rates using a 60-minute “hit” window.

5.1.7 Phenomena-based model performance

The impact of atmospheric phenomena on model performance can be inferred from diurnal and seasonal evaluation of forecast error. In the present analysis, we focus on two aspects of power forecast modeling: 1) the ability of individual atmospheric model members to accurately predict instantaneous wind speeds, and 2) the power generation forecast of the optimized ensembles.

Shown in Figure 5-17 is the diurnal forecast speed mean absolute error (MAE) of the WFIP ensemble members for various forecast hours. Subsets of this data are presented for three seasons: NDJF (November – February), MAMJ (March – June), and JASO (July – October). These three seasons correspond to the cool season, spring transition, and warm season (selected for the dominant modes of characteristic weather disturbances). In general, we see a maximum of MAE during the overnight hours 04 UTC – 12 UTC (10 PM – 06 AM CST), with a period of rapid decline 12 – 18 UTC (06 AM – 12 PM CST). This pattern is markedly exaggerated during the transition and warm seasons.

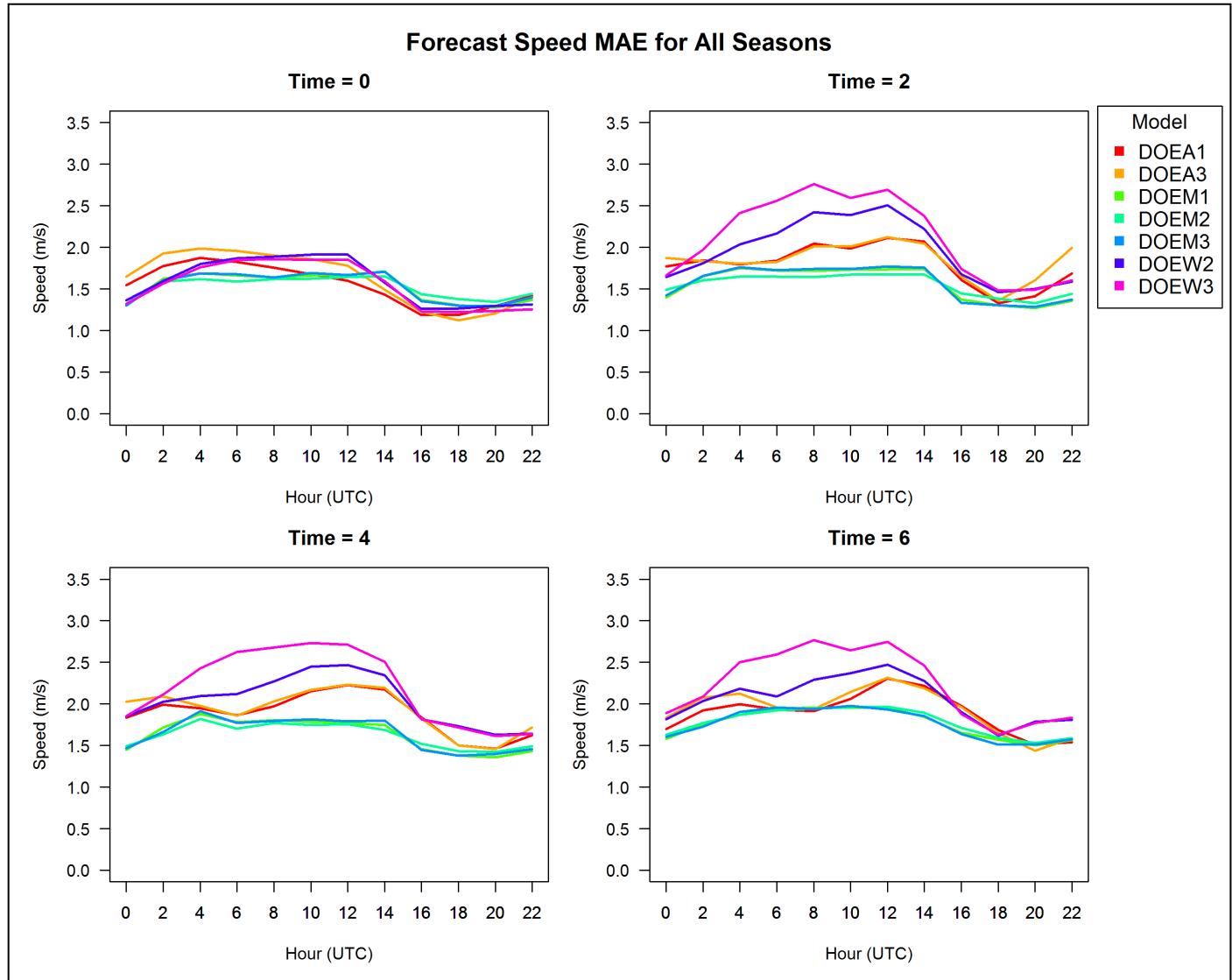


Figure 5-17. Diurnal forecast speed mean absolute error (MAE) of WFIP ensemble forecast members for varying time-ahead forecast hours.

The observed and predicted instantaneous power generation was compared for the WFIP and baseline forecasts, as is depicted in Figure 5-18 and Figure 5-19. Shown in Figure 5-18 is the model capacity factor bias by time of day and season for forecast hour 2. These results are presented for the WFIP aggregate, McCamey, Sweetwater East, and East Texas regions. Additional results of the forecast capacity factor for the WFIP aggregate region as predicted by the WFIP and baseline forecasts for various times ahead are shown in Figure 5-19. The diurnal model errors vary between the WFIP and baseline forecasts, both in absolute magnitude and distribution. The WFIP optimized forecast has overall bias closer to near-zero values, as seen by the lower interquartile range and lesser outlier distribution. Although both optimized models show a general over prediction of regional capacity factor values 10 – 14 UTC (4 – 8 CST), the baseline forecast also exhibits an under prediction of capacity factor values 0 – 4 UTC (particularly during the transition season, 18 – 22 CST). In addition, the baseline forecast tends to have extreme outliers in the bias distribution around noon and during the evening hours (MAMJ and NDJF).

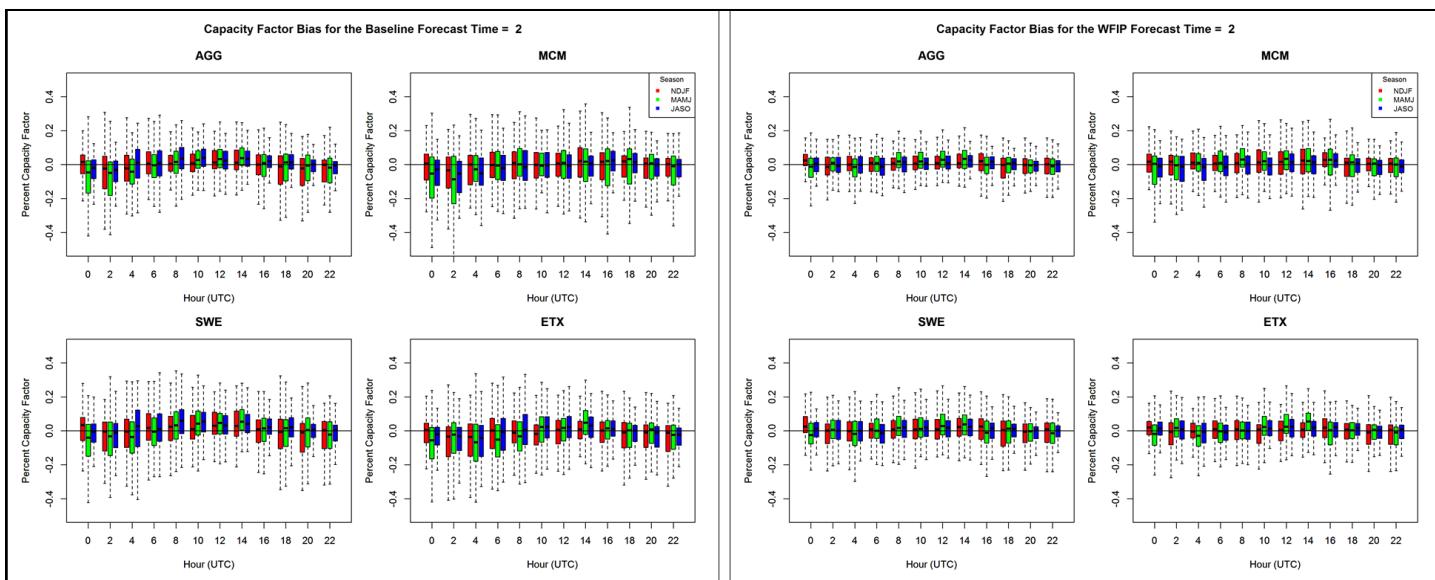


Figure 5-18. Diurnal capacity factor bias for the WFIP (left) and baseline (right) forecasts, decomposed by region and season. Seasons are as follows: November, December, January, and February (NDJF, red), March, April, May, and June, (MAMJ, green), and July, August, September, and October (JASO, blue).

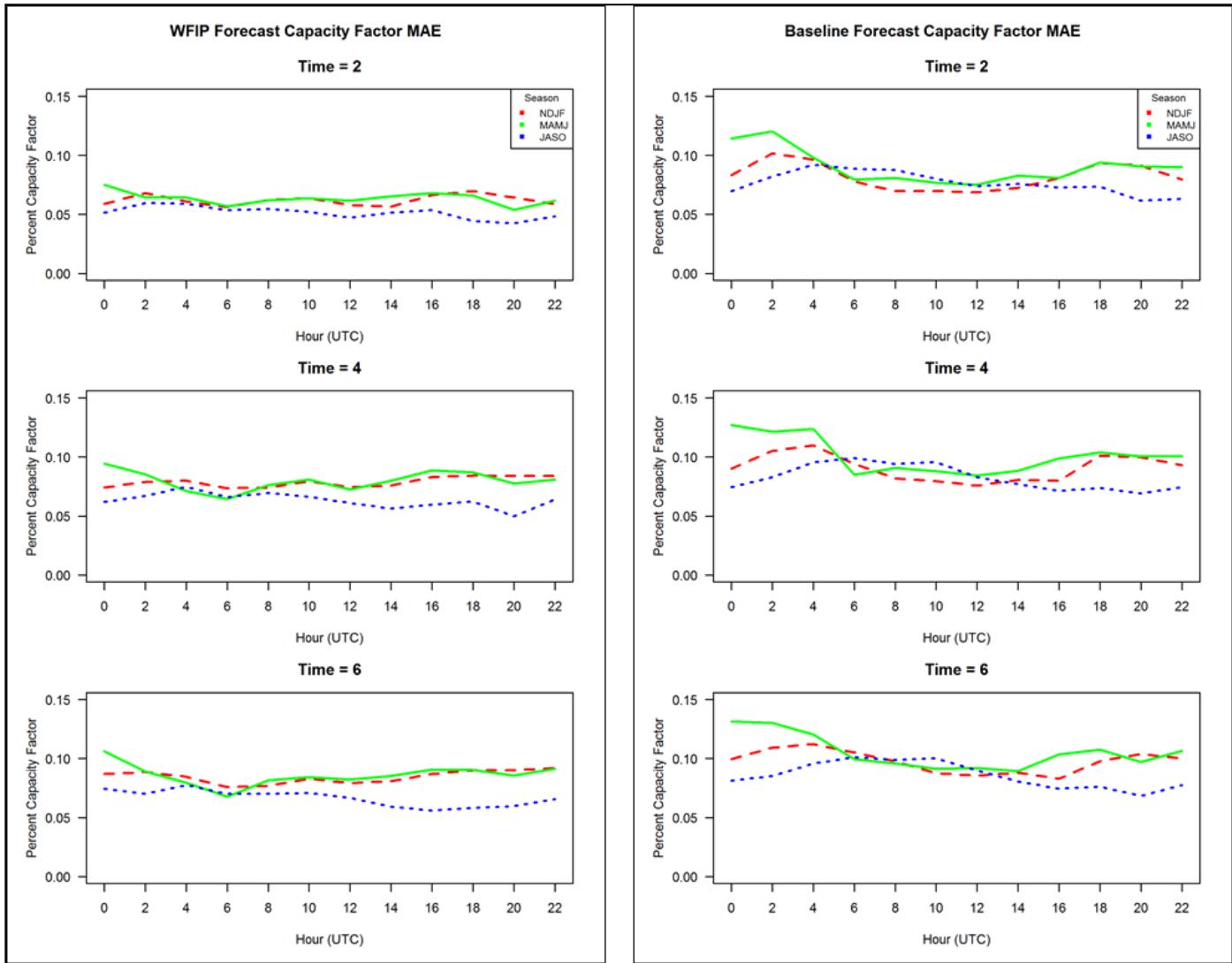


Figure 5-19. Diurnal mean absolute error (MAE) of capacity factor for the WFIP (left) and baseline (right) forecasts, decomposed by season and time-ahead forecast hour. Seasons are as in Figure 5-18: NDJF (dashed red), MAMJ (solid green), JASO (dotted blue).

The 15-minute change in capacity factor was also evaluated for the WFIP and baseline forecasts, shown in Figure 5-20 and Figure 5-21. The diurnal distribution of model errors varies between the WFIP and baseline forecasts. The baseline forecast shows a clear diurnal signal, with MAE values varying between 2-5 %. In the baseline forecast, there is strong over prediction of 15-minute ramp values at 02 – 04 UTC (08 PM – 10 PM CST) and under prediction at 14 – 16 UTC (08 AM – 10 AM CST). The WFIP forecast exhibits near-zero median bias for all hours. A relatively small and randomly distributed bias allows for less than 2% mean absolute error in predicted capacity factor, with the largest errors during the late morning, as evidenced by the peak values at 14 – 16 UTC (08 AM – 10 AM CST).

Through analysis of the diurnal and seasonal patterns of wind speed and capacity factor bias described above, one can deduce the boundary layer atmospheric processes responsible for coherent patterns in model forecast underperformance. The analysis below will focus on the following diurnal periods: overnight, mid-morning, and evening.

- Overnight. During the overnight hours, there is an increase in model speed MAE (particularly during the warm season). The time of day and seasonal fluctuation suggest that this pattern is likely associated with low-level jet development. The pattern is less readily seen in the distribution of model capacity factor from the ensemble forecasts, likely due to model output statistics (MOS) applied to optimize the forecast system.

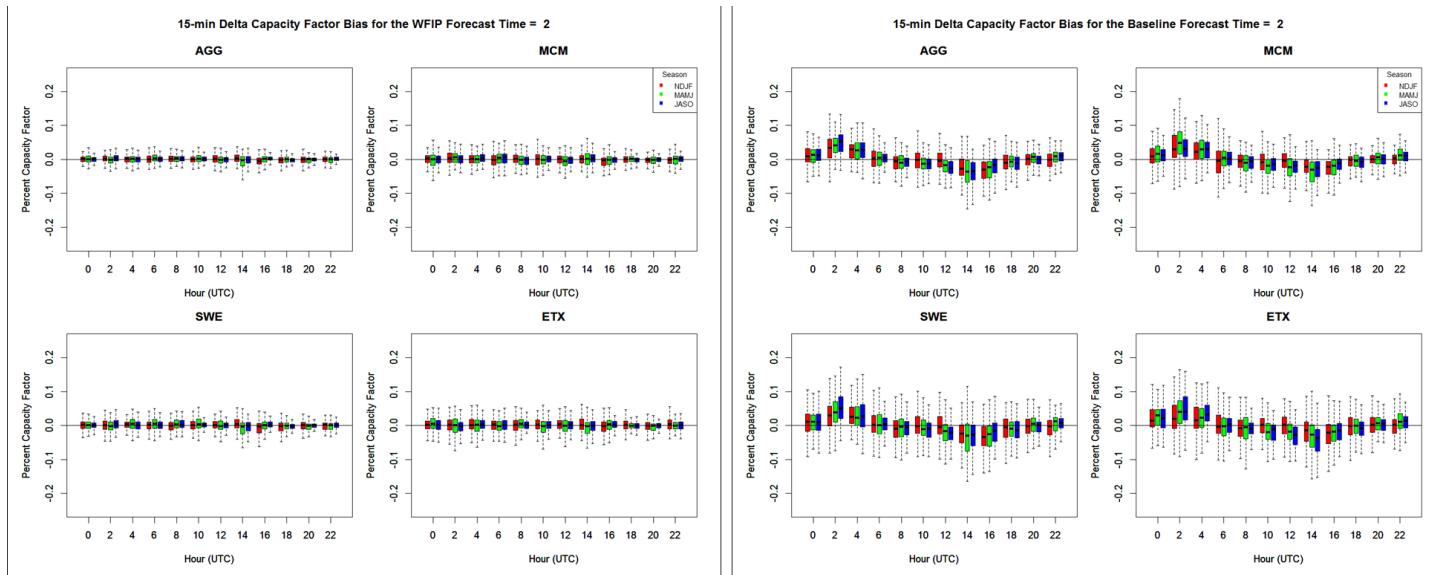


Figure 5-20. Diurnal bias of the 15-minute change in capacity factor for the WFIP (left) and baseline (right) forecasts, decomposed by region and season. Seasons are as follows: NDJF (red), MAMJ (green), JASO (blue).

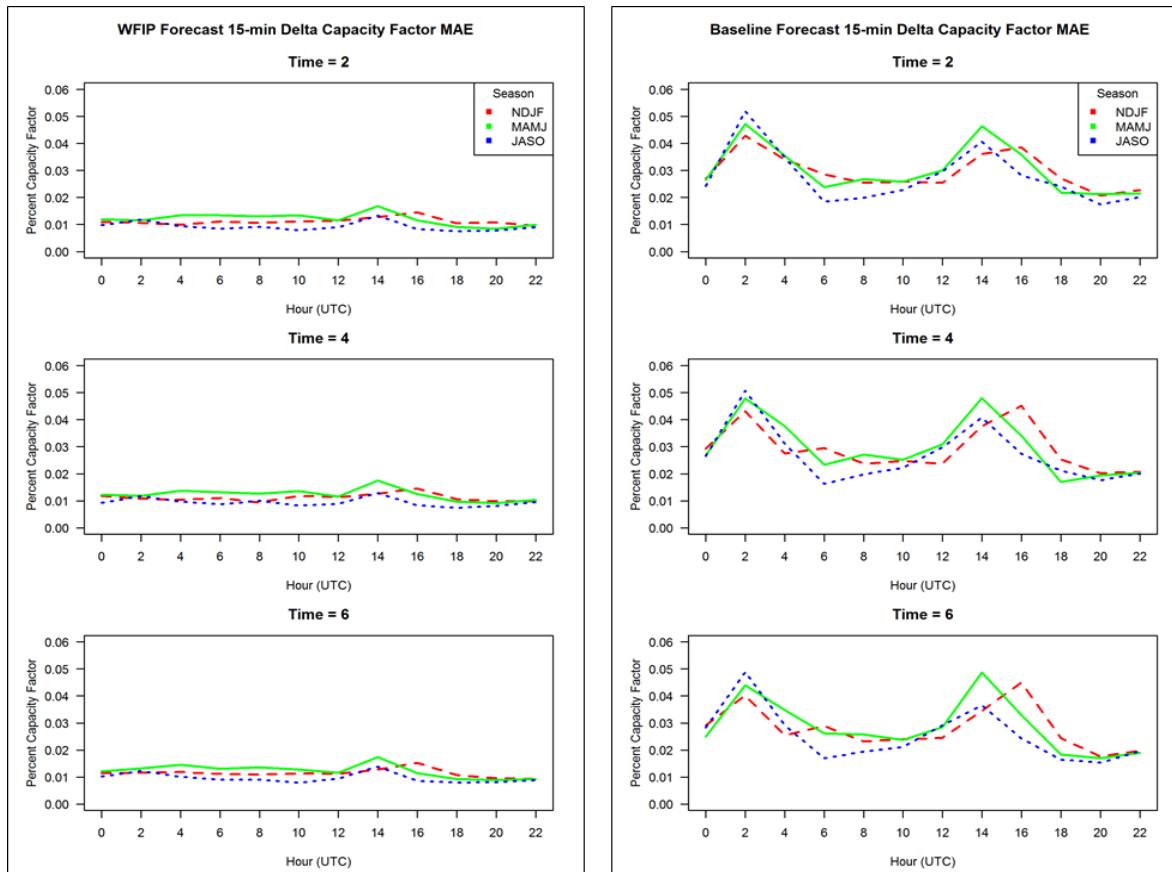


Figure 5-21. Diurnal mean absolute error (MAE) of the 15-minute change in capacity factor for the WFIP (left) and baseline (right) forecasts, decomposed by season and time-ahead forecast hour. Seasons are as follows: NDJF (dashed red), MAMJ (solid green), JASO (dotted blue).

- Mid-morning. During the mid to late morning hours, there is a sharp decrease in the forecast speed MAE, and concurrent peak in the 15-minute change in model capacity factor MAE. Also evident during this time frame is an overall high bias of model instantaneous capacity factor values. This period corresponds to the morning transition. During the morning transition, the boundary layer evolves from a thermally stable, high-shear environment to one with increasing instability and convective mixing. During this time, momentum is transferred from higher levels to nearer the surface and surface wind speeds increase. In the cold season, this boundary layer transition is shifted later in the morning hours due to a number of factors, e.g., lower sun insolation angle, later sunrise, and stronger nighttime surface inversions. These peak values of MAE during the morning transition hours show a clear seasonal trend in both the WFIP and baseline forecasts, demonstrating that the diurnal pattern of model error is phenomena-based.
- Evening. The nighttime under prediction occurs during the hours of the evening transition, a time during the first few hours after sunset in which the rate of radiative cooling is maximized. During this time, turbulent mixing becomes confined to a shallow, stable layer near the surface. It is surmised that the under prediction in capacity factor during these hours may be due to insufficient model resolution of the stable layer, within which low-level winds are still greatly affected by surface frictional effects.

, Various atmospheric processes contribute independently or in concert to the forecast bias. Overall, for temporal and seasonal regimes, the WFIP optimized forecast shows great improvement over the baseline forecast, as evinced by the dampening of forecast error magnitudes. Future work would further evaluate the model performance bias under varying physical conditions such as mean wind speed, stability, terrain complexity, and land cover.

5.2 OU CAPS

OU CAPS analysis focuses on a set of sensitivity studies for selected ramp cases. The experiment set included a configuration withholding the special WFIP surface and sounding data, and forecasts using less dense vertical grid spacing, spacing that is more typical for convection-allowing models run at this horizontal resolution. The reduced vertical spacing had 53 levels, 400 m average grid spacing and a cubic stretching function (see right pane of Figure 4-5). All forecasts in the experiment set have 3 km horizontal grid spacing, used the same operational weather observations, including radar data, and used the same data assimilation strategy as the production forecasts.

Verification for the ramp study was done using data from the TTU Tower and at the project wind SoDAR sites and to measure and visualize high-resolution vertical structure in the boundary layer important for the wind power forecasts. Verification was done at 80-m AGL, corresponding to the typical wind turbine hub height.

The analysis of the results begins by considering some sample ramp events covering the most typical types, and then examines the overall statistics over all the WFIP-selected ramp cases. In most ramp cases the results for the data denial experiment tracked closely to the production forecast, so for clarity the result of that particular run is not included in many of the figures.

An example of a ramp due to synoptic scale front is shown in Figure 5-22. This case from 30-Dec-2011 shows a large-scale down ramp over 12 hours that was fairly well predicted by the production configuration (denoted

“P”, red squares) and the reduced vertical grid spacing experiment (denoted “R”, green, circles). Data from the TTU Tower interpolated to 80 m AGL is in black. Although, in many cases the reduced-vertical resolution experiment produces slightly slower wind speeds, in this case there is variation from run-to-run, and the forecasts with the stronger wind speed seems to be correlated with a stronger initialization. The down ramp is followed by an up ramp, apparent at 18 UTC and beyond. The increase in wind speed is somewhat under forecasted by all the forecasts in the experiment set.

A sample spring case from 7-May-2012 showing multiple up- and down-ramps, shows that the observed initial small ramp well is very well predicted (Figure 5-23). There is a down ramp at the 15-hr mark that is over forecasted in the models, then there is an up-ramp that is depicted well in the forecasts initialized at 18 UTC, but was poorly handled by the 12 UTC forecasts. In operational use of such multiple forecasts (known as a time-lagged ensemble) the variation among forecasts with differing initial times could be used as an indicator of the level of uncertainty in the forecasts.

A summer thunderstorm case from 16-June-2012 is shown in Figure 5-24. This case shows a well-forecasted up-ramp that rises to more than 20 ms^{-1} in the observations. The forecasts handle short-term wind fluctuations between the 20- and 30-hr marks remarkably well, although a short-term up-ramp at 16 hours is missed. The production run with higher vertical resolution produces forecasts of higher wind speeds at peaks.

Figure 5-25 shows that the CAPS ARPS production run had the best scores over the length of the forecast period, with the 80-m wind speed MAE for all ramp cases increasing from about 2.25 ms^{-1} in the first few hours to around 3.0 ms^{-1} at 10 hr. Withholding the WFIP surface and profile data produced slightly larger average errors, about 0.05 ms^{-1} greater in the first 6 hours, then nearly the same, and the reduced vertical resolution forecasts had errors about 0.15 ms^{-1} more than the production, with the difference decreasing after about 5 hours. Similar errors and trends were measured at the SODAR sites in the network (not shown), though at some of the sites the reduced vertical resolution actually showed lower errors than the production runs in the longer forecast times (8 - 12 hr). This may be due to the distribution of vertical levels above the boundary layer being superior in the reduced vertical resolution cases thus able to handle better mid-level waves.

An examination of the *vertical extent* of forecast errors hints at possible causes in the data denial experiment. Figure 5-26 and Figure 5-27 show the mean absolute wind speed forecast error (ms^{-1}) for the CAPS ARPS WFIP run and the ARPS run with reduced vertical resolution, respectively. The verification shows there is a gradient in the error near the 80 m AGL level with greater errors at and above that level. The increase in errors aloft is reduced in the production run compared to the more typical vertical grid spacing. The more significant difference there may be due to the fact that the vertical separation of grid levels in the typical configuration (Figure 5-27) increases more rapidly there and may lose the ability to properly resolve of the top of the nocturnal boundary layer.

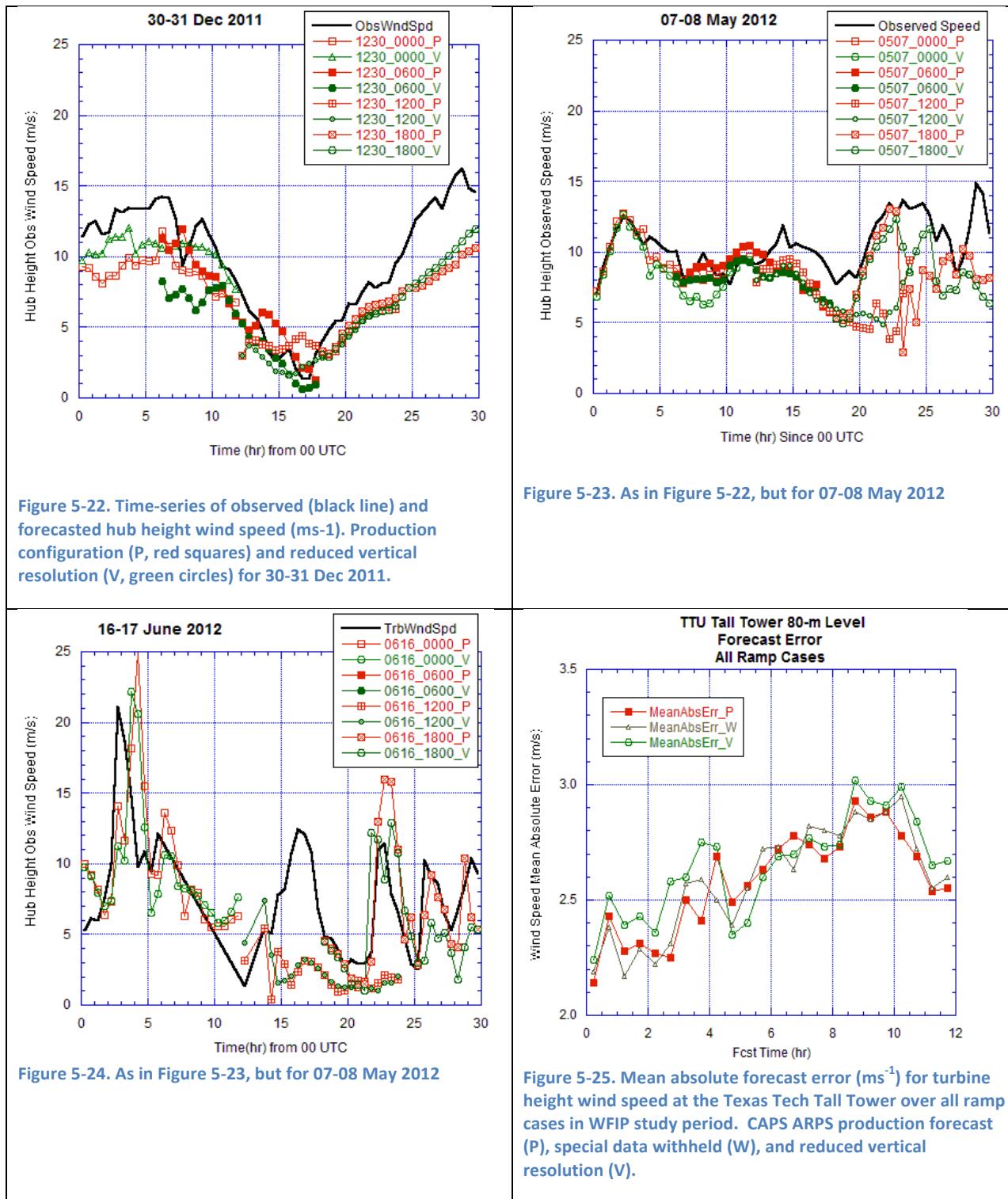


Figure 5-22. Time-series of observed (black line) and forecasted hub height wind speed (ms⁻¹). Production configuration (P, red squares) and reduced vertical resolution (V, green circles) for 30-31 Dec 2011.

Figure 5-23. As in Figure 5-22, but for 07-08 May 2012

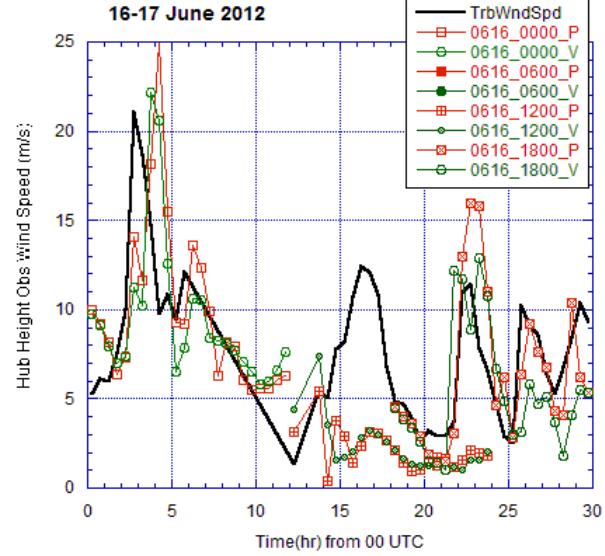


Figure 5-24. As in Figure 5-23, but for 07-08 May 2012

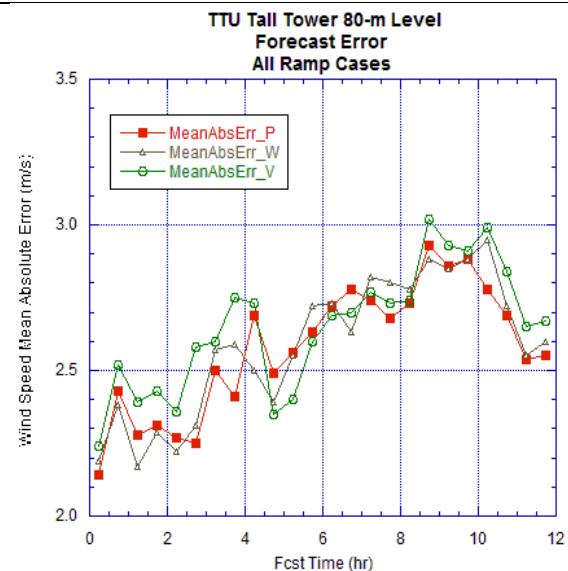


Figure 5-25. Mean absolute forecast error (ms⁻¹) for turbine height wind speed at the Texas Tech Tall Tower over all ramp cases in WFIP study period. CAPS ARPS production forecast (P), special data withheld (W), and reduced vertical resolution (V).

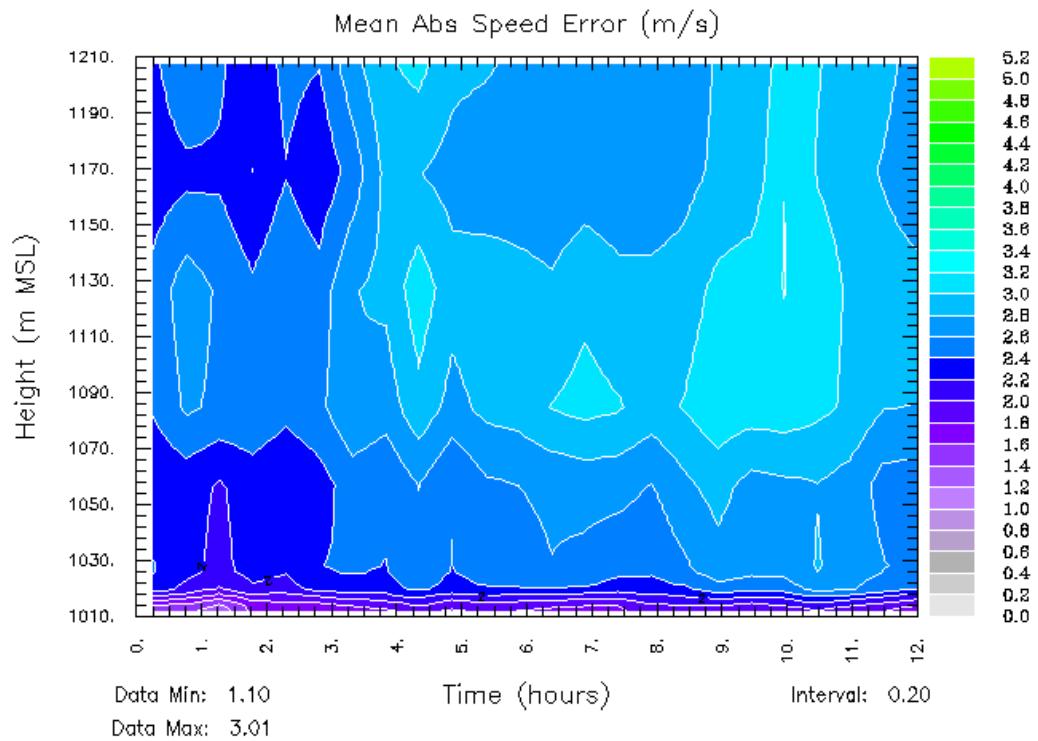


Figure 5-26. Mean Absolute Speed Error (ms^{-1}) as a function of height and forecast time for all ramp cases for the CAPS Production forecasts.

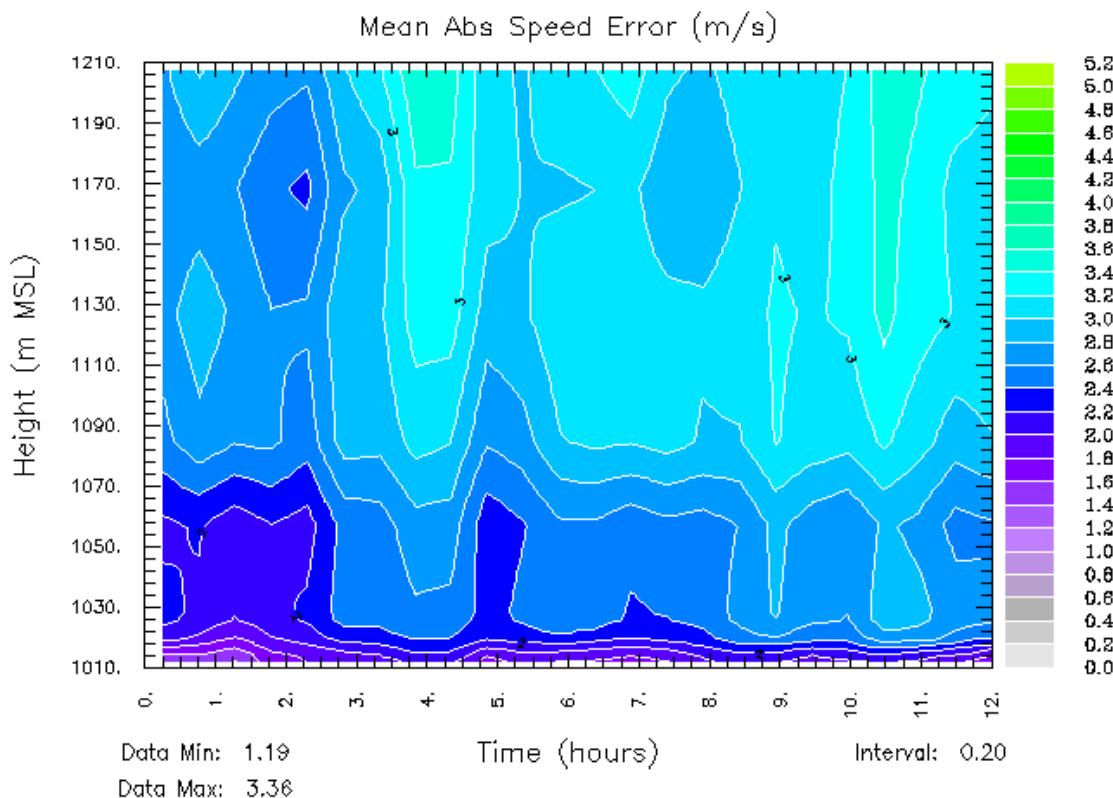


Figure 5-27. Mean Absolute Speed Error (ms^{-1}) as a function of height and forecast time for all ramp cases for the reduced vertical resolution forecasts.

5.3 TTU

5.3.1 Month-long period sensitivity study

The mean absolute 0-24hr wind speed errors measured against METAR surface observations for the EnKF and 3DVAR month-long experiments, averaged over 120 assimilation cycles, are shown in Figure 5-28. Analysis errors (0 hour) reflect the fit to assimilated observations prescribed by the data assimilation parameters of each system and should not be considered an appropriate measure of analysis quality. Throughout the remainder of the forecast window, however, neither system consistently performs better. In the early forecast hours prior to 12hr forecast time, the EnKF outperforms 3DVAR.

This difference is largest at forecast hour 7 when the EnKF produces a 3.3% improvement over 3DVAR (1.47 m/s compared with 1.52 m/s, statistically significant at the 85% confidence level using a one-sided Student T test). In the middle of the forecast window from roughly 12-18 hr, the two systems perform essentially the same, and by 24-hr forecast time, 3DVAR is slightly better (1.58 m/s vs 1.61 m/s). This improvement late in the forecast window is less prevalent than the differences found earlier, and is statistically significant at only the 70% confidence level. Nonetheless, these results suggest the EnKF performs better with regard to surface winds at 0-12 hr lead times, 3DVAR performs better at lead times greater than around hour 18, and a transition period exists from about 12-18 hr where neither system outperforms the other.

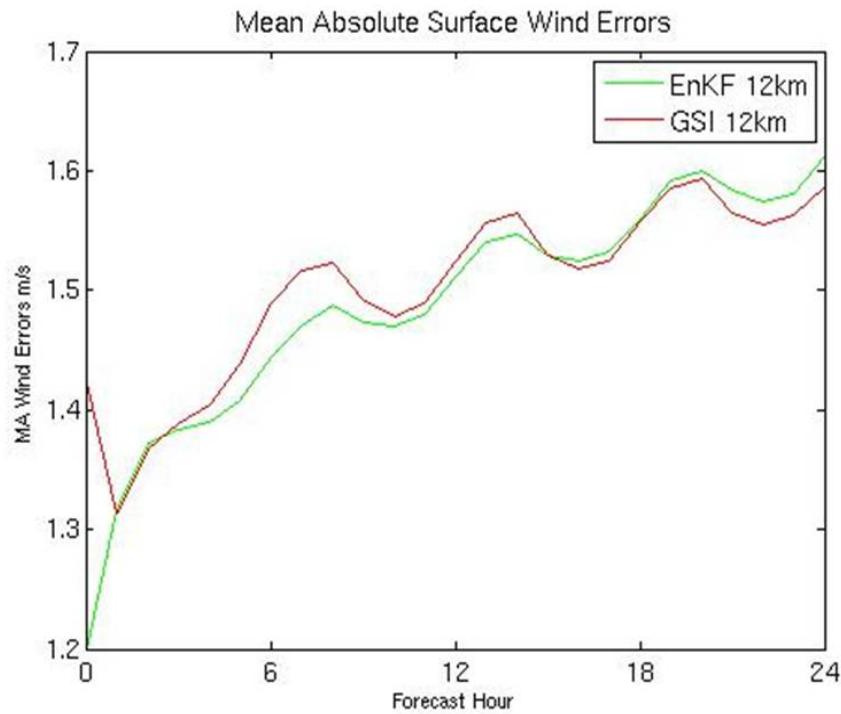


Figure 5-28. The mean absolute 0-24hr wind speed errors for the EnKF (green line) and 3DVAR (red line) month-long routine observation experiments averaged over 120 assimilation cycles in December 2011.

The improved performance at short lead times of the EnKF is likely caused by the flow dependent covariances it uses during assimilation. Figure 5-29 shows the background field and the analysis increments of both 500-hPa geopotential height (GPH) and 80-meter meridional (V) wind component for the 1200 UTC December 1, 2011 initialization for EnKF and 3DVAR. At that time an upper-level trough had just moved ashore over the southwest coast of the U.S., with broad cyclogenesis occurring in the lee of the Rocky mountains in the Texas panhandle. An arctic surface cold front can be seen surging southward along the east side of the Rocky Mountains (marked by the wind shift from northerly to southwesterly). This synoptic pattern is described in Colle and Mass (1995) as one that is commonly associated with strong southward moving cold fronts into the west Texas region. As expected, EnKF flow-dependent analysis increments exhibit features that are aligned with both the strong 500-hPa geopotential height gradient at the base of the trough, and along the ridge axis

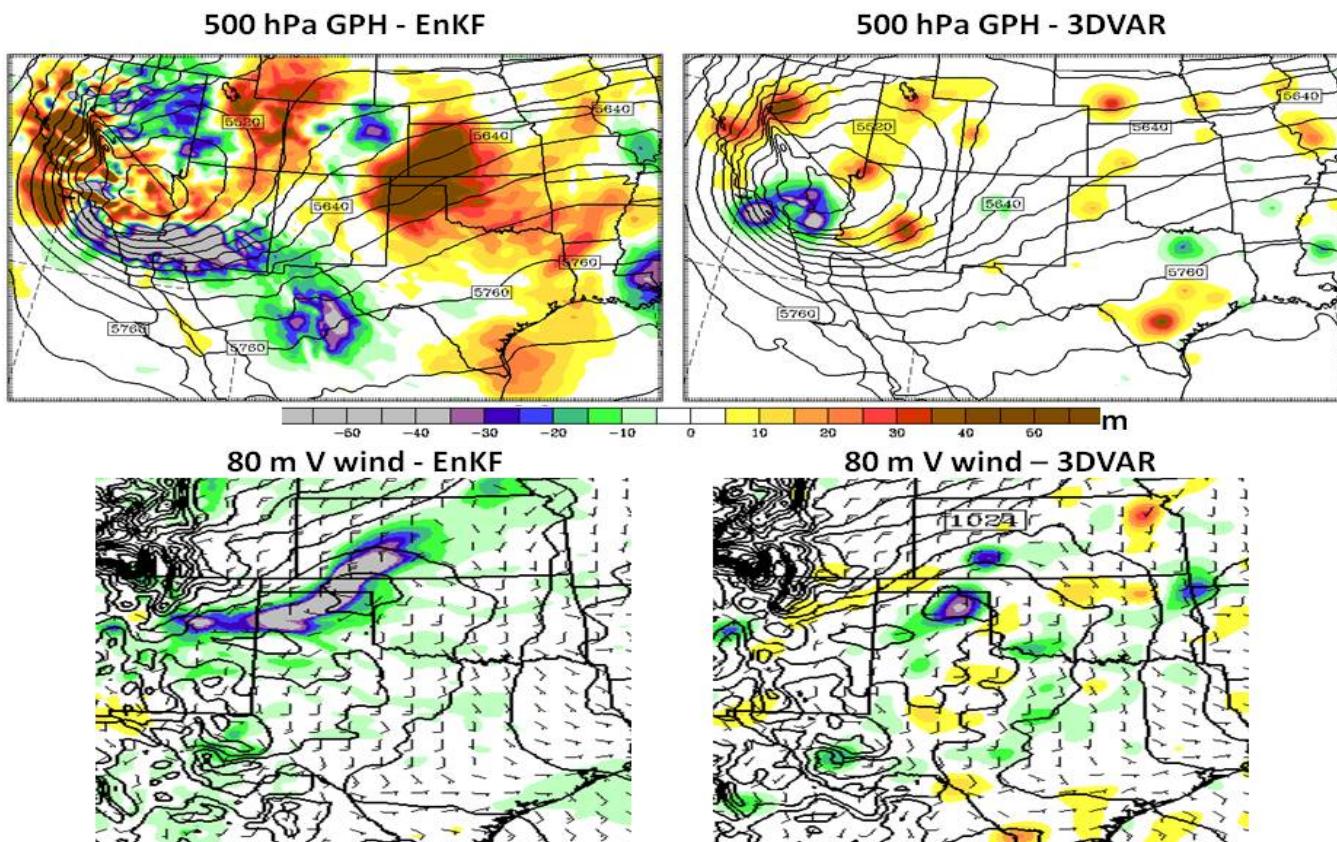


Figure 5-29. The 500-hPa background field (black contours, contour interval is 30 m) and analysis increments (shaded) in the top row, and the sea-level pressure background field (black contours, contour interval is 2 hPa), background 80-m wind barbs, and 80-m meridional (V wind) analysis increments (shaded) in the bottom row for the EnKF and 3DVAR analyses valid at 1200 UTC December 1, 2011.

extending from the Texas panhandle into Wyoming. EnKF analysis increments at the surface show a clear signal along the southward surging cold front with values of roughly -7 m s^{-1} . The negative values indicate that the assimilation procedure significantly strengthened the northerly flow behind the front. Also as expected, 3DVAR analysis increments at 500-hPa are in similar locations and possess similar maximum magnitudes, but are clearly more isotropic in nature and less aligned with flow features at analysis time. This behavior extends to the surface where 3DVAR is unable to achieve the along-front increment of the EnKF from the assimilated observation in the north panhandle of Texas. It follows that the EnKF analysis increments that are clearly more realistic evolve in time to improve early forecast performance over that of 3DVAR.

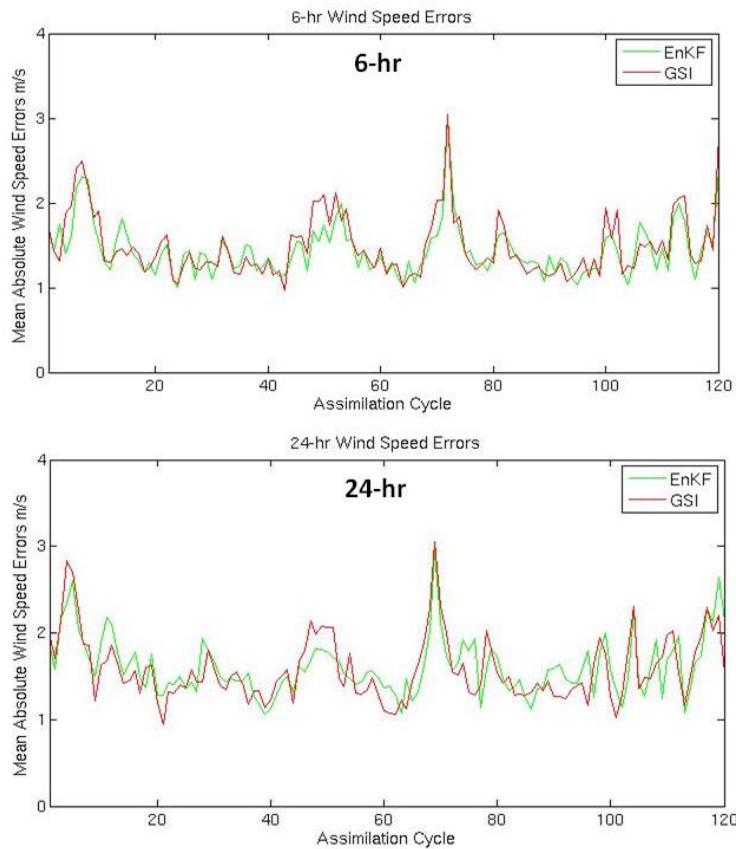


Figure 5-30. The mean absolute EnKF (green line) and 3DVAR (red line) wind speed errors for all 120 assimilation cycles in December 2011 at both 6hr and 24hr forecast time for the month-long routine observation experiments

The improved performance of 3DVAR by the end of the forecast window, however, suggests it is more efficient than the EnKF in improving the larger scale features in the analysis. Ancell (2012) also showed roughly a 12-hr window of improvement from forecasts integrated from analyses produced with a high-resolution EnKF over forecasts down-scaled from coarser EnKF analyses. That study speculated that the improvements in the higher-resolution assimilation run were diminishing near the end of the 12 hour timeframe because down-scaled forecasts begun from improved synoptic-scale analyses aloft from coarser assimilation were beginning to positively force the near-surface fields. We speculate that the same behavior is occurring here, and that around 12-18 hr forecast time the improved surface forecasts of the 3DVAR system must be emerging as the improved governing synoptic-scale flow begins driving the surface fields, dominating any positive forecast effects of flow dependence achieved by the EnKF. This idea is supported by the inspection of mean absolute 500-hPa geopotential height errors measured against rawinsonde data in the month-long verification area (not shown). The 3DVAR run possesses smaller 500-hPa errors at all forecast times, and whereas errors are slightly smaller from 6 - 12hr (a few meters), they grow throughout the remainder of the forecast window.

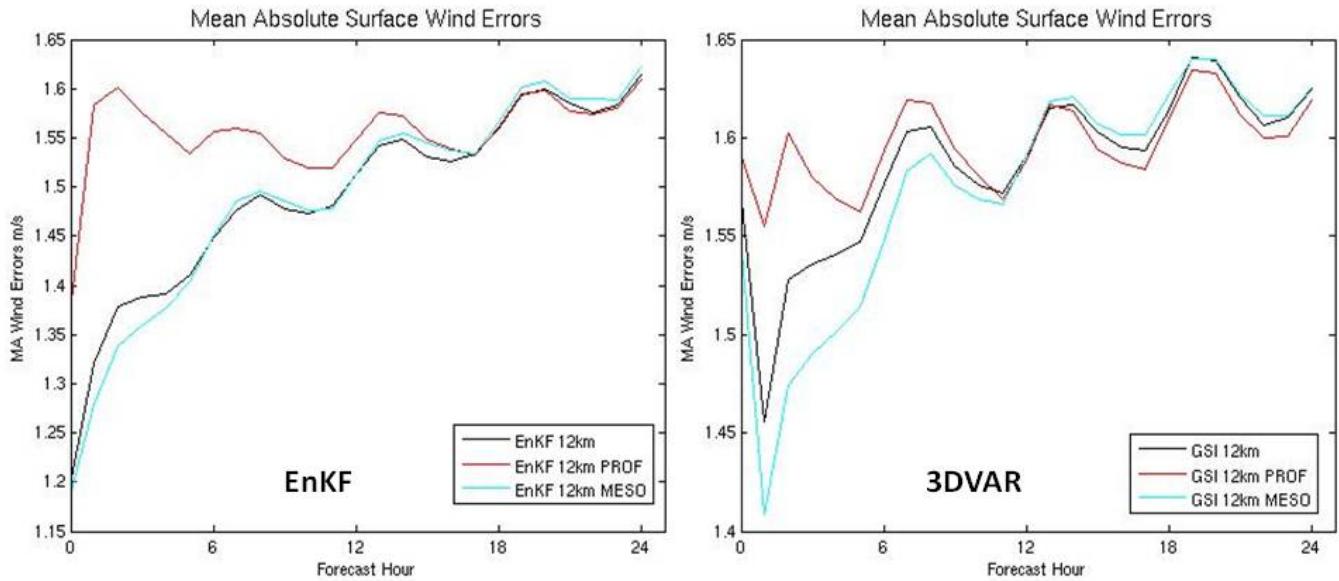


Figure 5-31. The mean absolute 0-24hr wind speed errors for the month-long control run (black line, routine observations only), the run that assimilates mesonet observations in addition to routine observations (cyan line), and the run that assimilates profiler/sodar observations in addition to the routine observations (red line) for both EnKF and 3DVAR.

Although the EnKF exhibits better performance on average at 6-12hr forecast time, and the 3DVAR system shows improved forecasts beyond 18hr, neither system consistently performs better. Figure 5-30 shows the mean absolute errors averaged over the verification area for all 120 assimilation cycles at both 6hr and 24hr forecast time. Note that both forecast times show cycles for which the EnKF was better than 3DVAR, and vice-versa. This provides a good next step to further research into the benefits of each system for surface wind forecasts. Since the different systems perform better at different analysis times, it is possible the success of each system depends on specific flow situations, and this will be further evaluated in the future.

Further motivation for this evaluation is warranted since synoptic-scale flow-dependent predictability was clearly shown in Ancell and McMurdie (2013), and may be manifesting itself here within the different data assimilation systems.

Figure 5-31 depicts the mean absolute errors for both the EnKF and 3DVAR for the month-long control run (no mesonet or profiler/SoDAR observations, only routine observations) as well as the runs that independently assimilate mesonet observations and profiler/SoDAR observations in addition to the routine observations. Both systems show the same result - mesonet data improves early forecasts, and SoDAR/profiler observations degrade them. There are essentially no statistically significant impacts from either observation type beyond roughly 12-hr forecast time in both systems. This lack of impacts beyond 12 hr for any observation type in either system, either positive or negative, supports the idea that the lower

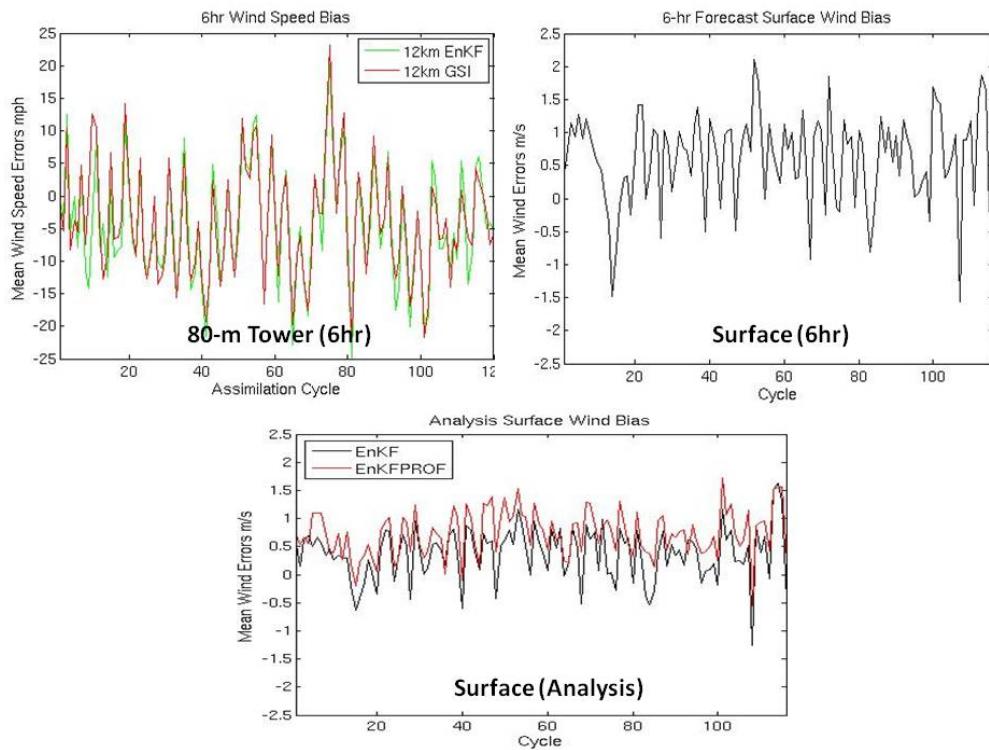


Figure 5-32. The background (6-hr forecast) mean wind speed errors for all assimilation cycles measured against 80-m tower data for EnKF (green line) and 3DVAR (red line), the mean EnKF surface wind speed errors for the same cycles measured against surface METAR data, and the surface analysis bias for the EnKF control run (black line) as well as the run assimilating profiler/sodar observations (red line).

atmospheric observation types tested here provide temporary benefits until the influence of driving synoptic flow becomes dominant.

For mesonet observations, positive impacts were realized in both systems up to about 6 hr forecast time. The largest improvement from mesonet observations within both systems is at 3hr forecast time, and for 3DVAR is 3.9% (1.47 m/s vs. 1.53 m/s), while improvements are more subtle in the EnKF (2.9%, 1.34 m/s vs. 1.38 m/s). These improvements are statistically significant at the 85% confidence level.

Since more assimilated observations generally improve forecasts, it is unexpected that the WFIP observations so clearly degrade early forecasts. However, this issue is likely due to a lower atmospheric slow wind bias. Figure 5-32 depicts the background (6-hr forecast) mean wind speed errors for all assimilation cycles measured against the 80-m tower data. For both assimilation systems, the average mean error, or bias, is at least -4 mph (-1.8 m/s). Figure 5-32 also shows the mean surface wind speed errors for the same cycles measured against surface METAR data, which shows a significantly smaller but fast wind speed bias of about 0.6 m/s. Lastly Figure 5-32 depicts the surface analysis bias for the EnKF control run as well as the run assimilating profiler/SoDAR observations. The EnKF experiment that assimilates profiler/SoDAR observations has a fast wind speed bias (0.73 m/s) that is about twice as large as that of the EnKF run without profiler/SoDAR observations. This strongly suggests that the removal of the lower atmospheric slow wind speed bias during assimilation of profiler/SoDAR winds is being spread to the surface through covariance relationships, producing surface wind speeds that are too large. This results in the degradation of early forecasts from both systems when profiler/SoDAR observations are assimilated. This underscores a critical

need to address lower atmospheric model wind bias during assimilation if profiler/SoDAR observations are to provide value to wind forecasts. Since Ancell (2012) showed simple wind bias removal schemes can improve analyses but degrade forecasts upon the return of bias during model integration, the development of good bias removal techniques should be a key focus for wind power forecasting in the coming years. It is also possible that other physics schemes, such as boundary layer parameterizations, may produce smaller near-surface wind bias, and this potential solution should also be explored.

5.3.2 Wind ramp forecasts

Forecasts from the EnKF and 3DVAR were compared for 10 wind ramp cases, and surface network and profiler/SoDAR observation impacts were evaluated in each system for the same ramp cases. Since this evaluation focuses on specific ramp forecasts, differences among the experimental runs can be quite large. Figure 5-33 shows an example of the differences between the two systems for one such wind ramp forecast initialized at 1200 UTC December 2, 2011 valid at a single 80-m meteorological tower. A very large upramp began around 4 hr into the forecast period, an onset that was captured very well by the EnKF. However, the EnKF achieved only about half the magnitude of the observed ramp, whereas the maximum winds during the ramp were well forecast by the 3DVAR run (but were forecast too early). Ultimately, these differences were a result of each system's evolution of the synoptic features creating the wind ramp. A midlatitude cyclone with a trailing Pacific cold front tracked over the location of the meteorological tower in the EnKF run, producing a smooth veering of winds with time (shown in Figure 5-33). The 3DVAR run, however, was associated with a cyclone track further north, allowing the cold front to track across the observation location and ramp up the winds more strongly. The 3DVAR wind direction forecast shown in Figure 5-33 reveals the frontal passage by the wind shift from southerly to westerly, a feature also captured by the observations. Thus, whereas the 3DVAR system seemed to locate the synoptic ramp-producing features better, the EnKF seemed to better capture their timing, at least with respect to the enhanced pressure gradients involved with higher wind

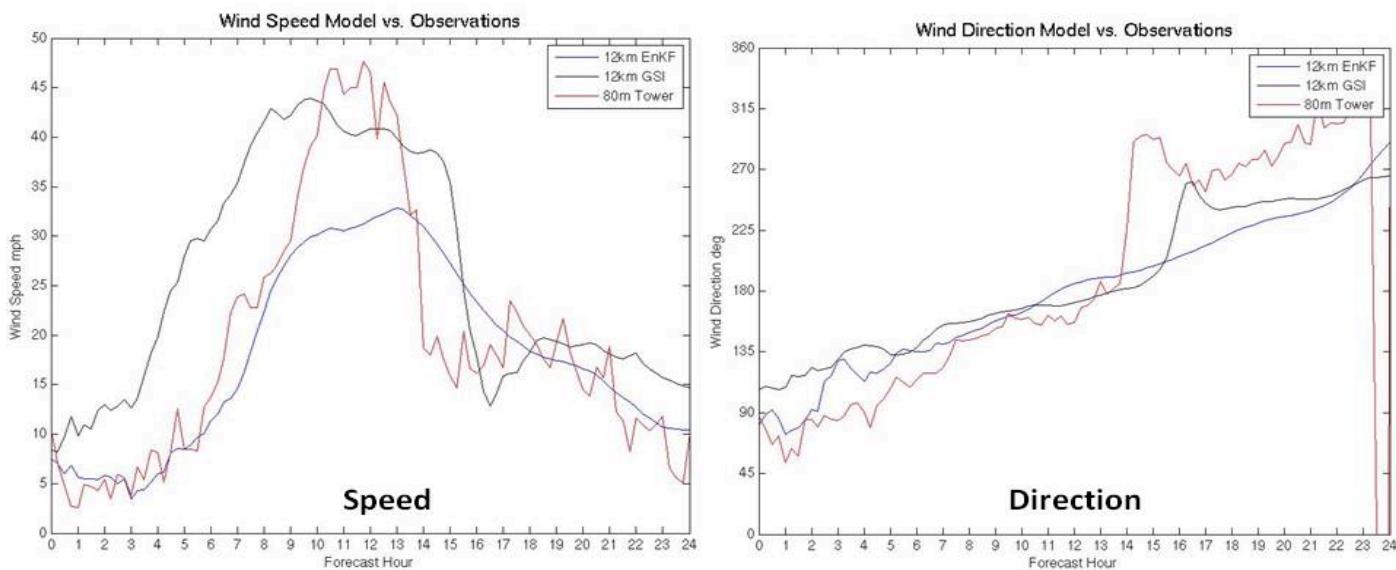


Figure 5-33. Example wind ramp speed and direction forecasts from the EnKF (blue line) and 3DVAR (black line) systems initialized at 1200 UTC December 2, 2011 valid at the location of a single 80 m meteorological tower. The observed wind speed and direction from the tower are also shown.

speeds. Similar differences were observed (not shown) between the runs from each system that assimilated the different sets of observations. In order to more generally understand the wind ramp forecast performance of each system, as well as that when the different observational assets are included, we now present composites of the errors for different ramp characteristics.

In terms of ramp magnitude and maximum winds, a consistent picture emerges with regard to the performance of the two assimilation systems. For all lead times, Figure 5-34 shows the performance of the EnKF versus 3DVAR, as well as the different observation impacts in each system, for absolute ramp onset (hr) and duration (hr) errors. Figure 5-35 shows the same comparisons for absolute ramp magnitude (m/s) and maximum wind speed (m/s) errors. The results are composited by three categories of forecast lead time (early: 0-9hr, middle: 9-15hr, and late: 15-24hr), and data was placed into each of these forecast lead time bins based on the forecast hours over which the majority of the observed ramp occurred. Results averaged over all lead times are also presented in the same figures. Data from all twenty 80-m towers was used for this composite verification, although there were some cases for which some of the towers had missing data. For this forecast wind ramp analysis, since statistical significance is very challenging to appropriately apply to a 10-member sample, we have chosen a rough set of guidelines to reveal the relative performance of the various experimental runs from the composite of cases used here. Onset and duration errors are viewed as

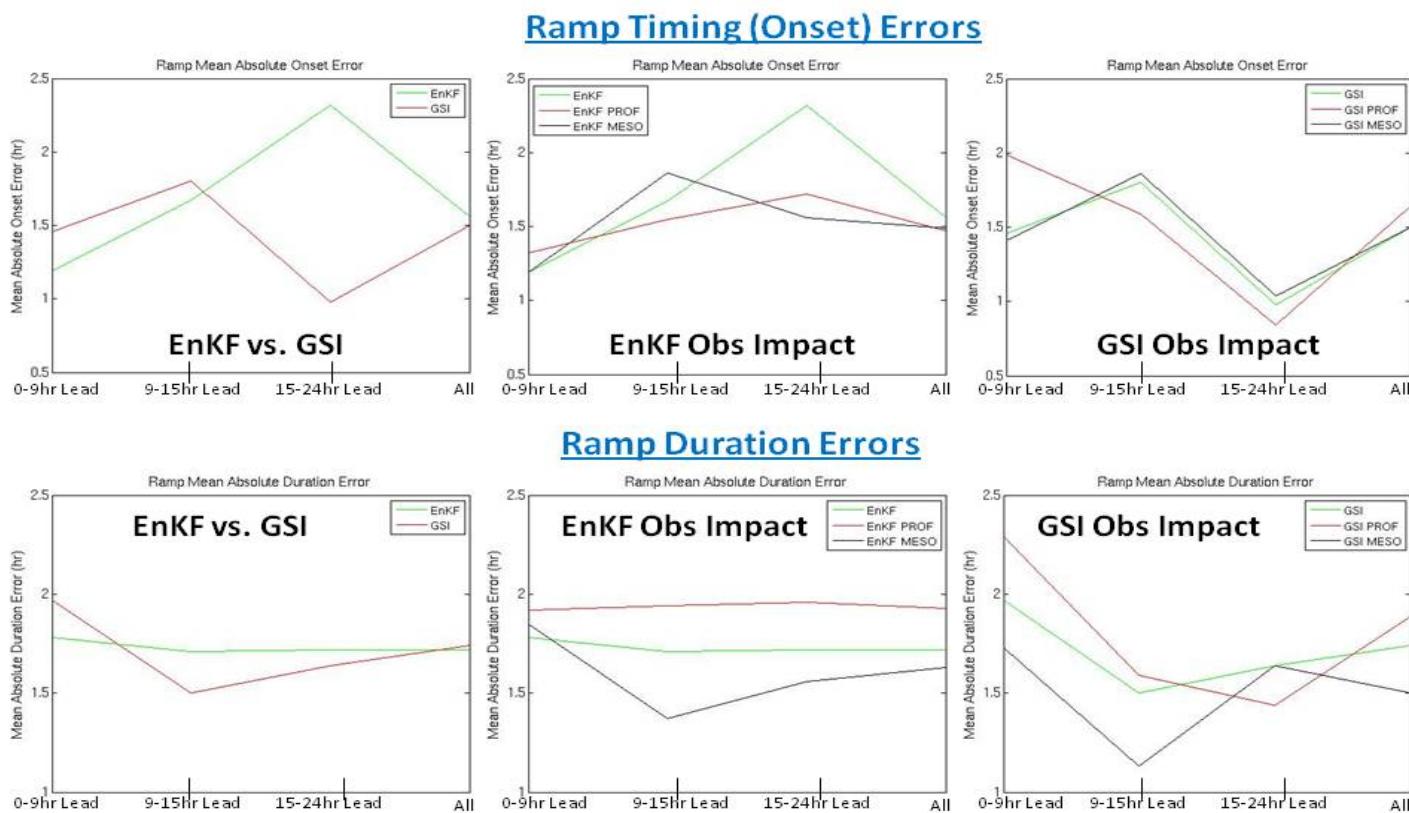


Figure 5-34. Example wind ramp speed and direction forecasts from the EnKF (blue line) and 3DVAR (black line) systems initialized at 1200 UTC December 2, 2011 valid at the location of a single 80-m meteorological tower. The observed wind speed and direction from the tower is also shown.

"significant" if they exceed 15 min, whereas ramp magnitude and maximum wind speed errors are viewed as significant if they exceed 0.5 m/s. These guidelines are also influenced by what we perceive would be important with regard to error to wind farm operators. It should be noted that without formal statistical testing, however, these results are only suggestive of the forecast performance of the various runs more generally.

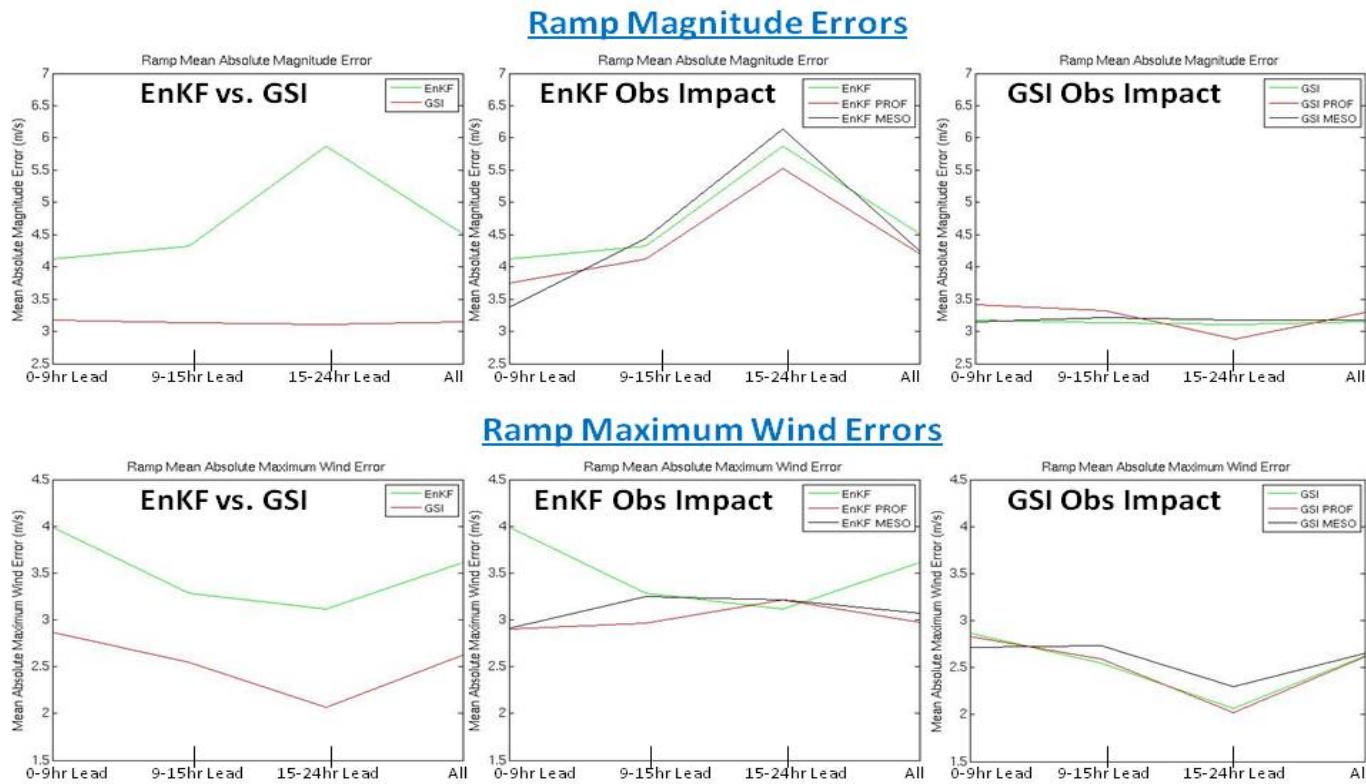


Figure 5-35. Ramp timing and duration errors averaged over all wind ramp cases for the EnKF and GSI routine and observation impact experiments.

Much like the month-long comparison, EnKF and 3DVAR ramp onset and duration errors averaged over all lead times were similar, but the EnKF showed roughly a 15-min advantage for early lead times, and 3DVAR showed an advantage at later forecast hours. This advantage for ramp duration was most prominent for middle forecast lead times at which point it was about 15 minutes, and didn't occur until late lead times for ramp onset at which point the improvement was over an hour. The ramp characteristics of onset and duration identify the timing and structure of the ramp events and the features controlling them (i.e. the front). Thus, it stands to reason that the same factors influencing the month-long results apply to the ramp cases as well: the EnKF provides early forecast improvements through flow-dependent adjustments that are outweighed by downscaling of improved larger-scale structure in 3DVAR at later forecast hours.

In terms of ramp magnitude and maximum winds, a consistent picture emerges with regard to the performance of the two assimilation systems. For all lead times, and particularly at late lead times, the 3DVAR system outperforms the EnKF. This likely is a result of the tendency of the mean EnKF forecast to underforecast wind speeds associated with significant events. This was examined in Ancell (2013), who

showed that wind speeds involved with synoptic-scale cyclones were underforecast by as much as 10 m/s over a 48-hr forecast window, and that this issue became worse as lead time increased. The reason for this issue is that the ensemble mean is not constrained to be on the model attractor, resulting in unrealistic behavior that favors reduced magnitudes. This suggests that best member techniques that select ensemble members in some way closest to the ensemble mean might improve EnKF wind ramp forecasts (a technique also shown to be successful in Ancell 2013). This will be explored in future work. In any case, results here suggest the deterministic 3DVAR forecasts will most likely be better than an ensemble mean for metrics that are based on ramp magnitudes at all forecast times.

The data denial experiments produced mixed results. For ramp onset and duration, observation impacts were negative at early forecast hours within the 3DVAR system for profiler/SoDAR observations, which perhaps is caused by the bias issue identified with near surface winds in the current model configuration. Mesonet network observations were beneficial for early and middle lead time 3DVAR forecasts, but otherwise impacts were negligible. Within the EnKF, there were improvements in ramp onset errors of about 45 min at late lead times, and roughly a 15-30 min improvement at middle and late lead times from mesonet network observations. Profiler/SoDAR observations also improved EnKF onset errors in late forecast hours by about 45 min but degraded forecasts of ramp duration by about 15 min throughout the forecast window. It is somewhat puzzling how assimilated observations can improve one feature of the ramp (timing) but degrade another (duration) when both aspects might be expected to be strongly correlated for a specific ramp-causing event. However, the results here suggest this is not the case on average, and that it is possible for an assimilation system to beneficially adjust the location of a front, for example, but not correct for the magnitude and intensity of the front in terms of the wind field surrounding it. One method to examine this behavior is through ensemble sensitivity analysis of different ramp aspects, and this is a planned future extension of this work. Since profiler/SoDAR observations are unexpectedly degrading duration forecasts in the EnKF, we also speculate that the lower atmospheric bias issue may play a role.

Observation sensitivity was subtler for ramp magnitude and maximum wind forecasts in both the EnKF and 3DVAR. In fact, 3DVAR essentially showed no significant effect from either observation type at any forecast time. In the EnKF, observation impacts were positive for both types of observations for short forecast lead times. This likely is a result of adjustments made during assimilation to account for the unrealistically slow wind speeds produced by the ensemble mean. No significant impacts were seen at middle or late forecast lead times within the EnKF. Interestingly, this suggests surface and near surface observations are more efficient at making adjustments to the timing and structure of ramp-inducing events (onset, timing) than to the magnitudes of the associated wind field (magnitude, maximum winds).

6 Accomplishments

WFIP represents the first large-scale comprehensive public-private collaborative endeavor to investigate how to improve short-term wind power production forecasting. For the SSA, the coordination between all project partners—DOE, the NOAA labs (ESRL and ARL), PNNL, NREL, and the SSA team—resulted in the project meeting all major objectives set forth in the original proposal. This includes

- 1) An NWP-based analysis to provide objective guidance for the selection of locations for sensor deployment for the field campaign;
- 2) The setup, testing, establishment, and operation of the full set of the 13 member AWST-based forecast model run cycles, the OU CAPS ARPS runs and sensitivity analysis, and the TTU EnKF 3DVAR data assimilation tests;
- 3) Deployment and operation of all surface (meteorological and flux measurements) and remote sensing (SoDAR, wind profilers) stations;
- 4) Completion of the initial¹⁹ economic analysis of cost savings attributable to improved wind power forecasts.

Final results include

- 1) Significant improvement in forecast power production performance for the AWST modeling ensemble, including
 - a. a decrease in mean absolute error (MAE) of aggregate power from the WFIP ensemble versus baseline forecasts with the largest (60%) at hour 1 and smallest (< 20%) after hour 3; and
 - b. enhanced project observations resulting in up to a 200% improvement in the critical success index (CSI) for deterministic ramp;
- 2) Identification and analysis of the principal phenomena responsible for significant ramp events affecting the ERCOT domain;
- 3) Insight into of data assimilation scheme performance (TTU) in the context of ramp predictions and data withholding (TTU);
- 4) The value of greater vertical resolution for wind speed and ramp forecasts (OU CAPS);
- 5) Successful development of gap-filling techniques for remote sensing instrument platforms; and
- 6) Preliminary findings showing significant annualized production cost savings and load payment savings resulting from the improved short-term forecasts.

Technology transfer: Two project User Group (composed of utilities, ISOs, and other interested stakeholders) meetings were held in Tucson AZ on 7 February 2012 and Salt Lake City on 27 February 2013. Presentations were given on project progress and preliminary results. Preliminary and final results and findings were also presented at other conferences and workshops (see Appendix B).

¹⁹ The Department of Energy (DOE) has decided to undertake additional studies to explore the complex interactions between wind forecasting and power system operations prior to publication of results.

7 Summary and Conclusions

7.1 AWST WFIP Forecast System

An AWST ensemble of nine high-resolution models and the NOAA ESRL HRRR model were used to produce 15-minute forecasts for the WFIP southern region. The AWST NWP ensemble featured multiple models (ARPS, MASS, WRF), different data assimilation methods, and physics packages. A power and wind forecast was then generated for each ensemble member and wind plant location by applying a statistical bias correction method. Finally, a single deterministic forecast was generated using an optimal weighting scheme based on model performance. The ensemble WFIP forecast was then compared to a baseline forecasting system to examine forecast performance. The baseline power forecast did not use any WFIP ensemble members or project observations as input, but did include the statistical bias correction and ensemble weighting scheme. The attributes of the NWP baseline and WFIP ensemble forecast are summarized in Table 7-1. The results showed a decrease in MAE of aggregate power from the WFIP ensemble versus baseline forecasts with the largest (60%) at hour 1 and smallest (< 20%) after hour 3 (Section 5.1.1).

The more accurate WFIP forecasts resulted from a combination of factors summarized in Table 7-1. To isolate the impact of project observations on forecast improvement, several data denial experiments were conducted (Table 7-2). The experiment with the MASS ensemble member ran for one year as part of the real-time AWST nine-member ensemble whereas the ARPS historical runs were of much shorter duration (Table 7-2) and examined several active weather periods.

Results from the real-time data denial experiments showed that assimilating project observations had minimal impact on *overall* forecast performance, reducing the aggregate forecasted power RMSE by 0.3 % of capacity. For historical forecasts, the assimilation of project observations showed little to no improvement in forecast performance, even when they were withheld from the RR model used for IC and BC. However, project observations did improve deterministic *ramp event forecasts* for specific cases with up to a 200% improvement in the CSI score during certain forecast hours (section 5.1.6). This result suggests that project observations have value in certain regimes that cannot be readily quantified using metrics such as MAE and RMSE over large samples.

To improve aggregate statistics of MAE or RMSE over many cases would require sampling the complete suite of relevant meteorological variables (i.e. temperature, moisture, and wind) over a larger volume of the atmosphere. A recent study summarizing results from NOAA's Winter Storm Program suggests that under sampling is one of several factors that can lead to minimal impact from targeted observing campaigns (Hamill et al. 2013). In fact, NOAA found that special WFIP observations in the NSA produced a greater impact than in the SSA. This result was attributed to the fact that more instrumentation was deployed over a wider geographic area in the northern versus SSA.

Ramp events can adversely impact system reliability. Therefore, accurate forecasts of ramp events can help system operators anticipate disruptive impacts from large changes in wind generation over short periods of time. Probabilistic ramp forecasts were produced using each WFIP ensemble member as input and compared to a baseline ELRAS system. The ELRAS system was composed of a single ARPS member that did not include IAU after assimilation, but included project observations (Table 7-3). Ramp forecasts using several NWP

models from the WFIP ensemble more accurately predicted the timing and magnitude of ramp events than those produced from a single NWP member in ELRAS (Section 5.1.3). The best performing single member ramp forecast was generated with the HRRR model as input and benchmarked using RPSS as the performance

Table 7-1: The NWP baseline and WFIP ensemble forecast attributes.

Attribute	Baseline Forecast	WFIP Ensemble Forecast	Proposed Sensitivity Experiment
Observations	Standard observations ¹ Satellite-derived sea surface temperatures	Standard observations, Non-project observations ² , Project observations ³	Baseline with project observations
Data assimilation method	Optimal interpolation	3D variational, nudging (warm start)	Change baseline assimilation scheme to nudging
Forecast models	MASS	ARPS, WRF, MASS	Baseline with model members
Model grid spacing	10 km	5 km	Decrease baseline grid spacing
Ensemble size	2 members	9 members	Increase baseline ensemble size
Initial and boundary conditions	GFS/NAM	Real-time RR	Use real-time RR boundary conditions with baseline
Initialization frequency	6 hours	2 hours	Increase baseline frequency

1 Rawinsondoss, automated surface observing system stations, buoys
 2 Mesonet surface stations, WSR-88D (radar reflectivity), satellite winds
 3 Sodars, profilers, Texas Tech tall tower, wind plant towers

Table 7-2: Comparison the of real-time and historical observation denial experiments.

Model	Period	Initial and Boundary Conditions	Forecast Product Validated	Metric ¹
MASS real-time	1 Year (Nov 2011-Sep 2012)	Real-time Rapid Refresh (RR) (warm start)	Aggregate power, wind speed Probabilistic ramp forecast	MAE/RMSE/ RPSS
ARPS historical	2 Months (Dec 2011, Apr 2012)	Real-time RR (cold start)	Aggregate power, wind speed	MAE/RMSE
ARPS historical	1 Week (1-7 Dec 2011)	Data denial RR (cold start)	Aggregate power, wind speed	MAE/RMSE
ARPS historical	Ramp Cases (Table 5-2)	Real-time RR (cold start)	Deterministic ramp event	CSI

1 Detailed metric descriptions in Section 5.1

Table 7-3: Baseline ELRAS and WFIP ensemble ramp forecast attributes.

Attribute	Baseline ELRAS Forecast	WFIP Ensemble Forecast
Data assimilation method	3D variational (IAU)	3D variational (no IAU), nudging (warm start)
Forecast models	ARPS	ARPS, WRF, MASS
Ensemble size	1 member	9 members
Initial and boundary conditions	NCEP RUC/RR	ESRL Real-time RR

metric. The improved ramp forecasts were likely due to the larger WFIP ensemble versus the single model used in ELRAS.

The observation targeting technique used in this study was based on a simplified correlation analysis of time-lagged forecasts from a single NWP model rather than a more advanced method using the EnKF. The EnKF uses regime-dependent instead of isotropic structure functions in the data assimilation systems. These weighting functions vary for specific weather events and more accurately represent how observations influence the analysis of model parameters at nearby grid points. However, even ensemble based methods can be limited by small sample (i.e. cases) and ensemble size.

Although the EnKF assimilation system was tested in this study, the limited computational resources precluded running many ensemble members at high resolution. In addition, the EnKF system as configured for the project could only assimilate a few types of available observational data at less frequent update cycles. As a result, the AWST models initialized using the EnKF hybrid method for WFIP produced no significant improvement in real-time power forecasts. With capability to assimilate all available observations at higher resolution over much larger numbers of ensemble members, such techniques have potential to improve model accuracy which would translate to improved wind power forecasts (Schwartz and Zhiqian 2014).

In addition to improved data assimilation and observation targeting techniques, several sensitivity experiments should be performed as part of future efforts to determine which component of the WFIP system contributed most to more accurate forecasts. A number of different parameters including the model used, grid spacing, and other factors (Table 1) should be varied to determine which attributes are most important for improving forecast accuracy. The current ERCOT operational forecast system is already running a larger ensemble of NWP models at higher resolution to leverage forecast improvements compared with the original STWFP.

7.2 OU CAPS sensitivity studies

For the duration of the WFIP field deployment and modeling campaign, OU CAPS contributed unique 12-hour real-time numerical weather forecasts that included high resolution vertical boundary layer structure, and assimilated all NEXRAD Doppler radar data in the region and the standard surface ASOS, mesonet, and upper air data. The CAPS forecasts also included the special observations from the project wind profilers and SoDARs. The number of vertical model levels in the lowest 200 m was increased to 5 levels from the more DE- EE0004420

typical 3 levels. A total of 58 vertical levels were used in the CAPS forecasts. The horizontal resolution of the CAPS forecasts was 3 km, which is considered the minimum resolution needed to support the explicit forecasting of thunderstorms.

In addition to providing the high-resolution forecasts to the WFIP real-time ensembles that were evaluated by AWST, CAPS ran some additional experiments on major ramp event days with two principal goals, 1) to determine the impact of the special project wind profiler and SoDAR data, and 2) to gauge the effect of the additional vertical resolution in the boundary layer. Our evaluation of these tests focused primarily on the Texas Tech University Tall Tower (TTU) in Lubbock, Texas.

At the TTU site we found that the best wind speed forecasts at the wind turbine height (80 m AGL) came from the CAPS operational forecasts, which included the special WFIP observation data (including the TTU tower data) and had the high resolution in the boundary layer. The main benefits among experiments were found in the first 5 hours of the forecasts. The CAPS operational 80-m wind speed forecasts during the ramp days had mean absolute errors of about 2.25 ms^{-1} in the first 2 hours, with MAE increasing to near 3.0 ms^{-1} at 10 hours. Removing the special observation data had a small negative effect on the forecasts at this site with approximately a 0.05 ms^{-1} increase in mean absolute wind speed error. Reducing the vertical resolution in the boundary layer had a greater impact, with a $0.15\text{--}0.20 \text{ ms}^{-1}$ increase in error in the first 5 hours. We also verified the forecasts using the WFIP SoDAR data at other locations. The results from those sites were similar although for some cases the reduced resolution forecasts had slightly better results than the higher resolution forecasts beyond 8 hours. This may be explained by the increased resolution in the boundary layer producing somewhat reduced resolution of features in the upper layers.

7.3 TTU data assimilation sensitivity studies

The first objective of this study was to compare lower atmospheric wind speed forecasts from a fully flow-dependent ensemble data assimilation/forecasting system (EnKF), to that of a deterministic 3DVAR system that uses static covariances. Verification of 24-hr forecasts was performed against surface observations over a large area including Texas and Oklahoma over a month-long period. Forecasts from each system for 10 wind ramp cases were also evaluated against 80-m tower data within a relatively small area of existing wind farms in central Texas. For the month-long period, it was found that 0-12 hr forecasts were better within the EnKF, likely a result of a more realistic analysis achieved through flow-dependent covariance relationships. Forecasts from 18-24 hr, however, were better in the 3DVAR system, suggesting that they benefit later in the forecast from their dynamical links to an improved synoptic-scale flow aloft. Neither system was better than the other consistently at any forecast hour, however, and understanding whether certain flow regimes are associated with advantages in the different systems is a logical extension of this work.

Both profiler/SoDAR and mesonetwork observations had little impact on later forecast hours. Mesonetwork observations tended to improve early forecasts in both the EnKF and 3DVAR and should be considered beneficial to lower-atmospheric wind prediction, but profiler/SoDAR observations degraded early forecasts in both systems, a likely result of assimilating lower-atmospheric wind observations in the presence of a lower-atmospheric slow wind bias. Mitigating this issue can be done by 1) testing different physics schemes, such as the PBL parameterization, in the hopes of discovering a physics configuration that produces sufficiently small biases, or 2) developing a bias removal technique that improves the analysis, yet doesn't degrade forecasts as it has been shown to do in previous work. Until these recommendations are tested, assimilating

profiler/SoDAR observations will likely continue to degrade near-surface wind forecasts within the current modeling configuration, and shouldn't be used for 0-12hr wind power prediction.

Wind ramp forecast performance showed some similar behavior to that of the month-long experiment. Ramp onset and duration, which reflect mostly the timing, location, and structure of the synoptic and mesoscale structures creating the ramps, was improved at early forecast hours with the EnKF, but was improved at late forecast hours with 3DVAR. In tandem with the month-long experimental results, this implies different data assimilation systems should be used as the best guidance at different forecast lead times for day-to-day wind prediction and ramp timing. The effects of nonlinearity, which can cause the ensemble mean to possess wind speeds that are unrealistically too slow, result in 3DVAR outperforming the EnKF with regard to forecasts of ramp aspects at all lead times that are associated with wind speed magnitudes (ramp magnitude and maximum winds during the ramp). Thus, the ensemble mean should not be used for wind prediction for these ramp characteristics. This will always be an inherent property of the ensemble mean, and "best member" techniques that select specific ensemble members (typically close to the mean in some way) should be employed to improve the value of the EnKF forecasts related to wind speed maximum magnitudes. Whether sufficient improvement in this way can be obtained to become competitive with the better 3DVAR forecasts shown here is a key question, and will be examined in future work.

Observation impacts on wind ramp forecasts varied. Profiler/SoDAR and mesonet network observations generally had very little impact on 3DVAR forecasts of any ramp aspect. The exception was for early and middle lead time forecasts of ramp duration, which were improved by mesonet network observations. Aside from early forecast improvements in the EnKF likely associated with the unrealistic behavior of the ensemble mean, no significant impacts were seen with regard to ramp magnitude and maximum winds. Both observation types improved ramp onset forecasts later in the forecast window. Mesonet network observations also improved longer lead time forecasts of ramp duration, but profiler/SoDAR observations degraded EnKF forecasts over the entire forecast window. A planned next step toward understanding the mixed results found here among different ramp characteristics, assimilation systems, and lead times, is to perform ensemble sensitivity analysis of the different ramp characteristics used here. It is hoped that such an examination reveals a more detailed picture of the mixed impacts exposed by the observations used in this study. Observation types and locations could then potentially be chosen to minimize forecast errors associated with wind ramp aspects deemed most important to wind power operators. Finally, it should be noted that the results here should be viewed as a benchmark for further investigation into more advanced assimilation systems, such as 4DVAR or hybrid techniques, which have the potential to improve forecast over that of both the EnKF and/or 3DVAR.

8 Recommendations

WFIP, a project with substantial breadth and depth in the context of field observations, modeling efforts, and economic analyses, generated a huge amount of observational data and model output. The analysis presented here (and in the companion reports presented by NOAA and WindLogics), although comprehensive, is not complete in the sense of investigating the science, economics, and policy incentives to further the application and facilitate the use of wind power forecasts in the stakeholder community. The results highlighted in the preceding sections demonstrate that improved model systems, data assimilation methods, statistical post-processing, and an enhanced (remote sensing) observation network can produce more accurate short-term

wind power forecasts. However there is still opportunity for significant improvement, and better understanding (and therefore better parameterizations or better model physics) of the phenomena that drive the wind and large changes (i.e. ramp events) in the wind field. In particular, **additional post-WFIP work** should include:

1. More comprehensive economic analyses inclusive of all forecasting time scales and indirect cost savings (some of this work is being considered by DOE as a contingency to release of the economic studies undertaken by the AWST and WindLogics teams);
2. Identification of the ultimate recipients of economic benefits;
3. Additional, more extensive analysis of phenomena responsible for ramp events/outliers by leveraging one-year's worth of continuous remote sensing/surface observation field data;
4. Further analysis of the forecast model performance, including
 - a. A more in-depth study of the enhanced observation network value, through data denial efforts focused on particular phenomena (i.e. convectively-driven and LLJs) and the largest ramp events;
 - b. An inter-comparison of the data denial experiments for the Northern and Southern Study Areas that further quantifies the differences (and forecast value) of having a denser network of remote sensing instrumentation (as in the NSA);
 - c. A regime-based analysis (i.e. cold season versus warm season);
5. Additional sensitivity experiments to determine which component (see Table 7-1) of the WFIP system contributed most to more accurate forecasts, in terms of overall improvement, regime-based performance, and selected ramp events (particular focus should be on data assimilation schemes such as 4DVAR).

General recommendations include:

1. Embrace efforts by DOE and NOAA to use information learned from this project to expand existing network capabilities or deploy new sensor networks in regions already populated by or undergoing an expansion of wind farms, or those areas set to deploy such facilities;
2. Apply lessons learned from modeling sensitivity studies to further enhance forecasting efforts and reduce power production forecast errors (and increase cost savings) on a variety of temporal and spatial scales;
3. Further engage stakeholders (BAs, utilities, ISOs, wind energy non-governmental organizations) in outreach efforts to better quantify the value of improved forecasts to their interests; and
4. Ensure that the lessons learned here (e.g. thoroughness of an economic analysis) are applied to future WFIPs.

9 References

Ancell, B.C., 2012: Examination of analysis and forecast errors of high-resolution assimilation, bias removal, and digital filter initialization with an ensemble Kalman filter. Mon. Wea. Rev., 140, 3992 – 4004.

Ancell, B.C., 2013: Nonlinear characteristics of ensemble perturbation evolution and their application to forecasting high-impact events. Wea. and Forc., in press.

Ancell, B.C., and L.A. McMurdie, 2013: Ensemble adaptive data assimilation techniques applied to land-falling North American cyclones. *Data Assimilation for Atmospheric, Oceanic, and Hydrologic Applications*, Volume 2, Springer, 555 - 575.

Anderson, J.L., and S.L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741 – 2758.

Anderson, J.L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884 – 2906.

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72 - 83.

Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The data assimilation research testbed: A community data assimilation facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283-1296.

AWEA 2013: American Wind Energy Association, "U.S. Wind Industry Annual Market Report Year Ending 2012," 2013.

Banta, R.M., R.K. Newsom, J.K. Lundquist, Y.L. Pichugina, R.L. Coulter, and L. Mahrt (2002), Nocturnal low-level jet characteristics over Kansas during CASES-99, *Bound.-Layer Meteor.* **105**, 221-252.

Barker, D., X.-Y. Huang, Z. Liu, T. Auligné, X. Zhang, S. Rugg, R. Ajjaji, A. Bourgeois, J. Bray, Y. Chen, M. Demirtas, Y.-R. Guo, T. Henderson, W. Huang, H.-C. Lin, J. Michalakes, S. Rizvi, and X. Zhang, 2012: The Weather Research and Forecasting model's community variational/ensemble data assimilation system: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 831 – 843.

Benoit, R., J. Côté, J. Mailhot, 1989: Inclusion of a TKE Boundary Layer Parameterization in the Canadian Regional Finite-Element Model. *Mon. Wea. Rev.*, **117**, 1726–1750.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using Incremental Analysis Updates. *Mon. Wea. Rev.*, **124**, 1256–1271

Bonner, W. D., 1968: Climatology of the low level jet. *Mon. Wea. Rev.*, **96**, 833-850.

Bossanyi, E.A., 1985: Short-Term Wind Prediction Using Kalman Filters. *Wind Engineering*, **9**(1), 1–8.

Bretherton, C. S., and S. Park, 2009: A new moist turbulence parameterization in the Community Atmosphere Model. *J. Climate*, **22**, 3422-3448

Brewster, K., 1996: Application of a Bratseth analysis scheme including Doppler radar data. Preprints, 15th Conference on Weather Analysis and Forecasting. Amer. Meteor. Soc., Norfolk, VA, 92-95.

Brewster, K. A., J. Brotzge, K. W. Thomas, Y. Wang, M. Xue, J. Gao, D. Weber, and K. Howe, 2008: High resolution assimilation of CASA and NEXRAD radar data in near-real time: Results from Spring 2007 and plans for Spring of 2008. 12th Conf. IOAS-AOLS, New Orleans, LA, Amer. Meteor. Soc

Brewster, K., M. Hu, M. Xue, and J. Gao, 2005: Efficient assimilation of radar data at high resolution for short range numerical weather prediction. World Weather Research Program Symposium and Nowcasting and Very Short-Range Forecasting WSN05, Toulouse, France. WMO World Weather Research Program, Geneva, Switzerland. Symposium CD, Paper 3.06.

Broersen, P. M. 2006: Automatic spectral analysis with missing data. *Digital Signal Processing*, **16**(6), 754-766.

Cardell, J.B.; Anderson, C.L. 2009: Estimating the system costs of wind power forecast uncertainty. Power & Energy Society General Meeting, 2009. PES '09. IEEE.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model description and implementation. *Mon. Wea. Rev.*, **129**, 569 – 585.

Cherkassky, V., X. Shao, F. Mulier, V. Vapnik, 1999: Model complexity control for regression using VC generalization bounds. *IEEE Transactions on Neural Networks*, **10**(5), 1075–1089.

Cherkassky, V., X. Shao, 2001: Signal estimation and denoising using VC-theory. *Neural Networks*, **14**, 1, 37-52

Cherkassky, V., F. Mulier, 2007: *Learning from data: Concepts, theory, and methods*, 2nd ed., John Wiley & Sons. Inc., Hoboken, New Jersey.

Demirtas, M., D. M. Barker, Y. Chen, J. Hacker, X.-Y. Huang, C. Snyder, and X. Wang, 2009: A hybrid data assimilation system (ensemble transform Kalman filter and WRF-VAR) based retrospective tests with real observations. Preprints, 19th Conf. on Numerical Weather Prediction, Omaha, NE, Amer. Meteor. Soc., 9A.5. Everitt B.S., 2002: Cambridge Dictionary of Statistics, CUP ISBN: 9780521690270

DOE 2013: 2012 Wind Technologies Market Report. Available at <http://emp.lbl.gov/sites/all/files/lbnl-6356e.pdf>.

DOE 2008: 20% Wind Energy by 2030: Increasing Wind Energy's Contribution to U.S. Electricity Supply. DOE/GO-102008-2567. Available at <http://www.nrel.gov/docs/fy08osti/41869.pdf>.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077 – 3107.

Fabbri, A., and T. G. San Rom, J. R. Abbad, and V. H. Méndez Quezada, 2005: Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market. *IEEE Trans. Power Systems*, **20**, 1440 – 1446.

Freedman, J. M. and D. R. Fitzjarrald, 2001: Post-frontal air mass modification. *J. Hydrometeorology*, **2**, 419-437.

- Freedman, J., M. Markus, and R. Penc, 2008: Analysis of West Texas wind plant ramp-up and ramp-down events. AWS Truewind Report. 250-278.
- Gage, K. S., and G. D. Naström, 1986: Theoretical interpretation of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft during GASP. *J. Atmos. Sci.*, 43, 729–740.
- Gao, J., M. Xue, K. Brewster, and K. K. Droegemeier 2004: A three-dimensional variational data assimilation method with recursive filter for single-Doppler radar, *J. Atmos. Oceanic Technol.* 457-469
- Gaspari, G., and S.E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125**, 723-757.
- Grachev, A. A., Fairall, C. W., Persson, P. O. G., Andreas, E. L., & Guest, P. S. 2005: Stable boundary-layer scaling regimes: The SHEBA data. *Boundary-layer meteorology*, **116**(2), 201-235.
- Grell, G. A., and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**(14), Article 1693.
- Hacker, J. P., and C. Snyder. 2005. Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL. *Mon. Wea. Rev.*, **133**, 3260-3275.
- Hong, S-Y., and J-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- Hamill, T., F. Yang, C. Cardinali, and S. Majumdar, 2013: Impact of targeted winter storm reconnaissance dropwindsonde data on mid-latitude numerical weather prediction. *Mon. Wea. Rev.*, **141**, 2058-2065.
- Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103 – 120.
- [Hu, M., M. Xue, and Keith Brewster](#), 2006: 3DVAR and cloud analysis with WSR-88D Level-II data for the prediction of the Fort Worth tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134**, 675-698.
- Iacono, M.J., J.S. Delamere, E.J. Mlawer, M.W. Shephard, S.A. Clough, and W.D. Collins, Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *J. Geophys. Res.*, **113**, D13103
- Kain, J. S. and J. M. Fritsch. 1990. A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.* **47**, 2784–2802
- Kain, J.S., and J.M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain-Fritsch scheme. The Representation of Cumulus Convection in Numerical Models, K.A. Emanuel and D.J. Raymond, American Meteorological Society, 246 pages

Kain, John S., 2004: The Kain–Fritsch Convective Parameterization: An Update. *J. Appl. Meteor.*, **43**, 170–181.

Koenker, R. W., 1994: Confidence Intervals for regression quantiles, in P. Mandl and M. Huskova (eds.), *Asymptotic Statistics*, 349–359, Springer-Verlag, New York.

Lambert, W. C., Merceret, F. J., Taylor, G. B., & Ward, J. G. 2003: Performance of five 915-MHz wind profilers and an associated automated quality control algorithm in an operational environment. *Journal of Atmospheric & Oceanic Technology*, **20**(11), 1488-1495

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1089

Lynch, Peter, Xiang-Yu Huang, 1992: Initialization of the HIRLAM Model Using a Digital Filter. *Mon. Wea. Rev.*, **120**, 1019–1034.

Makarov, Y.V., R. T. Guttromson, Z. Huan, K. Subbarao, P. B. Etingov, B. B. Chakrabarti and J. Ma, 2010: Wind Energy Management System Integration Project. PNNL-19189.

Manobianco, J., J.W. Zack and G.E. Taylor. 1996. Workstation-based real-time mesoscale modeling designed for weather support to operations at the Kennedy Space Center and Cape Canaveral Air Station. *Bull. Amer. Meteor. Soc.*, **77**, 653-672

Mlawer, E.J., S.J. Taubman, P.D. Brown, M.J. Iacono, and S.A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663 – 16682

Murphy, A. H., 1969: On the Ranked Probability Score. *J. Appl. Meteor.*, **8**, 988–989.

Nakanishi, M. and H. Niino, 2006: An improved Mellor-Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397-407

Naström, G. D., and K. S. Gage, 1985: A climatology of atmospheric wavenumber spectra observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950–960.

Noilhan, J. and S. Planton, 1989: A Simple Parameterization of Land Surface Processes for Meteorological Models. *Mon. Wea. Rev.*, **117**, 536–549.

Oncley, S. P., Friehe, C. A., Larue, J. C., Businger, J. A., Itsweire, E. C., & Chang, S. S. 1996: Surface-layer fluxes, profiles, and turbulence measurements over uniform terrain under near-neutral conditions. *Journal of the Atmospheric Sci.*, **53**(7), 1029-1044

Parrish, D.F., and J.C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis system. *Mon. Wea. Rev.*, **120**, 1747 – 1763.

- Press W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery 2002: Numerical Recipes in C: The Art of Scientific Computing: 2nd Edition. Cambridge University Press, pp. 671-681.
- Rutledge, Steven A., Peterv Hobbs, 1983: The Mesoscale and Microscale Structure and Organization of Clouds and Precipitation in Midlatitude Cyclones. VIII: A Model for the “Seeder-Feeder” Process in Warm-Frontal Rainbands. *J. Atmos. Sci.*, **40**, 1185–1206.
- Savijärvi, Hannu, 1990: Fast Radiation Parameterization Schemes for Mesoscale and Short-Range Forecast Models. *J. Appl. Meteor.*, **29**, 437–447.
- Schreck, S., J. Lundquist, and W. Shaw 2008: U.S. Department of Energy Workshop Report: Research Needs for Wind Resource Characterization. Technical Report NREL/TP-500-43521. Available at <http://www.nrel.gov/docs/fy08osti/43521.pdf>.
- Schroeder, J. L., W. S. Burgett, K. B. Haynie, I. Sonmez, G. D. Skwira, A. L. Doggett, and J. W. Lipe, 2005: The West Texas Mesonet: A Technical Overview. *J. Atmos. Oceanic Technol.*, **22**, 211-222.
- Schwartz, C. S., and L. Zhiqian, 2014: Convection-permitting forecasts initialized with continuously cycling limited-area 3DVAR, ensemble kalman filter, and “hybrid” variational–ensemble data assimilation systems. *Mon. Wea. Rev.*, **142**, 716–738.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers. 2005. A description of the Advanced Research WRF Version 2. NCAR Tech Notes-468+STR
- Smolarkiewicz, P. K. and L. G. Margolin, 1998: MPDATA: A Finite-Difference Solver for Geophysical Flows, *J. Comp. Phy.*, **140**, 459-480
- Song, J., K. Liao, R. L. Coulter, B. M. Lesht, 2005: Climatology of the Low-Level Jet at the Southern Great Plains Atmospheric Boundary Layer Experiments Site. *J. Appl. Meteor.*, **44**, 1593–1606.
- Stauffer, D. R., and N. L. Seaman. 1990. Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Mon. Wea. Rev.*, **118**, 1250-1277.
- Stephens, G. L. and P. J. Webster, 1979: Sensitivity of Radiative Forcing to Variable Cloud and Moisture. *J. Atmos. Sci.*, **36**, 1542-1556.
- Sun, W.Y. and C-Z Chang, 1986: Diffusion model for a convective layer: Part I. Numerical simulation of a convective boundary layer. *J. Climate Appl. Meteor.* **25**, 1445-1453.
- Tao, W.-K., and J. Simpson (1993), The Goddard Cumulus Ensemble Model. Part I: Model description, *Terr. Atmos. Oceanic Sci.*, **4**, 19–54.
- Taylor, K.E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183-7192.

- Therry, G., and P. Lacarrère, 1983: Improving the eddy kinetic energy model for planetary boundary layer description. *Bound.-Layer Meteor.*, **25**, 63–88.
- Thompson, G., R.M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519 – 542.
- Torn, R.D., G.J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490 – 2502
- Torn, R. D., and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663-677.
- Wang, Xuguang, David Parrish, Daryl Kleist, Jeffrey Whitaker, 2013: GSI 3DVar-Based Ensemble–Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution Experiments. *Mon. Wea. Rev.*, **141**, 4098–4117.
- Weather Research and Forecasting (WRF) Developmental Testbed Center, 2010: Gridpoint Statistical Interpolation (GSI), Version 2.0 User's Guide. 90 pages.
- Wendell, L.L, H. L Wegley, and M. G. Verholek, 1978: *Report from a working group meeting on wind forecasts for WECS operation*. doi:10.2172/6548011
- Wilks, Daniel S., 1995: Statistical Methods in the Atmospheric Sciences: An Introduction. Academic Press, 467pp.
- Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Physics*. **75**, 161-193.
- Xue, M., K. Droegemeier, V. Wong, A Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu and D. Wang. 2001. The Ad Advanced Regional Prediction System (ARPS) - A multi-scale nonhydrostatic atmospheric simulation and prediction model. Part II: Model physics and applications. *Meteorol. Atmos. Phys.*, **76**, 143-165
- Xue, M., D.-H. Wang, J. Gao, K. Brewster, and K.K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS) – storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Physics*, **82**, 139-170.
- Xue, M., J. Zong, and K. K. Droegemeier, 1996: Parameterization of PBL turbulence in a multi-scale non-hydrostatic model. Preprint, 11th AMS conference on Numerical Weather Prediction. Amer. Metero. Soc., Norfolk, VA., 363-365.
- Zack, J., S. Young, and E. Natenberg, 2011: Evaluation of wind ramp forecasts from a rapid-update dynamical-statistical ramp prediction system. European Wind Energy Association, Annual Event, Brussels, Belgium.

Zhu, K., Y. Pan, M. Xue, X. Wang, J. S. Whitaker, S. G. Benjamin, S. S. Weygandt, and M. Hu, 2013: A regional GSI-based EnKF system for the Rapid Refresh configuration: Results with a single, reduced resolution. *Mon. Wea. Rev.*, 141, 4118-4139.

Appendices

9.1 Appendix A - Publications

Ancell, B.C., 2012: Examination of analysis and forecast errors of high-resolution assimilation, bias removal, and digital filter initialization with an ensemble Kalman filter. *Monthly Weather Review*, Vol. 140, No. 12, pages 3992-4004.

Ancell, B.C., E. Kashawlic, and J.L. Schroeder, 2014: Evaluation of wind forecasts and observation impacts from variational and ensemble data assimilation for wind energy applications. Submitted to *Monthly Weather Review*, in review.

Kiliyanpilkil, V. P. and Basu, S.: An algorithm based on the statistical learning theory for denoising and gap-filling of SoDAR wind data, *Wind Energy*, (to be submitted).

Kiliyanpilkil, V. P., and Basu, S.: Mesoscale wind speed spectra of lower atmospheric boundary layer, *Environmental Research Letters*, (to be submitted).

Wilczak, J. et al., 2014: The Wind Forecast Improvement Project (WFIP): A Public-Private Partnership Addressing Wind Energy Forecast Needs

9.2 Appendix B - Conferences, Workshops, and User Group Meetings

Utility Variable Integration Group (UVIG) Fall Workshop, Maui, Hawaii, 12 - 14 October 2011
American Meteorological Society (AMS) 3rd Conference on Weather, Climate, and the New Energy Economy, New Orleans LA, 22 - 26 January 2012
WFIP Southern Study Area 1st User Group Meeting, Tucson, AZ, 7 February 2012
UVIG Workshop on Forecasting Applications, Tucson AZ, 8 - 9 February 2012
Western Electric Coordinating Council, March 2012
American Wind Energy Association (AWEA), Wind Power 2012, 3 – 6 June 2012 Atlanta, GA
AMS 20th Symposium on Boundary Layer and Turbulence, Boston MA, 9 – 13 July 2012
16th International Symposium for the Advancement of Boundary-Layer Remote Sensing (ISARS), Boulder, Colorado, USA, 5–8 June, 2012.
11th International Workshop on Large Scale Integration of Wind Power into Power Systems, 13 – 15 November 2012, Lisbon, Portugal
AMS 4th Conference on Weather, Climate, and the New Energy Economy, Austin TX, 6 – 10 January 2013
AMS 6th Symposium on LiDAR Atmospheric Applications, Austin TX, 6 - 10 January 2013
WFIP Southern Study Area 2nd User Group Meeting, Salt Lake City, UT, 27 February 2012
UVIG Workshop on Forecasting Applications, Salt Lake City, UT 27 – 28 February 2013
AWEA Wind Power 2013, Chicago IL, 6 – 8 May 2013
AMS 5th Conference on Weather, Climate, and the New Energy Economy, Atlanta GA, 3 – 6 February 2014
UVIG Forecasting Workshop, Tucson AZ, 26 – 27 February 2014

List of Tables

Table 2-1. WFIP Southern Study Region Team members: capabilities, qualifications, and expertise.....	9
Table 3-1. Ramp (MW) thresholds for 15-, 60-, and 180-minute periods for WFIP wind farm regions and the WFIP aggregate.....	11
Table 3-2. WFIP instrument deployment (July 2011 – September 2012). Shaded rows indicate grouping by IOS.....	22
Table 3-3 Comparison of NCSU SoDAR-based and TTU tower-based wind data for the period 1 July 2011 – 12 September 2012.....	26
Table 4-1. AWST model members and static configuration.....	36
Table 4-2. Table of AWST model members and run configuration parameters, ARPS members IDs are denoted as DOEA1-3, MASS members are denoted, DOEM1-3, and WRF members are denoted as DOEW1-3, with the last integer representing each member of the ensemble and model.....	37
Table 4-3: List of ensemble forecast methods for the WFIP project by type and included members. The real-time WFIP forecast method is optimized ensemble (OPTENS) and the baseline forecast method is OPTENS_STWP (short term wind power forecast). To compare sensitivity to model members; ENS_WHRR, ENS_NHRRR, ENS_BWHRRR, and ENS_BNHRRR were run but without time varying weighting as used in the OPTENS method.....	39
Table 4-4: Table of observations assimilated into the AWST ensemble members. Project data are highlighted in grey.....	41
Table 4-5: List of model members for the data denial sensitivity analysis.....	42
Table 4-6: List of AWST and ESRL data denial periods.....	42
Table 4-7. Parameter values for the DART EnKF and GSI 3DVAR systems.....	47
Table 5-1. Time averaged aggregate ensemble member performance for three forecast methods. Error is RMSE as a percent of nameplate capacity of the aggregate. Larger errors are denoted with darker red shading, The RF MYNN PA method was unavailable during the forecast project due to an error in the configuration software.....	56
Table 5-2: List of days covering each ramp event during the winter and summer periods. An asterisk (*) denotes that more than one ramp event occurred during the listed date.....	66
Table 7-1: The NWP baseline and WFIP ensemble forecast attributes.....	88
Table 7-2: Comparison of real-time and historical observation denial experiments.....	88
Table 7-3: Baseline ELRAS and WFIP ensemble ramp forecast attributes.....	89

List of Figures

Figure 3-1. Observed wind power production (expressed by capacity factor in %) for wind farm aggregates (varying colors and line types) and all WFIP wind farms (solid black line) during a large up-ramp event on 7 - 8 September 2012.....	12
Figure 3-2. Radar imagery of a thunderstorm complex near Sweetwater aggregates on 17 June 2012 2340 UTC. Markers depicting points of interest include: instrument sites (light blue), and ERCOT wind farms (red). Note the line of convection-induced outflow to the north east of Sweetwater.....	13
Figure 3-3. Time-height cross-section of 10-minute horizontal wind speed from the AWST SoDAR unit at Reagan on 5 May 2012. Note the low-level jet feature (warmer colors--higher wind speeds) 03 UTC - 12 UTC (9PM - 6AM CST)._____	14
Figure 3-4. Wind speeds for the TTU 200-m tower (large red circles), the NCSU SoDAR (large black diamonds), and the TTU 915 MHz wind profiler (green circles) for 4 January 2012 0400 UTC. Other profiles (symbols and lines) are from the individual MESO WFIP forecast system members. Dot-dash lines represent neutral (black) and stable (red) profiles for $z_0 = 0.1$ and $u^* = 0.4$. Y-axis is logarithmic._____	14
Figure 3-5. Observed power output by capacity factor (%) for regional wind farm aggregates and all WFIP for days 178 -183 (26 June - 1 July) 2012. Horizontal red dash-dot line depicts average capacity factor (35%) for the period._____	15
Figure 3-6. Existing surface observation networks within and adjacent to the region of interest (outlined in thick blue line). Dashed lines represent approximately 200 km grid squares._____	16
Figure 3-7. The mean predicted fractional wind speed RMSE reduction (color shading) for all seasons, forecast sites, and look-ahead times for a site with a collocated profiler, sodar, and flux station. Existing and recommended sensor locations area also shown. Recommended sensor sites were chosen based on the overall predicted error reduction, predicted error reduction for specific look-ahead times and seasons, and location of existing sensors. Only a representative sample of existing met towers and surface stations is shown. Background color contours depict the mean predicted fractional wind speed RMSE reduction for all seasons, forecast sites, and look-ahead times for a site with a collocated profiler, SoDAR, and flux station._____	18

Figure 3-8. DNOAA/PNNL/AWST WFIP deployment with existing profiler, NEXRAD, and upper air (rawinsonde) networks. IOSS suites are referenced in Table 3-2. Solid red line (outlined rectangle) shows the ERCOT ROI and center of the model domain. Figure courtesy of J. Wilczak, NOAA ESRL.	19
Figure 3-9. Regional aggregates of wind farms in the WFIP SSA (ERCOT domain). Aggregates are defined as NTX: North Texas; SWN: Sweetwater North; SWE: Sweetwater East; SWW: Sweetwater West; ETX: East Texas; WTX: West Texas; MCM: McAmey; SGL: San Angelo.	20
Figure 3-10 (left): Texas Tech 200-m tall tower at Reese Technology Center; (above right): lowest three levels of instrumentation on the 200 m tower showing the various instrumentation on each boom arm; (above left): TTU Vaisala LAP-3000 Boundary Layer Profiler	23
Figure 3-11 (left): the deployment locations of NCSU SoDAR, TTU 200 m meteorological tower, and TTU wind profiler; (right): an image of the NCSU SoDAR (red box). The 200 m tall tower is visible in the background.	23
Figure 3-12 (top left): AWST ART-1 SoDAR commissioning at Reagan TX; (top right) SoDAR commissioning at Ozona TX;(bottom left): PNNL surface met station at Cleburne TX with AWST ART-1 SoDAR in background; (bottom right): NOAA flux station at Colorado City TX.	24
Figure 3-13. Deployment of Leosphere Windcube LiDARs at a wind farm in Texas.	25
Figure 3-14. Comparison of NCSU SoDAR-based (left panel) and TTU tower-based (right panel) wind data. From top to bottom, the panels represent data availability, mean wind speed, histograms, and wind roses, respectively. Time frame: 1 July 2011 – 12 September 2012.	27
Figure 3-15. Monthly occurrence of ramp events in the WFIP aggregate region meeting or exceeding the medium ramp rate threshold, separated by duration of event: 180-minute (blue), 60-minute (green), and 15-minute (red).	28
Figure 3-16. Power spectral density using FFT (red lines), WT (blue lines with circles), and ARMAsel (black lines) for SoDAR wind speed from 1 July, 2011 to 12 September 2012. The top and bottom panels correspond to altitudes at 18m and 123m above ground level, respectively. The dashed orange lines represent a slope of -5/3.	29
Figure 3-17. Proposed gap-filling and de-noising algorithm (see text for details).	30
Figure 3-18. Time-height plots of average wind speed from a wind profiler (top-left panel), a SoDAR (bottom-left panel), and a meteorological tower (bottom-right panel). The gap-filled SoDAR-profiler wind data are shown in top-right panel. Time of interest: 30 January 2012.	31
Figure 3-19. Time-height plots of data availability (top panels) and average wind speeds (bottom panel). Time frame: 30 November 2011 – 2 April 2012. See text for details.	32
Figure 4-1: Overview of the WFIP forecast process.	35
Figure 4-2: Domain of the WRF two-way nested and MASS one-way nested runs. The outer grid has 15-km grid space while the inner grid has 5-km grid spacing. For ARPS, only the inner grid shown was run.	37
Figure 4-3: Assimilation and forecast timeline for one member of the 0000 and 0200 UTC ARPS, MASS and WRF WFIP cycle, the hatched area denotes the part of the forecast that was unavailable for real-time applications.	38
Figure 4-4, Map of the CAPS 3-km forecast domain (1200 x 1080 km), with the WFIP-South Study Area in the inner box.	43
Figure 4-5. Vertical spacing of mass and horizontal wind levels in the lowest 500 m. Left: the ARPS production runs for WFIP, right: typical vertical spacing in real-time high-resolution models.	44
Figure 4-6: Schematic of data assimilation and forecast timeline for a sample run time (1200 UTC) showing IAU applied with a triangular time weighting scheme in the 15-min pre-forecast period and a 12-h forecast.	44
Figure 4-7. The 12km modeling domain used in this study. The black circle indicates the ramp verification area containing the 80-m meteorological towers, and the black box shows the verification area for the month-long experiments. A typical distribution of routine assimilated observations is also shown by the colored circles.	45
Figure 4-8. Two examples of tuning runs for DART EnKF vertical localization and GSI 3DVAR horizontal localization performed over the first week of December 2011.	46
Figure 5-1. Forecast system sample size as a percent of maximum number of possible model runs for the each of the 10 high-resolution WFIP NWP members.	51
Figure 5-2. Root Mean Square Error (RMSE) of 0- to 13-hour power production forecasts for the WFIP system-wide aggregate as a percent of nameplate capacity for the each of the 10 high-resolution WFIP NWP members. The “zero” represents the time of model initialization,	52
Figure 5-3. Root Mean Square Error (RMSE) in m/s of 0- to 13-hour capacity-weighted forecasts of average 80-m wind speed for the WFIP system-wide aggregate for the each of the 10 high-resolution WFIP NWP members. The “zero” time represents model initialization.	52
Figure 5-4: Bias of the 0- to 13-hour power production forecasts for the WFIP system-wide aggregate as a percent of nameplate capacity for the each of the 10 high-resolution WFIP NWP members. The “zero” time represents model initialization,	53

Figure 5-5: Forecast system aggregate power RMSE as a percent of nameplate capacity for unadjusted (top) and MOS (bottom) 15-minute forecast methods for each of the high-resolution WFIP NWP members.	54
Figure 5-6: Forecast system aggregate power bias as a percent of nameplate capacity for unadjusted (top) and MOS (bottom) 15-minute forecast methods for each of the 10 high-resolution WFIP NWP members.	55
Figure 5-7: Comparison of the percent improvement in forecast performance as a function of forecast look-ahead time using the OPTENS versus OPTENS_STWPF (baseline) method. The values represent the percent reduction in RMSE of the OPTENS_WFIP over that of the OPTENS_STWPF method.	57
Figure 5-8: RMSE (% capacity) of six ensemble forecasts of the WFIP system-wide aggregate power production (discussed in Chapter 4.1.1 and Table 4-3).	57
Figure 5-9. Taylor diagram showing individual ensemble member, STWPF (open upside down black triangle), and WFIP (optimized ensemble, depicted by the open green triangle) 3-hr forecast performance as compared with observations (black asterisk). Thin grey solid lines represent the centered RMSE, black dotted lines depict the coefficient of determination (R^2), and blue dotted lines show the standard deviation.	58
Figure 5-10: Reliability diagram for all individual WFIP member ramp forecasts, and ramp forecasts generated from a combination of WFIP forecast members (all, best 3, all w/o HRRR). For reference, a perfect forecast is shown in orange.	59
Figure 5-11: Ranked Probability Skill Score (RPSS) for all individual WFIP member ramp forecasts, and ramp forecasts generated from a combination of WFIP forecast members (all, best 3, all w/o HRRR).	61
Figure 5-12: Ranked Probability Skill Score (RPSS) contribution averaged over all forecast times shown by 60-minute ramp rate. RPSS contribution is shown for all individual WFIP member ramp forecasts and ramp forecasts generated from a combination of WFIP forecast members (all, best 3, all w/o HRRR). Observed ramp magnitude values along the x-axis are in MW/hour units.	62
Figure 5-13: The RMSE of power production forecasts for the WFIP system-wide aggregate for the AWST data denial experiments for a 30-day period (top) December and (bottom) April.	63
Figure 5-14: The RMSE of capacity weighted average wind speed forecasts for the WFIP system-wide aggregate for the AWST data denial experiments for a 30-day period (top) December and (bottom) April.	64
Figure 5-15: RMSE of forecasts of the WFIP system-wide aggregate (a) power generation (% of capacity) and (b) wind speed (m/s) for data denial experiments that included denial of project data to the ESRL RR data assimilation process for one week in December 2011.	65
Figure 5-16: Critical Success Index (CSI) for ramp event forecasts for the November 2011 – April 2012 (panels a and b) and April 2012 – September 2012 (panels c and d) periods. Panels (a) and (c) are for 60 minute ramp rates using a 30-minute “hit” window and panels (b) and (d) are for 180 minute ramp rates using a 60-minute “hit” window.	68
Figure 5-17. Diurnal forecast speed mean absolute error (MAE) of WFIP ensemble forecast members for varying time-ahead forecast hours.	69
Figure 5-18. Diurnal capacity factor bias for the WFIP (left) and baseline (right) forecasts, decomposed by region and season. Seasons are as follows: NDJF (red), MAMJ (green), JASO (blue).	70
Figure 5-19. Diurnal mean absolute error (MAE) of capacity factor for the WFIP (left) and baseline (right) forecasts, decomposed by season and time-ahead forecast hour. Seasons are as follows: NDJF (dashed red), MAMJ (solid green), JASO (dotted blue).	71
Figure 5-20. Diurnal bias of the 15-minute change in capacity factor for the WFIP (left) and baseline (right) forecasts, decomposed by region and season. Seasons are as follows: NDJF (red), MAMJ (green), JASO (blue).	72
Figure 5-21. Diurnal mean absolute error (MAE) of the 15-minute change in capacity factor for the WFIP (left) and baseline (right) forecasts, decomposed by season and time-ahead forecast hour. Seasons are as follows: NDJF (dashed red), MAMJ (solid green), JASO (dotted blue).	72
Figure 5-22. Time-series of observed (black line) and forecasted hub height wind speed (ms-1). Production configuration (P, red squares) and reduced vertical resolution (V, green circles) for 30-31 Dec 2011.	75
Figure 5-23. As in Figure 5-22, but for 07-08 May 2012	75
Figure 5-24. As in Figure 5-23, but for 07-08 May 2012	75
Figure 5-25. Mean absolute forecast error (ms ⁻¹) for turbine height wind speed at the Texas Tech Tall Tower over all ramp cases in WFIP study period. CAPS ARPS production forecast (P), special data withheld (W), and reduced vertical resolution (V).	75
Figure 5-26. Mean Absolute Speed Error (ms ⁻¹) as a function of height and forecast time for all ramp cases for the CAPS Production forecasts.	76
Figure 5-27. Mean Absolute Speed Error (ms ⁻¹) as a function of height and forecast time for all ramp cases for the reduced vertical resolution forecasts.	76
Figure 5-28. The mean absolute 0-24hr wind speed errors for the EnKF (green line) and 3DVAR (red line) month-long routine observation experiments averaged over 120 assimilation cycles in December 2011.	77
Figure 5-29. The 500-hPa background field (black contours, contour interval is 30 m) and analysis increments (shaded) in the top row, and the sea-level pressure background field (black contours, contour interval is 2 hPa), background 80-m wind barbs, and DE- EE0004420	77

<i>80-m meridional (V wind) analysis increments (shaded) in the bottom row for the EnKF and 3DVAR analyses valid at 1200 UTC December 1, 2011.</i>	78
<i>Figure 5-30. The mean absolute EnKF (green line) and 3DVAR (red line) wind speed errors for all 120 assimilation cycles in December 2011 at both 6hr and 24hr forecast time for the month-long routine observation experiments</i>	79
<i>Figure 5-31. The mean absolute 0-24hr wind speed errors for the month-long control run (black line, routine observations only), the run that assimilates mesonet observations in addition to routine observations (cyan line), and the run that assimilates profiler/sodar observations in addition to the routine observations (red line) for both EnKF and 3DVAR.</i>	80
<i>Figure 5-32. The background (6-hr forecast) mean wind speed errors for all assimilation cycles measured against 80-m tower data for EnKF (green line) and 3DVAR (red line), the mean EnKF surface wind speed errors for the same cycles measured against surface METAR data, and the surface analysis bias for the EnKF control run (black line) as well as the run assimilating profiler/sodar observations (red line).</i>	81
<i>Figure 5-33. Example wind ramp speed and direction forecasts from the EnKF (blue line) and 3DVAR (black line) systems initialized at 1200 UTC December 2, 2011 valid at the location of a single 80 m meteorological tower. The observed wind speed and direction from the tower are also shown.</i>	82
<i>Figure 5-34. Example wind ramp speed and direction forecasts from the EnKF (blue line) and 3DVAR (black line) systems initialized at 1200 UTC December 2, 2011 valid at the location of a single 80-m meteorological tower. The observed wind speed and direction from the tower is also shown.</i>	83
<i>Figure 5-35. Ramp timing and duration errors averaged over all wind ramp cases for the EnKF and GSI routine and observation impact experiments.</i>	84

List of Acronyms

ANL – DOE/Argonne National Laboratory
ARL – NOAA/Air Resources Laboratory
APRS - Advanced Regional Prediction System
ASOS – Automated Surface Observing System
AWST – AWS Truepower LLC.
CSI – Critical Success Index
CONUSNEST -NAM contiguous United States 4 km nest
DART - Data Assimilation Research Testbed
EnKF – Ensemble Kalman Filter
ERCOT - Energy Reliability Council of Texas
ESRL – NOAA/Earth Systems Research Laboratory
FFT – Fast Fourier Transform
GDAS – Global Data Assimilation System
GFS – Global Forecast System
GSI – Gridpoint Statistical Interpolation
HRRR – High Resolution Rapid Refresh model
LiDAR – Light Detection And Radar
LLNL – DOE/Lawrence Livermore National Laboratory
MAE – Mean Absolute Error
MASE – Mean Absolute Speed Error
MASS – Mesoscale Atmospheric Simulation System
METAR – Météorologique Aviation Régulières (translated to English: Meteorological Terminal Aviation Routine Weather Report)
MHz – Mega-Hertz
MISO - Midwest Independent System Operator
MOS – Model Output Statistics
MPP – Multi-Peak Picking
NAM – North American Mesoscale forecast system
NCEP – NOAA/NWS/National Centers for Environmental Prediction
NCSU – North Carolina State University
NDAS – NAM Data Assimilation System
NMMB – Non-hydrostatic Multiscale Model on the B grid
NWP – Numerical Weather Prediction
NSA – Northern Study Area (WindLogics domain)
NREL – DOE/National Renewable Energy Laboratory
NWP – Numerical Weather Prediction
NWS – NOAA/National Weather Service
OU CAPS – Oklahoma University - Center for Analysis and Prediction of Storms
PNNL – DOE/Pacific Northwest National Laboratory
RAP – Rapid Refresh model (run at NOAA/NCEP)
RASS – Radio Acoustic Sounding System
RFI – Radio Frequency Interference
RMSE – Root Mean Squared Error

RPSS – Ranked Probability Skill Scor
RR – Rapid Refresh model (run at NOAA/ESRL)
RUC – Rapid Update Cycle model
SDSU – South Dakota State University
SoDAR – Sound Detection And Radar
SNR – Signal-to-Noise Ratio
SSA – Southern Study Area (AWS Truepower domain)
STWPF – Short Term Wind Power Forecast
TTU – Texas Tech University
QC – Quality Control
WFIP – Wind Forecast Improvement Project
WPR – Wind Profiling Radar
WRF – Weather Forecasting and Research
3DVar – Three-dimensional variational data assimilation