Final Technical Report
Cover Page

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

Recipient: WindLogics Inc. 607951092

Award Number: DE-EE0004421

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 Northern Study Area.

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

Principle Investigator: Cathy Finley, Senior Scientist, cfinley@windlogics.com, 651-556-4283

Report Submitted by: Lübbert Kruizenga, Controller, luebbert.kruizenga@windlogics.com, 651-556-4274

Date of Report: April 30, 2014

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

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

NREL (Gregory Brinkman, Energy Analysis Engineer, Gregory.Brinkman@nrel.gov, 303-384-7390)

SDSU (Dennis Todey, Associate Professor, Dennis.Todey@sdstate.edu, 605-688-5678)

MISO (Michael McMullen, Director of West Regional Operations, mmcmullen@misoenergy.org, 651-632-8404)

Cost-Sharing Partners: None

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 – Martha Amador

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-EE0004421. 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

Accurate wind energy forecasting is a key component to integrating large amounts of wind energy into the electric grid system, and the National Oceanic and Atmospheric Administration's (NOAA's) suite of weather forecast models is the current backbone of most wind energy forecasting activities in the U.S. Wind energy forecast vendors (including WindLogics) use these forecasts either directly in their wind power forecasts or as inputs into statistical or higher resolution mesoscale models for wind energy forecasting purposes. Thus if wind energy forecasting is to be improved, it is critical to improve the NOAA weather model forecasts both through additional data assimilation and research.

This report contains the results from research aimed at improving short-range (0-6 hour) hub-height wind forecasts in the NOAA weather forecast models through additional data assimilation and model physics improvements for use in wind energy forecasting. Additional meteorological observing platforms including wind profilers, sodars, and surface stations were deployed for this study by NOAA and DOE, and additional meteorological data at or near wind turbine hub height were provided by South Dakota State University and WindLogics/NextEra Energy Resources over a large geographical area (approximately 800km x 800km) in the U.S. Northern Plains for assimilation into NOAA research weather forecast models. The resulting improvements in wind energy forecasts were examined in many different ways to quantify the forecast improvements important to power grid system operators and wind plant owners/operators participating in energy markets.

Two operational weather forecast models (OP_RUC, OP_RAP) and two research weather forecast models (ESRL_RAP, HRRR) were used as the base wind forecasts for generating several different wind power forecasts for the NextEra Energy wind plants in the study area. Power forecasts were generated from the wind forecasts in a variety of ways, from very simple to quite sophisticated, as they might be used by a wide range of both general users and commercial wind energy forecast vendors: a 'raw' forecast generated by directly converting the weather forecast model wind speed to wind power, a 'bias-corrected raw' forecast that used a simple rolling two week average wind speed bias as a first order correction to the wind speed before converting the forecasted wind speed to power, a 'trained' forecast that used sophisticated machine learning techniques and local wind plant data to statistically correct the wind power forecasts, and a 'trained ensemble' which combined data from multiple weather forecast models and local wind plant data to statistically correct the wind power forecasts. The error characteristics of each of these types of forecasts were examined and quantified using bulk error statistics at both for the local wind plant and the system aggregate forecasts. The wind power forecast accuracy was also evaluated separately for high-impact wind energy ramp events.

The overall bulk error statistics calculated over the first six hours of the forecasts at both the individual wind plant and at the system-wide aggregate level over the one year study period showed that the ESRL_RAP-based power forecasts (all types) had lower overall error rates than the other weather forecast model-based power forecasts. At the wind plant level, the HRRR-based power forecasts also had lower error rates than the operational model-based power forecasts at the majority of sites, which is also important to the system operator and of great significance to the wind plant operator. The training process reduced the bulk forecast errors in the power forecasts for each of the model-based power forecasts, both for the individual wind plant forecasts and for the system-wide aggregate forecasts. At most sites, the forecast improvement provided by the training was significant. The trained ensemble power forecasts further improved the error statistics beyond the individual forecast model-based trained power forecasts at each wind plant location, but since the trained ensemble forecasts include the North American Model (NAM), which was not the focus of this study, the results of the trained ensemble forecasts for the system aggregate power forecasts are only briefly presented.

The bulk error statistics of the various model-based power forecasts were also calculated by season and model runtime/forecast hour as power system operations are more sensitive to wind energy

forecast errors during certain times of the year and certain times of day. The results showed that there were significant differences in seasonal forecast errors between the various model-based power forecasts. This was true at both the individual site level, and the system aggregate level. The results from the analysis of the various model-based power forecast errors by model runtime and forecast hour showed that the forecast errors were largest during the times of day that have increased significance to power system operators (the overnight hours and the morning/evening boundary layer transition periods). The HRRR-based power forecasts tend to have smaller forecast errors than the other model-based power forecasts during the higher-error time periods that are also critical time periods for the system operators. Results from the various trained model-based power forecasts also showed that the training process tends to reduce the forecast error the most during the forecast hours that generally have higher forecast error. These analyses collectively show that when it comes to power system operations, the level of complexity in identifying the best weather model for a particular seasonal/daytime regime is yet another argument for using advanced, multi-model ensemble wind power forecasting systems in actual production.

The number of large hourly forecast errors exceeding a significant threshold of rated capacity was also calculated for all model-based power forecasts over the entire year of the study as power grid operators are especially concerned about the extreme events (i.e. the tails of the forecast error distribution). Although these events occur rarely, they can influence the reserves and operating practices used to ensure system reliability. The results showed that the research model-based power forecasts reduce the number of large forecast errors at most wind plant locations compared to operational model-based power forecasts. The training process greatly reduces the number of large forecast errors in all the model-based forecasts, particularly in the ESRL_RAP-based forecasts at the larger forecast errors. This is likely because the training process tends to create smoother forecasts than the raw power forecasts, which reduces the magnitude of the ‘peaks’ and ‘valleys’ in the forecasts where the largest forecast errors often occur.

Wind energy ramp events are also of concern to power system operators as very large wind power forecast errors can occur during these events creating challenges for electric grid system operations. The rate of ramping is of particular concern as there are limits to how quickly conventional generation can be ramped up (or down) to offset the changes in wind generation in order to keep the system balanced. A comprehensive analysis of wind energy forecast errors for the various model-based power forecasts was presented for a suite of wind energy ramp definitions. The results compiled over the year-long study period showed that the bias-corrected raw power forecasts based on the research models (ESRL_RAP, HRRR) more accurately predict wind energy ramp events than the current operational forecast model, both at the system aggregate level and at the local wind plant level. At the system level, the ESRL_RAP-based forecasts most accurately predict both the total number of ramp events and the occurrence of the events themselves, but the HRRR-based forecasts more accurately predict the ramp rate. There was little difference between the various model-based power forecasts in the accuracy of the timing of the ramp events. In general, up-ramp events were more accurately predicted than down-ramp events, both at the system level and individual wind plant level for all model-based forecasts. At the individual site level, the HRRR-based forecasts most accurately predicted the actual ramp occurrence, the total number of ramps and the ramp rates (40-60% improvement in ramp rates over the coarser resolution forecasts). The HRRR-based forecasts also most accurately predicted both the number and occurrence of potentially convective high amplitude/short duration events during the convective season, but tended to over-predict the number of these events leading to a higher number of false alarms than the other model-based power forecasts.

The advantages of geographic dispersion of wind plants on the reduction of system-wide aggregate wind power forecast errors and the number and severity of wind energy ramp events was also demonstrated. The results showed that the system-wide wind energy forecast errors (as defined by the NextEra wind plants in the study area) were roughly half of what would be expected at a typical

individual wind plant. This is a significant reduction in forecast error that results from forecast errors tending to cancel each other out to some extent as the geographic dispersion of the wind plants increases. The number of observed system-wide wind energy ramps was also calculated for varying amounts of geographic dispersion using two different ramp definitions. For each of the ramp definitions, the number of observed ramps in the forecast time series decreased significantly as the geographical diversity of the wind plant locations increased.

The relative importance of the additional data provided in this study on wind energy forecast accuracy was assessed through several data denial experiments in which NOAA 're-forecasted' select time periods using the ESRL_RAP weather forecast model. Two sets of forecasts were generated for each data denial period: one set of forecasts was generated by initializing the forecast using only the operational meteorological data currently available (the 'control forecast'), and one set was generated by initializing the forecast model using the currently available operational meteorological data plus the additional data provided for this project (the 'experimental forecast'). All other aspects of the forecasts were identical. The results showed that there was a clear improvement in forecast accuracy at almost all NextEra wind plant sites in the study area due to the assimilation of the additional meteorological data provided in this study. The greatest improvements (both in the 0-6 hour timeframe, and the 0-15 hour timeframe) were seen in the northern half of the study area, where the operational meteorological observing networks are very sparse compared to the southern portion of the study area.

Unfortunately, the future of the operational NOAA wind profiler network is uncertain due to budget constraints. If the operational network provided similar improvements to wind energy forecasts as was seen in the northern portions of the study area (where the only wind profilers present were those deployed in this study), we would expect to see short-range (0-6 hour) wind energy forecast error increase by 2-5% of energy produced as a result of the loss of this observational network based on the results from this study. While the impacts of the wind profiler network may be somewhat smaller in the southern portions of the study area because of other available data sources, the loss of this important observational network would invariably lead to decreased accuracy in wind energy (and other) forecasts.

1. Introduction

As home to nearly 50% of the country's installed wind generation, the Great Plains represent a critical region for programs that seek to improve meteorological prediction of the wind resource on both large and small time scales. The northern portion of the Great Plains, in particular, has an excellent wind resource and expansive open areas that make it a prime region for current and future wind energy development. Much of this area lies within the footprint of the Midcontinent Independent System Operator (MISO), and MISO already had 10,586 MW of installed wind capacity in North Dakota, Minnesota, South Dakota and Iowa as of the end of 2012 (www.windpoweringamerica.gov).

Accurate wind energy forecasting is a key component to integrating large amounts of wind energy into the electric grid system, and the National Oceanic and Atmospheric Administration's (NOAA's) suite of weather forecast models is the current backbone of most wind energy forecasting activities in the U.S. Wind energy forecast vendors (including WindLogics) use these forecasts either directly in their wind power forecasts or as inputs into statistical or higher resolution mesoscale models for wind energy forecasting purposes. Thus if wind energy forecasting is to be improved, it is critical to improve the NOAA weather model forecasts both through additional data assimilation and research. The focus of this project is to provide data at both the front end and back end of the forecast process in an effort to help NOAA improve this critical backbone in wind energy forecasting, and to show the economic justification for continued wind forecast improvement efforts by quantifying the economic benefit of improved wind energy forecasts on power grid system operations.

The Upper Midwest/Northern Plains area was chosen for this study because:

- It is a region of extensive current (and future) wind energy development.
- The meteorology and forecasting issues found in the region are similar to most of the central U.S. where roughly half of all installed wind capacity currently exists.
- The large penetration of wind in the northern portion of the MISO footprint makes this an ideal region for quantifying the economic impact of wind energy forecast accuracy on utility operations.

Forecasting is challenging in the Upper Midwest and Northern Plains as the region experiences the extremes of both winter and summer weather systems. Ramp events are particularly challenging to forecast in this area because they can be created from a wide variety of weather phenomenon ranging from mid-latitude cyclones and their attendant fronts to smaller scale thunderstorm outflows. Vertical mixing and low-level static stability also can play a major role in creating (or preventing) wind energy ramps, especially during periods of low-level jet activity when the winds at turbine hub-height can be quite strong while the winds at the surface are relatively weak. This is also an area in the U.S. where cold fronts can slow down and stall, turning into slowly meandering "stationary" fronts. Wind energy ramps can occur even during relatively quiescent weather conditions depending on the position of the stationary front relative to the wind plant. In addition to wind energy ramps, other high impact weather events that must be forecasted include turbine icing events, high wind speed turbine cut-out events, and both high and low temperature turbine cut-out events.

The overarching outcome of this project is to identify and quantify the benefits of improved short-term foundational weather forecast models and hub-height level wind forecasts for use in wind energy forecasting. The demonstrated forecast improvement and more efficient wind energy integration that results could also be used to build support within the wind industry for sharing proprietary data with NOAA to improve foundational boundary layer wind forecasts.

This study is part of a larger project and constitutes the northern study area in a broader study titled the 'The Wind Forecast Improvement Project (WFIP): A Public/Private Partnership for Improving Short Term Wind Energy Forecasts and Quantifying the Benefits of Utility Operations' (Wilczak et al., 2014), with a companion study conducted in an area of the southern Great Plains (Freedman et al., 2014).

2. Project Description and Goals

The overall goal of this project is to identify and quantify the benefits of improvements in NOAA foundational weather forecasts to wind energy forecasting. By quantifying the benefits, including those that result from more efficient integration of wind energy into electric power systems as a result of the improved weather forecasts from NOAA, we hope to provide guidance and justification for appropriate investments in additional weather observations, data sharing and forecast models.

The three main objectives of this project are to:

1. Provide additional meteorological data at or near wind turbine hub height to NOAA over a large geographical area in the upper Midwest/Northern Plains for assimilation into NOAA forecast models with the aim of improving short-term wind (energy) forecasts.

2. Evaluate the impacts of enhanced NOAA weather forecast model improvements on short-term wind speed and wind power forecasts, particularly for the Earth System Research Laboratory version of the Rapid Refresh Model (ESRL_RAP) and the High Resolution Rapid Refresh (HRRR) model. This evaluation is both in an overall sense and during potentially high impact events on electric grid operation such as wind energy ramp periods.

3. Quantify the economic benefits of forecast improvements on power system operations in the Midcontinent Independent System Operator (MISO) operating area and power system operations in general.

While an economic analysis was conducted as part of the Wind Forecast Improvement Project (WFIP), 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 these 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.

A project team was assembled consisting of a major wind plant developer/owner/operator (NextEra Energy Resources), wind energy forecaster (WindLogics), university (South Dakota State University - SDSU), national laboratory (National Renewable Energy Laboratory) and an independent system operator (MISO) with extensive combined experience in data collection, wind plant operation, wind energy forecasting, and utility grid operation and planning. The team worked in partnership with NOAA and DOE through the duration of the project to accomplish the project objectives.

An overview of the study area is shown in Figure 1, which includes the locations of the NextEra wind plants, four tall meteorological tower locations provided SDSU, two WindLogics sodar units, and additional NOAA and DOE observational platforms that were deployed for use in the project. (For more information about the additional NOAA and DOE observational instruments, please refer to the NOAA

report (Wilczak et al., 2014)). NextEra Energy Resources/WindLogics also provided real-time data from 42 tall towers scattered throughout the study area (not shown in Figure 1), and 411 nacelles from operating wind plants. These additional data sources were supplied to NOAA in real time for assimilation into the enhanced research forecast models. Data from approximately 75 additional NextEra tall towers were provided (not in real time) for forecast verification. The box outlined in Figure 1 is roughly an 800 km x 800 km area that contains approximately 2347 MW of NextEra wind generation (as of the beginning of the project) spread across 25 wind plant locations that are scattered throughout the area. The region of interest is generally characterized by gently rolling terrain with several moderate terrain features (such as the Buffalo Ridge in southwest Minnesota and northeast South Dakota) that impact the regional wind.

It should be noted that NOAA has a fairly dense surface observation network in place in Minnesota, Iowa and eastern Nebraska, but the surface observation network becomes quite sparse in parts of North Dakota and South Dakota where many of NextEra's wind assets are located. In regions such as the Great Plains, three-dimensional volumetric data of atmospheric variables are likely needed in order to significantly improve the forecasts from numerical weather prediction models. However, there is also a lack of wind profiler measurements in the northern portion of the study region with the current operational NOAA wind profiler network. Several non-federal surface mesonets provide additional surface data coverage, but North Dakota, South Dakota and western Nebraska are still relatively "data poor" regions for both surface and upper air measurements. Since weather systems that affect the wind resource across the Midwest generally move in a west to east direction, the lower density in both the surface and upper air observing networks in North Dakota and South Dakota (extending into Montana and Wyoming) likely impacts the quality of the near-surface wind forecasts across the Midwest. The additional upper air and near-surface observations shown in Figure 1 were collected and provided to NOAA in real-time, and the impact of the additional data on forecast accuracy was explored through several data denial experiments conducted by NOAA. The impacts of the additional data on the accuracy of wind energy forecasts in the study area will be discussed in Section 4.6.

In addition to the current short-term operational weather forecast models [Rapid Update Cycle (OP_RUC), Rapid Refresh (OP_RAP) and the North American Model (NAM)] that are run at the National Center for Environmental Prediction (NCEP), NOAA provided access to the research forecast models for use in the project [ESRL_RAP and HRRR]. The research models are the potential future generations of forecast models being developed and tested at the Earth System Research Laboratory (ESRL), and include improvements to the representation of atmospheric and land surface processes. The additional data being provided in this study was also assimilated into the research model forecast initial conditions in addition to the data sources that are assimilated into the operational weather forecast models. All available short-term weather forecast models were used by WindLogics to create wind power forecasts for 23 NextEra wind plants (Table 1) in the study area. Power forecasts were generated both with wind forecasts directly from the weather forecast models converted to power through plant-derived power curves, and with optimization methods that use the real-time plant power data along with weather model forecast data to train a machine learning system to make more accurate site-specific wind power forecasts for each wind plant. The power forecasting procedure is discussed in more detail in Section 3. An analysis of the forecast error improvements both in the bulk error statistics and during wind energy ramp events is presented in Section 4. A summary of the major findings of the study and suggestions for future work are given in Sections 6 and 7.

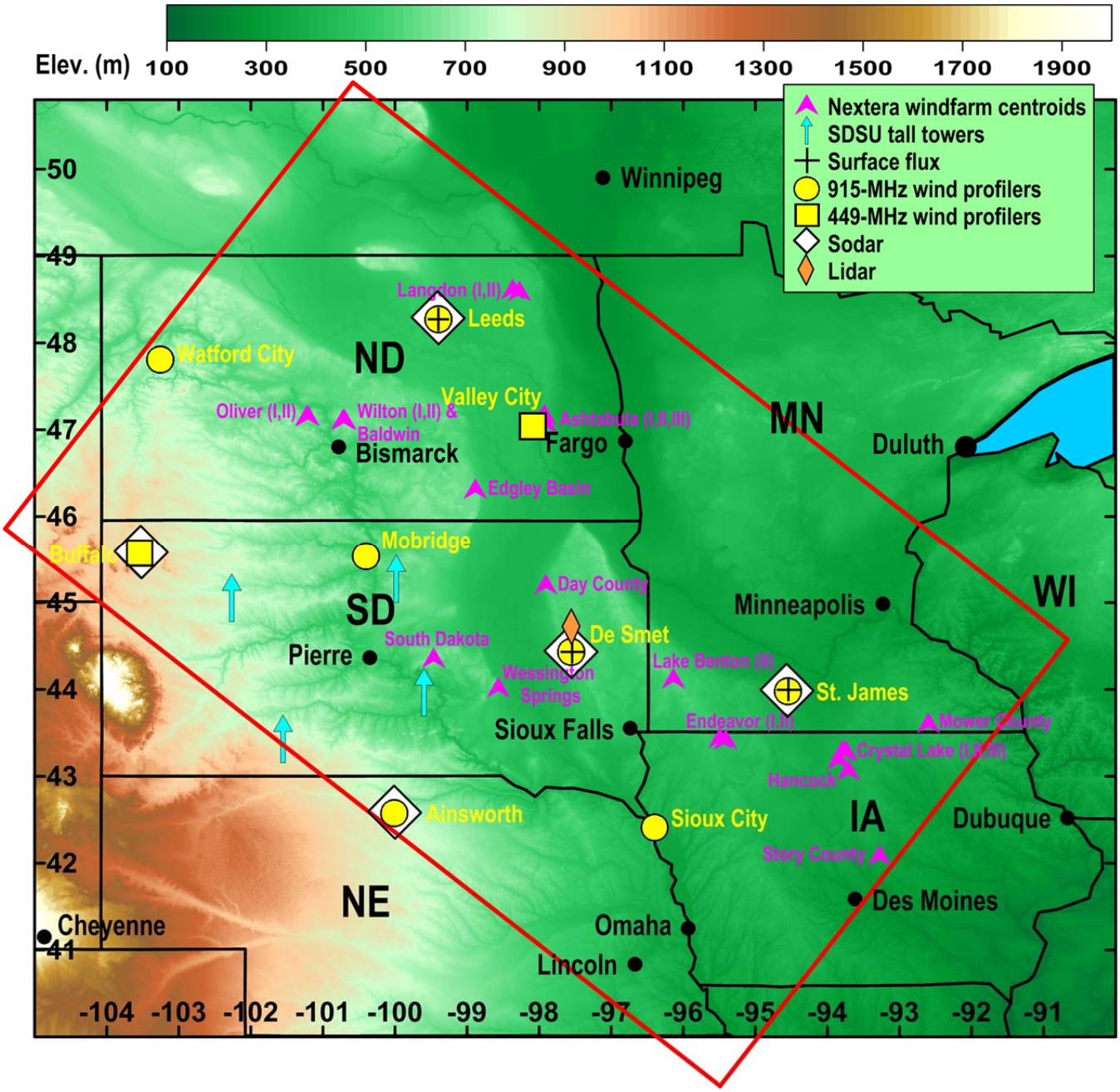

Figure 1: Overview of the locations of additional data sources and operating NextEra wind plants used in this study. The red box denotes a rough outline of the study area and colored shading indicates elevation. NextEra Energy Resources also provided real-time data from 42 tall meteorological towers (not shown). The pink icons denote the locations of the NextEra Energy operating wind plants which were used for forecasting in this study. Image courtesy of NOAA.

Plant Name	Location	Hub Height	Rated Capacity
Langdon 1, 2	Cavalier County, ND	80m	159 MW
Edgeley Basin	LaMoure County, ND	80m	61.5 MW
Ashtabula 1, 2, 3	Barnes, Griggs and Steele Counties, ND	80m	268.5 MW
Wilton 1, 2	Burleigh County, ND	80m	99 MW
Oliver 1, 2	Oliver County, ND	80m	99.2 MW
South Dakota Basin	Hyde County, SD	65m	40.5 MW
Wessington Springs	Jerauld County, SD	80m	51 MW
Lake Benton II	Pipestone County, MN	50m	102.8 MW
Mower	Mower County, MN	80m	98.9 MW
Endeavor 1, 2	Osceola County, IA	80m	150 MW
Crystal Lake 1, 2, 3	Hancock and Winnebago Counties, IA	80m	416 MW
Hancock	Hancock County, IA	65m	97.7 MW
Day County	Day County, SD	80m	99 MW
Story County	Story and Hardin Counties, IA	80m	150 MW
Baldwin	Burleigh County, ND	80m	102.4 MW
AGGREGATE TOTAL			2155.48 MW

Table 1: Locations and rated capacities of the NextEra Energy wind plants used for wind power forecasting in this study.

3. Power Forecasting Procedure

There are several approaches one can take to make a wind (power) forecast, depending on the timescale and application. It should be noted that making a wind power forecast involves two distinct processes, each with their own potential sources of error: 1) making an accurate hub-height wind speed forecast, and 2) converting that wind speed forecast into an accurate power forecast. The sources of error in the hub-height wind speed forecasts are the same as the sources of forecast error for weather forecasting in general – imperfect knowledge of the state of the atmosphere at the start of the forecast period, and errors in the way physical atmospheric processes are represented in the numerical forecast models. There are also many potential sources of error in converting the wind speed to wind power at operational wind plants which can include: lack of accurate knowledge of turbine availability at any given time, lack of information about plant curtailments, turbines performing on a degraded power curve due to icing or maintenance issues, imperfect knowledge of turbine waking and electrical losses between the turbines and the plant metering point. Since one of the main goals of this project is to quantify forecast improvements in the NOAA weather forecast models for wind energy forecasting purposes, we strove to eliminate as many of the potential sources of error in converting wind speed to power as was possible by assuming 'perfect' knowledge of turbine availability (within the limits of the turbine data accuracy), removing periods of curtailment and potential icing periods from the forecasting

dataset, and using plant power at the meter point to make forecasts (i.e. electrical losses accounted for). The lack of accurate turbine nacelle directional information (compared to 'true north') did not allow directional plant wake losses to be taken into account. It should also be noted that in real-world forecasting operations, one only has knowledge of turbine availability (if this information is available) at the start of the forecast. However, the assumption of 'perfect' availability knowledge at all times had little impact on the overall accuracy of the forecasts over the entire one year of the study period since curtailment and icing periods were filtered out of the forecasting dataset.

Several different methodologies were used in this study to create wind power forecasts from weather forecast model data. These methods ranged from basic power forecasts created with as little interference as possible in the forecasted wind speeds from the weather models (i.e., using the "raw" data from the forecasts) to methods using sophisticated statistical algorithms to make corrections to more basic power forecasts using available wind plant production data. The various methodologies are described in more detail below.

3.1 Raw Forecasts

Since one the of goals of the project was to evaluate improvements in the NOAA weather models for wind energy forecasting purposes, we created 'raw' power forecasts in which there was as little manipulation of the raw weather forecast model data as possible in the process of creating a wind power forecast. This allows us to evaluate power forecast improvements that translate to weather model wind forecast improvements as directly as possible. These power forecasts were created at each individual wind plant by calculating a site average wind speed (using all model points covering the site) and then using a plant power curve derived from wind turbine nacelle data to convert wind speed to power.

3.2 Bias Correct Raw Forecasts

The bias corrected raw forecasts are similar to the raw power forecasts discussed above, but since we utilized the average of the nacelle wind speeds to create a plant power curve (some of which are waked) and we are using the model wind speeds in the forecasts (which are free-stream), this likely introduces some bias in the forecasts. To correct for this, we calculated a rolling two week hub-height wind speed bias for every forecast hour based on the past two weeks of previous weather model forecasts, and then used the calculated bias to correct the wind speed before making the power forecast. Note that in doing so, we are also removing some of the forecast error due to bias at the later forecast hours. This is a simple statistical correction to create more accurate wind power forecasts at each site.

3.3 Trained Forecasts

Most commercial wind energy forecast vendors use some sort of statistical or artificial intelligence method to further improve the accuracy of wind energy forecasts delivered to their customers. WindLogics has developed a combined approach of using numerical forecast models coupled with a machine learning system to make more accurate wind forecasts for a specific site. Machine learning systems are sophisticated methods for creating classification or regression functions from a set of training data. Such systems are able to 'learn' complex nonlinear relationships between hypothesis variables and target variables based on the training data provided. The machine learning system is trained for each individual wind farm, providing more accurate site-specific wind (power) forecasts than

would be provided by using a numerical forecast model alone. One drawback to using a machine learning system is that a minimum of several months of data (both model forecasts and plant production data) are needed to start the training process.

The system was trained at the end of December 2011 (initially using data from September – December 2011) and trained forecasts were generated starting in January 2012. The system was retrained at the end of each month (using data from September 2011 – end of the current month) to generate trained forecasts for the following month. As with any statistics-based correction process (such as machine learning), one must choose what statistical quantity or quantities to optimize for. WindLogics employs a dual-optimization approach to its machine learning system which attempts to choose parameters that minimize overall forecast Root Mean Square Errors (RMSE) while simultaneously trying to minimize errors in 'bin RMSE' (i.e. the error in a binned power histogram). Even with this dual optimization approach, the process tends to minimize the bulk RMSE of the forecasts more so than the bin RMSE, which results in significant improvements in the forecast bulk RMSE, but at the expense of some 'sharpness' in the forecast features. Thus the trained forecasts are not ideally suited for ramp forecasting purposes.

3.4 Trained Ensemble Forecasts

These are the standard short-term wind power forecasts issued to WindLogics customers if all data needed to make the forecast is available. The trained ensemble forecasts are created in a second stage training process in which a trained short-range model-based power forecast is combined with the trained power forecast based on the operational NCEP North American Model (NAM) and wind plant observations at the start of the forecast in order to further improve the accuracy of individual wind plant power forecasts. These forecasts are also trained in such way as to minimize the Root Mean Square Error in the forecasts, and are included here to evaluate the impact of short-term weather forecast model improvements on state-of-the-art wind power forecasts.

3.5 Experiment Design

Each of the types of power forecasts described above were generated for each NextEra wind plant in the study area using a suite of short-range numerical weather prediction models – two operational forecast models and two research (or experimental) forecast models. These forecast models are run every hour incorporating the most recent weather observations at the start of the forecast, and are used to generate weather forecasts for 0-15 or 0-18 hours into the future.

The operational forecast models are those that are run in an operational setting at the National Center for Environmental Prediction (NCEP), and output from these forecast models is publically distributed for all weather forecasting purposes. Since the short-term operational forecast model at the start of the forecasting experiment -- the Rapid Update Cycle or RUC (Benjamin et al. 2004) -- was scheduled to be replaced by the next generation short-term forecast model, the Rapid Refresh model or RAP, WindLogics was granted access to the RAP output at the start of the forecasting experiment while the model was being run quasi-operationally at NCEP prior to operational implementation. The RAP replaced the RUC as the operational short-term forecast model on May 1, 2012.

The research models include a research version of the Rapid Update Cycle, and the High Resolution Rapid Refresh (HRRR). The HRRR has higher spatial grid resolution than the RAP (3 km vs. 13 km) and is being developed for better convective storms forecasting and other applications that require higher spatial grid resolution.

A table of the various short-range model-based power forecasts made in this experiment and their shorthand designations is given in Table 2. In addition to the short-range forecast models, the North American Model (NAM -- the mid-range operational model run at NCEP) was used to generate forecasts for use in the trained ensemble forecasts. The 'operational' forecasting phase of the project ran from September 1, 2011 - September 1, 2012, and the time periods throughout the project during which the various power forecasts were made are shown in Figure 2.

Base Model Forecast/Forecast Type	**Raw**	**Bias Corrected Raw**	**Trained**	**Trained Ensemble**
Operational Rapid Update Cycle (OP_RUC)	OP_RUC_RW	OP_RUC_BCRW	OP_RUC_TR	OP_RUC_ENS
Operational Rapid Refresh (OP_RAP))	OP_RAP_RW	OP_RAP_BCRW	OP_RAP_TR	OP_RAP_ENS
Research Rapid Refresh (ESRL_RAP)	ESRL_RAP_RW	ESRL_RAP_BCRW	ESRL_RAP_TR	ESRL_RAP_ENS
High Resolution Rapid Refresh (HRRR)	HRRR_RW	HRRR_BCRW	HRRR_TR	HRRR_ENS

Table 2: Reference list for the short-hand designations of the various types of model-based wind power forecasts used in this study as described above.

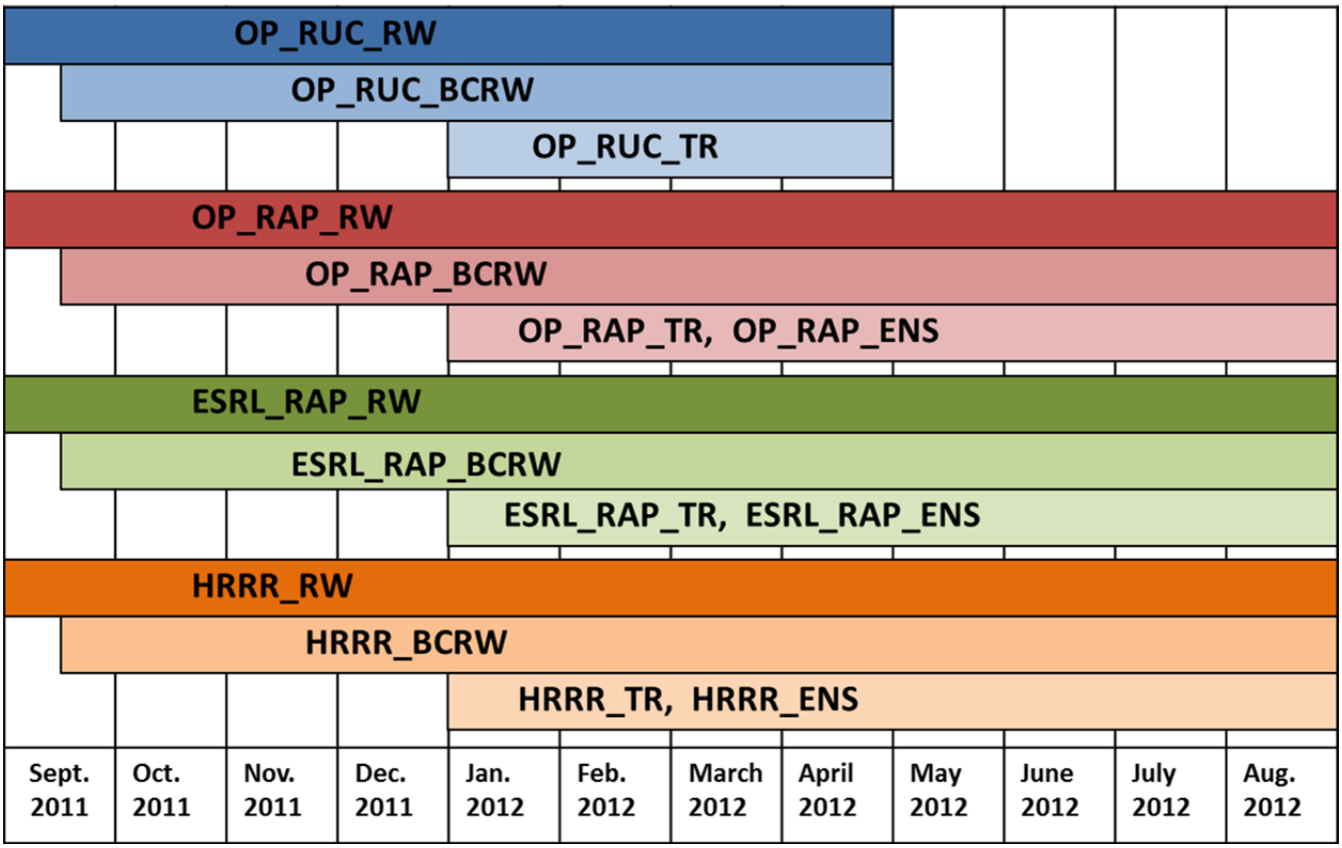

Figure 2: Time periods over which the various power forecasts were made during the course of the project. Since the training process requires several months of training data, trained forecasts were only generated from January 2012 – August 2012.

Since we are creating 'operational' forecasts from experimental research models (which are not run in operational forecasting environments), several assumptions and allowances needed to be made to make the forecast comparisons more representative of what one would likely see in an operational environment:

- Because of the computational requirements for generating HRRR forecasts, the 0Z and 12Z model forecast cycles could not be delayed long enough to include radiosonde measurements (an extremely important upper air dataset collected twice daily at 0Z and 12Z). Given the critical importance of this data set on forecast accuracy, the 0Z and 12Z runs of all model forecasts were excluded from the forecast analysis.

- We are also making the assumption that in an operational setting, HRRR data would be available in the same timeframe as RUC or RAP. (The experimental version of the HRRR is currently available an hour later than the operational RUC or RAP.)

- Unless explicitly stated otherwise, the 'forecast hour' is the hour relative to when the power forecast was issued. Because of the time requirements for running the weather forecast models and for distributing the model output, the power forecasts for a specific hour are based on the weather model runtime from the previous hour.

It should also be noted that the training process corrects for systematic forecast errors, and when significant changes to the weather forecast models occur (such as extensive changes to the

representation of physical processes), the system is normally retrained using only weather model data output generated after the change in an operational environment. During this experiment, changes were made to the research forecast models over the first 6-7 months of the forecasting phase of the project (as routinely occurs during work to improve the research model forecasts). Given that the time-line for the forecasting phase of the project was 1 year, it was not possible to withhold the first 6-7 months of model forecast output from the training process and then generate a long enough time period of trained forecasts to make statistically significant comparisons with the trained and untrained forecasts. Thus the trained forecasts generated from the research weather forecast models (ESRL_RAP and HRRR) during the course of the project are not optimal. This will be discussed further in Section 4.

In addition to individual wind plant forecasts, we also created 'system aggregate' forecasts in which all forecasts for the individual wind plants were aggregated together. These aggregate forecasts are the forecasts that would often be used by system operators. It is important to note that this 'system aggregate' is NOT the MISO system aggregate wind power forecast, as the system aggregate forecast in our case only includes the NextEra operating wind plants in the study area as outlined in Table 1.

New short-range power forecasts (0-15 hours) were made every hour as new weather model forecasts became available. Various forecast error statistics were calculated (both in a bulk sense and for wind energy ramp events) and the error metrics were compared to gauge the impact of the additional weather data and forecast model improvements on wind energy forecast accuracy. The results of this analysis are discussed below.

4. Forecasting Results

4.1 Bulk Error Statistics for Individual Site Forecasts

The two most commonly used bulk error statistics for measuring forecast accuracy are the Mean Absolute Error (MAE) and the Root Mean Square Absolute Error (RMSE). The Mean Absolute Error is defined as:

$$\boldsymbol{Mean\ Absolute\ Error\ (MAE) = average}\left(\sum_{\#\ fx\ values} \boldsymbol{abs(forecast\ power - actual\ power)}\right)$$

Equation 1

Here 'fx' is shorthand for forecast. The mean absolute error is a standard measure of how close the forecasts are to the actuals power values we are trying to forecast.

The Root Mean Square Error is defined as:

$$\boldsymbol{Root\ Mean\ Square\ Error\ (RMSE)} = \sqrt[2]{\boldsymbol{average}\left[\sum_{\#\ fx\ values} \boldsymbol{(forecast\ power - actual\ power)^2}\right]}$$

Equation 2

Root mean square error is often a more telling error metric of forecast performance than mean absolute error since it penalizes large forecast errors more than small errors, and is a better measure of how closely the forecast 'follows' the actual power production.

Since one of the goals of this project is to measure forecast improvements from additional data assimilation and model improvements, we need to compare error metrics between different forecasts to quantify these improvements. If the differences in the errors are normalized by average actual energy, they can be viewed as the percent improvement in one model-based power forecast compared to another:

$$\boldsymbol{Percent\ Improvement\ (PI) = \frac{RMSE(forecast\ 1) - RMSE(forecast\ 2)}{average\ actual\ energy\ produced} \times 100}$$

Equation 3

In wind energy industry, the common practice is to normalize forecast errors by the rated capacity of the wind plant so as to avoid overemphasizing relative small errors in power output during periods of low wind energy production:

$$\boldsymbol{Percent\ Capacity\ Improvement\ (PCI) = \frac{RMSE(forecast\ 1) - RMSE(forecast\ 2)}{rated\ capcity} \times 100}$$

Equation 4

While these types of capacity normalized metrics are helpful from the perspective that the metrics are anchored to a fixed value (the rated plant or aggregate capacity), they can be somewhat misleading in that a given forecast error will be given the same 'weight' regardless of where the plant is actually operating relative to its capacity during the time period of evaluation. It should be noted that we could define similar metrics to those shown in Equation 3 and Equation 4 using Mean Absolute Error (MAE) in place of RMSE.

4.1.1 Annual

We compiled the error statistics above for several different time periods over which the various forecasts overlap as shown in Figure 2. In the interests of brevity, only the results from the bias-corrected raw power forecasts will be shown at the individual wind plants. Recall that these forecasts have already been 'improved' in the sense that the model generated wind speeds have been statistically altered to remove consistent model bias before generating a power forecast, but since all commercial wind power forecast vendors make some kind of statistical correction to the weather model output in the process of generating wind power forecasts, these forecasts likely represent the best compromise for still being very closely aligned to the weather forecasts models, but with a simple statistical correction applied. It should also be noted that the results for the Day County site (NE South Dakota) only contain data from Sept. 2011 – Aug. 10, 2012 due to catastrophic failure of the data archiving mechanism at the site which occurred on August 10, and the data from August 10-31 was not recoverable.

Figure 3 shows the average Percent Improvement (as defined in Equation 3) during the first six hours of the forecasts between the OP_RUC_BCRW (recall the RUC was the operational short-range weather forecast model for the first 8 months of the forecasting experiment) and the HRRR_BCRW power forecasts at each of the NextEra wind plants in the study area. In areas where the wind plants are very close to each other, they have been aggregated together. Warm colors indicate locations where the HRRR-based bias-corrected power forecasts are performing better, and cool colors indicate

locations where the OP_RUC-based bias-corrected power forecasts are performing better. As can be seen in Figure 3, there is some geographic variability in the performance of the two forecasts, but the HRRR_BCRW forecasts perform better than the OP_RUC_BCRW forecasts at 11 of the 13 NextEra wind plant locations. A similar analysis is shown in Figure 4 for the Percent Improvement between the OP_RUC_BCRW power forecasts and the ERSL_RAP_BCRW forecasts (the ESRL_RAP has the additional data assimilated and model physics improvements). As can be seen in Figure 4, the ESRL-RAP-based bias correct power forecasts perform better at all NextEra wind plant locations in the study area compared to the operational model-based bias corrected power forecasts. The Percent Improvement ranged from 1.7% to 5.3% (with several locations close to 5%) and an average over all sites of approximately 3%.

The RUC model was replaced with the RAP model as the short-range weather model forecast run operationally at NCEP on May 1, 2012. As was discussed above, WindLogics was granted early access to the RAP forecasts for use in this project as they were being run quasi-operationally prior to operational implementation. Thus we have a full year of the operational RAP-based power forecasts with which to compare with the research model-based power forecasts. The Percent Improvement over the year-long forecast experiment between the OP_RAP_BCRW power forecasts and the HRRR_BCRW power forecasts averaged over the first 6 hours of the forecasts at the NextEra wind plant locations is shown in Figure 5. Over the year-long period of the forecasting project, the use of the HRRR weather forecast model improved power forecasts over the current operational model at 9 of the 13 wind plant (or wind plant aggregate) locations. The HRRR_BCRW power forecasts performed worse (in the bulk statistical sense) than the OP_RAP_BCRW power forecasts at only 3 of the 13 locations with one of the locations showing a statistical ‘tie’ between the two forecasts. As was the case with the operational RUC-based bias corrected power forecasts, the ERSL_RAP_BCRW power forecasts also perform better that the OP_RAP_BCRW power forecasts at all NextEra wind plant locations over the first six hours of the forecasts as shown in Figure 6. The improvements were not quite as large as the improvements seen over the OP_RUC_BC power forecasts, with Percent Improvements ranging from 0.5% to 5% and an overall average at the local site level of approximately 2.3%. These results suggest that the additional data assimilation and model physics advances are improving the overall power forecast accuracy across most of the Northern Study Area domain, although the exact impacts of the additional data assimilation on the forecast accuracy (as opposed to improvements in the model physics) will be better quantified in the data denial experiments discussed in Section 4.6.

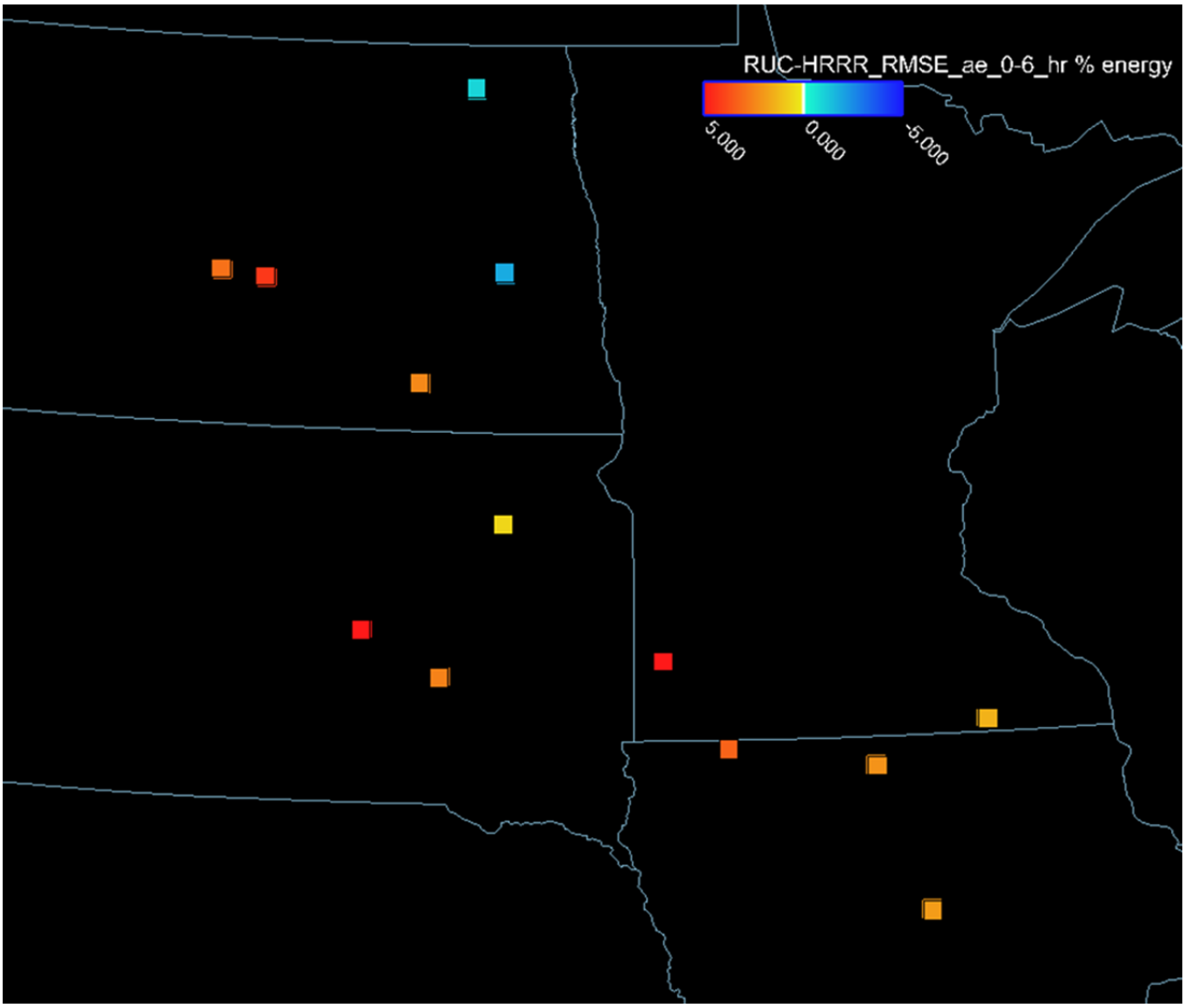

Figure 3: Percent Improvement between the OP_RUC_BCRW and the HRRR_BCRW power forecasts over the first 6 hours of the forecasts. Warm (cool) colors indicate where the HRRR-based (OP_RUC-based) bias-corrected raw forecasts are performing better on average. Results are valid over the period from Sept. 15, 2011 – April 30, 2012.

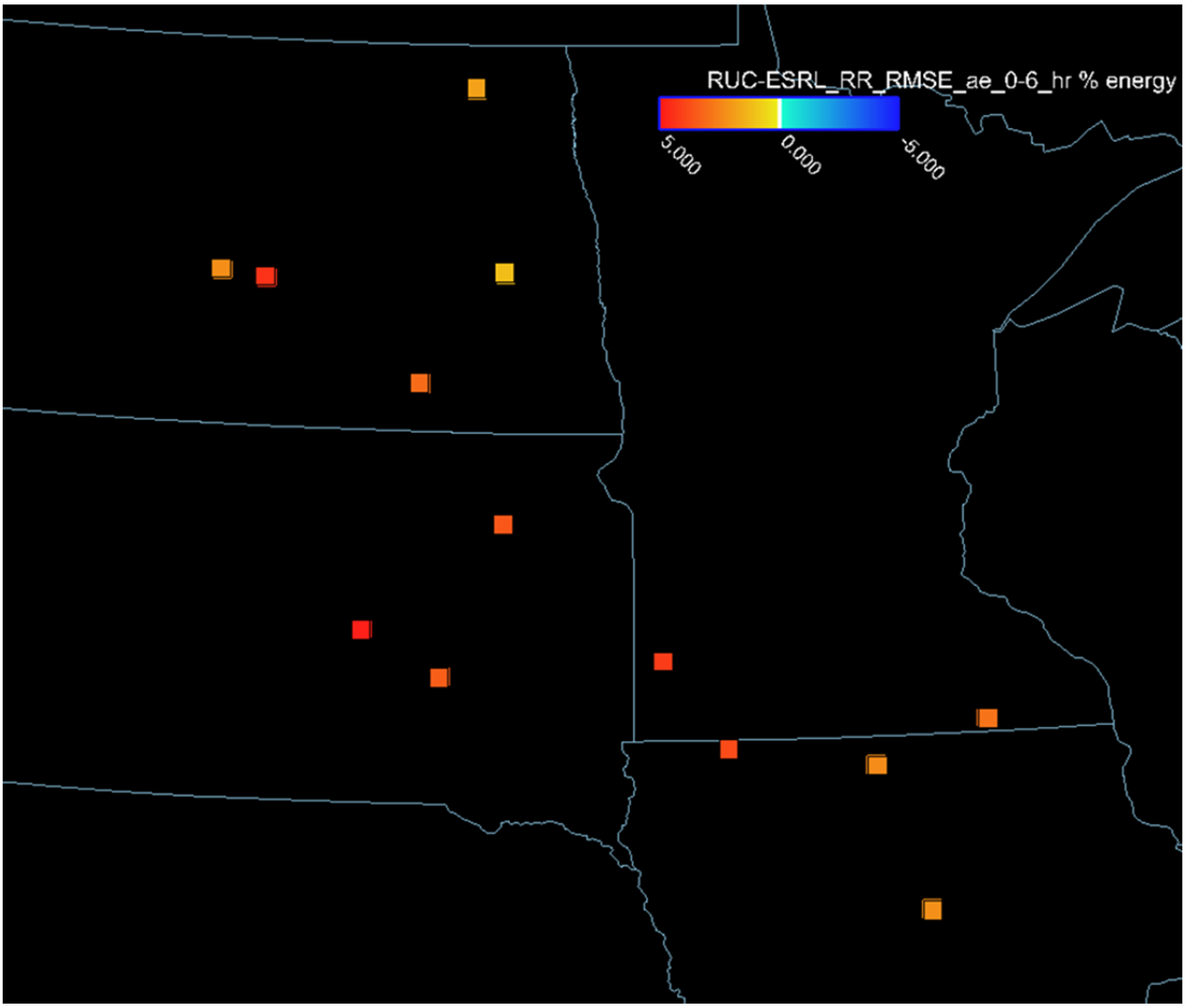

Figure 4: Percent Improvement between the OP_RUC_BCRW and the ESRL_RAP_BCRW power forecasts over the first 6 hours of the forecasts. Warm (cool) colors indicate where the ESRL_RAP-based (OP_RUC-based) bias-corrected raw forecasts are performing better on average. Results are valid over the period from Sept. 15, 2011 – April 30, 2012.

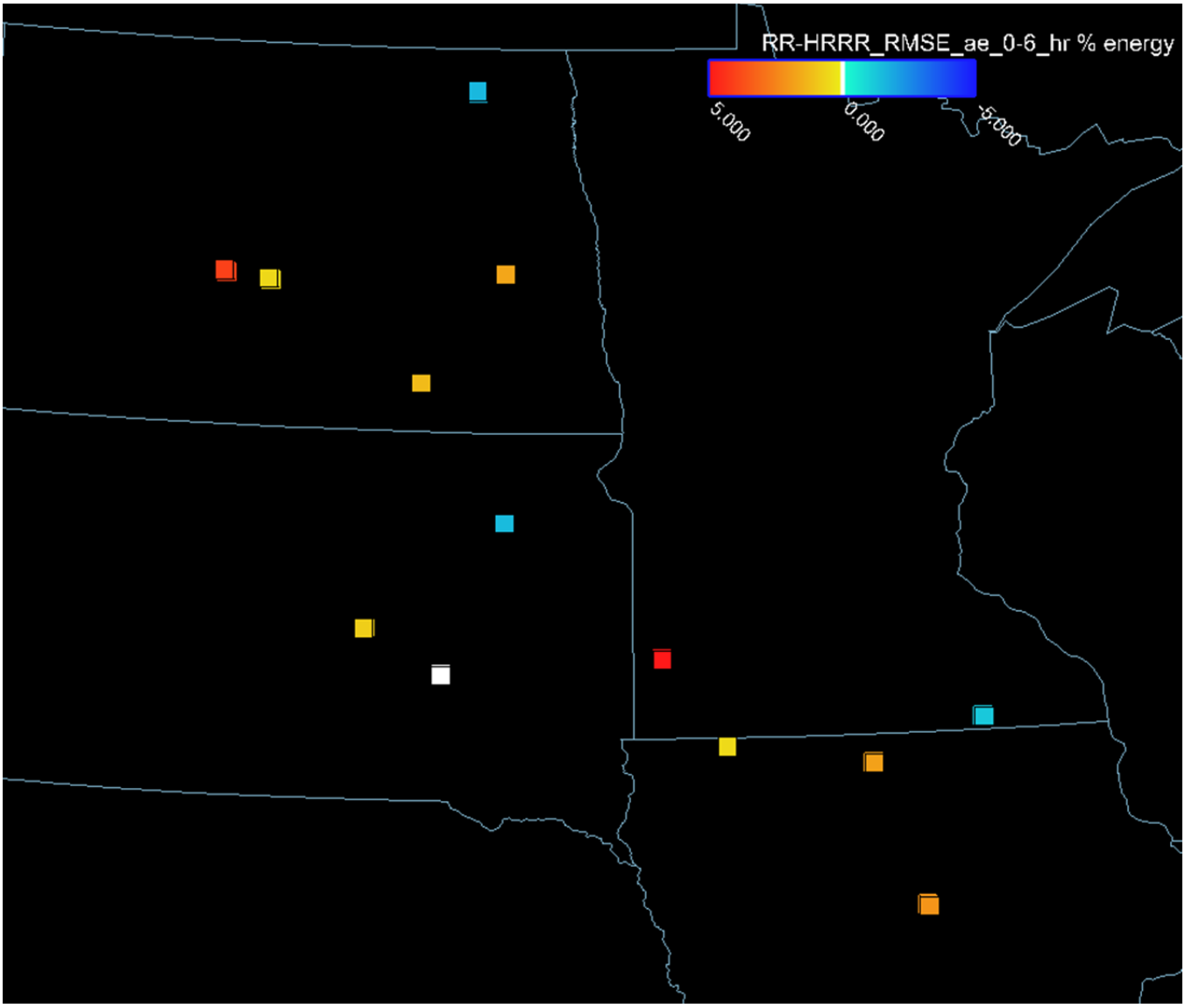

Figure 5: Percent Improvement between the OP_RAP_BCRW and the HRRR_BCRW power forecasts over the first 6 hours of the forecasts. Warm (cool) colors indicate where the HRRR-based (OP_RAP-based) bias-corrected raw forecasts are performing better on average. It should be noted that the results for the Day County site (NE South Dakota) only contain data from Sept. 2011 – Aug. 10, 2012 due to catastrophic failure of the data archiving mechanism leading to loss of data after August 10.

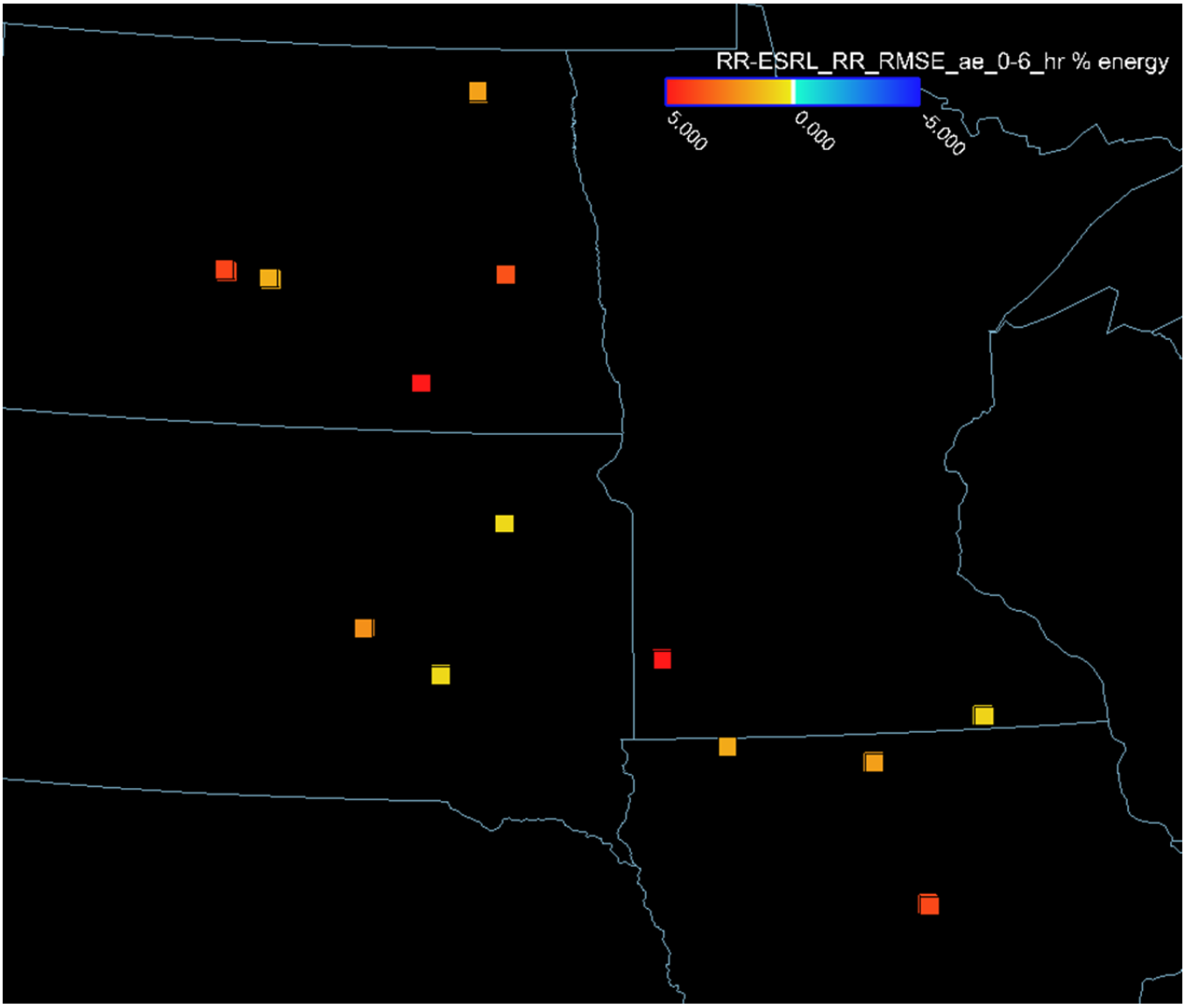

Figure 6: Percent Improvement between the OP_RAP_BCRW and the ESRL_RAP_BCRW power forecasts over the first 6 hours of the forecasts. Warm (cool) colors indicate where the ESRL_RAP-based (OP_RAP-based) bias-corrected raw forecasts are performing better on average. It should be noted that the results for the Day County site (NE South Dakota) only contain data from Sept. 2011 – Aug. 10, 2012 due to catastrophic failure of the data archiving mechanism leading to loss of data after August 10.

The bulk error statistics shown above for the bias-corrected raw power forecasts indicate that the coarser resolution ERSL_RAP-based bias-corrected power forecasts perform better than the HRRR-based bias-corrected power forecasts even though the HRRR model has better horizontal grid resolution than the ERSL_RAP (3 km grid spacing vs. 13 km grid spacing). The only exceptions are the Lake Benton 2 site which sits on top of the Buffalo Ridge, the most prominent topographic feature in the study area, and the Oliver aggregate site in the butte country of central North Dakota. One would expect the HRRR-based forecasts to have an advantage over the other forecast models in areas of complex terrain because the HRRR model better resolves the local terrain features.

There could be several reasons for the HRRR-based power forecast bulk error statistics not performing as well as the coarser resolution ESRL_RAP-based power forecasts. For example, the high resolution HRRR is now capable of resolving mesoscale flow features which our current observational network

cannot always detect accurately. This could lead to inaccurate representation of these flow features, which may make the forecasts less accurate. It is also possible that the HRRR model is resolving small-scale features that are present in the observations, but the timing of the features impacting the wind plant is inaccurate which leads to less accurate bulk error statistics. However, these forecasts can still provide more value in alerting the forecast user to characteristics of potential wind energy ramp events as will be discussed in Section 4.4.

SITE	HRRR RMSE difference (bias-corrected raw - trained) % ae	OP_RAP RMSE difference (bias-corrected raw-trained) % ae	ESRL_RAP RMSE difference (bias-corrected raw-trained) % ae
Day County*	6.409504	4.023678	4.854585
Edgeley Basin	5.940571	8.159186	4.665076
Lake Benton 2	9.510061	15.60519	11.47338
Mower County	4.258922	3.452599	4.136737
South Dakota Basin	7.234905	7.563802	6.692912
Story County 1	4.375384	6.885237	4.527707
Wessington Springs	5.912535	5.56463	5.819415
AGGREGATES			
Ashtabula	6.456688	8.188639	6.852883
Crystal Lake /Hancock	9.110165	10.63111	12.25722
Endeavor	4.77346	4.941143	4.849246
Langdon	2.918298	2.795649	2.401154
Oliver	6.692291	8.664703	7.795691
Wilton/Baldwin	8.205423	7.559012	8.913332
*only through 8/10/2012			

Table 3: RMSE difference between the bias-corrected raw power forecasts and the trained forecasts (normalized by the actual energy produced) over the first six hours of the forecasts for the period from January 1, 2012 – August 31, 2012 for each of the model-based power forecasts. The forecasts based on the HRRR, ESRL_RAP, and OP_RAP are shown in columns 1, 2, and 3 respectively. Positive values indicate that the trained forecasts have lower RMSE values than the bias-corrected raw forecasts.

In addition to the bulk error statistics for the bias-corrected raw power forecasts generated from the suite of weather forecast models, we have compiled the bulk error statics for the trained power forecasts generated from each of the weather forecast models for comparison. Since several months of data are required for the training process, the trained forecasts were generated starting in January 2012 through the end of the forecasting project. The RMSE difference between the bias-corrected raw power forecasts and the trained forecasts normalized by the actual energy produced for each of the model- based forecasts over the first 6 hours of the forecast is shown in Table 3.

As can be seen in Table 3, the training process reduces the RMSE in the power forecasts at all sites for all model-based power forecasts. There is quite a bit of variability in the forecast accuracy improvements between the various wind plant locations and base weather model forecast being used. This could be due to the fact that some sites are inherently more difficult to forecast for than others, and the systematic forecast errors that the training process normally corrects will vary between the various weather forecast models.

At most sites, the forecast improvement provided by the training is significant. The average improvement is also larger for the OP_RAP-based bias-corrected power forecasts than for the ESRL_RAP-based power forecasts, so the training process tends to reduce the differences in forecast errors somewhat since the ESRL_RAP_BCRW power forecast have lower error rates than the OP_RAP_BCRW power forecasts as was shown in Figure 6. The training process also tends to improve the coarser-resolution model-based forecasts (OP_RAP, ESRL_RAP) more than the higher resolution model-based forecasts (HRRR). This could be due to the fact that the HRRR model resolves smaller-scale weather features, which creates forecast errors that may appear more 'random' in nature than the errors in the coarser resolution model-based forecasts. Improvements in the forecasts incurred through the training process are also not uniformly distributed across all forecast hours which has implications for power system operations and will be discussed further in Section 4.1.3.

The trained ensemble power forecasts further improve the errors statistics beyond the individual forecast model-based trained power forecasts at each wind plant location. Since the trained ensemble forecasts include the North American Model (NAM), which is not the focus of this study, we will only briefly present the results of the trained ensemble forecasts for the system aggregate power forecasts in Section 4.2 below.

4.1.2 Seasonal

Wind energy forecasts errors could be more significant to MISO operations during certain seasons of the year. For example, the MISO system tends to see more "minimum generation" periods during the spring and fall when loads are low. The system may also have less flexibility during these periods since planned outages of generation and transmission facilities would often be scheduled during these times. We therefore looked at the forecast error metrics on a seasonal basis to see if there were significant differences.

To better understand the performance of the forecasts during different times of year when different weather conditions are present, the relative performance of the various forecasts was broken down by season is shown in Figure 7 - Figure 9. Figure 7 shows the Percent Improvement (as defined in Equation 3) between the OP_RAP_BCRW and the HRRR_BCRW power forecasts for the first 6 hours of the forecasts for each of the seasons. The HRRR-based bias-corrected raw power forecasts perform better (in the bulk statistics) at most sites compared to the OP_RAP-based bias-corrected power forecasts during the summer and fall months (Figure 7a and Figure 7d), but performs better than the RAP-based forecasts at only half the sites in winter and spring. The region in which the HRRR_BCRW power forecasts generally perform better also shifts from the southern portion of the study area in winter to the northern portion of the study area in spring (Figure 7b and Figure 7c). It's also interesting to note that while the HRRR_BCRW power forecasts perform better than the OP_RAP_BCRW forecasts overall during summer, they don't perform as well at any of the South Dakota sites. Note also that the error differences as a percentage of actual energy produced are much larger in the summer than in the other months. Thus it appears that the HRRR is providing substantial improvement to the forecasts in the summer months compared to the current operational short-term forecast model.

A similar comparison between the ESRL_RAP_BCRW and the HRRR_BCRW power forecasts broken down by season is shown in Figure 8. It's evident that the ESRL_RAP-based bias-corrected power forecasts perform better than the HRRR-based forecasts (in the bulk statistics) at almost all sites in the study region in all seasons. There are a few sites at which the HRRR_BCRW power forecasts perform better, but the locations vary from season to season. The only site at which HRRR_BCRW forecasts clearly perform better is the Lake Benton 2 site which is on top of the Buffalo Ridge as was discussed above. The HRRR_BCRW forecasts also perform better at the Oliver aggregate site in central North

Dakota during the transition seasons. This site is also located in an area of more complex terrain (for the northern plains) along the edge of the Missouri River valley. However the Wilton/Baldwin site to its east is also located along the edge of the Missouri River valley, and the HRRR_BCRW power forecasts only perform better at this site (compared to the ESRL_RAP_BCRW) in the fall, so the link between more complex topography and the HRRR-based forecast performance is not straightforward and is likely flow dependent.

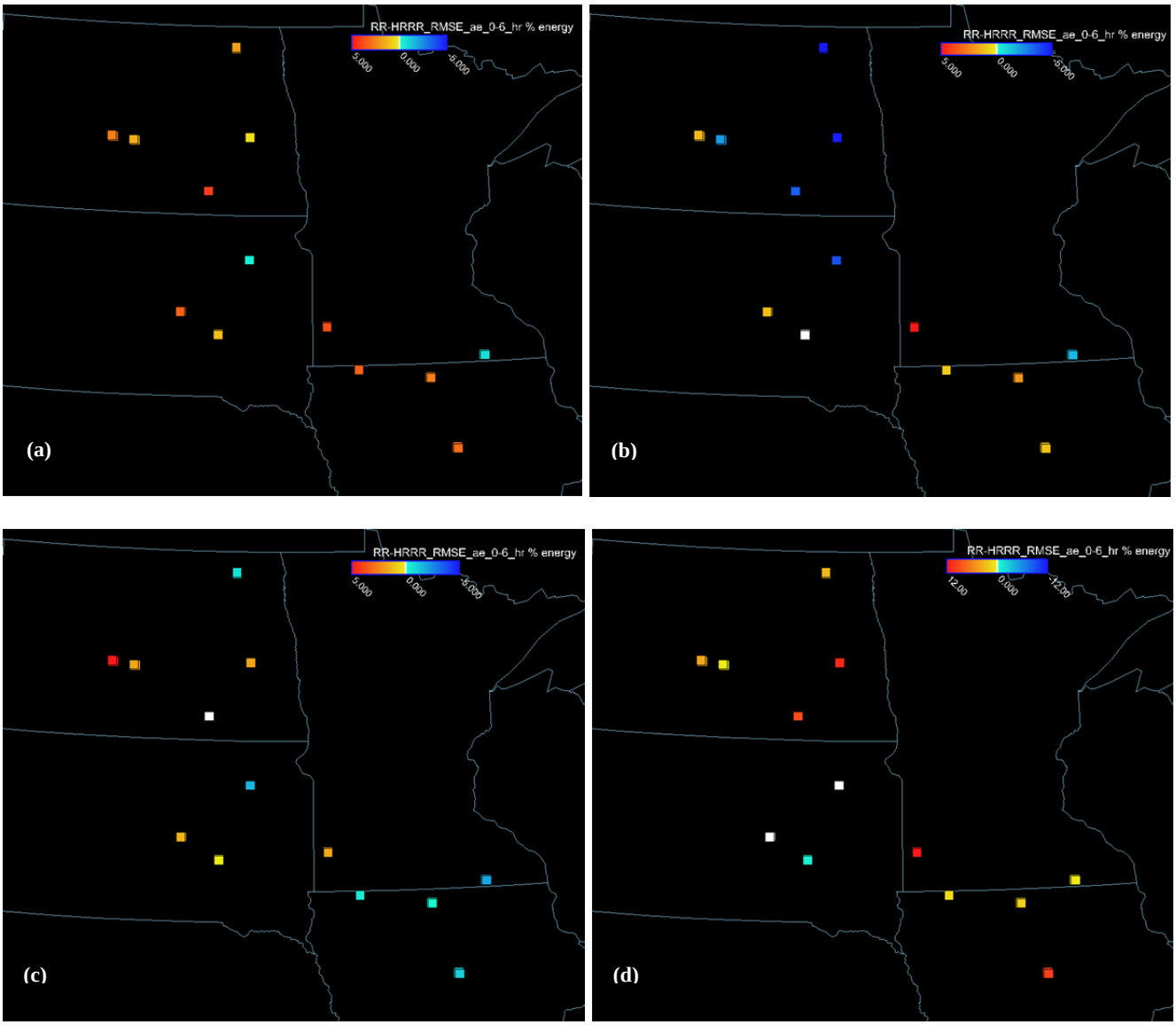

Figure 7: The Percent Improvement (PI) between the HRRR_BCRW power forecasts and the OP_RAP_BCRW power forecasts over the first six hours of the forecasts for (a) fall (S-O-N), (b) winter (D-J-F), (c) spring (M-A-M) and (d) summer (J-J-A). Warm (cool) colors indicate where the HRRR-based (OP_RAP-based) bias-corrected raw forecasts are performing better on average. It should be noted that the results for the Day County site (NE South Dakota) in (d) only contain data through Aug. 10, 2012. Note also that the scale on the forecast error difference is different in the summer than the other seasons (panel d).

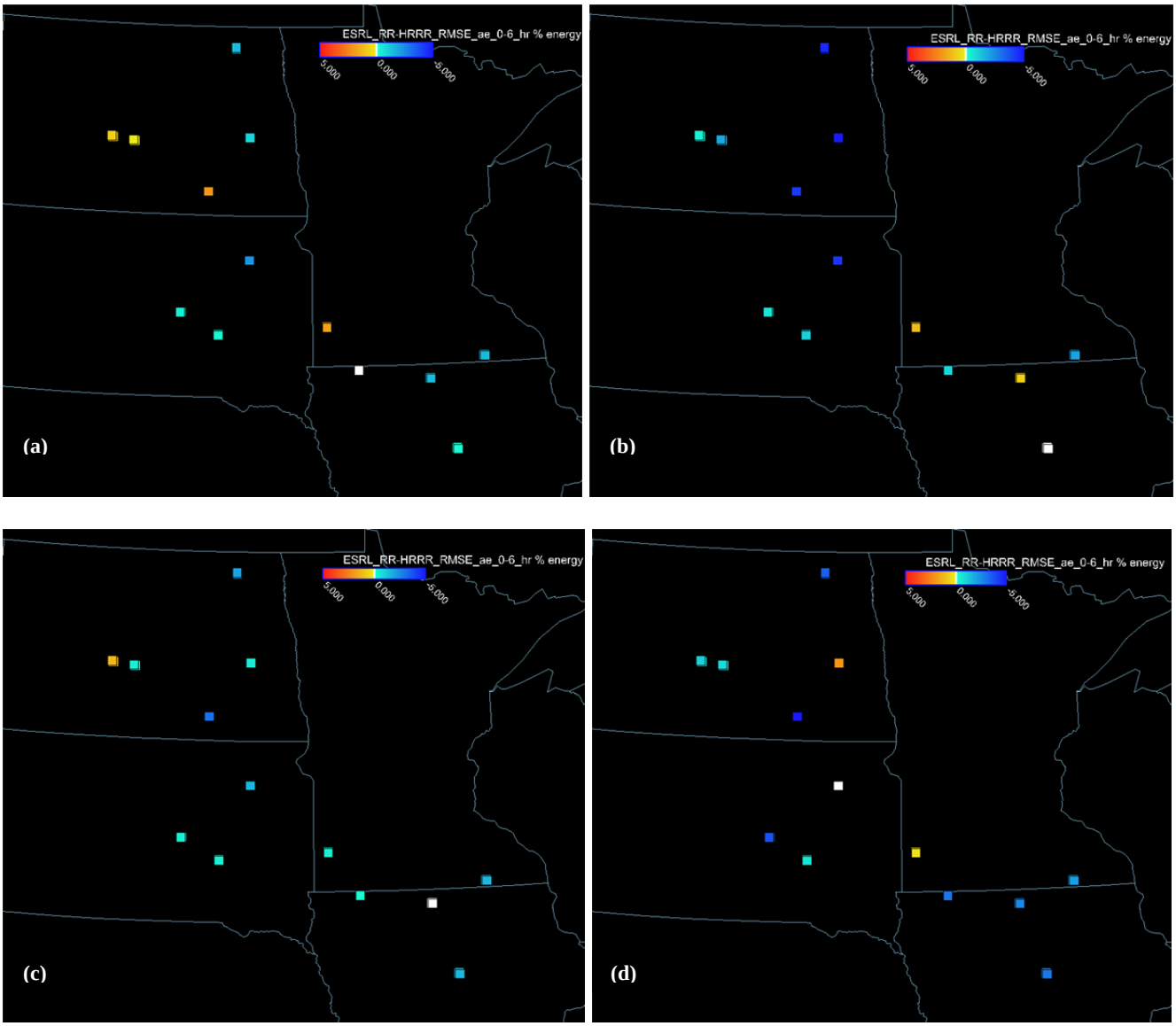

Figure 8: The Percent Improvement (PI) between the HRRR_BCRW power forecasts and the ESRL_RAP_BCRW power forecasts over the first six hours of the forecasts for (a) fall (S-O-N), (b) winter (D-J-F), (c) spring (M-A-M) and (d) summer (J-J-A). Warm (cool) colors indicate where the HRRR-based (ESRL_RAP-based) bias-corrected raw forecasts are performing better on average. It should be noted that the results for the Day County site (NE South Dakota) in (d) only contain data through Aug. 10, 2012.

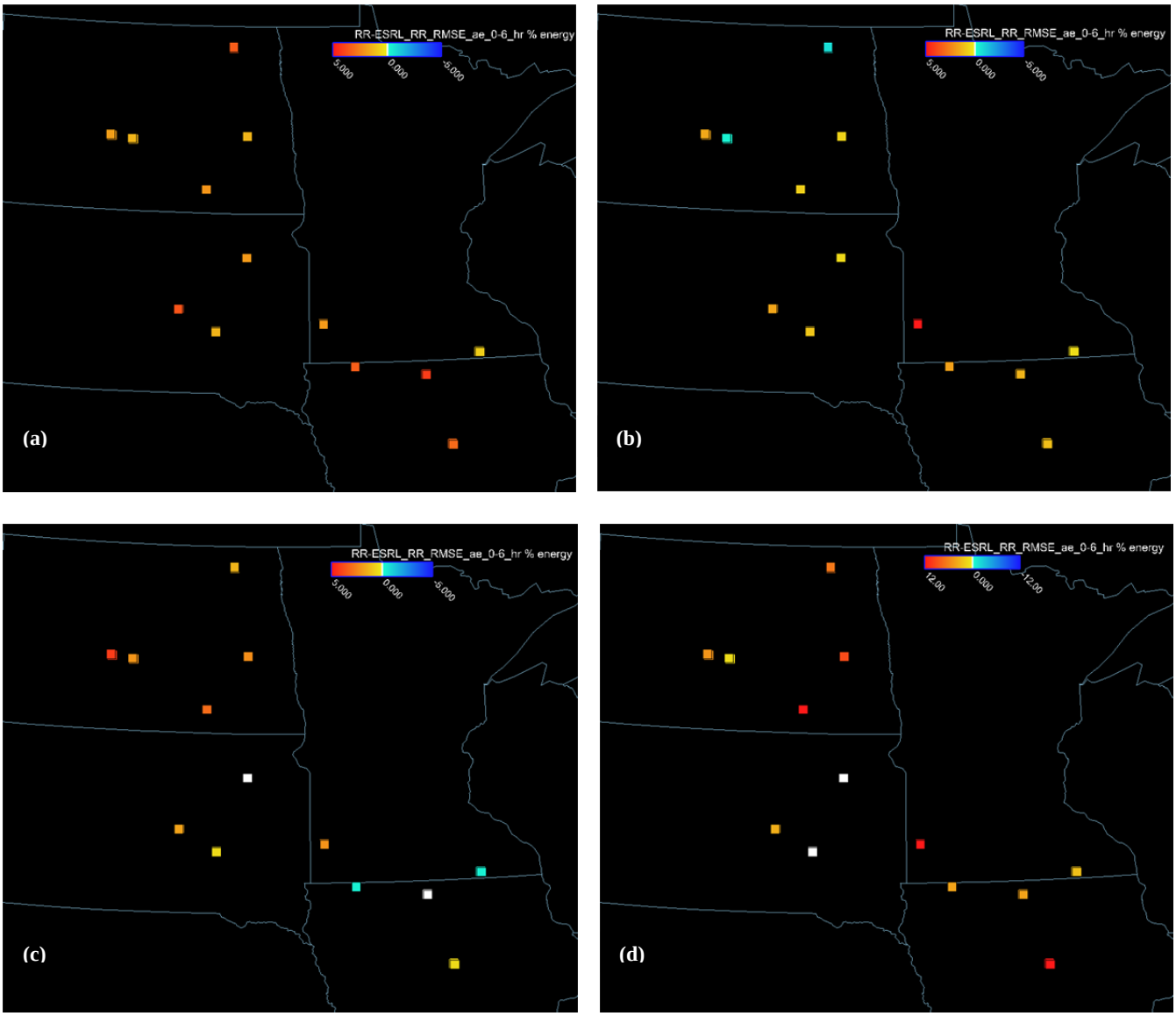

Figure 9: The Percent Improvement (PI) between the ESRL_RAP_BCRW power forecasts and the OP_RAP_BCRW power forecasts over the first six hours of the forecasts for (a) fall (S-O-N), (b) winter (D-J-F), (c) spring (M-A-M) and (d) summer (J-J-A). Warm (cool) colors indicate where the ESRL_RAP-based (OP_RAP-based) bias-corrected raw forecasts are performing better on average. It should be noted that the results for the Day County site (NE South Dakota) in (d) only contain data through Aug. 10, 2012. Note also that the scale on the forecast error difference is different in the summer than the other seasons (panel d).

The Percent Improvement between the operational RAP-based bias-corrected power forecasts and the research ESRL_RAP-based bias-corrected power forecasts is shown in Figure 9. The ESRL_RAP_BCRW power forecasts perform better than the OP_RAP_BCRW power forecasts (in the bulk statistics) at most sites in all seasons. Comparing the results from the various seasons, the ESRL_RAP_BCRW power forecasts don't perform quite as well as the OP_RAP_BCRW power forecasts in the northern most study area during winter (Figure 9b), the southern part of the study domain in spring (Figure 9c), and South Dakota in the summer (Figure 9d). This pattern in forecast performance during different seasons is similar to the pattern seen in the comparison between the HRRR_BCRW and the OP_RAP_BCRW power forecasts (Figure 7), but is significantly less pronounced as that seen in the HRRR-based forecast comparison. The relative impacts of data

assimilation and forecast model improvements on the seasonal improvements in the research model-based raw forecasts will be discussed in more detail in Section 4.6.

4.1.3 Analysis by runtime and forecast hour

Overall bulk wind energy forecast error statistics at individual wind plant locations were shown in the previous sections, but forecast error as a function of time of day is also of interest to system operators as the challenges in balancing the system may be more sensitive to wind energy forecast errors during certain times of the day (for example, those periods where load is changing rapidly in the morning and evening) or during the overnight hours when load is low. The results below show the Mean Absolute Error (MAE) normalized by plant rated capacity as a function of forecast 'runtime' (i.e., the time in when the forecast is generated) and forecast hour for the first 12 hours of the forecasts for the two largest aggregate sites in study area. The statistics were calculated from Sept. 15, 2011 – the end of August 2012 for the bias-corrected raw model-based power forecasts.

Figure 10 shows the MAE as a function of runtime and forecast hour for the Ashtabula plant aggregate in eastern North Dakota. The results shown are for the ESRL_RAP_BCRW power forecasts, although the diurnal pattern in the forecast error is similar for all the model-based bias-corrected raw power forecasts. As can be seen in Figure 10, forecast error is largest during the late night/early morning hours, and is smallest early in the forecasts when forecasts models are initialized during mid-morning to late afternoon hours. These results are consistent with the results of other studies (Banta et al., 2013; Storm et al., 2009; Dabberdt et al., 2004; Zhong and Fast, 2003;, Seaman, 2000) which showed that forecast models have more difficultly accurately representing near-surface weather parameters during stable atmospheric conditions (most prevalent during the nighttime hours) than during the daytime when the boundary layer is well mixed and more homogeneous.

There is also a secondary maximum in forecast error that occurs during the hours around sunrise when the boundary layer is transitioning from stable to unstable conditions. This time period corresponds to the ramp-up in operating system load as the business day begins, so it could have additional significance to power system operators. The development of more accurate boundary layer representations in forecast models during stable atmospheric conditions and boundary layer transition periods is an area where more research and work is needed, and would greatly benefit the wind energy industry (as well as all other forecast users). Also of note is that the wind energy forecast error is significant even at the beginning of forecasts.

The results comparing the hourly Mean Absolute Error of the ESRL_RAP _BCRW and HRRR_BCRW power forecasts vs. the OP_RAP_BCRW power forecasts at the two largest aggregate wind plants are shown in Figure 11 - Figure 14. Figure 11 shows the difference in MAE between the OP_RAP_BCRW and ESRL_RAP_BCRW power forecasts as a function of model runtime and forecast hour at the Ashtabula site in North Dakota. A similar plot of the results at the Crystal Lake aggregate site in Iowa is shown in Figure 13. Red and green shading indicates where the ESRL_RAP-based bias-corrected power forecasts are performing better, and blue shading indicates where the OP_RAP-based bias-corrected power forecasts are performing better. As can be seen in both figures, the ESRL_RAP-based forecasts are more accurate than the OP_RAP-based forecasts at most forecast hours at both sites, although the ESRL_RAP-based forecasts struggle a bit more in the very early morning hours at the Iowa site compared to the OP_RAP-based forecasts. This difference could be due to differences in the way the boundary layer representations in the two models handle nocturnal situations in more moist (Iowa) vs. drier (North Dakota) conditions, or differences in the representation of land surface processes.

Figure 12 shows the difference in MAE between the OP_RAP and the HRRR-based bias-corrected power forecasts as a function of the forecast 'runtime' (i.e. the starting time of the weather model forecast) and forecast hour at the Ashtabula site in North Dakota. A similar plot of the results at the Crystal Lake aggregate site in Iowa is shown in Figure 14. Red and green shading indicates where the HRRR-based forecasts are more accurate, and blue shading indicates where the OP_RAP-based forecasts are more accurate. As can be seen in both figures, the HRRR_BCRW power forecasts are performing better than the OP_RAP_BCRW power during the late night/early morning hours (i.e., during the hours where, in general, forecast models tend to have the highest errors as shown in Figure 10). This pattern is more pronounced at the North Dakota site as compared to the Iowa site, perhaps for a the same reasons that were discussed above. The HRRR_BCRW power forecasts are also more accurate than the OP_RAP_BCRW power forecast at the start of the forecasts for most of the forecast runtimes.

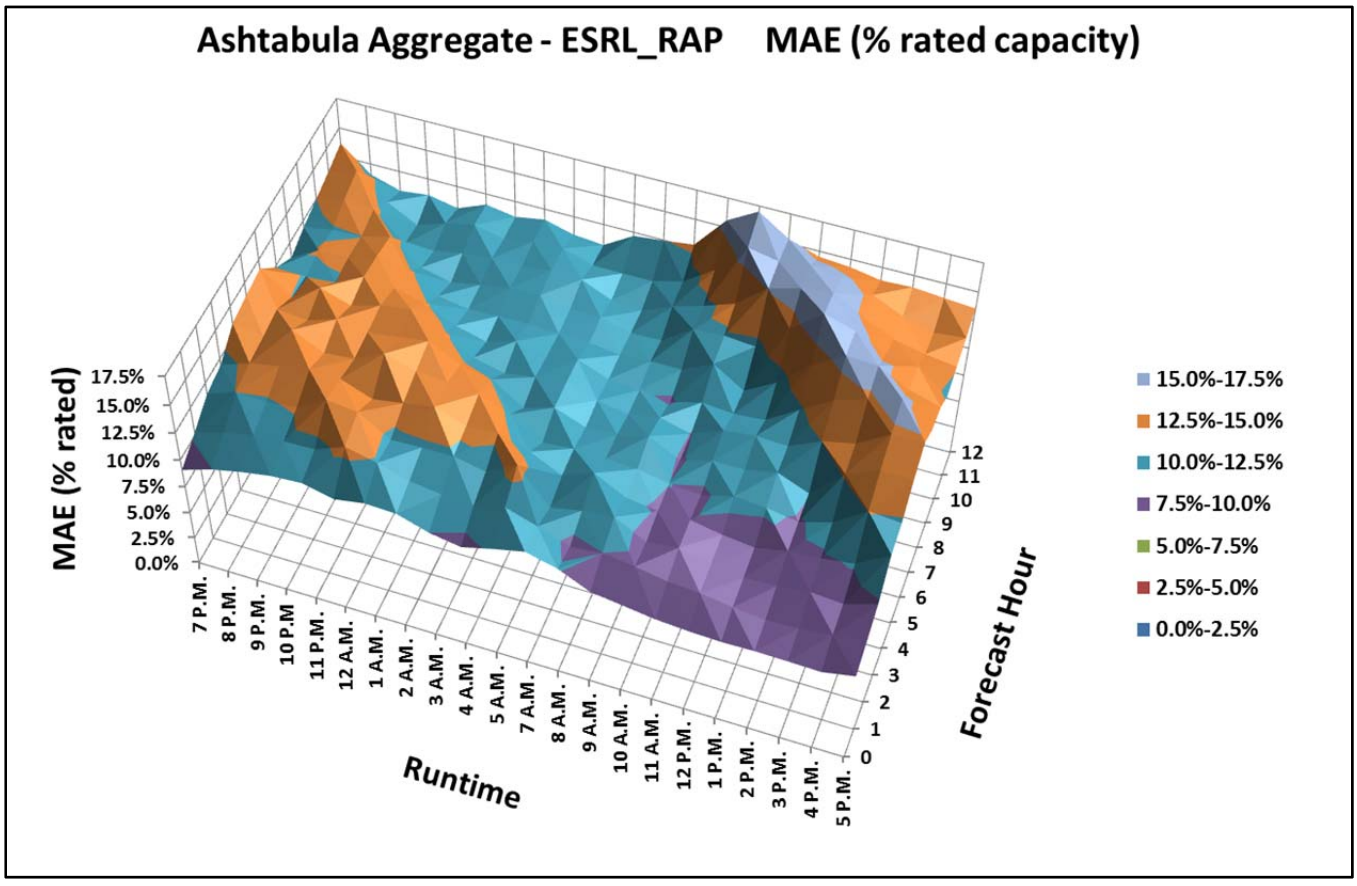

Figure 10: Mean Absolute Error (MAE) normalized by plant rated capacity as a function of model runtime (in CST) and forecast hour for the Ashtabula aggregate site in eastern North Dakota for the ESRL_RAP-based bias-corrected power forecasts. The error scale is shown at right. The statistics were calculated from September 15, 2011 – end of August 2012.

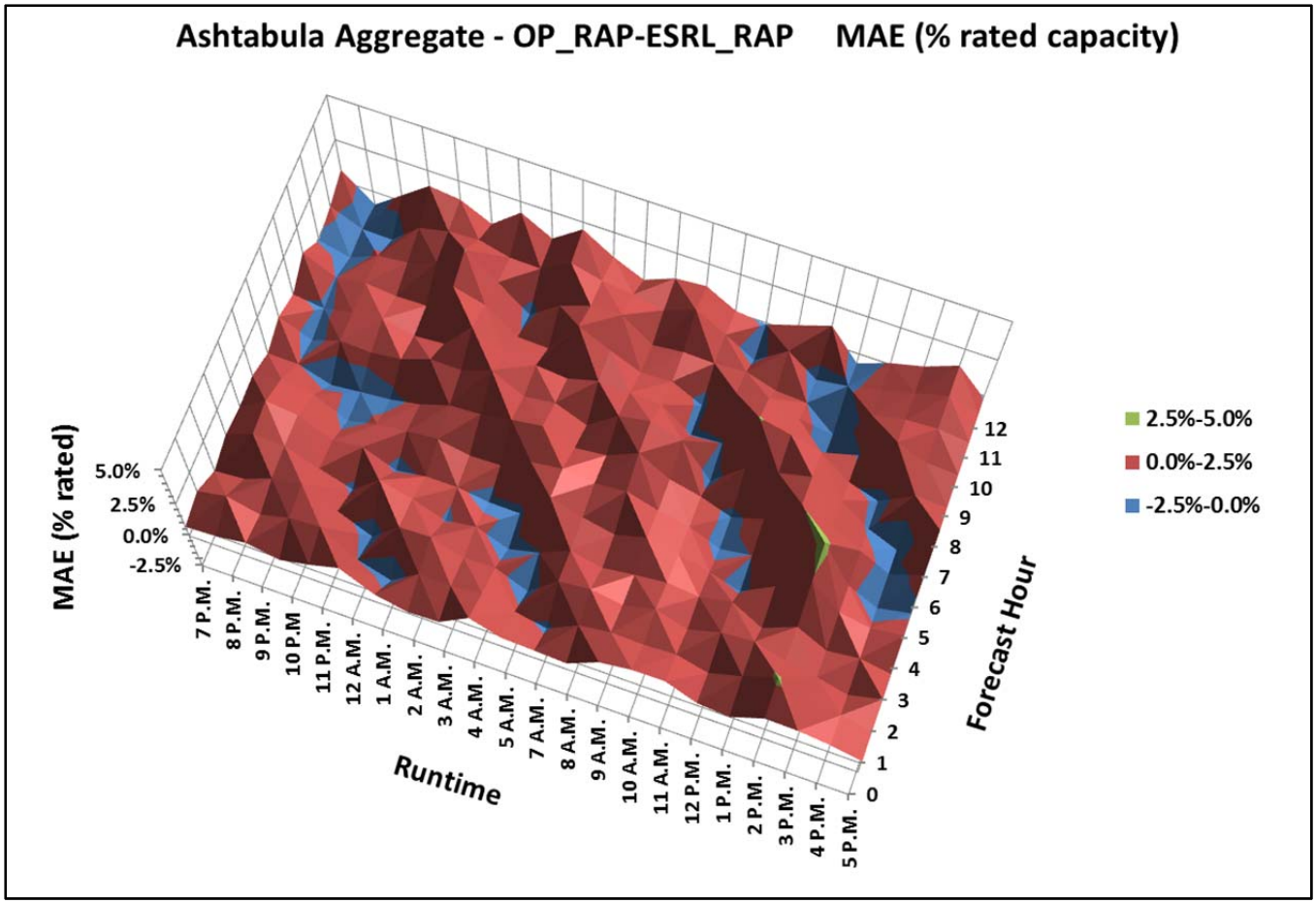

Figure 11: Difference in MAE between the OP_RAP_BCRW and ESRL_RAP_BCRW power forecasts ((MAE(OP_RAP_BCRW) – MAE(ESRL_RAP_BCRW)) expressed as a percentage of plant rated capacity) as a function of forecast runtime (i.e. the time the forecast is started in CST) and forecast hour for the first 12 hours of the forecasts. Red (blue) colors indicate where the ESRL_RAP-based (OP_RAP-based) bias-corrected raw power forecasts are performing better on average. The statistics were calculated from September 15, 2011 – end of August 2012 for the Ashtabula aggregate site in eastern North Dakota.

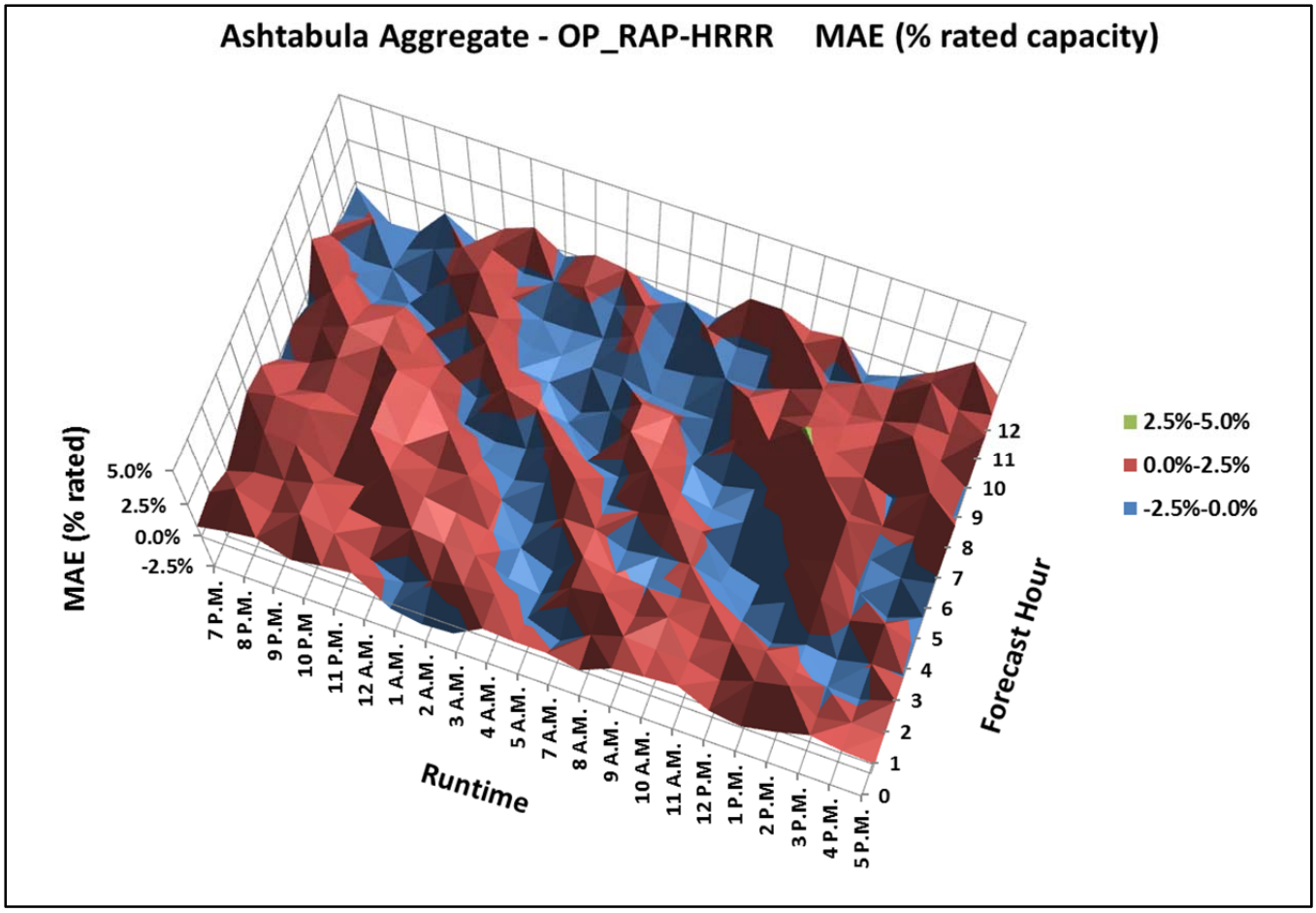

Figure 12: Difference in MAE between the OP_RAP_BCRW and HRRR_BCRW power forecasts ((MAE(OP_RAP_BCRW) – MAE(HRRR_BCRW)) expressed as a percentage of plant rated capacity) as a function of forecast runtime (i.e. the time the forecast is started in CST) and forecast hour for the first 12 hours of the forecasts. Red (blue) colors indicate where the HRRR-based (OP_RAP-based) bias-corrected raw power forecasts are performing better on average. The statistics were calculated from September 15, 2011 – end of August 2012 for the Ashtabula aggregate site in eastern North Dakota.

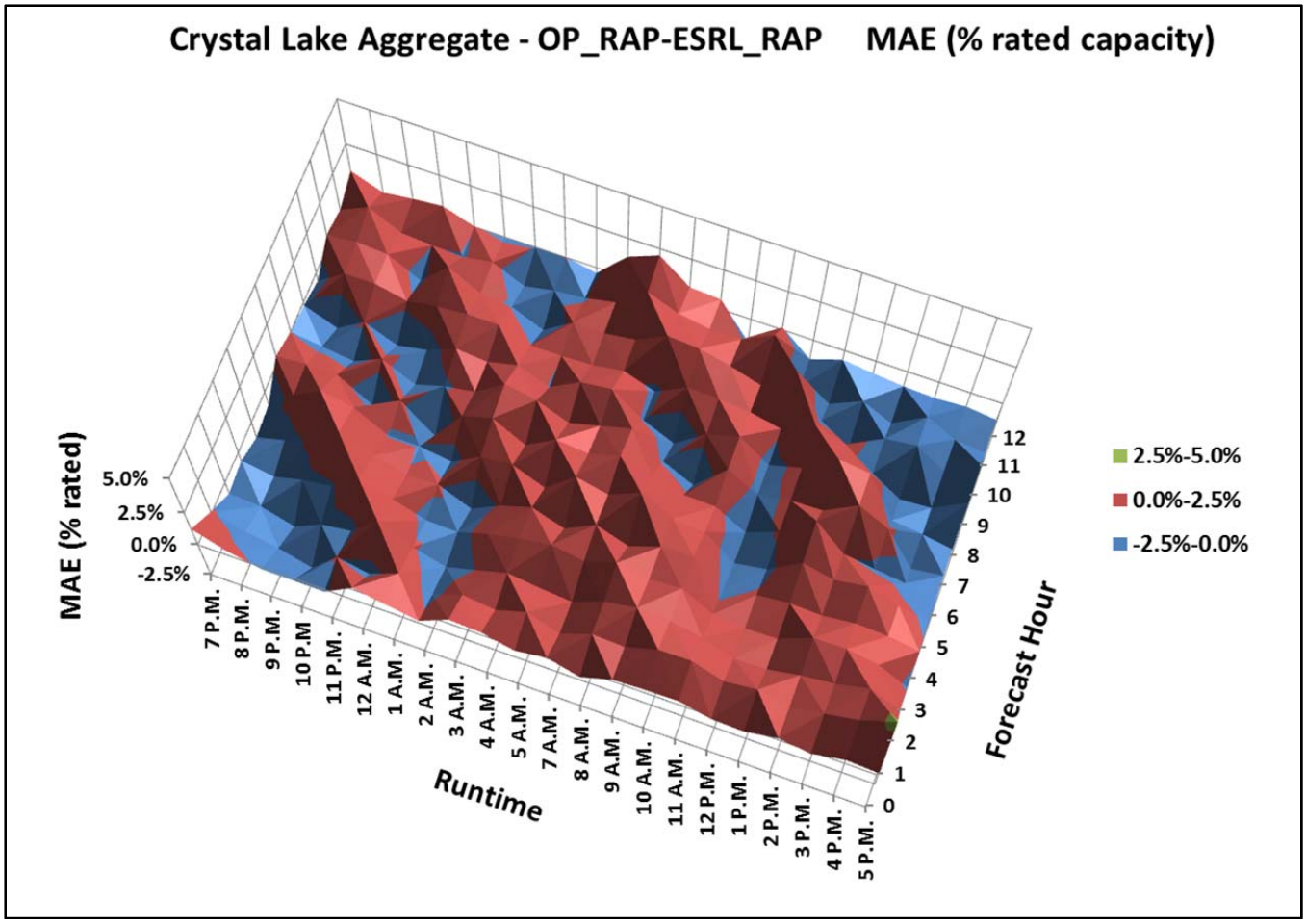

Figure 13: Difference in MAE between the OP_RAP_BCRW and ESRL_RAP_BCRW power forecasts ((MAE(OP_RAP_BCRW) – MAE(ESRL_RAP_BCRW)) expressed as a percentage of plant rated capacity) as a function of forecast runtime (i.e. the time the forecast is started in CST) and forecast hour for the first 12 hours of the forecasts. Red (blue) colors indicate where the ESRL_RAP-based (OP_RAP-based) bias-corrected raw power forecasts are performing better on average. The statistics were calculated from September 15, 2011 – end of August 2012 for the Crystal Lake aggregate site in north central Iowa.

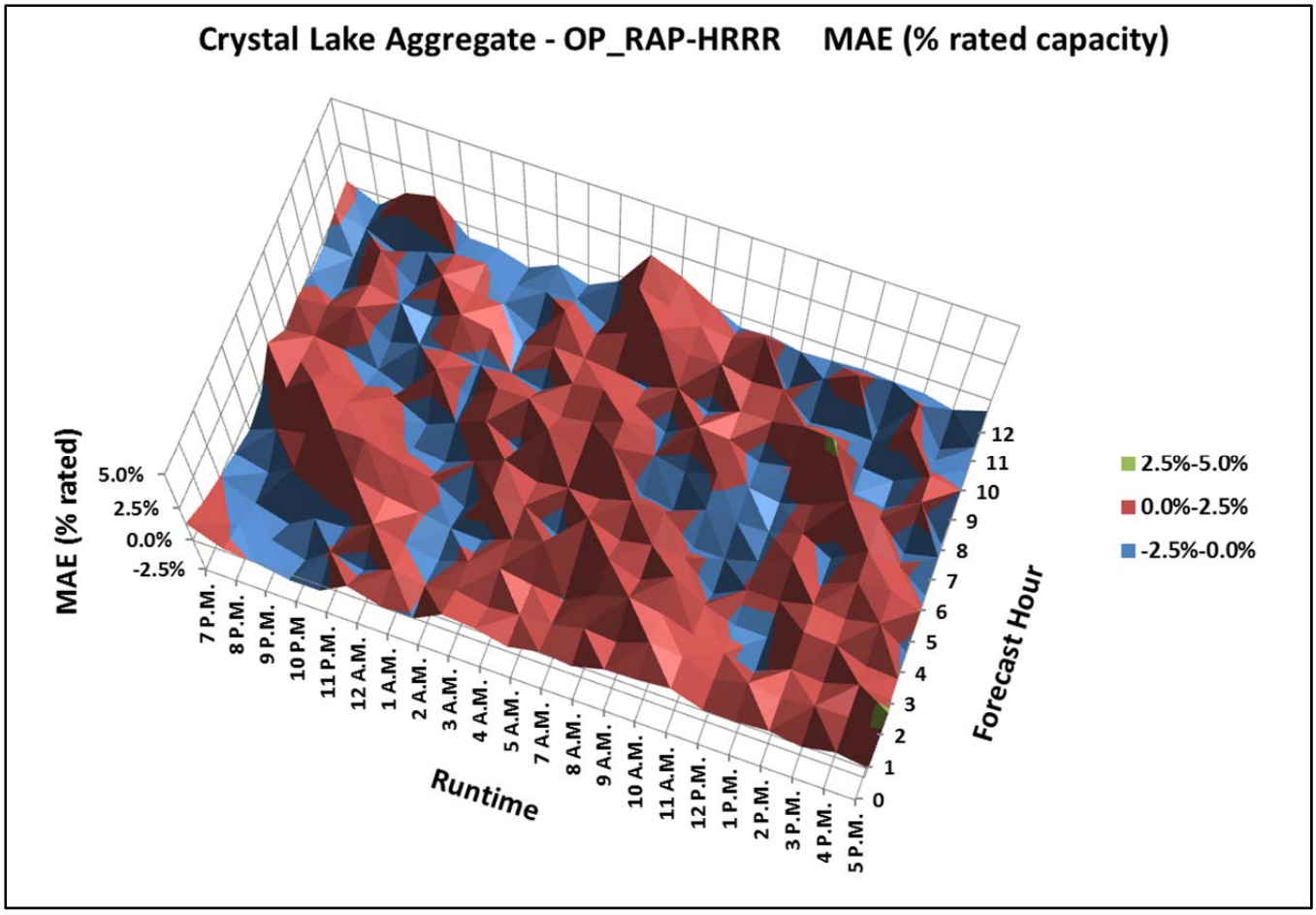

Figure 14: Difference in MAE between the OP_RAP_BCRW and HRRR_BCRW power forecasts ((MAE(OP_RAP_BCRW) – MAE(HRRR_BCRW)) expressed as a percentage of plant rated capacity) as a function of forecast runtime (i.e. the time the forecast is started in CST) and forecast hour for the first 12 hours of the forecasts. Red (blue) colors indicate where the HRRR-based (OP_RAP-based) bias-corrected raw power forecasts are performing better on average. The statistics were calculated from September 15, 2011 – end of August 2012 for the Crystal Lake aggregate site in north central Iowa.

As was discussed above, the training process improves all forecasts overall, but it improves the forecasts at some hours more than others as shown in Figure 15 and Figure 16 below. Figure 15 shows the difference in MAE between the ESRL_RAP bias-corrected raw power forecasts and the ESRL_RAP trained forecasts for the Ashtabula Aggregate site, while Figure 16 shows the same statistics for the Crystal Lake Aggregate. The training process provides the most improvement to the forecasts during the same time periods that have the most forecast error in the raw power forecasts – the late evening/early morning hours when there is a stable boundary layer, and during the few hours around sunrise as the boundary layer transitions between stable and unstable. The improvement is larger at the Crystal Lake site in Iowa than the Ashtabula site in North Dakota, perhaps because the untrained Crystal Lake forecast errors were larger than at Ashtabula. The training process provides the least improvement at the very start of the forecasts (when errors are generally smallest) and during the afternoon and early evening hours when the boundary layer is unstable and vertically well-mixed.

These are rather interesting results. The HRRR model tends to help reduce forecast error during the hours when forecasts generally have the most difficulty, and these are hours that may also have increased significance to power system operations. The training process also tends to reduce the error

the most during the hours that generally have higher forecast error, so both of these characteristics are favorable in terms of providing value to power system operators.

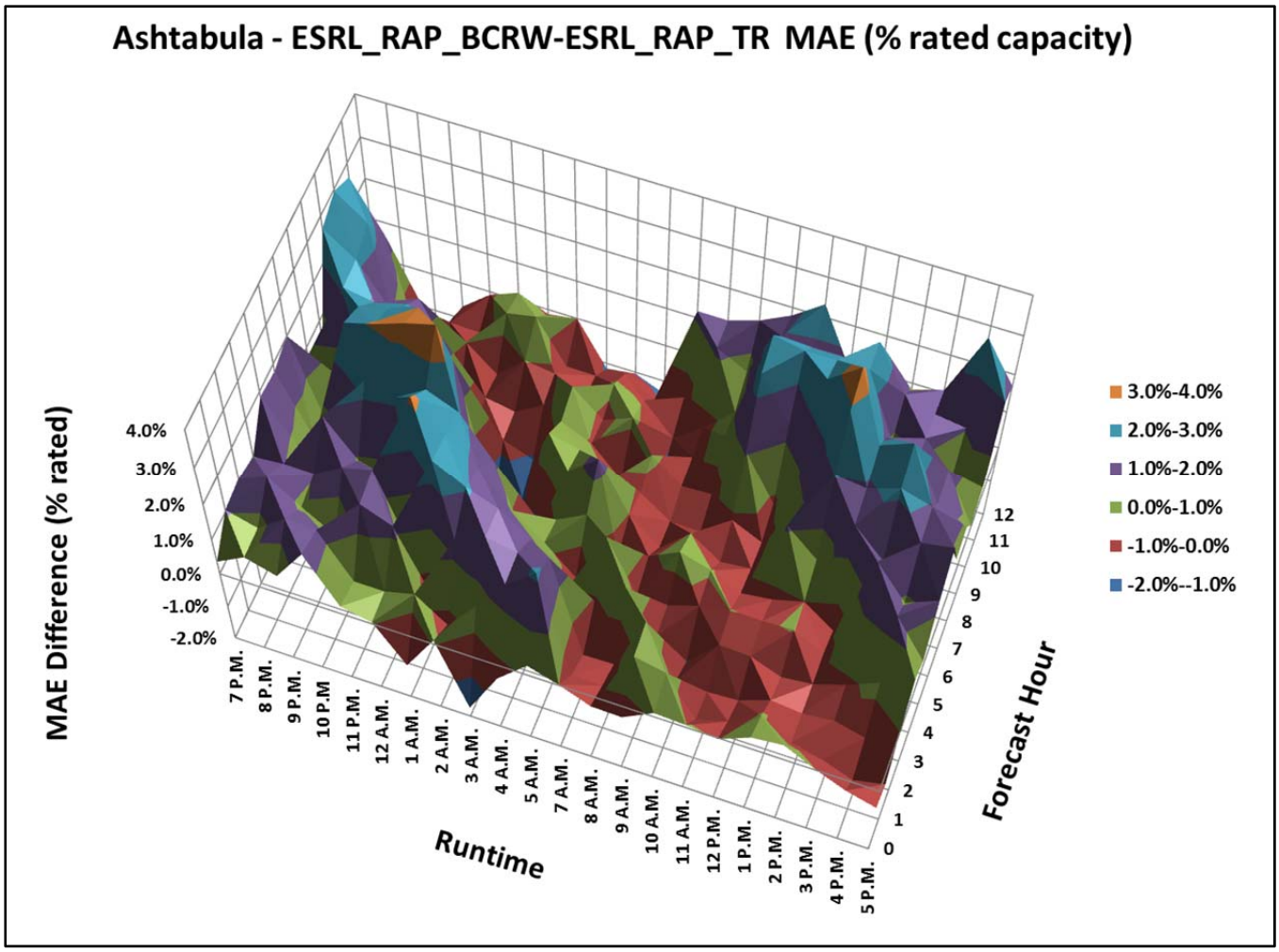

Figure 15: Difference in MAE between the ESRL_RAP_BCRW and ESRL_RAP_TR power forecasts ((MAE(ESRL_RAP_BCRW) – MAE(ESRL_RAP_TR)) expressed as a percentage of plant rated capacity) as a function of forecast runtime (i.e. the time the forecast is started in CST) and forecast hour for the first 12 hours of the forecasts. The statistics were calculated from January 1, 2012 – end of August 2012 for the Ashtabula aggregate site in east central North Dakota.

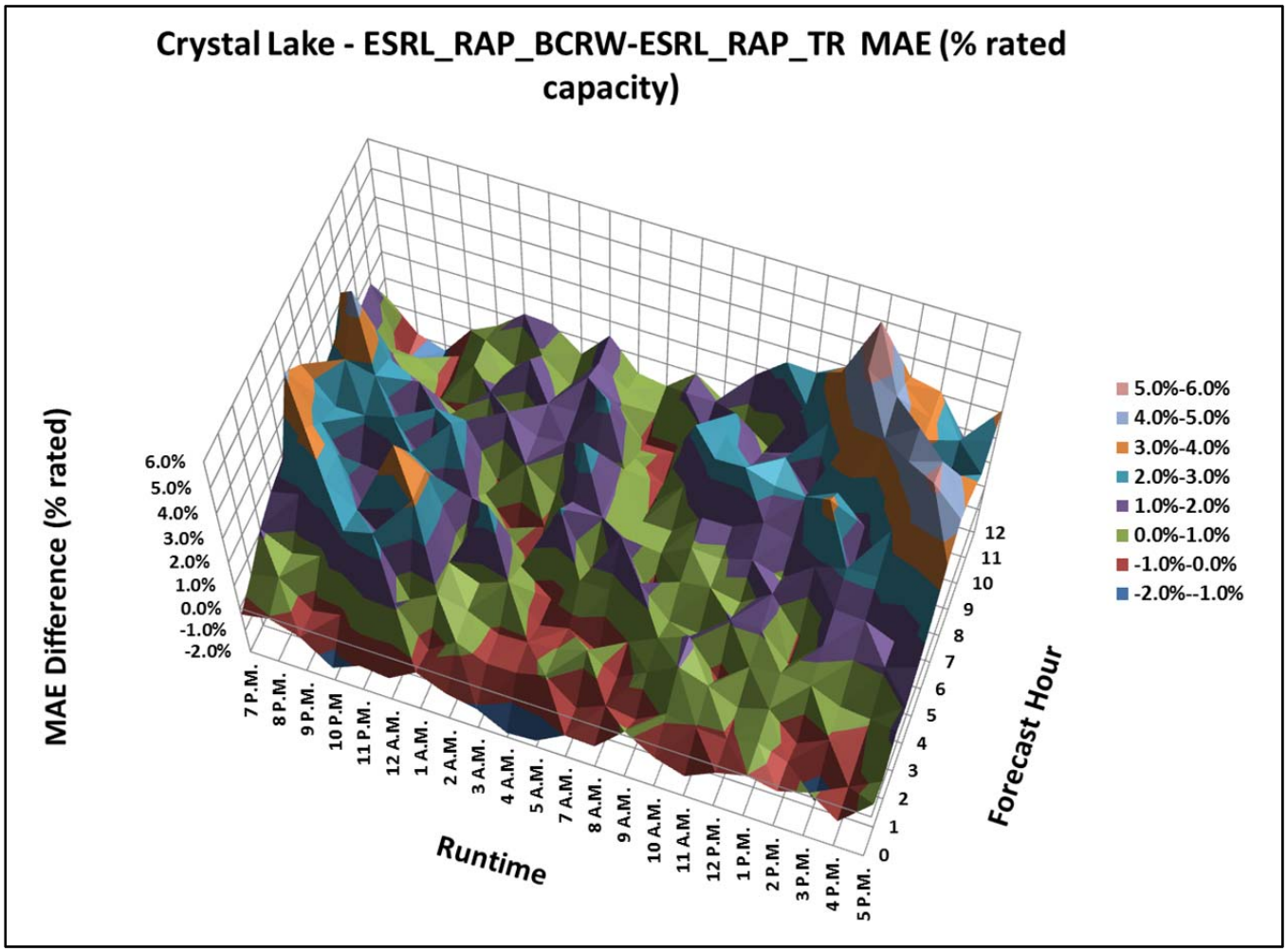

Figure 16: Difference in MAE between the ESRL_RAP_BCRW and ESRL_RAP_TR power forecasts ((MAE(ESRL_RAP_BCRW) – MAE(ESRL_RAP_TR)) expressed as a percentage of plant rated capacity) as a function of forecast runtime (i.e. the time the forecast is started in CST) and forecast hour for the first 12 hours of the forecasts. The statistics were calculated from January 1, 2012 – end of August 2012 for the Crystal Lake aggregate site in north central Iowa.

4.2 Bulk Error Statistics for System Aggregate Forecasts

While individual wind plant errors are important for wind plant owners and useful from a forecast performance perspective, system operators often make decisions based on the total forecasted wind generation for a larger operating area. System aggregate forecasts are therefore important, too, and we looked at the combined error of all the NextEra wind plant forecasts in the study area to assess forecast performance at the operating system level.

The system aggregate forecasts were created by summing the forecasts for each individual site. Note that through this process, forecast error is typically reduced. Forecast errors at different wind plants may be of differing signs (i.e., at some locations the power may be over-forecasted, while at other locations it could be under-forecasted), so that some of the forecast errors at various wind plants will tend to cancel each other out in the aggregate forecast. (This will be demonstrated and discussed in

more detail in Section 4.5.1.) It also raises the possibility that a specific model-based wind power forecast may perform somewhat differently at the local wind plant level and the system aggregate level due to the distribution of forecast errors across the operating area.

Curtailment periods pose a larger challenge in evaluating forecasts at the system aggregate level since one or more wind plants are often curtailed during windy periods. While these periods can be removed from the evaluation dataset at the local plant level with little loss of data, at the system level the curtailment of any one wind plant would require the data point be removed. This can lead to a significant loss in the number of evaluation data points during windy periods. To work around this issue, the individual wind plant production was estimated during curtailment periods by using the nacelle wind speeds and plant derived power curves prior to aggregating up to the system level. All results shown below at the system level are calculated using aggregates of actual production values during non-curtailed periods and the estimated production during curtailed periods.

4.2.1 Overall Error Statistics

As was discussed in Section 3, the Rapid Update Cycle (OP_RUC) was the operational short-range weather forecast model at the start of the forecast experiment, but was replaced with the Rapid Refresh (OP_RAP) on May 1, 2012. Since all four model-based forecasts were available for the period from September 2011 – April 2012, a four-way comparison of the RMSE as a percent of the rated capacity for the bias-corrected system aggregate raw power forecasts was made during this period (Figure 17). As can be seen in Figure 17, forecast error generally increases with forecast hour (as should be expected) and the ESRL_RAP_BCRW power forecasts have the lowest overall RMSE for all forecast hours. Both the ESRL_RAP_BCRW and OP_RAP_BCRW power forecasts have lower overall RMSE values than the OP_RUC_BCRW power forecasts, indicating that the new operational forecast model (OP_RAP) is already an improvement over the previous operational forecast model (OP_RUC) for wind energy forecasting purposes.

An analysis over the entire year of the forecasting project is shown in Figure 18, which also includes both the raw power forecasts (i.e., power values calculated directly from the forecast model winds without any statistical correction applied) and the bias-corrected raw power forecasts. As can be seen in Figure 18, bias correcting the model wind speeds prior to calculating power improves the power forecasts overall, particularly for the HRRR-based power forecasts. Over the entire year of the forecasting project, the ESRL_RAP bias-corrected raw power forecasts have the lowest RMSE values at all forecast hours, and the gap between the forecast errors between the ESRL_RAP and OP_RAP–based bias-corrected power forecasts has widened (and the gap between the OP_RAP and HRRR-based bias corrected power forecasts has been reduced) compared to the period from September 2011 – April 2012 shown in Figure 17. This is due to the significant differences in system-wide seasonal forecast errors between the various model-based forecasts which will be discussed further below.

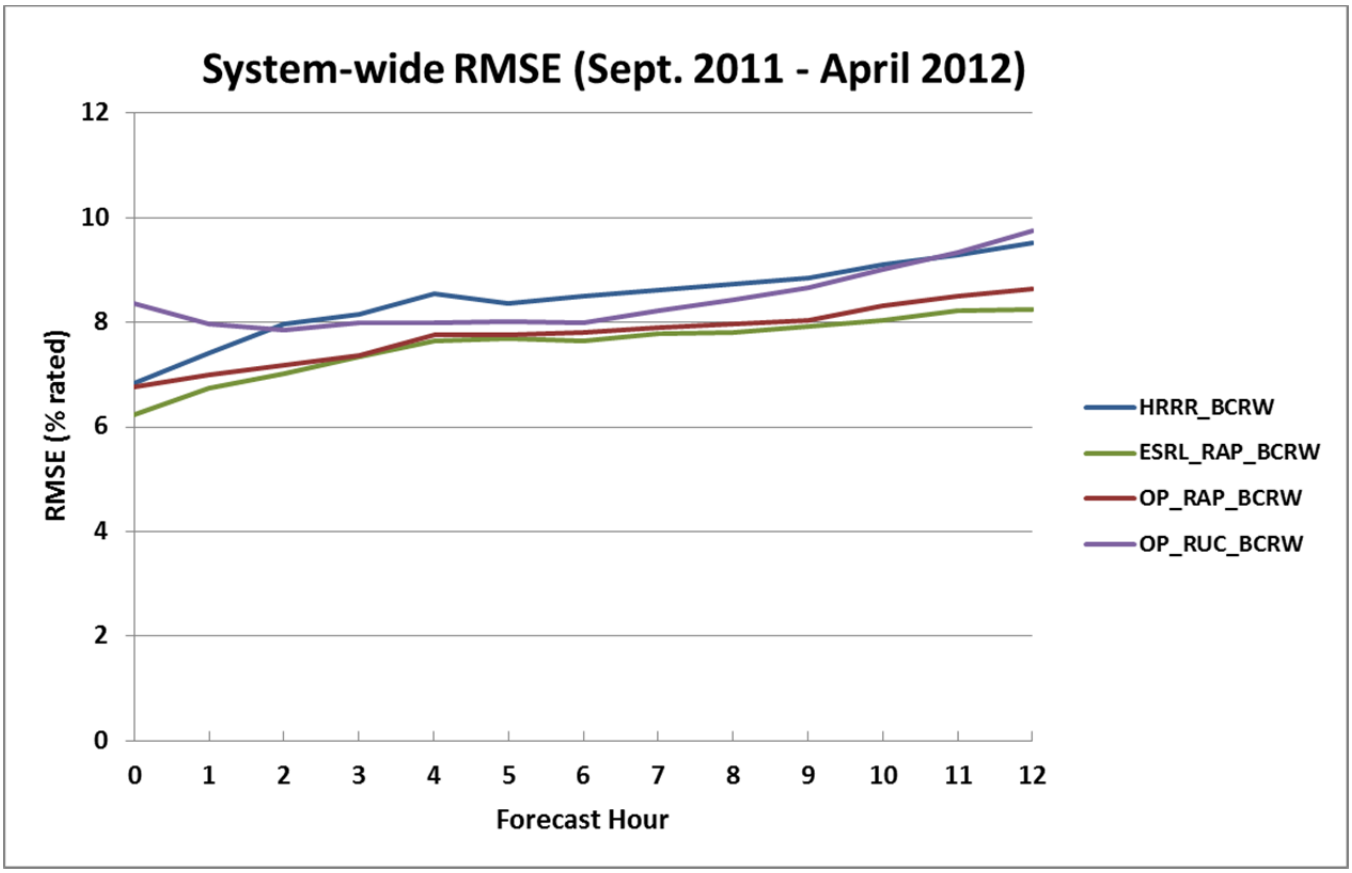

Figure 17: Root Mean Square Error (as a percent of rated capacity) as a function of forecast hour for the various model-based bias-corrected raw system-wide aggregate power forecasts. The RMSE was calculated over the period from September 2011 through April 2012.

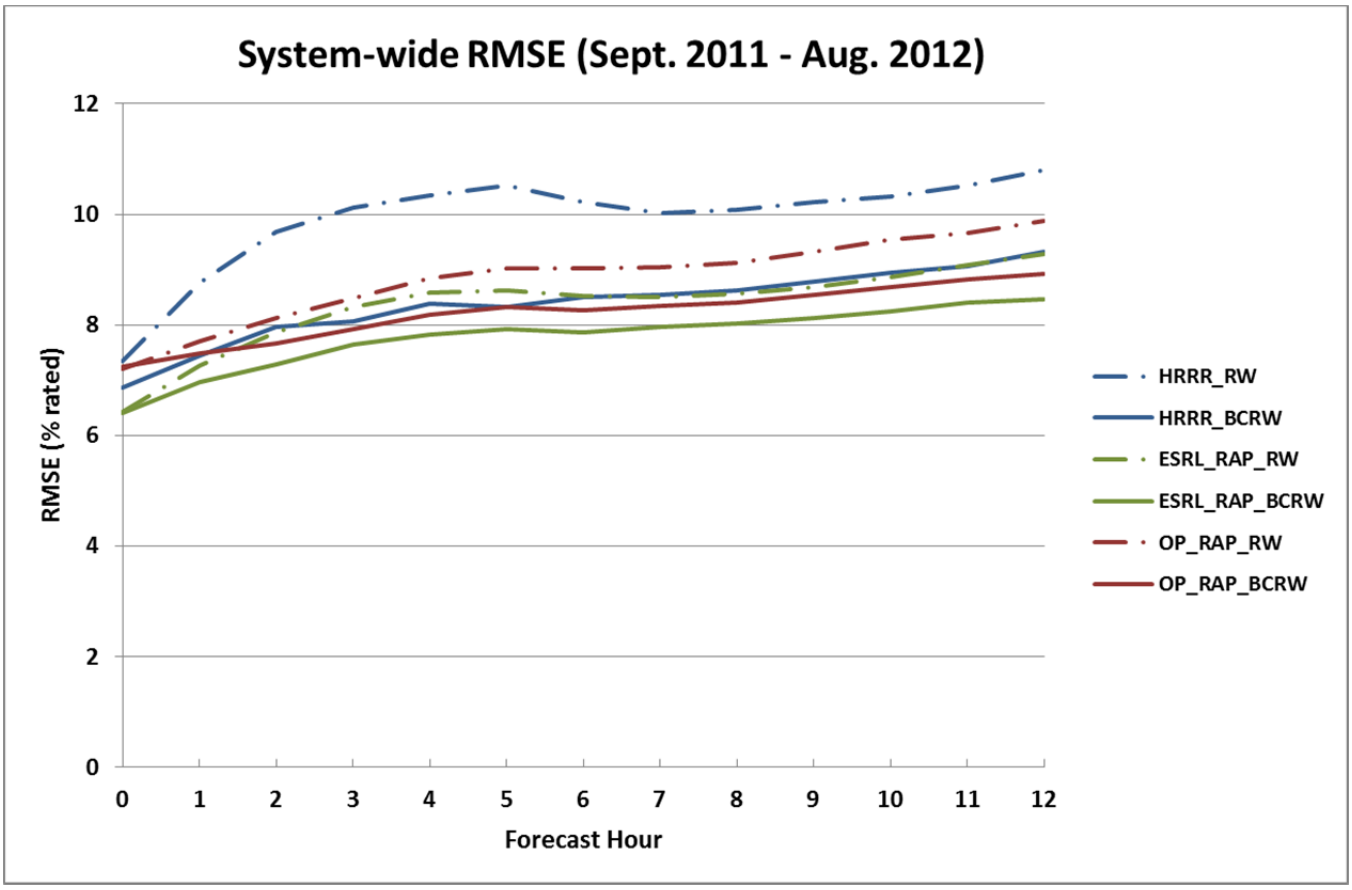

Figure 18: Root Mean Square Error (as a percent of rated capacity) as a function of forecast hour for the various model-based raw and bias-corrected raw system-wide aggregate power forecasts over the entire year of the forecasting project (September 2011 through August 2012).

The impacts of the training process on the system-wide RMSE for the various model-based power forecasts are shown in Figure 19. The training process improves the overall power forecast RMSE for all model-based forecasts compared to the bias-corrected raw power forecasts, but the improvement is larger for the coarser-resolution model-based forecasts (0.32% for ESRL_RAP and 0.53% OP_RAP over the first 6 hours of the forecasts) than for the higher resolution model-based forecast (0.02% for HRRR over the first 6 hours). (Note however that the training process starts with the forecast variables directly from the model, so the improvement in the trained forecasts relative to the raw power forecasts is much larger for the HRRR-based power forecasts than the coarser resolution model-based forecasts, similar to the improvement seen between the raw and bias-correct raw power forecasts in Figure 18.) Possible reasons for this were discussed above in Section 3.5. As with the bias-corrected raw power forecasts, the RMSE values for the trained ESRL_RAP-based power forecasts are smaller than the OP_RAP and HRRR-based trained power forecasts.

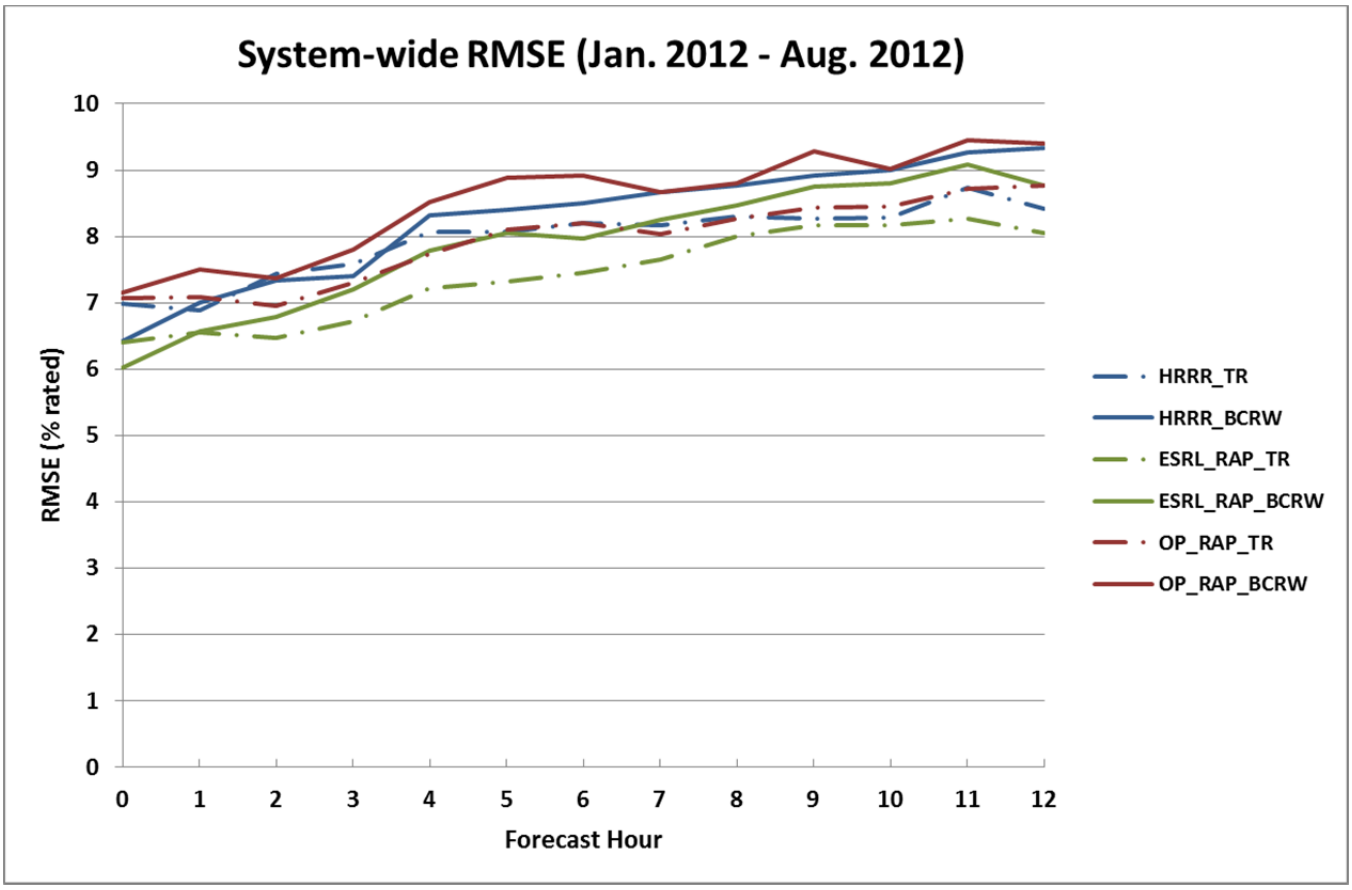

Figure 19: Root Mean Square Error (as a percent of rated capacity) as a function of forecast hour for the various model-based bias-corrected system-wide aggregate raw and trained power forecasts over the entire time period over which trained forecasts were generated (January 2012 through August 2012).

Although not the focus of this study, the trained ensemble power forecasts (a trained combination of a short-range model-based power forecast and the 'day-ahead' power forecast generated from the NAM forecast model) further improve the overall power error statistics as shown in Figure 20. The trained ensemble power forecasts have the lowest overall RMSE at all forecast hours, particularly at the start of the forecast where the persistence forecast is weighted most heavily in the forecast generation process (often termed 'smart persistence'). Although there is some variability between the various weather forecast models in the amount of power forecast improvement seen through the use of statistical techniques, Figure 20 serves as an illustration of the power forecast improvements that can be seen from the various statistical techniques based on operational weather forecast models. Since the goal of this study is to evaluate power forecast improvements through improvements to short-range weather forecast models and additional data assimilation, the inclusion of a longer-range forecast model in the forecasts will tend to make this evaluation more difficult and thus the trained ensemble forecasts will not be included in the rest of the analysis that follows.

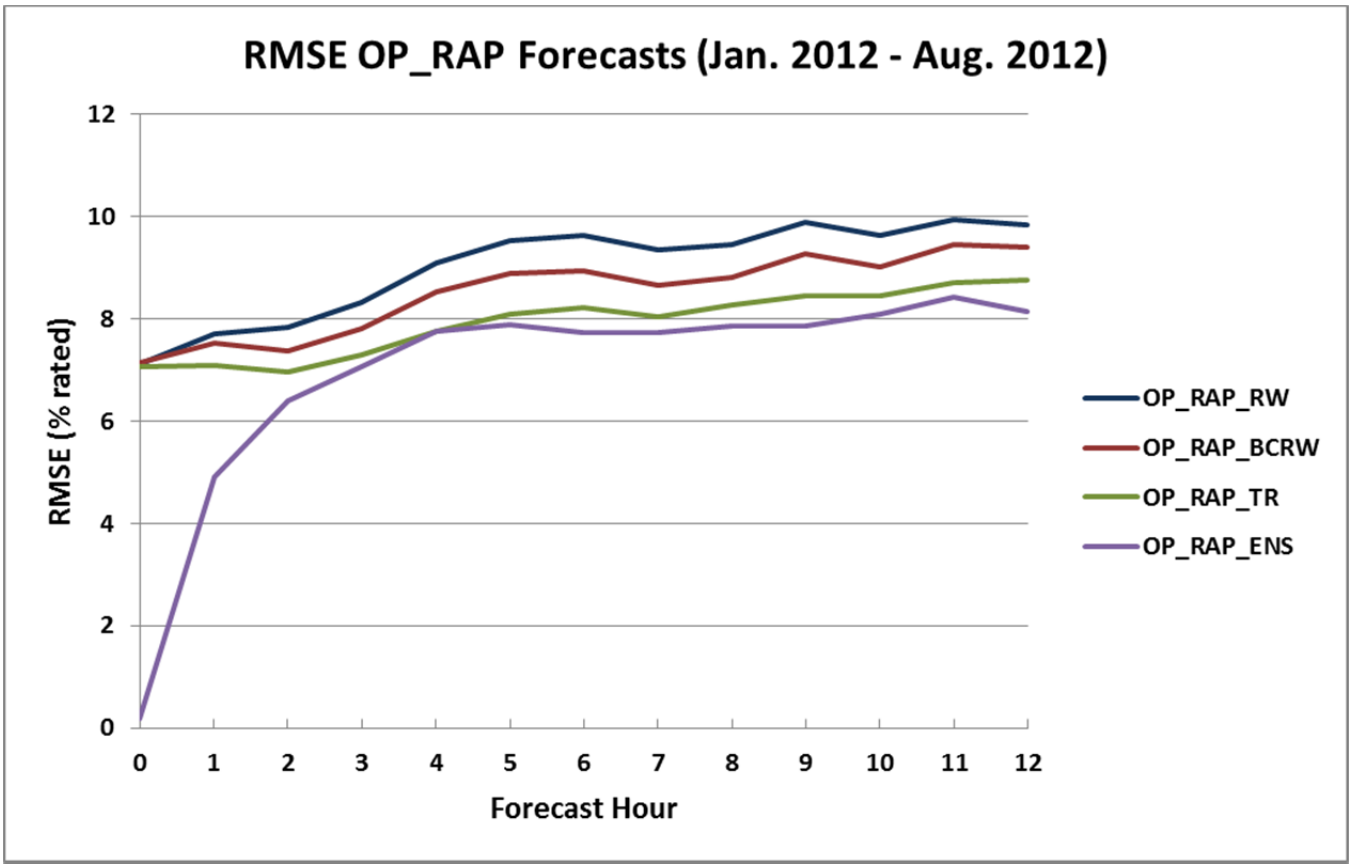

Figure 20: Root Mean Square Error (as a percent of rated capacity) as a function of forecast hour for the system-wide aggregate raw, bias-corrected raw, trained and trained ensemble OP_RAP-based power forecasts over the time period which trained forecasts were generated (January 2012 through August 2012).

4.2.2 Seasonal

As was discussed previously, forecast performance can be impacted by many different aspects of atmospheric flows, including seasonal variations in the types of meteorological features that impact local weather (and winds), and the fact the some weather patterns are more predictable than others (Morss et al., 2009; Ngan et al., 2009; Boer, 1994). A seasonal analysis similar to that presented for individual wind plants was done at the system aggregate forecast level to assess the system-wide seasonal forecast performance and improvements.

The RMSE for the raw and bias-corrected raw model-based power forecasts for the OP_RAP, ESRL_RAP and HRRR, broken down by each season, are shown in Figure 21 - Figure 24. As can be seen by comparing the results from each season, there is quite a bit of variability in forecast performance from season to season, with the overall system aggregate RMSE lowest for the model-based raw forecasts in the winter months, and highest in the summer months. Forecasting winds in the northern plains during the summer months is more challenging than other seasons due to weaker dynamical forcing mechanisms responsible for the development and evolution of larger-scale weather

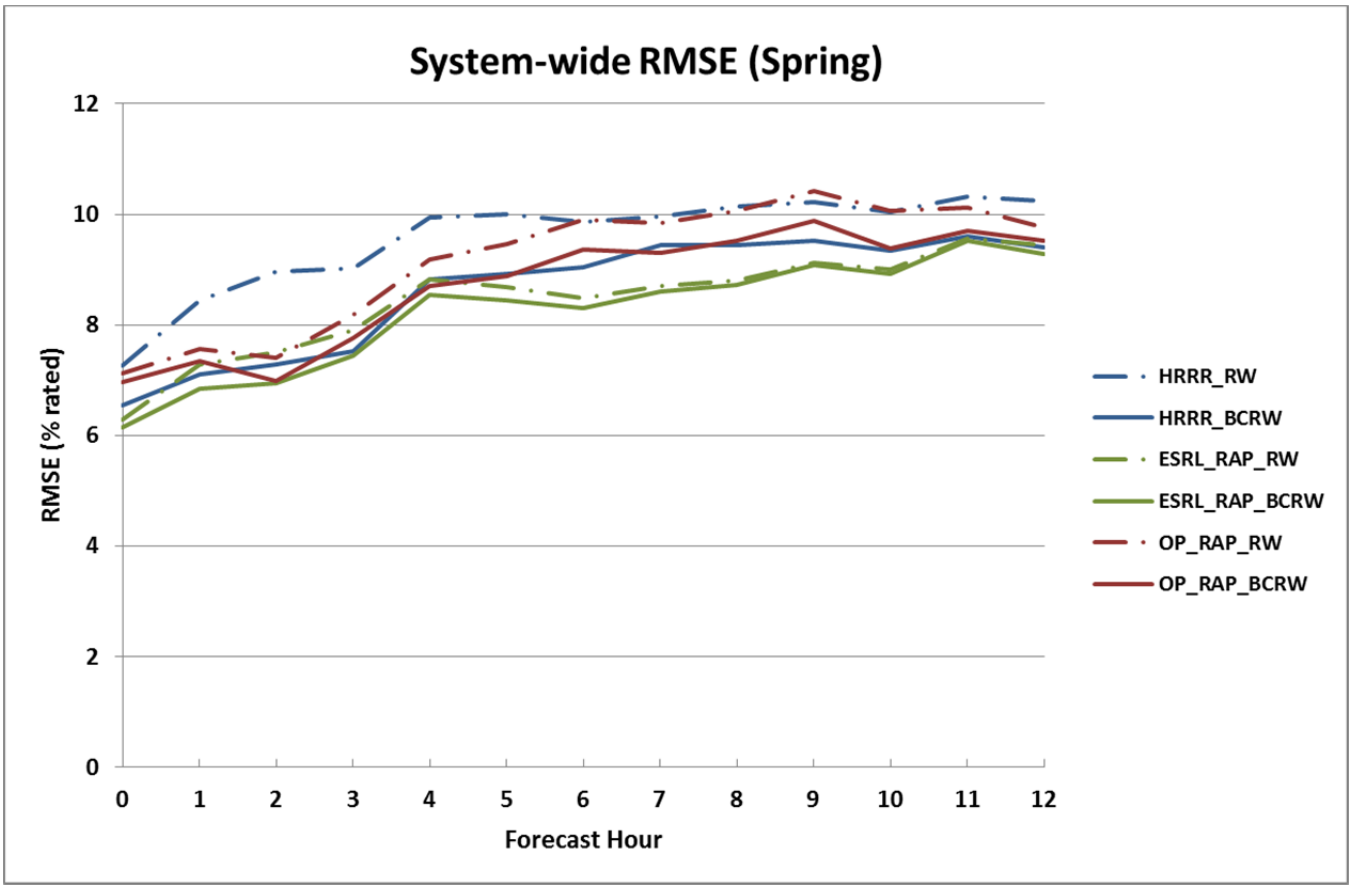

Figure 21: Root Mean Square Error (as a percent of rated capacity) as a function of forecast hour for the various model-based bias-corrected system-wide aggregate raw and bias-corrected raw power forecasts over the spring months (March, April and May 2012).

systems, and the complex windflows generated by thunderstorms. Bias-correcting the wind speeds before converting to power reduces the RMSE for the HRRR-based forecasts by 34% and the ESRL_RAP-based forecasts by 25% over the first 12 hours of the forecasts during the summer months as can be seen in Figure 22, indicating that a significant portion of the power forecast error during the summer is due to wind speed bias in these forecasts. Note also that during the summer months, the OP_RAP_BCRW power forecasts have significantly higher error rates than the HRRR_BCRW and ESRL_RAP_BCRW power forecasts (Figure 22). It should also be noted that the summer of 2012 was anomalous in that temperatures across the study area were well above average, and the southern half of the study area experienced extreme drought.

However, a similar bias correction does not improve the power forecasts in the winter months as can be seen in Figure 24, particularly for the coarser resolution model-based power forecasts that already have relatively small RMSE values in winter. The winds in the northern plains during the winter months are driven by large-scale weather systems (i.e. areas of low and high pressure), which are usually better predicted by the weather forecast models than smaller-scale flow features. There is also a significant gap between the performance of the HRRR-based power forecast and the coarser-resolution model-based power forecasts in the winter months, and the winter RMSE values for the HRRR-based forecasts are larger than in all other seasons. An in-depth look at the HRRR, ESRL_RAP and OP_RAP model wind profiles would be required to determine the root cause of this performance gap.

Winds during the transition seasons (spring and fall) are challenging to forecast because the meteorological conditions can transition between winter and summer-type flow patterns. During the transition seasons (Figure 21 and Figure 23) the ESRL_RAP_BCRW power forecasts perform better than OP_RAP_BCRW and HRRR_BCRW power forecasts, but there is less spread among the various model-based forecasts than in the summer and winter months. The forecast error rates during the fall and spring are also higher on average than the summer and winter months (with the exception of the OP_RAP-based forecasts, which have the highest error rates in summer, and the HRRR-based forecasts, which have the highest error rates in winter) for reasons that were discussed above.

When it comes to power system operations, this level of complexity in identifying the best weather model for a particular seasonal regime is yet another argument for using advanced, multi-model ensemble wind power forecasting systems in actual production. With proper ensemble training methods, such systems should be able to implicitly extract the skill from different models for the various regimes. The results here clearly show that a single model is unlikely to be the best choice at least for some seasons of the year.

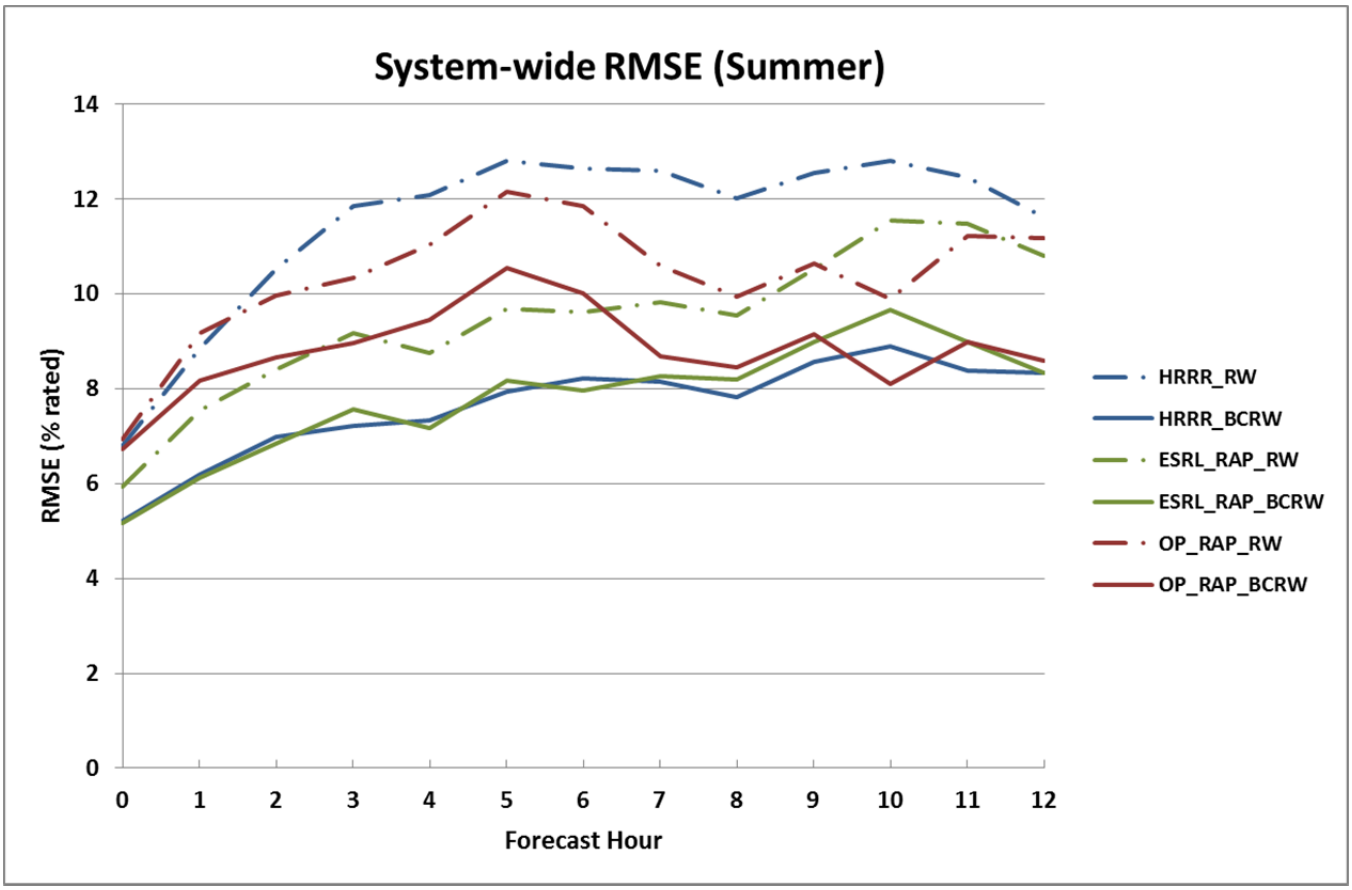

Figure 22: Root Mean Square Error (as a percent of rated capacity) as a function of forecast hour for the various model-based bias-corrected system-wide aggregate raw and bias-corrected raw power forecasts over the summer months (June, July and August 2012).

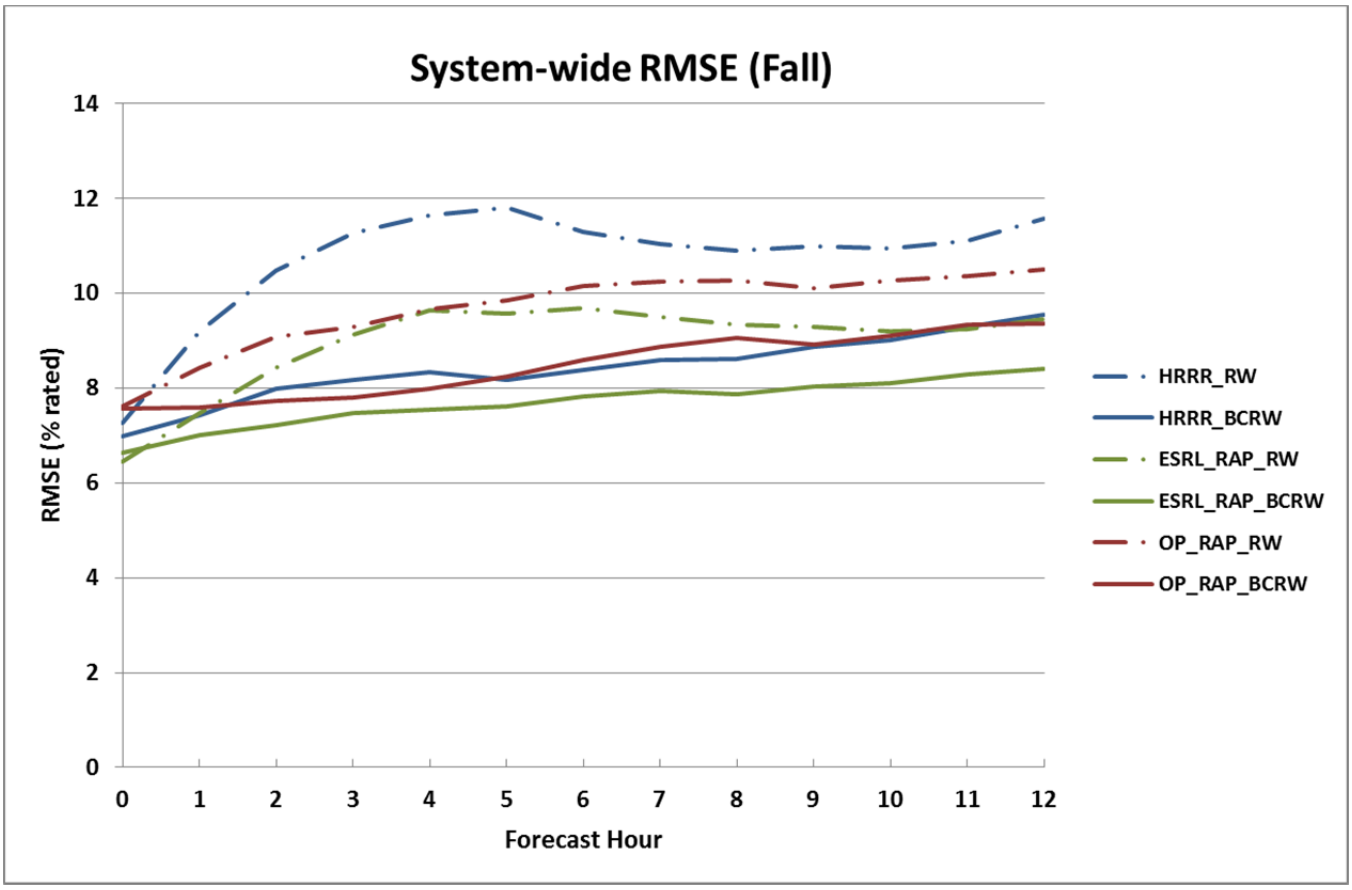

Figure 23: Root Mean Square Error (as a percent of rated capacity) as a function of forecast hour for the various model-based bias-corrected system-wide aggregate raw and bias-corrected raw power forecasts over the fall months (September, October and November 2011).

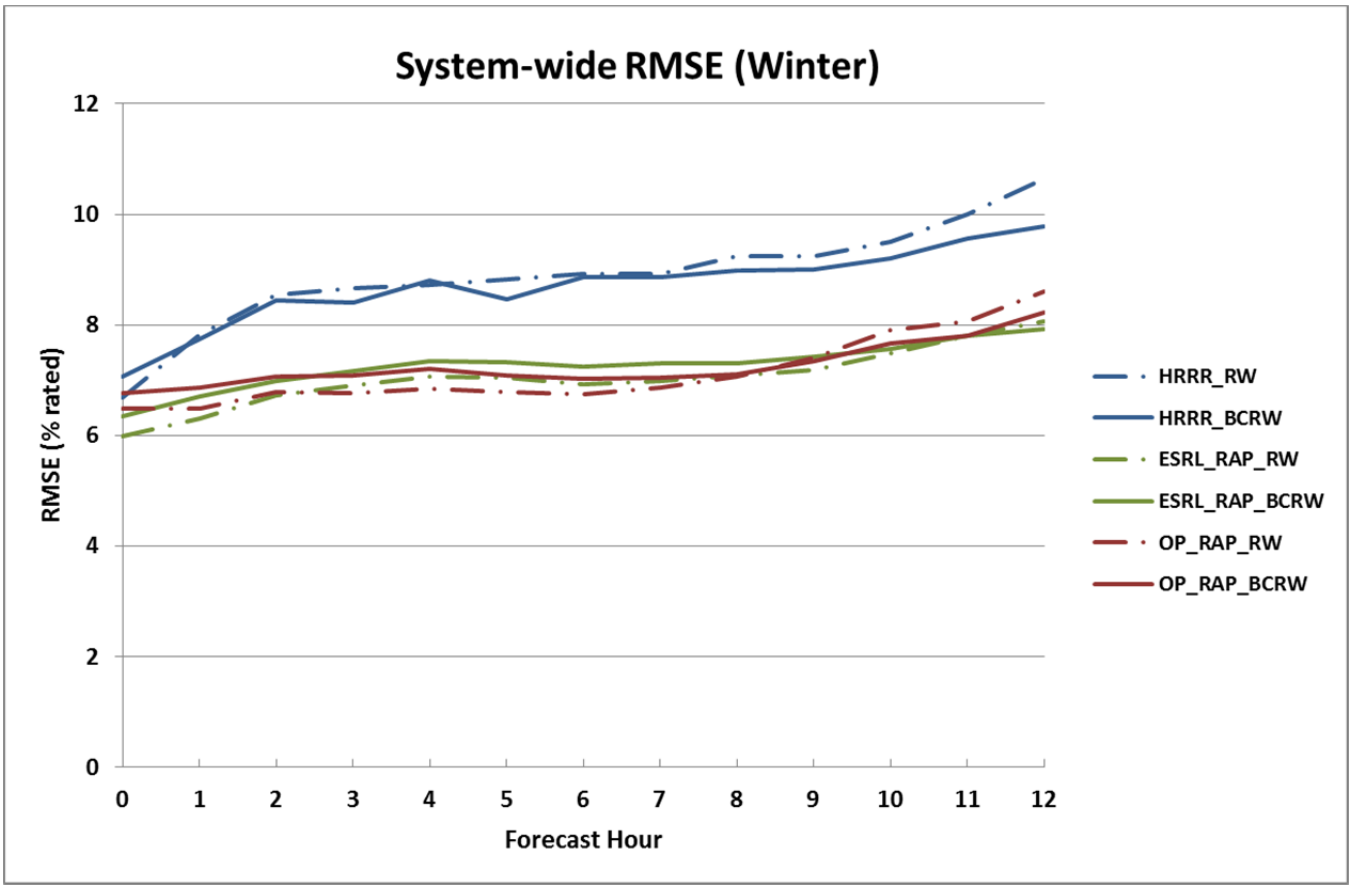

Figure 24: Root Mean Square Error (as a percent of rated capacity) as a function of forecast hour for the various model-based bias-corrected system-wide aggregate raw and bias-corrected raw power forecasts over the winter months (December 2011, January and February 2012).

4.3 Large Forecast Error Analysis

4.3.1. Definition of Large Forecast Errors

While overall error metrics are useful for gauging the quality of the forecasts, they often don't tell the complete story of forecast quality in the eyes of forecast users. This is particularly true for power system operations because reliability of the grid is of primary importance. Grid operators are concerned about the extreme events (the tails of the forecast error distribution), even if they occur only rarely, because that can influence the reserves and operating practices that they use to ensure reliability. These large forecast errors often occur as a result of actual or predicted "ramp events" when wind power production is changing rapidly. Ramp events will be specifically addressed in Section 4.4, but it is useful to start by looking at the occurrence of large forecast errors in a more general sense.

An analysis of the forecast errors at the tails of the forecast error distributions was done by calculating *the hourly forecast errors* over the period from September. 2011 – April 2012 for OP_RUC-based forecast comparisons, and from September 2011 – August 2012 for the current operational forecast model (OP_RAP) comparisons, and counting the number of forecast hours that had errors that equaled or exceeded some threshold value during the first six hours of the forecasts. At the wind plant level, threshold values of 50% or 80% of the wind plant capacity were chosen. At the system aggregate

level, thresholds of 15%, 20%, 23.2% (which corresponds to a 500 MW error) and 25% were chosen. A relative 'percent improvement' metric for measuring the differences in the large forecast errors (LFE) was calculated as follows:

$$\boldsymbol{LFE(Percent\ Improvement) = \frac{\#\ forecast\ 1 - \#\ forecast\ 2}{\#\ forecast\ 1}\ x\ 100}$$

Equation 5

where '# forecast 1' is the number of hourly forecast errors equal to or exceeding the magnitude of the threshold in the 'control' forecasts, and '# forecast 2' is the number of hourly forecast errors equal to or exceeding the magnitude of the threshold in the comparison forecasts. Note that '# forecast 1' and '# forecast 2' are *not* the number of ramp events, although the largest hourly forecast errors usually occur during ramp events. Wind plant events such as curtailment and icing, events that might create large forecast errors due to factors other than the accuracy of the wind forecasts, have been filtered out of the analysis dataset as was discussed in Section 3.

4.3.2 Results

Table 4 shows the results from the percent improvement of the large forecast errors calculated by comparing the bias-corrected raw HRRR, ESRL_RAP and OP_RAP-based power forecasts with the bias-corrected raw OP_RUC-based power forecasts for the time period from September 2011 – April 2012 for the individual wind plant locations. As can be seen in Table 4, the new operational model-based forecasts (OP_RAP) and the research model-based forecasts significantly reduce the number of hourly large forecast errors at most wind plant locations. The improvement seen at some locations exceeds 40%, although some care must be taking in interpreting the results in the 80% capacity threshold columns as there are significantly fewer samples of errors exceeding this threshold than the 50% threshold (calculations containing fewer than ten samples of either forecast are denoted with a '*'). The most consistent improvements (in terms of the number of locations with a positive improvement) are found in the ESRL_RAP_BCRW forecasts, where the forecasts reduce the numbers of large forecast errors greater than 50% of capacity at all wind plant locations, and 80% of capacity at nine of the thirteen wind plant locations. The current operational short-term forecast model (OP_RAP) and HRRR-based bias corrected raw forecasts also make a significant reduction in the number of large wind power forecast errors compared to the OP_RUC-based forecasts.

A similar comparison between the current operational model-based (OP_RAP) bias-corrected wind power forecasts and the research model-based forecasts (ESRL_RAP and HRRR) for the entire year of the forecasting experiment is shown in Table 5. The left-most two columns in Table 5 show the results of the large forecast error comparison between the bias-corrected raw OP_RAP and HRRR-based forecasts. The HRRR-based forecasts have fewer large forecast errors exceeding 50% (80%) of the plant rated capacity at eight (six) of the thirteen locations, showing a substantial improvement at some locations. Note that this is similar to what was found the OP_RUC-HRRR comparison shown in Table 4, although the percent improvement is somewhat larger in the OP_RUC-HRRR comparisons at most sites.

The comparison between the bias-corrected raw OP_RAP and the ESRL_RAP-based power forecasts are shown in the right-most two columns of Table 5. The ESRL_RAP-based forecasts reduce the number of large forecast errors at almost all sites compared to the OP_RAP-based forecasts. This is again similar to what was seen in the ESRL_RAP, OP_RUC-based forecast comparisons in Table 4, but the percent improvement in the ESRL_RAP-OP_RAP comparison is smaller at most locations

compared to the percent improvements seen in the ESRL_RAP-OP_RUC comparison. A similar comparison between the bias-corrected raw ESRL_RAP and the HRRR-based power forecasts is shown the middle two columns of Table 5. The higher resolution HRRR-based forecasts only reduce the large forecast errors compared to the coarser resolution ESRL_RAP-based forecasts at two of the thirteen sites. The possible reasons for this will be explored further in the more detailed ramp event analysis in Section 4.4.

SITES	OP_RUC-HRRR 50%	OP_RUC-HRRR 80%	OP_RUC-ESRL_RAP 50%	OP_RUC-ESRL_RAP 80%	OP_RUC-OP_RAP 50%	OP_RUC-OP_RAP 80%
Day County	-7.65	23.68	19.95	50.00	14.75	26.32
Edgeley Basin	28.93	39.25	15.29	11.21	9.50	26.17
Lake Benton 2	40.13	53.85*	41.45	61.54*	1.64	53.85*
Mower County	11.84	-50.00*	29.28	33.33*	18.42	-50.00*
South Dakota Basin	31.68	3.85	32.06	19.23	19.21	19.23
Story County 1	12.88	-35.29	27.36	35.29	21.33	47.06*
Wessington Springs	20.63	10.75	28.18	33.33	15.45	33.33
AGGREGATES						
Ashtabula	-28.27	0.00*	6.55	16.67*	-1.19	-83.33*
Crystal Lake/Hancock	20.44	0.00*	11.05	-25.00*	6.63	62.50*
Endeavor	24.25	-23.81	27.84	-52.38	1.50	-76.19
Langdon	-25.87	-242.86*	6.93	-171.43*	-5.08	-214.29*
Oliver	12.23	0.00	10.06	31.25	-17.75	-53.13
Wilton/Baldwin	28.89	10.00*	34.00	10.00*	29.11	-80.00

Table 4: Percent improvement in number of large hourly forecast errors (as calculated in Equation 5) for a 50% and 80% of plant capacity threshold. The two left-most columns show the comparison between the bias-corrected raw OP_RUC and HRRR forecasts (green shaded boxes indicate the HRRR has fewer large errors), the middle two columns show the comparison between the bias-corrected raw RUC and ESRL_RAP forecasts (pink shaded boxes indicate the ESRL_RAP has fewer large errors), and the two right-most columns are the comparison between the OP_RUC and the OP_RAP forecasts (yellow shaded boxes indicate the OP_RAP has fewer large errors). Values denoted with (*) indicated one or both of the comparison forecasts has less than 10 error samples that meet or exceed the threshold, so the values may not be statistically significant.

SITES	OP_RAP-HRRR 50%	OP_RAP-HRRR 80%	ESRL_RAP-HRRR 50%	ESRL_RAP-HRRR 80%	OP_RAP-ESRL_RAP 50%	OP_RAP-ESRL_RAP 80%
Day County	-13.49	2.21	-17.43	-23.15	3.35	20.59
Edgeley Basin	26.24	19.75	10.19	31.58	17.87	-17.28
Lake Benton 2	34.48	-44.44*	-2.56	-85.71*	36.11	22.22*
Mower County	-11.76	-52.78	-19.92	-77.42	6.80	13.89
South Dakota Basin	5.37	1.85	-14.00	-29.27	16.99	24.07
Story County 1	10.63	-1.85	-22.24	-120.00	26.89	53.70
Wessington Springs	2.59	-19.86	-7.23	-43.22	9.16	16.31
AGGREGATES						
Ashtabula	8.31	45.45	-16.79	-33.33	21.49	59.09
Crystal Lake/Hancock	2.15	-46.15	3.45	-11.76*	-1.34	-30.77
Endeavor	6.47	2.13	-12.97	-2.22	17.21	4.26
Langdon	-14.93	9.09	-30.38	-42.86	11.85	36.36
Oliver	20.36	-1.19	-3.12	-34.92	22.77	25.00
Wilton/Baldwin	-3.76	-40.00	-8.43	-44.83	4.31	3.33

Table 5: Percent improvement in number of large hourly forecast errors (as calculated in Equation 5) for a 50% and 80% of plant capacity threshold. The two left-most columns show the comparison between the bias-corrected raw OP_RAP and HRRR forecasts (green shaded boxes indicate the HRRR has fewer large errors), the middle two columns show the comparison between the bias-corrected raw ESRL_RAP and the HRRR forecasts (pink shaded boxes indicate the HRRR has fewer large errors), and the two right-most columns are the comparison between the OP_RAP and the ESRL_RAP forecasts (yellow shaded boxes indicate the ESRL_RAP has fewer large errors). Values denoted with (*) indicated one or both of the comparison forecasts has less than 10 error samples that meet or exceed the threshold, so the values may not be statistically significant.

Similar analyses were preformed between the various bias-corrected raw model-based forecasts at the system aggregate level. Recall that the definitions for what constitutes a 'large forecast error' changes between the individual site level and the system aggregate level due to the fact that geographic dispersion of the wind plants helps smooth out changes to both the aggregate power production and the aggregate power forecasts compared to the local plant level.

The comparisons between OP_RUC-based bias corrected raw power forecasts and the current operational model-based and research model-based forecasts are shown in Table 6. At the system aggregate level, both the coarser resolution ESRL_RAP and OP_RAP-based forecasts show a significant improvement in reducing large wind energy forecasts errors compared to the OP_RUC. Note also that the difference in the improvement seen between the OP_RUC-ESRL_RAP is larger for the larger forecast errors compared to the OP_RUC-OP_RAP comparisons. This result is consistent with the comparisons seen at the individual wind plant level shown in Table 4.

Note however that the bias corrected raw HRRR-based power forecasts do not show an improvement in reducing large forecast errors at the system aggregate level compared to the OP_RUC-based power forecasts, despite the fact that the HRRR-based forecasts had fewer large forecast errors at most of the individual wind plants (see Table 4). A similar result was found in the comparison of large forecast errors between the current OP_RAP_BCRW power forecasts and the ESRL_RAP_BCRW and

HRRR_BCRW power forecasts shown in Table 7. Once again the ESRL_RAP-based power forecasts show a significant reduction in the number of large forecast errors compared to the OP_RAP-based power forecasts, but the HRRR-based forecasts do not reduce the number of large forecast errors at the system aggregate level compared to the OP_RAP model-based forecasts, despite having done so at most of the individual wind plant locations (see Table 5). This result could be due to a higher correlation between the large forecast errors between various site forecasts in the HRRR-based forecasts compared to the coarser resolution model-based forecasts and will be explored further in the following sections.

error (% rated)	15	20	23.2	25
OP_RUC-HRRR	-2.984	-6.646	-22.656	-21.687
OP_RUC-ESRL_RAP	27.238	40.506	46.094	51.807
OP_RUC-OP_RAP	26.853	36.709	37.500	40.964

Table 6: Percent improvement in number of large hourly forecast errors (as calculated in Equation 5) for a 15%, 20%, 23.2% (500 MW) and 25% of the system-wide aggregate capacity threshold for the bias-corrected raw model-based forecasts from September 2011 – April 2012. Negative values indicate the OP_RUC-based bias corrected raw forecasts have fewer large forecast errors, positive values indicate the new operational model or research model-based forecasts have fewer large forecast errors.

error (% rated)	15	20	23.2	25
OP_RAP-HRRR	-6.202	-3.532	-3.879	-11.429
OP_RAP-ESRL_RAP	15.535	34.015	37.500	42.143

Table 7: Percent improvement in number of large hourly forecast errors (as calculated in Equation 5) for a 15%, 20%, 23.2% (500 MW) and 25% of the system-wide aggregate capacity threshold comparing the OP_RAP bias-corrected raw forecasts with the research model-based forecasts from September 2011 – August 2012. Negative values indicate the OP_RAP-based bias corrected raw forecasts have fewer large forecast errors, positive values indicate the research model-based forecasts have fewer large forecast errors.

Table 8 shows the percent improvement in the number of large forecast errors between the bias-corrected raw model-based power forecasts and the trained model-based forecasts for the OP_RAP, ESRL_RAP and HRRR forecasts for the period over which trained forecasts were generated (January 2012 – August 2012). As can be seen in Table 8, the training process greatly reduces the number of large forecast errors in all the system aggregate forecasts, particularly for the ESRL_RAP-based forecasts at the larger forecast errors. This is likely because the training process tends to create smoother forecasts than the raw power forecasts, which reduces the magnitude of the 'peaks' and 'valleys' in the forecasts where the largest forecast errors often occur. While this helps reduce large forecast errors in the overall system aggregate forecast (and may have impacts on the economics of the forecasts in power system operations for the many "normal" operating hours), the potential magnitude of the forecast extremes is one of the concerns during potential wind energy ramp events, which is why the trained forecasts may not be the forecast of choice when considering ramp event as was discussed previously. In other words, because the normal training objective is to reduce the average error, it may also smooth ramp events in the forecast.

error (% rated)	15	20	23.2	25
HRRR	7.89	28.80	23.19	34.83
ESRL_RAP	22.36	33.49	55.29	55.32
OP_RAP	19.45	32.85	39.29	34.85

Table 8: Percent improvement in number of large hourly forecast errors (as calculated in Equation 5) for a 15%, 20%, 23.2% (500 MW) and 25% of the system-wide aggregate capacity threshold comparing the bias-corrected raw model-based forecasts and trained forecasts for each model-based forecasts from January 2012 – August 2012. Negative values indicate the model-based raw bias corrected raw forecasts have fewer large forecast errors, positive values indicate the model-based trained forecasts have fewer large forecast errors.

It's clear from the above analysis that the ESRL_RAP-based power forecasts are producing fewer large forecast errors than all other model-based forecasts at both the local site and system aggregate levels when only the hourly forecast errors are considered. However, hourly forecast errors alone don't take into account things such as the potential timing of the passage of a particular weather feature (although the feature is still in the forecast). Wind ramp timing errors and other forecast errors of interest to system operators will be addressed and evaluated in a more thorough ramp event analysis which follows below.

4.4 Ramp Event Error Analysis

4.4.1 General Discussion and Significance to System Operations

Very large forecast errors often occur when the wind power production is changing rapidly (i.e., wind energy 'ramp events'). These errors can create challenges for electric grid system operations, especially if the event is unanticipated**.** Not only are the timing and magnitude of wind energy ramp events of concern to system operations, but the ramp rate is also of concern since there are limits on how quickly conventional types of generation can be ramped up and down.

Our ability to predict ramp events at various time horizons greatly depends on what meteorological feature is causing the ramp. Generally speaking, the larger and longer-lived a particular weather feature, the better it can be predicted in the 2-48 hour timeframe. Wind energy ramp events are a forecasting challenge since there is the potential for rapid changes in wind speed at a given location from a wide variety of meteorological phenomenon, or even from complex combinations of such phenomena.

The main weather drivers of up-ramp events in the northern Great Plains include:

- Cold frontal passage – The strongest winds tend to be behind the front and can persist for many hours following frontal passage. This can create a large, sudden wind ramp at any given wind plant location in some cases, and can lead to a longer duration wind ramp at the system level as the front sweeps across the operating area.

- Thunderstorms and thunderstorm outflow – The extent to which this will create a significant system-wide ramp will depend on the size of the thunderstorm complex and the geographical dispersion of the wind farms. These events often create ramp events of short duration but very rapid rate at the wind plant level, and are probably the most difficult to accurately predict of the all the weather events that can create wind energy ramps.

- Rapid intensification of an area of low pressure – These are larger-scale features that are usually forecast pretty well within 12-24 hours of occurrence, although the exact timing of these events is still a challenge to predict accurately. These events create wind energy ramps both at the local wind plant level, and at the system level.

- Local topographic flows – Although the Northern Plains is not an area of large topographic variation, there are some topographic features (such as the Buffalo Ridge in southwest Minnesota, and rolling terrain in southeast Minnesota and northeast Iowa) that can create local topographic flows such as standing gravity waves that can enhance winds under certain atmospheric conditions.

- Vertical mixing processes – Vertical mixing will erode the stable near-surface boundary layer in the morning (sometime in the few hours after sunrise) often leading to increased wind speeds at hub height. The stabilization of the near-surface boundary layer in the evening/nighttime hours can determine how high above the surface the nocturnal low-level jet develops. The extent to which these processes occur depends on what type of land use/snow is covering the surface, the amount of clouds, and the strength of the winds in the lowest 1-2 km of the atmosphere. These processes occur daily, and may or may not produce wind changes that qualify as a 'ramp event' at the local plant level depending on the broader atmospheric conditions.

Likewise, the weather events/situations that can cause down-ramp events in the northern Great Plains include:

- Near-surface boundary layer stabilization at sunset/nightfall – Again, the extent to which this occurs depends on what type of land use/snow is covering the surface, the amount of clouds, and the strength of the winds in the lowest 1-2 km of the atmosphere.

- Relaxation of pressure gradient as high pressure moves in following cold front passage, or as an area of low pressure moves away – This can create down-ramp events at both the local plant and system aggregate levels.

- Pressure changes following the passage of thunderstorm complexes – These will largely affect the local wind plant level, and are very difficult to predict accurately.

- A decrease in wind speed as a warm front passes through – Warm fronts tend to be very slow moving, and the winds right along the front tend to be weaker than the winds both north and south of the front. This can create a down-ramp/up-ramp event at the local wind plant level as the front passes.

Down-ramps in general may be more difficult to forecast because they are usually not directly associated with sharply defined meteorological features (thunderstorm complexes may be the exception) – although little work has been done on the determining the accuracy of up-ramp predictions versus down-ramp predictions to date.

4.4.2 Definitions and Metrics

4.4.2.1 Ramp Event Definition

The definition of a wind energy ramp event will vary from one operating area to another depending on the penetration of wind on the system and the other types of generation available. When the Midwest Independent System Operator (MISO) was asked about the set of conditions that they would use to define a wind energy ramp event, they said they currently didn't have any since they have not seen a change in wind generation to date that they had difficulty handling (although they were quick to point out that this could change as the wind penetration increases further on the system). MISO has a very large operating area with the wind generation dispersed throughout the region which helps smooth out system-wide wind ramps. (The effects of geographic dispersion on system-wide wind energy ramps will be discussed further in Section 4.5 below.) They also have many different types of conventional generation available and a well-designed energy market which makes the system quite flexible. Since any definition of a wind energy ramp event for the MISO system is currently fairly arbitrary and the definition of a ramp event will vary by forecast user, a more general approach is to consider a 'suite' of ramp definitions. This approach will potentially benefit more forecast users and is more informative from a forecast performance perspective.

In the analysis to follow, we have defined a suite of ramp definitions as indicated below:

- The power changes X% (of rated capacity) over a Y hour period (or less).
- The event can be longer than Y hours if d(power)/dt $\geq$ (X% capacity)/(Y hours) occurs at some point during the event.
- We consider the entire period during which power is changing in a (nearly) continuous way to be an event.

Here 'X' and 'Y' are allowed to vary. Reasonable ranges of X and Y were chosen from scatter plots of observed ramp amplitude versus ramp duration for a site (or the entire system) where all ramps that equaled or exceeded 15% of the rated capacity were considered, regardless of the duration. An example scatter plot of the system-wide ramps found in the forecast observational time series (i.e. the aggregate of all the NextEra wind plants in the study area) is shown in Figure 25.

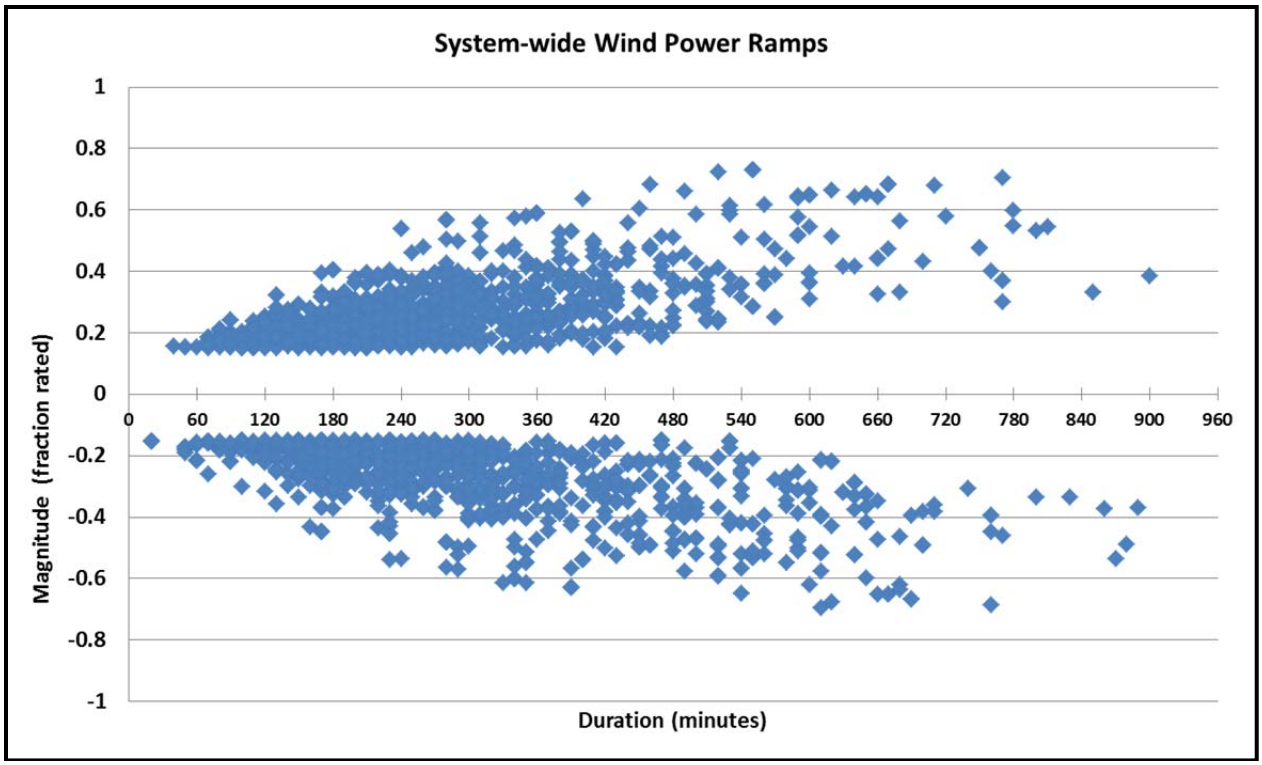

Figure 25: Scatter plot of observed ramp magnitude vs. ramp duration from the system-wide aggregate forecast time series. Ramp events with a magnitude of 15% (or larger) of the system-wide rated capacity are plotted.

4.4.2.2 Description of Ramp Detection Algorithm

WindLogics has developed an algorithm to detect wind energy ramps in both forecasts and actual plant power using 10-minute observed power values and forecasts interpolated to 10-minute intervals. At first glance this may seem like an unfair comparison since forecast power values (based on the weather forecast model output) are only available every hour. However, system operators must balance the grid system on a minute-to-minute bases, so the 10-minute interval used here is already relatively crude in this context.

The algorithm uses the 10-minute power values to calculate a 10-minute power time derivative, and then strings together consecutive time intervals during which the derivative is of the same sign. An iterative process is then used to filter out small/short-term power fluctuations that may be embedded in larger potential ramp events by evaluating nearby like-sign derivative periods and combining them into one longer 'like signed' period if user-defined thresholds are met. Once changes are no longer made to the start/end points of the 'like signed' derivative periods, the changes in power and time intervals over which is the power is changing in a (nearly) continuous way are calculated. The start/end points of potential ramp events are further refined by calculating the average time rate of change over the entire potential ramp event, and then removing any points at the beginning/end of the potential ramp period where the local power derivative is less than a threshold percentage of the overall ramp rate. This eliminates time periods at the beginning/end of the potential ramp events where the power is still changing in continuous way, but at a much smaller rate than the central portion of the event. Only those events that meet user defined criteria for a 'ramp event' are kept for further analysis. This ramp detection algorithm is very flexible and easily allows evaluation of different user-defined ramp definitions.

Examples of the ramps events detected by the algorithm are shown in Figure 26 and Figure 27. The 'top hat' line structures on these figures indicate where the ramp detection algorithm is indicating a ramp based on user-defined criteria. The height of the 'top hat' indicates the magnitude of the ramp, and the width of the 'top hat' indicates the duration of the ramp. There is an up-ramp in the observations shown in Figure 26, and the forecast does a good job of capturing it in the 04Z and 06Z model runs. However, the 05Z forecast run has an upward trend in the forecast power during the observed ramp period which does not meet the user-defined ramp criterion. There is also a ramp forecasted in the 04Z run which does not occur in the observations. Figure 27 shows a similar forecast series for a down ramp that is not well forecasted.

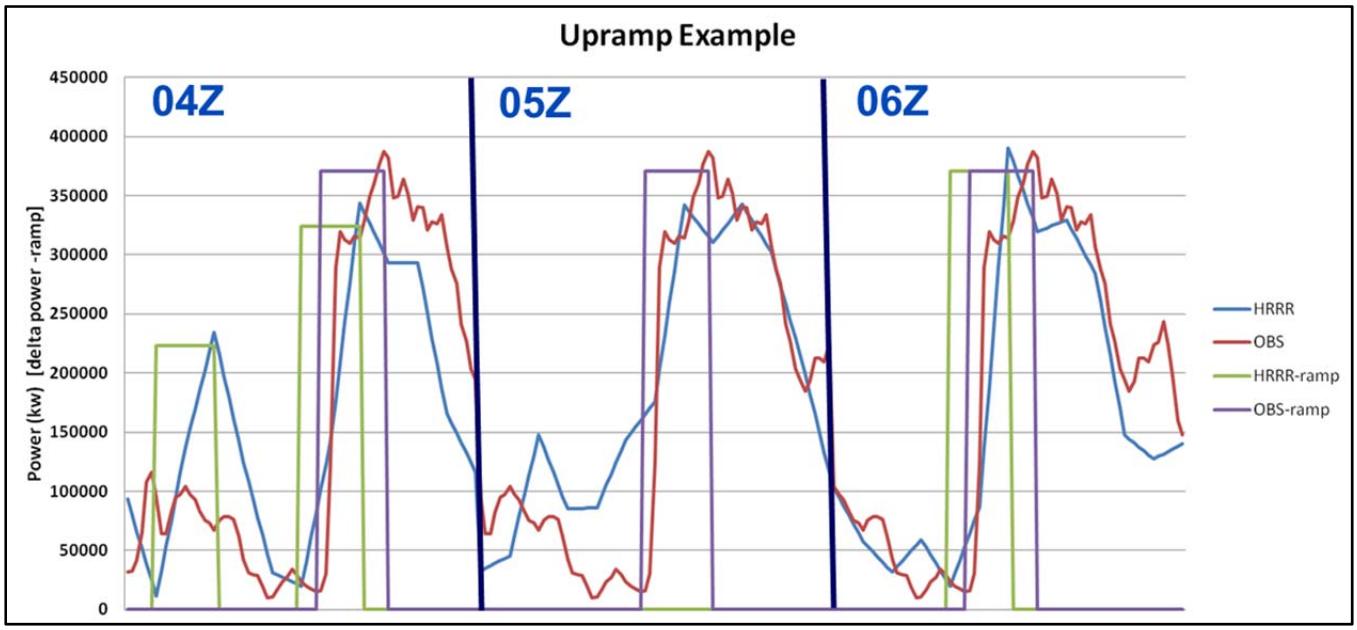

Figure 26: Three consecutive bias-corrected raw HRRR-based power forecasts (0-12 hours) on 6 May 2012 at a wind plant in North Dakota showing an 'up-ramp' event. The observed power is in red, the forecast power is in blue. The green (purple) 'top hat' line structures indicate the magnitude and duration of the forecast (observed) ramp events as detected by the ramp algorithm.

.

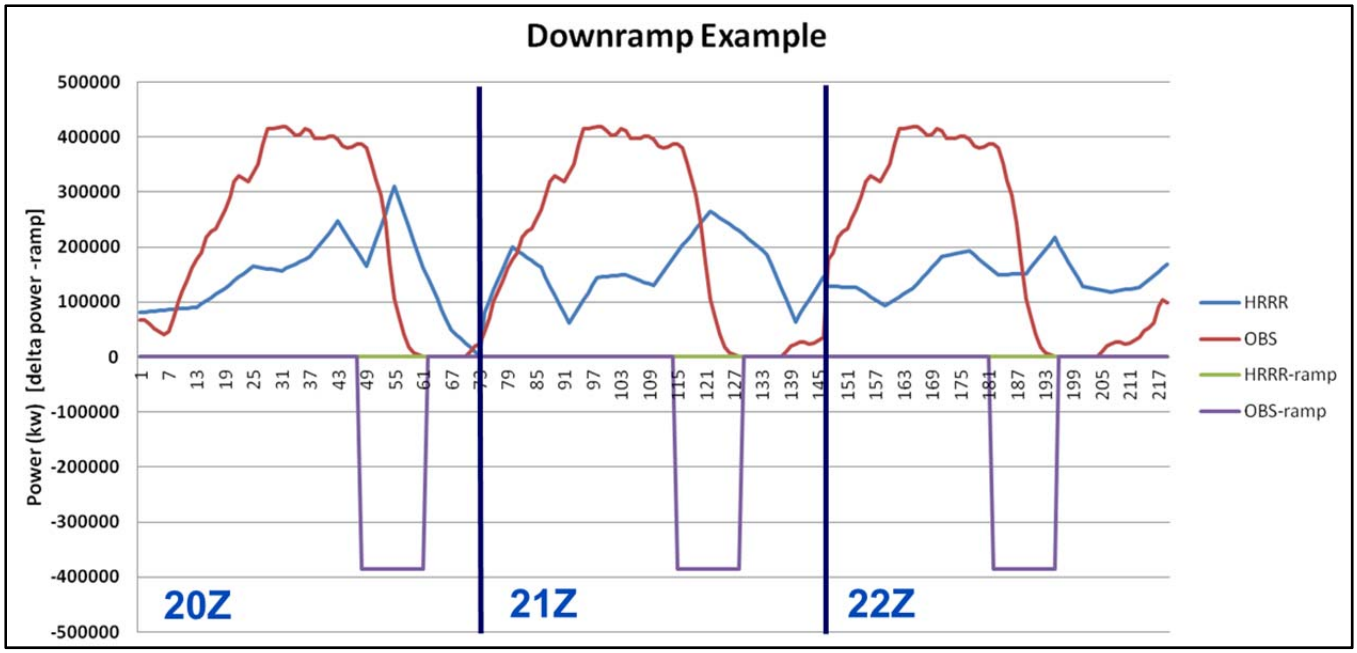

Figure 27: Three consecutive bias-corrected raw HRRR-based power forecasts (0-12 hours) on 27 May 2012 at a wind plant in North Dakota showing a 'down-ramp' event. The observed power is in red, the forecast power is in blue. The green (purple) 'top hat' line structures indicate the magnitude and duration of the forecast (observed) ramp events as detected by the ramp algorithm.

4.4.2.3 Metric Definitions

When calculating error metrics for discreet events (such as wind energy ramps), one is required to classify whether or not a forecast was successful. To this end, forecast ramp events are stratified into three different classifications: "hits" (the forecast successfully forecast an observed event), "misses" (the forecast did not successfully forecast an observed event), and "false alarms" (the forecasted event was not observed). For the wind energy ramp event metrics and statistics shown below, we are using the following definitions:

- *Hit* :
    - the midpoint of the predicted ramp is within a narrow window (± 2 hours) of the midpoint of the observed ramp
    - the magnitude of the predicted ramp is within ± 50% of observed magnitude (and of correct sign)
- *Miss*:
    - there is no predicted ramp in an extended window (± 4 hours) of the midpoint of an observed ramp
    - the midpoint of the predicted ramp is inside the extended window of the midpoint of the observed ramp, but outside the narrow window

    - the midpoint of the predicted ramp is within the narrow window of the midpoint of the observed ramp, but does not meet the criteria for a 'hit' (wrong magnitude or sign)
    - 'wrong sign' is a special case of a miss – the sign of predicted ramp is opposite that of the observed ramp within extended window
- *False Alarm:*
    - there is a predicted ramp, but no observed ramp within the extended window of the midpoint of the predicted ramp

Note that the requirements for a forecast ramp event to be considered a 'hit' are quite stringent, and that these requirements will likely also vary depending on the forecast user.

To make comparisons between the various model-based forecasts, only the first 15 hours of the forecasts for the ESRL_RAP and OP_RAP were used to make the forecast periods consistent with the HRRR. Ramp timing was defined by the mid-point of the detected ramp. By using time windows to make the determination about forecast ramp classification, this precludes in the inclusion of some of the forecast hours in the analysis. Observed ramps detected during the first two forecast hours (0,1) and last 4 forecast hours (12-15) are excluded from the analysis that follows to allow for the time windows defined above.

One of the most fundamental error metrics for event-based forecasts is the frequency bias. Frequency bias is defined as:

$$\boldsymbol{Frequncy\ Bias} = \frac{\#\ \boldsymbol{forecasted\ events}}{\#\ \boldsymbol{observered\ events}}$$

Equation 6

Frequency bias can vary between zero and infinity, with values greater than 1 indicating that the forecasts over-predict the number of events, and numbers less than 1 indicating that the forecasts under-predict the number of events.

Using the above definitions, we have calculated three additional ramp event error metrics. The first is the Critical Success Index (CSI – also sometimes called the Threat Score) which is the number of forecast events that are considered a successful forecast compared to all classified events. The value of CSI varies from 0 to 1 (with 1 being a perfect forecast), and is often considered a more balanced scoring metric in that it takes misses into account, as well as hits and false alarms. Critical Success Index is defined as:

$$\boldsymbol{Critical\ Success\ Index}\ (\boldsymbol{CSI}) = \frac{\#\ \boldsymbol{hits}}{(\#\ \boldsymbol{hits} + \#\ \boldsymbol{misses} + \#\ \boldsymbol{false\ alarms})}$$

Equation 7

The second error metric is the Recall, which is the fraction of all observed events that were successfully forecasted:

$$\boldsymbol{Recall} = \frac{\boldsymbol{\#\ hits}}{\boldsymbol{\#\ observed\ events}}$$

Equation 8

The third error metric is the False Alarm Ratio, which is the fraction of forecasted events that fall into the False Alarm category:

$$\boldsymbol{False\ Alarm\ Ratio\ (FAR)} = \frac{\boldsymbol{\#\ false\ alarms}}{\boldsymbol{\#\ forecasted\ events}}$$

Equation 9

In addition to the error metrics above, ramp forecast errors based on the ramp event timing, magnitude, duration and ramp rate were calculated. All additional forecast error statistics were calculated by subtracting the actual value from the forecast value. The ramp event timing was calculated based on the time of the center point of the ramp. For example, positive timing errors indicate that the forecast ramp event occurred *later* than the actual event (i.e. the time of the forecast event was greater). Both the average absolute timing errors (a measure the average timing error, whether positive or negative) and average timing error (or timing bias) were calculated. Average ramp rate errors were calculated by first taking the absolute value of the ramp rate and then subtracting (adding) the observed value from the forecast value if the ramps were of the same (opposite) sign, so negative ramp rate errors indicate that the forecast ramp rate was less than the actual ramp rate regardless of the sign of the ramp event.

4.4.3 Results

The error metrics and ramp event error statistics discussed in the previous section were calculated for each wind plant and for the entire system aggregate for the bias-corrected raw model-based power forecasts over the entire one year forecasting period, and were compared for relative accuracy. The results are discussed in more detail below.

4.4.3.1 System Aggregate

The frequency bias (as defined in Equation 6) for the system-wide forecasts over a subset of ramp definitions is shown in Figure 28. Included in this analysis are the bias-corrected raw power forecasts based on the HRRR, ESRL_RAP (the research Rapid Refresh), and OP_RAP (the operational Rapid Refresh) forecast models over the entire year of the project. As can be seen in Figure 28, the bias-corrected raw ESRL_RAP-based power forecasts most accurately predict the total number of ramp events. (A Frequency Bias value of 1 indicates that the model predicts the same number of the events as was observed.) At the system level, the HRRR-based bias-corrected raw forecasts tend to over-predict the number of ramp events by approximately 9%, and the OP_RAP-based bias-corrected raw forecasts tend to under-predict the number of events by approximately 10%.

A similar analysis considering only the up-ramp events and down-ramp events is shown in Figures Figure 29 and Figure 30 respectively. For the up-ramp events, the HRRR-based bias-corrected raw forecasts most accurately capture the number of up-ramp events (over-predicts by approximately 4%), while the coarser resolution model-based bias-corrected raw power forecasts both under-predict the number of system-wide up-ramp events by approximately 5% (ESRL_RAP) and 13% (OP_RAP). For down-ramp events, both research model-based power forecasts over-predict the number of down-ramp

events (approximately 17% for HRRR-based forecasts, approximately 9.5% for ESRL_RAP-based forecasts) while the operational model-based bias-corrected raw power forecasts under-predict the number of down-ramps by approximately 6%.

System-wide ramps in the MISO operating area will be driven by larger weather systems that impact many of the wind plants within a relatively short period of time. This means that system-wide ramps will most often be caused by the development, propagation and decay of synoptic-scale weather systems (i.e. high and low pressure systems) and associated frontal features as was discussed above. It's possible that weakening and propagation of these systems tends to occur too quickly in the HRRR forecasts (and too slowly in the OP_RAP forecasts), but this requires further investigation.

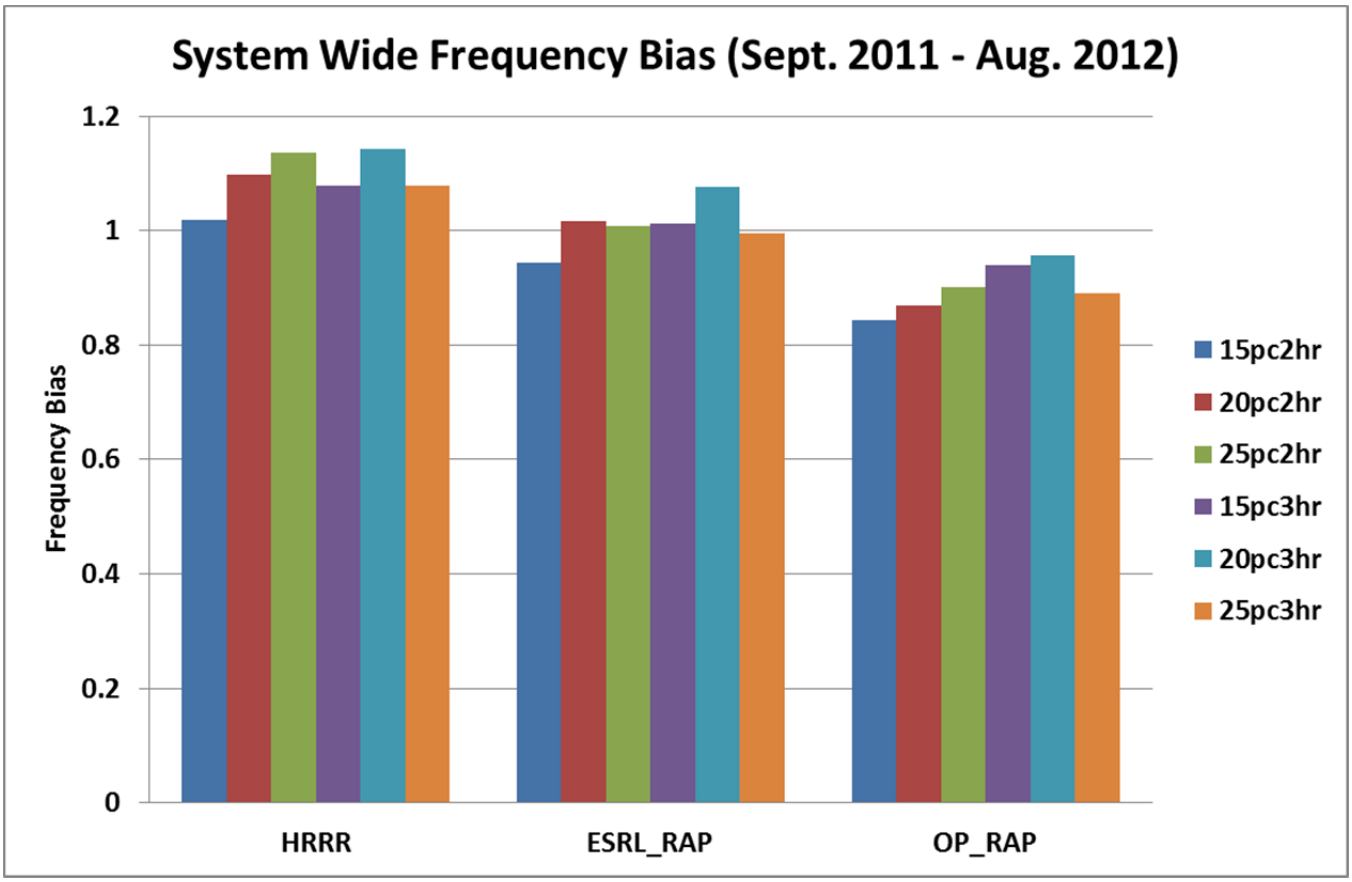

Figure 28: System-wide Frequency Bias for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 6 different ramp definitions as defined above.

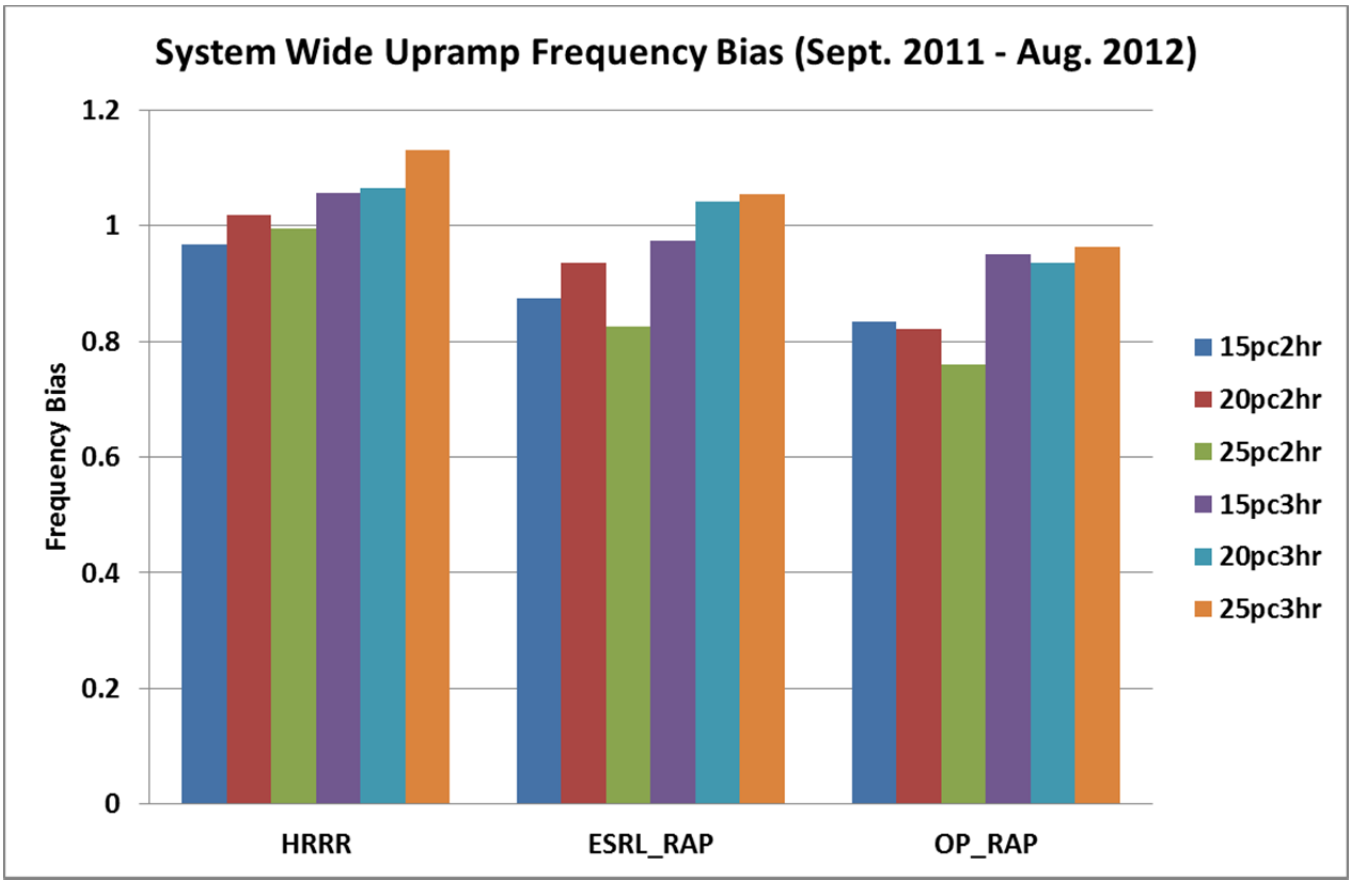

Figure 29: System-wide Frequency Bias for all wind power up-ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 6 different ramp definitions as defined above.

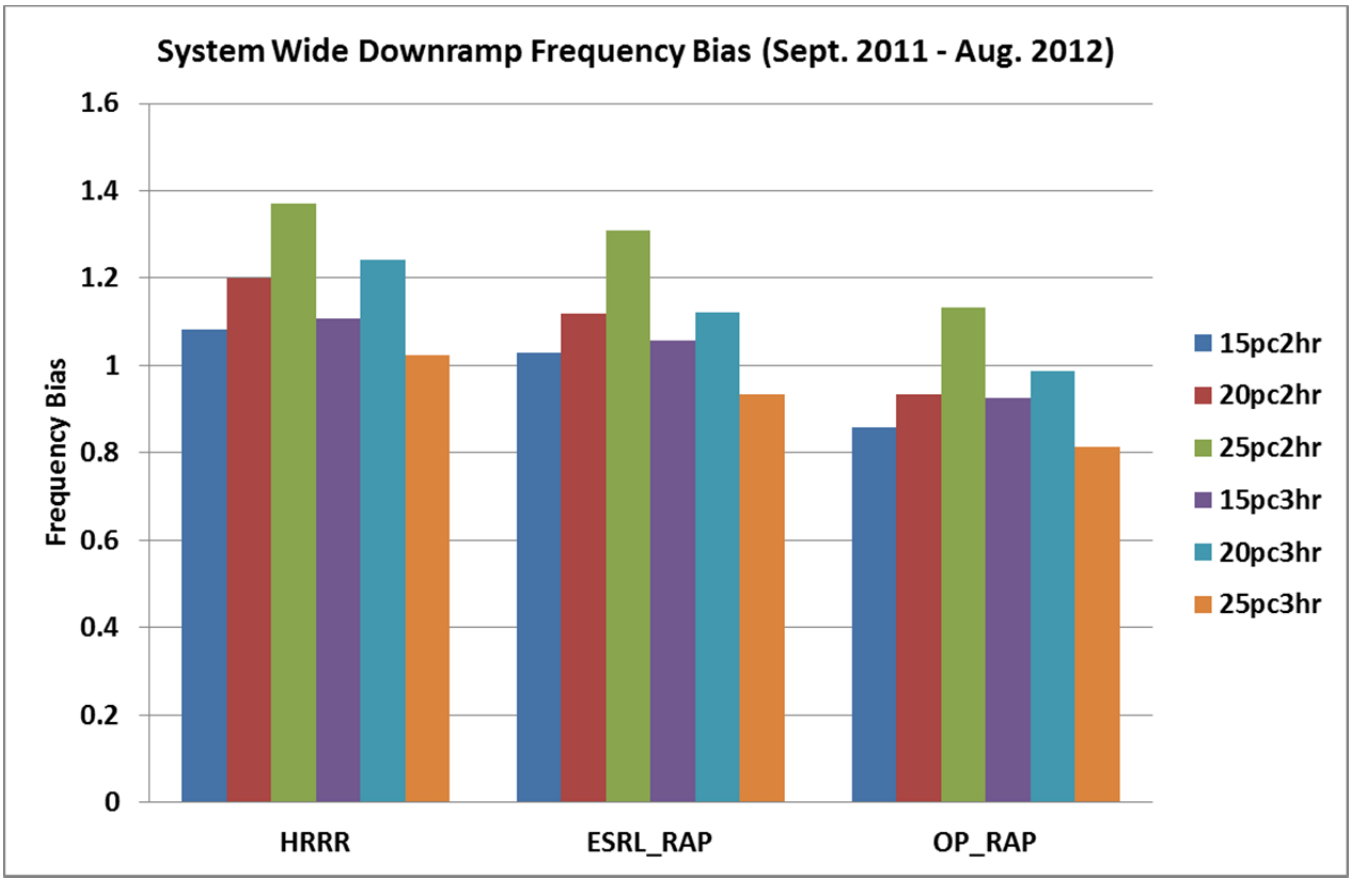

Figure 30: System-wide Frequency Bias for all wind power down-ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 6 different ramp definitions as defined above.

Figure 31 shows the system-wide Recall (as defined in Equation 8) for the ESRL_RAP-based bias-corrected raw power forecasts for the entire suite of ramp definitions tested. The trends and patterns in the system-wide Recall values with respect to the suite of ramp definitions is similar for all the bias-corrected raw model-based power forecasts, so only one example is shown. It should be noted that at the system level, no system-wide ramp events were observed for ramp definitions of 25-50 % rated capacity over a duration of 1 hour, or 50% rated capacity over a 2-hour period (the Recall for these events is zero because no events occurred). There were also very few events observed for definitions of 40% rated capacity over a 2-hour period, and 50% over a 3-hour period, so the Recall values for these ramp definitions should be viewed with some caution. As can be seen in Figure 31, smaller amplitude, longer duration events are predicted more accurately by the forecasts than shorter duration, higher amplitude events with a maximum Recall value of 0.55 occurring for a ramp definition of 20% rated capacity over a 2-hour period. Recall values drop off slowly at first from the peak value for nearby ramp definitions, and then drops off more quickly for shorter duration, higher amplitude events, likely because rapidly evolving weather features and more extreme events are more difficult to accurately predict.

A comparison of the Recall values for all three bias-corrected raw model-based power forecasts for a subset of ramp event definitions is shown in Figure 32. At the system level, both research model-based forecasts more accurately predict ramp events than the operational model-based forecasts, with the ESRL_RAP-based forecasts more accurately predicting ramp events than the HRRR and OP_RAP-

based forecasts for most ramp definitions by approximately 1% and 4% respectively. The only exceptions are for the larger magnitude, shorter duration events that are more accurately predicted by the bias-corrected raw HRRR forecasts as shown in Figure 33, although caution is warranted in these cases as there are usually relatively few events for these ramp definitions compared to other definitions. These larger magnitude, shorter duration events are more difficult for system operators to handle – especially if the event is unanticipated – as these events have the largest ramp rates making it more challenging to balance the system. For the range of ramp definitions shown in Figure 32, the ability of the forecasts to accurately predict ramp events appears to be more sensitive to ramp duration rather than ramp magnitude.

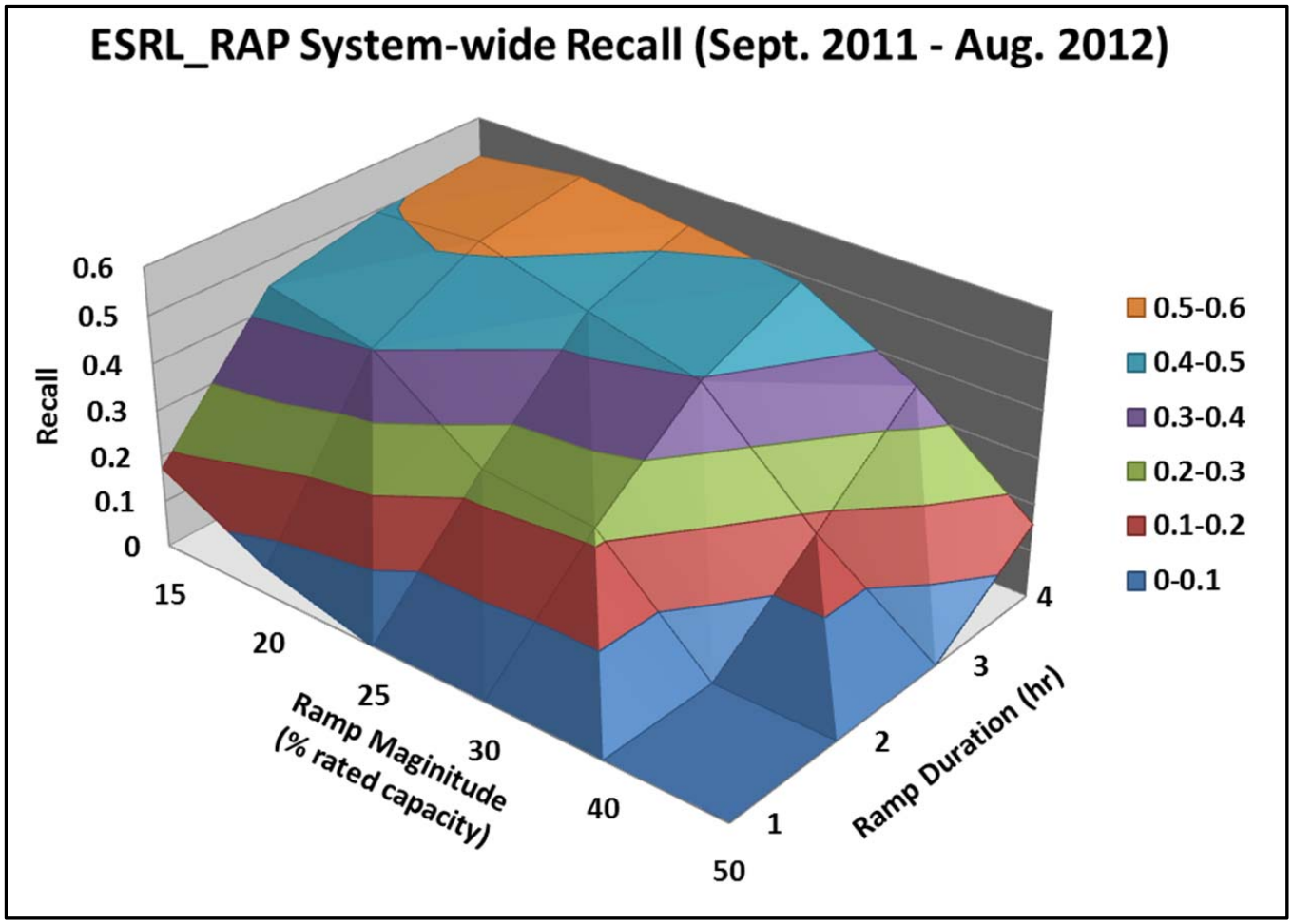

Figure 31: System-wide Recall for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw ESRL_RAP -based power forecasts for a suite of ramp definitions as defined in the text. The Recall values are zero for ramp definitions of 25-50 % rated capacity over a duration of 1 hour, and 50% rated capacity over a 2-hour period because no system-wide ramp events were observed for these definitions.

Similar analyses considering up-ramp and down-ramp events separately are shown in Figure 34 and Figure 35 respectively. Up-ramp events are more accurately predicted by the HRRR-based bias-corrected raw forecast at the system level, although the ESRL_RAP-based forecasts are a very close second. Both experimental model-based forecasts show an improvement of 4-5% on average over the operational model-based forecasts at accurately predicting up-ramp events the system level. All forecasts more accurately predict up-ramp events versus down-ramp events as can be seen by

comparing Figure 34 and Figure 35. Note that there is also less spread in Recall values for the various ramp definitions for the down-ramp vs. up-ramp events. The ESRL_RAP-based forecasts most accurately predict down-ramp events at the system level, with an approximately 2.8% improvement over the HRRR-based forecasts and approximately 4% improvement over the OP_RAP-based forecasts over the year-long forecast period.

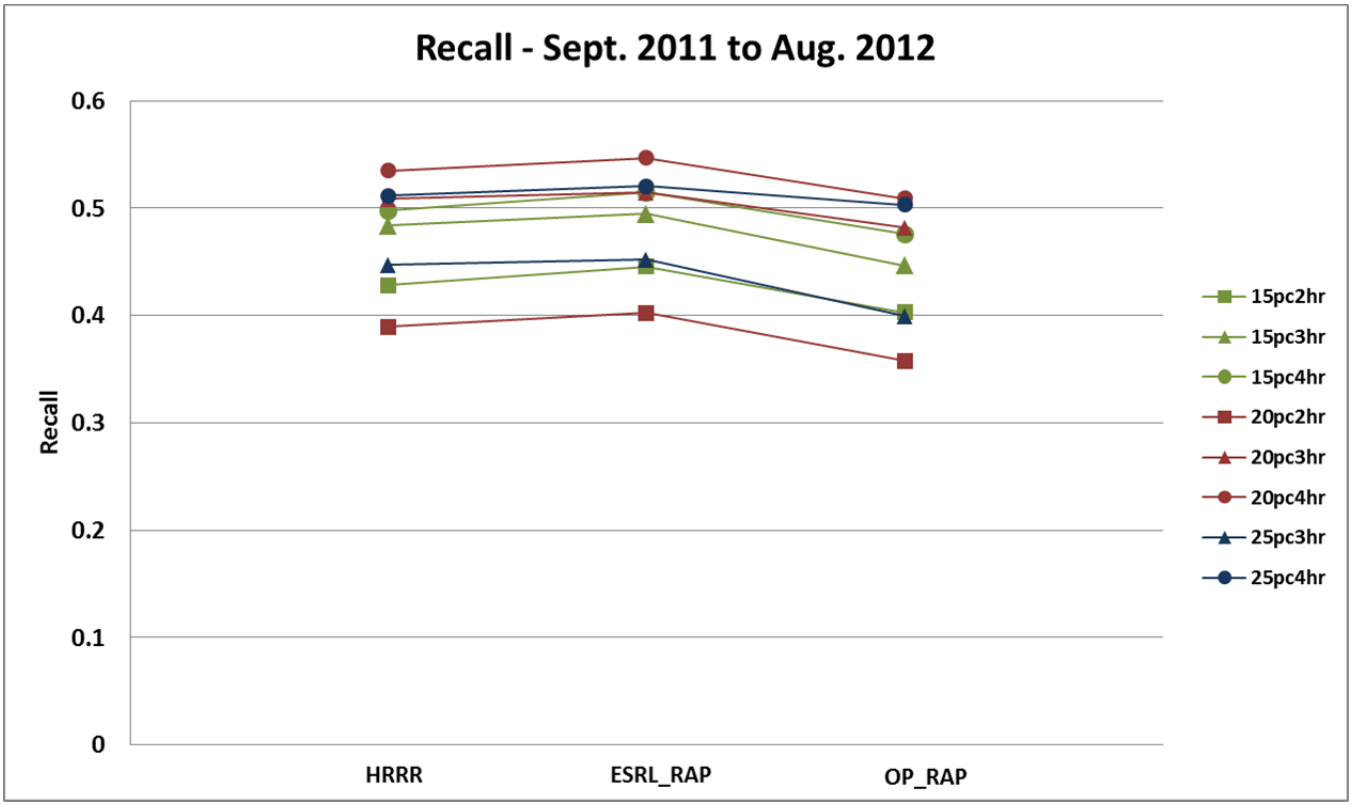

Figure 32: Recall for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

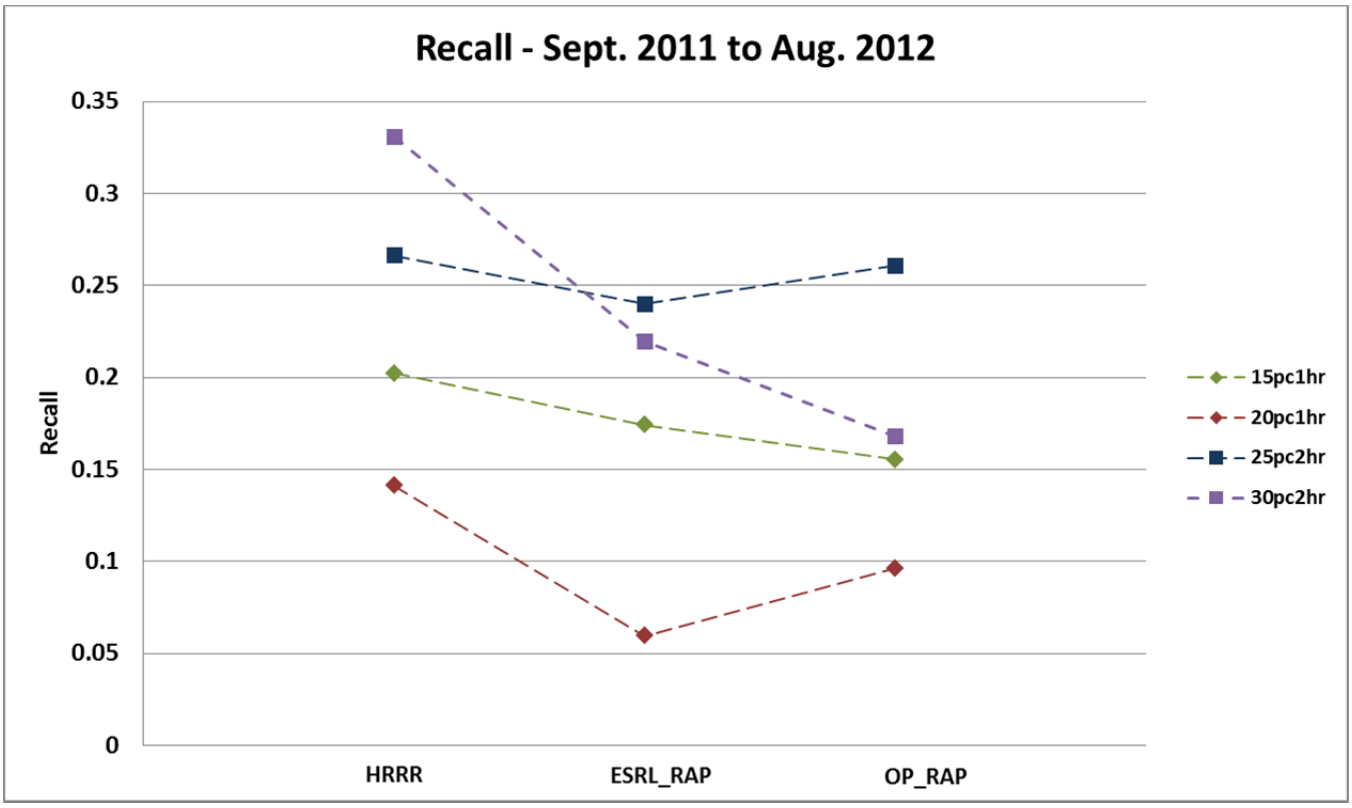

Figure 33: Recall for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 4 different shot duration ramp definitions as defined in the text. The dotted lines denote that fewer than 100 samples were available for one or more of the model-based bias-corrected power forecasts.

Figure 36 shows the system-wide Critical Success Index (CSI) scores for the aggregate HRRR, ESRL_RAP, and OP_ RAP-based bias-corrected raw power forecasts for the entire one-year forecast period for subset of the suite of ramp definitions. The ESRL_RAP-based bias-corrected raw forecasts have the highest Critical Success Index scores for most ramp definitions, likely because it has the highest Recall scores and fewer total misses than the HRRR and OP_RAP forecasts. Overall for all the model-based power forecasts, the CSI scores tend to be higher for the smaller amplitude, longer duration events where the Recall scores are higher and False Alarm Ratios are lower.

The role of misses in the CSI also cannot be overlooked. For larger magnitude events, as many as 86% (70-80% values were typical) of all misses were classified as 'clean misses', meaning that a forecast ramp was either outside the longer window of $\pm$ 4 hours of the center point of an observed ramp, or the forecast ramp (if present) did not meet the minimum criteria based on the ramp definition above. For smaller magnitude events, the percentage of 'clean misses' was typically in the 50-60% range. The HRRR-based bias-corrected raw forecasts had lower percentages of 'clean misses' than the coarser resolution model-based forecasts for all ramp definitions, but the HRRR-based forecasts also tended to over-predicted the number of events leading to larger false alarm ratios and resulting in the lowest CSI scores.

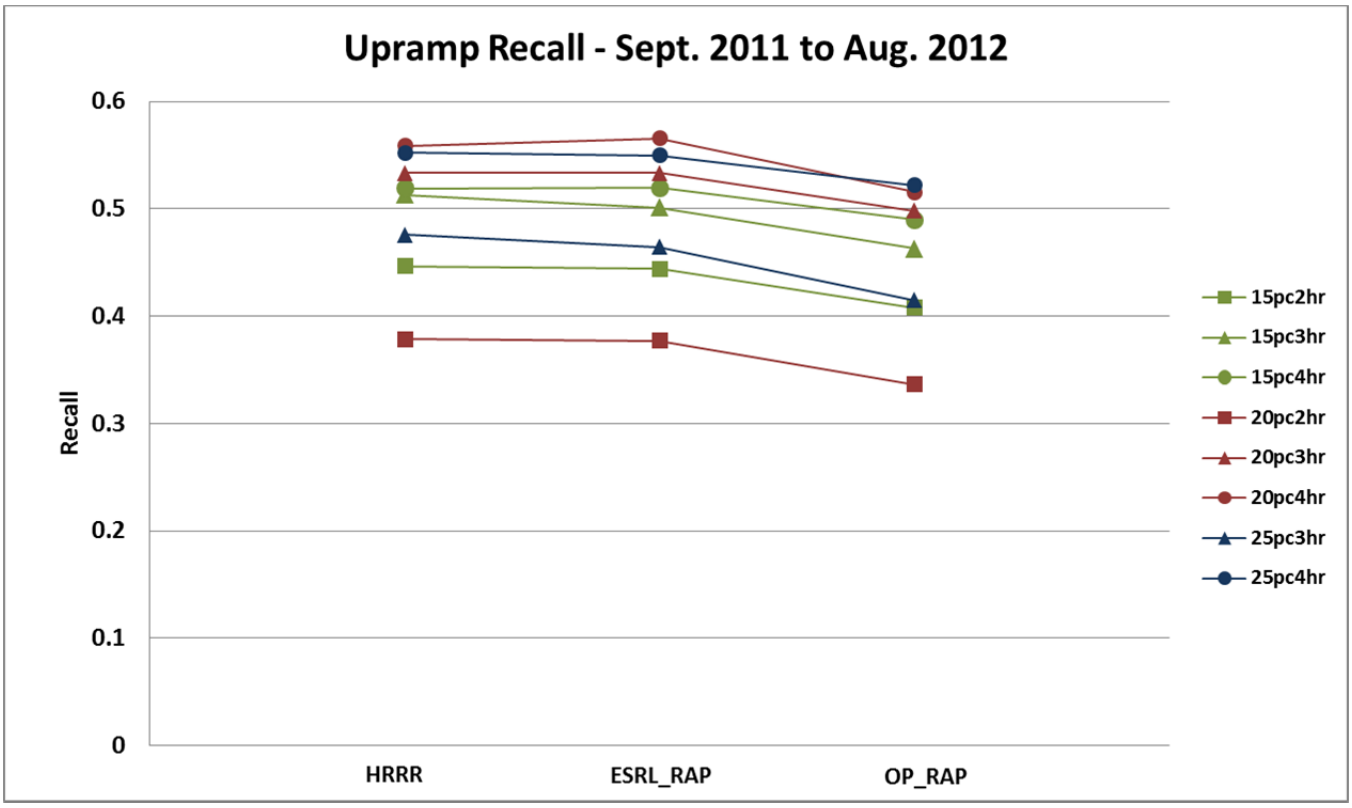

Figure 34: Recall for all system-wide wind power up-ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

The system-wide False Alarm Ratios (FAR) for the three model-based bias-corrected raw forecasts are shown in Figure 37, and for up-ramps and down-ramps broken out separately in Figure 38 and Figure 39, respectively. The HRRR-based bias-corrected raw forecasts have the highest overall FAR for most ramp definitions, while the OP_RAP forecasts have the lowest overall False Alarm Ratios. This is likely a result of the HRRR-based forecasts over-prediction of the number of ramp events, and the under-prediction of the number of ramp events by the OP_RAP-based forecasts. The overall FAR values appear to be more sensitive to ramp magnitude, with higher magnitude events tending to have higher FAR scores for all forecasts. For any given ramp magnitude, the shorter-duration events have higher FAR scores than longer-duration events. These general patterns hold true for both up-ramp and down-ramp events, although there is less variation in the FAR scores for up-ramp events (Figure 38) as compared to down-ramp events (Figure 39). The FAR for down-ramp events is also higher than up-ramp events for all forecasts over most of the ramp definitions.

Timing, magnitude, duration and ramp rate errors were also calculated for all system-wide ramp events classified as 'hits', and 'misses' where the 'misses' where such that errors could be calculated (i.e. those misses occurring within the extended window). Duration and magnitude errors are best expressed in terms of average ramp rate errors. The rate of ramping is of concern to system operators in that there are limits as to how quickly conventional generation can be ramped up (or down) to offset the changes in wind generation in order to keep the system balanced. While both up-ramp and down-ramp rates in wind generation are of a concern, down-ramp events are a larger potential issue for system operators as the only recourse for offsetting a down-ramp in wind energy is to ramp-up conventional generation (or shed load if that option is available). In the event of wind energy up-ramp,

system operators also have the option of curtailing the wind production in order to keep the system balanced.

The system-wide average ramp rate errors for all events classified as ‘hits’ are shown in Figure 40. Negative values indicate that the forecast ramp rate is smaller than observed (for both up-ramp and down-ramp events). As can be seen in Figure 40, all forecasts under-predict the ramp rate for all ramp definitions, but the HRRR-based bias-corrected raw forecasts have significantly smaller ramp rate errors (50% to 60% less for most ramp definitions) than the coarser resolution model-based bias-corrected raw forecasts. This is likely because the HRRR-based forecasts significantly outperformed the coarser-resolution forecasts in accurately forecasting ramp duration. As with the average timing errors, there is a clear demarcation in forecast ramp rate errors as a function of ramp definition, with all the forecasts more accurately forecasting ramp rate for the smaller magnitude (15% rated) ramp events. For these events, the HRRR-based bias-corrected raw forecasts are extremely accurate at correctly predicting the average ramp rate, with errors as much as 80-90% smaller than the coarser resolution forecasts. A similar picture emergences when only ‘misses’ occurring within the extended window are considered, although the difference in ramp error rates between the HRRR-based forecasts and coarser resolution model-based forecasts is not quite as large (Figure 41). The HRRR_BCRW power forecasts still perform substantially better at predicting ramp rate (40% to 60% improvement over the coarser resolution forecasts), but there is also more variability in the ramp rate errors for reasons that were discussed previously.

The average absolute timing error (as described above) for ‘hits’ differed little between the three model-based bias-corrected raw forecasts, ranging from 45-52 minutes with the ESRL_RAP_BCRW forecasts having errors of about 5 minutes less than the HRRR-based and OP_RAP-based forecasts (on average) as can be seen in Figure 42. The average absolute timing errors also varied little across the suite of ramp definitions, and were a few minutes larger for up-ramp events for the HRRR and OP_RAP-based bias-corrected raw forecasts compared to the timing errors for all events. The ERSRL_RAP-based forecasts had slightly smaller timing errors for up-ramp events compared to all events (not shown). For down-ramp events there was almost no difference in absolute average timing errors between the three forecasts which all had errors of approximately 48 minutes on average.

The average absolute timing errors for ‘misses’ were more dependent on ramp definition than the errors for ‘hits’, with the smaller magnitude events of 15% rated capacity having smaller timing errors for both up-ramp and down-ramp ‘misses’ as can be seen in Figure 43. For the smaller-amplitude events, the HRRR_BCRW power forecasts have the smallest errors and OP_RAP-based forecasts tend to have the largest timing errors. Note that the average absolute timing errors are sometimes less than 120 minutes (2 hours), which indicates that a fraction of these events were classified as misses because they did not meet the amplitude criteria to be classified as a ‘hit’. (Magnitude error statistics indicated that the bias-corrected raw forecasts tended to over-predict the magnitude of these events.) For the larger-amplitude ‘misses’, the OP_RAP_BCRW forecasts had the smallest absolute average timing errors, while the ESRL_RAP-based forecasts had the largest errors (approximately 13 minutes larger than the OP_RAP on average).

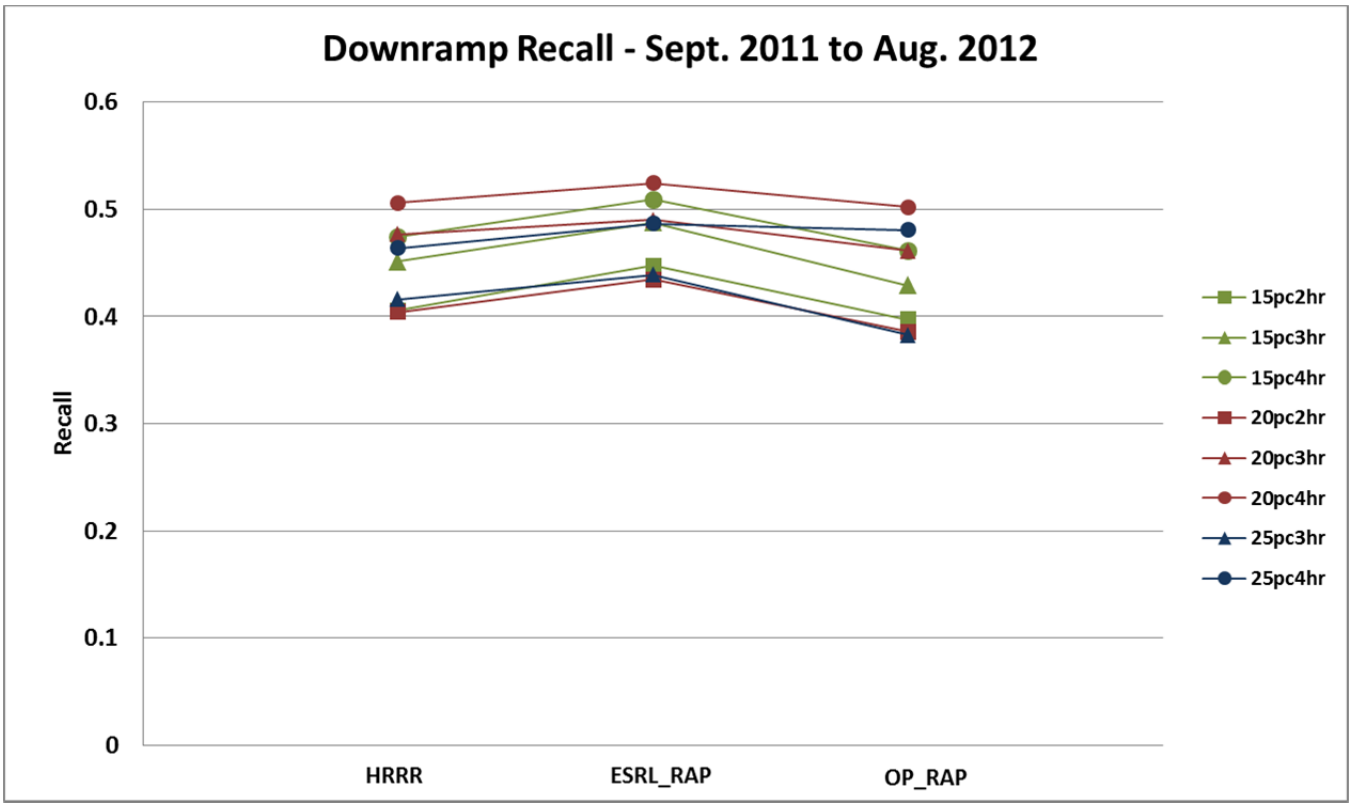

Figure 35: Recall for all system-wide wind power down-ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

Average timing errors were also calculated to gauge (on average) whether or not the forecasted ramp events occurred earlier or later than the observed events. The forecasted ramp events in all three model-based bias-corrected raw power forecasts occurred 5-10 minutes later than the observed events on average, with the HRRR-based forecasts having roughly half the average timing errors than the coarser resolution model- based forecasts as shown Figure 44. Average timing errors were slightly larger and more variable (based on ramp definition) for up-ramp events, particularly for the ESRL_RAP and OP_RAP-based forecasts which had average timing errors as large as 20 minutes (timing errors of 10-15 minutes were more typical). Average timing errors for down-ramp events were similar for all model-based forecasts and were less than 5 minutes for most ramp definitions. The average timing errors for 'misses' was much more variable than those for 'hits' as can be seen in Figure 45; however, it should be noted that the number of 'misses' in the extended verification window was often significantly less than the number of 'hits' for any given ramp definition. Thus, some of this variability could be due to smaller sample sizes. Some general trends can be still be seen, as the experimental model-based bias-corrected raw forecast 'misses' tended to occur earlier than the observed ramp (negative average timing errors), while the operational model-based forecast 'misses' tend to occur later than the observed ramp (on average). Some caution must be used in this case when comparing the actual numbers as a few outliers could skew the results with the smaller sample size available for these calculations.

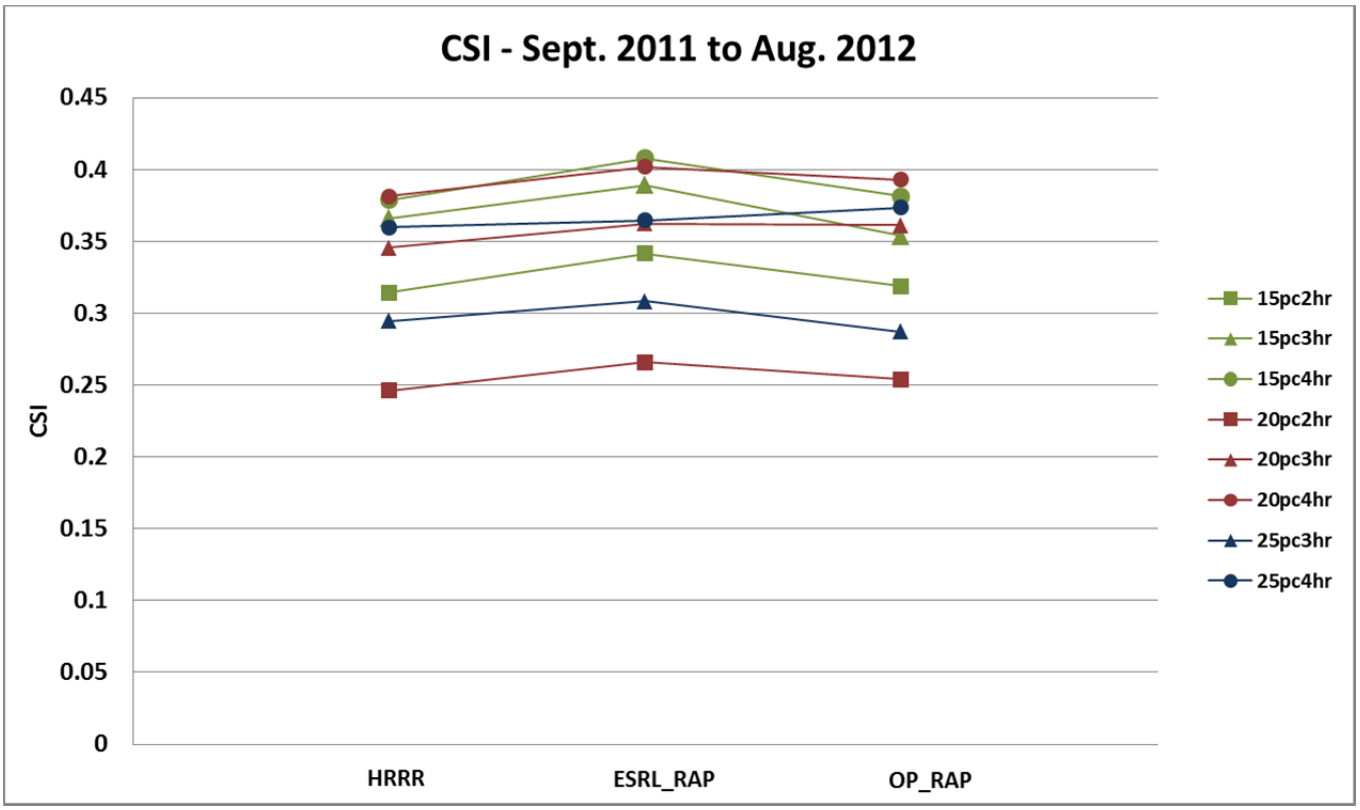

Figure 36: Critical Success Index (CSI) for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

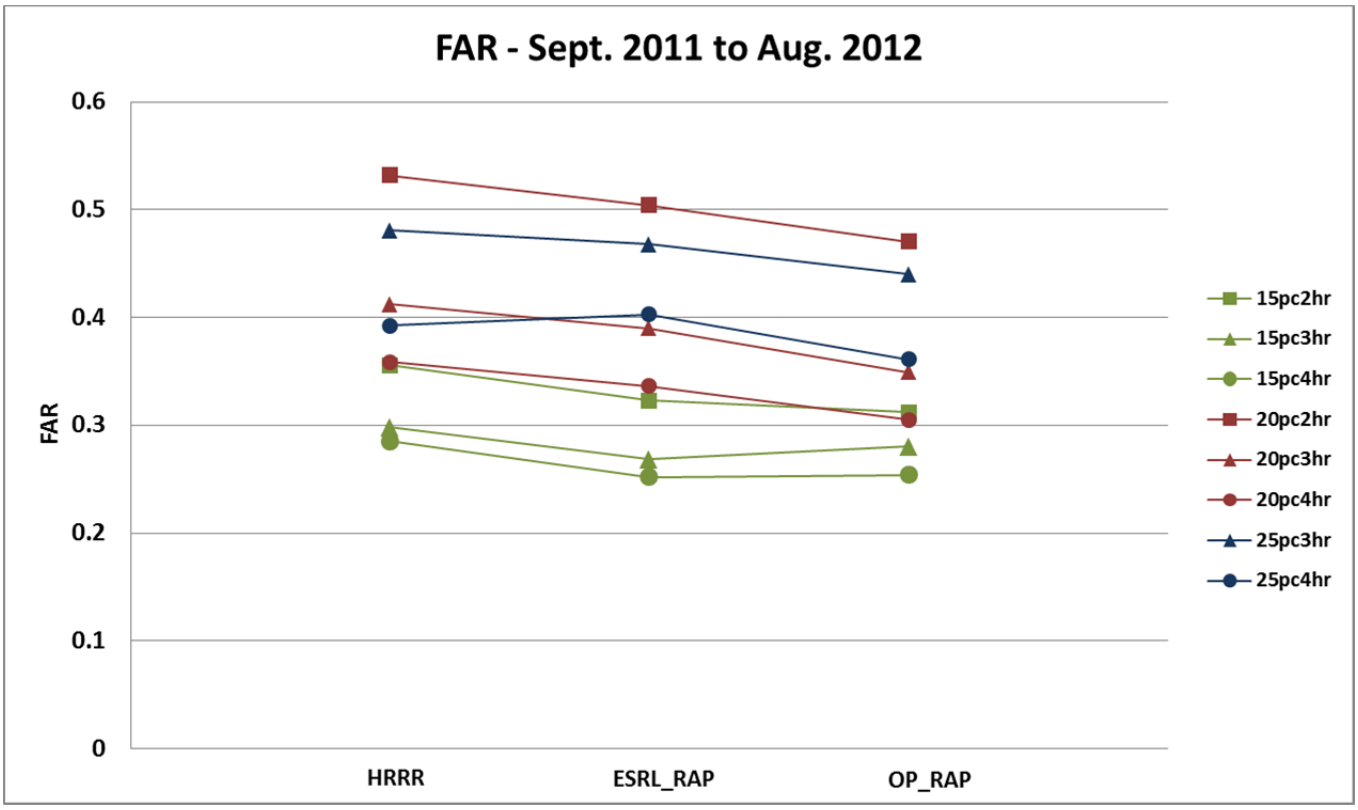

Figure 37: False Alarm Ratio (FAR) for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

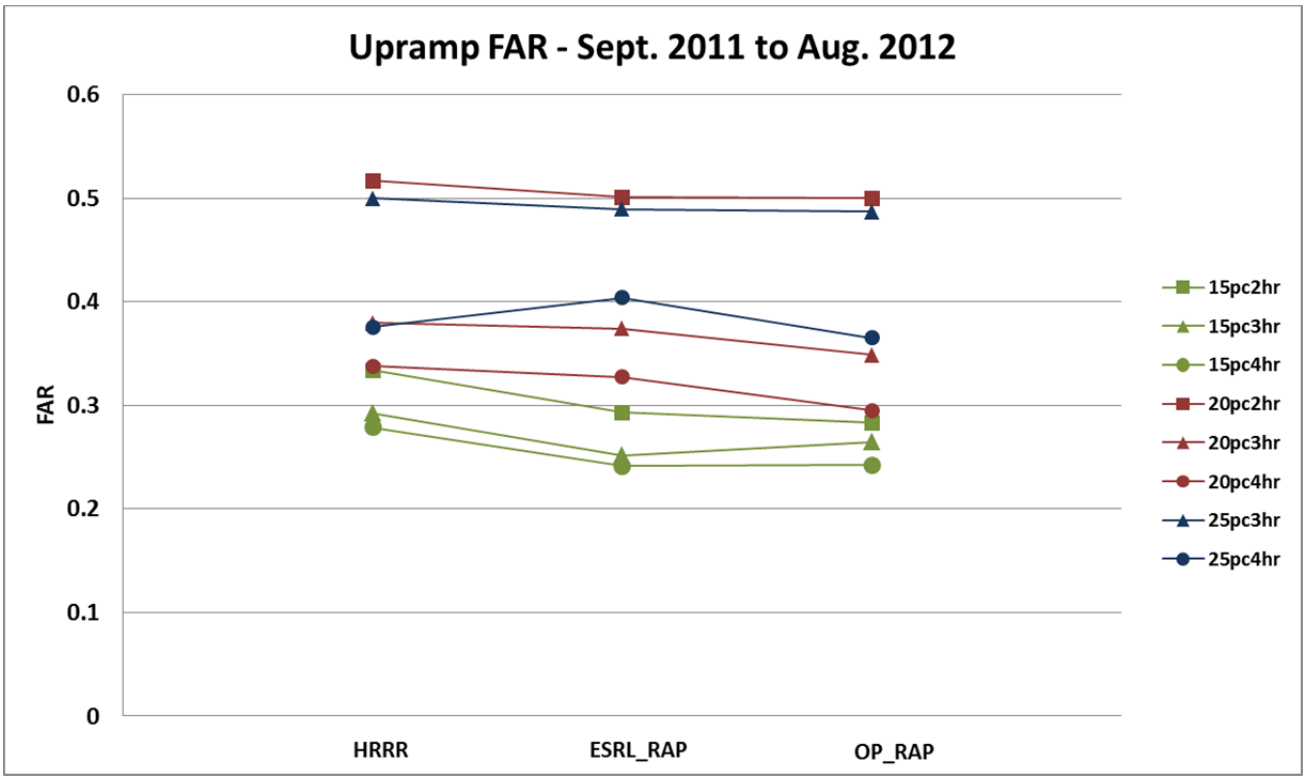

Figure 38: False Alarm Ratio (FAR) for all system-wide wind power up-ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

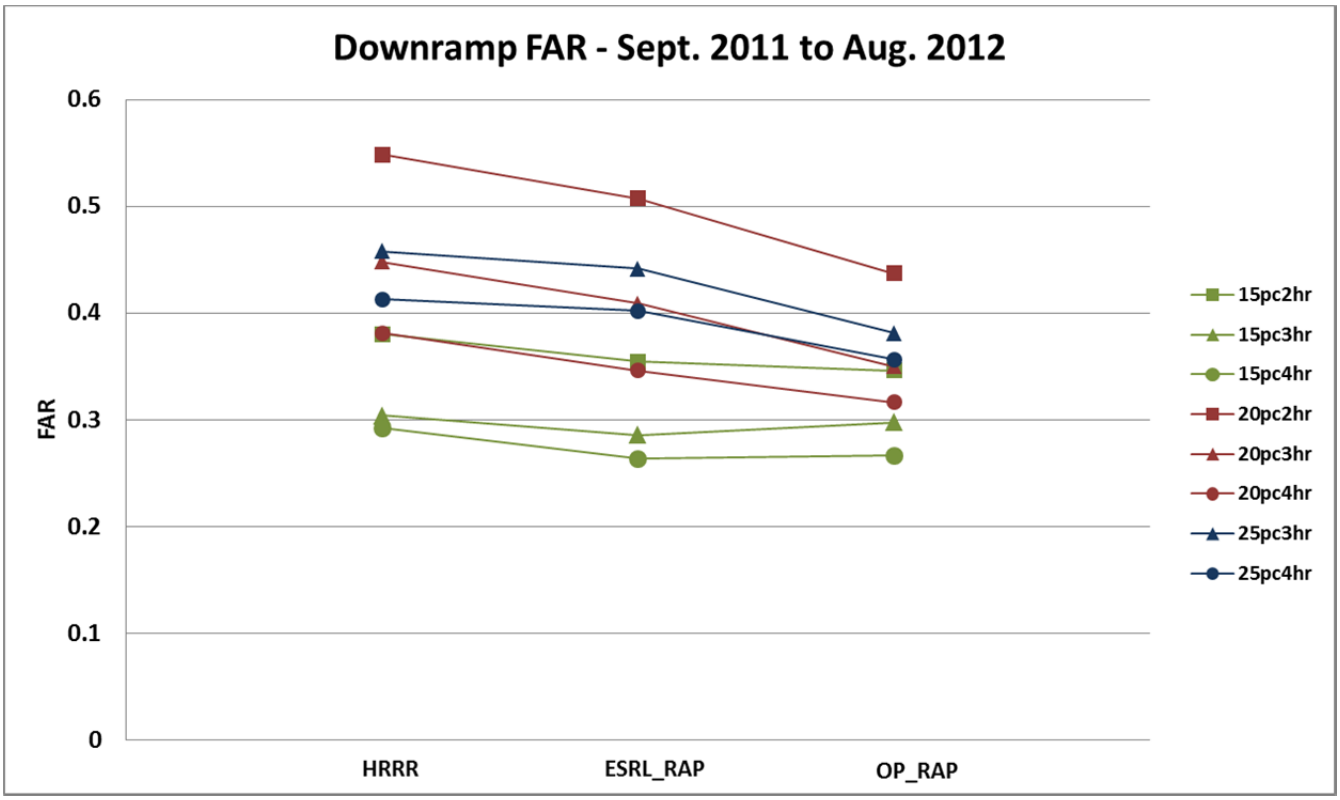

Figure 39: False Alarm Ratio (FAR) for all system-wide wind power down -ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

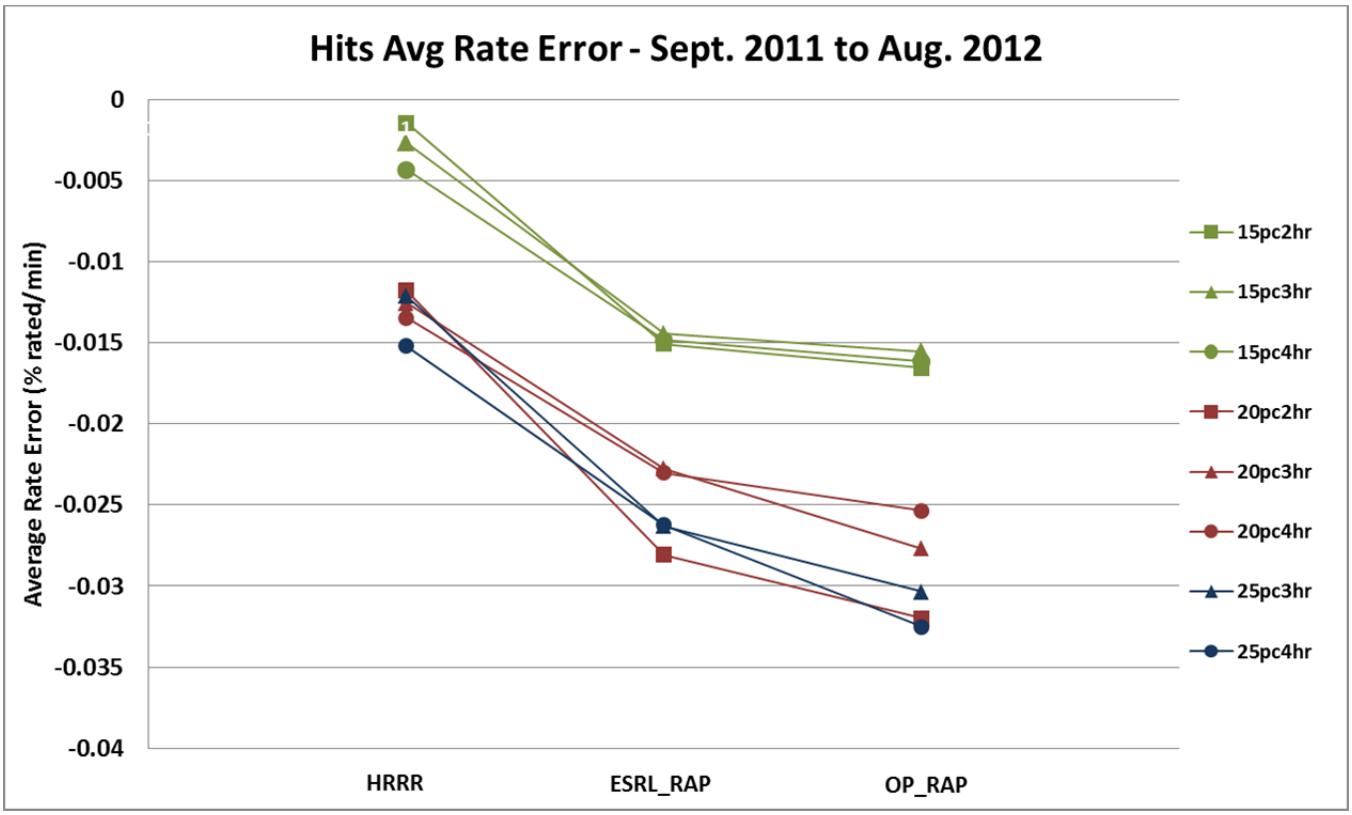

Figure 40: Average Ramp Rate Errors (as defined in the text) for all ramp events classified as 'hits' for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

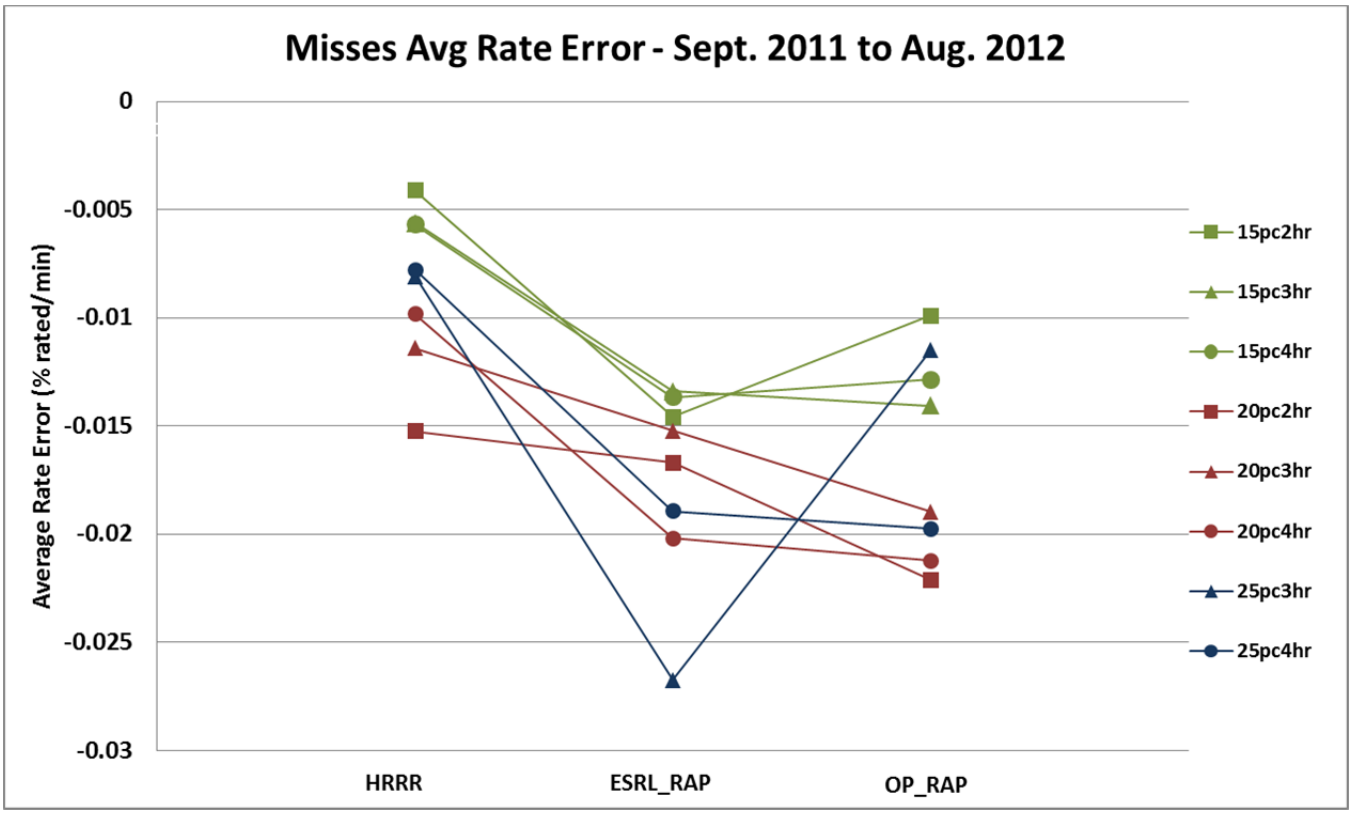

Figure 41: Average Ramp Rate Errors (as defined in the text) for all ramp events classified as 'misses' occurring in the extended time window for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

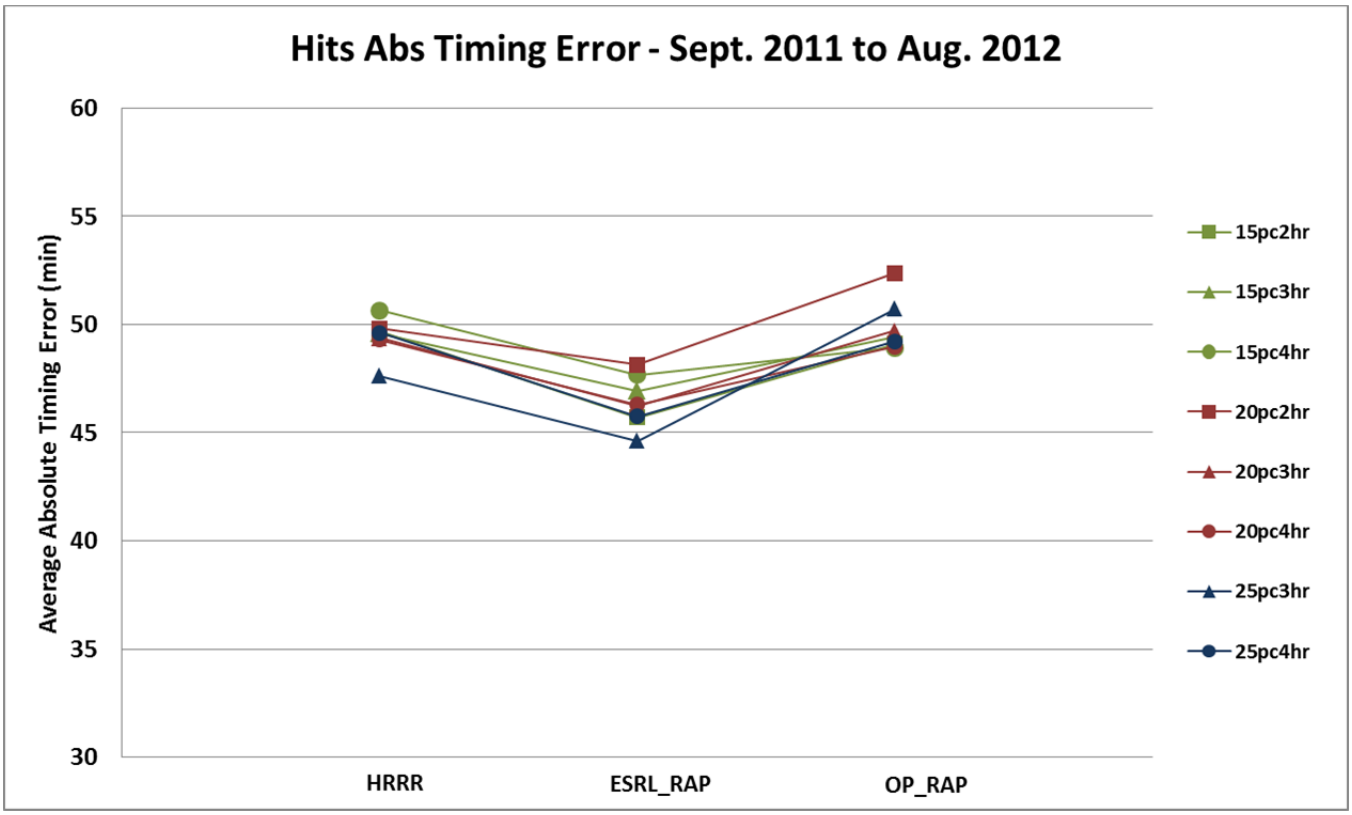

Figure 42: Average Absolute Timing Errors for all ramp events classified as 'hits' for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

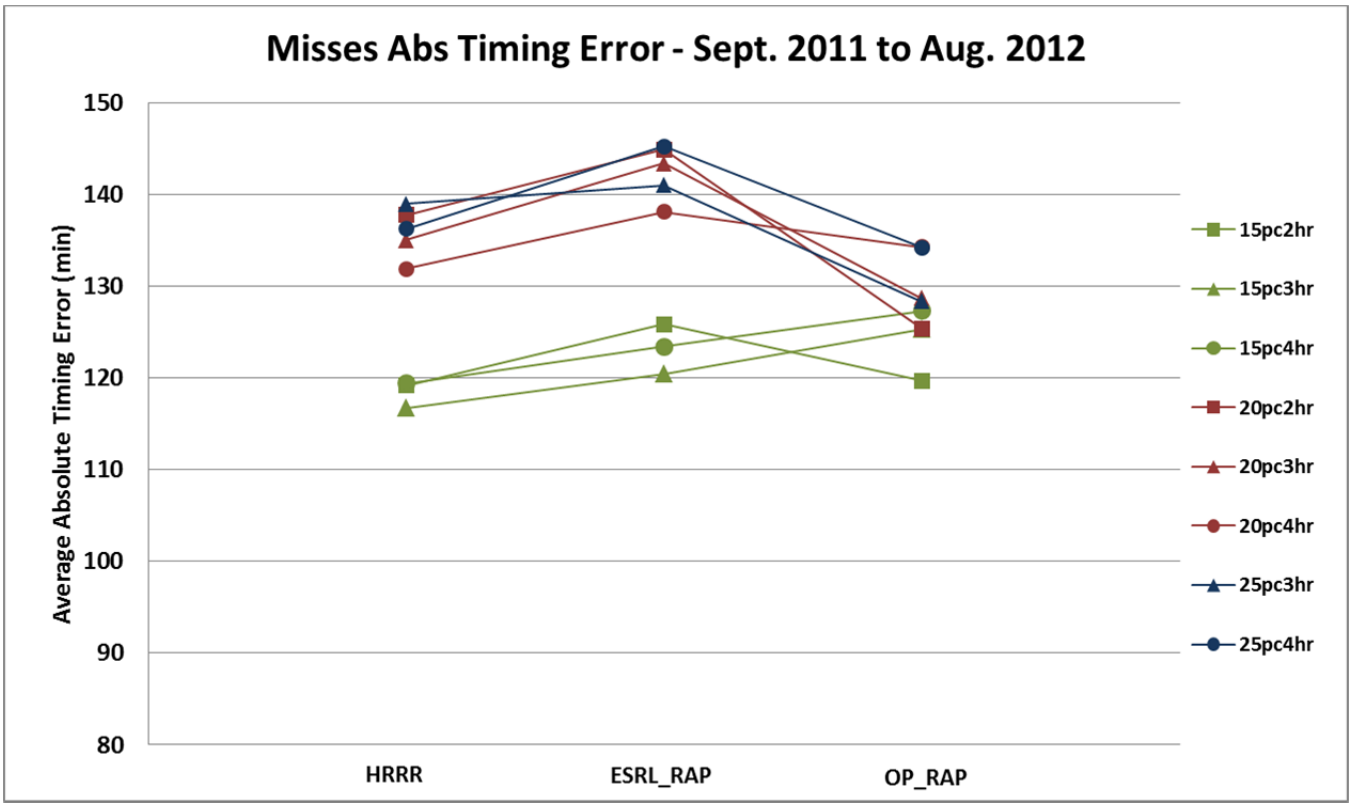

Figure 43: Average Absolute Timing Errors for all ramp events classified as 'misses' occurring in the extended time window for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

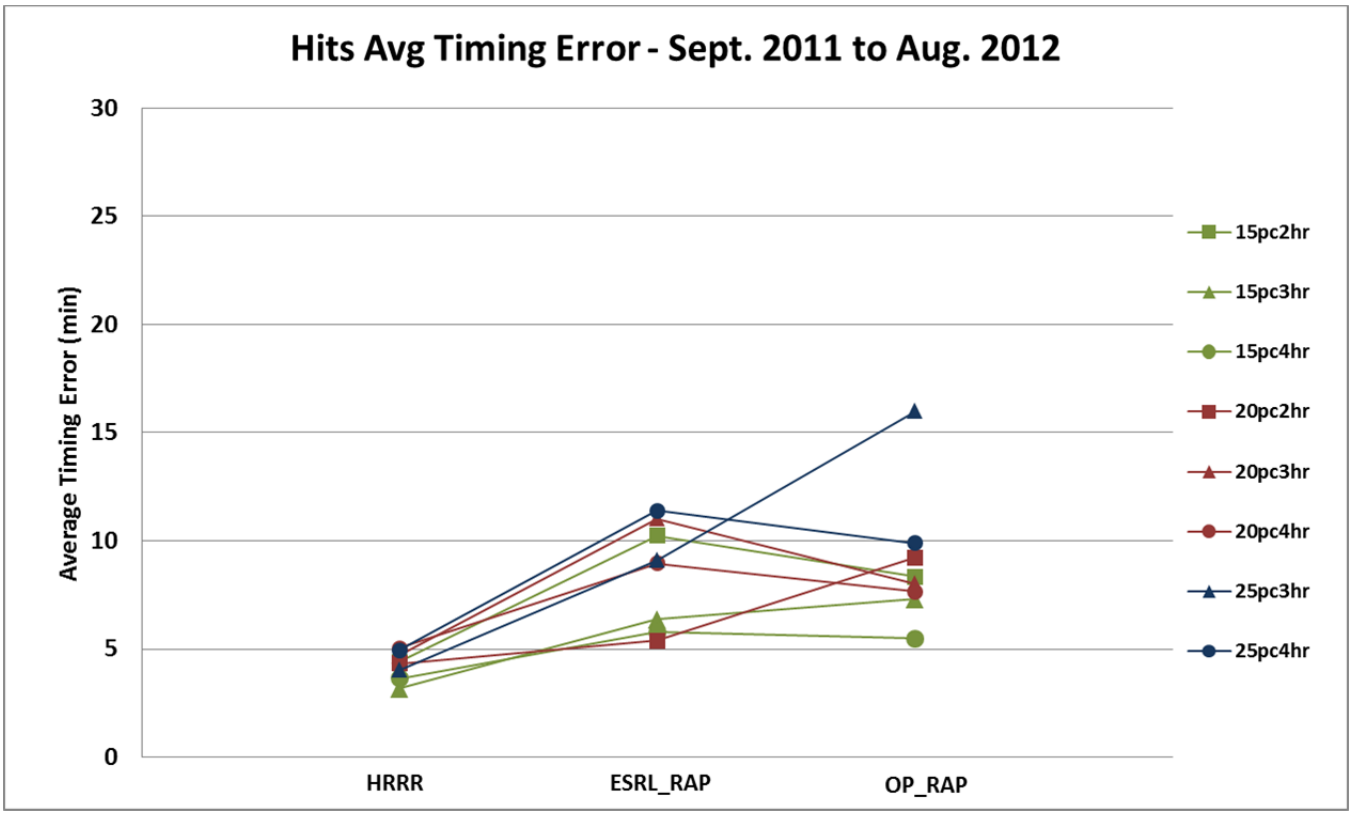

Figure 44: Average Timing Errors for all ramp events classified as 'hits' for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

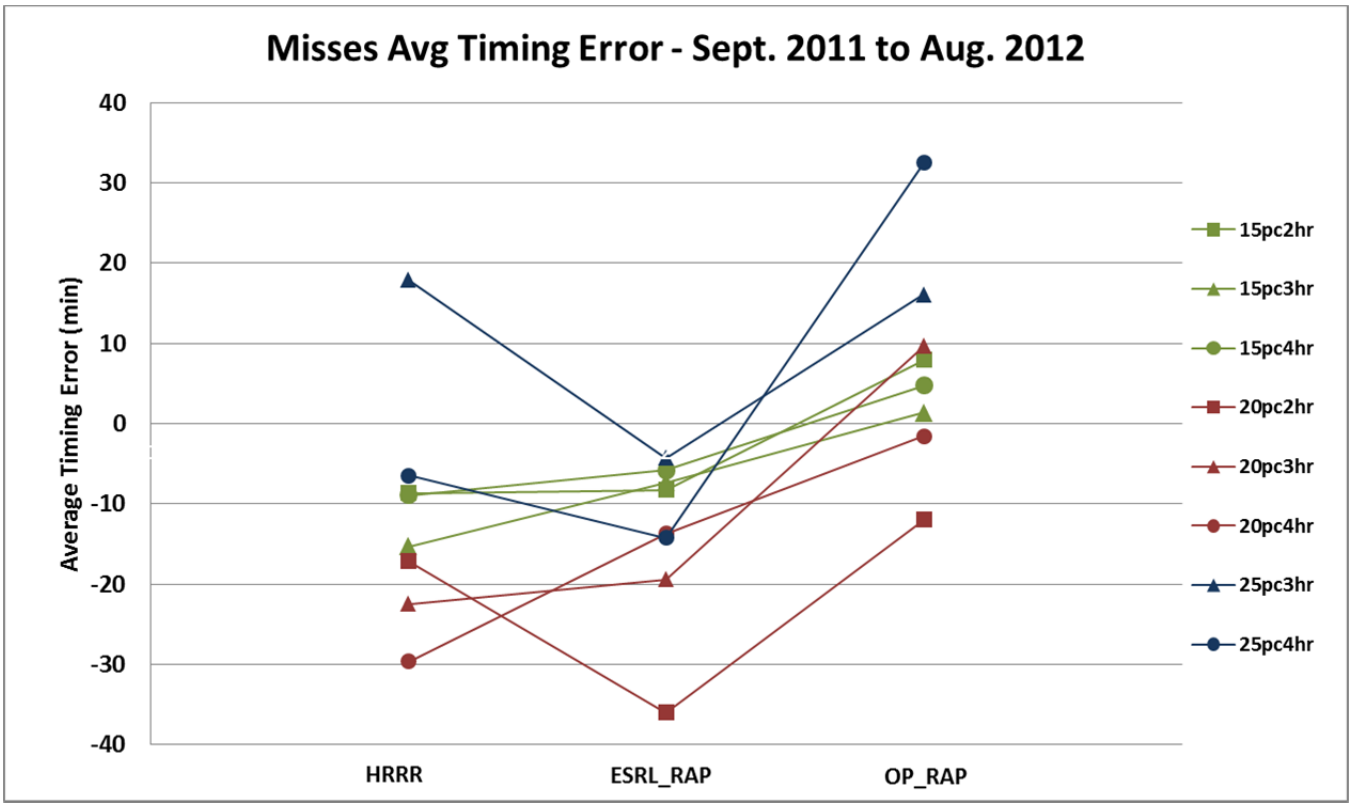

Figure 45: Average Timing Errors for all ramp events classified as 'misses' occurring in the extended time window for all system-wide wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined in the text.

4.4.3.2 Individual Sites

While system operators are most concerned about what happens at the system level, MISO is a nodal market system, which means that MISO is also using more local data to identify constraints and set nodal prices, so market participants are affected by what happens on the local/regional level since the price of electricity can vary from node to node. A similar analysis using a suite of ramp definitions was also performed at each wind plant shown in Figure 1 to gauge forecast performance during ramp events at the local level. Localized ramp events are more challenging to predict since they can be driven by smaller-scale weather features (such as thunderstorm outflows, local topographic flows, frontal passages, etc.) that are more difficult to accurately forecast than the larger synoptic-scale features that drive the system-wide wind energy ramps.

At the individual site level, the general results between the various forecast sources are similar from site to site, but there is some geographic dependency in the ramp event error statistics (as one might expect). The results presented below will focus on four representative sites: the Ashtabula Aggregate in east central North Dakota, the Crystal Lake/Hancock Aggregate in north central Iowa, the South Dakota Basin site in central South Dakota, and the Oliver Aggregate in central North Dakota. The Ashtabula and Crystal Lake sites were chosen because they are the largest sites (in terms of capacity) in the northern and southern portions of the study area, respectively. The Oliver Aggregate and the South Dakota Basin sites were chosen to give a limited representative sampling of the geographical diversity in the results. (It should also be noted that the Crystal Lake/Hancock Aggregate site and the

South Dakota Basin site are the largest and smallest wind plants in the study area, respectively.) In the previous section, the results indicated that the ramp metrics for the ESRL_RAP-based bias-corrected raw forecasts were better at the system aggregate level than the HRRR-based bias-corrected raw forecasts over the study period (and both forecasts were better than the operational model-based forecasts). However, as will be seen below, the HRRR-based forecasts more accurately predict ramp events at individual sites than the other model-based forecasts.

The Frequency Bias for the Ashtabula, Crystal Lake, Oliver, and South Dakota sites is shown in Figure 46 - Figure 49 for a subset of ramp definitions. (Note that the suite of ramp definitions at the local site level is somewhat different than at the system level, as larger amplitude/shorter duration events can occur at the local level compared to the system level.) While all forecasts tend to under-predict the number of events for most ramp definitions, the HRRR_BCRW power forecasts do a significantly better job at forecasting the number of events for all ramp definitions as compared to the other forecasts. The only exceptions are the Crystal Lake site where the HRRR_BCRW forecasts tend to slightly over predict the number of ramp events for all definitions, and the large amplitude ramp events at the Ashtabula site where the HRRR-based forecasts over predict the number of events. This is opposite to what was seen at the system level, and it is possible that many of these large local ramp events are convective in nature. At all sites (including those not shown), the bias-corrected raw HRRR-based forecasts better predict the total number of ramp events than the other forecasts (and the ESRL_RAP-based forecasts better predict the number of ramps compared to the OP_RAP-based forecasts).

Figure 50 - Figure 53 show the distribution of the Recall metric (i.e. the fraction of events that are predicted correctly) for a suite of ramp definitions for the chosen subset of wind plants over the entire year of the forecast period. In addition to the smallest frequency bias, the HRRR-based forecasts also more accurately predict ramp events than the coarser resolution model forecasts at the individual site (or aggregate site) level. Unlike at the system level where Recall values peak at the smallest magnitude/longest duration events, at the wind plant level the Recall values peak for longer duration ramps which have magnitudes of approximately 30% to 40% of the plant rated capacity. Recall values are also quite stable at the individual plant level across a wide range of ramp magnitudes for ramp events with durations greater than two hours, unlike the results at the system level where prediction accuracy dropped off rapidly as ramp magnitude increased and ramp duration decreased.

A comparison of the Recall metric for the various model-based bias-corrected raw power forecasts for a suite of ramp definitions at each site is shown in Figure 54 - Figure 57. For all ramp definitions, the HRRR-based power forecasts more accurately predict ramp events than the coarser-resolution forecasts, while the ESRL_RAP-based forecasts predict events more accurately than the operational RAP-based forecasts. In general, moderate amplitude (approximately 30% rated capacity), longer duration events are predicted better by all the bias-corrected raw forecasts than are larger amplitude, and shorter duration events, although the ramp duration is the more decisive factor. Recall values drop significantly for the short duration ramp events for all model-based forecasts, but there is a greater difference in Recall values between the HRRR-based forecast and the coarser resolution model-based forecasts for the shorter duration events (less than 2 hours), meaning the HRRR-based forecasts are capturing these events much better than the coarser resolution forecasts. These are the events that are of greatest concern to system operators even if they occur at the local level, as they create the largest challenges for balancing the system in the case of existing transmission constraints.

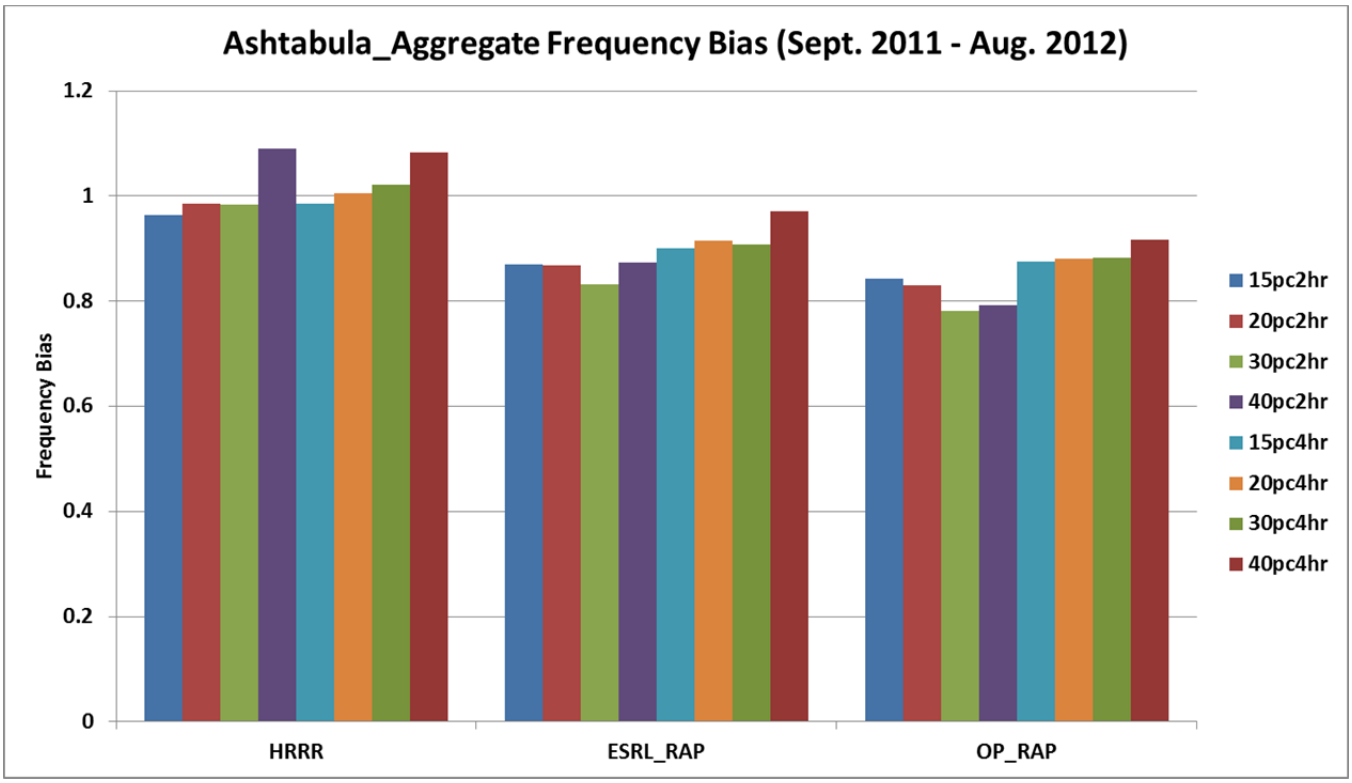

Figure 46: Ashtabula Aggregate Frequency Bias for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined above.

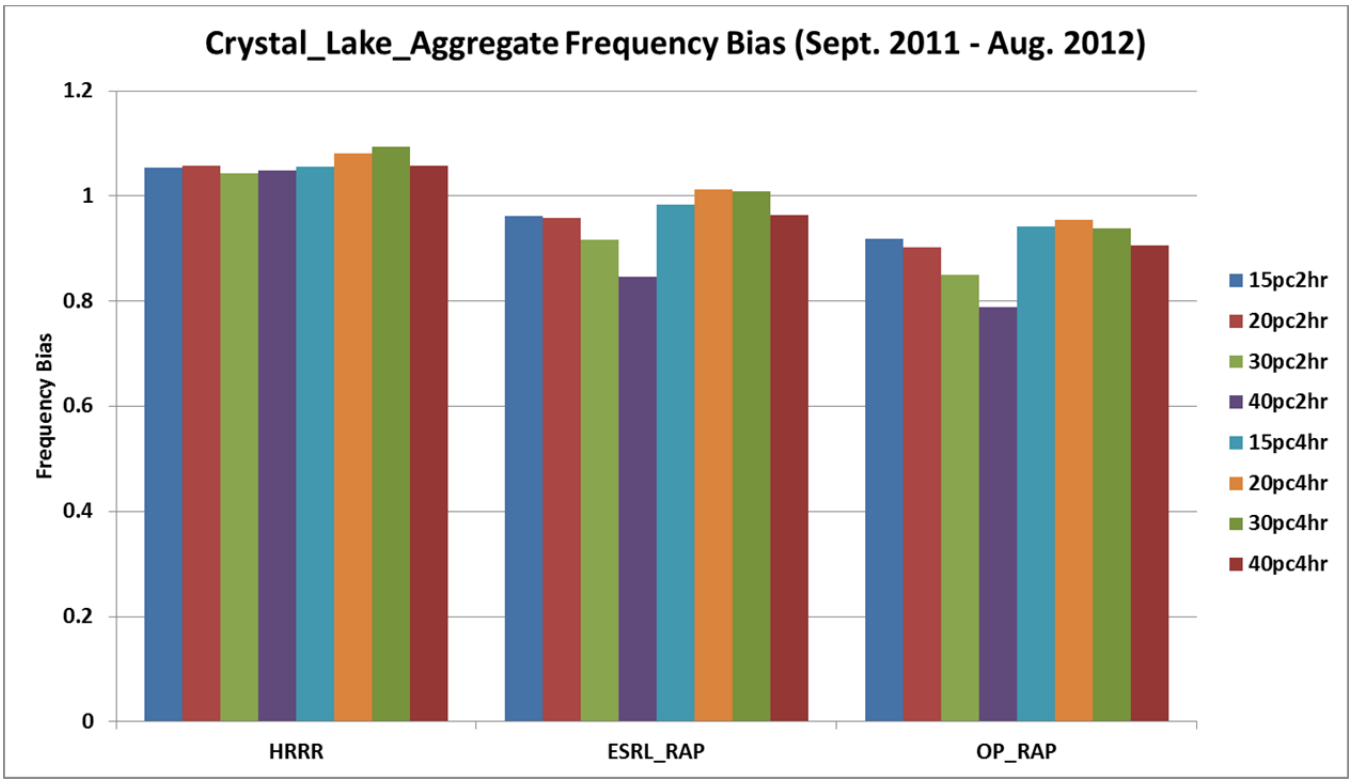

Figure 47: Crystal Lake Aggregate Frequency Bias for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined above.

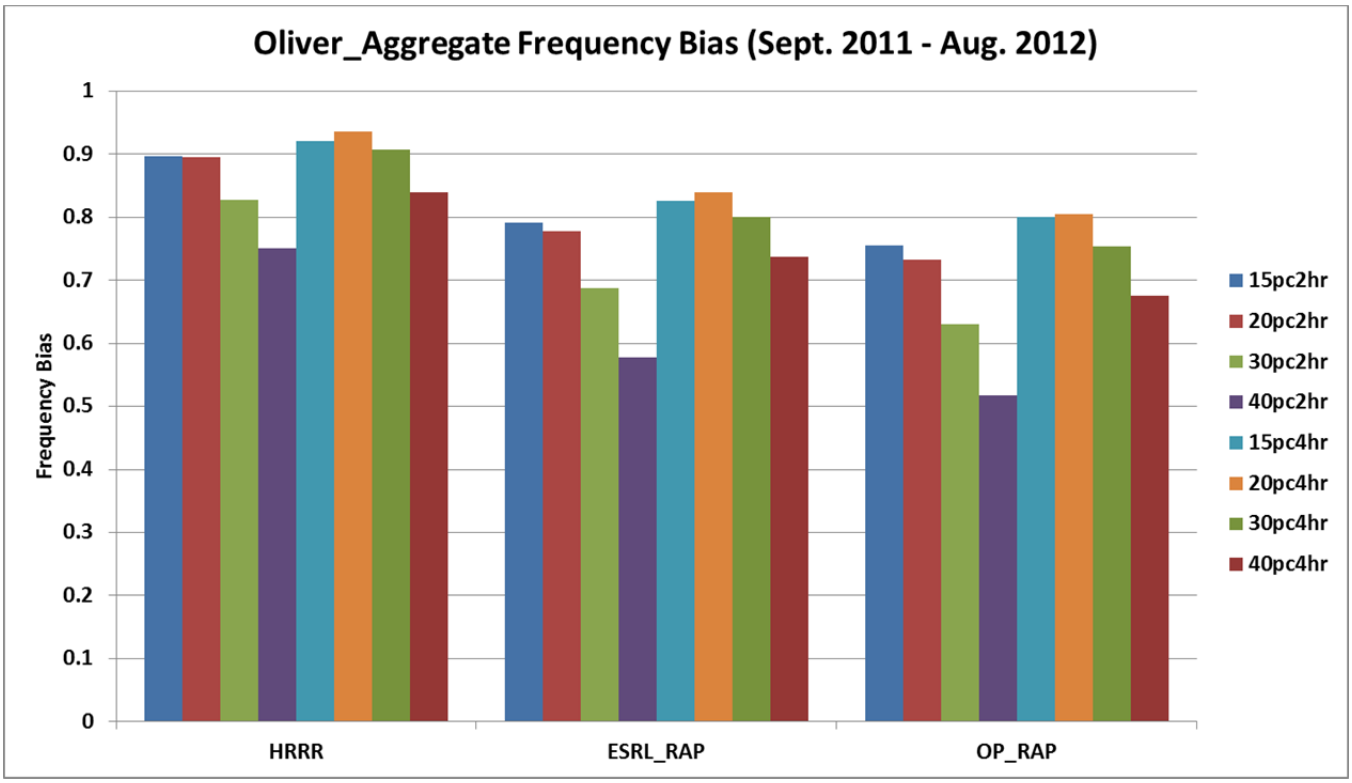

Figure 48: Oliver Aggregate Frequency Bias for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined above.

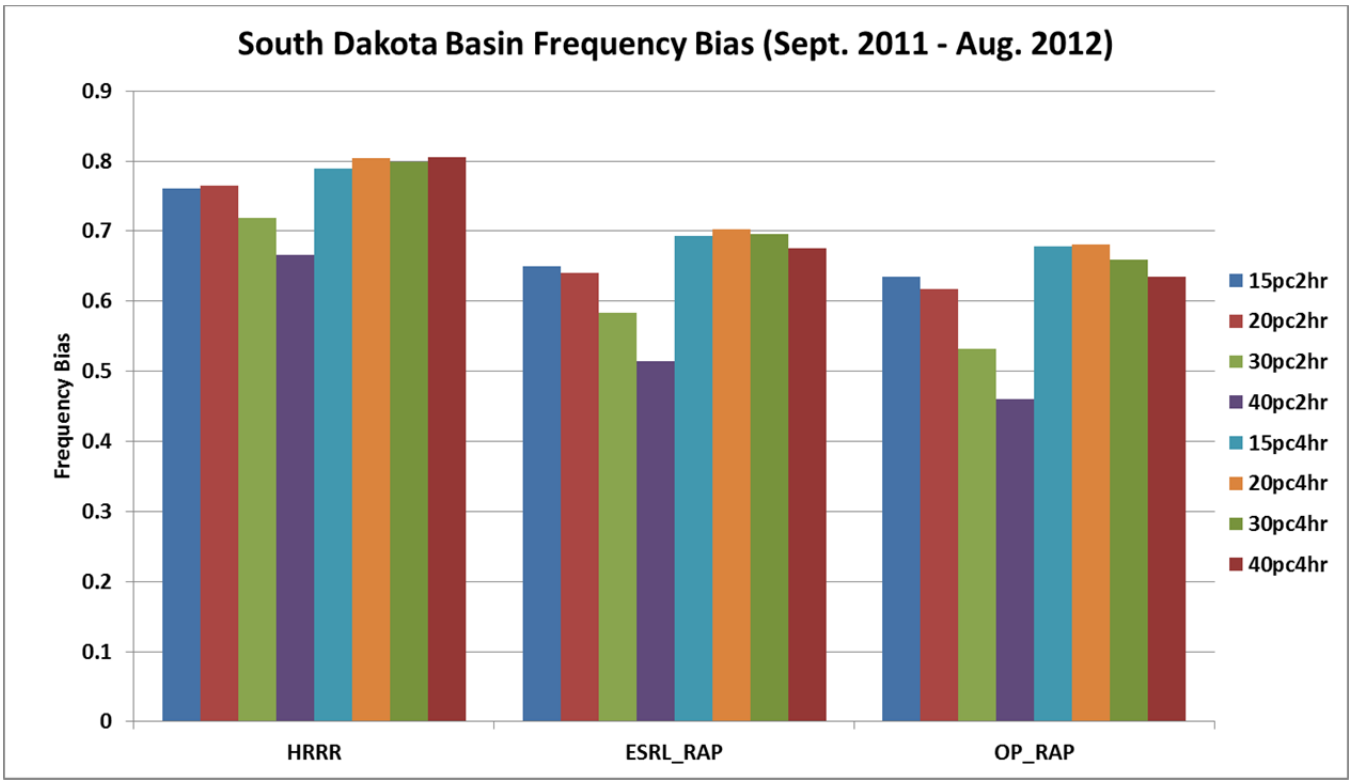

Figure 49: South Dakota Basin Frequency Bias for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 8 different ramp definitions as defined above.

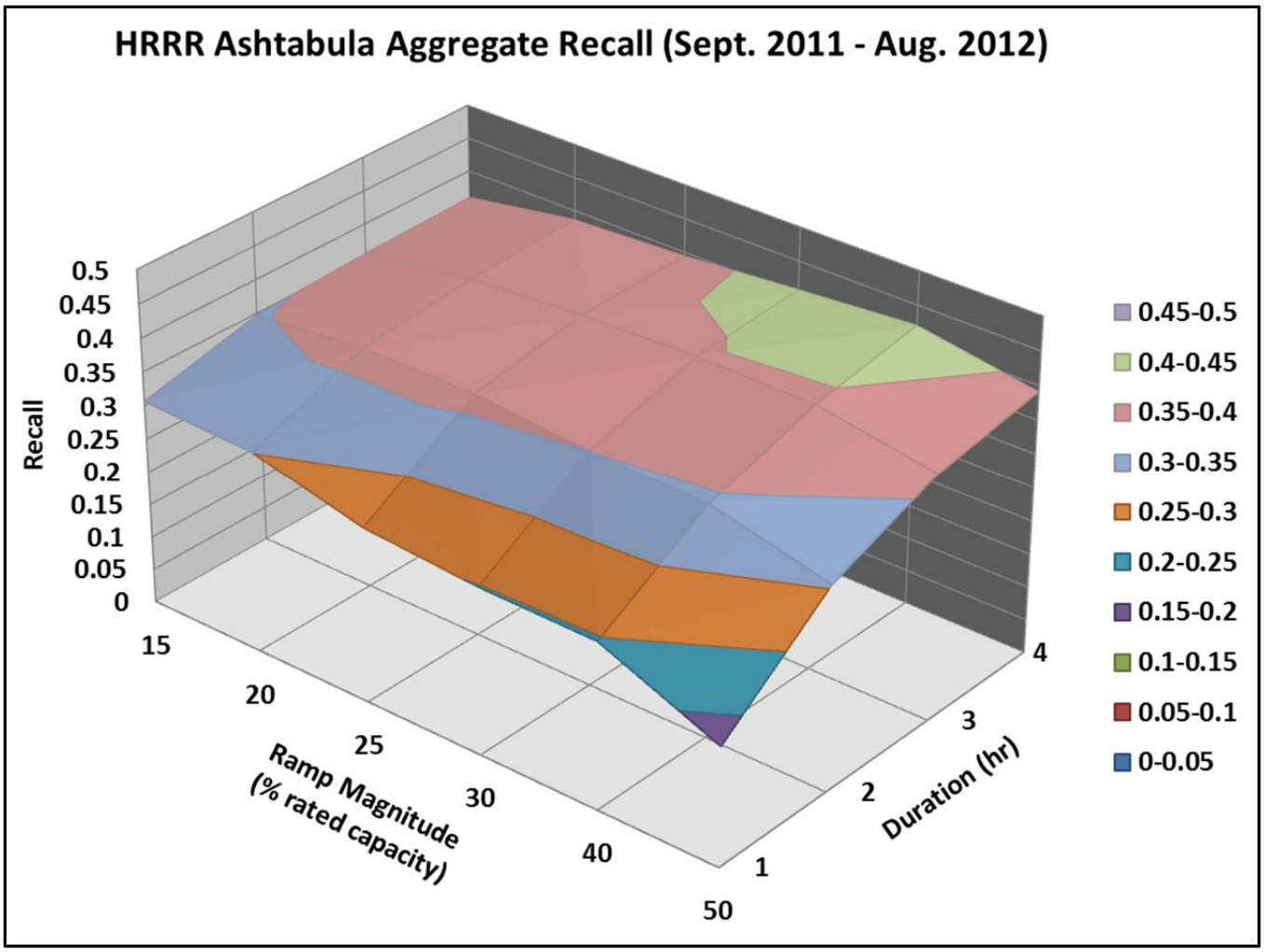

Figure 50: Ashtabula Aggregate Recall for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR-based power forecasts for a suite of ramp definitions as defined in the text.

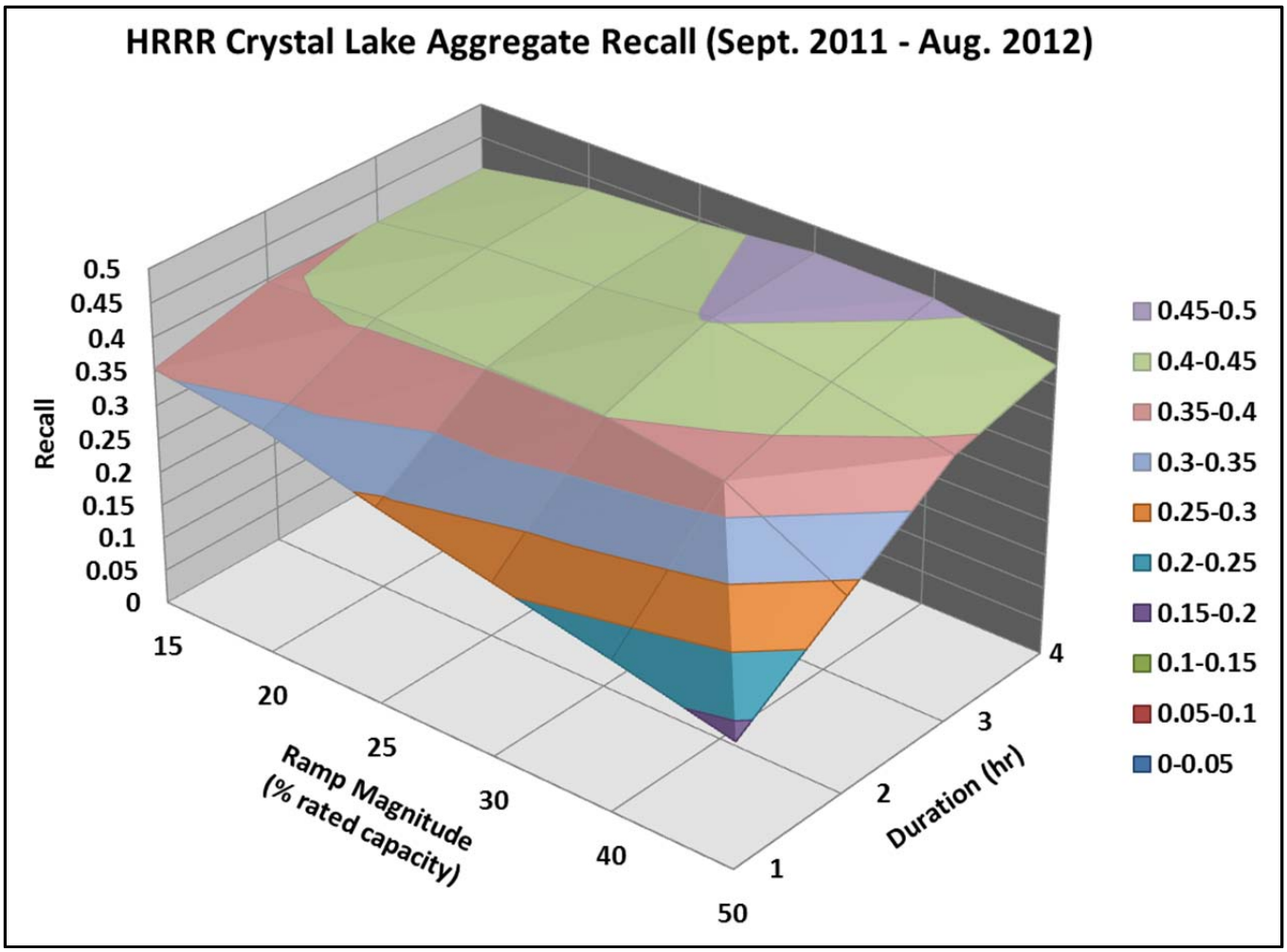

Figure 51: Crystal Lake Aggregate Recall for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR-based power forecasts for a suite of ramp definitions as defined in the text.

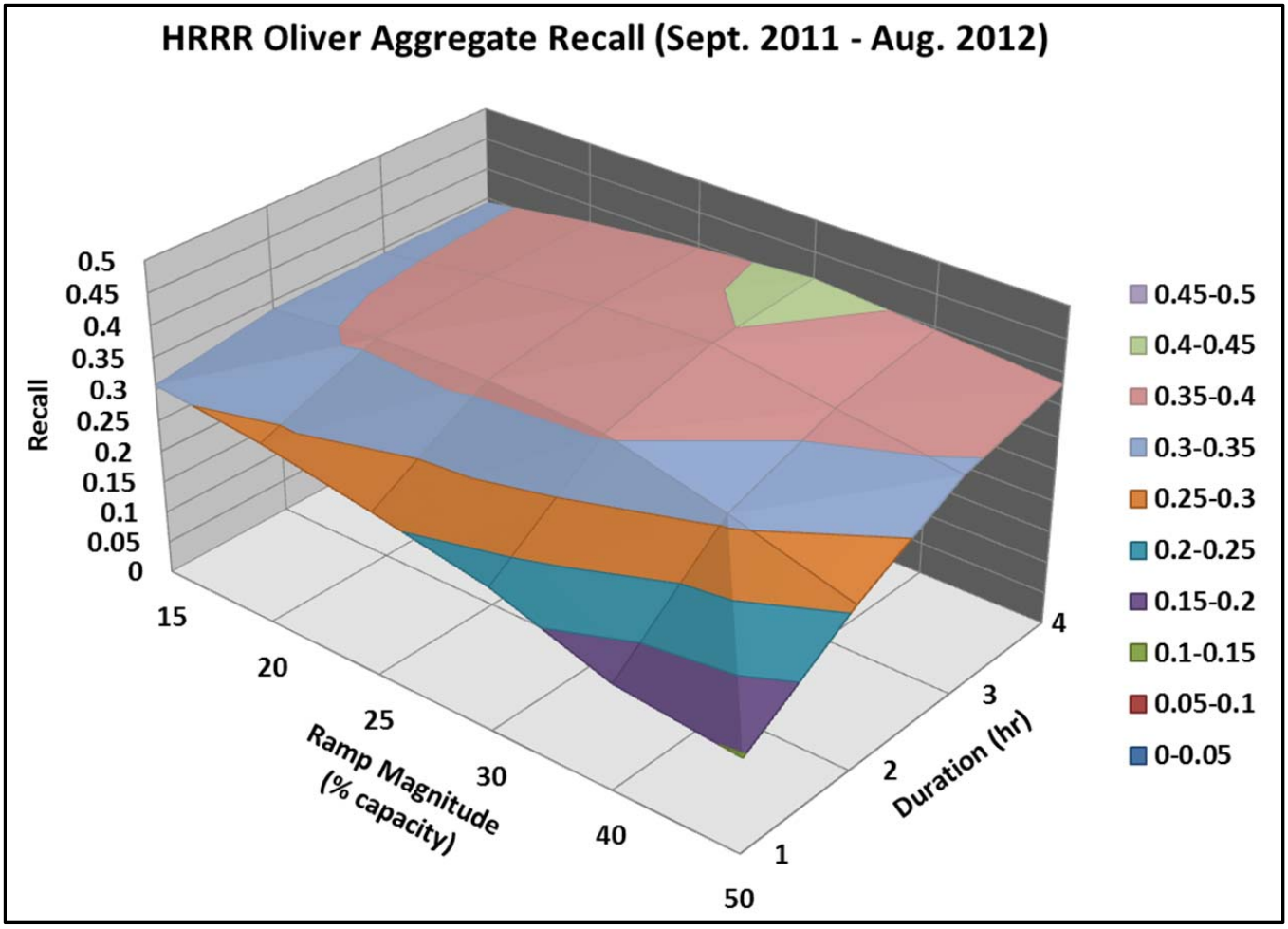

Figure 52: Oliver Aggregate Recall for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR-based power forecasts for a suite of ramp definitions as defined in the text.

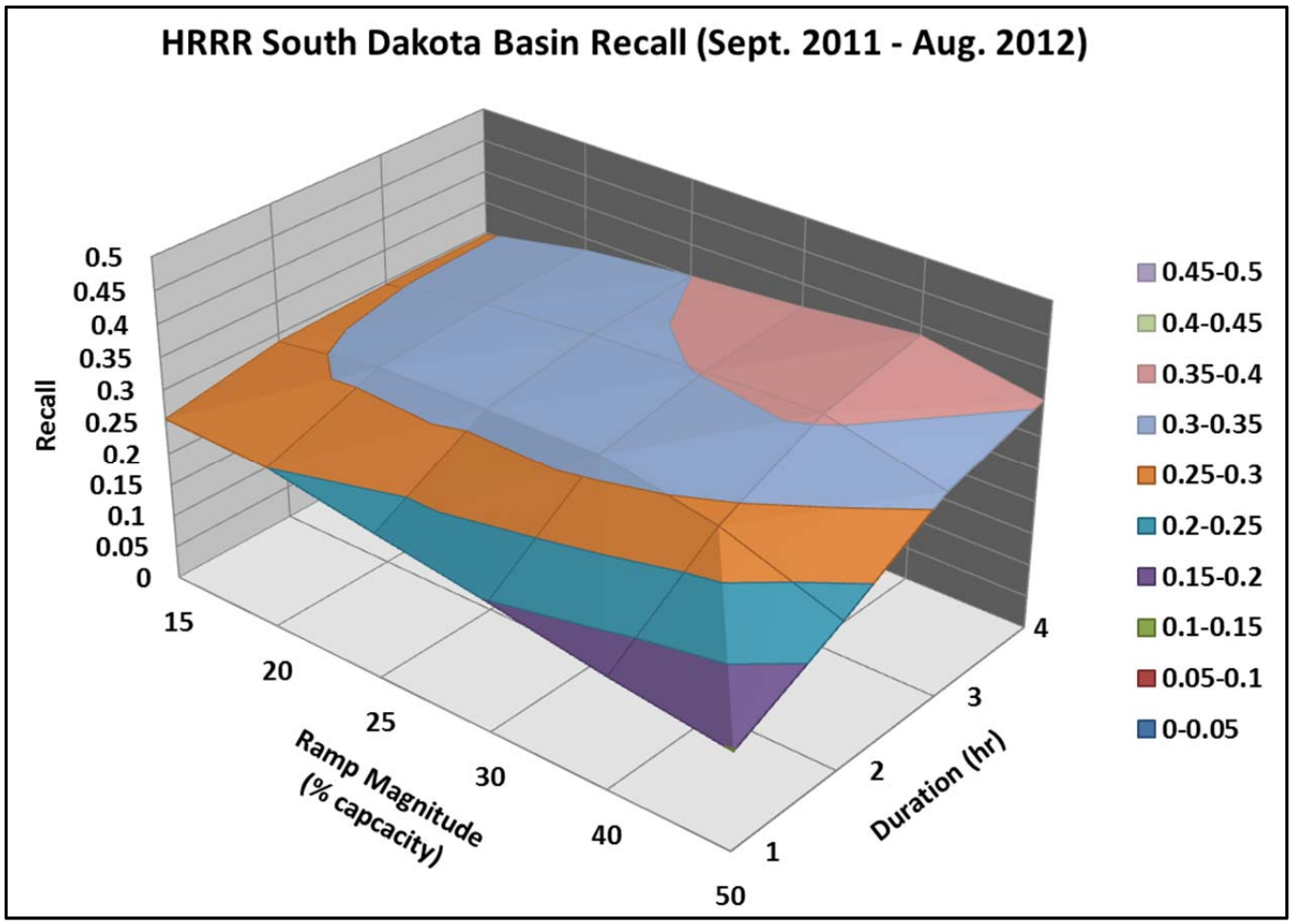

Figure 53: South Dakota Basin Recall for all wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR-based power forecasts for a suite of ramp definitions as defined in the text.

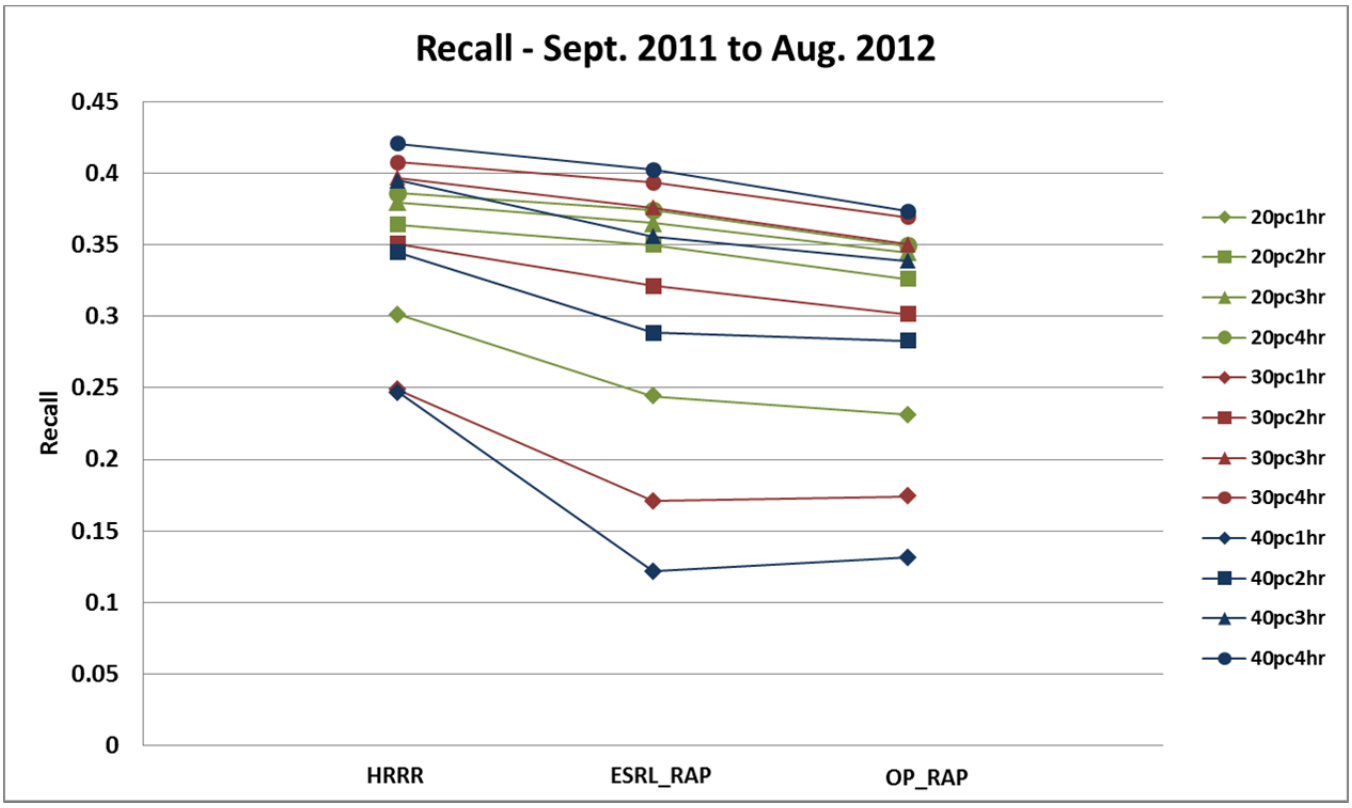

Figure 54: Recall for all Ashtabula Aggregate wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

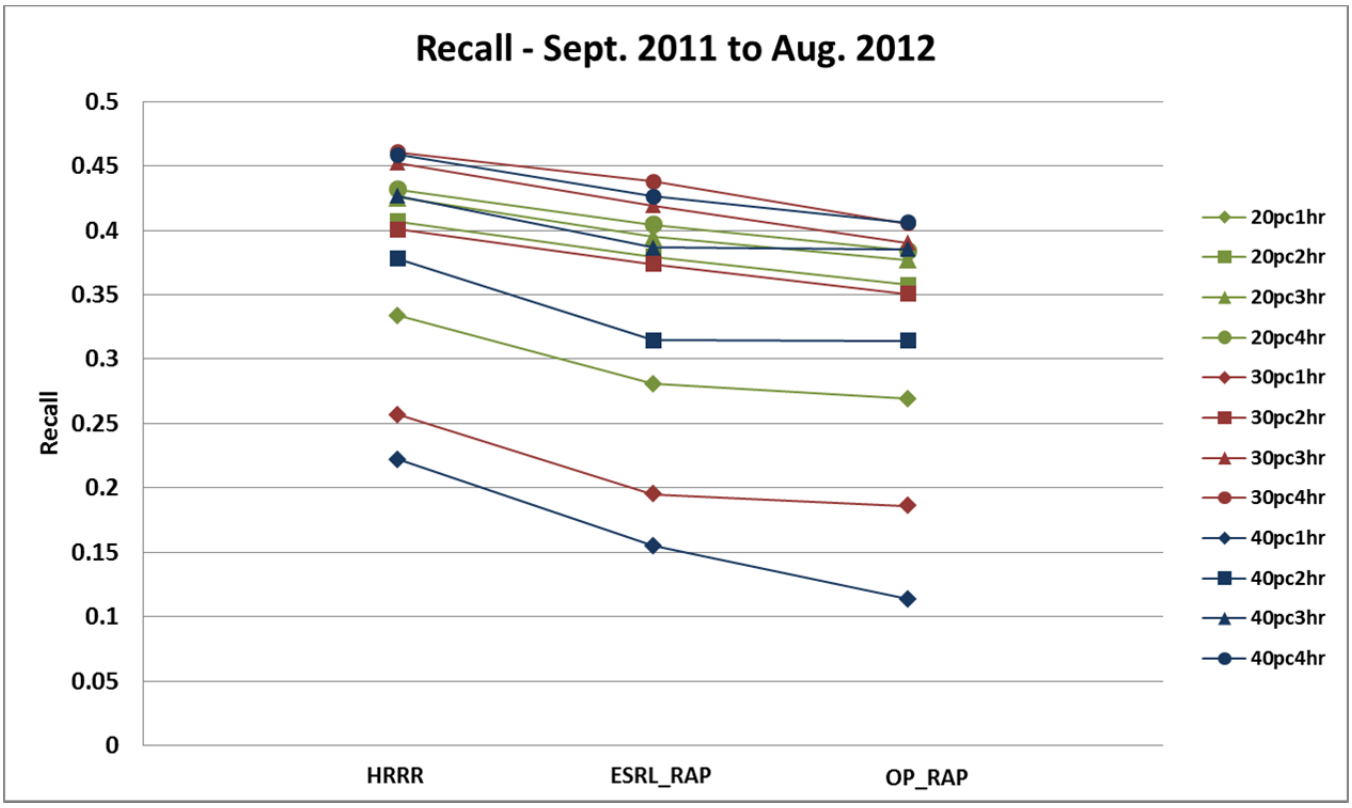

Figure 55: Recall for all Crystal Lake Aggregate wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

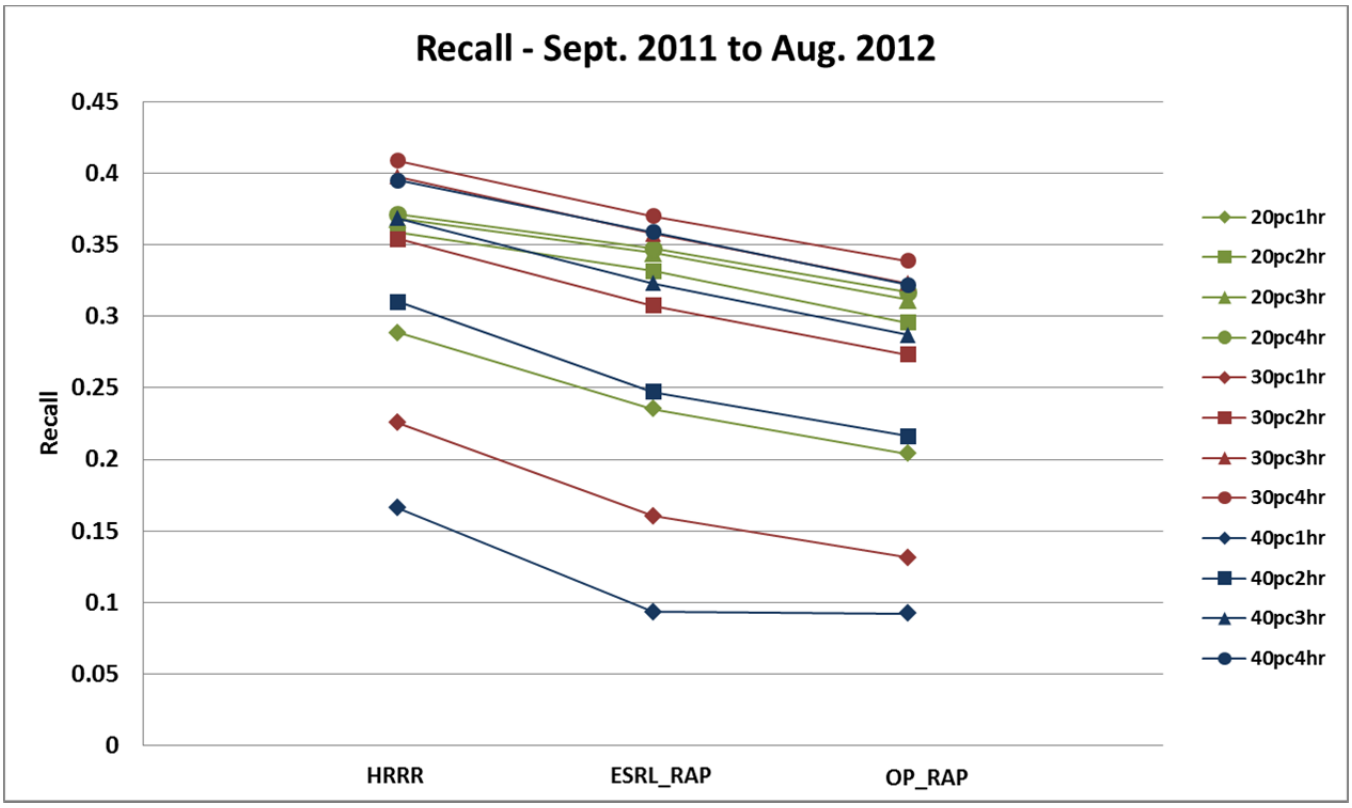

Figure 56: Recall for all Oliver Aggregate wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

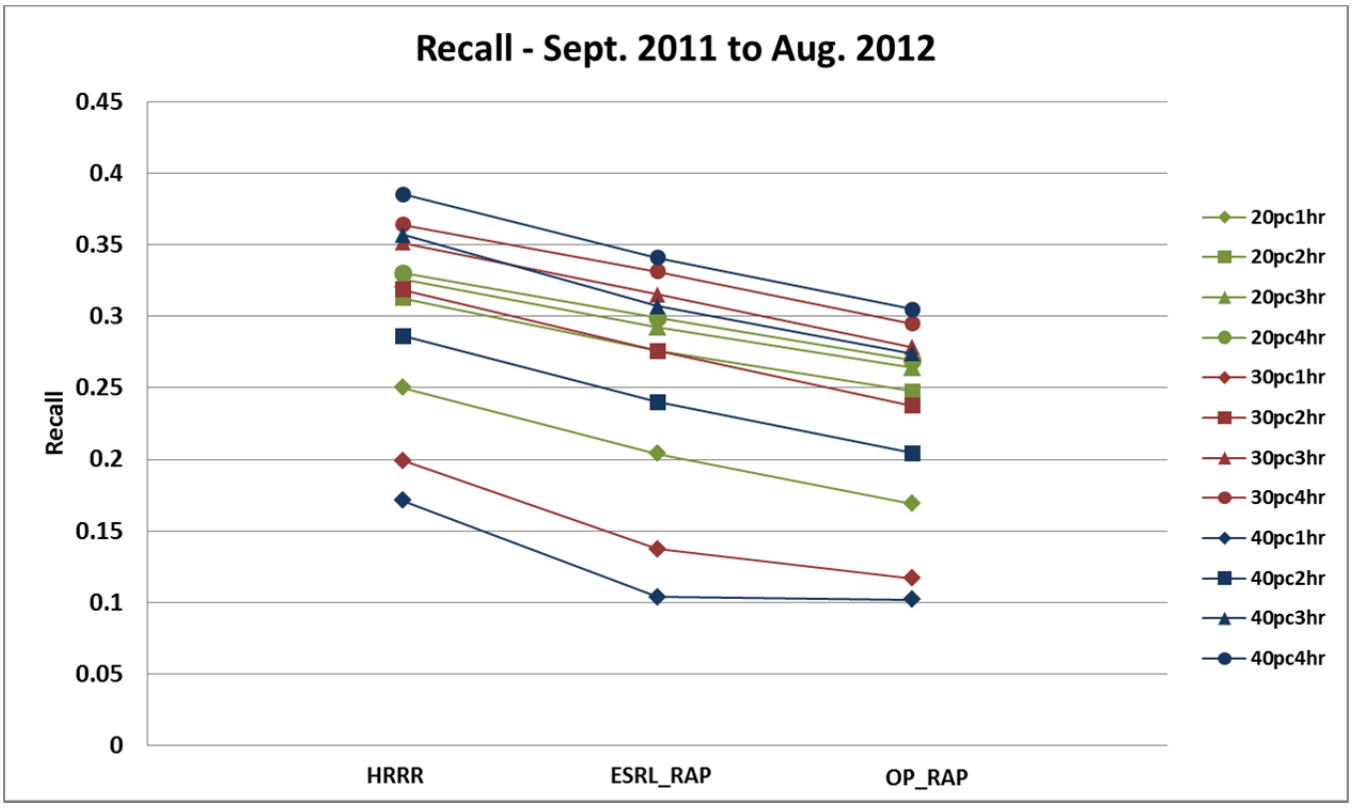

Figure 57: Recall for all South Dakota Basin wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

At the individual site level, the bias-corrected raw forecasts based on the research models have the highest Critical Success Index (CSI) scores (as defined in Equation 7) for most ramp definitions (see Figure 58 - Figure 61). At all sites, the longer duration ramp event definitions have higher CSI scores for all forecasts than the shorter duration definitions. There is more site-to-site variability in the CSI results than with the Frequency Bias and the Recall metrics, with the HRRR-based forecasts have better CSI scores than the other model-based bias-corrected raw power forecasts for all ramp definitions at some locations (Oliver, South Dakota Basin, and for most ramp definitions at Crystal Lake). At other locations, HRRR-based forecasts have better CSI scores for the larger amplitude, shorter duration ramps, and the ESRL_RAP-based forecasts have slightly better CSI scores for smaller amplitude, longer duration events (Ashtabula and Crystal Lake). There are a few exceptions where the OP_RAP-based forecasts have better CSI scores than the ESRL_RAP-based forecasts at the Ashtabula and Crystal Lake sites. This might be attributed to the lower False Alarm Ratios for the operational forecasts at those particular ramp definitions (compare Figure 59 - Figure 60 with Figure 63 - Figure 64), although the operational model-based forecasts do not capture as many of the observed events as the research model-based forecasts.

While the HRRR_BCRW power forecasts do well at predicting the number of events fairly accurately, they also have the highest False Alarm Ratios (FAR) for almost all ramp definitions at all plant sites (Figure 62 - Figure 65). Unlike the Recall and the CSI metrics, the FAR is largely stratified by ramp magnitude (rather than ramp duration), with FAR values higher for the larger amplitude, shorter duration events for all the bias-corrected raw power forecasts. This result is counterintuitive, as the number of predicted smaller amplitude ramp events is much larger than the predicted number of larger amplitude events, and the frequency bias for the smaller magnitude events is larger than the frequency

bias for the larger magnitude ramp events for a given ramp duration (though in many cases both large and small magnitude events are under predicted). It could be that timing of the large magnitude events is more difficult to predict, and hence these events are counted as false alarms because the ramp center point is outside of ± 4 hour accuracy window. It also might be that the types of weather events creating these large local ramp events (thunderstorm outflows, rapid intensifying/weakening weather systems) are difficult to accurately predict or are too often predicted, but a full understanding of this result would require a detailed analysis of the types of meteorological phenomenon that were causing the larger ramp events during the forecast period.

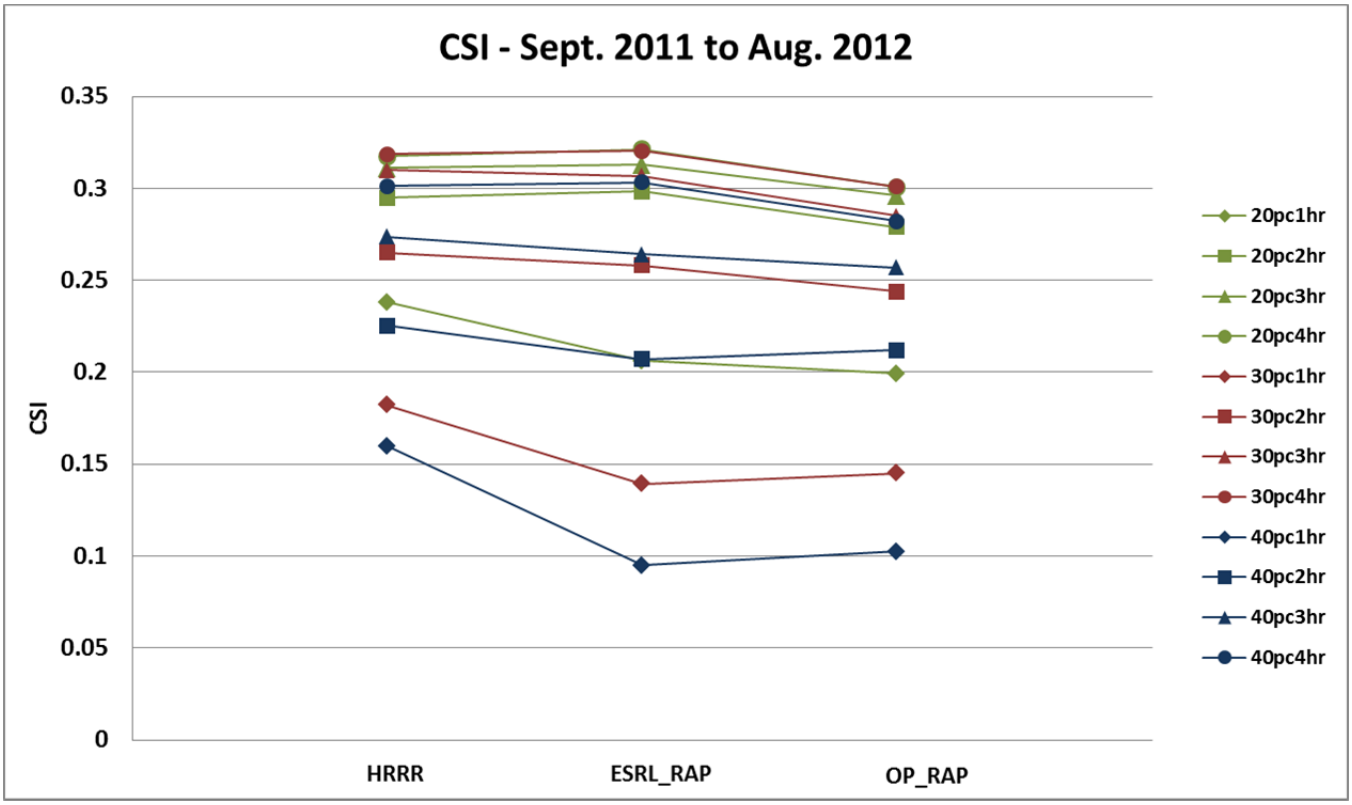

Figure 58: Critical Success Index (CSI) for all Ashtabula Aggregate wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

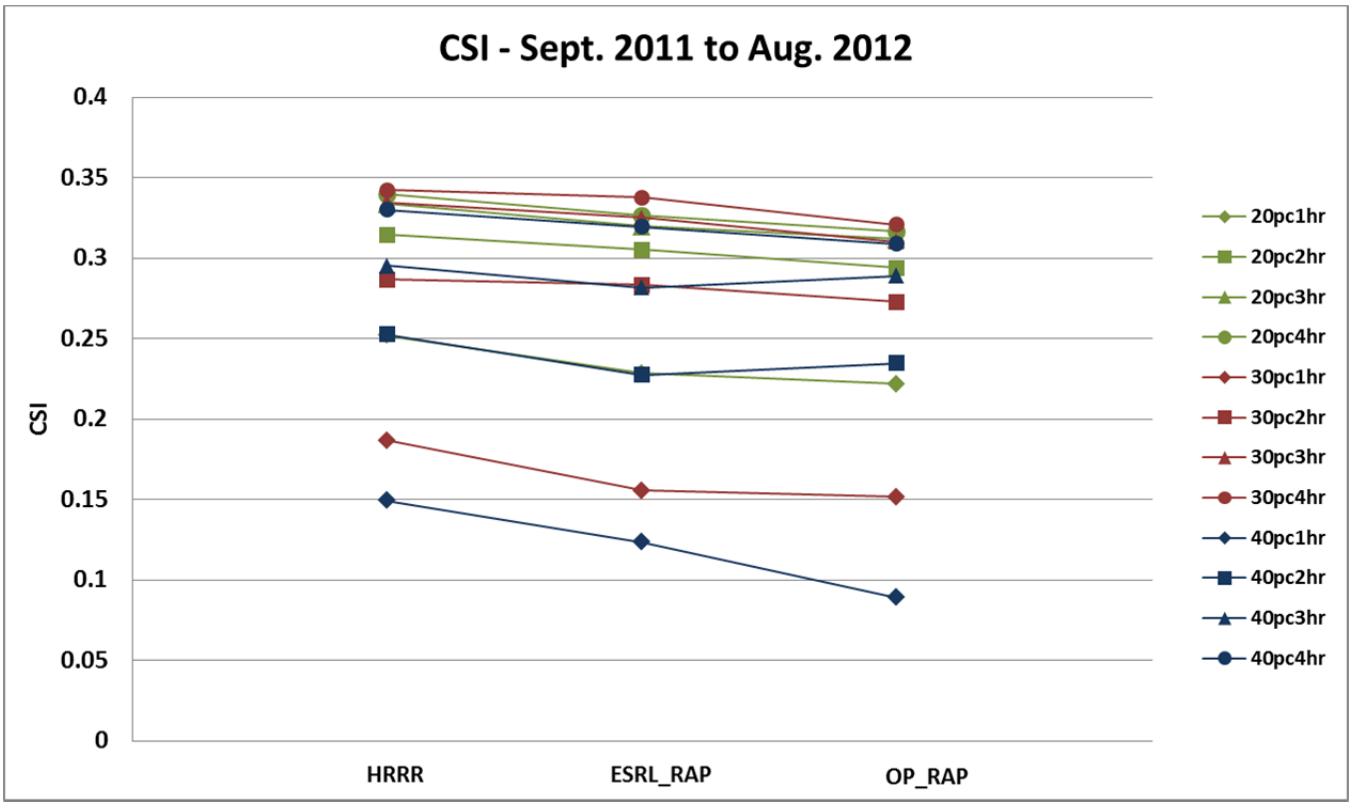

Figure 59: Critical Success Index (CSI) for all Crystal Lake Aggregate wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

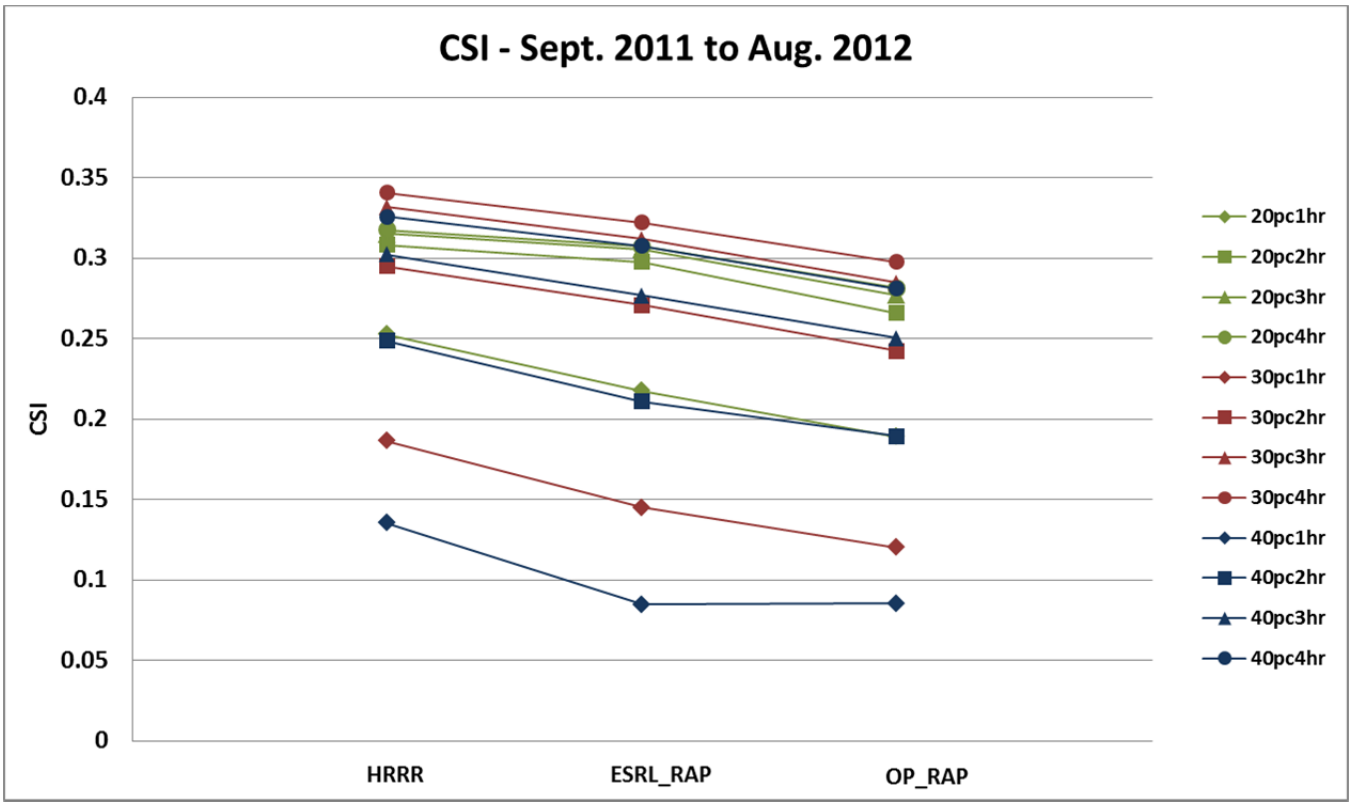

Figure 60: Critical Success Index (CSI) for all Oliver Aggregate wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

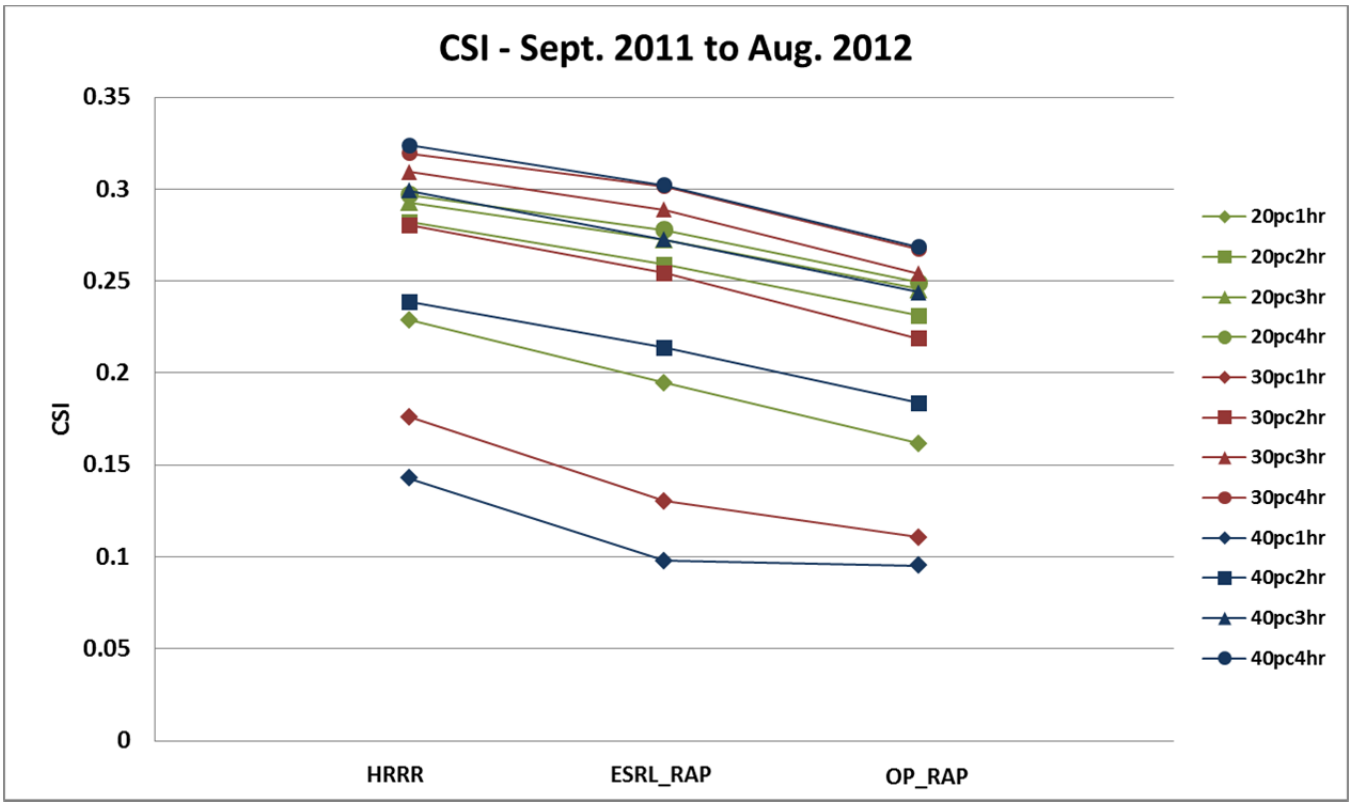

Figure 61: Critical Success Index (CSI) for all South Dakota Basin wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

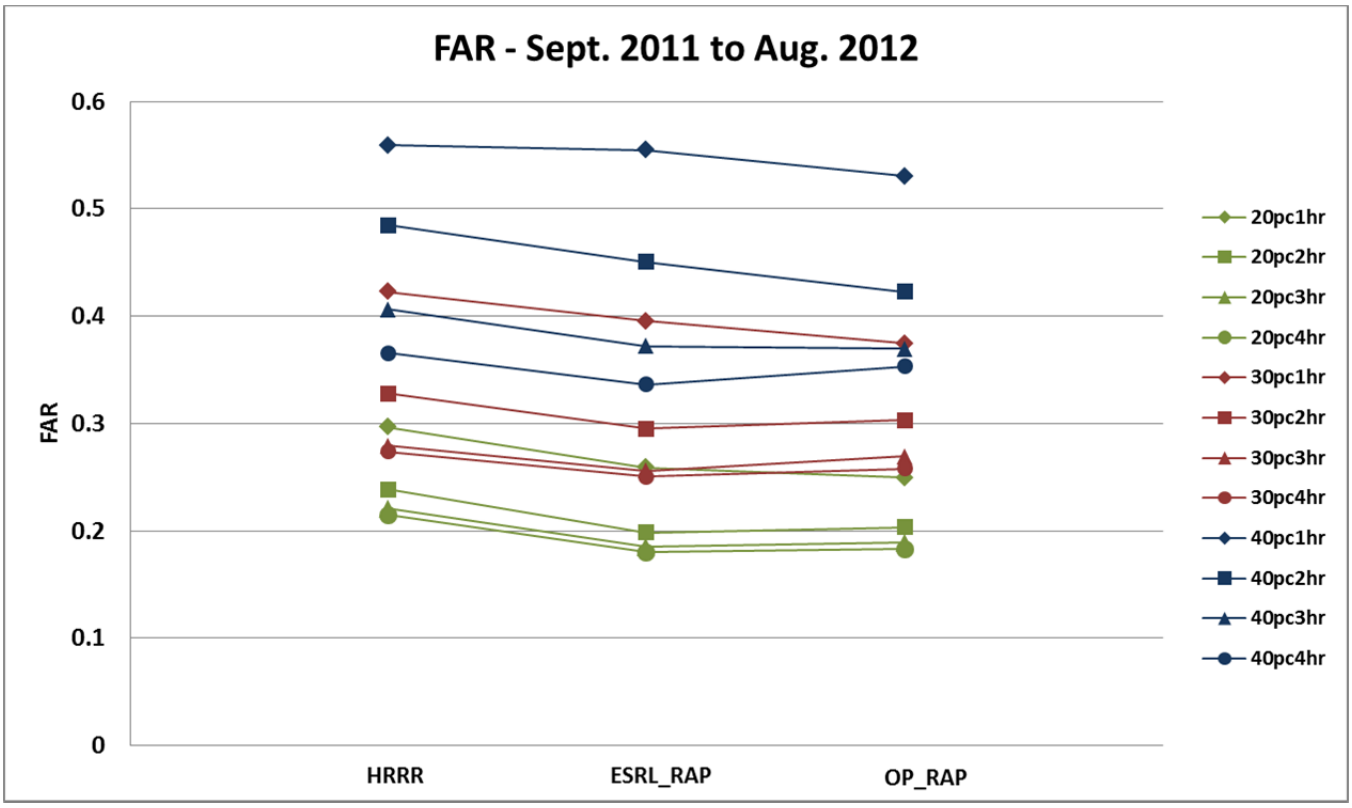

Figure 62: False Alarm Ratio (FAR) for all Ashtabula Aggregate wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

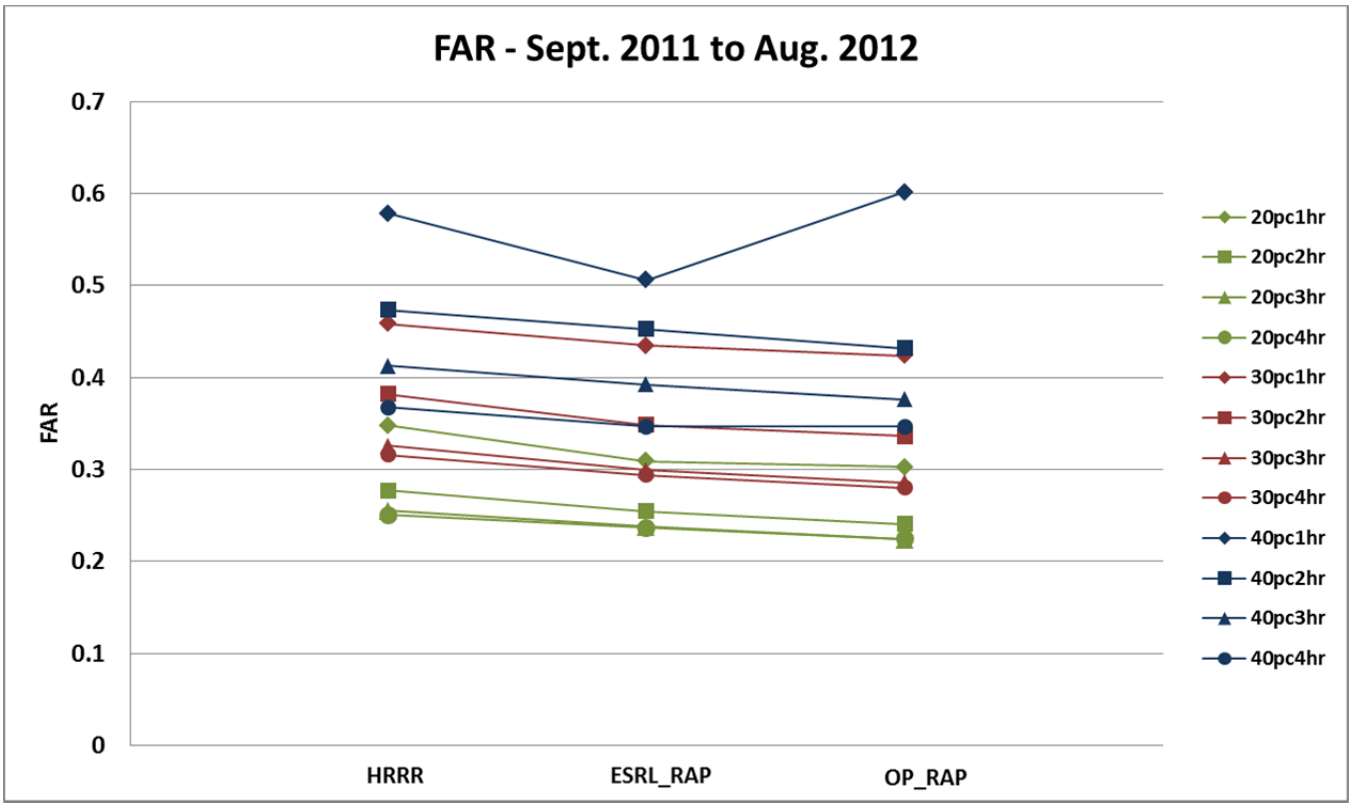

Figure 63: False Alarm Ratio (FAR) for all Crystal Lake Aggregate wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

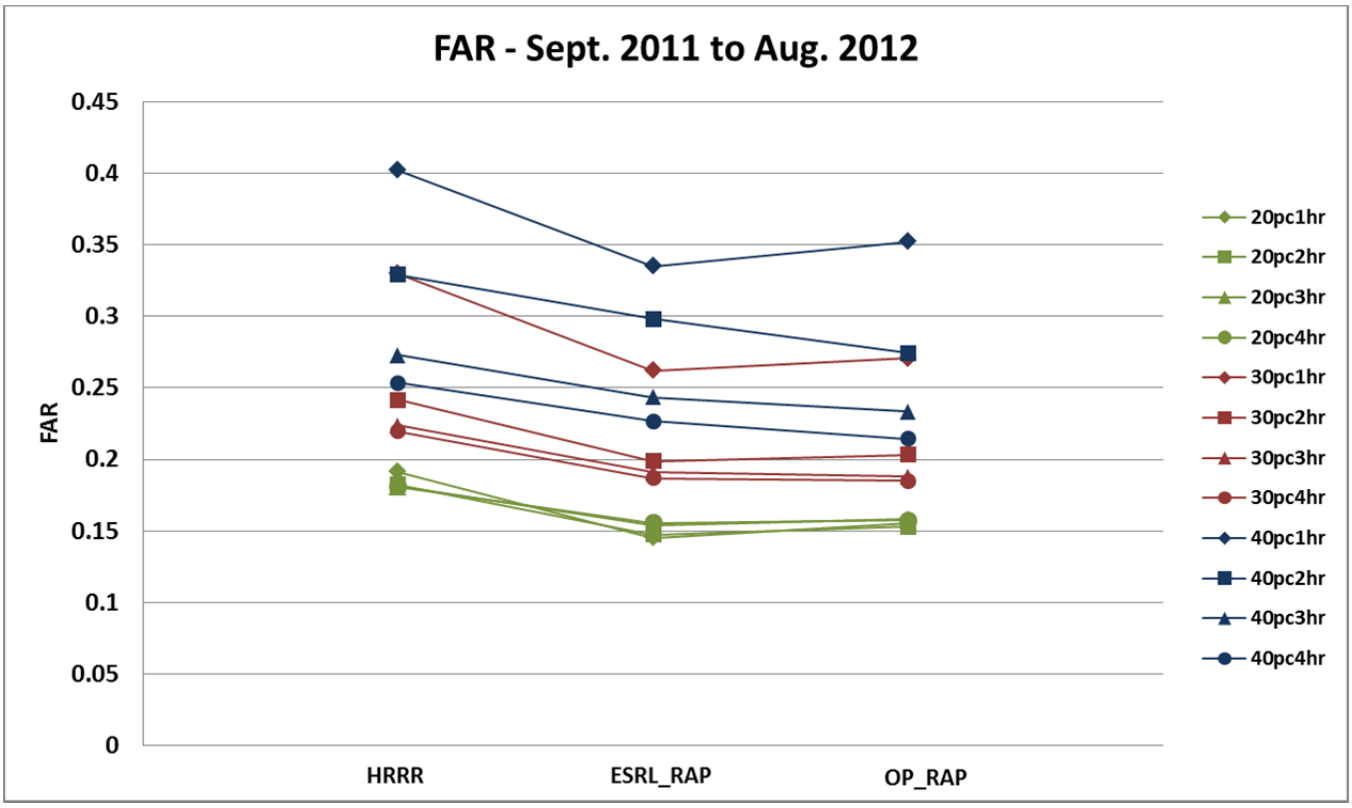

Figure 64: False Alarm Ratio (FAR) for all Oliver Aggregate wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

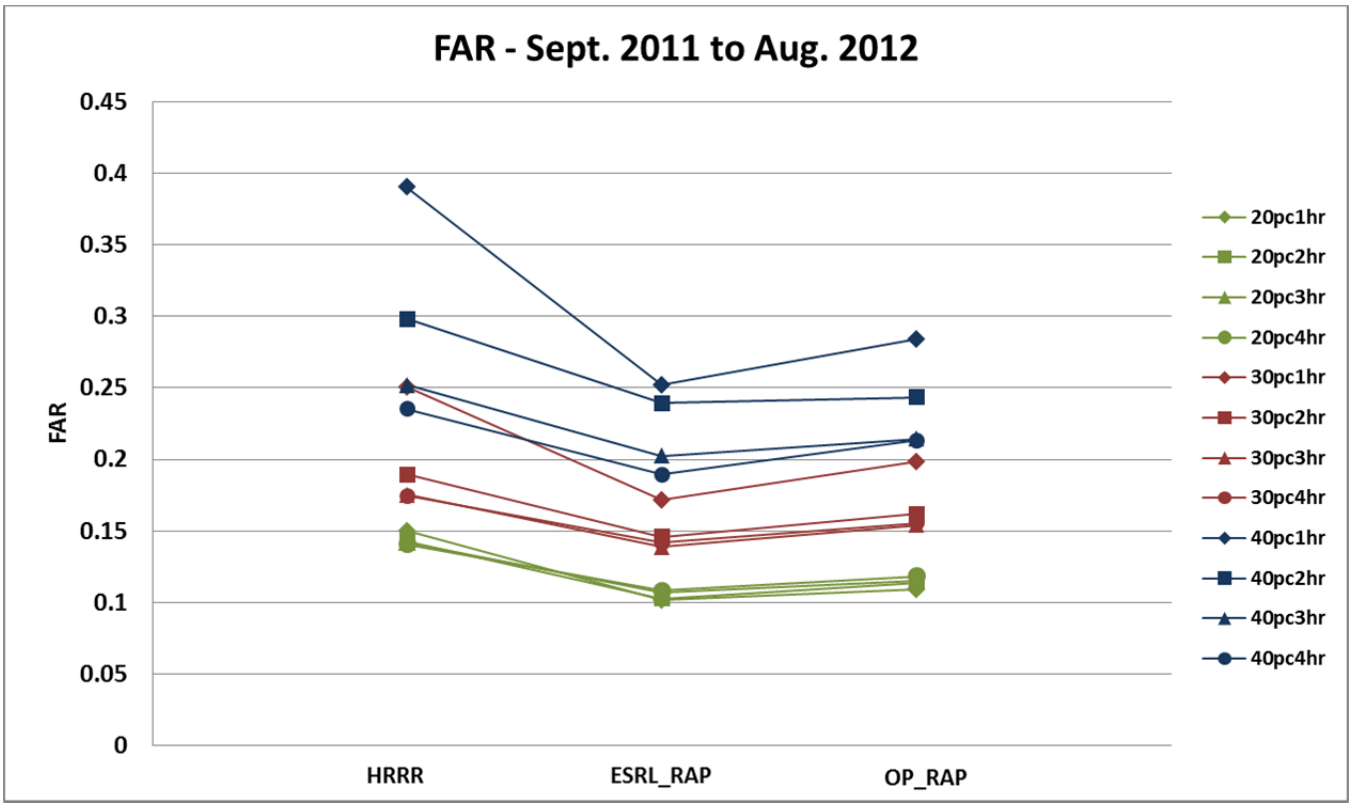

Figure 65: False Alarm Ratio (FAR) for all South Dakota Basin wind power ramp events during the entire forecast period (September 2011 – August 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined in the text.

Convective ramp events, in particular, are extremely difficult to predict accurately. To better gauge the accuracy of the various forecasts in predicting convective ramp events, an analysis similar to the one discussed above was performed for a time period only spanning the convective months (May – July, 2012). August is normally a convective month in the northern plains, but 2012 was a drought year in the central U.S., with all of the northern plains being affected by drought through the second half of July and August.

The Frequency Bias for a range of ramp definitions for the Ashtabula and Crystal Lake Aggregates for the convective season are shown in Figure 66 and Figure 67, respectively. A comparison of these results with the Frequency Bias calculated over the entire year (Figure 46 - Figure 47) shows that the Frequency Bias is generally higher (or nearly equal) for all model-based bias-corrected raw power forecasts for the convective months as compared to the entire year, but the difference is much larger for the HRRR-based forecasts than for the coarser resolution model forecasts, particularly for the large amplitude events (in fact the HRRR-based forecasts tend to overpredict the number of these events at many locations). Since the HRRR model can resolve convection to some extent (while the coarser resolution models only parameterize it), this is another indication that the larger magnitude, shorter duration events are likely convective in nature.

Another indicator that the HRRR-based forecasts more accurately predict convective ramp events than the coarser resolution model-based forecasts as can be seen in Figure 68 - Figure 71, which show a comparison of the Recall metric for various high amplitude/short duration ramp definitions at the four wind plant sites discussed previously. The Recall values (i.e., the fraction of observed events that are predicted correctly) for the HRRR-based forecasts are larger for the convective months compared to

the entire year time period at most sites, while the Recall values for the coarser resolution model-based forecasts often decrease compared to the entire year time period for most high amplitude/short duration ramp definitions. It should be noted that these results are not universally true at all sites (the South Dakota Basin site shown in Figure 71 being the one these examples), but the results do hold at most sites for the majority of the potential convective ramp event definitions. While the HRRR_BCRW power forecasts also have the highest Critical Success Index scores of all the power forecasts for the high amplitude/short duration events, improvements in the ramp event forecast accuracy seen in the convective season don't always translate into better CSI scores as the False Alarm Ratio (FAR) usually also increases during the convective months for high amplitude/short duration ramp events. This can be seen by comparing the CSI and FAR values for the Ashtabula and Crystal Lake sites (Figure 72 with Figure 73, and Figure 74 with Figure 75). Notice that all metrics tend to be poorer for the high amplitude/short duration ramp events compared to the metrics for smaller amplitude/longer duration events, indicating that shorter-duration events (in which convective events will fall) are still extremely difficult to forecast, but the HRRR-based forecasts provide a substantial improvement in the forecasts of these events.

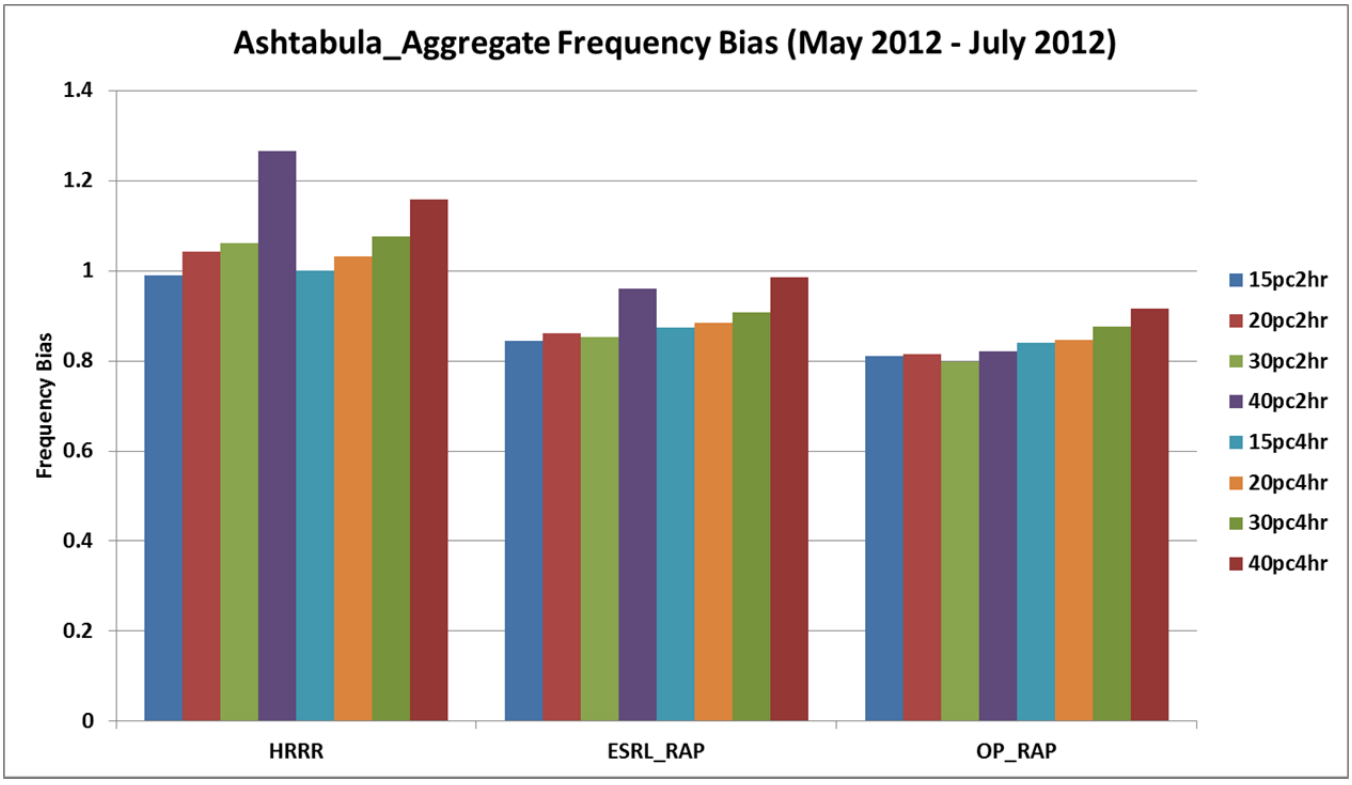

Figure 66: Ashtabula Aggregate Frequency Bias for all wind power ramp events during the warm season convective time period (May 2012 – July 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined above.

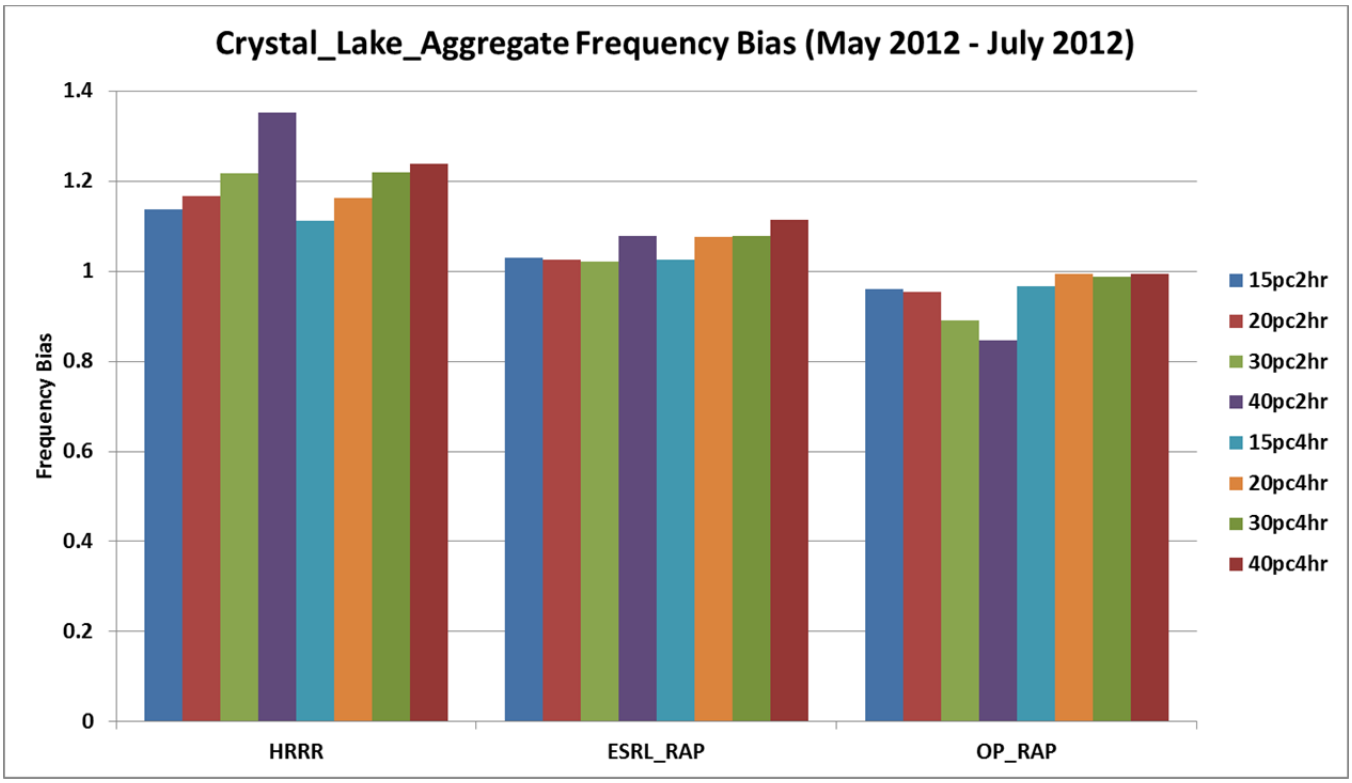

Figure 67: Crystal Lake Aggregate Frequency Bias for all wind power ramp events during the warm season convective time period (May 2012 – July 2012) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 12 different ramp definitions as defined above.

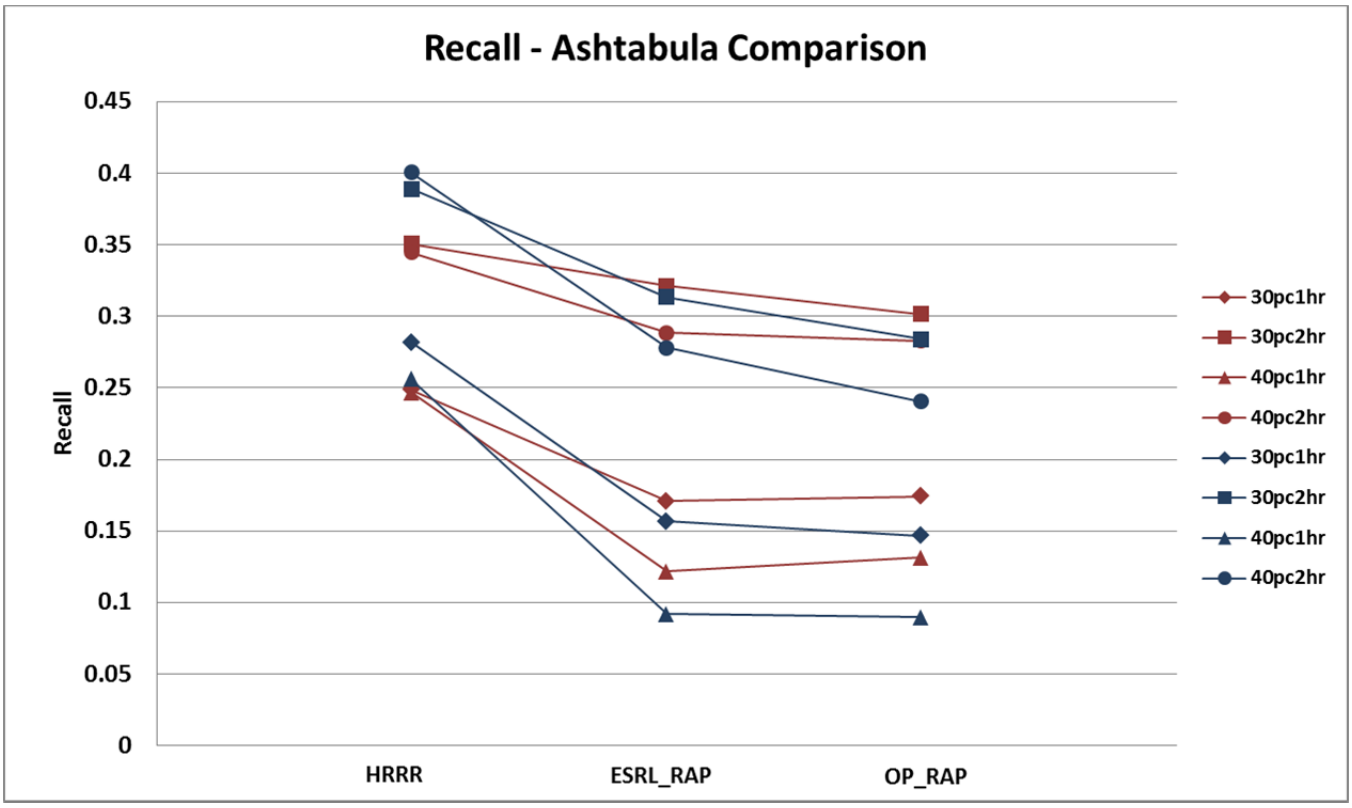

Figure 68: Recall comparison for Ashtabula Aggregate wind power ramp events during the entire year (red) and the warm season convective time period (May 2012 – July 2012, blue) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 4 different high amplitude/short duration ramp definitions as defined in the text.

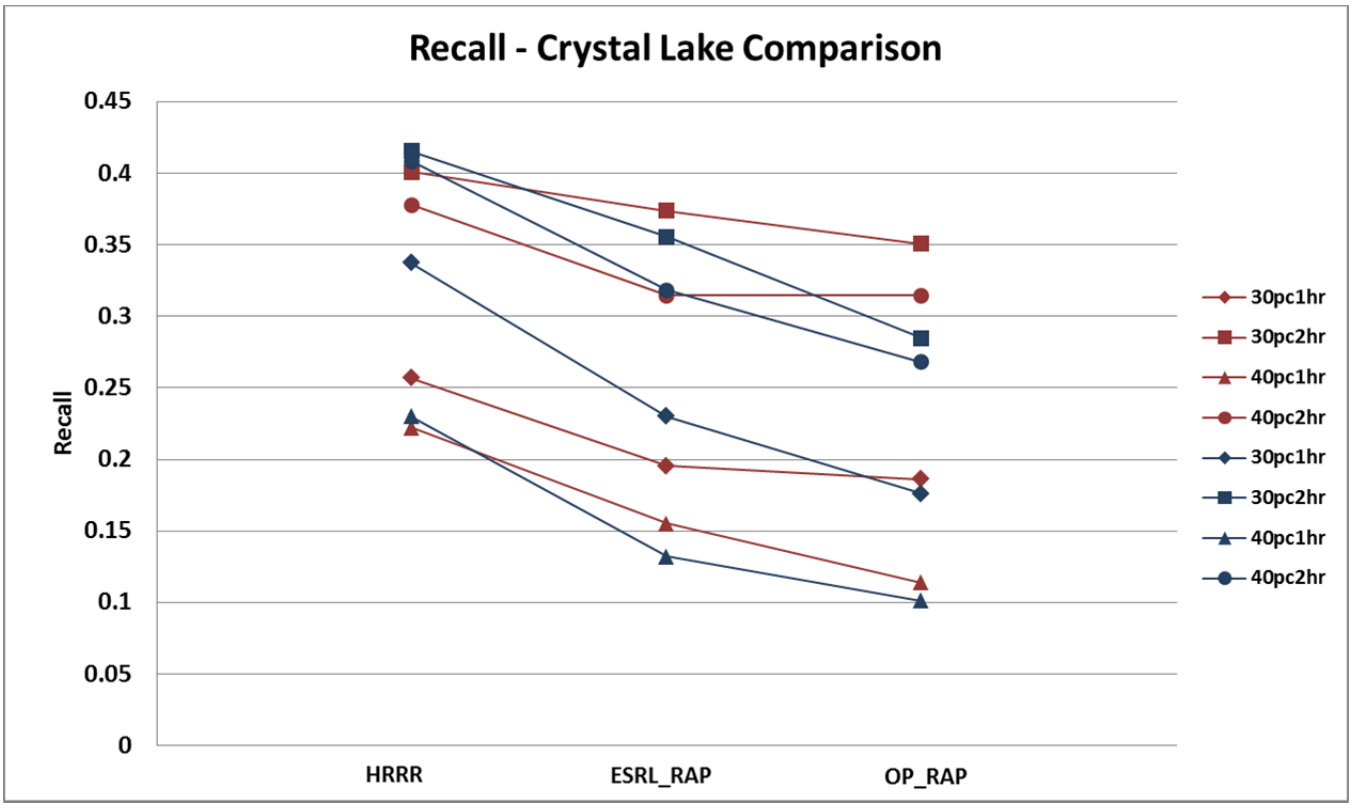

Figure 69: Recall comparison for Crystal Lake/Hancock Aggregate wind power ramp events during the entire year (red) and the warm season convective time period (May 2012 – July 2012, blue) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 4 different high amplitude/short duration ramp definitions as defined in the text.

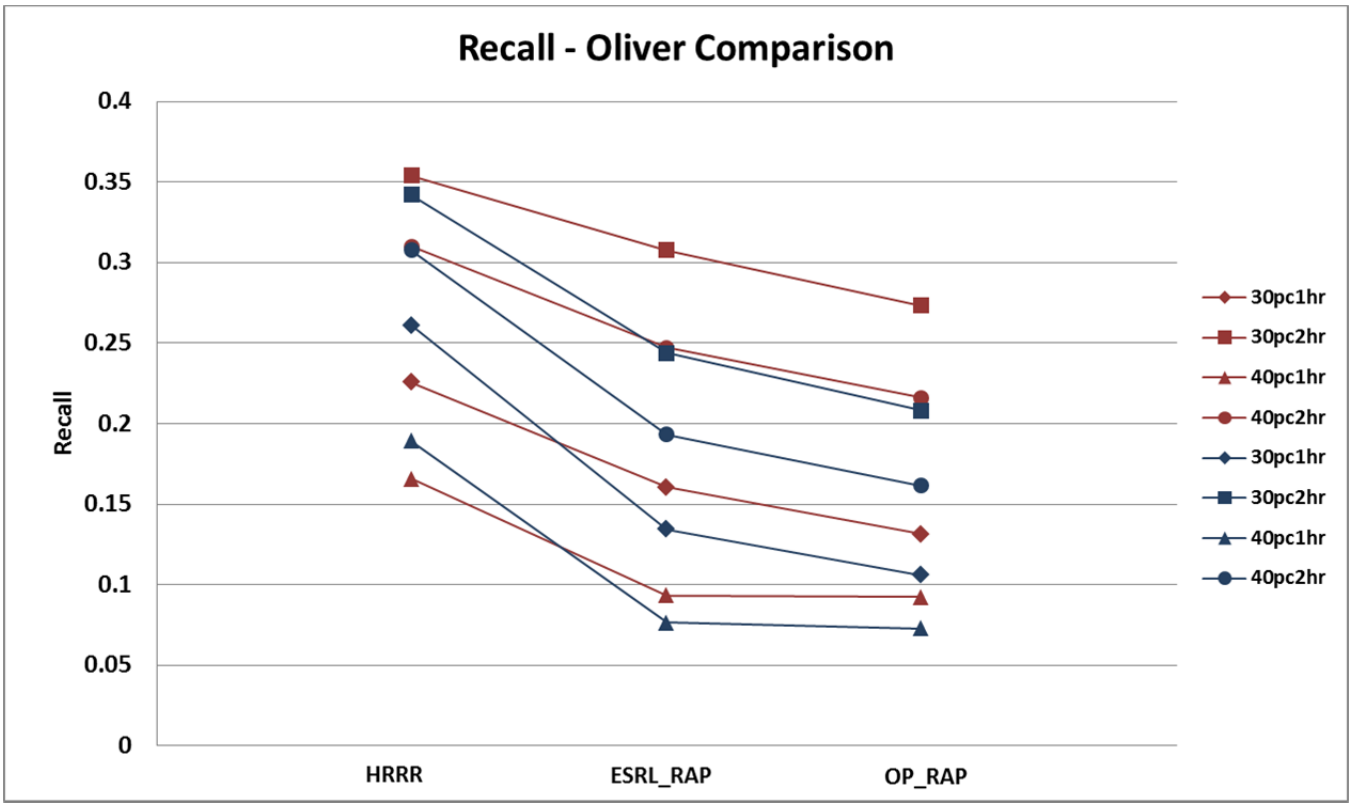

Figure 70: Recall comparison for Oliver Aggregate wind power ramp events during the entire year (red) and the warm season convective time period (May 2012 – July 2012, blue) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 4 different high amplitude/short duration ramp definitions as defined in the text.

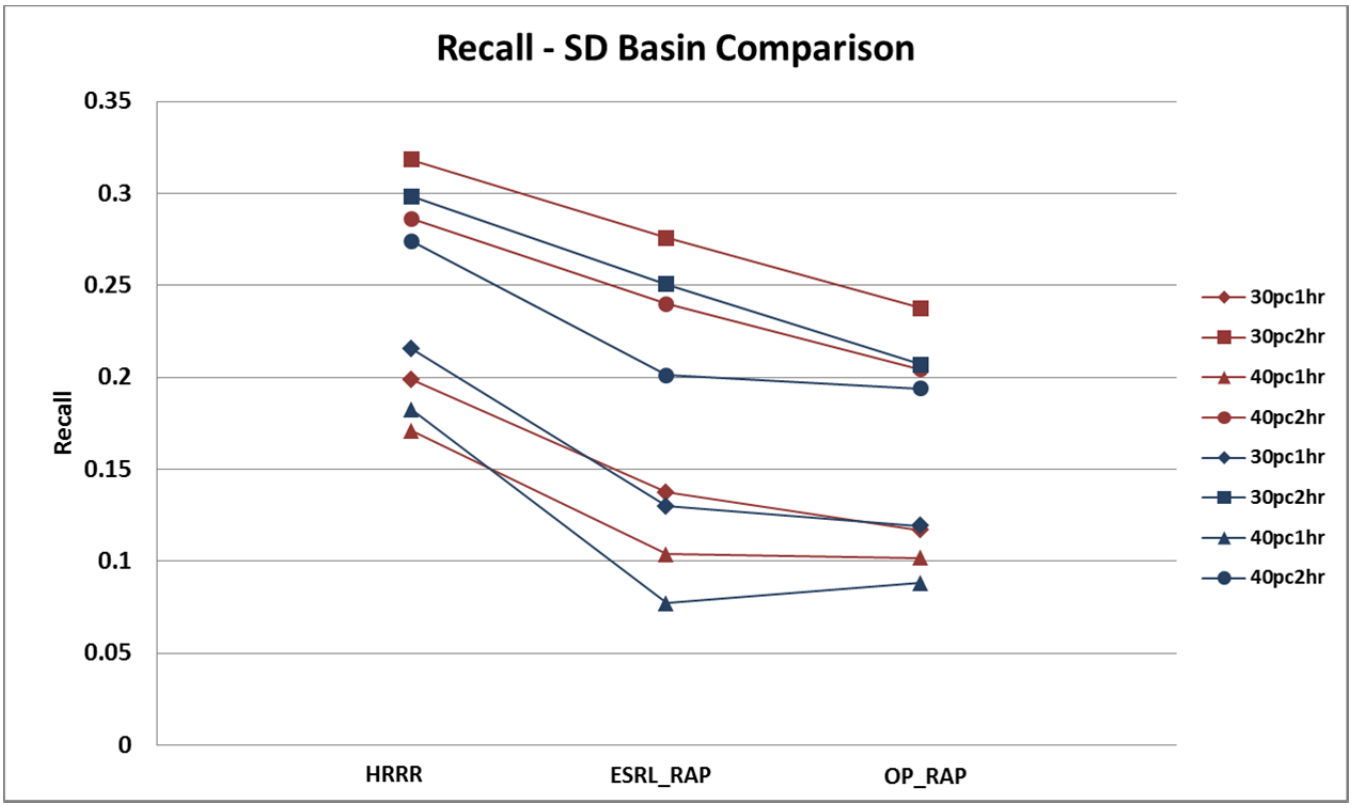

Figure 71: Recall comparison for South Dakota Basin wind power ramp events during the entire year (red) and the warm season convective time period (May 2012 – July 2012, blue) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 4 different high amplitude/short duration ramp definitions as defined in the text.

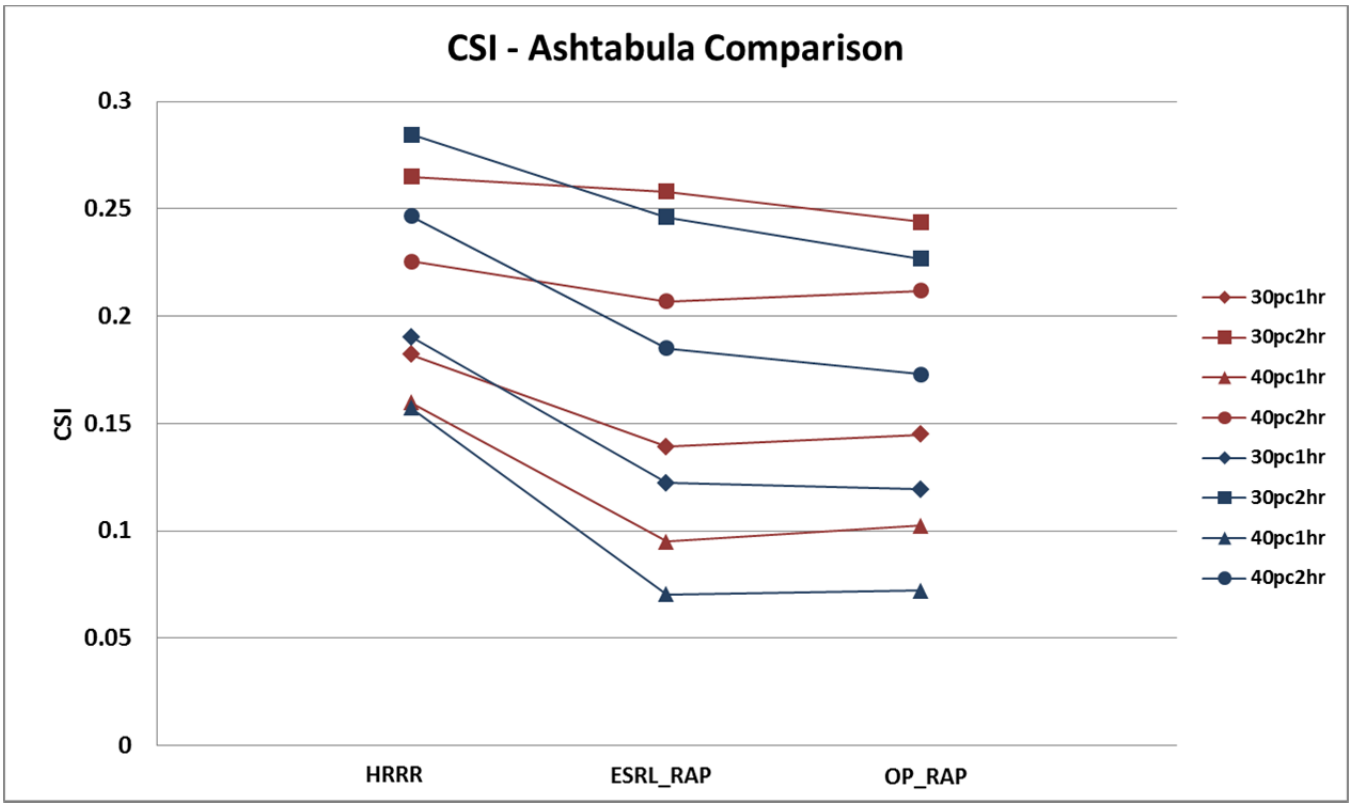

Figure 72: Critical Success Index (CSI) comparison for Ashtabula Aggregate wind power ramp events during the entire year (red) and the warm season convective time period (May 2012 – July 2012, blue) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 4 different high amplitude/short duration ramp definitions as defined in the text.

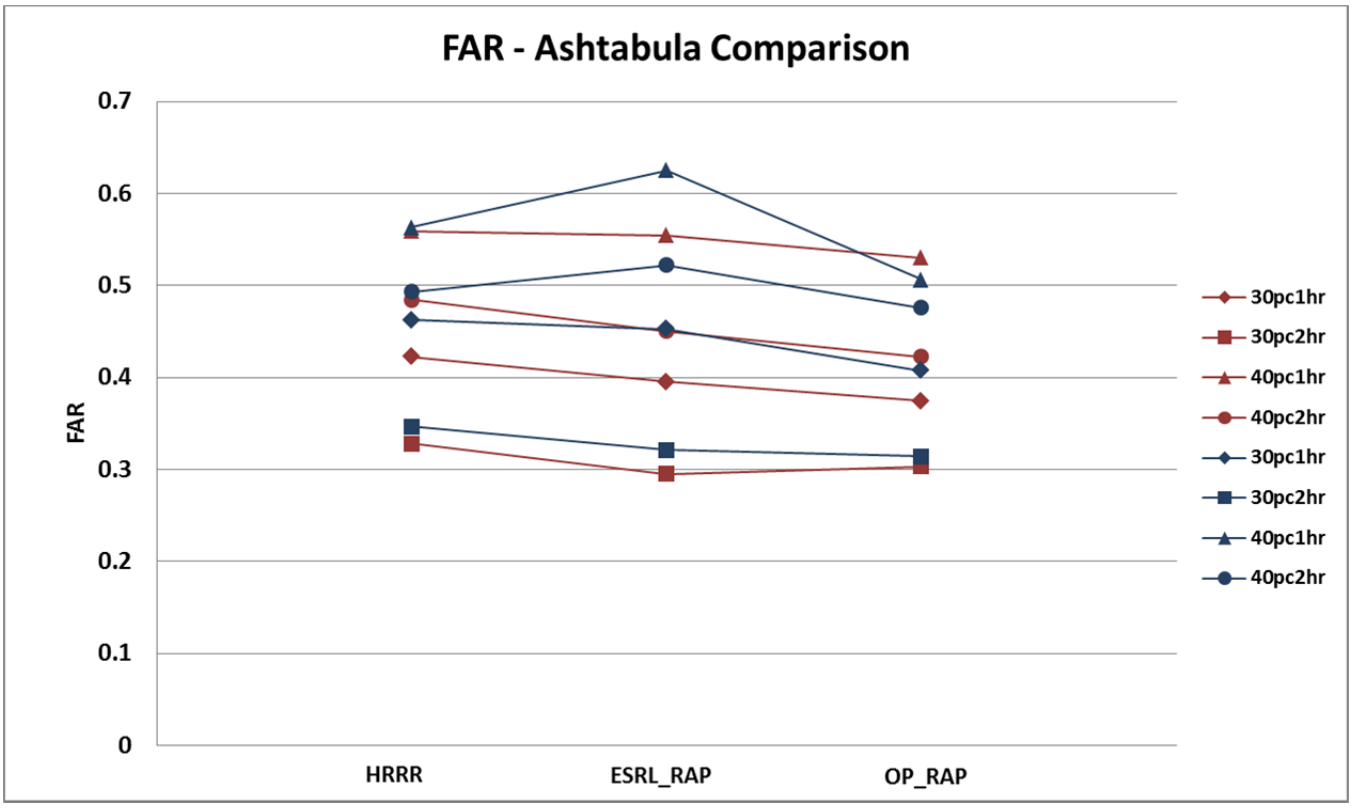

Figure 73: False Alarm Ratio (FAR) comparison for Ashtabula Aggregate wind power ramp events during the entire year (red) and the warm season convective time period (May 2012 – July 2012, blue) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 4 different high amplitude/short duration ramp definitions as defined in the text.

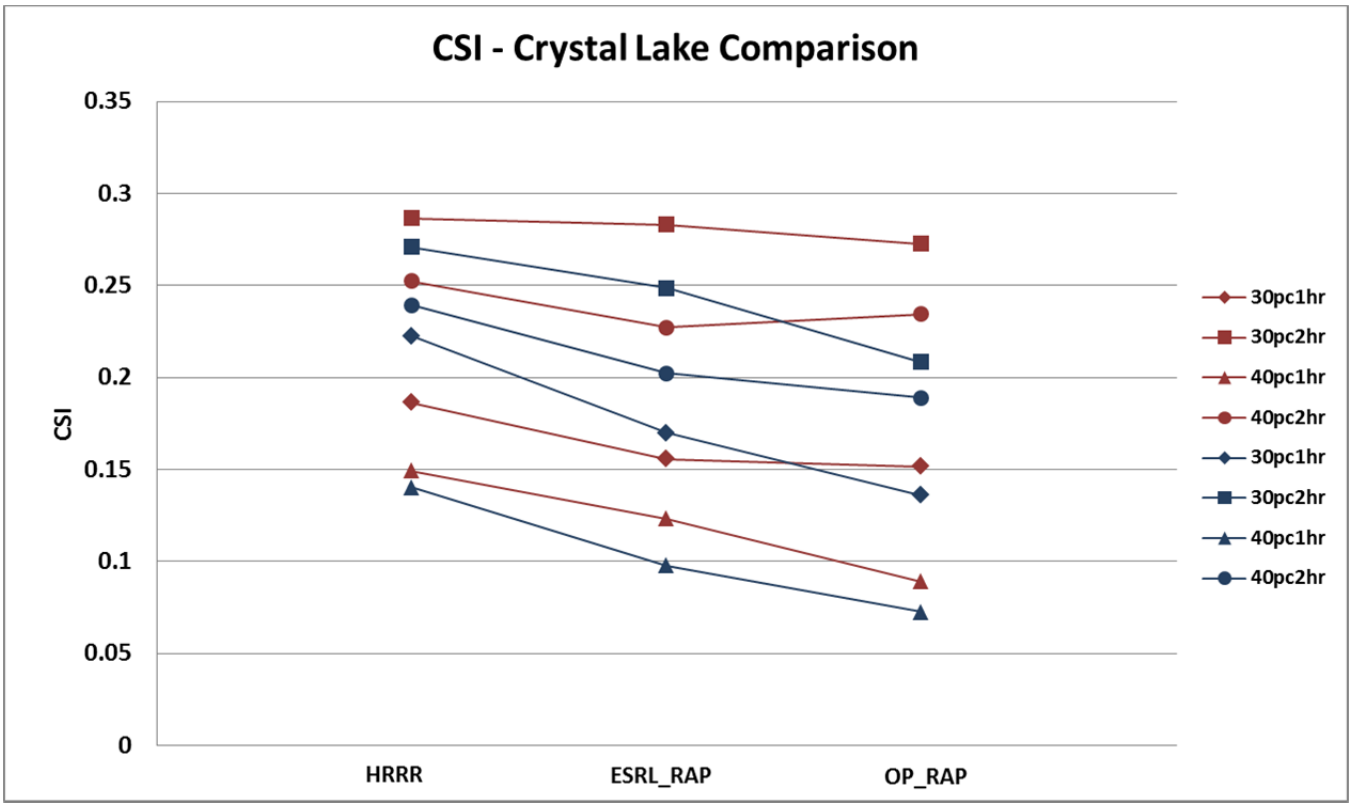

Figure 74: Critical Success Index (CSI) comparison for Crystal Lake/Hancock Aggregate wind power ramp events during the entire year (red) and the warm season convective time period (May 2012 – July 2012, blue) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 4 different high amplitude/short duration ramp definitions as defined in the text.

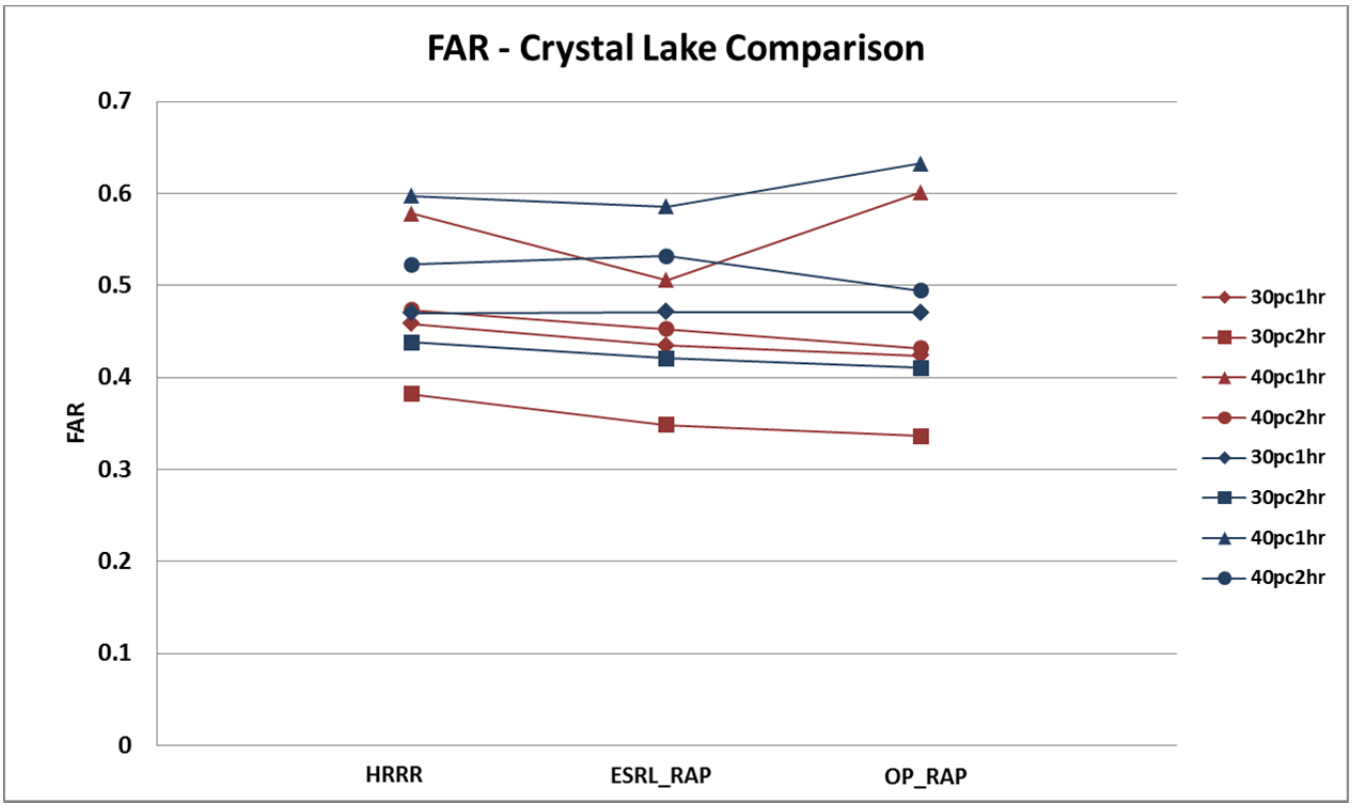

Figure 75: False Alarm Ratio (FAR) comparison for Crystal Lake/Hancock Aggregate wind power ramp events during the entire year (red) and the warm season convective time period (May 2012 – July 2012, blue) for the bias-corrected raw HRRR, ESRL_RAP, and OP_RAP-based power forecasts. Results are shown for 4 different high amplitude/short duration ramp definitions as defined in the text.

The results presented above show that the bias-corrected raw power forecasts based on the research models (ESRL_RAP, HRRR) more accurately predict wind energy ramp events than the current operational forecast model, both at the system aggregate level and at the local wind plant level. At the system level, the ESRL_RAP-based forecasts most accurately predict both the number of ramp events, and the occurrence (within a time window of the actual event) of the ramps themselves, but the HRRR-based forecasts more accurately predict the ramp rate which is of particular concern to system operators. There was little difference between the various model-based power forecasts in the accuracy of the timing of the ramp events. In general, up-ramp events are more accurately predicted than down-ramp events, both at the system level and individual wind plant level (plant-level results were not shown) for all model-based forecasts. At the individual site (or local aggregate site) level, the HRRR-based forecasts most accurately predict the actual ramp occurrence (again within a time window of the actual event), the total number of ramps and the ramp rates. The HRRR-based forecasts also most accurately predict both the number and occurrence of potentially convective high amplitude/short duration events during the convective season, but tend to over-predict the number of these potentially convective events leading to a higher number of false alarms than the other model-based power forecasts. However, even though the research models show improvement in predicting wind energy ramp events, the results presented above indicate that continued work is needed to improve wind energy ramp forecasts.

4.5 Effects of Geographic Dispersion of Wind Plants on Forecast Accuracy

When individual wind plant forecasts are aggregated together, there is the potential for forecast errors to 'cancel out' to some extent, thus providing a more accurate aggregate forecast. Although this is not a direct measure of forecast accuracy, it is of relevance to system operators who make planning decisions based on the total expected wind plant output in various regions and the entire operating area. The extent to which forecasts errors potentially cancel between wind plants depends greatly on the geographic dispersion of the wind generation (i.e., how close the wind plants are to each other), the meteorology of the region, and the complexity of the terrain. The large footprint of the MISO system and the large geographic area over which the wind generation is spread has the potential for a significant cancelation of forecast errors when aggregating all the individual wind plant forecasts together.

To investigate the extent to which forecast errors are reduced in the aggregate wind power forecast compared to individual wind plant forecasts, six different aggregates were created with varying levels of aggregation from the local level up to the full system level (as represented by the NextEra wind plants) as shown in Table 9. The Ashtabula 1 wind plant was used as a starting location, and additional wind plants were added by the dual considerations of geographic proximity to the existing aggregate wind plants and capacity. In each aggregate, additional generation of between 300 and 400 MW was generally added. It should be noted that this exercise is not meant to be a rigorous scientific experiment, but rather to show how aggregation can reduce aggregate error in a real-world setting.

Aggregate	Wind Plants	Rated Capacity (MW)
Ashtabula1	Ashtabula1	196.5
Ashtabula_Agg	Ashtabula 1 + Ashtabula 2,3	428.4
Aggregate1	Ashtabula_Agg + Langdon 1,2 + Edgeley Basin + Day County	747.9
Aggregate2	A1 + Wilton 1,2 + Baldwin + Oliver 1,2	1047.9
Aggregate3	A2 + S.D. Basin + Wessington + Lake Benton 2 + Endeavor 1,2	1392.9
Aggregate4	A3 + Crystal Lake 1,2,3	1808.9
System_Agg	A4 + Hancock + Mower + Story County 1	2155.48

Table 9: Wind plants used to create aggregate forecasts for various levels of wind plant aggregation within the study area.

4.5.1 Effects on Forecast Bulk Error Statistics

To look at the effects of wind plant forecast aggregation on the forecast error statistics, we calculated a 'Mean Absolute Error' (MAE) and a 'Mean Total Absolute Error' (MTAE) for each of the wind plant aggregates and compared them. 'Mean Absolute Error' and a 'Mean Total Absolute Error' are defined as follows:

$$\boldsymbol{Mean\ Absolute\ Error\ (aggregate) = avg\left(\sum_{\#\ fx\ values} abs\left(\sum_{\#\ sites} (forecast - actual)\right)\right)}$$

Equation 10

$$\boldsymbol{Mean\ Absolute\ Total\ Error(aggregate) = avg}\left(\sum_{\#\ \boldsymbol{fx\ values}}\left(\sum_{\#\ \boldsymbol{sites}}\boldsymbol{abs(forecast - actual)}\right)\right)$$

Equation 11

The MAE is calculated the way one normally calculates forecast MAE, but when individual wind plant forecasts are aggregated together, the forecast errors can cancel each other out in the aggregate forecast. The Mean Total Absolute Error (MTAE) is calculated by taking the absolute value of the individual wind plant error first, and then summing, so there is no potential for the forecast errors to cancel out when summed together. Note that MAE will equal MTAE for a single wind plant, and the ratio of MAE/MTAE a measure of forecast error cancelation when multiple wind plants are aggregated together. The calculations were done for the entire year of forecasts (forecast hours 0-15) and the results for the various model-based bias-corrected raw power forecasts are shown in Table 10.

MAE/MTAE	HRRR_BCRW	ESRL_RAP_BCRW	OP_RAP_BCRW
Ashtabula1	1.00	1.00	1.00
Ashtabula_Aggregate	0.95	0.95	0.95
Aggregate1	0.73	0.72	0.73
Aggregate2	0.64	0.62	0.63
Aggregate3	0.54	0.52	0.53
Aggregate4	0.52	0.50	0.51
System_Aggregate	0.49	0.47	0.49

Table 10: The ratio between the Mean Absolute Error (MAE) and Mean Total Absolute Error (MTEA) for each the model-based bias-corrected raw wind power forecast. The ratio of MAE/MTAE is a measure of forecast error cancelation in the aggregate wind power forecasts as compared to individual wind plant forecasts.

As can be seen in Table 10, the results are very similar between the various model-based bias-corrected power forecasts at the various levels of aggregation. There is some benefit in aggregating even at the local level as can be seen comparing the results for Ashtabula1 and the Ashtabula aggregate, where there is a 6% reduction in forecast error from aggregation, although the reduction in forecast error from local aggregation will likely vary by location.

The largest reduction in forecast error from aggregation occurs when the geographical diversity of the added wind plant locations is largest. At the system aggregate level (defined by the NextEra wind plants in the study area), the ratio of MAE/MTAE is approximately 0.47 to 0.49, so the system-wide wind energy forecast errors are roughly half of what would be expected at a typical wind plant. This is a significant reduction in forecast error that results from the geographic dispersion of the wind plants (the further apart the wind plants are located from each other, the less correlated the forecast errors are likely to be). Forecast error also grows in time as the forecast progresses, and the forecast reduction does vary somewhat depending on forecast hour, with a 3% to 5% greater reduction in the aggregate forecast error at forecast hour 15 compared to the start of the forecast in the system aggregate forecasts. It should also be noted that the amount of reduction in forecast error due to aggregation is geographically dependent, and one would expect a greater reduction in aggregate forecast error in regions where the forecast errors at the individual wind plants are less correlated as long as the forecasts are unbiased. In regions such as the trade wind belts where there is little geographic

variability in the winds (unless terrain is a factor), one would expect to see less reduction of forecast error due to aggregation as compared to midlatitudes for a given distribution of geographic diversity in wind plant locations.

4.5.2 Effects on Frequency and Magnitude of Ramp Events and Ramp Forecast Accuracy

In addition to the potential cancelation of forecast errors as individual wind plants are aggregated up to the system level, the geographic dispersion of wind plants in the operating area also plays a significant role in the number and the characteristics of system-wide wind energy ramp events. In small operating areas where the wind plants are all located in close proximity, both smaller-scale weather features (such as a thunderstorm outflows or fronts) and larger-scale weather features (such as development and movement of low and high pressure systems) can create significant system-wide ramp events since all the wind plants will be experiencing similar atmospheric conditions. In larger operating systems where there is significant geographic diversity in the wind production, smaller-scale atmospheric features will only affect a portion of the wind production at any given time, and even if these atmospheric features sweep across the operating area, the system-wide production will change at a slower rate (as a fraction of the capacity) than for systems where wind production is geographically concentrated. In general, the larger and longer-lived a particular atmospheric feature, the better it can be predicted, so although large-scale atmospheric features can create system-wide wind energy ramps even in power systems with wide geographic dispersion of the wind production, these events will likely be more predictable.

To illustrate the impact of geographic diversity of wind production on the number of wind energy ramp events, the number of ramp events was calculated (as seen in the observed power in the forecast time series) for each of the aggregates listed in Table 9 for two different ramp definitions as discussed in Section 4.4.2.1. The results over the entire year time period are shown in Figure 76. For each of the ramp definitions, the number of observed ramps in the forecast time series initially decreases significantly as the geographical diversity of the wind plant locations increases, then incremental additions eventually have a smaller impact as new wind plants are added. For the smaller of the ramp definitions (20 percent of capacity change over a two hour period or less), the reduction in number of ramps between the individual wind plant and the system aggregate is 88%, while for the larger ramp definition (30 percent capacity change over a two hour period or less) the number of ramps is reduced by 97% between the wind plant and system aggregate. For even larger amplitude ramp definitions, the number of ramps experienced quickly drops to zero as greater geographical diversity is added to the aggregate wind power production (not shown). While larger wind energy ramps are more difficult to predict at the local level, these results illustrate the that there are advantages to having a larger balancing area with geographic diversity in wind energy production in that large system-wide ramp events are less likely than they are in a smaller balancing area with less geographic diversity in wind production.

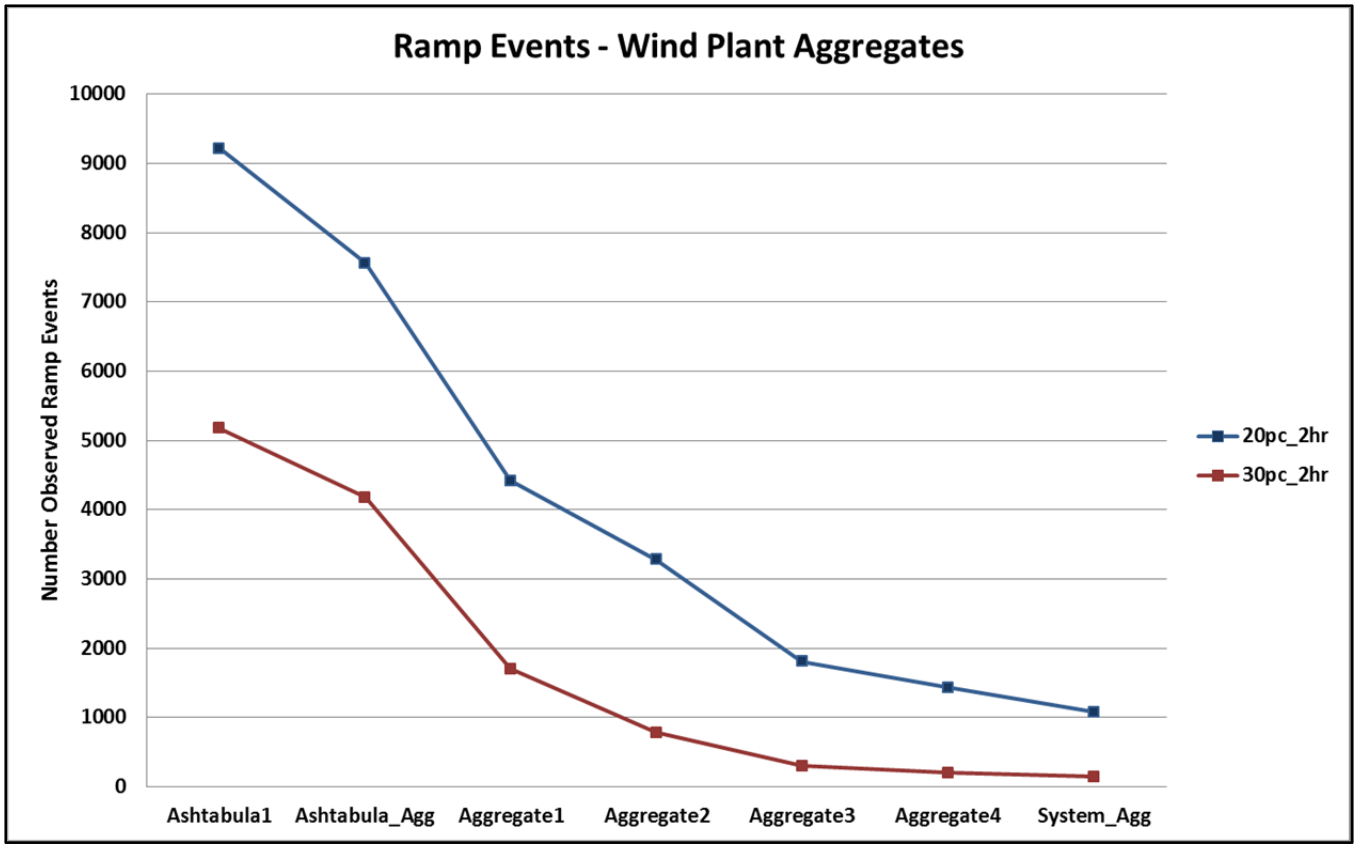

Figure 76: Number of observed ramp events in the forecast time series for two ramp definitions for each of the wind plant/aggregates defined in Table 9 over the one year duration of the forecasts.

4.6 Data Denial Experiments

4.6.1 Brief Description of Experiments

To assess the relative impacts of the additional data provided in this study being assimilated into the research weather forecast models separately from the improvements to the model physics, NOAA ran several data denial experiments in which select periods were 're-forecasted' using the ESRL_RAP weather forecast model. Two sets of forecasts were generated for each data denial period: one set of the forecasts was generated by initializing the forecast only using the operational meteorological data currently available (hereafter denoted the 'control forecast'), and one set was generated by initializing the forecast model using the currently available operational meteorological data plus the additional data collected as part of this project (hereafter denoted the 'experimental forecast'). All other aspects of the forecasts were identical.

Interesting ten day meteorological periods were identified in each season for the 're-forecasting' periods in the experiments. These time periods included:

Winter: 00Z November 30 – 23Z December 6, 2011 and 00Z January 7 – 23Z January 15, 2012
Spring: 00Z April 14 – 23Z April 25, 2012
Summer: 00Z June 9 – 22Z June 17, 2012
Fall: 00Z September 16 – 23Z September 25, 2011 and 00Z October 13 – 23Z October 21, 2011

For more a more detailed description of the data denial experiments, please refer to the companion NOAA WFIP report (Wilczak et al., 2014).

NOAA provided the weather forecast data for both the control and experimental forecasts, and WindLogics generated raw and bias-corrected raw power forecasts for both forecasts at all NextEra wind plants in the study area. An analysis of the relative impact of the additional data on the wind energy forecasts follows.

4.6.2 Forecast Error Results

To measure the impacts of the additional data provided in this study on the wind power forecast accuracy, a 'Percent Improvement' as defined in Equation 3 was calculated by comparing the RMSE of the control and experimental ESRL_RAP raw model-based power forecasts. The *raw power forecasts* were used in this analysis since the same base model was used to generate both forecasts, and the 0z and 12Z runs were included in the analysis since the same operational meteorological data sets were used for both the control and experimental forecasts. These results also differ from those shown previously in that the first six hours of the forecast are relative to when the forecast was generated (i.e. includes the model initialization time through hour six), not relative to when the forecasts were available for use.

The Percent Improvement (i.e. the average RMSE difference between the control and experimental forecast normalized by the average energy produced) over the first six hours of the forecast for the NextEra wind plants in the study area for all data denial periods is shown in Figure 77. Also shown in Figure 77 are the locations of the operational wind profilers, and the additional wind profilers, sodars, and SDSU meteorological towers provided in this study. The results show that the assimilation of the additional data improves the wind power forecasts at all locations (with the exception of Mower County at the eastern edge of the study area), and that the forecasts are most improved in North Dakota and South Dakota where the operational observational network is sparse. The average improvement at sites in North Dakota is 3.41%, with 1.41% average improvement for sites in South Dakota, and 0.57% average improvement for sites in Iowa and southern Minnesota (1.06% without including Mower County in the calculation). The largest improvement of 5.5% occurred at the Ashtabula site which is within approximately 15 miles of one of the additional tropospheric wind profilers deployed during the project. It's not clear why the error statistics at the Mower County site do not show improvement, but there are several possible reasons for this. The site is at the eastern edge of both the operational and experimental wind profiler networks. Depending on the weather patterns during the data denial study periods, one would expect the observations to have less impact on the forecasts at the edges of the study domain, particularly if the area is not directly downstream from the observational network. The Mower County site is also surrounded to the north, west, and south by other wind plants in very close proximity, which introduces the possibility of waking effects from the surrounding plants reducing the wind speed at the site. It's possible under such conditions that by *improving* forecasts of free-stream winds, it could make the power forecasts at Mower County site worse if the original forecasts had a low wind speed bias to start with. To conclusively test this idea, one would need to have access to free-stream (i.e. outside the possible influences of all surrounding turbine wakes) meteorological towers.

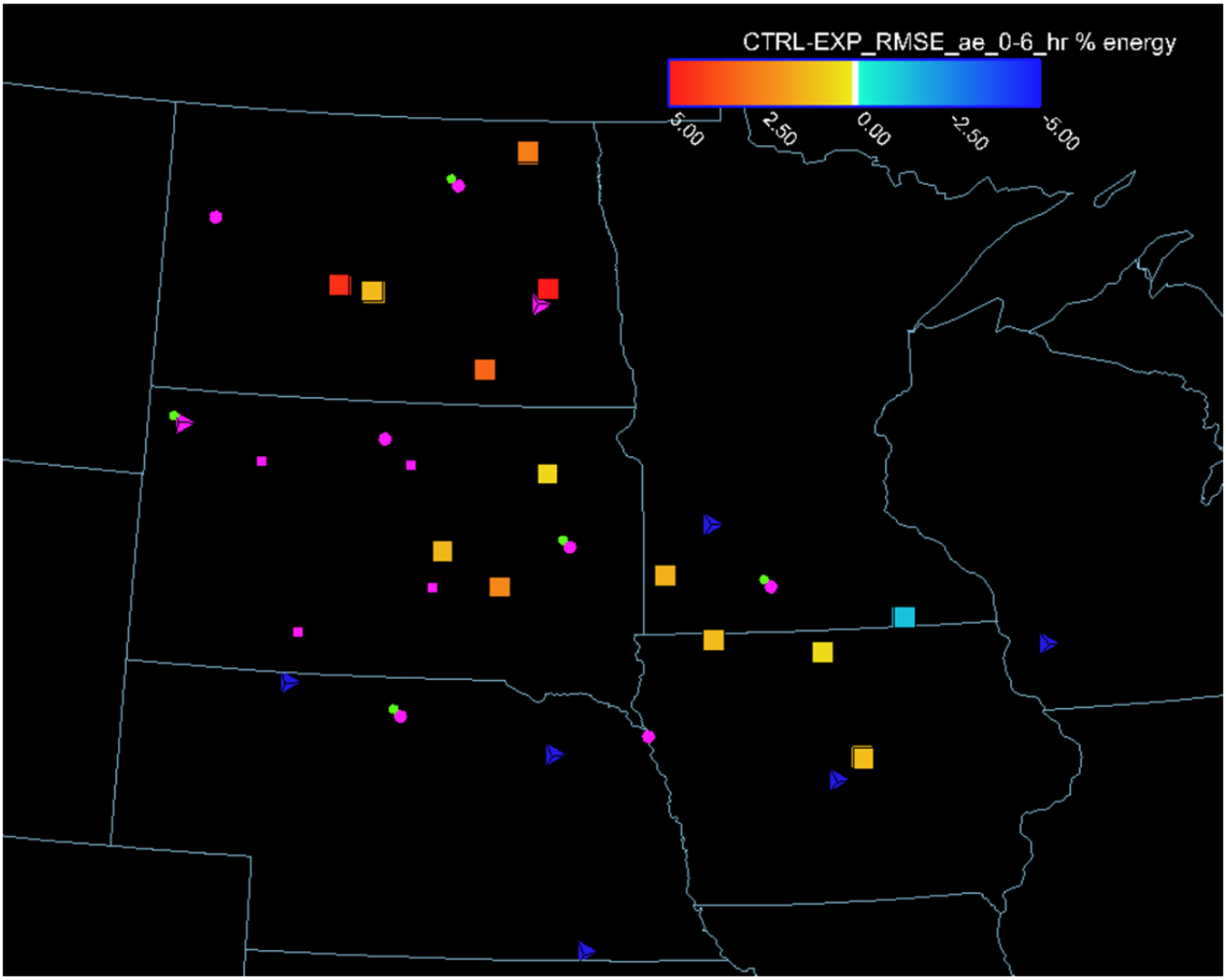

Figure 77: RMSE difference between the control and experimental ESRL_RAP raw power forecasts normalized by the average power produced over the first six hours of the forecast at the NextEra wind plant locations (large squares) for all data denial periods. Warm (cool) colors indicate that the experimental (control) forecast is performing better. The operational wind profiler locations are shown as blue tetrahedrons. The tropospheric (boundary layer) profilers deployed for this study are shown as bright pink tetrahedrons (circles), sodar locations are shown as green circles (located at the same sites as the wind profilers), and the SDSU tall towers are shown as bright pink small squares.

Results from the individual seasons are shown in Figure 78 - Figure 81. As can be seen from a comparison of these figures, most of the sites with greatest gains in forecast accuracy in each of the seasons due to the assimilation of the additional data are in the northern half of the study area. In each data denial study period, there appear to be one or two 'outliers' in the general pattern of forecast improvements. It's important to note that with the relatively small sample size in each of the data denial periods for the individual seasons, one short period of relatively large forecast error can skew the bulk error statistics which may account for some of the inconsistencies in the general overall patterns.

The largest overall forecast improvements from the additional data assimilation over the study area occur during the fall study period as shown in Figure 78. All sites (with the exception of the Wilton/Baldwin aggregate site which has a 'Percent Improvement' of -0.7%) show forecast improvements over the first six hours of the forecasts. The largest average improvements are seen at

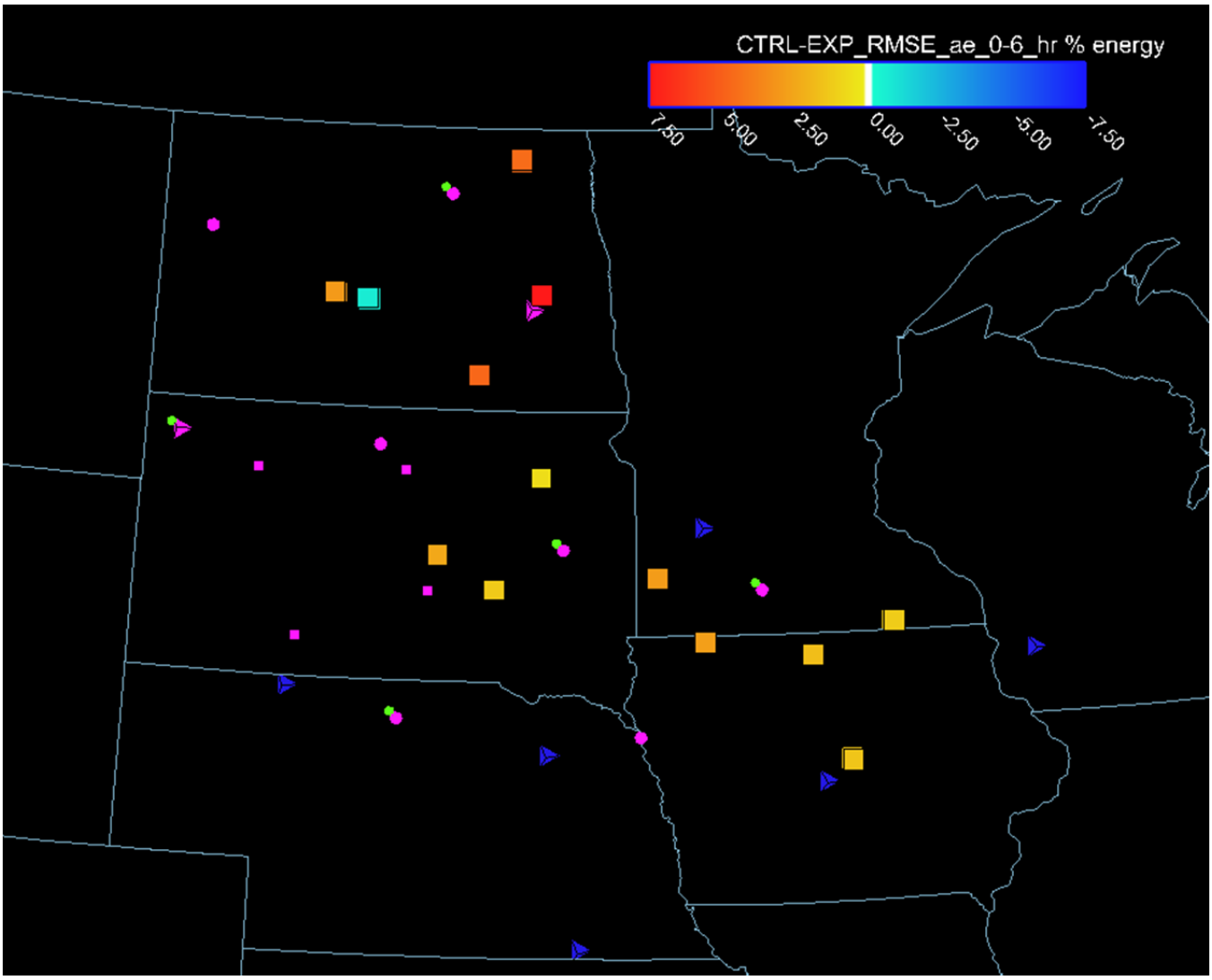

Figure 78: Fall data denial period RMSE difference between the control and experimental ESRL_RAP raw power forecasts normalized by the average power produced over the first six hours of the forecast at the NextEra wind plant locations (large squares). Warm (cool) colors indicate that the experimental (control) forecast is performing better. The operational wind profiler locations are shown as blue tetrahedrons. The tropospheric (boundary layer) profilers deployed for this study are shown as bright pink tetrahedrons (circles), sodar locations are shown as green circles (located at the same sites as the wind profilers), and the SDSU tall towers are shown as bright pink small squares.

the northern locations with an average improvement of 3.8% over all North Dakota sites. The largest improvement at any site occurred at the Ashtabula Aggregate site which shows an improvement of 7.5%.

A significant improvement in forecast accuracy is also seen at most sites during the winter data denial period (Figure 79), the only exceptions being Edgeley Basin in North Dakota (-0.7%), the Crystal Lake aggregate in northern Iowa (-0.3%) and Mower County in southeast Minnesota (-2.4%). The largest improvements again are concentrated in the northern portions of the study area, with the largest improvement occurring at the Ashtabula aggregate site with an improvement of 5.3%.

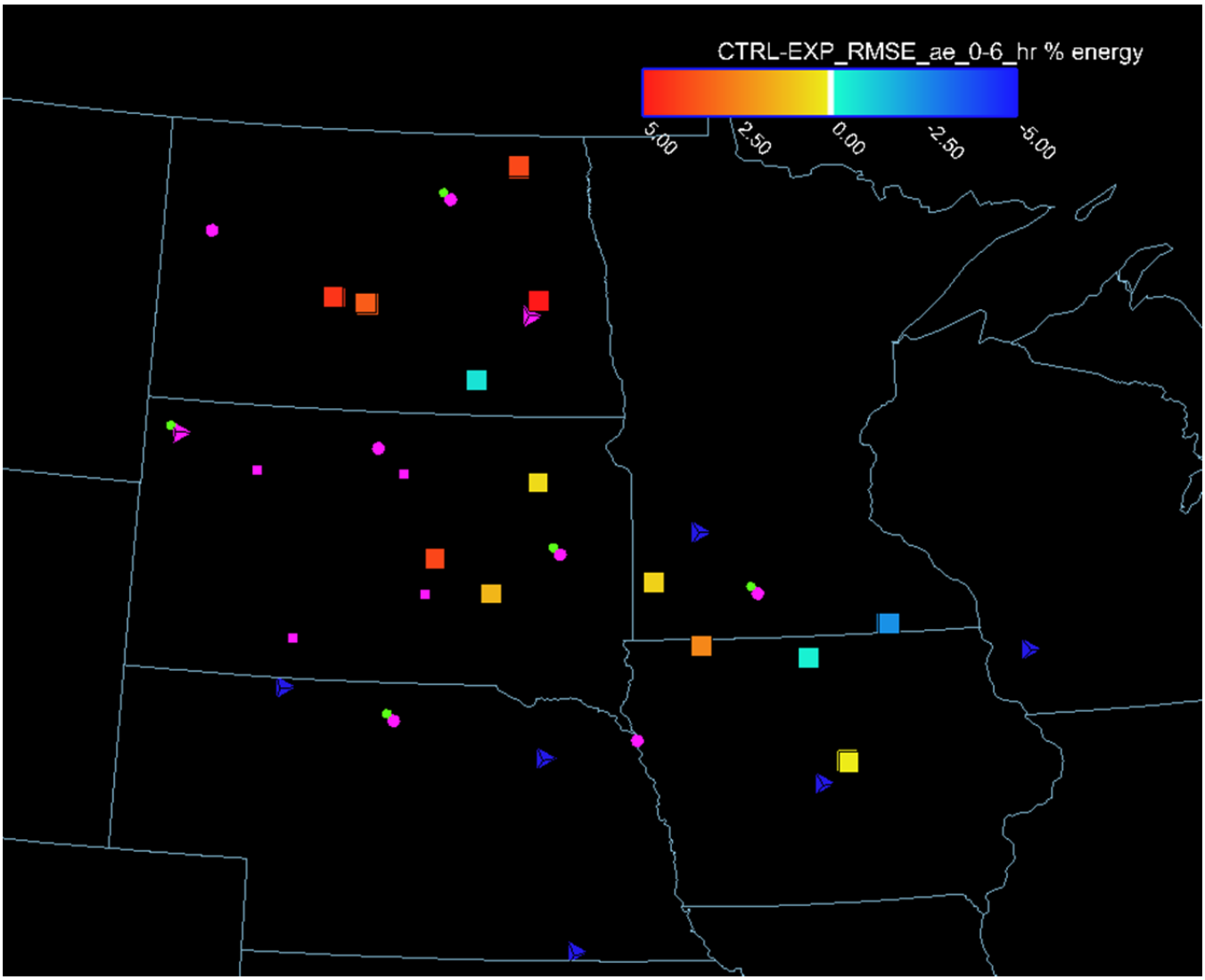

Figure 79: Winter data denial period RMSE difference between the control and experimental ESRL_RAP raw power forecasts normalized by the average power produced over the first six hours of the forecast at the NextEra wind plant locations (large squares). Warm (cool) colors indicate that the experimental (control) forecast is performing better. The operational wind profiler locations are shown as blue tetrahedrons. The tropospheric (boundary layer) profilers deployed for this study are shown as bright pink tetrahedrons (circles), sodar locations are shown as green circles (located at the same sites as the wind profilers), and the SDSU tall towers are shown as bright pink small squares.

The Percent Improvement at the wind plant sites for the spring data denial period is shown in Figure 80. During the spring period, the assimilation of the additional data improves the wind power forecasts at eight of the thirteen sites, with two sites a statistical 'tie'. While fewer sites show improvements in the first six hours of the forecasts than in the fall and winter periods, the improvements at some sites are significant, with the forecasts at the Oliver aggregate site in central North Dakota having an average improvement of 8.8%, the largest of any of the data denial periods. Once again, the greatest gains in forecast accuracy are in the northern half of the study domain, with the North Dakota sites showing an average improvement of 4.1%. However, the other areas show the smallest improvements of any of the data denial periods, with the South Dakota sites having an average improvement of 0.5%, and the southern Minnesota/Iowa sites having a small average negative Percent Improvement (-0.5%).

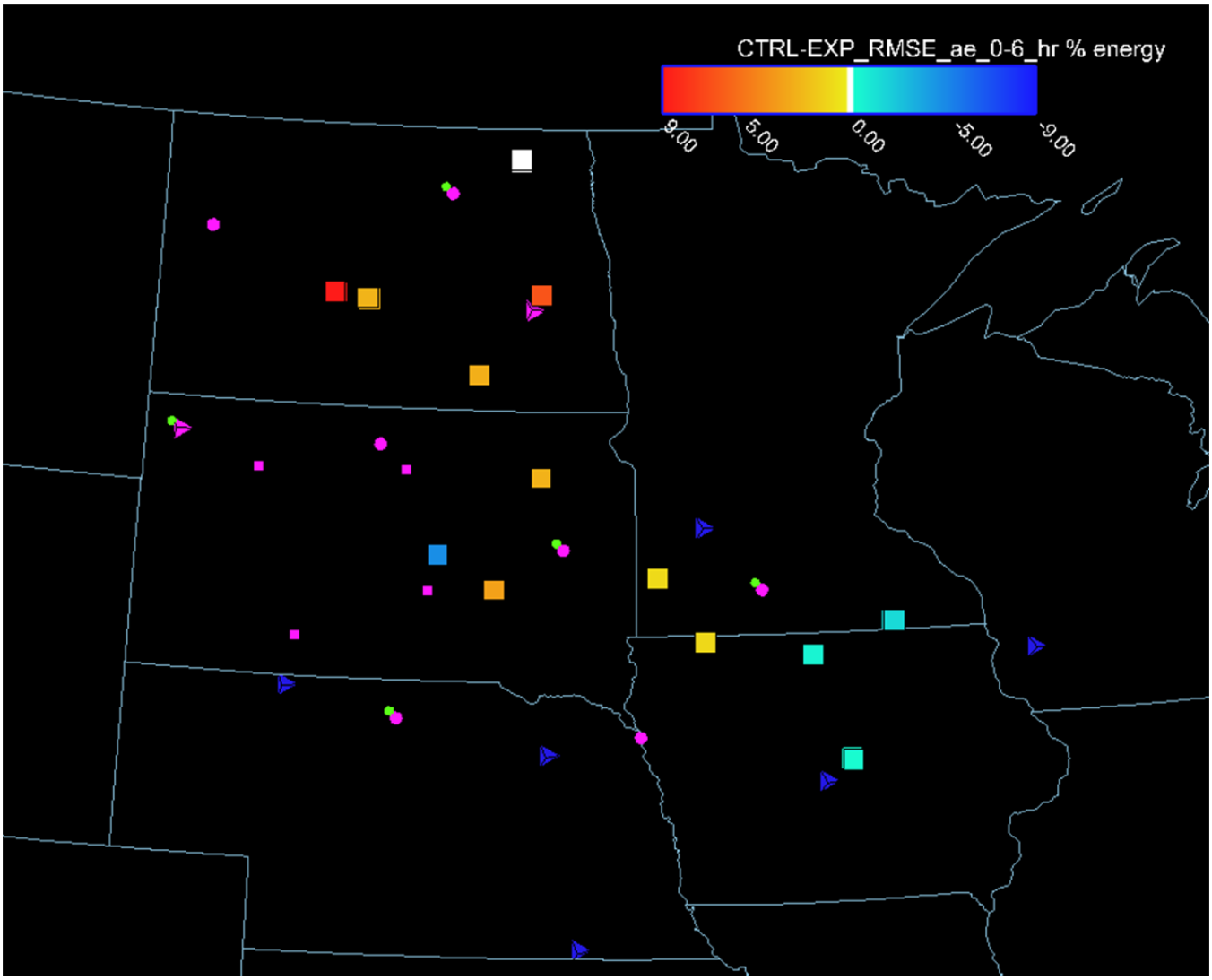

Figure 80: Spring data denial period RMSE difference between the control and experimental ESRL_RAP raw power forecasts normalized by the average power produced over the first six hours of the forecast at the NextEra wind plant locations (large squares). Warm (cool) colors indicate that the experimental (control) forecast is performing better. The operational wind profiler locations are shown as blue tetrahedrons. The tropospheric (boundary layer) profilers deployed for this study are shown as bright pink tetrahedrons (circles), sodar locations are shown as green circles (located at the same sites as the wind profilers), and the SDSU tall towers are shown as bright pink small squares.

As was discussed earlier in this report, the winds during the summer months can be very challenging to forecast accurately because of weaker dynamical forcing mechanisms responsible for the development and evolution of larger-scale weather systems, and complex windflows generated by thunderstorms. Clouds in general (and thunderstorms in particular) are also extremely difficult to represent properly in initial conditions in weather forecast models. Despite these challenges, the results from the summer data denial period show an improvement at 10 of the 13 wind plant sites (Figure 81). As was seen in the other data denial periods, the improvements are largest in the northern portion of the study area, with the North Dakota, South Dakota, and Minnesota/Iowa sites having an average Percent Improvement of 4.9%, 1.8% and 0.97% respectively. There is also significant geographic variability in the sites with the largest improvements, which could be a result of more successful local convective forecasts due to the additional data assimilation.

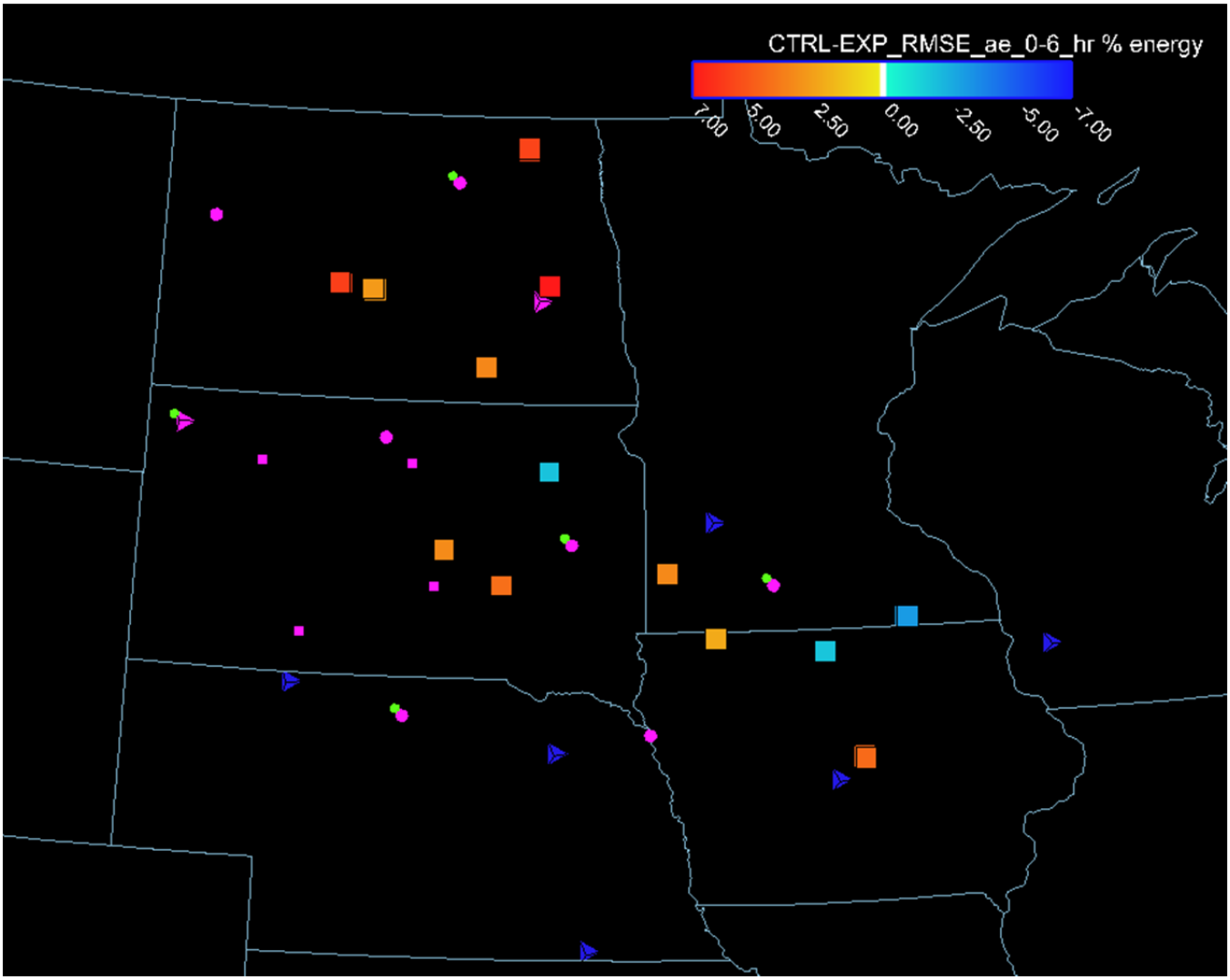

Figure 81: Summer data denial period RMSE difference between the control and experimental ESRL_RAP raw power forecasts normalized by the average power produced over the first six hours of the forecast at the NextEra wind plant locations (large squares). Warm (cool) colors indicate that the experimental (control) forecast is performing better. The operational wind profiler locations are shown as blue tetrahedrons. The tropospheric (boundary layer) profilers deployed for this study are shown as bright pink tetrahedrons (circles), sodar locations are shown as green circles (located at the same sites as the wind profilers), and the SDSU tall towers are shown as bright pink small squares.

As seen in the Percent Improvement results from each of the data denial periods shown above, the additional data assimilation made the largest improvements to forecast accuracy during the first six hours of the forecasts at the sites in the northern half of the study area where the operational observational network is extremely sparse. The Percent Improvement calculated over the first twelve hours of the forecasts revealed that the assimilation of the additional data also had a larger impact on the forecasts at later forecast hours in the northern portion of the study area as shown in Figure 82. The average improvement over the first twelve hours of the forecasts at the North Dakota sites is 2.1%, with an average improvement of 0.76% at the South Dakota sites, and 0.23% at the southern Minnesota/Iowa sites.

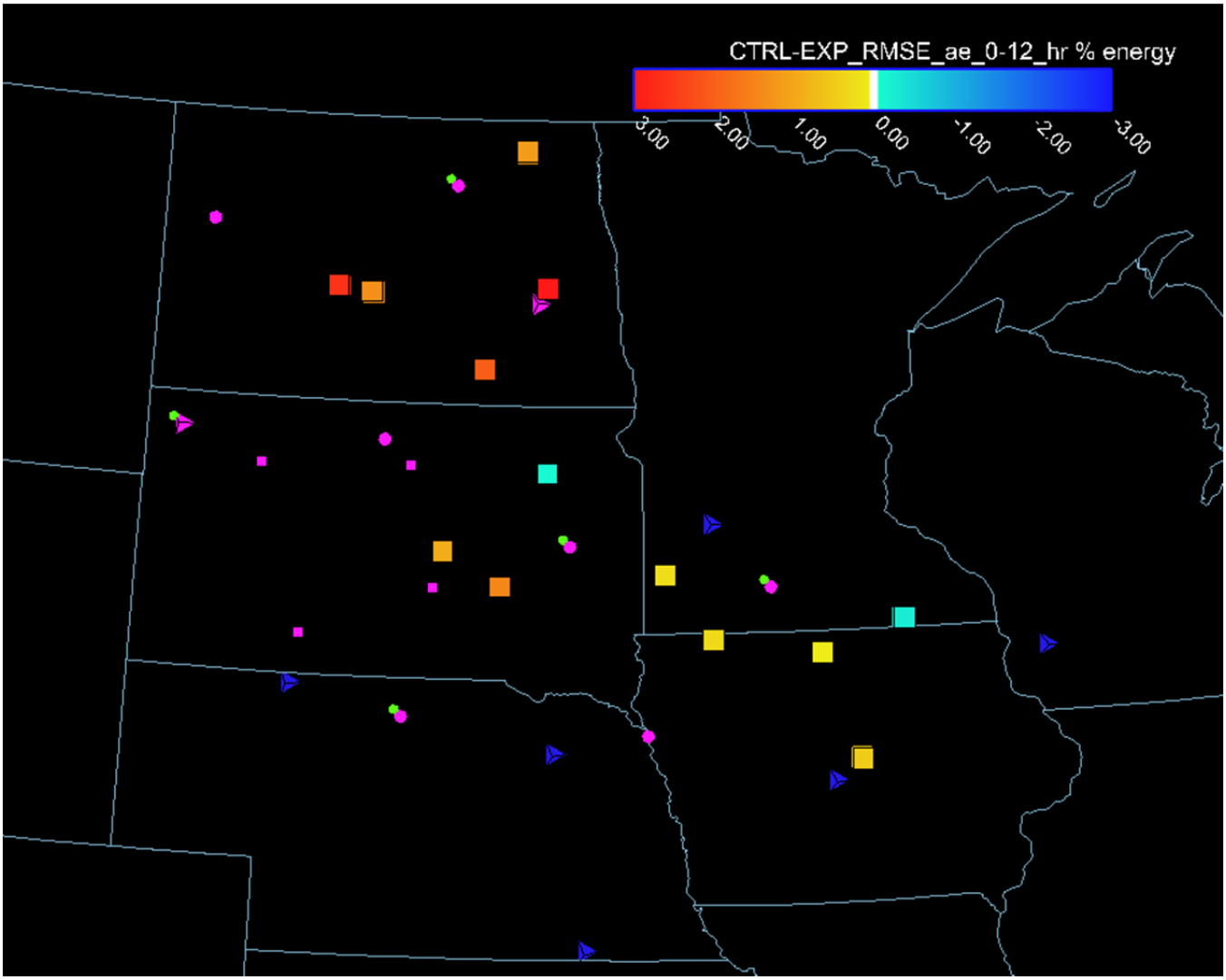

Figure 82: RMSE difference between the control and experimental ESRL_RAP raw power forecasts normalized by the average power produced over the first twelve hours of the forecast at the NextEra wind plant locations (large squares) for all data denial periods. Warm (cool) colors indicate that the experimental (control) forecast is performing better. The operational wind profiler locations are shown as blue tetrahedrons. The tropospheric (boundary layer) profilers deployed for this study are shown as bright pink tetrahedrons (circles), sodar locations are shown as green circles (located at the same sites as the wind profilers), and the SDSU tall towers are shown as bright pink small squares.

Time series of the RMSE normalized by the average energy produced as a function of forecast hour for the ESRL_RAP control, experimental, and the OP_RAP raw power forecasts for all data denial periods at the Ashtabula Aggregate and Crystal Lake Aggregate sites are shown Figure 83 and Figure 84, respectively. The Ashtabula Aggregate site has the largest overall improvements in forecast accuracy from the additional data assimilation. As can be seen in Figure 83, the largest forecast improvements occur during the first six hours of the forecasts, but lesser improvements are seen through forecast hour eleven. At other sites in northern portions of the study area, the forecast improvement is seen throughout the first fifteen hours of the forecasts, although the average improvement is smaller than at the Ashtabula site. The results from the Crystal Lake Aggregate site (Figure 84) show what is typically seen in the southern portions of the study area; the experimental forecasts have better accuracy during the first 1-2 hours of the forecasts, then have slightly higher RMSE values compared to the control

forecasts over the next 6-8 hours. At some sites in the southern portion of the study area, the experimental forecasts are again more accurate than the control forecasts during the later forecast hours, so the average RMSE in the experimental forecasts is slightly lower over the first 15 hours of the forecast than in the control forecasts. (This is not the case at the Crystal Lake and Mower County sites.)

At the system level, although the results are affected by improvements at the site level, they are also affected by the extent to which forecast errors cancel out when the sites are aggregated and the fact that the capacity varies by site so each of the sites are not 'weighted' equally when aggregated. A time series of the RMSE normalized by the average energy produced over the first fifteen hours of the system aggregate forecasts is shown in Figure 85. The experimental raw power forecasts show a clear improvement in system-wide forecast accuracy compared to the control forecasts through the first three hours of the forecasts, then show a very small improvement at most forecast hours through forecast hour twelve.

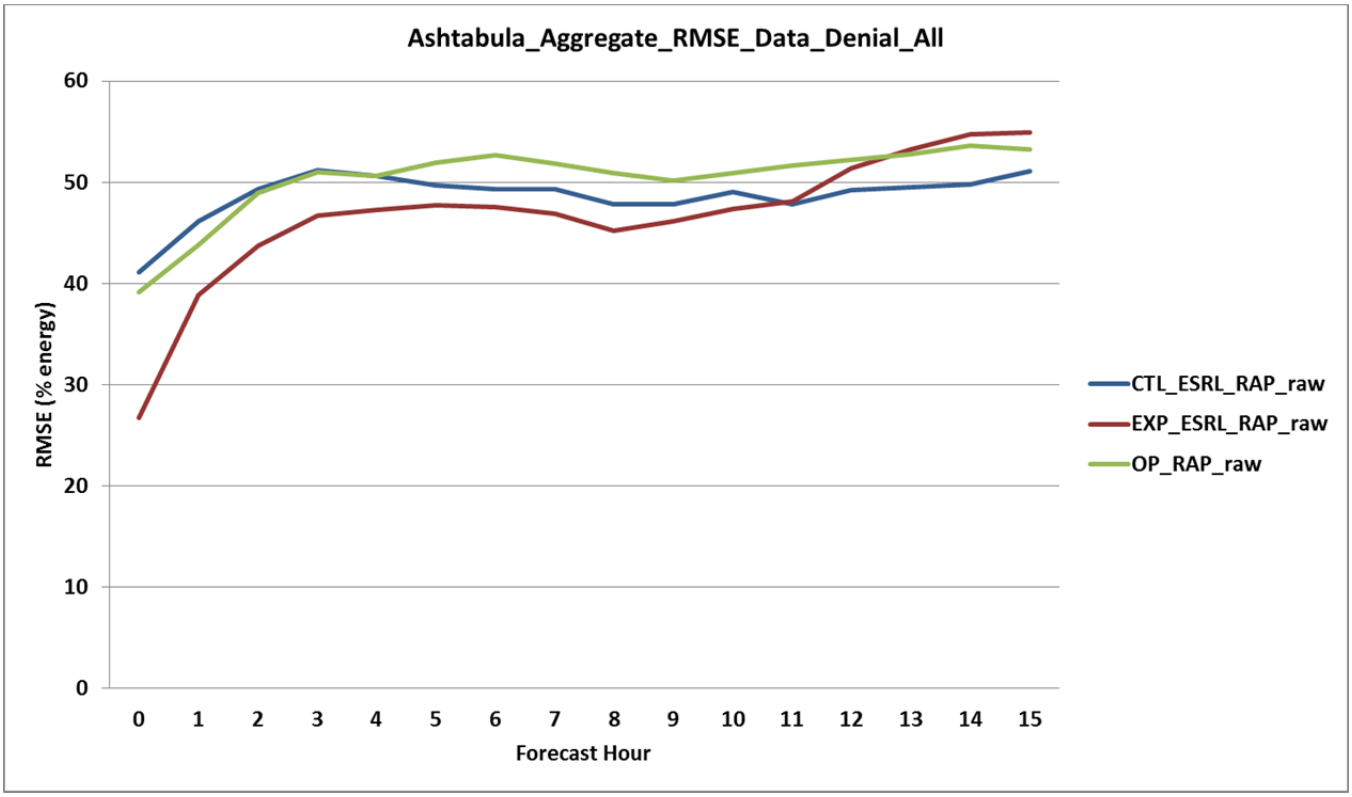

Figure 83: RMSE normalized by the average energy produced as a function of forecast hour for the control and experimental ESRL_RAP and the OP_RAP raw power forecasts for all data denial periods at the Ashtabula aggregate site in eastern North Dakota.

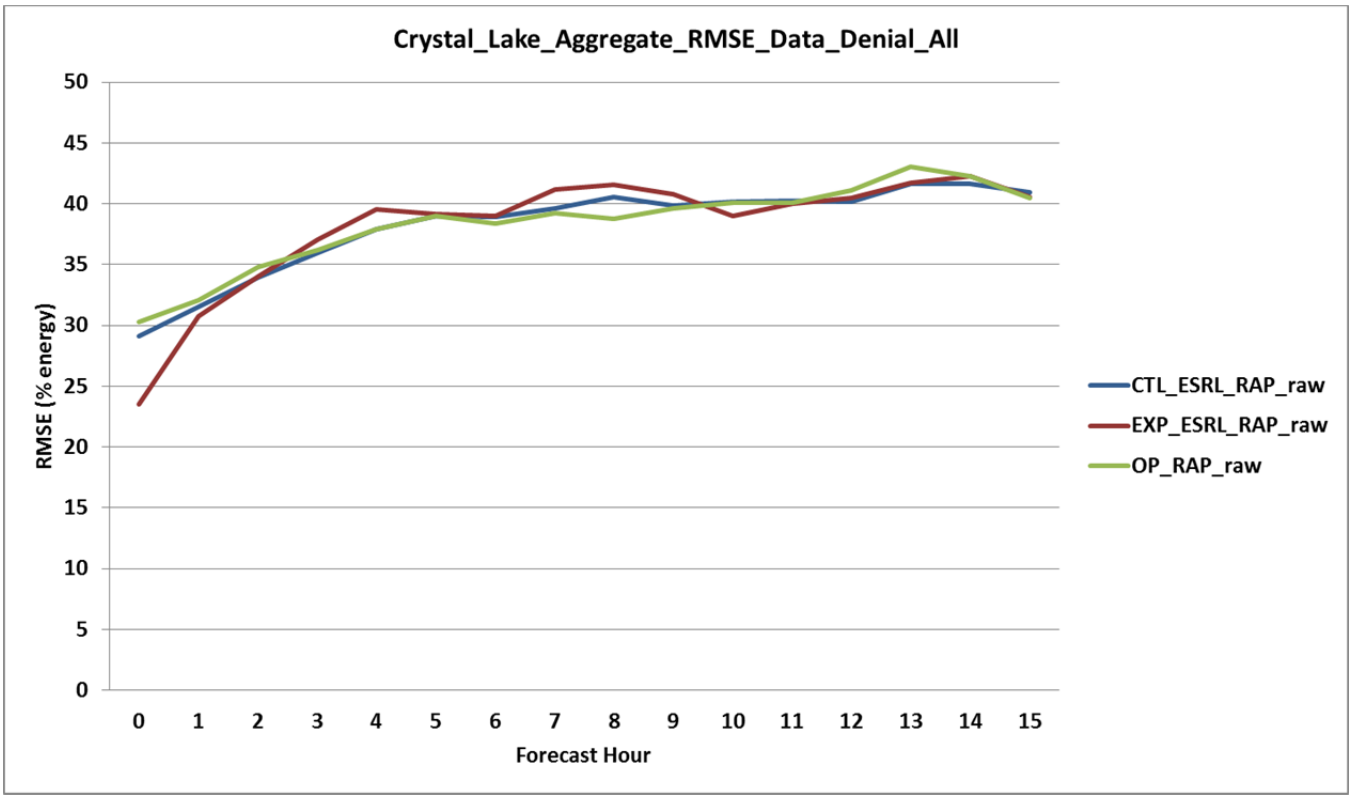

Figure 84: RMSE normalized by the average energy produced as a function of forecast hour for the control and experimental ESRL_RAP and the OP_RAP raw power forecasts for all data denial periods at the Crystal Lake aggregate site in north central Iowa.

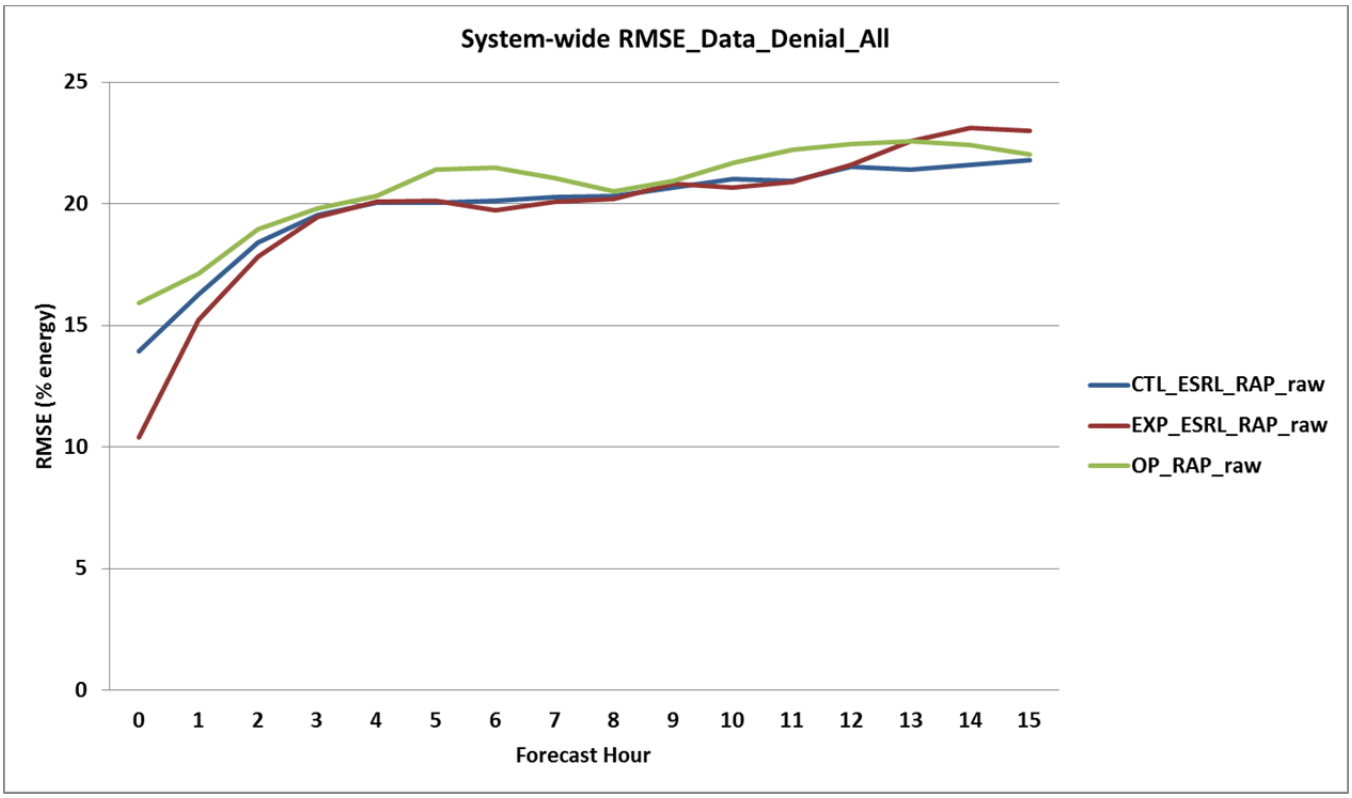

Figure 85: System-wide RMSE normalized by the average energy produced as a function of forecast hour for the control and experimental ESRL_RAP and the OP_RAP raw power forecasts for all data denial periods.

The results from the data denial experiments described above show that there is a clear improvement in forecast accuracy at almost all NextEra wind plant sites in the study area due to the assimilation of the additional meteorological data provided in this study into the forecast model initial conditions. The greatest improvements are seen in the northern half of the study area, where both the operational upper air and surface meteorological observing networks are very sparse compared to the southern portion of the study area. The largest forecast accuracy improvements were seen at the Ashtabula Aggregate site in eastern North Dakota which was located approximately fifteen miles from one the of 449 MHz tropospheric wind profilers deployed for use in this study. The forecast improvements in the northern portion of the study area also extended over a longer period within the forecasts compared to the improvements seen in the southern portion of the study area, with forecast improvements extending through the first twelve-fifteen hours of the forecasts at many northern sites.

The forecast improvement due to the additional data assimilation at the system aggregate level was smaller than was seen at most of the individual wind plant locations, likely due to forecast error cancelation as site-level forecasts were aggregated together, and the fact that wind plant capacities were relatively large at a few locations where the forecast improvements were relatively small (or were negative). However, it's important to consider forecast performance both at the individual site level and the system aggregate level as MISO is uses local data to identify system constraints and set nodal electricity prices.

5. Accomplishments

The following conference presentations were given in order to share the results of this study with the Meteorological and Power Grid System Operator communities:

"WFIP: Northern Study Area". Wind Power Peer Review Meeting, Department of Energy, Washington, D.C. March 2014

"Overview of the Final Results from the Wind Forecast Improvement Project (WFIP): Northern Study Area". Utility Variable-Generation Integration Group Forecasting Workshop, Tucson, AZ. (Invited Talk) February 2014

"Overview of the Final Results from the Wind Forecast Improvement Project (WFIP): Northern Study Area". *94th Annual Meeting of the American Meteorological Society,* Atlanta, GA. February 2014

"Short-Term Wind Energy Forecast Improvements Resulting from the Wind Forecast Improvement Project (WFIP): Northern Study Area". Utility Variable Generation Integration Group Forecasting Workshop, Salt Lake City, UT. (Invited Talk) February 2013

"Short-Term Wind Energy Forecast Improvements Resulting from the Wind Forecast Improvement Project (WFIP): Northern Study Area". *93rd Annual Meeting of the American Meteorological Society*, Austin, TX. January 2013

"WFIP: Northern Study Area". Wind Power Peer Review Meeting, Department of Energy, Washington, D.C. June 2012

"Update on the DOE/NOAA Wind Forecasting Improvement Project – Northern Plains Study Area". Utility Variable Generation Integration Group Forecasting Workshop, Tucson, AZ. (Invited Talk) February 2012

"Improving Short-Term Wind (Power) Forecasts, and Assessing their Value for the Midwest ISO". *92rd Annual Meeting of the American Meteorological Society*, New Orleans, LA. January 2012

An overview paper of this study was submitted to the Bulletin of the American Meteorological Society.

Some of the improvements to the NOAA RAP forecast model resulting from this study were implemented in the operational RAP v2 forecast model at NCEP.

6. Summary and Conclusions

We have presented the results from research aimed at improving hub-height wind forecasts in the short-range (0-6 hour) NOAA weather forecast models through additional data assimilation and improved model physics over the U.S. Northern Plains for use in wind energy forecasting. Additional meteorological observing platforms including wind profilers, sodars, and surface stations were deployed for this study by NOAA and DOE, and additional meteorological data at or near wind turbine hub height were provided by SDSU and WindLogics/NextEra Energy Resources to NOAA over a large geographical area in the upper Midwest/Northern Plains for assimilation into the NOAA research weather forecast models. The improvements in the wind energy forecasts that resulted were examined

in many different ways in order to quantify the improvements in the forecasts as they would be used by power grid system operators, and by wind plant owners/operators participating in the energy markets.

Two operational weather forecast models (OP_RUC, OP_RAP) and two research weather forecast models (ESRL_RAP, HRRR) were used as the base wind forecasts for generating several different wind power forecasts as would be generated by commercial wind energy forecast vendors: a 'raw' forecast generated by directly converting the weather forecast model wind speed to wind power using a plant-derived power curve, a 'bias-corrected raw' forecast that used a simple rolling two week average wind speed bias as a first order correction to the wind speed before converting the forecasted wind speed to power, a 'trained' forecast that used sophisticated machine learning techniques and local wind plant data to statistically correct the wind power forecasts, and a 'trained ensemble' which combined data from multiple weather forecast models and local wind plant data to statistically correct the wind power forecasts. The error characteristics of each of these types of forecasts were examined and quantified using bulk error statistics for both the local wind plant and the system aggregate wind power forecasts.

The overall bulk error statistics calculated over the first six hours of the forecasts at both the individual wind plants and at the system-wide aggregate level over the one year study period showed that the ESRL_RAP-based power forecasts (all types) had lower overall errors rates than the other forecast model-based power forecasts. At the wind plant level, the HRRR-based power forecasts also had lower error rates than the operational model-based power forecasts at the majority of sites, which is also important to the system operator and of great significance to the wind plant operator. Even though the HRRR model has better horizontal grid resolution than the ERSL_RAP model (3 km grid spacing vs. 13 km grid spacing), this did not translate into better overall bulk error statistics in the power forecasts (with a few exceptions in regions of more complex terrain). There could be several possible reasons for this non-intuitive result. For example, the high resolution HRRR is now capable of resolving mesoscale flow features which our current observational network cannot always detect accurately. This could lead to inaccurate representation of these flow features, which may make the forecasts less accurate. It is also possible that the HRRR model is resolving small-scale features that are present in the observations, but the timing of the features impacting the wind plant is inaccurate which leads to less accurate bulk error statistics. Further investigation would be required to determine the exact cause of the forecast error differences.

The training process reduced the bulk forecast errors for all model-based power forecasts, both for the individual wind plant forecasts and for the system-wide aggregate forecasts. At most sites, the forecast improvement provided by the training was significant. As with any statistical correction, the process is optimized for reducing (or maximizing) certain average forecast characteristics. In this case, the training process reduces the overall error (RMSE) of the forecasts at the expense of some sharpness in the forecast features (i.e. the trained power forecasts are smoother than the raw power forecasts). The trained ensemble power forecasts further improved the error statistics beyond the individual forecast model-based trained power forecasts at each wind plant location, but since the trained ensemble forecasts include the North American Model (NAM), which was not the focus of this study, the results of the trained ensemble forecasts for the system aggregate power forecasts were only briefly presented.

The number of large forecast errors (hourly forecast errors exceeding a large threshold of rated capacity) were also calculated for all model-based power forecasts over the entire year of the study as power grid operators are especially concerned about the extreme events (i.e. the tails of the forecast error distribution), even if they occur only rarely, because these errors can influence the reserves and operating practices that they use to ensure system reliability. The results showed that the research model-based power forecasts reduce the number of large forecast errors at most wind plant locations compared to operational model-based power forecasts (particularly the operational RUC). The training process greatly reduced the number of large forecast errors in all the model-based forecasts,

particularly in the ESRL_RAP-based forecasts at the larger forecast errors. This is likely because the training process tends to create smoother forecasts than the raw power forecasts, which reduces the magnitude of the 'peaks' and 'valleys' in the forecasts where the largest forecast errors often occur. Overall, the results showed that the ESRL_RAP-based power forecasts had fewer large hourly forecast errors than all other model-based forecasts at both the local site and system aggregate levels.

The bulk error statistics of the various model-based power forecasts were also calculated by season and model runtime/forecast hour as power system operations are more sensitive to wind energy forecast errors during certain times of the year and certain times of day. The results showed that there were significant differences in seasonal forecast errors between the various model-based forecasts. This was true at both the individual wind plant level, and the system aggregate level. The results from the analysis of the various model-based power forecast errors by model runtime and forecast hour showed that the forecast errors were largest during the times of day that have increased significance to power system operators (the overnight hours and the morning/evening boundary layer transition periods). The HRRR-based power forecasts tend to have smaller forecast errors than the other model-based power forecasts during the higher-error time periods that are also critical time periods for the system operators. Results from the various trained model-based power forecasts also showed that the training process tends to reduce the forecast error the most during the forecast hours that generally have higher forecast error. These analyses collectively show that when it comes to power system operations, the level of complexity in identifying the best weather model for a particular seasonal/daytime regime is yet another argument for using advanced, multi-model ensemble wind power forecasting systems in actual production. With proper ensemble training methods, such systems should be able to implicitly extract the skill from different models for the various regimes. The results here clearly show that a single model is unlikely to be the best choice at least for some seasons of the year or times of day. So while higher forecast model resolution is not the simple answer to providing better overall wind energy forecasts, the HRRR model would be an important member in a state-of-the-art multi-model ensemble in terms of providing value to power system operators.

Wind energy ramp events, in which the amount of wind energy produced changes very rapidly in time, are also of concern to power system operators as very large wind power forecast errors can occur during these events, creating challenges for electric grid system operations. The rate of ramping is of particular concern to system operators in that there are limits as to how quickly conventional generation can be ramped up (or down) to offset the changes in wind generation in order to keep the system balanced. While both up-ramp and down-ramp rates in wind generation are of a concern, down-ramp events are a larger potential issue for system operators as the only recourse for offsetting a down-ramp in wind energy is to ramp-up conventional generation (or shed load if that option is available). A comprehensive analysis of wind energy forecast errors for the various model-based power forecasts was presented for a suite of wind energy ramp definitions. The results compiled over the year-long study period showed that the bias-corrected raw power forecasts based on the research models (ESRL_RAP, HRRR) more accurately predict wind energy ramp events than the forecasts based on the current operational forecast model, both at the system aggregate level and at the local wind plant level. At the system level, the ESRL_RAP-based forecasts most accurately predict both the total number of ramp events and the occurrence of the events themselves, but the HRRR-based forecasts more accurately predict the ramp rate. There was little difference between the various model-based power forecasts in the accuracy of the timing of the ramp events. In general, up-ramp events were more accurately predicted than down-ramp events, both at the system level and individual wind plant level for all model-based forecasts. At the individual wind plant level, the HRRR-based forecasts most accurately predict the actual ramp occurrence, the total number of ramps and the ramp rates (40 to 60% improvement in ramp rates over the coarser resolution forecasts). The HRRR-based forecasts also most accurately predict both the number and occurrence of potentially convective high amplitude/short duration events during the convective season, but tends to over-predict the number of these events leading to a higher number of false alarms than the other model-based power forecasts.

The advantages of geographic dispersion of wind plants on the reduction of system-wide aggregate wind power forecast errors and the number and severity of wind energy ramp events was also demonstrated. The results showed that the system-wide wind energy forecast errors (as defined by the NextEra wind plants in the study area) were roughly half of what would be expected at a typical individual wind plant. This is a significant reduction in forecast error that results from forecast errors tending to cancel each other out to some extent as the geographic dispersion of the wind plants increases (the further apart the wind plants are located from each other, the less correlated the forecast errors are likely to be). The number of observed system-wide wind energy ramps was also calculated for varying amounts of geographic dispersion using two different ramp definitions. For each of the ramp definitions, the number of observed ramps in the forecast time series initially decreased significantly as the geographical diversity of the wind plant locations increased, then incremental additions eventually had a smaller impact as new wind plants are added.

To assess the relative impacts of the additional data provided in this study being assimilated into the research weather forecast models separately from the improvements to the model physics, NOAA ran several data denial experiments in which select periods were ‘re-forecasted’ using the ESRL_RAP weather forecast model. Two sets of forecasts were generated for each data denial period: one set of the forecasts was generated by initializing the forecast only using the operational meteorological data currently available (the ‘control forecast’), and one set was generated by initializing the forecast model using the currently available operational meteorological data plus the additional data collected as part of this project (the ‘experimental forecast’). All other aspects of the forecasts were identical. The results from the data denial experiments showed that there is a clear improvement in wind energy forecast accuracy at almost all NextEra wind plant sites in the study area due to the assimilation of the additional meteorological data provided in this study into the forecast model initial conditions. The greatest improvements were seen in the northern half of the study area, where both the operational upper air and surface meteorological observing networks are very sparse compared to the southern portion of the study area. The forecast improvements in the northern portion of the study area also extended over a longer time period within the forecasts compared to the improvements seen in the southern portion of the study area, with forecast improvements extending through the first twelve-fifteen hours of the forecasts at many northern sites. The forecast improvement at the system aggregate level was smaller than was seen at most of the individual site locations, likely due to forecast error cancelation as site-level forecasts are aggregated together, and the fact that the capacities of the wind plants at a few locations are relatively large at the locations where the forecast improvements were smaller. However, forecast performance at both the individual site level and the system aggregate level is important as MISO uses local data to identify system constraints and set nodal electricity prices.

Unfortunately, the future of the operational NOAA wind profiler network is uncertain due to budget constraints. If the operational network provided similar improvements to wind energy forecasts as was seen in the northern portions of the study area (where the only wind profilers present were those deployed in this study), we would expect to see short-range (0-6 hour) wind energy forecast error increase by 2-5% of energy produced as a result of the loss of this observational network based on the results from this study. While the impacts of the wind profiler network may be somewhat smaller in the southern portions of the study area because of other available data sources, the loss of this important observational network would invariably lead to decreased accuracy in wind energy (and other) forecasts.

7. Recommendations

As is the case in any research study, the results often lead to additional questions and uncover additional issues that need to be addressed in future research. Below are some suggestions for future work that could expand upon the work and results from this study.

A key physical process in weather forecast models that effects wind energy forecasting every day is the model representation of the land surface processes and the evolution of the planetary boundary layer, particularly under stable conditions. One of the problems with improving these processes for wind energy forecasting purposes is that very few hub-height wind observations (either public or proprietary) are available for testing and model forecast verification. Proprietary hub-height wind data provided by private industry to NOAA, and the publicly available hub-height wind observations provided by SDSU and others, could be used to fill that gap. Research that takes advantage of this data to improve model boundary layer processes is already underway at NOAA (Olson et al. 2012), and it is hoped that this work will continue following the completion of this project. Additional studies that aim to improve our general physical understanding of the atmosphere and the correct representation of physical processes within our weather forecast models are key components to improving our wind energy forecasts.

Wind energy ramp events are extremely difficult to forecast accurately, and although the next generation of NOAA short-range weather forecast models used in this study showed clear improvement in forecasting these events compared with the current operational weather forecast models, the results presented in this study indicate that continued work is needed to improve wind energy ramp forecasts. If would also be extremely valuable to investigate the forecast accuracy of ramp events broken down by the meteorological feature responsible for the ramp event (i.e. front, thunderstorm, etc.) to help quantify the forecast issues associated with ramp events in order to help focus the efforts of future research on the most significant events.

More detailed data denial studies would be helpful in order to quantify the contribution of the various individual data sources (wind profilers, tall meteorological towers, etc.) to the forecast improvements resulting from the additional data sources provided in this study. These additional studies would help provide additional focus to the types of data sources that have the largest impact on wind energy forecast accuracy.

8. References

Banta, Robert M., Y.L. Pichugina, N.D. Kelley, R.M. Hardesty, and W.A. Brewer, 2013: Wind Energy Meteorology: Insight into Wind Properties in the Turbine-Rotor Layer of the Atmosphere from High Resolution Doppler Lidar. *Bull. Amer. Meteor. Soc* ., **94**, 883-902.

Benjamin, Stanley G., and Coauthors, 2004: An Hourly Assimilation–Forecast Cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518.

Boer, G.J., 1994: Predictability Regimes in Atmospheric Flow. *Mon. Wea. Rev*., **122**, 2285-2295.

Dabberdt, W.F., and Coauthors, 2004: Meteorological research needs for improved air quality forecasting: Report of the 11th Prospectus Development Team of the U.S. Weather Research Program. *Bull. Amer. Meteor. Soc*., **85**, 563-586.

Freedman, Jefferey, et.al., 2014: 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. DOE final report, award number DE- EE0004420.

Morss, R. E., C. Snyder, and R. Rotunno, 2009: Spectra, Spatial Scales, and Predictability in a Quasigeostrophic Model. *J. Atmos. Sci.*, **66**, 3115–3130

Ngan, K., P. Bartello, D. N. Straub, 2009: Predictability of Rotating Stratified Turbulence. *J. Atmos. Sci.*, **66**, 1384–1400.

Olson, Joseph B., J.M. Wilczak, R.M. Banta, and Y Pichugina, 2012: An Investigation of Simulated Boundary layer Winds in Low-Level Jet Cases Observed during the Wind Forecast Improvement Project. Electronic Proceedings of the 20th Symposium on Boundary Layers and Turbulence, American Meteorological Society, Boston, MA. (https://ams.confex.com/ams/20BLT18AirSea/webprogram/Paper209735.html)

Seamen, N., 2000: Meteorological modeling for air-quality assessments. *Atmos. Environ*., **34**, 2231-2259.

Storm, B.J., J. Dudhia, S. Basu, A. Swift and I. Giammanco, 2009: Evaluation of the Weather Research of Forecasting model on forecasting low-level jets: Implications for wind energy. *Wind Energy*, **12**, 81-90.

Wilczak, James M., et al., 2014: The Wind Forecast Improvement Project (WFIP): A Public/Private Partnership for Improving Short Term Wind Energy Forecasts and Quantifying the Benefits of Utility Operations. DOE final report, award number DE-EE0003080.

Zhong, S. and J.D. Fast, 2003: An evaluation of the MM5, RAMS, and Meso-Eta models at subkilometer resolution using field campaign data in the Salt Lake Valley, *Mon. Wea. Rev*., **131**, 1301-1322.

List of Acronyms

CSI: Critical Success Index

ESRL: Earth Systems Research Laboratory

ESRL_RAP: Earth System Research Laboratory version of the Rapid Refresh model

ESRL_RAP_BCRW: Bias-corrected raw wind energy forecast based on the research Rapid Refresh model

ESRL_RAP_ENS: Trained ensemble wind energy forecast based on the research Rapid Refresh model and the North American Model

ESRL_RAP_RW: Raw wind energy forecast based on the research Rapid Refresh model

ESRL_RAP_TR: Trained wind energy forecast based on the research Rapid Refresh model

FAR: False Alarm Ratio

HRRR: High Resolution Rapid Refresh model

HRRR_BCRW: Bias-corrected raw wind energy forecast based on the High Resolution Rapid Refresh model

HRRR_ENS: Trained ensemble wind energy forecast based on the High Resolution Rapid Refresh model and the North American Model

HRRR_RW: Raw wind energy forecast based on the High Resolution Rapid Refresh model

HRRR_TR: Trained wind energy forecast based on the High Resolution Rapid Refresh model

MAE: Mean Absolute Error

MTEA: Mean Total Absolute Error

MISO: Midcontinent Independent System Operator

NAM: North American Model

NCEP: National Center for Environmental Prediction

NOAA: National Oceanic and Atmospheric Administration

NREL: National Renewable Energy Laboratory

OP_RAP: Operational Rapid Refresh model

OP_RAP_BCRW: Bias-corrected raw wind energy forecast based on the Operational Rapid Refresh model

OP_RAP_ENS: Trained ensemble wind energy forecast based on the Operational Rapid Refresh model and the North American Model

OP_RAP_RW: Raw wind energy forecast based on the Operational Rapid Refresh model

OP_RAP_TR: Trained wind energy forecast based on the Operational Rapid Refresh model

OP_RUC: Operational Rapid Update Cycle model

OP_RUC_BCRW: Bias-corrected raw wind energy forecast based on the Operational Rapid Update Cycle model

OP_RUC_ENS: Trained ensemble wind energy forecast based on the Operational Rapid Update Cycle model and the North American Model

OP_RUC_RW: Raw wind energy forecast based on the Operational Rapid Update Cycle model

OP_RUC_TR: Trained wind energy forecast based on the Operational Rapid Update Cycle model

RMSE: Root Mean Square Error

SDSU: South Dakota State University

WFIP: Wind Forecast Improvement Project