NOAA Study to Inform Meteorological Observation for Offshore Wind

Positioning of Offshore Wind Energy Resources (POWER)

DOC/National Oceanic and Atmospheric Administration

NOAA Study to Inform Meteorological Observation for Offshore Wind

Positioning of Offshore Wind Energy Resources (POWER)

Final Report: April 1, 2014

Co-sponsor
U.S. Department of Energy
Energy Efficiency and Renewable Energy Office
Wind and Water Power Program
Award Number: DE-EE0003080

Submitted by Melinda Marquis, NOAA/ESRL Renewable Energy Manager
NOAA Principle Investigators:
Melinda Marquis
Stan Benjamin
Joseph Olson
Eric James
Bob Banta
Yelena Pichugina
Jim Wilczak
Irina Djalalova
Laura Bianco
Jacob Carley

U.S. Department of Energy Team:
Jese Zayas, HQ Program Manager
Joel Cline, Wind and Water Program
Brad Ring, Field Project Officer

Notice and Disclaimer

This report is based upon work supported by the U. S. Department of Energy under Award No. DE-EE0003080. 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

List of Acronyms

CFSR - Climate Forecast System Reanalysis
CONUSNEST – NAMRR contiguous United States 4 km nest
ESRL – Earth System Research Laboratory (within NOAA Research)
EMC – Environmental Modeling Center (within NOAA National Weather Service)
GDAS – Global Data Assimilation System
GFS – Global Forecast System
GSD – Global Systems Division
GSI – Gridpoint Statistical Interpolation
HRDL – High-Resolution Doppler Lidar
HRRR – High-Resolution Rapid Refresh
NAM – North American Mesoscale forecast system
NAMRR – North American Mesoscale Rapid Refresh assimilation/forecast system
NCEP – National Centers for Environmental Prediction
NDAS – NAM Data Assimilation System
NMMB – Nonhydrostatic Multiscale Model on the B grid
NWP – Numerical Weather Prediction
RAP – Rapid Refresh hourly assimilation/forecast system
RTG – Real-time global
SST – Sea surface temperature
TM** – Time minus ** (where ** indicates a two-digit hour)
WFIP – Wind Forecast Improvement Project
WPR – Wind-profiling radar
WRF-ARW – Weather Research Forecast Model - Advanced Research WRF version
WSR-88D – Weather Surveillance Radar - 88D
3DVar – Three-dimensional variational data assimilation

Executive Summary - POWER

This research addresses the well-known lack of reliable wind information at the heights of turbine rotors over the ocean, to inform critical decisions that must be made about the deployment of offshore wind farms—such issues as siting, required hardware robustness, predictability of the winds aloft, and many others. Available marine wind-profile measurements indicate that vertical extrapolation of near-surface wind data can lead to large errors in estimating winds at hub height. Numerical weather prediction (NWP) models, often used to provide wind information in inaccessible locations, have not yet been extensively verified over the ocean. Therefore, obtaining the most reliable rotor-level wind information will require field measurement campaigns, in which accurate wind profiles through the rotor layer are measured. Design and planning for these measurement campaigns would benefit greatly from any prior information on wind properties in the rotor layer. In this project, we use an existing dataset that includes wind-profile data from a research cruise in the Gulf of Maine. The shipboard profile data were measured by a Doppler lidar and wind-profiling radar while the ship was underway. This unique dataset was used to characterize the rotor-layer winds, including their spatial and temporal variability and vertical structure.

Another key objective of this research is to verify hub-height winds predicted by two different NOAA NWP forecast models. The verification performed here is an evaluation of model skill in predicting hub-height winds over the ocean, using a total of 12 model configurations based on two modeling systems, the Rapid Refresh (RAP) system and the North America Mesoscale Rapid Refresh (NAMRR) forecast system. The two models are run at both normal resolution ('parent" model) and nested down to high-resolution versions (called High-Resolution Rapid Refresh or HRRR and the CONUS-NEST). To test the impact of assimilating upper-air wind data into these models, all four of these versions were run with and without assimilation of hourly wind data from the land-based array of 915-MHz wind-profiling radars. For the four remaining configurations, output from the two modeling systems were averaged to form a two-member ensemble for both the parent- and the nested-grid models, and each of these was run both with and without profiler assimilation. Because ensembling of short-term model forecasts is an area of active development in NOAA, these results provide a glimpse of future capabilities.

Overall results from this POWER study include:

An initial *observed* characterization of offshore wind behavior at turbine-level, based on High-Resolution Doppler Lidar profile data

An initial error assessment of short-range model forecasts for offshore turbine-level winds, showing roughly similar skill to those recently demonstrated for land-based turbine-level winds

More accurate offshore turbine-level wind forecasts, in general, by assimilation of land-based regional wind-profiling radar observations

A 2-year 80-m composite wind-speed map based on an hourly updated 3-km-grid model, to recommend areas for long-term offshore wind observations

More specifically:

Predictions from the HRRR model, which assimilates measurement data in real time from a variety of sources every hour, have been archived for past two years, 2012-2013. This dataset has been used to make horizontal maps of one of the key wind-energy (WE) variables, the wind speed at 80 m height (approximating turbine hub height), averaged over the 2 years (and subsets), for the U.S. and coastal waters. These maps reveal many interesting aspects of spatial variations in the offshore wind fields, such as the variability of the cross-shore gradient of wind-speed. Once verified by offshore measurements of winds at this level, such composite maps are potentially a very useful tool for addressing offshore spatial variability.

The HRDL and *RHB* profiler measurements themselves illustrate spatial and temporal variability of the wind field aloft over the ocean, finding the winds within the turbine rotor layer to be highly variable in time and space. Flow features, such as low-level jets, ramps in wind speed, and diurnal variations with stronger winds at night were noted. Considerable structure was evident in the HRDL-measured profiles below 100 m, and significant changes were noted over the 15-min sampling intervals, indicating that vertical resolutions of 10 m and profile-sampling frequencies of 15 min or less are required to adequately document the structure and variability of flow phenomena affecting rotor-layer winds.

Forecast model winds at marine wind-turbine levels were directly compared in several ways with winds through that layer measured on the *RHB* and then land-based profiler array. The HRDL comparisons included time-height cross sections of wind speed from the sea surface to 500 m, profiles of error statistics (including bias, RMSE, and correlation coefficient R^2), and the diurnal behavior of model skill. The verifications in general showed model winds to be biased slow by 0.5 to 2 m s^{-1}, model agreement with the measurements to within 2 m s^{-1} (RMSE) above 100 m ASL for the initial conditions, and R^2 of 0.8-0.9. Below 100 m ASL, these discrepancies all increased in value. Diurnal behavior of model skill for 100-m wind speeds indicated good agreement during the daytime, but underpredictions of 2 m s^{-1} at night for the initial and short-term forecasts (to 3 hr).

Wind-speed RMSEs were plotted against forecast valid time for each of the basic model configurations to study their relative skill, including the comparisons of model skill with and without assimilation of the land-based profiler data. In general, assimilating the profiler data improved the model forecasts for several hours over the ocean at the *RHB* location: the experimental runs with profiler data all had lower errors than the control runs by up to 0.2 m s^{-1} (8%) for at least the first 3-4 hr. The model forecasts were also verified against the data from the land-based profilers—the data that had also been assimilated into the model runs. These comparisons also reveal how long the assimilated data improved forecast skill at those sites; such a regional profiling network is being considered as a candidate for improving future offshore

wind forecasts. The initial and short-term improvements were qualitatively similar to those seen in the *RHB* verification, but were quantitatively more dramatic, because of the strong constraint on the modeled winds by the measurements. The simple two-member ensemble average was also evaluated to demonstrate the effects of ensembling, and the ensemble mean values showed improvements over the individual-model runs.

The critical horizontal scales of wind variability offshore are unknown, so they must be determined by measurement, which could then be used to verify whether the NWP models are also capable of reproducing the spatial flow variability observed. Long-term arrays of wind-profiling offshore buoys are recommended as an important component of a measurement strategy to understand offshore flow variability. A useful offshore wind-energy measurement network would consist of an appropriate mixture of cross-shore transects of buoy-mounted wind-profiling sensors and along-shore lines of these sensors, and several options are presented, depending on the problem to be addressed. For example, is it the cross-shore or along-shore variability, or is it the variability over 1000 km or over 100 km that is the reason for taking the measurements? It would most likely be impractical to deploy a high enough density of profiling devices over large enough areas to understand offshore flow variability, especially in the along-shore direction. Therefore mobile-sensing platforms, such as the *RHB* and/or aircraft, will be required along with the long-term arrays to characterize along-shore variability and discover recurrent areas of stronger and weaker flows. Intensive Observational Periods (IOP's) are recommended to focus resources and understand meteorological processes driving the flow variability in the along- and cross-shore directions.

An augmentation of boundary-layer profilers inland is recommended for better characterization of the regional meteorology and improved NWP assimilation results. The new HRRR-based 80-m wind composite map for the U.S. including offshore areas should be extended beyond the current 2-year period. This is because significant differences have been noted even between two consecutive years (2012-1013) or two different summers (2002 and 2004), so stable statistics—means, extremes, and others—should be based on as long of a sampling period as possible. The HRRR model forecasts cover the entire U.S., so the initial 2-year HRRR offshore wind composite maps available now, or updated longer-term versions as they become available, can be applied to other geographical regions. The NEAQS-2004 dataset used in this study has provided important insight into the nature of offshore flow, how to sample it, and how to model it. Today, new routine measurement systems, such as weather radar and commercial aircraft data, as well as improved use of updated satellite capabilities, combined with new NWP modeling capabilities that can exploit the new observations, should provide improved forecasting capabilities over offshore waters of the U.S. A measurement campaign similar to NEAQS-04, except designed specifically to address offshore wind energy needs, is recommended to take advantage of today's measurement and modeling capabilities and to assess the improvements in predictive skill.

1. INTRODUCTION

Today's global economy depends to a large extent upon a reliable electricity generation and distribution system. In the United States, and indeed around the world, traditional power generation technologies such as coal, natural gas, and nuclear energy are gradually being supplemented, and in some cases replaced, by renewable-energy generation systems. Advancing technology, as well as increasing concern over anthropogenic global climate change, is accelerating this transition to increasingly affordable new systems.

Although more desirable in terms of its smaller anthropogenic emissions footprint, renewable energy generation such as wind, solar, tidal, and geothermal power relies on geophysical phenomena that are variable in space and time. Since power generation infrastructure must generally be fixed in location, it is important for utility managers and decision makers to have access to the most accurate information on wind characteristics at heights required for wind energy. During resource assessment, for example, decision makers can then take into account distance to transmission lines as well as local maxima in the specific resource to make final decisions regarding the placement of new renewable energy power plants.

Utility companies and resource managers have typically used measurement-based datasets for guidance on promising sites for wind and solar power installations. However, these datasets are typically brief in duration, and more importantly are very sparse in their spatial coverage. Waters off the coast of the United States (U.S.) have great potential for generating electrical power from the wind. Because of the lack of appropriate measurements offshore, current estimates of this resource are obtained from numerical weather prediction (NWP) model output (e.g., Schwartz et al. 2010; Musial et al., 2010, http://www.nrel.gov/wind/pdfs/40745.pdf), by extrapolation of shoreline measurements outward into the ocean or other water body (such as the Great Lakes), or by vertical extrapolation from measurements near the water surface. Direct measurement of wind properties within the turbine-rotor layer of the atmosphere over the ocean are rare and have not contributed to attempts to estimate the U.S. offshore wind resource or to describe offshore wind characteristics.

Intelligent development of the offshore wind-energy (WE) capability will require high-quality measurements through and including the turbine layer. These measurements are needed directly for estimating the resource and understanding the flows producing it, but also for evaluating the capabilities of current NWP models to characterize the wind resource, for quantifying the errors involved using such models, and for finding ways to improve the models and reduce the errors. If models could be shown to produce accurate 4-D wind fields, the need to deploy costly instrumentation to the offshore environment could be reduced. For example, a new two-year mean-wind composite or "2-yr climatology" based on an archive of the 3-km-grid High-Resolution Rapid Refresh (HRRR) model runs has been compiled as part of this project, using latest improvements in model physics and hourly data assimilation, and is described in

this report. Obtaining maximum benefit from offshore measurement deployments will require careful consideration of issues such as required sampling frequency, required vertical resolution of the measurements, necessary measurement precision, and horizontal sampling density of the sensors. In the absence of offshore measurements these specifications are unknown.

Measurements from several offshore, ship-borne campaigns are available that use National Oceanic and Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL)'s High-Resolution Doppler Lidar (HRDL). HRDL is a remote-sensing lidar system equipped with a motion-compensation system that has been shown to produce the kind of high-quality profile measurements from shipboard needed to determine properties of the flow in the rotor layer of offshore wind turbines (Pichugina et al 2012; Grund et al. 2001). Recently Pichugina et al. (2012) have presented analyses of HRDL profiles and time-height cross-sections using data from a field campaign, the New England Air Quality Study (NEAQS-2004), designed to study air quality off the New England coast in the Gulf of Maine, a region slated for potential offshore WE development. Measurements were taken during a nearly six-week research cruise as the ship, the Research Vessel *Ronald H. Brown* (*RHB*), sailed around off the New England Coast sampling pollution plumes from cities of the northeastern U.S. These analyses showed strong spatial and temporal variability in the winds aloft over the sea, and such features as low-level jets (LLJs), strong vertical shear through the rotor layer, and stronger winds at night than during daytime.

In this study, called the *Positioning of Offshore Wind Energy Resources*, or POWER, project, the HRDL dataset is used to investigate properties of the flow in and through the turbine rotor layer in the offshore environment, as needed to address sampling and modeling issues critical to WE. An important asset of the NEAQS dataset is the availability of an array of 11 shoreline and inland 915-MHz wind-profiling radars (WPRs), used in this study to assimilate into the models runs, and a profiler also mounted on the *RHB* (which we will refer to as the *RHB* profiler). Runs using two different NOAA models, with and without assimilation of data from the 11 land-based profilers, demonstrate the effectiveness of the ingested data on model forecast skill. HRDL profiles were analyzed for 15-min intervals, and the *RHB* profiler profiles were taken at 60-min intervals as the ship sailed around the Gulf of Maine. This unique data resource is used to evaluate the spatial and temporal wind variability offshore in and above the turbine rotor layer and to verify model output at these heights. The *RHB* profiler dataset is used in this study to validate numerical forecast model output over a layer of the atmosphere up to 2000 m.

Many unanswered questions and many unknowns exist about the offshore environment, especially above the ocean surface at the heights of the turbine blades. Resolving these issues is important for the range of offshore WE requirements, such as resource assessment, real-time forecasting for activities such as power-generation operations and maintenance scheduling, characteristics of the flow to be encountered by the turbines, NWP skill, and many others. To address these will require measurement campaigns and long-term measurement arrays to pro-

vide rotor-layer data on wind properties. But what kind of instrumentation will be required, what kind of measurement density, and how should the campaigns and arrays be designed? What should be the appropriate role for NWP models? It would also be useful, as analysis of the data and results from these campaigns become available, to know how they compare with other measurement-based results over the ocean.

The purpose of the present research program is to use the unique Doppler-lidar dataset from the 2004 New England research cruise, as well as the data from the wind-profiling radars, to get a sense of what the answers to these questions are likely to be. The HRDL and *RHB* profiler measurements themselves are used to illustrate spatial and temporal variability of the wind field aloft over the ocean and thus provide an initial look at the spatial variability seen in the available dataset. An important aspect of this research is to verify the model output against the hub-height-level winds measured by the Doppler lidar over the water as the *RHB* cruised around the Gulf of Maine.

The model verification performed here is an evaluation of model skill in predicting hub-height winds over the ocean, using a total of 12 model configurations based on two modeling systems, the hourly-updated Rapid Refresh (RAP) system and a new hourly-updated version of North America Mesoscale (NAM) forecast system. The two models are run at both normal resolution ("parent" model) and nested down to high-resolution versions (called High-Resolution Rapid Refresh or HRRR and the CONUS-Nest). To test the impact of assimilating upper-air wind data into these models, all four of these versions were run with and without assimilation of hourly wind data from the land-based array of 915-MHz wind-profiling radars. As a check of the robustness of the result of the HRDL-model and the *RHB* profiler-model comparisons, the model forecasts were also verified against the data from the land-based profilers—the data that had been assimilated into the model runs. These comparisons also provide insight into how long the assimilated data improved the forecast model skill at those sites, since such a regional profiling network is being considered as a candidate for improving future offshore wind forecasts. For the four remaining configurations, output from the two modeling systems were averaged to form a two-member ensemble that included simulations with and without profiler assimilation. Since ensembling of short-term model forecasts is an area of active development in NOAA, these results provide a glimpse of future capabilities.

In the next chapter we discuss the NEAQS-2004 *RHB* research cruise, the instrumentation used in the project, and the modeling systems and approaches employed. Chapter 3, the Results chapter, is divided into four sections. In the first section, the horizontal distribution of mean wind speeds at 80-m height for a full two-year period (2012-2013) over the coast of the northeastern U.S. was calculated, based on archived output from the real-time version of the HRRR model currently being run at ESRL. In the second section the Doppler lidar data analyses and lidar-model comparisons are presented. Detailed comparisons of the output of the various fore-

cast model configurations with the land-based and *RHB* profiler data are presented in the third section. A second model-profiler comparison for the different modeling configurations using existing routine NCEP verification software is presented in the fourth section, and these comparisons are then used to assess the effects on model forecast skill of forming the 2-member ensemble. Section 4 is a summary and discussion section, and Section 5 presents conclusions and recommendations.

Major efforts accomplished under funding for this research project were:

- Evaluation of HRRR and CONUS-nest model skill at simulating winds was performed by comparing model output to HRDL and wind-profiler data taken on the *RHB*.
- Assessment of spatial variability of winds (in both horizontal and vertical dimensions) was performed using lidar and wind-profiler data taken on the *RHB* and gridded HRRR and CONUS-nest model output during the study periods.
- Impact of additional observations on wind simulations from HRRR and CONUS-nest model were assessed by running numerical forecast experiments with and without assimilation of wind profiling radar data from NEAQS.
- A two-year "climatology" for offshore wind resource was developed using the real-time HRRR models with hourly data assimilation.
- The above results are used to provide an assessment of the requirements for an offshore observational network for wind resource characterization and to inform future research in this area.
- An hourly-updated version of the NAM forecast system, which cycles both parent and CONUS-nest domains, was developed. This system is known as the NAM Rapid Refresh, or NAMRR.

In completing this research several research tasks were accomplished under this funding. These include:

a. A literature survey of relevant studies was performed to ensure coordination and leveraging of all existing and planned research and data.
b. 2004 wind-profiling radar data were collected, reprocessed with up-to-date software, and quality controlled.
c. Datasets required for running RAP, HRRR, NAMRR and NAMRR CONUS-NEST data assimilations and model predictions were collected and prepared, including sea states, global fields and observations.
d. Data from Navy ships, Quikscat (satellite estimate of marine surface wind speeds), and ESR-1 during July 2004 were collected.
e. Wind profiler data were prepared and included into the Gridpoint Statistical Interpolation (GSI) data assimilation system used by both RAP+HRRR and NAM+CONUS nest systems.

f. The RAP, HRRR, NAMRR and NAMRR CONUS-NEST models were run with and without assimilation of experimental land-based wind-profiler data.
g. Statistical evaluation of model wind simulation skill of RAP, HRRR, NAMRR and NAMRR CONUS-NEST models were performed against conventional observations (onshore as well as offshore shallow and deep water), including preliminary assessment of ensemble means.
h. Statistical evaluation of model wind simulation skill of RAP and HRRR and NAMRR and NAMRR CONUS-NEST models were performed against RV *RHB* observations (lidar and wind profiler).
i. Data from a longer 6-week period of available shipborne Doppler lidar and other sources were analyzed to estimate spatial and temporal scales of offshore wind variability.
j. Model fidelity at simulating meteorological phenomena of importance for shallow- and deep-water offshore wind energy captured in the *RHB* observations were evaluated.
k. Spatial variability of the winds (in both the horizontal and vertical) were assessed, first using the *RHB* ship track data, and then using the high-resolution model simulations that have been validated using the limited offshore data. From this spatial variability analysis, an assessment of the requirements for an offshore observational network for wind resource characterization is provided, including the recommended height range, vertical resolution, and horizontal spacing of the measurement data. In addition, the experience gained in completing this analysis is used to provide an assessment of additional research approaches that could be pursued to quantitatively address the issue of optimal sensor spacing for data assimilation purposes.
l. The following written report includes a summary of the impact of the wind profiling array on offshore hub-height wind simulation; assesses the horizontal and vertical spatial scales, temporal sampling frequency, and measurement precision needed from offshore measurements to resolve meteorological phenomena associated with wind variability; and assesses how these phenomena vary from shallow to deep water.

Completion of these tasks has allowed us to gain new insight into the nature of rotor-level winds in the marine environment as well as into the performance of two forecast models used operationally at NCEP. These models provide 'foundational' forecasts that are downscaled and used by wind-energy practitioners to address industry needs.

2. RESOURCES AND PROCEDURES

a. The Dataset: NEAQS/ICARTT 2004 Measurement Campaign

The offshore, coastal, and inland dataset used in the present study was obtained during the NEAQS-04 field campaign of July-August 2004. NEAQS-04 was the second of two air-quality field studies, following the NEAQS-02 campaign in summer 2002 (Darby et al. 2007, Angevine et al 2006), which was aimed at characterizing local pollution sources in the New England region. In addition to further investigations of local emission sources, NEAQS-04 was also part of a larger research effort, the International Consortium for Atmospheric Research on Transport and Transformations-2004 (ICARTT-04), the major goal of which was to characterize continental outflow of pollution from North America, which may then be transported to Europe. It was therefore necessary to deploy both atmospheric chemistry and meteorological instrumentation to study the key processes producing these transports. Land-based, airborne, and ship-borne instrumentation all contributed to the dataset, as described by Fehsenfeld et al. (2006).

Offshore the major measurement platform was the NOAA *Research Vessel* (*R/V*) *Ronald H. Brown* (*RHB*), which cruised around the Gulf of Maine taking meteorological, air chemistry, and some oceanographic data from 9 July to 12 August 2004. The tracks where the ship steamed during this period are shown in Figure 2.1 *RHB* instrumentation important to the present study included NOAA's High-Resolution Doppler Lidar (HRDL), a NOAA 915-MHz wind-profiling radar, and NOAA's surface-flux measurement package, mounted on a boom protruding out from the bow (Fairall et al. 2006).

White et al. (2007) studied the meteorological conditions over New England relative to the climatology of the region. According to this study, the summer of 2004 was abnormally cool, cloudy, and wet. The mean synoptic flow pattern just below 1 km ASL indicated troughing over the northeastern U.S., and the frequency of cold-frontal passages was also abnormally high. Figure 2.2 shows the rainfall associated with these fronts during July and early August. Of interest to the pollution aspect of NEAQS-04, these conditions produced very good air quality, with very few high-ozone pollution days in the northeastern U.S.

The periods selected for study were based primarily on the availability of several days in a row of HRDL data, because one of the prime objectives of the POWER project was to directly compare modeled vs. measured winds in the turbine-rotor layer over the ocean, which is what the HRDL shipboard system provides. Because of the frequent frontal passages during summer 2004 and the air-quality objectives of the cruise, the NEAQS-04 campaign provided one stretch of more than a week when HRDL data were continuous and several other periods of 3-4 consecutive days. The first study period selected was 6-12 August, corresponding to the longest lull

between frontal passages according to White et al. (2007). For the second study period (10-17 July), two of the shorter periods separated by a day of rain (14 July: see Figure 2.2) were chosen, so that model runs could be performed for a week straight without having to restart. Although the intervening rainy period was not originally contemplated for analysis, we did include this period, which turned out to be interesting, in the model evaluations.

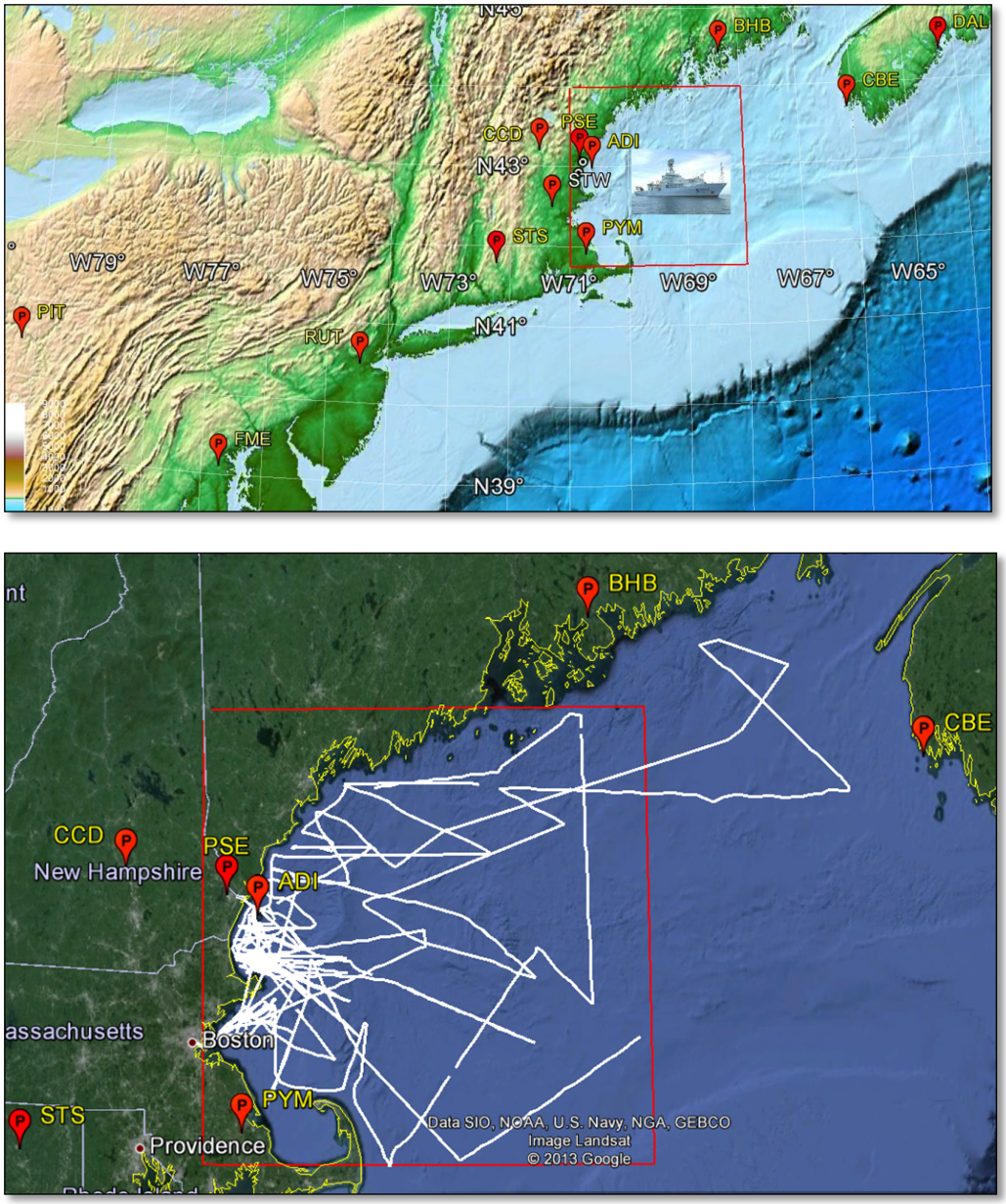


Figure 2.1. (Top) Google earth image of northeastern U.S., showing study area (red rectangle) and profiler locations (red pins); (bottom) Ship tracks for the entire NEAQS-2004 campaign, from 9 July to 12 August 2004.

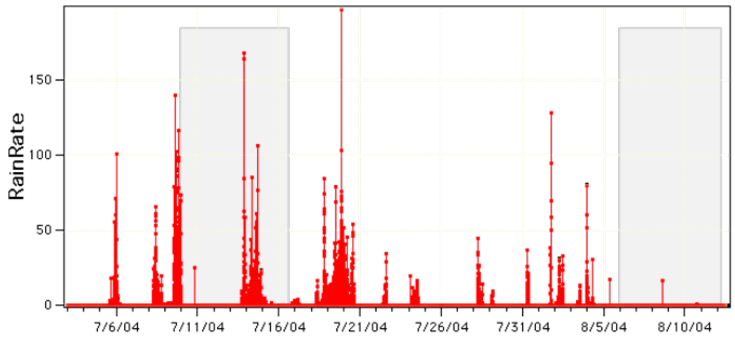


Figure 2.2. Rain rate during the NEAQS-04 experiment as determined from ship-borne measurements. The gray-shaded areas indicate the two periods selected for study. The two gray-shaded boxes indicate periods selected for model validation. Analysis of NCEP/HPC daily weather maps and surface meteorological data recorded on Appledore Island (White et al., 2007) on 13-14 July shows an eastward-propagating cold frontal passage at surface levels.

b. Measurement Systems

1) High-Resolution Doppler Lidar

The High-Resolution Doppler Lidar (HRDL) is a scanning, coherent, pulsed Doppler lidar designed and operated by NOAA/ESRL for atmospheric boundary-layer research, as described by Grund et al. (2001). Deployed on board the NOAA *R/V RHB*, HRDL was operated over the Gulf of Maine 24-h per day during the NEAQS 2004 field campaign (Pichugina et al. 2012).

Doppler lidar is similar in concept to Doppler weather radar, although the scattering targets for its near-IR signal are aerosol particles rather than hydrometeors. This makes the Doppler lidar useful for mapping the wind field in clear air, since aerosol particles are widely distributed in the lowest 2-4 km of the atmosphere, and near the ocean surface, sea-salt particles are especially effective backscatter targets. HRDL provides range-resolved measurements of the radial or "line-of-sight" wind, i.e., the component of the velocity parallel to the beam, and aerosol backscatter, at a range resolution of 30 m. The minimum range of the instrument along the beam is <200 m. Technical attributes of this lidar system are given in Table 1. HRDL employs a well collimated optical beam, so its signal does not suffer from interference from side-lobe returns.

Table 1. Technical characteristics of NOAA/ESRL Doppler Lidar

Characteristics	HRDL
Wavelength	**2.02 μm**
Pulse energy	**2.0 mJ**
Pulse rate	**200 Hz**
Range resolution	**30 m**
Velocity precision	**~10 cm s^{-1}**
Time resolution	**0.5 s**
Minimum range	**189 m**
Maximum range	**3-8 km**

A major obstacle to obtaining accurate wind profiles from the high-precision lidar measurements using these techniques is compensating for the pointing error and along-beam platform velocity that are due to ship motions, including those induced by wave action. To accomplish this, the lidar is equipped with a system that 1) determines the orientation and motion of the platform and then actively stabilizes the pointing of the scanner and 2) corrects the lidar velocity measurement by estimating and removing the platform motion component along the lidar pointing direction. Several tests are used to determine the accuracy of the pointing-angle corrections (Pichugina et al. 2012) both in the laboratory using specially constructed platform to simulate pitch, roll and other ocean wave movements, and in the open ocean. These tests indicated that implementation of the real-time motion-compensation system allowed maintaining the chosen scan parameters in the world frame to within ~1^0 or less under conditions normally encountered. For more recent deployments, additional improvements have been implemented that allow compensation for the ship's motions to within 0.5^0.

The accuracy of lidar-measured wind profiles can be assessed by comparison with profiles measured by other instruments also operated from the ship. A strong correlation between rawinsondes and HDRL horizontal and lateral wind components with correlation coefficients of 0.97 and 0.98 respectively combined for all heights above 100 m was reported in Wolfe et al. (2007). We compared lidar and rawinsondes at each height and the correlation-coefficient profile is shown in the left panel of Figure 2.3. The right panel shows the number of reciprocal points used for the comparison at each height from the surface to 1500 m. Slightly reduced correlation with $R^2 < 0.96$ observed at 3 lowest heights of 30, 60, and 90 m is a result of the influence of the ship's atmospheric wake on the rawinsonde measurements at these levels.

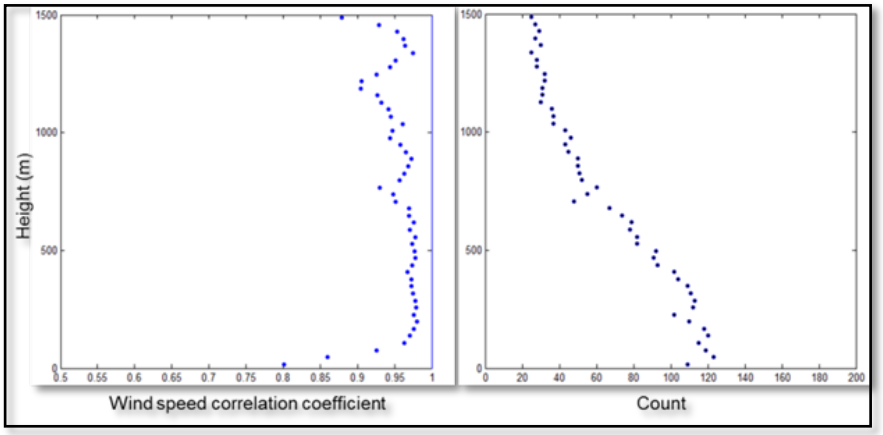


Figure 2.3. Correlation coefficients (left) between lidar and rawinsonde wind speed at 30 heights above the water surface and (right) number of points at each height used in calculations.

Comparison of motion-compensated, sonic-anemometer ''flux-wind'' (Fairall et al. 2006) measurements at 17m and lidar measurements at the closest height of 12.9 m (Figure2.4), demonstrates high agreement between the two instruments with correlation coefficient of 0.92 for wind speed and 0.98 for wind direction.

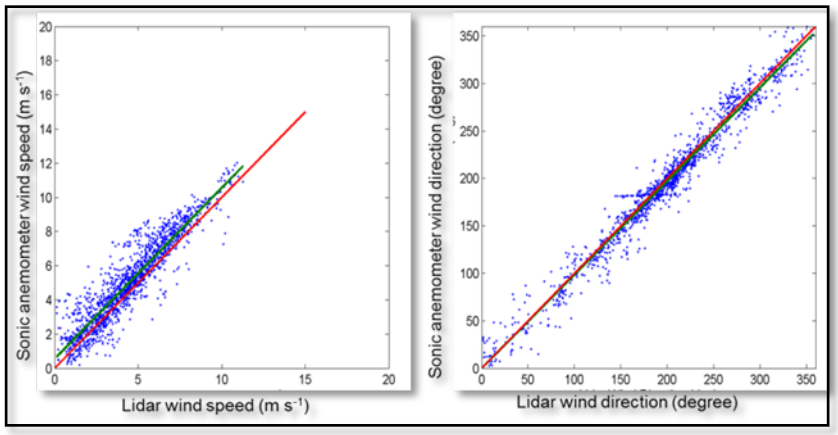


Figure 2.4. Scatter diagrams show comparisons between (left) wind speed and (right) wind direction measured by lidar and sonic anemometer. Green solid line is the best fit with the linear regression slope of 1.00±0.02 (wind speed) and 0.99±0.01 (wind direction). The bias of the regression is 0.55±0.11 and 0.15 ± 2.02 for wind speed and wind direction respectively. The red line is the 1:1 line.

As an example, lidar profiles were color coded and plotted on a Google Earth map (Figure 2.5) by Pichugina et al. (2012) to demonstrate the nature of spatial changes noted during a day of the research cruise along the ship's track.

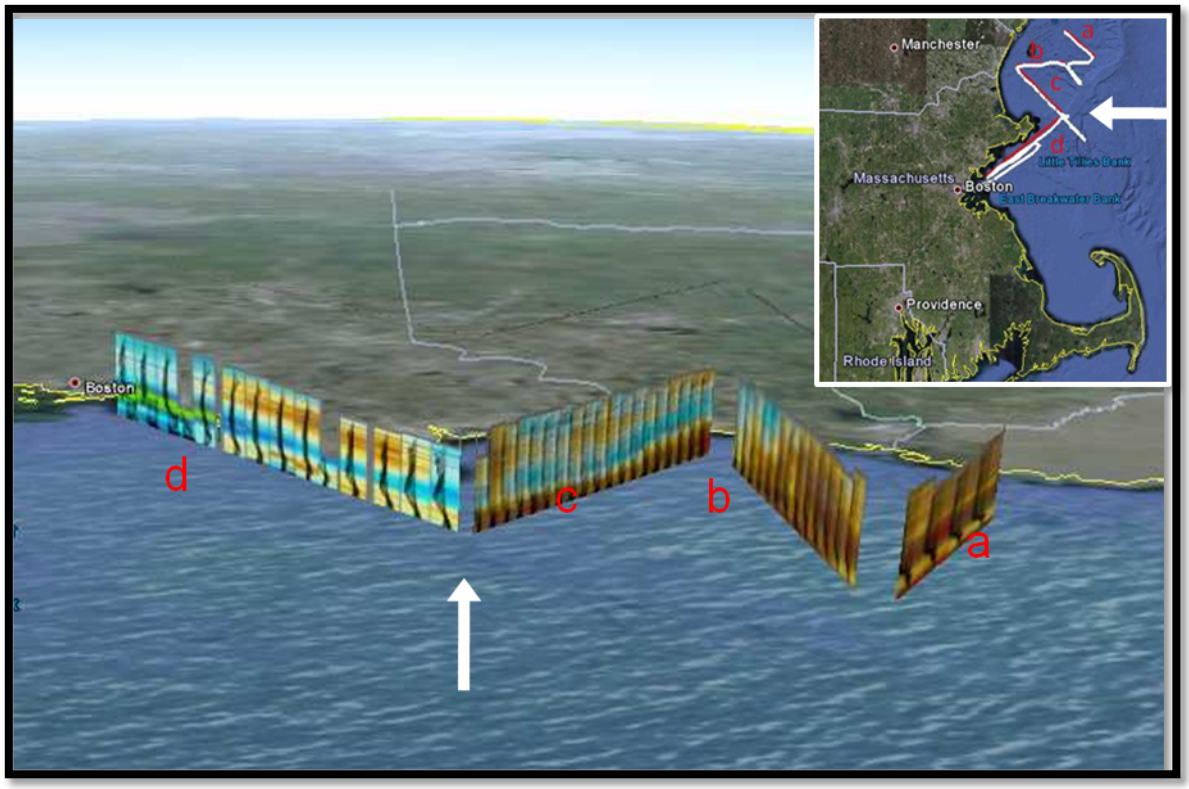


Figure 2.5. Time-height cross sections of color-coded wind speed for four segments of the ship track on 21 July 2004, superimposed on a Google-Earth image of the Gulf of Maine region looking toward the west. Color coding of wind speed runs from 0 (green) to 10 (red) m s^{-1}, and 15-min wind-speed profiles are shown from 10 to 1500 m above the water surface. Insert shows location of the four ship-track segments, plotted on an image of the New-England coastline, where north is up and white arrow indicates point of view. Image provides an overview of the spatial and temporal variability of the wind along the track.

2) Wind-Profiling Radars

The instrument data set involved in this project included a network of 11 inland wind-profiling radars (WPRs) (Strauch et al., 1984; Wilczak et al., 1996) used for NEAQS during the summer of 2004 (Figure 2.6) and one WPR located on the *RHB* (Wolfe et al., 2007).

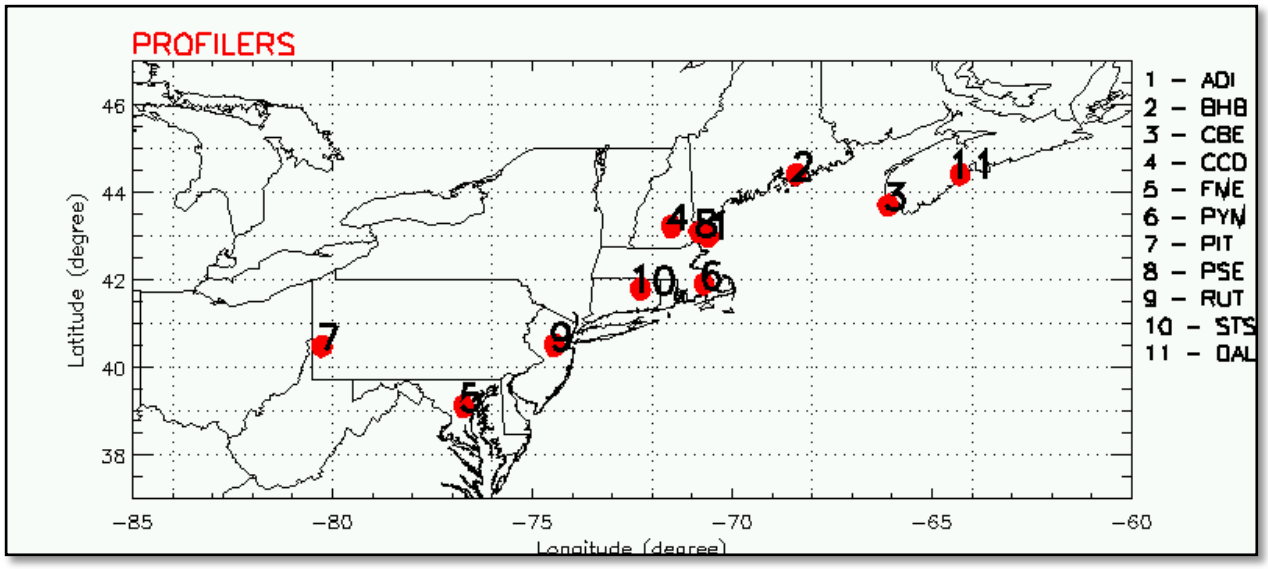


Figure 2.6. Location of the inland wind-profiling radars in New England.

Wind profilers are Doppler radars that do not require a hard target to receive a signal. In the clear atmosphere wind profilers receive backscattered signals from refractive index turbulence (Carter el al., 1995). All WPRs used in NEAQS-2004 were 915-MHz radars (33-cm wavelength) with the lowest range gate around 100 m. Two sampling modes are used by the instruments: the high-resolution mode uses a resolution of 60 m and the low-resolution mode uses a resolution of 100m. The maximum height with detectable signal varies with atmospheric conditions (a stronger signal occurs in a moister atmosphere), but the coverage typically ranges from the lowest level up to around 1.5 km above ground level (agl) for the high-resolution mode and up to around 4 km agl for the low-resolution mode. The different operational specifications used for the various NEAQS-04 radar wind profilers are summarized in Table 2.

Table 2. Operational specification for the radar wind profilers deployed during NEAQS-2004.

Site	High resolution specifications			Low resolution specifications		
	1st height (m)	Gate spacing (m)	# of gates	1st height (m)	Gate spacing (m)	# of gates
ADI	123	58	38	109	101	38
BHB	123	58	38	130	100	38
CBE	123	58	38	130	102	38
CCD	122	58	38	129	102	38
FME	124	55	36	282	96	40
PYM	138	58	38	129	102	38
PIT	114	58	45	39	96	58
PSE	85	60	72			
RUT	124	55	36	158	96	59
STS	71	52	95	272	110	71
DAL	199	96	33	275	192	30

Figure 2.7. displays a 24-hour time-height cross section of data from the Pease (NH) 915-MHz WPR (PSE), on August 12th, 2004, showing the onset and cessation of a low-level jet and its vertical structure. The depth of the atmosphere that the WPR was able to observe on this day was approximately 3km agl.

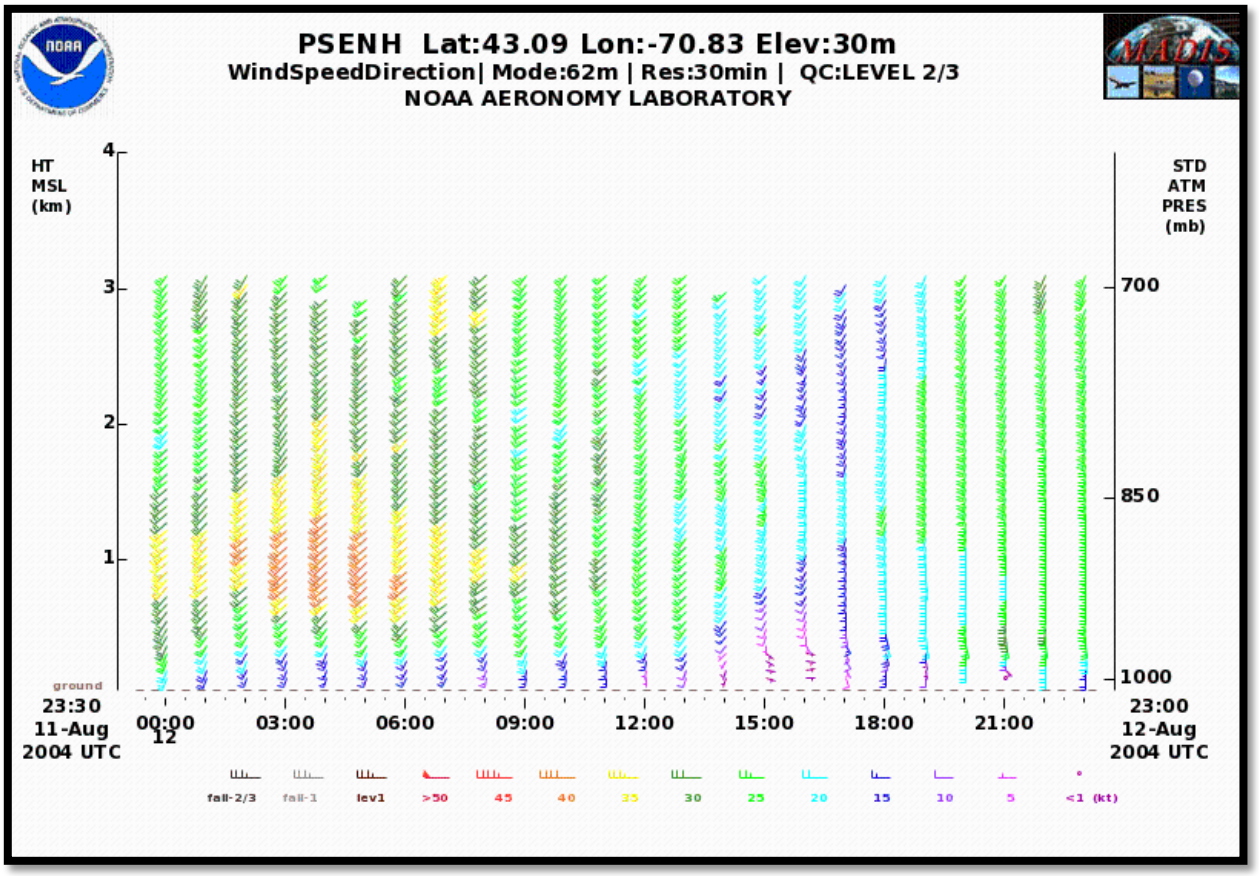


Figure 2.7. 24- hour time-height cross-section of hourly averaged winds from the 915-MHz Pease (PSE) Wind-Profiling Radar. The onset of a nocturnal low-level jet occurs between 00-01 UTC (20-21 LT), and ends between 08-09 UTC (04-06 LT) the next morning.

c. NOAA Model Forecast Systems

Four numerical weather prediction (NWP) systems run at NOAA/National Centers for Environmental Prediction (NCEP) and ESRL/Global Systems Divison (GSD) provide foundational meteorological predictions at time and space scales useful to the WE industry; these are the Rapid Refresh, or *RAP*, model and its associated nested High-Resolution Rapid Refresh, or *HRRR*, which are based on the Advanced Research version of the Weather Research and Forecasting (WRF-ARW) model, and the North American Mesoscale forecast system, or *NAM*, model and its associated CONUS Nest, which are based on the Nonhydrostatic Multiscale Model on the B grid (NMMB). The RAP and the NAM are run at horizontal grid intervals of 13 and 12 km, respectively. Each of these models is also nested down to a finer mesh size, with the NAMRR CONUS-NEST having a grid interval of 4 km, and the HRRR having 3 km grid spacing.

These models are continually undergoing improvement and are tested in experimental versions both at NCEP (for the NAM) and at ESRL/GSD (for the RAP and HRRR). An experimental version of the NAM (which produces 6-hourly forecasts out to 84 h) is the NAM Rapid Refresh, or *NAMRR*, developed to provide hourly assimilation and hourly forecasts, similar to the RAP capability. The NAMRR version is used in this study. Both RAP and NAMRR also ran their nested components, which included the HRRR as well as the NAMRR CONUS-NEST domains (Figure 2.9).

For the POWER study, *experimental* versions of the four models have been run for verification against the Doppler lidar and WPR measurements in and around the Gulf of Maine. These profiles were taken by the HRDL aboard the *RHB* in the atmospheric layer that would be occupied by turbine rotor blades, and by the WPRs both on the *RHB* and deployed on land along the coast.

In addition to the experimental retrospective model simulations conducted within POWER, we also used an archive of hourly-updated real-time HRRR forecasts over a two-year period (2012-13) to characterize the 3-km spatial variability of the low-level wind field. This portion of the study uses the real-time HRRR with its associated full CONUS domain (Figure 2.8).

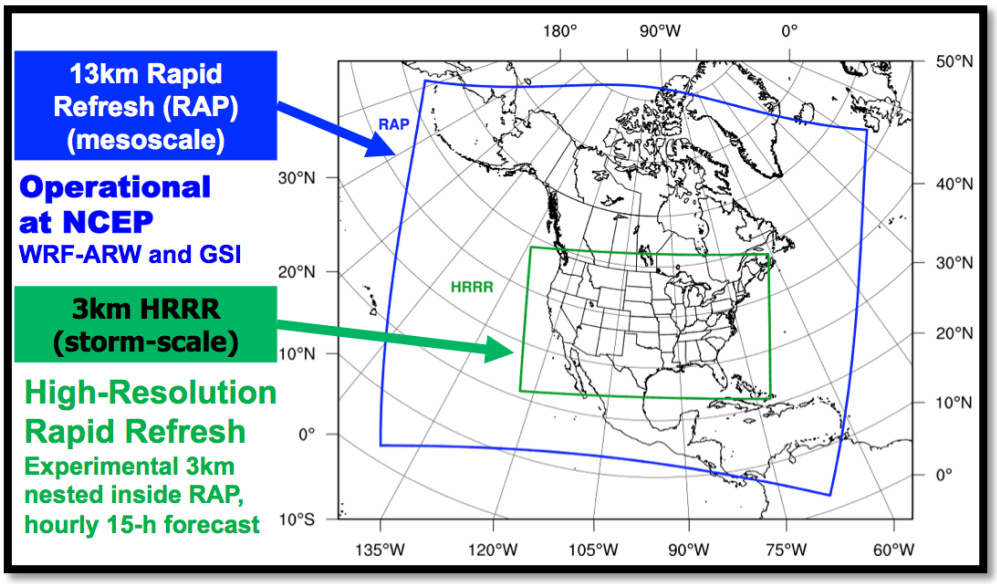


Figure 2.8. Domains for RAP and HRRR models used in POWER program: operational/real-time RAP and HRRR domains. HRRR covers entire lower 48 U.S.

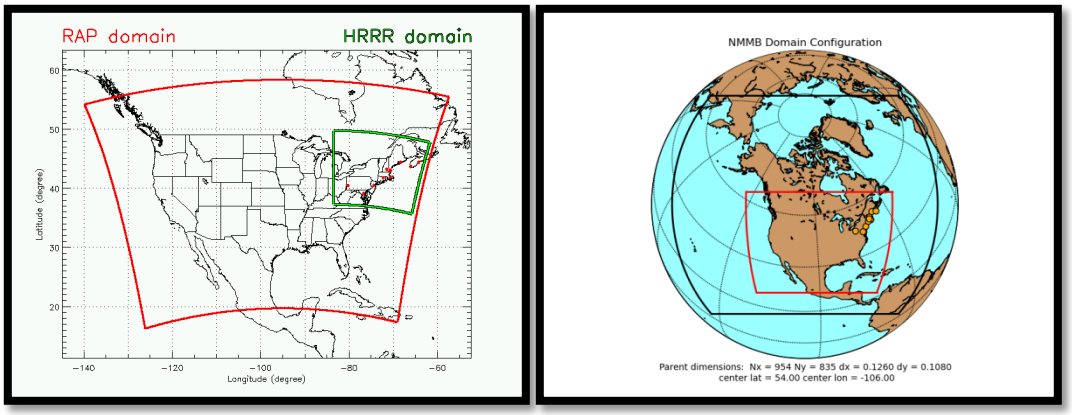


Figure 2.9. Domains for NWP models used in POWER program: left) Smaller experimental domains for RAP and HRRR for POWER retrospective experiments - RAP in red and the HRRR domain for POWER in green. (right) Domains for NAMRR parent in black and NAMRR CONUS-NEST in red on the right used in POWER and in operations.

1) Descriptions of models used and procedures

a) RAP and HRRR description

The RAP serves as NCEP's regional short-range rapidly updating forecast system, which provides hourly updated forecasts out to 18 hours but was only run out to 12 hours for the POWER study. The RAP replaced the operational Rapid Update Cycle (RUC) model at NCEP in May 2012. The model component of the RAP is based upon the WRF-ARW model (Skamarock 2008), and the data assimilation component uses the 3D-variational GSI (Wu et al. 2002).

The version of the RAP implemented for POWER data denial experiments features a 13 km C-grid domain covering a subset of North America, and CONUS (130-grib) subsets were saved for the POWER study (Figure 2a). The RAP model code used for the POWER project used a similar WRF model configuration and GSI assimilation configuration as that used in the RAP(v1) run operationally at NCEP from May 2012 until Feb 2014 when it was replaced by RAPv2 (not used for POWER). Boundary conditions for the RAP were obtained from the previous cycle's forecast from the CFSR system (Saha et al. 2010). The RAPv1 version used in POWER is described also at http://ruc.noaa.gov/pdf/RAPbrief.NCEP-Dir-20Mar2012.pdf .

The RAP (http://rapidrefresh.noaa.gov) uses a modified version of the WRF-ARW with explicit mixed-phase bulk cloud microphysics originally described by Thompson et al. (2008) to parameterize the effects of moist processes. Deep sub-grid scale convection is parameterized by the Grell 3D scheme and shallow-convective processes are parameterized by the Grell shallow-cumulus scheme. Land-surface physics are parameterized by the RUC LSM (Smirnova et al. 1997, 2000b). The radiative transfer in the shortwave spectrum is parameterized by the Goddard scheme (Chow and Suarez 1994) and the longwave spectrum is parameterized by the Rapid Ra-

diative Transfer Model (RRTM; Mlawer et al. 1997). The Mellor-Yamada-Janjic (MYJ; level 2.5) boundary layer scheme (Janjic 2001) parameterizes the turbulent mixing. The exchange coefficients regulating the fluxes of heat, moisture, and momentum between the land and atmosphere/water are prescribed by Janjic (1994). Note that since the POWER project began, both the RAP and HRRR have switched to the Mellor-Yamada-Nakanishi-Niino (MYNN) PBL and surface layer scheme. This new PBL scheme has resulted in improved low-level wind forecasts over land but has not yet been thoroughly tested over water. Table 3 summarizes the RAP model configuration used for POWER.

The versions the RAP implemented for POWER, as well as in operations, used the so-called "partial cycling" procedure. Partial cycling for the RAP involves a twice-daily 6-hr spin-up cycle from an initial condition taken from a 3 hr GFS forecast, valid at 03 and 15 UTC. The 1-hr forecast from the 6th cycle is then injected into the regular hourly cycled system at 09 and 21 UTC. Note that for the RAP, only the atmosphere component uses this partial cycling technique, whereas the soil moisture and temperature fields are continuously cycled. This is done so that the soil state is kept physically consistent with the RUC LSM physics and the soil vertical level configuration.

The RAP uses a digital filter initialization with a filter window length of 40 minutes, which is invoked at the beginning of every hourly cycle. The RAP also includes a cloud analysis procedure, using satellite data and surface ceiling observations, to initialize an accurate three-dimensional cloud field. This aspect was also of limited value since there was very little satellite data available during the POWER study period.

The HRRR (http://ruc.noaa.gov/hrrr) features 3-km grid spacing with a domain covering the contiguous U.S. but given severe data storage limitations, a truncated version covering the northeastern U.S. was used for the retrospective experimental portion of the POWER study (Fig 2.9). The HRRR is not yet run operationally at NCEP but is planned for implementation there in 2014. However, the HRRR is run year-around in a quasi-operational 24/7 experimental mode at NOAA/ESRL and already has a wide user base, including NOAA Weather Forecast Offices and private sector organizations. A primary purpose of the HRRR is to improve the operational capability of forecasting high-impact convective storms, which play an important role in the ramping of low-level winds. The version of the HRRR in development during POWER did not perform additional data assimilation on the 3-km grid. The initial and boundary conditions were obtained by direct interpolation from the RAP. The HRRR was run hourly, out to 12 hours, within ESRL's high-performance computing facility.

The forecast model component of the HRRR, like the RAP, uses a modified version of the WRF-ARW. Since most convective processes can be adequately resolved at 3-km grid scales, no deep- or shallow-convection schemes are used. The rest of the physical processes are parameterized using the same schemes employed in the RAP (above). Table 4 summarizes the HRRR model configuration.

For the 2012-2013 HRRR "climatology" developed for the POWER offshore wind resource, the HRRR was run over the full CONUS HRRR domain as shown in Figure2.8 This dataset consists of hourly 2-h HRRR forecasts, very close to analysis time and consistent with latest observations but also allowing local details (e.g., coastal circulations) to develop at 3-km resolution. The 2-year HRRR "climatology" data set consists of approximately 15,000 realizations (~85% of hourly runs over the 2-year period).

Table 3. 13-km Rapid Refresh domain configuration for POWER profiler data impact experiments.

13 km RR/RAP Description	Configuration
Points in x, y, z directions	758, 567, 51
Microphysics parameterization	Thompson et al. (2008)
Boundary layer parameterization	Janjic (2001)
Convective parameterization	Grell 3D/Grell shallow-cumulus scheme
Long/short wave radiation parameterization	Chow and Suarez (1994)/Mlawer et al. (1997)
Land surface model	Smirnova et al. (1997, 2000b)

Table 4. 3-km HRRR domain configuration for POWER profiler data impact experiments.

3 km HRRR Description	Configuration
Points in x, y, z directions	(Truncated to) 520, 450, 51
Microphysics parameterization	Thompson et al. (2008)
Boundary layer parameterization	Janjic (2001)
Convective parameterization	Turned off
Long/short wave radiation parameterization	Chow and Suarez (1994)/Mlawer et al. (1997)
Land surface model	Smirnova et al. (1997, 2000b)

b) NAMRR and NAMRR CONUS-NEST: Description

The NAM serves as the National Weather Service's regional short-range NWP system which provides forecasts out to 84 hours four times at day at 00, 06, 12, and 18 UTC. The current configuration of the operational NAM was implemented in October of 2011 and is based upon the Nonhydrostatic Multiscale Model on the B grid (NMMB; Janjic, 2003; Janjic, 2005; Janjic and Black, 2007; Janjic and Gall, 2012).

The version of the NAM implemented for POWER, known as the NAM-Rapid Refresh (NAMRR), was a part of an ongoing effort to transition the operational NAM system from one that provides forecasts 4 times per day at 00, 06, 12, 18 UTC to one that produces an hourly analysis *and* forecast (Figure 2.10). Substantial development during the POWER project was undertaken to develop this capability. The NAMRR featured two domains (Figure 2b), a parent 12-km domain and a one-way nested 4-km domain covering the contiguous United States (CONUS). The coverage of both of these domains is identical to that covered by the operational NAM[1]. The NMMB model code used for the POWER project was a more advanced version of the operational NMMB and came from NCEP/EMC's real-time, parallel-test developmental code. The general configuration of the parameterization schemes used in both the 12-km and 4-km domains may be found in Tables 5 and 6, respectively.

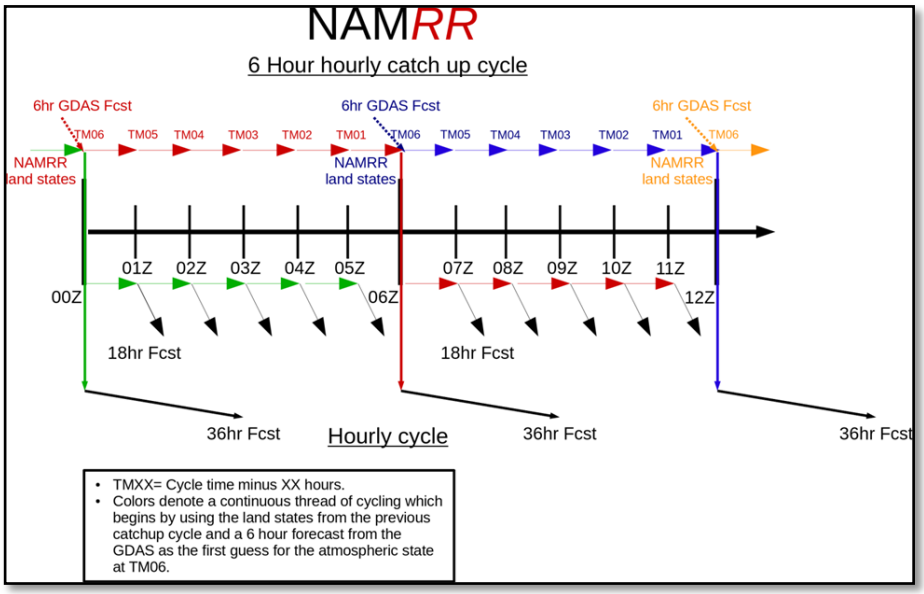


Figure 2.10. NAMRR data assimilation cycling diagram for POWER. Note that in the case of POWER one may replace the GDAS 6 hour forecast with the 6 hour forecast from the CFSR data assimilation system, which is similar to the GDAS.

[1] The operational NAM runs with 6 domains in total. The NAMRR only runs with the parent 12 km and the 4 km CONUSNEST domains.

Table 5. 12-km NAMRR domain configuration.

12 km NAMRR Parent Description	Configuration
Points in x, y, z directions	954, 835, 60
Microphysics parameterization	Ferrier et al. (2002, 2011)
Boundary layer parameterization	Janjic (2001)
Convective parameterization	Janjic (1994)
Long/short wave radiation parameterization	Iacono et al. (2008), Mlawer et al. (1997
Land surface model	Ek et al. (2003)
Gravity wave drag parameterization	Alpert (2004)

Table 6. The NAMRR 4 km CONUSNEST configuration is the same as the NAM 12 km parent domain configuration (Table 3) except with the differences noted here.

4 km CONUSNEST Description	Configuration
Points in x, y, z directions	1371, 1100, 60
Convective parameterization	Janjic (1994): Modified to be less active for higher resolution
Gravity wave drag parameterization	None

The NAMRR's configuration is broken up into two run-types, 'catch-up' and 'hourly' (Figure 2.10). Catch-up types occur at 00, 06, 12, and 18 UTC and are similar to the idea of the "partial cycling" procedure that was introduced in the December 2008 NAM implementation (Rogers et al. 2009). In POWER each 'catch-up' step starts at TM06 using a 6 hour forecast from the CFSR data assimilation system (Saha et al. 2010) for the first guess atmospheric state for the analysis while cycling the model land states from the previous catch-up cycle's 1-hr NAMRR forecast from TM01. Following the initial analysis at TM06, a series of 1-hr forecast and analysis cycles is conducted until the final analysis is finished at TM00 (e.g. the series of red arrows in Figure 2.10). Once the catch-up's analysis/forecast cycling has finished (e.g. at 06 UTC) a 36-hr forecast with the 12-km and 4-km domains was conducted for POWER. For the 'hourly' run-type (occurring at all other hours, e.g. 13 UTC, 14 UTC, etc.) an analysis was performed based upon a 1-hr forecast from the previous cycle. Following the analysis an 18-hr forecast was conducted with both the nest and the parent domain. It should be noted that, unlike the operational NAM, both the 12-km parent domain and the 4-km CONUSNEST are cycled in this system. For this

retrospective experiment, boundary conditions for the 12-km parent domain were obtained from analyses generated by the CFSR system.

The issue of initialization and reducing the effects of imbalances introduced during the data assimilation step has been a longstanding challenge in NWP (e.g. Daley, 1991). Therefore a digital filter initialization technique was adopted in the NAMRR for its hourly data assimilation, similar to that used in the RAP. Such an approach can help to produce an initial atmospheric state that is balanced within the context of the model's dynamics and can help mitigate the accumulation of excessive noise while accelerating the model adjustment process (Huang and Lynch, 1993). Furthermore the use of digital filter initialization has also been considered a necessity for the mitigation of noise in the implementation of hourly, rapidly updating forecast models (e.g. Benjamin, 2004a). In the POWER version of the NAMRR system a diabatic digital filter (Lynch et al., 1997) was applied immediately after each analysis, for both domains, using a filter window length of 40 minutes.

Special output from the NMMB forecast's nearest grid points at POWER profiler and *RHB* ship locations was also provided from the NAMRR's 12 km and 4 km domains in Binary Universal Form for the Representation of meteorological data. These data were shared with ESRL/PSD for use in forecast verification performed at POWER profiler and *RHB* locations.

All POWER experiments were started 24 hr prior to each case-study period to allow the NAMRR's land states to adjust from those initially specified by the much coarser CFSR system, which has an approximate grid-spacing of 35 km.

c) Common sea-surface temperature and snow analysis datasets for RAP/HRRR and NAMRR/CONUS-NEST

For both the RAP/HRRR and NAMRR/CONUS-NEST models, SSTs were updated once daily at 18 UTC based upon the most recent Real-time global 0.5 degree SST analysis created by the NCEP/EMC Marine Modeling and Analysis Branch (Thiébaux et al. 2003), which was considered the best available analysis to use at the time (Grumbine 2012, personal communication). At the same time snow cover and sea ice were updated from the most recent data available from the Interactive Multisensor Snow and Ice Mapping System. The northern hemispheric snow depth was also updated based upon analyses provided by Air Force Weather Agency.

d) HPC and data storage requirements

All RAP/HRRR and NAMRR/CONUS-NEST simulations were run on Zeus, the NOAA research and development high-performance computing system. NAMRR/CONUS-NEST model forecast jobs used 1224 processors to run the nested configuration of the 12-km parent and 4-km CONUSNEST domains. The GSI used 240 processors for each data assimilation step that was run for the 12 km and 4 km domains. All forecasts initialized at 00, 06, 12, and 18 UTC ran out to 36 hours. All other forecasts issued at other times, e.g. 01 UTC, ran to 18 hours. Both 12- and 4-km domains produced hourly forecast output.

The NAMRR simulations are currently occupying approximately 148 Terabytes of archived disk space. This exceptionally high disk usage reflects the retention of several very large files, which have been saved to restart model forecasts if necessary. Furthermore, the addition of the 4 km CONUSNEST domain leads to a substantial increase in the amount of required disk space.

2) Data Assimilation Procedure

NWP is an initial value problem and the atmosphere is a nonlinear, chaotic system which, when modeled, exhibits a strong sensitive dependence on initial conditions (Lorenz 1963). Therefore it is important that the best available initial conditions be used to initialize NWP models in order to yield the best possible forecasts. The process in which the best available initial conditions are obtained is through a procedure known as data assimilation, which combines a model forecast with observations to provide an estimate of the current state of the atmosphere. This estimate, i.e. the analysis, is then used as the initial state from which forecasts are initialized.

Both the RAP and NAMRR forecast systems use the NOAA Gridpoint Statistical Interpolation system, under the context of 3DVar, for data assimilation (GSI; Wu et al. 2002). The GSI analyzes the following variables: streamfunction, velocity potential, surface pressure, temperature, and normalized-relative humidity [a multivariate relation involving specific humidity, temperature, and pressure (Holm et al., 2002)]. The GSI is a complex variational data assimilation system, which is capable of assimilating a diverse set of observations. Such observations include, but are not limited to, radiosondes, wind profilers, Doppler radar radial velocities, satellite radiances, surface observations, wind turbine nacelle, and tall tower observations. For more detailed information about the GSI system, please see Appendix A.

The GSI settings and configuration used for both RAP and NAMRR are described in the following subsections.

a) RAP/HRRR – GSI settings

The RAP data assimilation system used for POWER is a developmental version of GSI, which includes updates from the operational RAPv1. RAP-GSI assimilates all standard observations as well as POWER-special wind observations.

Observation error statistics and background error statistics used for the RAP GSI analysis were configured to match those used in operations, with the exception of reduced observation errors for highly quality-controlled wind profilers. The observation errors used in operations were specified following the technique of Desroziers and Ivanov (2001).

The 2013-2014 real-time developmental version of the HRRR does perform data assimilation on the 3-km grid, but for the POWER experiments, no 3-km data assimilation was performed since no radar observations were available. The HRRR was initialized solely from the data assimilation performed within the RAP by means of interpolated RAP analyses used to generate initial and boundary conditions.

b) NAMRR – *GSI settings*

The implementation of GSI used during POWER for the NAMRR was a developmental version of the operational version of the GSI data assimilation code. The NAMRR performed full data assimilation on *both* 12-km and 4-km domains throughout the duration of the POWER project.

The background error statistics used in NAMRR-GSI were the same as those used in the operational NAM for cycles at TM05-TM00 (Figure 2.10). These background error statistics were derived using 60, 3-hr forecast pairs based upon the method of Houtekamer et al. (1996). At TM06, the beginning of the NAMRR catch-up window (Figure 2.10), the background error statistics from the GDAS system are used since the first guess forecast at this time corresponds to a six hour forecast from the CFSR system, which has its roots in the same forecast model and assimilation framework as the GDAS.

Observation errors for all conventional observations assimilated in POWER were set to be identical to the errors used in the operational NAM, which specified the errors following Desroziers and Ivanov (2001). More specifically, this approach estimates the observation error standard deviations adaptively through the application of successive analyses using observations with random perturbations. However, in tests assimilating experimental profiler observations as a part of the DOE-sponsored WFIP project, it was found that the observation errors used in operations for these data should be reduced due to the extra quality control that the WFIP field observations underwent. Therefore, based upon previous experience with the WFIP project, the observation error standard deviations for the multi-agency profiler observations for NAMRR were set to 2 m s^{-1} at the surface and up to 700 hPa. At 700 hPa the error was specified to increase by 0.2 m s^{-1} every 50 hPa up to a maximum of 5 m s^{-1}.

In addition to the reduction of observation errors for profiler observations, the gross error check for these was also modified relative to that used by the operational NAM. The gross error check in GSI takes the following form, note that the non-boldface notation indicates the application of the check for an individual observation

$$| y_{obs} - Hx^{f} | > \sigma_{o}\omega$$

where $\sigma_o$is the observation error standard deviation and $\omega$is the gross error rejection threshold. If the absolute value of the innovation for a single observation meets the specification of the inequality above then the observation is rejected. This was found to be problematic in early tests; the GSI would reject many observations that were known to be good. Therefore ω for the multi-agency profiler observations was set such that observations which differ from the background by more than 14 m s^{-1} are rejected.

Finally, a third modification to the treatment of RASS and multi-agency profiler observations was made. In operations a very short time window of ±6 minutes of the analysis time is used to select the observations to assimilate. For POWER a larger time window of ±21 minutes of the analysis time was used to ensure successful ingest of the POWER observations into the NAMRR. Aside from the modification of the time-window, no other adjustments were made to the treatment of RASS observations.

3) Simulations with and without profiler assimilation

a) CFSR observation processing for data assimilation

Both conventional and satellite observations from the CFSR project were used here in the POWER project (satellite radiances were only used in the NAMRR). Since these observational datasets were produced at six hourly intervals for the CFSR, substantial work to re-process these observations into an hourly output format compatible with the cycling frequency of the NAMRR and RAP was required. Furthermore, additional marine data from the International Comprehensive Ocean-Atmosphere Data Set were also included into the conventional observation data sets for assimilation. Here conventional simply means non-satellite radiance based observations.

Since the POWER retrospective study takes place in the summer of 2004, obtaining correct satellite radiance bias correction coefficients for use in the NAMRR-GSI data-assimilation system was exceptionally problematic, as these terms are necessary for the successful assimilation of radiance observations but can take months for the system to spin up properly. Fortunately the POWER project coincided with a recent advancement in the GSI system which allows for the spin up of radiance bias correction terms in as little as a day of hourly cycling (Zhu et al. 2013). Therefore, the NAMRR assimilated all available satellite radiance observations (subject to quality control procedures within the GSI), on both the 12- and 4-km domains.

The inclusion of radar reflectivity and radial velocity observations for this project was also investigated for the POWER 2004 study period. However, gridded mosaics of radar reflectivity from the WSR-88D network suitable for the cloud analysis and digital filter initialization schemes employed by both the NAMRR and the RAP were unavailable during POWER. Similarly, suitable level II Doppler radar radial velocity observations from the WSR-88D network were also unavailable. Therefore, aside from velocity azimuth display (VAD) profile winds, observations from the WSR-88D network were not assimilated in either RAP or NAMRR systems.

NCEP operational observation processing was used for encoding the POWER profiler and RASS data into the file format used by the GSI assimilation system, known as prepBUFR (prepared Binary Universal Form for the Representation of meteorological data). The encoding pro-

cess followed the method used in operations at NCEP and involved tanking the data, making dump files in raw BUFR format, and then encoding the observations into prepBUFR.

Once these observations were encoded into prepBUFR, the new prepBUFR files containing all conventional observations in addition to the special POWER profiler and RASS observations were shared with ESRL/GSD/PSD. These prepBUFR data were then used for assimilation for the NAMRR and RAP forecast systems.

3. RESULTS

a. Horizontal Composite Maps for Offshore Wind Speed, Using 2-year HRRR Archived Dataset

NOAA/ESRL, within its Global Systems Division (GSD), has been running an experimental, hourly updated 3-km-mesh High-Resolution Rapid Refresh (HRRR) model, nested within the RAP since 2011. This grid spacing (3 km) allows the depiction of many mesoscale phenomena, including mesoscale convective systems, cold fronts, orographic flows, and land/sea breeze circulations. The HRRR domain covers the entire contiguous United States (Figure 2.8), in contrast to the smaller northeast U.S. domain used for the numerical profiler data impact experiments run for this study (Figure 2.9, left). The data assimilation component of the system is GSI, and the base model is the WRF–ARW.

In January 2012, ESRL began to archive the two-dimensional (grib2) output from the HRRR, resulting in a two-year archive of HRRR forecasts. The model is run out to 15 forecast hours. In this section we use this archive to generate 2-yr average plots of the 80-m wind-speed field as an example of the kind of product that can be obtained from operational model output, in which hourly assimilation of available data sources is routine. Properly validated, such data sources may potentially be useful to wind-energy industry interests for resource assessment.

In this offshore wind resource analysis, we selected two-hour forecasts from the HRRR. The 0-h HRRR analysis fields are strongly dependent on the 13-km RAP and therefore, lacking in full 3-km physical consistency (adjusting to 3-km coastlines and terrain, etc.). To address this, we reasoned that 2-h forecasts would be a best compromise in that they are close enough to the initial time that the model error is small and still strongly constrained by observations only 2h old, but also far enough removed from the analysis that the predicted variables have had time to adjust on the 3-km grid. Using this dataset, we can calculate many statistics and plot maps of various measures of the low-level wind resource (80-m wind speed) in the offshore ar-

ea. Although winds at levels higher than 80 m may be of interest, at this time our HRRR archive contains only 10- and 80-m winds.[2]

The plot of the 2-yr average 80-m wind offshore of the East Coast (Figure 3.1. left) reveals some interesting patterns. The average wind speed generally increases with distance from the coast, but areas where higher average winds approach the coast are also apparent. These areas tend to be farther north, and particularly along sections of coast that are oriented more west-to-east than south-to-north.

[2]Before April of 2013, no three-dimensional variational data assimilation was done on the 3-km HRRR grid, but the initial conditions were simply inherited from the parent 13-km RAP. Since April 2013, the HRRR now includes data assimilation on its 3-km grid, including a 15-minute radar reflectivity assimilation procedure. Prior to the incorporation of data assimilation on the 3-km grid, the 13-km RAP was simply downscaled by interpolation for the HRRR analysis (0-h forecast). Since the character of 3-d wind fields from HRRR forecasts of 2-h duration are dominated by the model physics, resolution, and land-surface fields, none of which were changed substantially during this 2-year period, we consider the 2-h HRRR forecasts as a relatively consistent dataset for this resource assessment over the 2012-2013 period.

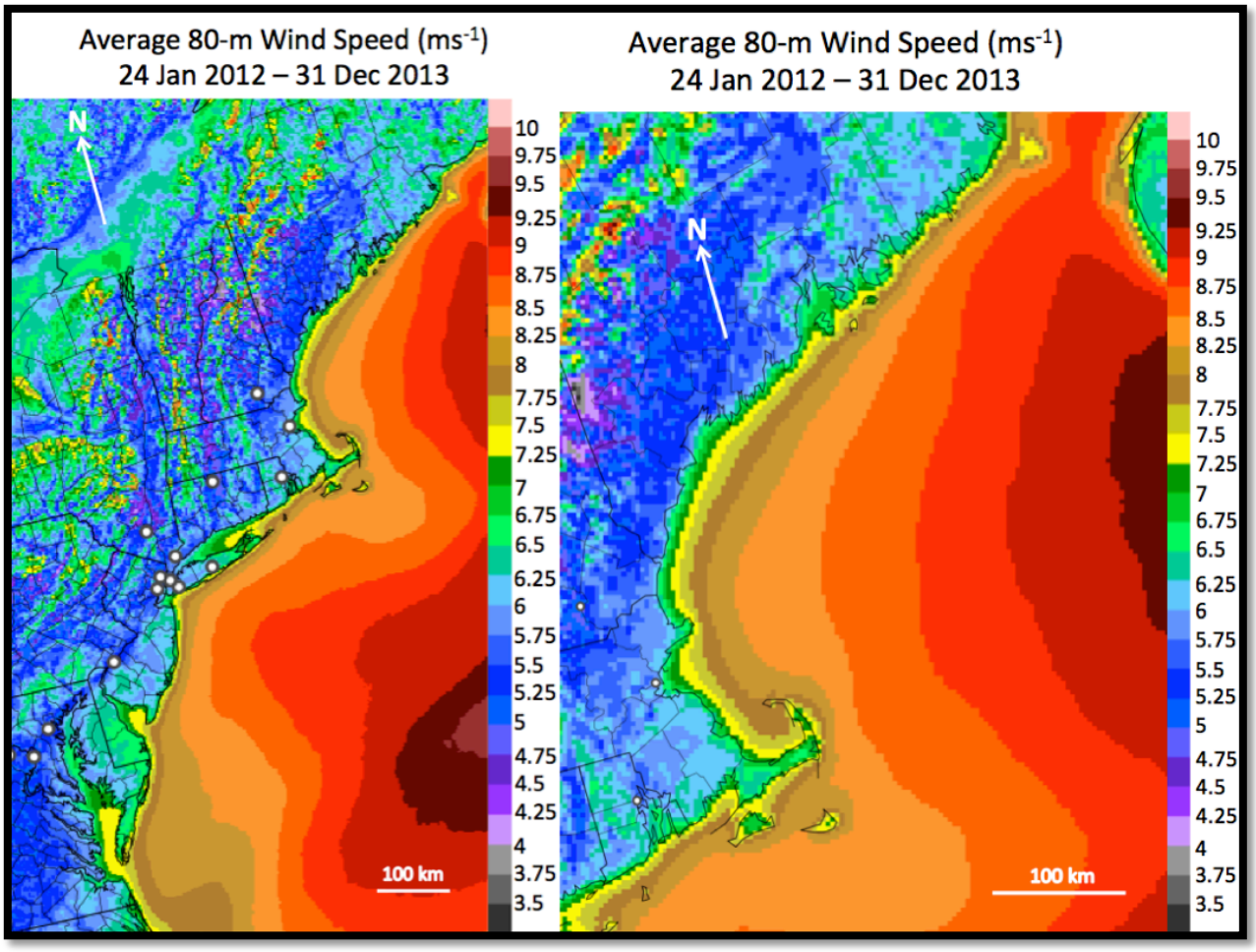


Figure 3.1. Average 80-m wind speed (m s^{-1}) during the period 24 Jan 2012 – 31 Dec 2013 according to 2-h HRRR forecasts for (left) the northeastern U.S. and (right) the Gulf of Maine.

Average winds greater than 8 m s^{-1} occur very close to the coast offshore of New Jersey, the Delmarva Peninsula, Long Island, and southeast of Rhode Island and Cape Cod. Figure 3.1 shows an enlarged view of the New England coast, including the Gulf of Maine. Here annual wind speeds exceeding 8 m s^{-1} extend into the Gulf of Maine, but slower wind speeds are evident in the vicinity of Cape Cod.

From this dataset, we can also plot the frequency of winds exceeding certain thresholds. For example, the 6 m s^{-1} threshold is considered to be near the lower end of what is required to turn a typical turbine blade. For the entire period, Figure 3.2. (left) shows areas within a few miles of the coast where the 80-m wind speed exceeds 6 m s^{-1} 60% of the time. Offshore areas near the southwestern tips of Martha's Vineyard and Nantucket Island, and the eastern tip of Long Island, have winds greater than 6 m s^{-1} more than 65% of the time.

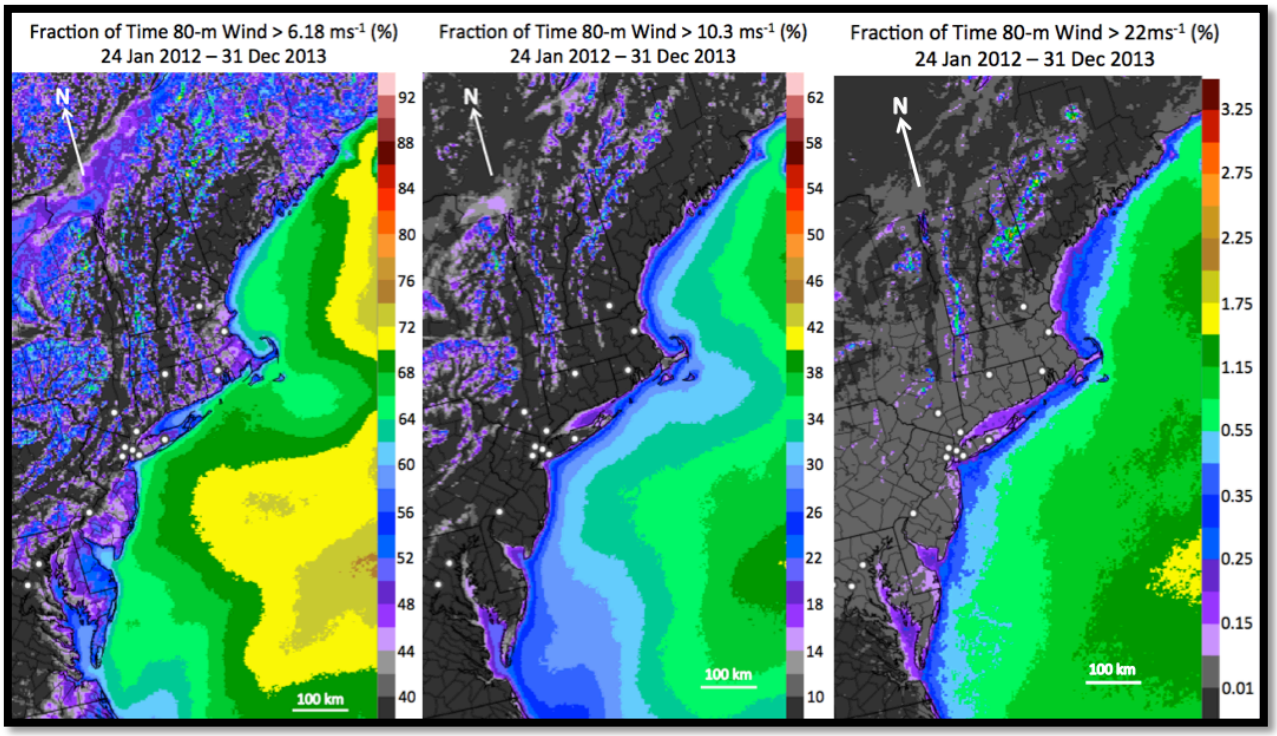


Figure 3.2. Percentage of time that 80-m wind exceeds (a) 6.18 m s^{-1} (12 kts), (b) 10.3 m s^{-1} (20 kts), and (c) 22 m s^{-1} (42.76 kts) during the period 24 Jan 2012 – 31 Dec 2013 according to 2-h HRRR forecasts.

Perhaps a more interesting threshold for wind energy generation is 10 m s^{-1}; we can also plot the frequency of exceedances of this larger threshold (Figure 3.2.,center). The aforementioned windy areas of eastern Long Island, Martha's Vineyard, and Nantucket Island experience winds greater than 10 m s^{-1} between 25% and 30% of the time. In the long-term averages, the coast of the Gulf of Maine appears fairly spatially homogeneous, with average wind speeds (and frequency of threshold exceedances) increasing with distance from the coast.

A final metric of interest for the entire two-year period is the frequency of exceedances of the 22 m s^{-1} threshold. This value is near the so-called "cutout speed" where turbines cannot generate any additional electricity and may actually begin to experience structural damage. For the entire period, winds of this magnitude only occur near coastal areas less than 1% of the time (Figure 3.2., right). Nantucket Island and Cape Cod have wind speeds greater than 22 m s^{-1} about 0.85% of the time.

We can also stratify the HRRR horizontal wind-speed composite by season and by time of day. As examples of an investigation of the diurnal cycle, in Appendix B we present some daytime (15-00 UTC) and nighttime (03-12 UTC) plots from the summers (June – August) of 2012 and 2013. Overall, summer of 2013 was windier than the summer of 2012. The winds exceeded

the 6- and 10-m s^{-1} thresholds more frequently during the summer of 2013 than 2012. This distinction is apparent in both the daytime and the nighttime composites.

Another way to visualize the spatial and time variability of the 80-m wind field in the HRRR is to take averages over interesting regions and then look at time series. Figure 3.4 shows a map of the sites picked for this study; two of them are DOE offshore wind energy project locations, and the other three are in different parts of the near-shore environment. We can use our dataset to plot time series of the fraction of the time above the 6-m s^{-1}, 10-m s^{-1}, and 22-m s^{-1} thresholds during 2012 and 2013 (Figs. 3.5-3.6).

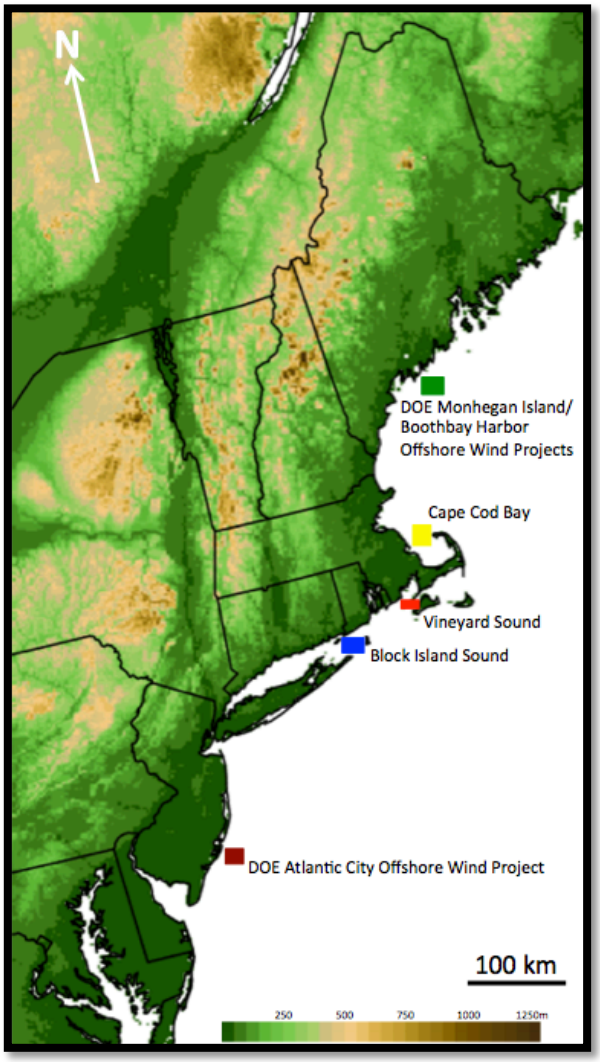


Figure 3.4. Map of potential offshore wind energy sites in the northeastern USA overlaid on inland terrain height (m).

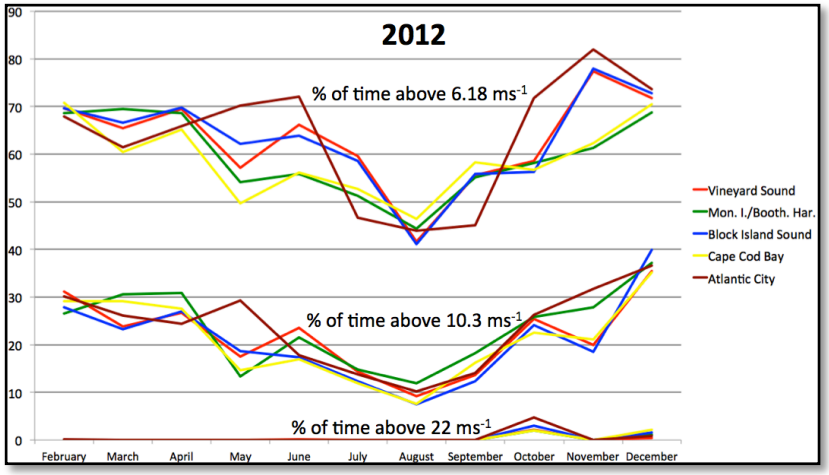


Figure 3.5. 2012 Monthly time series of percentage of time that 80-m wind exceeds three thresholds of 6.18, 10.3, and 22 m s^{-1} at each of the five potential offshore wind energy sites according to 2-h HRRR forecasts: Vineyard Sound, Monhegan Island/Boothbay Harbour, Block Island Sound, Cape Cod Bay, and Atlantic City.

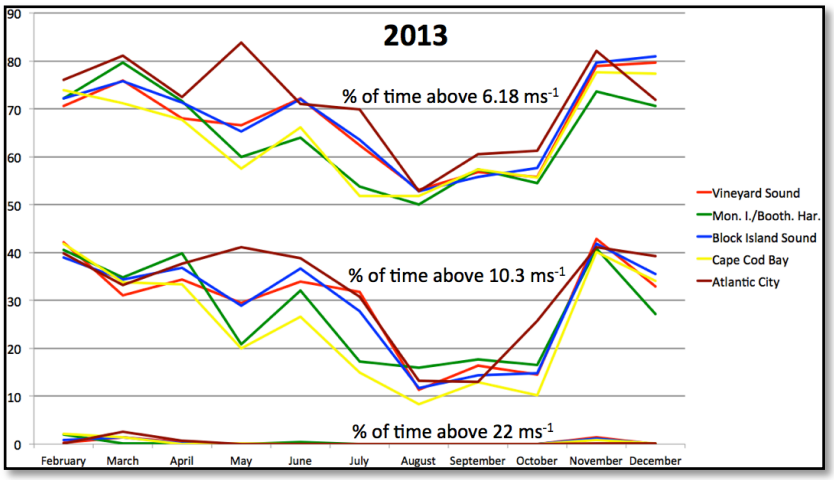


Figure 3.6. 2013 Monthly time series of percentage of time that 80-m wind exceeds three thresholds of 6.18, 10.3, and 22 m s^{-1} at each of the five potential offshore wind energy sites according to 2-h HRRR forecasts: Vineyard Sound, Monhegan Island/Boothbay Harbour, Block Island Sound, Cape Cod Bay, and Atlantic City.

In these plots a significant annual cycle is evident in the fractions of time above the 6- and 10-m s^{-1} thresholds. The 80-m winds are most frequently high in the winter and spring at all

sites, and lowest in the summer (with a minimum generally in August). Wind speeds were higher during 2013 than during 2012 in most months, except in the autumn. It is apparent that the higher threshold of 22 m s^{-1} hardly is ever exceeded in the HRRR model. Some of the regions display very similar behavior, such as the Cape Cod Bay region and the Monhegan Island/Boothbay Harbor regions of coastal Maine. Monhegan Island/Boothbay Harbor seems to have slightly more frequent 10-m s^{-1} winds year round, and both regions have a comparable frequency of 6 m s^{-1} winds.

Two other regions that display very similar behavior are the Vineyard Sound and Block Island Sound regions. These areas are windier than the Cape Cod Bay and Monhegan Island/Boothbay Harbor regions, despite the fact that they are farther south. These regions show a more delayed decrease in 6-m s^{-1} frequencies in the spring and summer months when compared with the more northern areas.

The Atlantic City region, off the coast of southern New Jersey, displays quite different behavior from the other four areas. During the summer and early autumn, the frequency of time above 10 m s^{-1} appears just as low or lower than the other four regions, but during the rest of the year the Atlantic City region has some of the highest frequency of high winds of all of the regions.

Although 2013 was significantly windier than 2012 for all thresholds, the 22 ms^{-1} threshold shows some interesting features. The peak in October of 2012 at all the regions is associated with a single high-impact event in the area: Hurricane Sandy. Percentages up to 3 or 4% later in the winter of 2012-2013 are associated with an anomalously large amount of cyclone activity off the east coast.

RAP MODEL WINDS – Study Period

To compare with the 2-year mean HRRR wind fields for offshore winds, we repeated a similar analysis for the two time periods chosen for this experiment (one week in August 2004 and 8 days in July 2004). The averaged wind speed at the possible turbine heights, between 50 and 150 meters over the sea and above ground level of the RAP model for these time periods shows the Gulf of Maine being a good source for offshore wind during these periods, as the mean wind speeds were around 8-9 m s^{-1} (Figure 3.7).

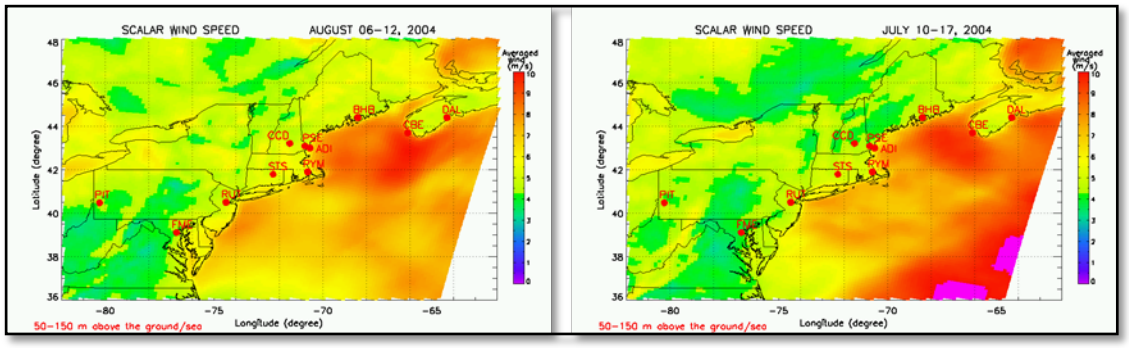


Figure 3.7. Scalar wind averaged at 50-150 meters above the ground/sea level for the RAP model over the two chosen time period: 6-12 August, on the left; and 10-17 July, on the right.

This high scalar wind speed over the ocean is evident for both time periods. For the August week the overall wind was higher, especially in the southeast area of the Gulf of Maine, where 95% of the time the speed was greater than 6 m s^{-1} (12 kt). We also notice that inside the Gulf of Maine mean nighttime wind speeds exceeded the daytime values (Figure 3.8).

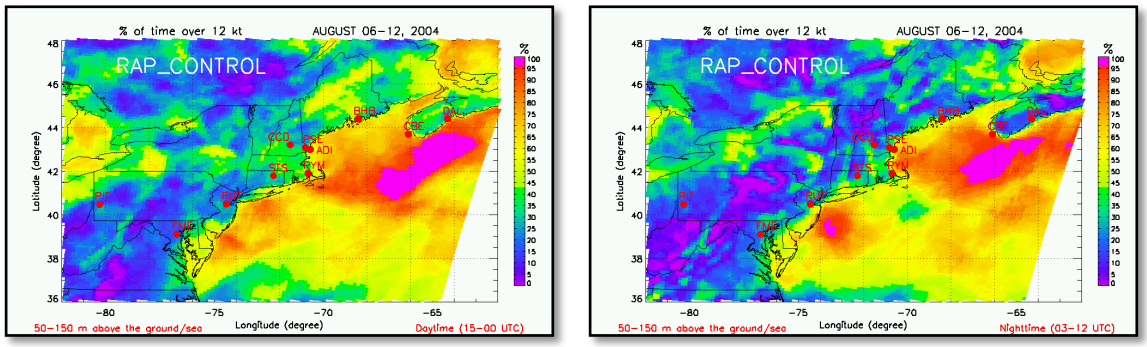


Figure 3.8. August 2004 week. Percent of time when the scalar wind speed averaged at 50-150 meters above the ground/sea level for the RAP model exceeds 12 knots (6.2 m s^{-1}). Daytime (15-00 UTC) is on the left and the nighttime (03-12 UTC) is on the right.

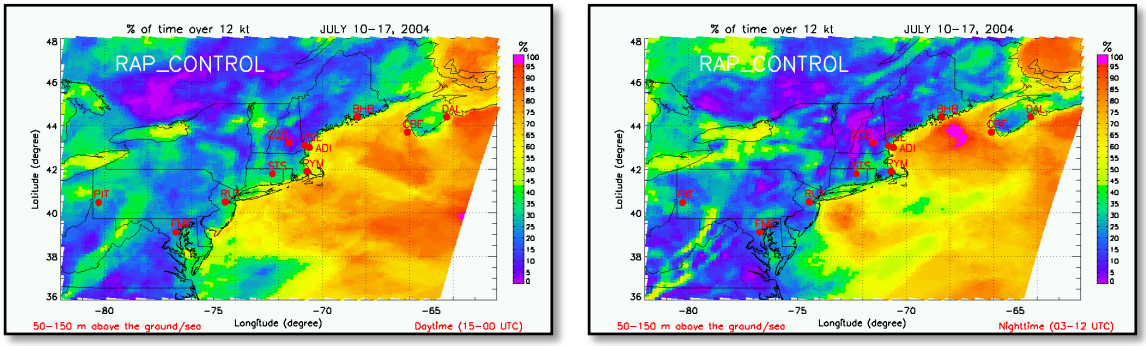


Figure 3.9. Same as Figure3.8 but for the July period.

b. High-Resolution Doppler Lidar

To address the objectives of the POWER study, ESRL's High-Resolution Doppler Lidar (HRDL), provides profiles of wind speed and direction from the deck of the *RHB* every 15 min. These profiles extend from near the ocean surface to heights of ~1500 m or more at vertical resolutions of < 10 m. The VAD analysis procedure further averaged these high-resolution data into vertical bins having gradually increasing vertical intervals, as shown later in Figure 3.15. The ship track for the entire 6-week NEAQS cruise in July and August 2004 was presented earlier in Figure2.1. (bottom panel). The tracks for the two intensive measurement-modeling study periods of the POWER analysis are shown in Figure3.10. In the following analysis, unless otherwise stated, all references to NAM and CONUSNEST are to the components of hourly-updated NAMRR system presented in Section 2c.

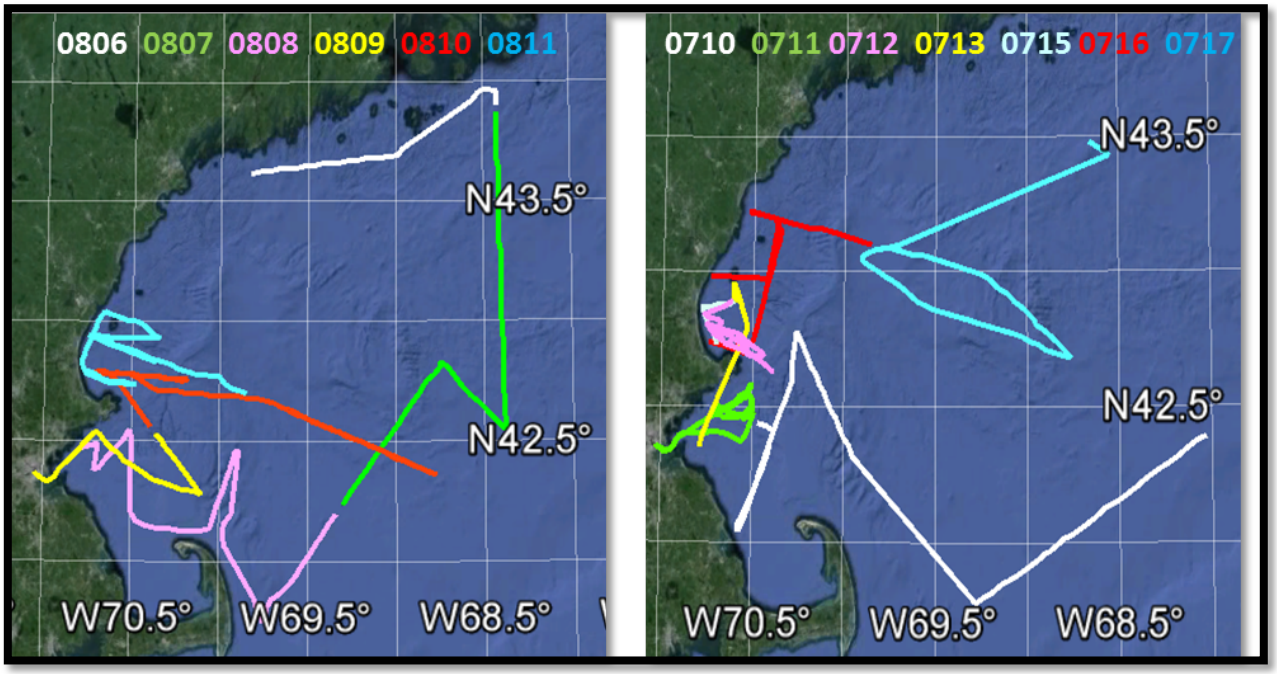


Figure 3.10. Ship tracks during two study periods: (left) August 6-12, (right) July 10-17. Ship for each day is shown by the color according the legend in the bottom of each plot. The red square represents area of 241 x 250 km with coordinates of the corners: NW (71N, 44W), SW (68N, 44W), NE (71N, 41.5W), and SE (68N, 41.5W). The scale on the left plot is 132 km (82 mi).

During both periods the ship was cruising in the Gulf of Maine, covering a wide 241 x 250 km (150 x 155 mi) area as indicated in the red squares in Figure3.10. Examples of 15-min average wind profiles are given in Figure 3.11 to illustrate the high vertical resolution of the data that allowed fast changing dynamic processes such as LLJs or sudden changes in wind speed or direction to be sampled.

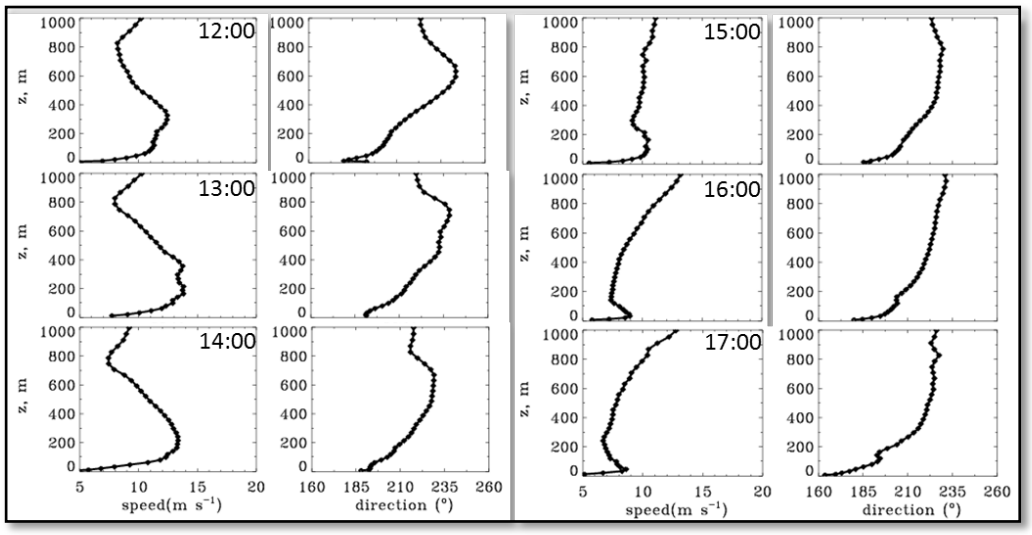


Figure 3.11. Samples of lidar measured wind speed and wind direction profiles for 6 hours in August 11. Time is shown in UTC. Symbols indicate height of measurements.

Both periods selected represent various wind-flow conditions as illustrated in the time-height plots of wind speed in Figure 3.12, which displays the lowest 300 m of the profile data for eight days of the POWER analysis periods. Wind-speed values are color coded and plotted as a function of time (UTC). The first half of each 24-h panel represents nighttime, and frequent periods when the peak wind speeds occur at heights within the cross section, indicative of low-level jet (LLJ) structure, are evident (see also Figure 3.11.).

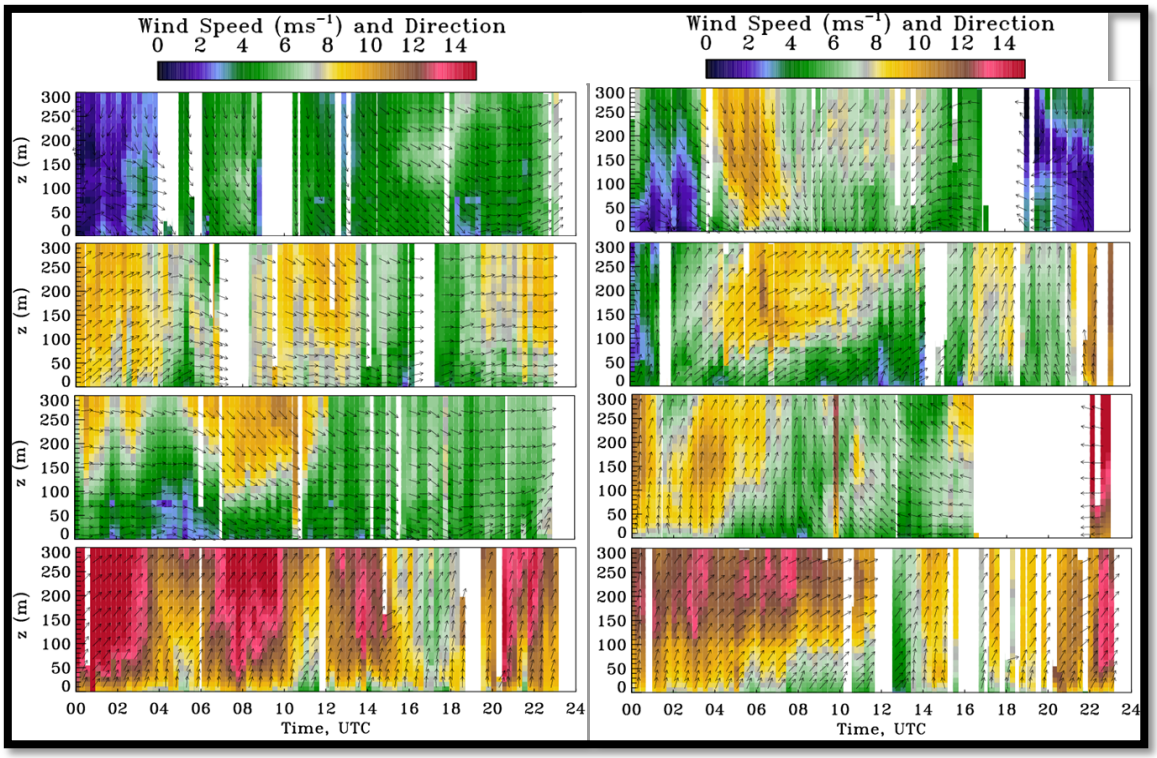


Figure 3.12. Time-height cross sections of 15-min averaged lidar-measured wind speed (color bar, scaled from 0 to 10 m s-1) and direction (arrows), computed from HRDL conical scans during (left, from top to bottom) August 7, 8, 9 11 and (right, from top to bottom), July 11,12,13,16. Vertical axis is height above sea level (m), and horizontal axis is time in UTC (local Eastern Standard Time lags UTC by 5 h).

Many episodes of high wind speed shear across the presumed turbine-rotor layer were observed during both periods, which could be identified on the cross-sections. Profiles of wind speed and direction on these plots were computed from the conical (VAD) scans only, so occasional gaps appear in the cross sections, when other scans (e.g., elevation or staring) were being performed over the entire 15-min period. Other blank (white) areas due to periods when data are unavailable as a result of fog, precipitation, or other outages can also be seen, such as on July 13 between 1700 and 2100 UTC.

Overall, synoptic meteorological conditions at the heights of the rotor-swept area (40-160m) of a presumptive turbine for the two selected periods were dominated by westerly through northwesterly wind directions and wide ranges (2-17 m s^{-1}) of wind speeds, with mean values of 7-8 m s^{-1} (See FigureC-4, Appendix C).

It is clear from the geographic time-height cross section of the winds presented in Figure 3.12 that HRDL data detect considerable variability in the winds along the ship track and with height. The ship tracks, color coded for wind speeds near turbine hub height (80 m) and at 10 m ASL for the entire cruise, are plotted in Figure 3.13.

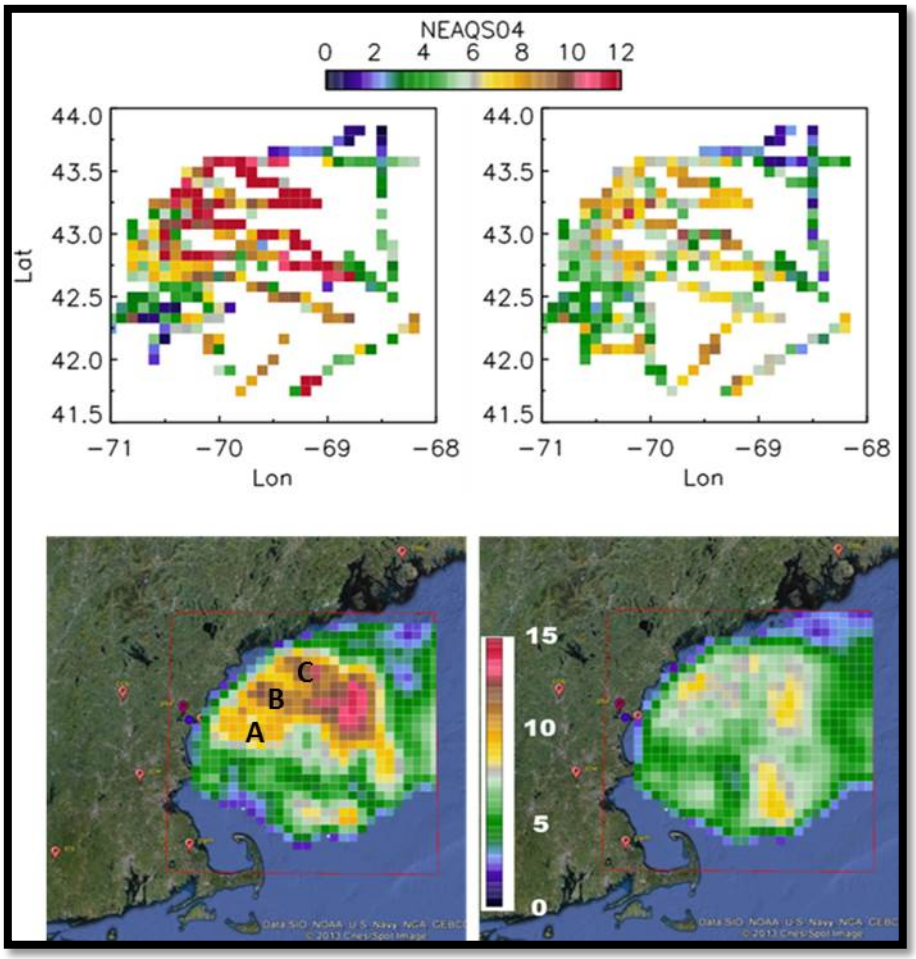


Figure 3.13. Top panels: Lidar-measured wind speed at (left) 80m and m and (right) 10 m along ship tracks during July 9- August 12. Bottom panels: Gridded wind speed at (left) 80 m and (right) 10 m superimposed on Google Earth map of the region.

For better visualization, these data can be interpolated to a Cartesian grid (Figure3.13. bottom panels) to show the horizontal distribution of wind speeds (left) at 80 m above the ocean surface, and (right) at 10 m. The distribution near the sea surface differs from that at hub height, indicating that extrapolating the distribution of near-surface winds to hub height using standardized profile shapes is not likely to produce a realistic horizontal distribution of wind speeds at hub height, consistent with the finding of Pichugina et al. 2012, that measured profile shapes are highly variable and do not conform to any standardized profile shapes, such as exponential or logarithmic. Such near-surface distributions are often obtained from buoy, ship, or satellite (e.g., synthetic aperture radar) data. A more focused analysis concentrating on the area of highest measurement density is given in Appendix C, Figure C-5. In addition to horizontal plots at 10 m and 80 m ASL, a horizontal plot of the difference between wind speed at the two levels is shown, indicating vertical differences of 5 m s^{-1} in some areas and 0 m s^{-1} in others.

The 80-m analysis (Figure3.13, left, bottom) shows a tendency for winds to be stronger with greater distance from shore out to about 100 km offshore for this period, and in the alongshore direction three maxima (A,B,C) are spaced at ~50 km in the west-central part of the dataset. Towards the southern extent of the plot, the winds were lighter in the vicinity of Massachusetts Bay, as also found in the HRRR 2-yr archive composite analyses in the previous section. Another region of weaker winds is noted to the northeast of the region of higher winds, which could be due to sparse sampling of this region during the cruise. If this were in fact a region where the wind speeds tended to be slower, this would imply an along-shore scale of variation between the two 'blue' regions in the plot of nearly 250 km. However, note that bottom plots in Figure 3.13 are not intended to be construed as representing any longer-term average horizontal distribution of wind speeds (for example, these maps should not be used for actual siting decisions). They represent a relatively short time period, during which the study area was not uniformly sampled. These plots are meant to illustrate the kind of analysis that could result from an extended field campaign involving mobile sampling. They demonstrate the advantages of such an analysis, if the area were adequately sampled (as could be built into the design of a field campaign customized to address wind-energy issues). Mean wind speeds, spatial gradients, and regions of strong and weak resource are all evident in this analysis. The figure also highlights potential pitfalls of such an analysis, especially if the sampling period is too short or the area is not uniformly sampled. For example, the area of weak wind speeds indicated in the northeast sector is a result of a few cruise legs which probably occurred on days when winds were light over the entire region (not just in this sector). This kind of uneven spatial sampling could be avoided in research cruises tailored to address WE issues.

The variability of wind speeds along the ship tracks can result from spatial variability of the flows, time-dependent changes in the flows, or a combination of both. Previously Pichugina et al. (2012) investigated this issue using a sampling period when the *RHB* sailed back and forth over the same ~40 km leg, sampling the leg 3 times, as shown in Figure 3.14a. Wind speed and direction at 80 m, and depth of the ocean shown as a function of distance from the shore in Figure 3.14.b, where 15-min averaged data are shown in a different color for each trace (leg). The resulting vertical cross sections (Figure 3.13.c), where the wind profiles are plotted as a function of longitude, show a persistent LLJ out to sea in the eastern portion of the cross section. Horizontal gradients in speed, especially in the turbine rotor layer (dotted lines in Figure3.13.c), created by this jet occupy about the eastern half of the cross section, over a distance of 20 km.

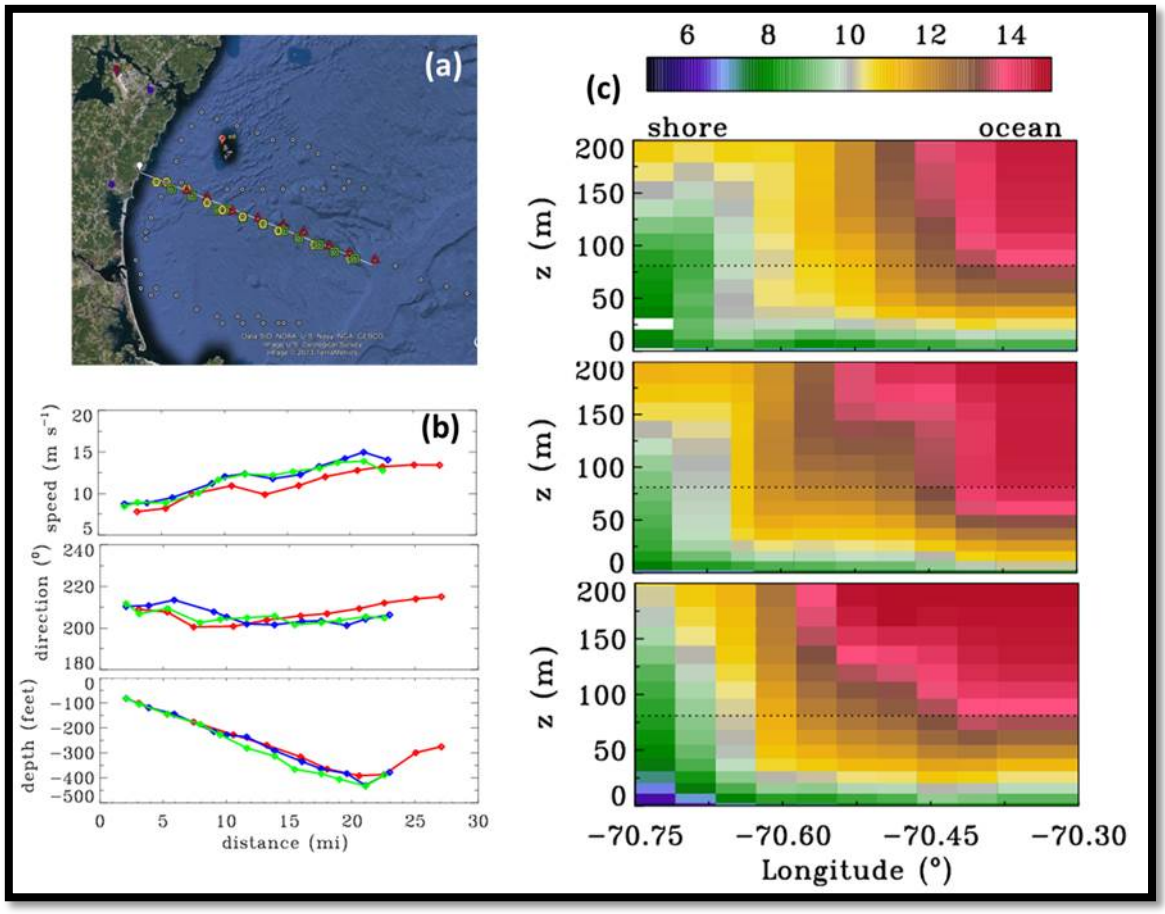


Figure 3.13. (a). Google-Earth map of ship tracks on 11 August, when the RHB retraced the same course three times (color symbols). (b) wind speed and direction at 80 m, and depth of the ocean shown as a function of distance from the shore. 15-min averaged data are shown in a different color for each trace (leg). (c). Vertical cross sections of mean *wind speeds for each leg as a function of longitude.*

Based on the flow structure and gradients found in this nearly 6-week dataset, we could define the sampling density of fixed wind-profiling sensors, such as buoy-based Doppler lidars, that would have been required over the ocean to adequately sample them. The LLJ sampling by the three repeated ship tracks found that in the cross-shore direction, strong horizontal gradients in wind speed occurred over a distance of 20 km. Sensors would need to be spaced at 10-km intervals perpendicular to the shore in the vicinity of this wind-speed gradient to be able to sample the location and strength of the strongest horizontal gradient. In the along-shore direction, peak-to-peak oscillations of 50 km and 250 km were noted. It is unclear, of course, whether these variations would show up in long-term averages, but if they were representative of relatively stationary patterns, a maximum spacing of ~20 km in the along-shore direction (parallel to the coast) would be required to sample the smaller-wavelength features and ~50 km for the longer wavelength features observed.

For temporal sampling, Pichugina et al. (2012) previously showed that the 15-min sampling revealed many sub-hourly periods having significant change in wind speed and direction, so to the extent that these are important to include in studies or if one might be interested in forming climatologies of such 'ramp' events, 15-min profile averaging (or less) is required. HRDL-profile data also reveal significant structure in the lowest 100 m with changes of several m s^{-1}, indicating that vertical resolution of 10 m is required to capture these features.

HRDL-Model Comparisons

The availability of high-quality HRDL wind-profile data through heights occupied by turbine rotors at offshore locations provides the opportunity to verify NWP model data in this oceanic region where few reliable data exist. For comparison with observations, the modeled wind predictions were extracted along the ship track at the exact location of lidar measurements. The 15-min HRDL profile data were averaged over 1-hour to fit the time-resolution of the models and interpolated to the model mean vertical grid. Figure 3.15 shows vertical data intervals of lidar measurements and model output.

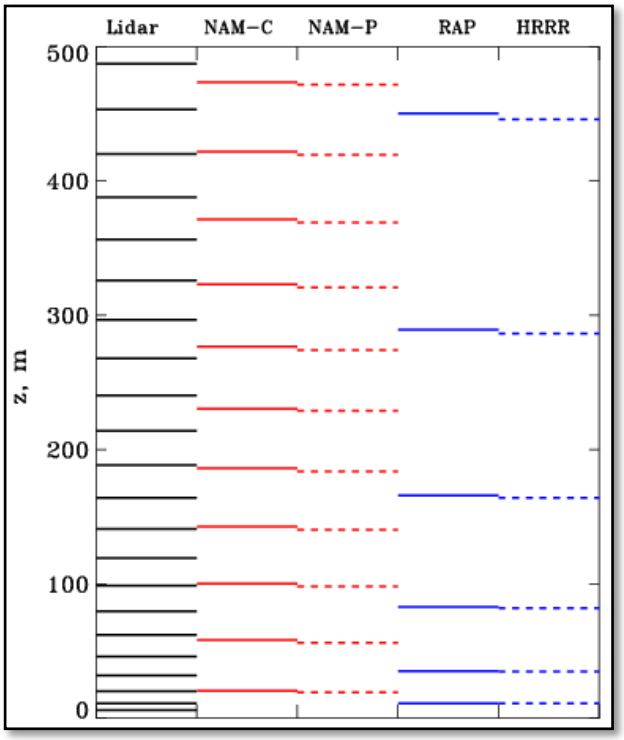


Figure 3.15. Heights of HRDL mean wind profile vertical data intervals and vertical levels of model output.

The quality and resolution of the HRDL data allow verification not only of the simulated mean wind at critical heights, but also of mean wind shear and turbulence (Pichugina et al.

2008), LLJ properties, characteristics of ramp events, and many other wind-energy-relevant aspects of the flow. Here we will focus on the initialization and prediction of the mean wind profile and the winds near hub height. An example of several direct comparisons of HRDL profile data with model profiles for the same time is given in Appendix C (Figure C-6). Figure3.16 shows an example of time-height cross sections of the HRDL-measured wind-speed data (bottom panel) plotted against model-initialization data ('forecast-0') for each of the four model configurations (top four panels) for two days of the project. This visual overview shows that the overall features and trends are represented in the models. Quantitative differences in flow strength and timing can be seen among the models, and between models and the HRDL values.

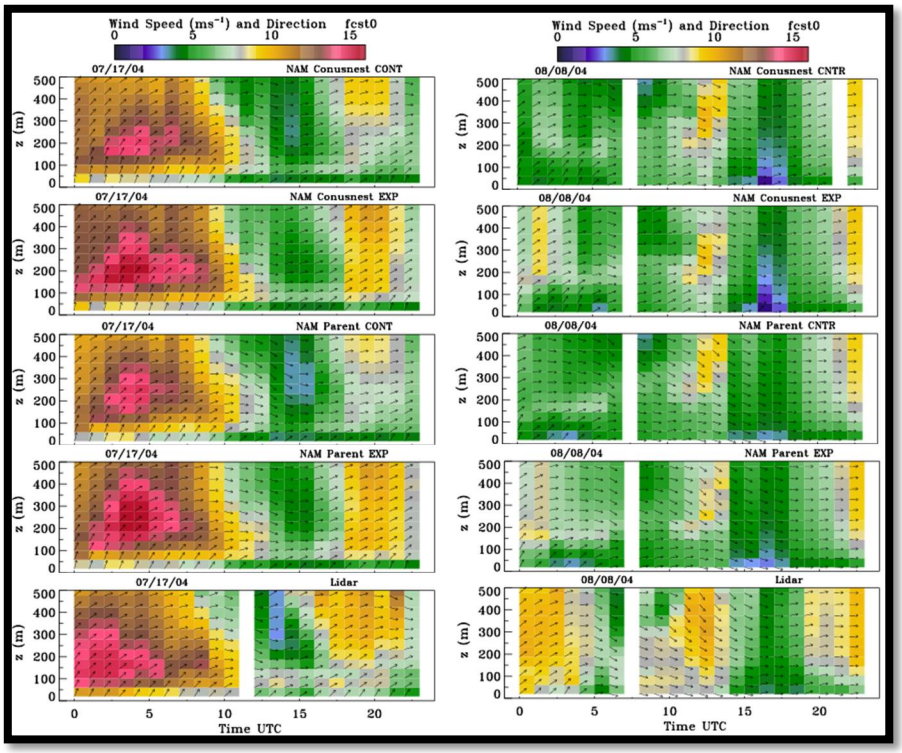


Figure 3.16. Time-height cross sections of 1-h averaged lidar-measured (bottom) and modeled (four top panels) wind speed and direction for (left) 17 July and (right) 8 August 2004. Two bottom panels are lidar measurements. The models are (from the top): NAM-CONUS control; NAM CONUS experimental; NAM PARENT control; NAM PARENT experimental. Color scale in m s^{-1}, vertical axis is height above sea level (m), and horizontal axis is time in UTC.

A quantitative statistical assessment of the mean-wind accuracy of the initial state of the models as a function of height is shown in Figure 3.17 and Figure 3.18 for the two intensive study periods. During the August period the RMSE for the finer-mesh models (NAMRR and its

CONUS-NEST) was just over 1.7 m s^{-1} above 100 m, but below 100 m this increased to 2.5 m s^{-1} near the surface. Above 100 m wind-speed slow biases were less than 1 m s^{-1} in general and correlation coefficients of 0.9 are seen, but again the statistics degrade below 100 m, as biases increase and the correlations become much smaller. These statistical profiles show that the assimilation procedures produce larger discrepancies near the ocean surface than above 100 m. We note that this is the opposite as was found in WFIP (NOAA WFIP final report, Wilczak et al., 2014) and may be due to the paucity of near-surface observations over the ocean. Similar comparisons of model ability to simulate diurnal cycles, strong and weak winds, LLJ-like profiles, and other time- and space-dependent features of the flow, can allow us to estimate areas of the most needed model improvement.

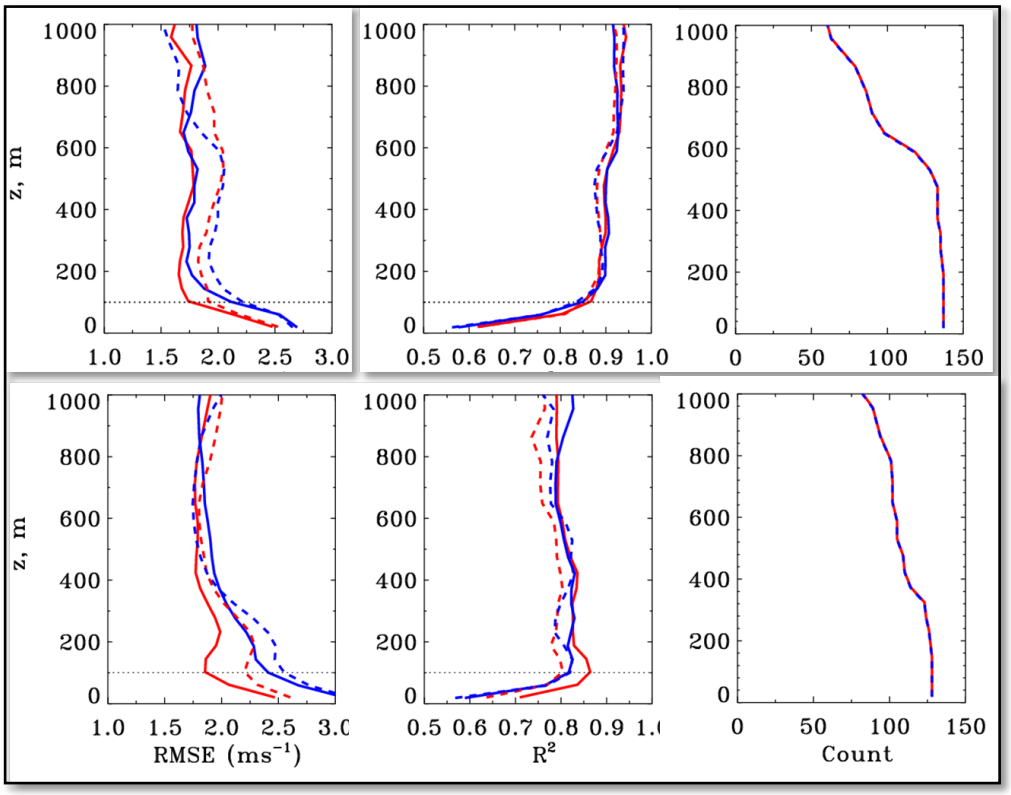


Figure 3.17. Profiles of RMSE and correlation coefficients between observed and initial-condition model wind speed averaged over (top) August and (bottom) July study periods. Red lines on all panels are for NAMRR-CONUS (solid) experimental and (dashed) control runs. Blue lines are for NAMRR parent (solid) experimental and (dashed) control runs. (Right panels) number of data points at each level.

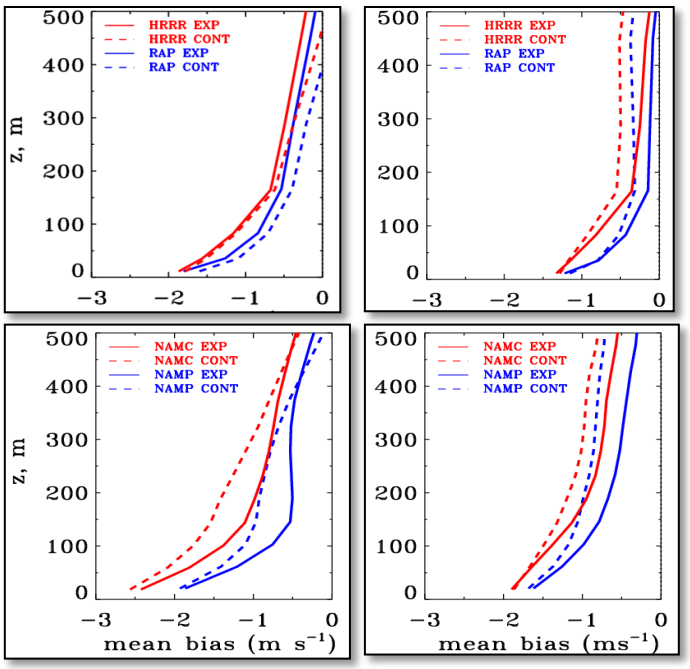


Figure 3.18. Profiles of mean bias (m s^{-1}, horizontal axis) as a function of height (m, vertical axis) between modeled and lidar-measured scalar winds for initial conditions for the (left) July and (right) August periods. The lines represent models according to the legend in the top right corner of each panel.

HRDL wind data were used to verify model forecasts out to 12 h lead time. To test overall model performance, HRDL data were averaged over the layer from 20 to 500 m ASL. Figure 3.19 shows the RMSE for the eight model configuration runs, including those with (experimental) and without (control) assimilation of profiler wind data, for the August period.

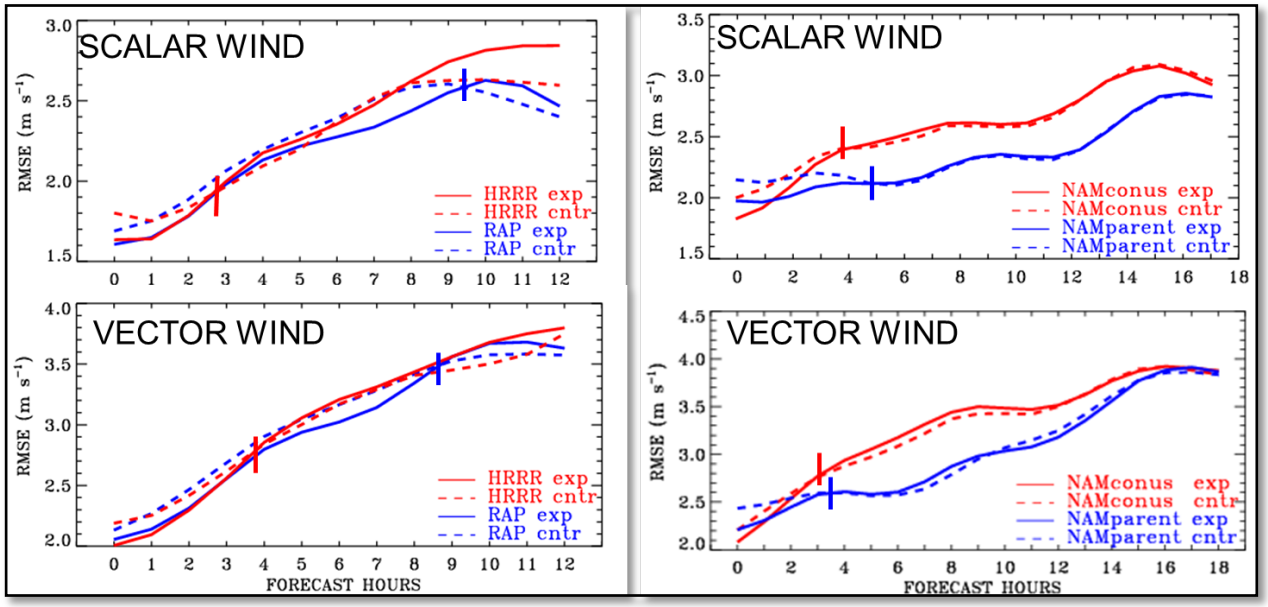


Figure 3.19. RMSE between observed and modeled wind in August 6-12 for data averaged over 20-500m layer. (Left) RAP and HRRR models, and (right) NAM CONUS and NAM PARENT models.

The plot shows the scalar-wind (i.e., mean wind-speed) deviation and the vector wind-speed error, which includes directional deviations for the August study period. In general, profiler assimilation produces an improved initialization by ~ 0.2 m s^{-1} and an improved forecast (i.e., the solid line lies beneath the dashed line) out to several hours, the crossover hour indicated by the vertical bar. The trace for the higher resolution models in red (HRRR and NAMconus) lies at or below that of the lower-resolution or 'parent' models for a short time after initialization, after which the coarser-resolution model seems to have the better error statistics.

As noted in later sections of this report, the July study period (Figure 3.20) exhibits some unexpected behavior in the model verification. For a brief period after initialization, the experimental assimilation (solid line) runs perform better than the control runs as anticipated, but after that period, the control runs have lower error much of the time. These apparently anomalous results are further discussed in later sections.

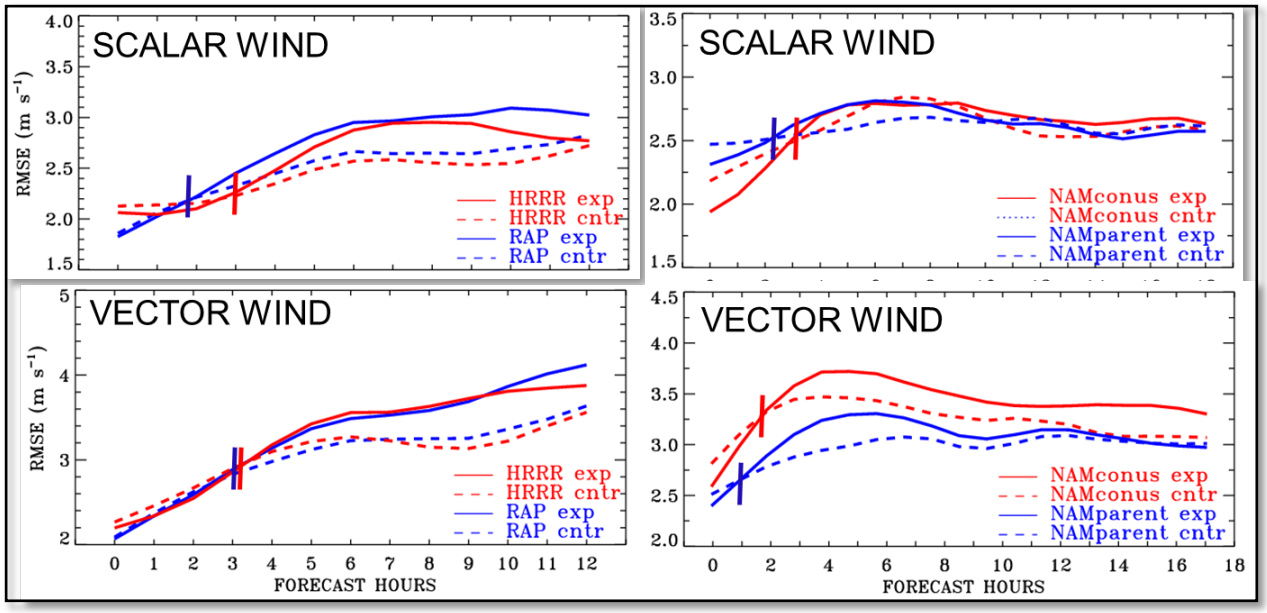


Figure 3.20. RMSE between observed and modeled wind in July study period for data averaged over 20-500m layer.

The behavior of each forecast as a function of time of day is shown in Figure 3.21 for the August time period. Plotted is the mean HRDL-measured wind at 100 m ASL for each hour of the day (black line, which remains the same from panel to panel), and the mean-wind forecast at ~100 m ASL valid at that hour of day, for each of the four forecast configurations of the NAMRR. In the top panel, the initial wind (forecast 0) is well represented in the initial model fields during the daytime (12-22 UTC), but underestimated at night, especially after midnight

(05 UTC). The 1-h forecasts (Forecast 1: second panel from top) valid after midnight still reflect the underestimates in the initial conditions, as do the 2- and 3-h forecasts. During daylight hours the 1-3-h forecasts are reasonable predictions of the measured winds, also in line with the skill of the initial conditions. The 12-h forecasts valid in the afternoon overpredict the measured winds, extending into the evening hours.

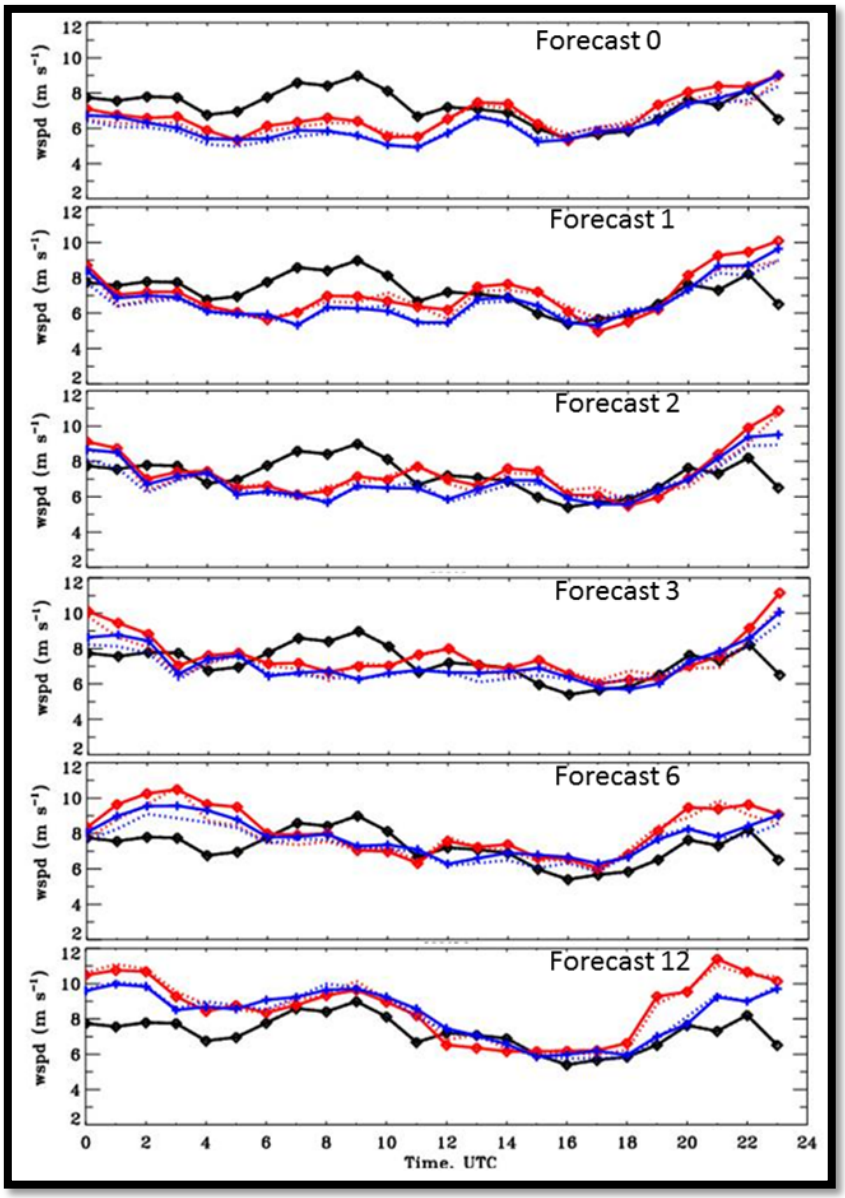


Figure 3.21. Time-series of lidar-measured and NAMCONUS and NAMRR modeled wind. Black trace, which is the same in all panels, is mean HRDL wind-speed data at 100 m ASL plotted against hour of the day (UTC, which leads Eastern Standard Time by 5 hr), averaged for each hour over the August time period. Top panel shows the model initial wind field (forecast-hour 0, or Forecast0) also plotted vs.

hour of the day, where blue lines are NAMparent runs and red are NAM-CONUS; the dotted lines are for the control runs, and solid, for experimental. Second panel shows the 1-hr forecast wind speed averaged for each hour and plotted at the valid time of the forecast, for each model run during the August period. Third panel shows the mean of the 2-hr model forecasts plotted at the valid time, fourth panel shows the 3-hr forecast, fifth panel shows the 6-hr forecast, and the bottom panel shows the 12-hr forecast.

Appendix C shows other lidar results, including time series at several levels within the rotor layer, wind-speed distributions for the two study periods, and the horizontal-distribution plots of wind speeds and vertical wind-speed differences over a smaller area of Figure 3.13, where the data coverage was relatively uniform. Also included are further comparisons of HRDL vs. model time series, and profiles of wind speed and mean model bias for forecast lead times from 0 to 11 hr. Scatter plots of lidar-measured vs. modeled wind speed show that the largest biases were at the highest wind speeds.

c. Wind-profiling Radars

For the profiler data comparison, the data from the models were extracted at the profiler site locations and interpolated to the vertical heights of the observed data. As described previously *control* runs are without profiler assimilation and the *experimental* runs do assimilate the land-based profiler data. An example of a time-series of the observed winds at the first level of the wind profiling radar and the corresponding model data is shown in Figure 3.22 for the Appledore Island (ADI) wind profiler site located 10 km off the coast. The data are extracted for the model runs at forecast hour 00.

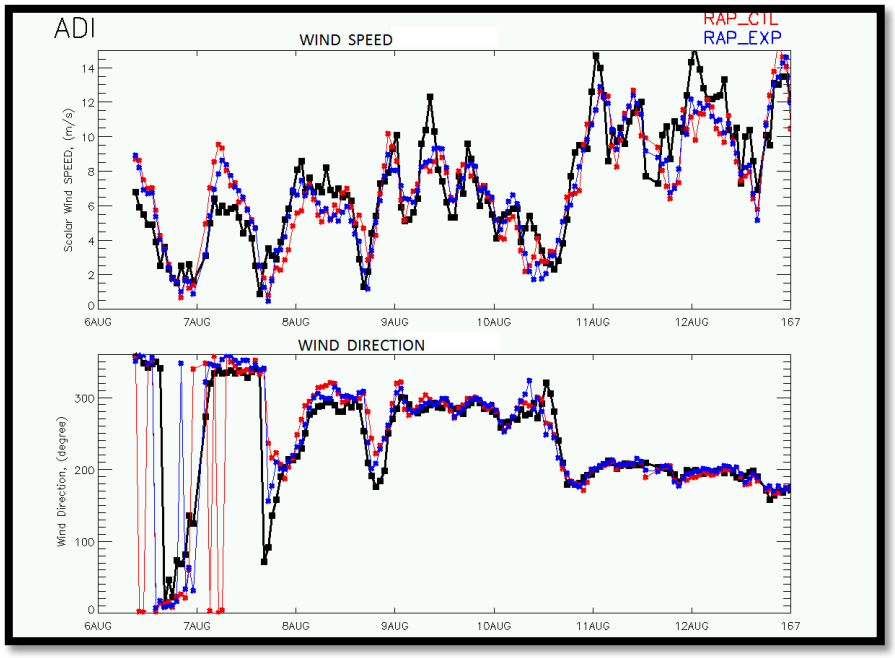


Figure 3.22. Wind speed (upper panel) and wind direction (lower panel) measured by the Appledore Island WPR at the first profiler level of 123 m agl during August study period. Observations are in black and RAP model runs in red (control) and blue (experimental) at the model initialization time.

We determine the impact of assimilation first by looking at the land-based profiler sites, and later by looking at the *RHB* profiler. The measure of the comparison we use is the RMSE of the scalar wind and the vector wind for the control and experimental simulations over the lowest 2 km of the atmosphere. Each model runs for 15 forecast hours, with the assimilation of the observed data at the initial time.

Figure 3.23 shows the RMSE vs. forecast hour for the August time period. In this section, control runs are in shades of red, and the experimental runs are in blues. The largest improvements in RMSE by the experimental runs over the control runs is at the initial time (forecast hour 00), decreasing at later forecast hours. For the August period, this improvement remains up to 5-7 forecast hours for all models (Figure 3.23). The assimilation of wind profiling radar data thus helps to improve the model forecasts for several hours. Similar results were found for the Wind Forecast Improvement Project (WFIP), as reported in Wilczak et al., 2014.

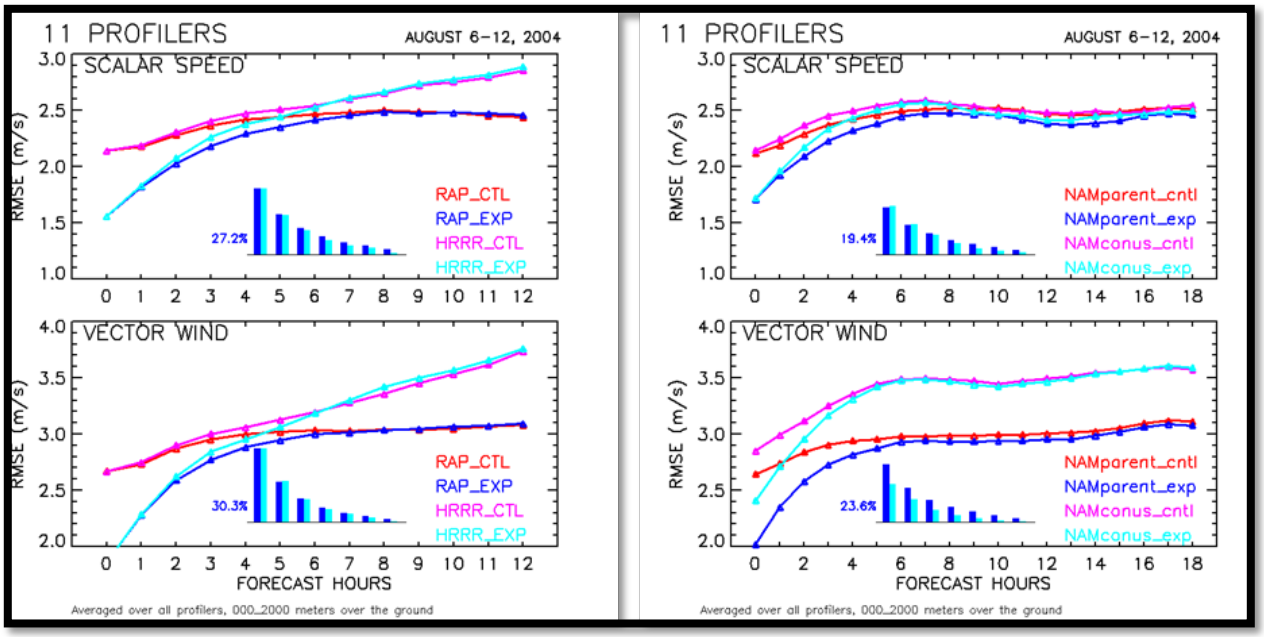


Figure 3.23. Land-based WPR vs. model comparisons for the August period. RMSE statistics of all RAP/HRRR models on the left and NAM models on the right as a function of forecast lead time, averaged over 100-2000 meters above the ground. The experimental model runs (blue and cyan) show lower RMSE compared to the control runs (red and magenta) at the initial time and up to 5-7 hours into the forecast. The inserts schematically show the model improvements, blue – for the low-resolution models and cyan – for the high-resolution models at the initial time plus at others 6 forecast hours. The value (%) of improvement relative to the first blue bar is shown on the left of the inset bar graph as a reference.

The statistics are less positive for the July period (Fig 3.24.). Both models showed significant differences between RMSE for the experimental and control runs, for all forecast hours. The model improvements were mostly negative except for the first 2 forecast hours. The middle of the July period (13-15 July) featured a storm vortex passing through the Gulf of Maine, and easterly flow north of the vortex.

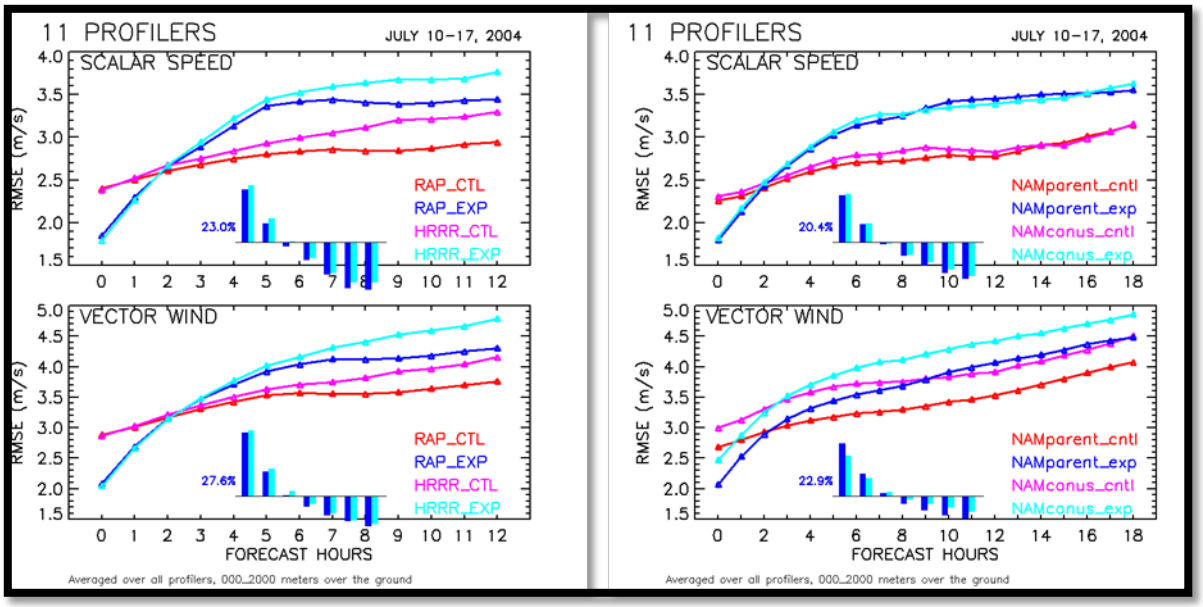


Figure 3.24. Same as Figure 3.23 but for the July period.

We found that the Appledore Island profiler data may be used as a good indicator of the changing weather (Figure 3.25). During the problematic time periods of 13-15 July, the wind direction changed to the south, southeast, and east directions, whereas for the other days the westerly or northerly wind directions prevailed. The model statistics did not perform as well during this period. As a matter of interest, we further analyze these July days, and a detailed analysis of this time period is given in Appendix D. The analysis shows that the anomalous behavior of the models for the July period was limited to that stormy precipitation period of 13-15 July and that period dominates the July verification statistics.

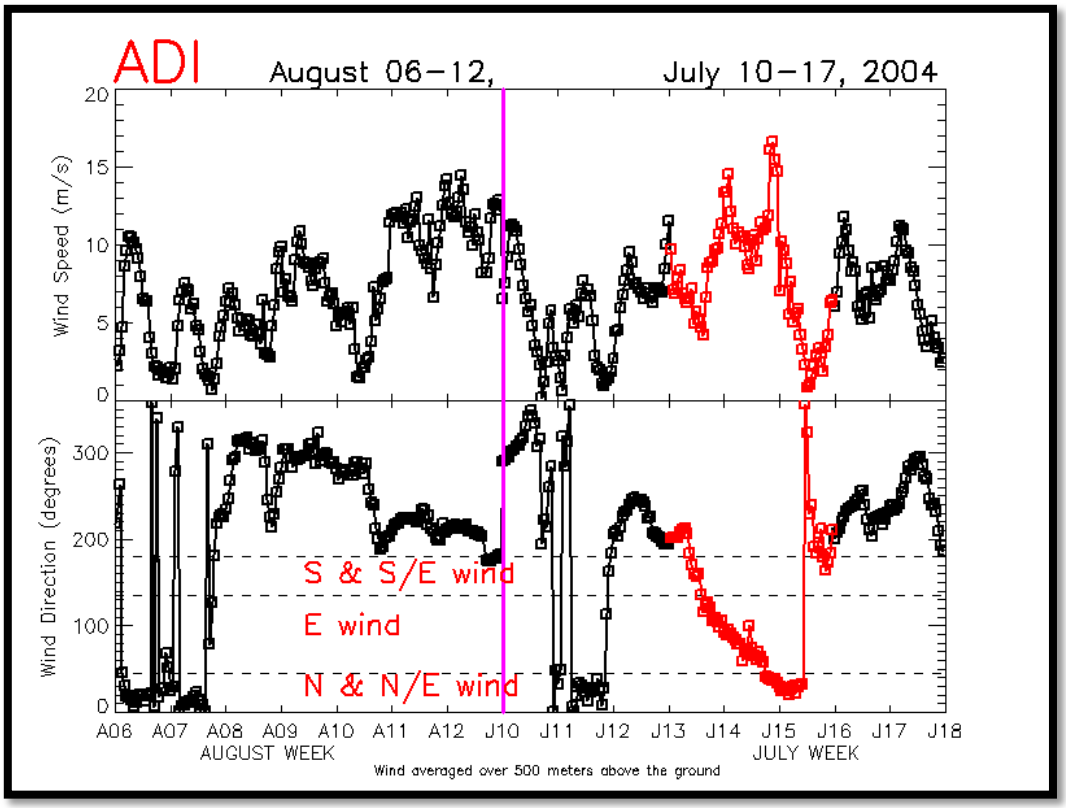


Figure 3.25. Appledore Island WPR data of the wind speed on the upper panel and direction on the lower panel. Wind is averaged over 500 m above the ground. The red color indicates the easterly flow for the problematic time period of 13-15 July 2004.

As part of this more detailed analysis, we compared daily time-height cross sections of the wind from the experimental versus control runs, and we located the profiler sites where the experimental and control models had larger differences. These sites were mostly located on the coast of the Gulf of Maine. We then re-ran the statistical analysis presented before but excluded those negatively affected sites and days, and the results are presented in Figure3.26. When those days were removed from the analysis dataset, the rest of the July days behaved very much like the August period. The models' behavior is now similar to that of the August time period, when the experimental runs showed larger RMSE improvement at the initialization time, remaining for 5-6 forecast hours for all models. The overall conclusions are that employing a finer mesh and assimilation of profiler data improved the model forecast for a few to several hours.

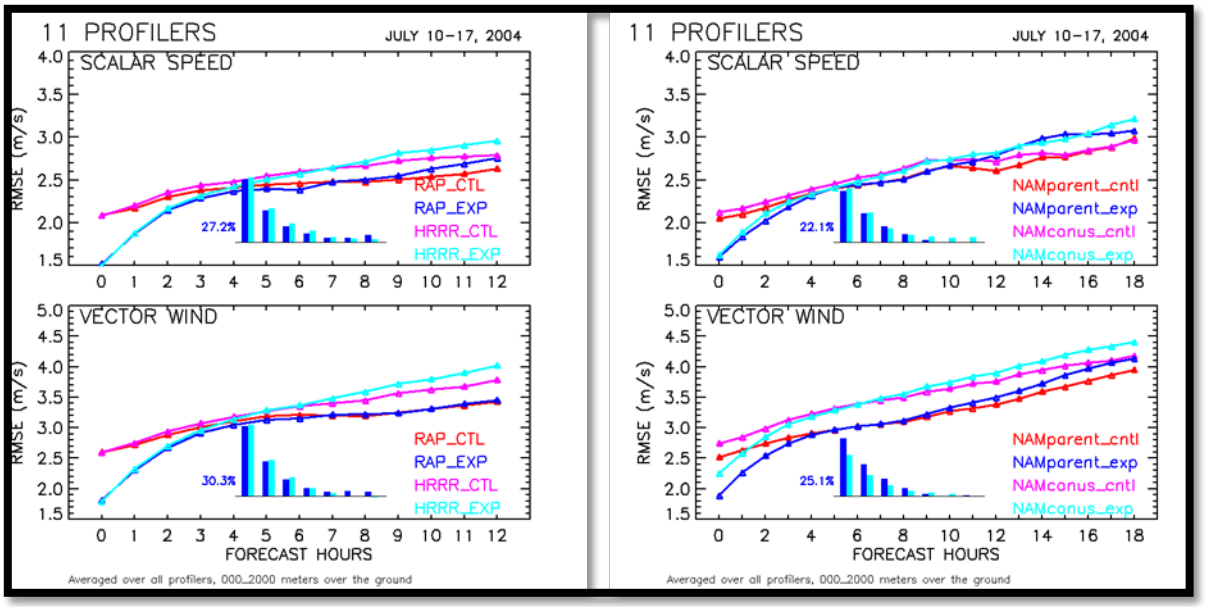


Figure 3.26. Same as Figure 3. 24 but excluding the following number of sites on the following days: 13 July – 7 sites; 14 July – 7 sites; 15 July – 4 sites.

To complete the analysis of how the assimilation of the land-based wind profiling radars data impacts the forecast, we next examine the profiler data collected from the wind profiler located on the *RHB* and compare them to the same models used in the preceding analysis. The track of the *RHB* was presented earlier in Figure 3.10 and the models' predictions are extracted at the moving location of the ship on an hourly basis. The data collected in 2004 by the wind profiling radar located on the *RHB* have been reprocessed for this experiment in particular, using the most up-to-date post-processing wind profiler algorithms in order to improve the data quality by reducing contamination from sea-clutter as much as possible (Weber et al., 1993). An example of a time-height cross-section of the observed wind together with those of all four model control-run outputs is shown in Figure 3.27. Even without assimilation, all models show good agreement with the observations: at the lower level of the atmosphere the wind direction changed from northeasterly at the beginning of the day to northwesterly in the middle of the day and finally southwesterly; low wind speeds were present in the first 5-6 hours of the day for the lowest 500 meters over the sea level, increasing in strength when the direction changed.

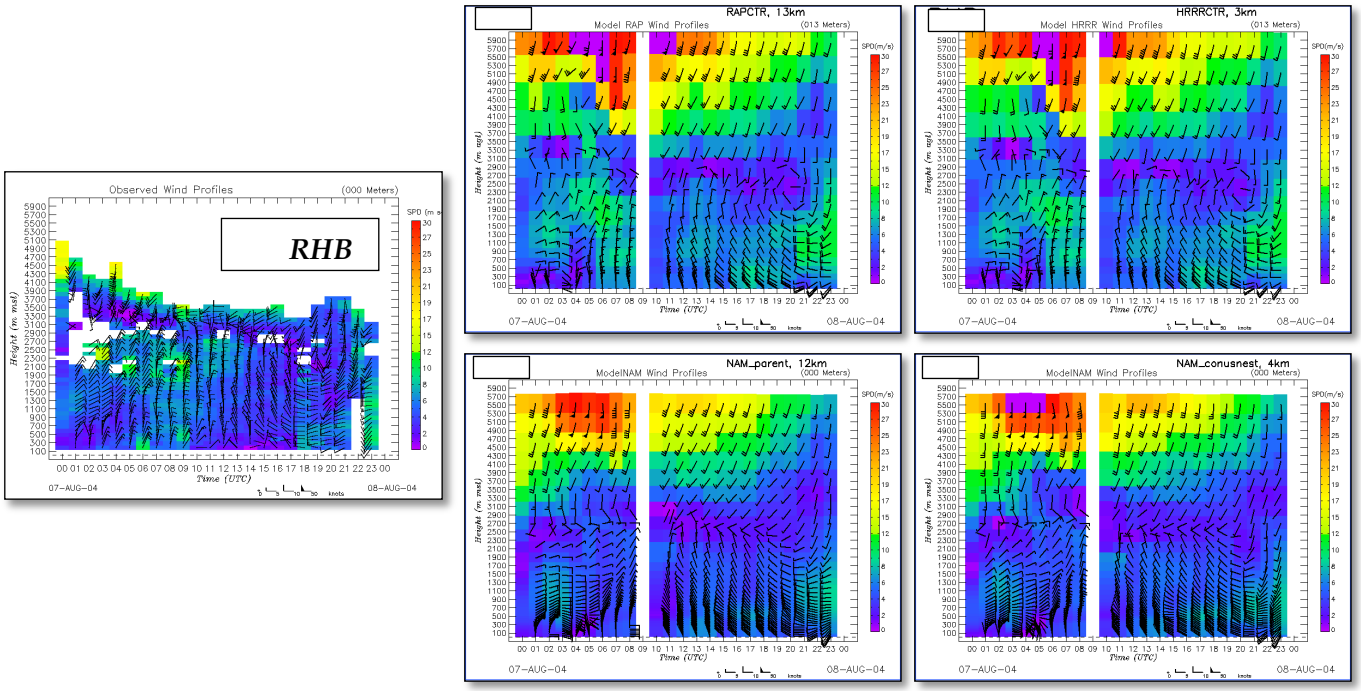


Figure 3.27. Time-height cross sections of WPR-measured and modeled winds. RHB profiler observed data on the left, RAP/HRRR data on the upper panels and NAM parent/CONUS nest data on the low panels. Data are shown for 7 August 2004. The model outputs are shown for forecast hour 00.

For the August and July analysis periods, we compare the mean scalar wind profiles of the *RHB* wind-profiling radar observations with the mean scalar wind profiles of the models up to 4 km. In Figure3.28 we show the comparison of the profiler observations versus the RAP/HRRR (control and experimental) models in the top panels and the comparison of the observations versus the NAM/CONUS (control and experimental) models in the bottom panels for the two time periods (the August period is presented on the left and July is on the right). The July data-set includes 5 days only, as we exclude the 13-15 July days from this analysis as well. We note that the experimental runs agree well with the observed scalar winds throughout the entire profile, but especially for the lowest 2 km of the atmosphere.

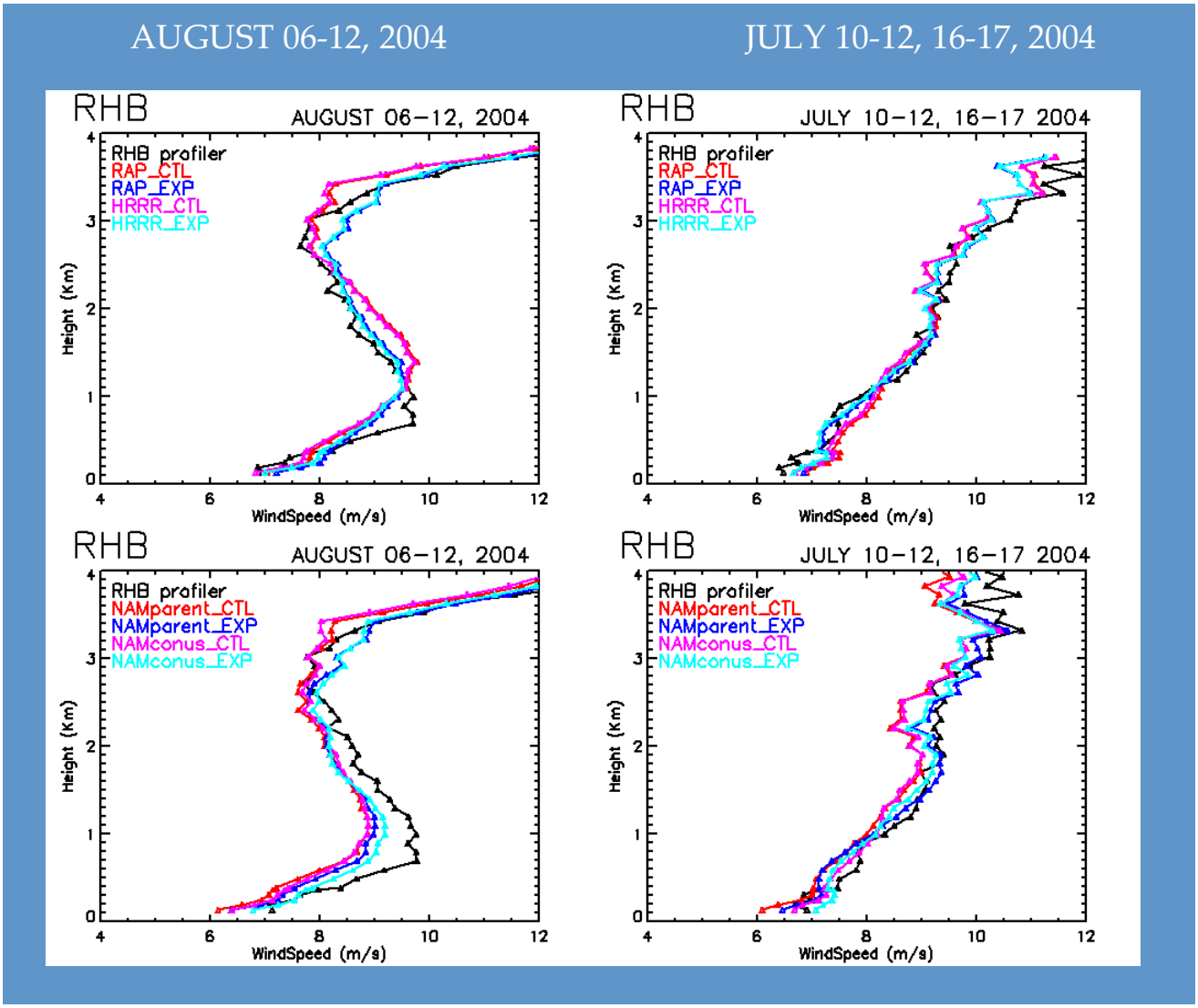


Figure 3.28. Scalar wind profiles of observations and models up to 4 km above the sea level, averaged over the two chosen periods (August on the left and July on the right). The model outputs are compared at forecast hour 00.

In Figure 3.29 the RMSE statistics for the RAP/HRRR (experimental and control) models are on the left and the NAM/CONUS (experimental and control) models are on the right. The data-set used here is again limited to 12 days only. RMSE values are computed over the first 2000 meters of the atmosphere. From this plot we notice that experimental runs improved the RMSE compared to the control runs, particularly over the first 5-7 forecast hours for the RAP/HRRR models and much longer for the NAM/CONUS. The fact that the improvement persisted at least for the first 7 forecast hours is further illustrated in the schematic bar-plots inserted in the figures, where the value of improvement relative to the first blue bar is shown on the left of it as a reference.

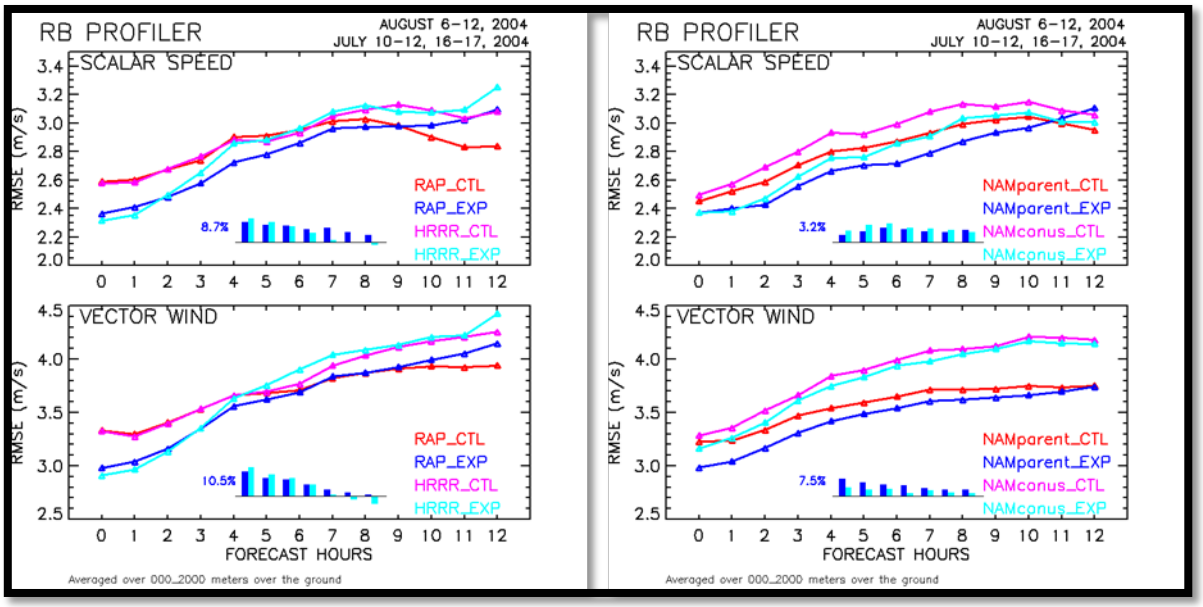


Figure 3.29. Same as Figure 3.23 but for RHB profiler, over 12 days from both chosen periods.

Profiler-model findings.

Twelve of the 15 simulated days showed improved experimental forecasts both onshore and compared to the *RHB* wind profiling radar observations. On 3 consecutive days, coincident with a passing mesoscale vortex, the experimental forecasts were worse than the control runs.

Land-basedWPR/model: For the all of August data-denial episode and the July episode except for the Gulf of Maine sites on days with onshore flow, the 0.1-2.0-km layer-averaged RMSE of the scalar wind speed of the RAP and HRRR models ranged from 2-3 m s^{-1} over the 12 forecast hours simulated. Assimilation of the land-based WPR's reduced the RMSE by up to 0.6 m s^{-1} (27%) at forecast hour 0, then decreases at later forecast hours, becoming insignificant by forecast hour 6.

RHB profiler/model: This profiler, which was not assimilated into the models, independently confirms the usefulness of the land-based profiling radar network. The scalar speed RMSE improvement for the August episode and offshore flow days in the July episode (12 days total) was near 8% at forecast hour 1, decreasing with time and becoming insignificant at forecast hour 8.

d. The RAP/NAMRR Ensemble, and further profiler comparisons with NAMRR and NAMRRconus

A simple, two-member ensemble mean (ENSMEAN) was generated by taking a simple average of the NAMRR and RAP forecasts. The next generation of operational NWP at NCEP will feature a merging of the NAMRR and RAP/HRRR systems toward an eventual North American Rapid-Refresh Ensemble (NARRE) system. Therefore, this simple two-member ensemble is meant to provide a preview of (and somewhat of an early insight into) the next generation of operational short-range NWP at NCEP.

In this section verification results of this ensemble vs. land-based profiler data are compared against results for each model separately. First, the NCEP forecast system has routine verification software which was run for an independent verification of the NAMRR models vs. the profiler data. Detailed verification of the 2-m temperature and 10-m winds were also performed. These analyses are presented in Appendix E.

a) Data impact results

The NAMRR system performed profiler-assimilation experiments for both summer 2004 study periods, 6– 12 August and 10 – 17 July. Prior to the beginning of each study period, a full day of cycling with the NAMRR was done to allow the system to spin-up.

The impact of the assimilated POWER profiler data on the NAMRR model forecast skill was evaluated by comparing forecasts against POWER land-based profiler observations. This was accomplished using the NCEP Forecast Verification System (Novak et al. 2013). This system uses model-predicted values from gridded binary format (GRIB) and interpolates the forecasts to measurement locations to calculate the verification statistics. Note that the forecast data in GRIB files have run through a post processor and use pressure as the vertical coordinate which has vertical spacing of 25 hPa for standard meteorological variables (e.g. wind, temperature, height, etc.). Statistical significance testing was done using the resampling technique for paired data described in Hamill (1999). In all verification experiments, resampling was done with 2000 samples. Any lines in the verification plots which lie outside the boxes are significant at the p=0.05 level. No bias correction has been applied to the NAMRR forecasts. The control simulations for the 12-km parent and 4-km CONUSNEST will be referred to as NAMRR and CONUSNEST respectively. The experimental simulations for the 12-km and 4-km domains, which assimilated the special POWER observations, will be referred as NAMRRX and CONUSNESTX.

i.) POWER profiler assimilation statistics

Prior to the evaluation of the forecasts, it is important to ensure that the POWER WPRs were successfully assimilated into the NAMRR system. One such, albeit brief, way of doing so is to look at the fits of the background forecast going into the analysis to the new observations /measurements (i.e. the POWER WPRs). Mean innovation statistics from each TM00 assimilation cycle, i.e. the *innovation* is the measurement minus the background forecast which goes into the analysis, for each domain shows distributions that are relatively Gaussian for both the July (Figure 3.30) and August (Figure 3.31) study periods. For the July study period, the NAMRRX system assimilated the same number of observations on its parent 12-km domain as it did on the 4-km CONUSNETX domain, which amounted to 112 364 individual observations. For the August case the parent assimilated 74 188 observations while the CONUSNEST assimilated one less. It should be noted that these sums do not include counts from the so-called catch-up cycles of the NAMRRX, merely the on-time analyses occurring at TM00 (Figure 2.10).

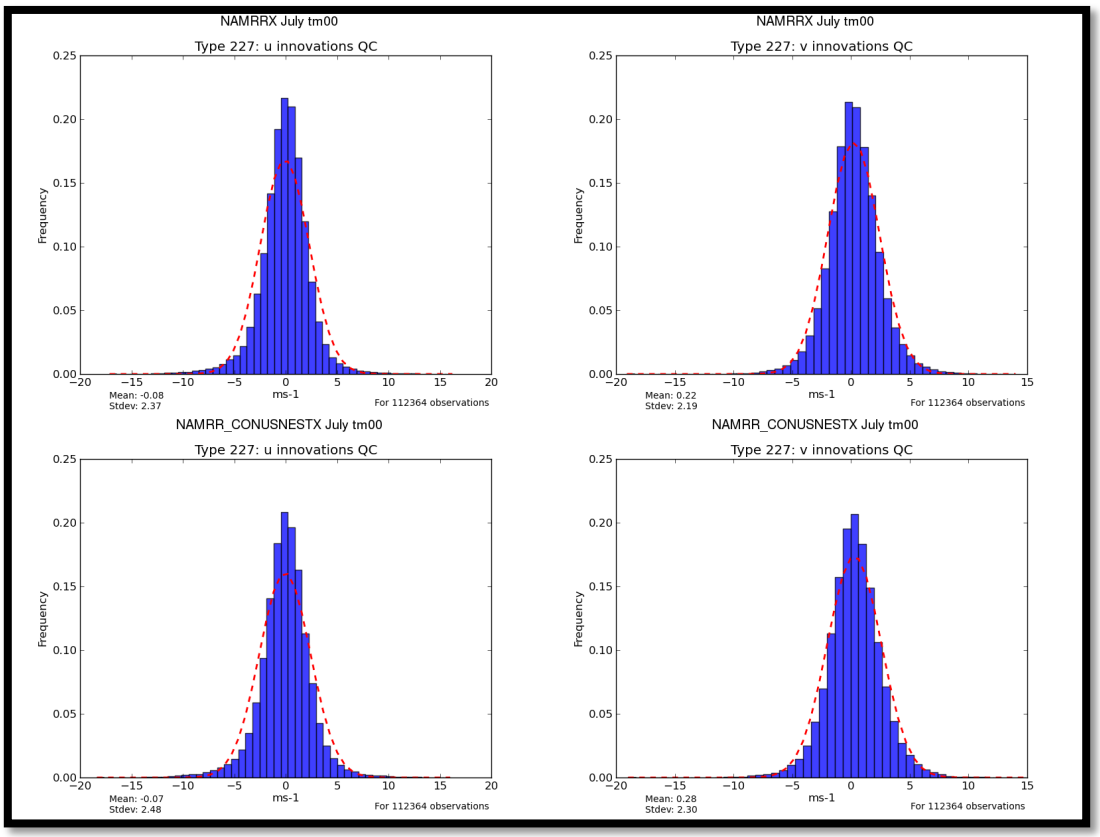


Figure 3.30. Wind profiler (left) zonal (u) and (right) meridional (v) wind observation innovations (observation-forecast) from all TM00 analysis steps during the July study period. Distributions featured along the top are from the NAMRRX while distributions along the bottom row are from the NAMRR CONUSNEST EXP. The red dotted lines correspond to Gaussian distributions.

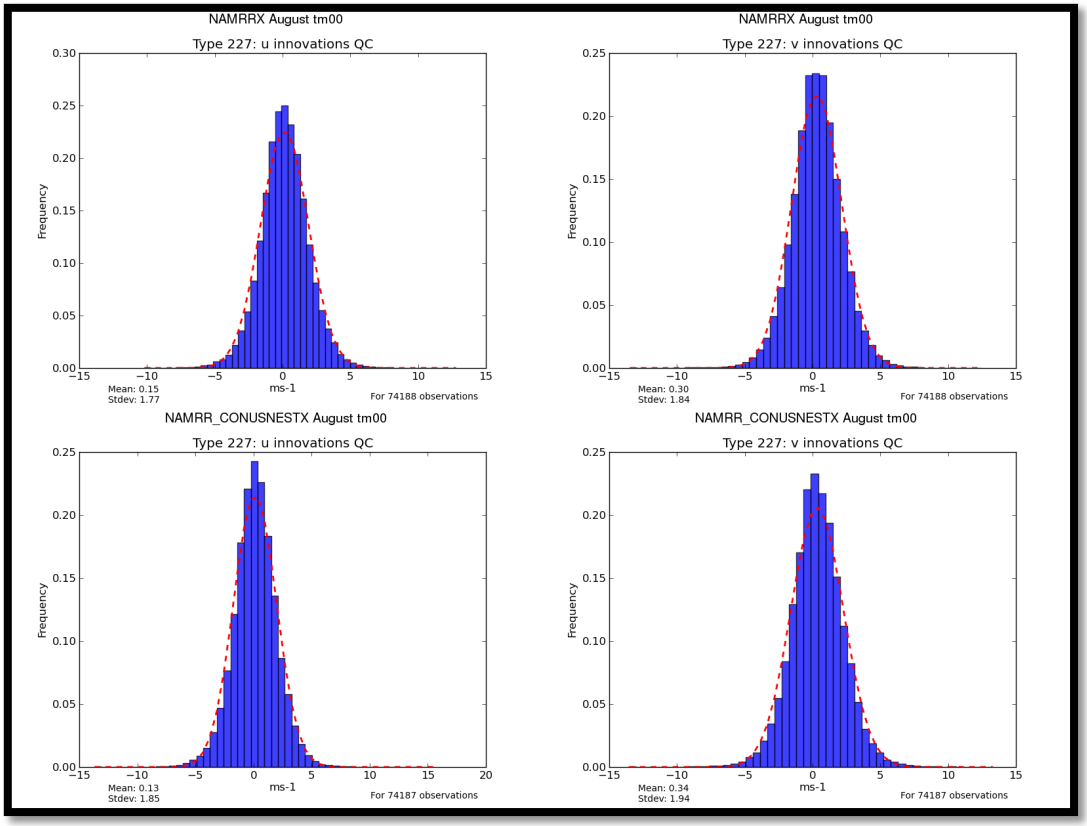


Figure 3.31. As in Figure 3.30 except for the August study period.

ii.) Bulk statistics against POWER profilers

Verification against POWER profilers was computed over all profilers within a 0.1-2km AGL layer.

1) 6 – 12 August, 2004

Figure 3.32 depicts the RMSE (top) and bias (bottom) for the August study period for both the 12-km NAMRR and 4-km CONUSNEST NAMRR domains. Verification statistics are nearly identical in behavior when comparing both domains, and we see that the assimilation of POWER profiler data has a lasting impact throughout the 12-h forecast period, as can be seen in the reduction in RMSE and improved wind speed bias. Impressively, these differences are statistically significant, or nearly so, through the entire forecast period for both domains.

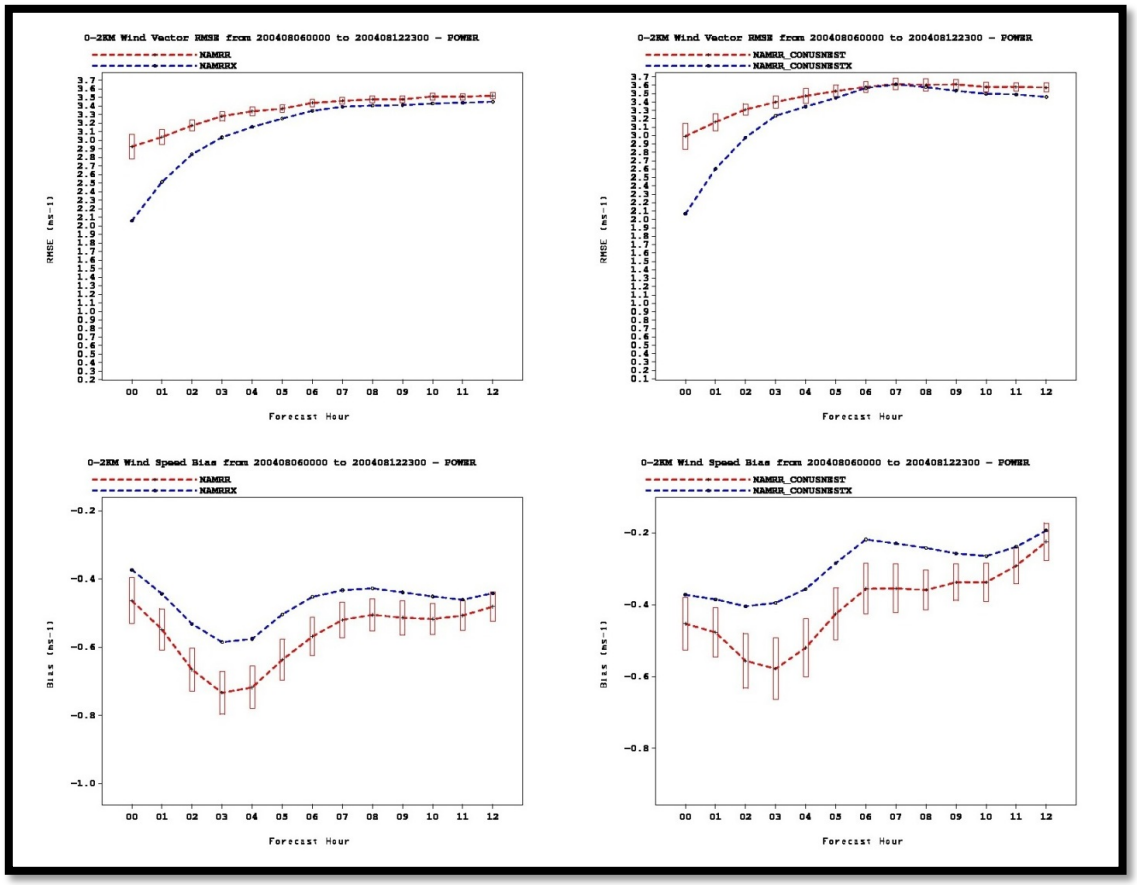


Figure 3.32. Vector wind RMSE (top) and wind speed bias (bottom) against WPR observations within the 0.1-2km AGL layer for the August period. Statistics from the 12-km parent domain occupy the left panels and forecasts from the 4-km CONUS nest domain. Red traces correspond to the control while blue traces correspond to the experiment.

Figure 3.33 is similar to Figure 3.32 except that it shows the effects of ensembling for the August runs, i.e. taking a simple two-member average of the RAP and NAMRR forecasts. The verification statistics from Figure 3.33 immediately show that the ENSMEAN forecast provides statistically significant improvements over the individual forecasts from the NAMRR and RAP systems, both when POWER profilers are not assimilated and when they are. Furthermore, note that when profiler observations are not assimilated, the forecast from the RAP system produces forecasts with lower RMSEs than that of the NAMRR. However when both systems assimilate the POWER observations the performance of the NAMRR and RAP systems, in terms of RMSE, are nearly identical. The biases indicated in Figure 3.33 show that both the NAMRR and RAP have wind speeds that are generally too slow through the forecast period, so it is no surprise that this is also the case with the ENSMEAN. When POWER observations are assimi-

lated, differences among NAMRRX, RAPX, and ENSMEANX remain approximately the same as those found without POWER profiler assimilation.

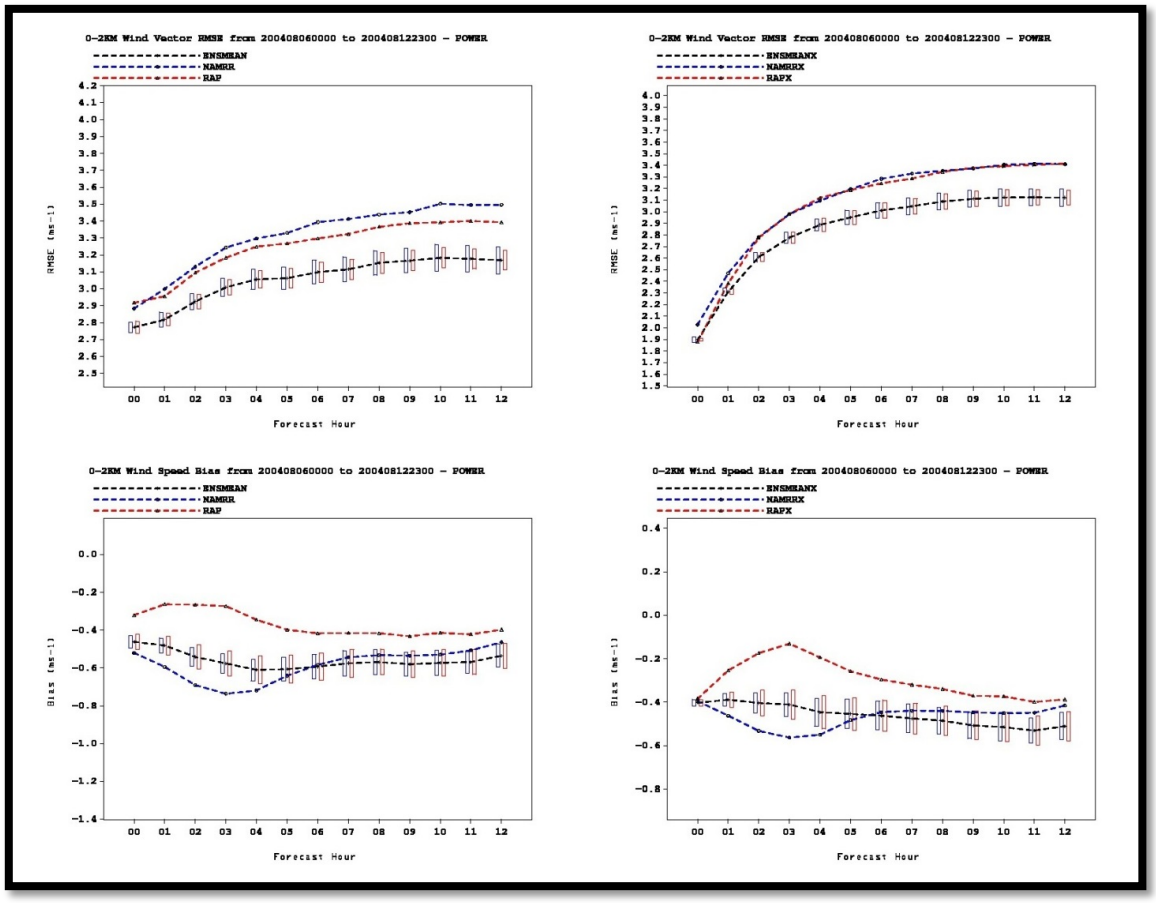


Figure 3.33. Vector wind RMSE (top) and wind speed bias (bottom) against WPR observations within the 0.1-2 km AGL layer for the August period. Each panel shows statistics comparing RAP (red), NAMRR (blue), and their mean (ENSMEAN; black). The left panel corresponds to the control simulations while the right panel corresponds to the experiment simulations.

2) 10 – 17 July, 2004

Figure 3.34 depicts the RMSE (top) and bias (bottom) for the July study period for both the 12-km and 4-km NAMRR domains. Verification statistics are nearly identical in behavior when comparing both domains. During the first 4 hr of the forecast, a generally statistically significant reduction in RMSE can be seen. However from 6 – 12 hr we see a statistically significant degradation in RMSE statistics. Bias shows degradation (decrease) during the first 6 hr and improvement during the 9 – 12 hour period; however the latter part of this period is not statistically significant.

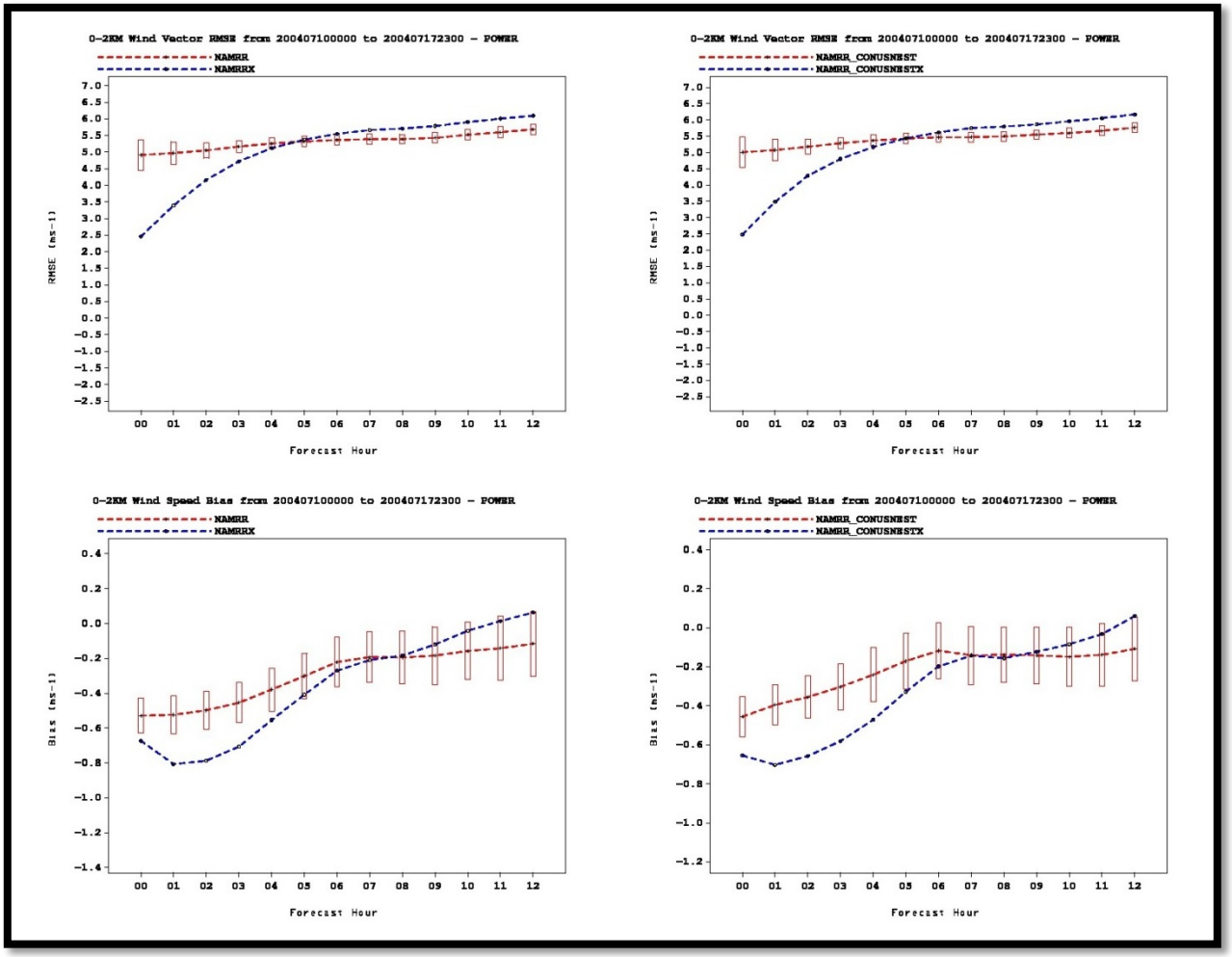


Figure 3.34. As in Figure 3.32 except valid for the July study period. Vector wind RMSE (top) and wind speed bias (bottom) against POWER wind profiler observations within the 0.1-2km AGL layer for the July period. Statistics from the 12-km parent domain occupy the left panels and forecasts from the 4-km nest domain. Red traces correspond to the control while blue traces correspond to the experiment.

The ensemble verification statistics for July from Figure 3.35 show that the ENSMEAN forecast provides an improvement in RMSE over either forecast from the NAMRR or the RAP systems, both when POWER profilers are not assimilated and when they are. Furthermore, note that when profiler observations are not assimilated, the forecast from the NAMRR system produces forecasts with lower RMSEs than that of the RAP. However when both systems assimilate the POWER observations the performance of the NAMRRX and RAPX systems, in terms of RMSE, are more similar but instead show the RAPX having slightly lower RMSE's than the NAMRRX for the first 5 hours of the forecast. Again, the NAMRR and RAP have wind speeds which are generally too slow through the forecast period, and so this is also the case with the

ENSMEAN. When POWER observations are assimilated, the negative bias remains and is slightly worse at the beginning of the forecast period for all systems. Also note that the assimilation of profiler observations for the RAPX in general had the same effect it did on the NAMRRX in Figure 3.32 – forecast degradation in RMSE was noted relative to RAP by about forecast hour 5.

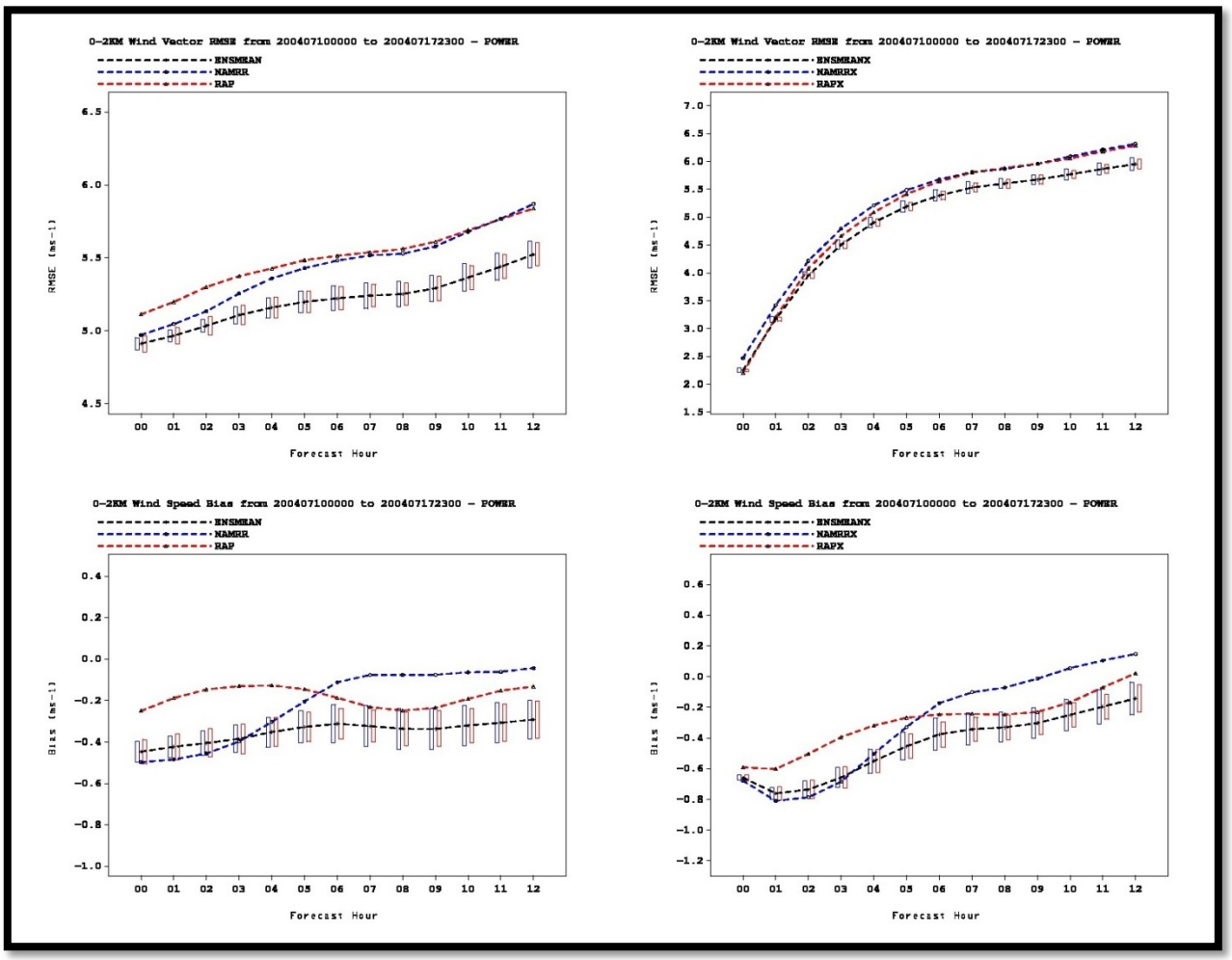


Figure 3.35. As in Figure 3.33 except valid for the July study period. Vector wind RMSE (top) and wind speed bias (bottom) against POWER wind profiler observations within the 0.1-2 km AGL layer for the July period. Each panel shows statistics comparing RAP (red), NAMRR (blue), and their mean (ENSMEAN; black). The left panel corresponds to the control simulations while the right panel corresponds to the experiment simulations.

b) July period

The overall modeling goals of the POWER project were to evaluate model skill for potential use in resource assessment as well as the forecasting implications. The behavior of the model validations during the days surrounding the 14 July frontal passage do not directly address these goals, but are interesting from the standpoint of current assimilation techniques, and what conditions may have led to this behavior. A detailed analysis of this time period (Appendix D) shows that a distinct flow feature, a cyclonic storm system, moved through the study area during this time period (see Figure 3.36 and Figure D.7-8), bringing precipitation and the easterly flow at Appledore Island shown in Figure 3.25. This low-pressure system moved through rather quickly, and would have a small impact on longer-term averages. We note that mistiming or misplacement of a flow feature by a model can lead to large error statistics using normal error metrics, such as RMSE. A detailed evaluation of when and at which profilers the largest errors occurred showed that the error was dominated by errors at the coastal profiler sites and for a limited time period (Appendix D). Similarly, verification of model data at available met-buoy sites showed that large errors occurred at only a few buoy sites—those farthest east in the Gulf of Maine—and for a limited time period. If the POWER error statistics are evaluated for profiler locations and time periods in August and July other than the brief period when the storm was present, the July results look very similar to the August results, and they can be analyzed together as in Figure 3.29.

Although the degraded RMSE for the July study period in the 6-12 hr range of the experimental runs in spite of the profiler-data assimilation appears somewhat surprising, this kind of effect has been noted before. Both Morss and Emanuel (2002) and Semple et al. (2012) found that it is not unusual for assimilated observations to have the potential to degrade the analysis and forecast on occasion. The impact of the data on the forecast is a function of the quality of the observations, the forecast model, the data assimilation methodology, and how well the atmospheric regime in question is represented by the data (e.g. Morss and Emanuel 2002). In the case of POWER we know the profiler observations have gone through substantial quality control. However it is worth noting that the NAMMRX, CONUSNESTX, and RAPX *all* depict forecast degradation in the 6-12 hour time range. This degradation, which exists amongst different modeling systems, is most likely a combined effect of the lack of flow-dependence in the background error term associated with 3DVar as well as insufficient sampling of the atmospheric regime.

The specification of the background error term in data assimilation is exceptionally important and determines how the observations are spread out into the analysis. With 3DVar a climatologically based background error term is used, which misses the so-called "errors of the day" (Lorenc 2003; Kalnay 2003). In other words, the background error term used in the 3DVar analysis system may have spread the information from the observations into the analysis in a manner that did not well reflect the current atmospheric regime. In addition, note that the pro-

filer observations are, generally, restricted to the northeast coast of the United States with none residing farther offshore (Figure 2.6). Therefore it is quite likely that the profiler observations would incompletely sample the onshore flow produced by synoptic systems that can heavily impact low-level wind forecasts. Such a case was observed during the July study period (Figure 3.36), which also coincided with a time during which the bulk of the poor forecast statistics occurred on 13 July 2004 (Figure 3.37). Future studies could explore the optimal positioning of offshore profilers, with consideration of relevant length scales, such as the Rossby radius of deformation, typically ~ 200-500 km for typical mid-latitude cyclone environments.

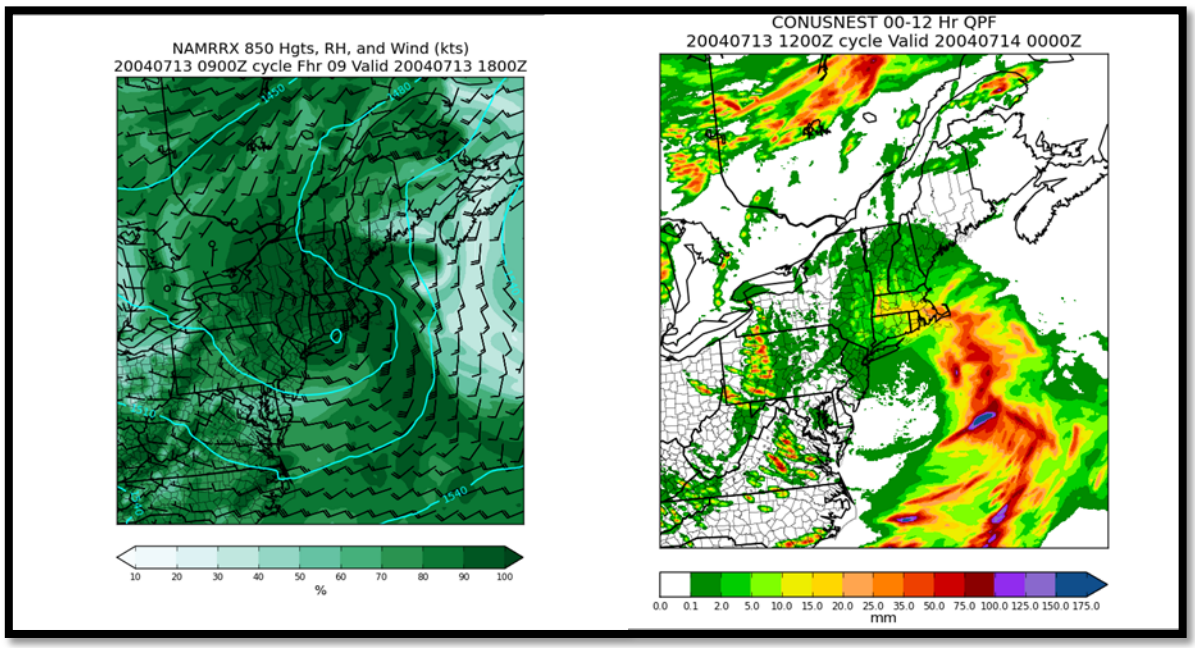


Figure 3.36. Left panel shows the 12-km NAMRRX 9-hr forecast 850-hPa heights (contoured), winds (barbs), and relative humidity (shaded). Right panel shows 12-hr accumulated precipitation from the CONUSNESTX initialized on 1200 UTC on 13 July 2004.

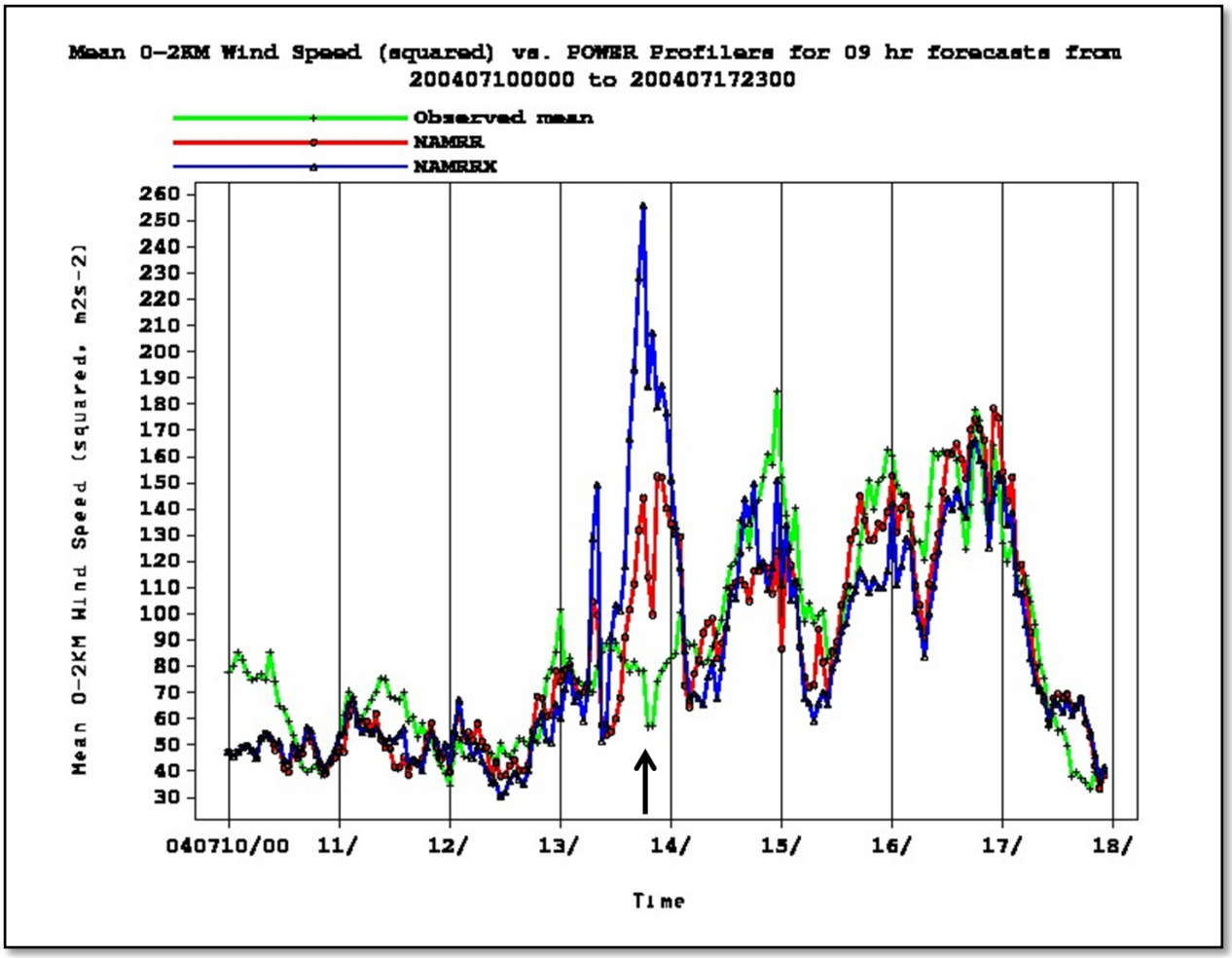


Figure 3.37. Forecast and observed mean 0.1-2km AGL wind speed (squared) for the duration of the July study period at POWER profiler locations. NAMRX is shown in blue, and control NAMRR is in red. Observations are shown in green. Note these are 9-hr forecasts. Arrow shows the approximate time period corresponding to the left panel of Figure 3.36, which also depicts marked forecast degradation.

c) Model evaluation synopsis

Overall we may conclude that the assimilation of the land-based wind-profiler data offered an improvement in the very short term, ~0–6 hour, wind forecasts in 0.1-2 km AGL layer. Improvements were seen in RMSE in both the NAMRR's parent 12-km domain and the 4-km CONUS-NEST domain, both of which assimilated these observations.

A degradation in RMSE statistics was noted during the July period beyond forecast hour 5 (Figure 3.32), which upon further investigation of time series and synoptic pattern, appeared likely to be a combined issue of data assimilation technique and observation representativeness, which is supported by current scientific literature (Morss and Emanuel 2002; Semple et al. 2012).

Additional work involving either more advanced assimilation techniques and/or more extensive observing networks is needed to help understand and answer this question.

The impact of averaging the RAP(X) and NAMRR(X) forecasts into a very simple two-member ensemble mean had the overall effect of reducing RMSEs for both POWER profiler and surface observation based verification. However, one might suggest that this could be result of the slight smoothing effect one would obtain from the averaging process. It is known that smoother forecasts are generally less susceptible to issues of the so-called 'double-penalty' than their less-smooth counterparts. For example, if a model has a forecast which completely misplaces a precipitation event by only a few grid points, it may be penalized once for not forecasting the observed event and penalized again for forecasting an event in the wrong location (even though it was spatially quite close). Subjectively, although the forecasts with finer detail may appear more visually favorable relative to coarser guidance, subtle timing/amplitude differences of big events, like ramps, can lead to large penalties when fine-resolution forecasts are verified using traditional metrics (e.g. RMSE). Smoother forecasts that exhibit comparatively small, large amplitude fluctuations in space and time can be less susceptible to this double-penalty (e.g. Rife et al., 2004). An approach to address this issue has very recently been introduced via a neighborhood approach and adopting probabilistic approaches to forecast verification at observing sites (Mittermaier, 2013). Such an approach to verification may be worth considering as a part of future work.

Brief comparisons between the RAP(X) and NAMRR(X) were also done as a part of evaluating the ENSMEAN forecasts against profiler and surface observations. Without assimilation of POWER profiler data, the RAP(X) and NAMRR(X) were mixed for both periods in terms of verification against the POWER wind profilers. In July the NAMRR had better performance as measured by RMSE (Figure 3.33), but in August the RAP had the lower RMSE (Figure 3.37). Most interestingly was that after both systems assimilated the POWER profiler observations their RMSEs became more nearly identical. However no statistical significance testing was done between these two systems.

Although the intent of this project was not aimed at comparing the forecasts of the RAP and NAMRR systems, it is nonetheless an exceptionally useful exercise. Comparisons of these systems is certainly beneficial to the development of NAMRR's rapid-update forecasting capability, since it is still highly developmental and needs testing against a system that has a long history of being run in a rapidly updated manner. For example, the results described in Appendix E demonstrated that the NAMRR's analysis may need some adjustment to more closely fit surface observations to improve its short term forecast. Furthermore, since the NAMRR and RAP systems will form the basis of a future rapidly updated ensemble prediction system; these kinds of comparisons will become even more useful into the future as we seek to understand each modeling system's characteristics. Finally, it is worth noting that following the completion of POW-

ER retrospective simulations, many changes were made to the NAMRR system, some of which address issues related to things like the NAMRR's 10 m AGL high wind speed bias.

4. SUMMARY and DISCUSSION

The POWER project was undertaken to address the deficiency of direct wind measurements at the heights of turbine rotor blades over the ocean, to use measured rotor-level winds to verify numerical forecast models offshore, and to provide information about the type, spatial density, and general location of measurements needed to support offshore wind energy in the United States. POWER leveraged a two-year collection of 3-km HRRR modeled wind speeds at turbine-height and the high-quality, nearly 6-week research dataset over the Gulf of Maine, which includes accurate profiles of the wind through the rotor layer that can be used to gain insight into wind properties aloft over the ocean and address WE issues. The objectives of the POWER project were to: 1. Focus future research aimed at informing requirements for an observation network to support U.S. offshore wind energy and for improving our understanding of the offshore wind resource. 2. Determine whether additional lower-tropospheric profilers can improve near-coastal and offshore wind resource assessments. 3. Evaluate the skill of several numerical weather prediction (NWP) models in the offshore area. 4. Evaluate spatial and temporal variability of offshore turbine-height winds.

Besides analyzing the existing dataset over the Gulf of Maine, for aid in designing and planning future research campaigns to inform WE, we also, for the first time, performed a study using high-quality winds aloft from a moving shipborne platform to test the skill of NWP forecast models over the water. The shipborne instrumentation featured HRDL measurements of wind properties, compensated for ship motions, and shown to be accurate against other measurement systems, and the *RHB* wind-profiling radar, which extended the profiling capability to 3-4 km ASL.

Flow Properties

The measurement systems aboard the *RHB* showed the winds within the turbine rotor layer to be highly variable in time and space. Strong shear and significant vertical structure were often present in the rotor layer below 200 m ASL, and the analyzed 15-min profile frequency revealed periods of significant changes in wind speeds in this layer at sub-hourly intervals. Low-level jet structure was evident often during the nocturnal portion of the diurnal cycle, contributing to a tendency for the winds to be stronger at night than during the day. Horizontal composites of HRDL-measured wind speeds at vertical levels near hub height and near the surface showed significant horizontal variability and significant differences in the horizontal distribution pattern of wind speeds at the two levels, as shown quantitatively in Fig. C.5. These differences indicate that **extrapolation of near-surface wind measurements to hub heights is**

likely to produce significant errors in the estimated horizontal distribution of the wind fields at those upper levels.

Model comparisons

Forecast model winds at marine wind-turbine levels were directly compared in several ways with HRDL-measured winds through these levels. Time-height cross sections from the sea surface to 500 m showed that the models captured overall values and trends reasonably well. The models at times differed quantitatively among themselves and from the HRDL-measured profiles in magnitude and timing. Vertical profiles of error statistics, including bias, RMSE, and correlation coefficient R^2, calculated for the August study period, showed **model slow biases of 0.5 to 2 m s^{-1} increasing downward to the ocean surface**, model initial-value agreement with the measurements above 100 m ASL to within 2 m s^{-1} (RMSE), and R^2 of 0.8-0.9 (recall that HRDL data were not assimilated into model initial conditions). But below 100 m, these degraded to RMSE values as large as 3 m s^{-1} in an individual case and R^2 as low as 0.5-0.6, indicating that the model initialization deviated more from the non-assimilated observations near the ocean surface.

The diurnal behavior of model skill was investigated by plotting mean model-predicted wind speeds at 100 m ASL for the August time period as a function of the hour of day when the forecast is valid, then comparing these with the mean HRDL wind speed for that hour (Fig. 3.21). The comparisons show that for both control and experimental runs, the **model initial wind conditions (Forecast 0) during daylight hours (12-22 UTC) were close to those measured at the *RHB*, but at night the modeled initial winds significantly underestimated the measured values, especially after local midnight**—by as much as 2.5 m s^{-1} for a 9 m s^{-1} wind measured at 0900 UTC. These initial-condition effects apparently carried over into the 1- to 3-hr forecasts, which showed similar daytime and nighttime behavior. In contrast, the 6- and 12-hr forecasts overestimated the 100-m wind speeds during afternoon and evening hours (1800-0600 UTC) by as much as 2-4 m s^{-1}.

To study the relative skill of the various model configurations, the error (RMSE) in wind speed was plotted against forecast valid time for each of the 8 basic configurations. It is expected that the experimental runs (with profiler assimilation) should perform better at least early in the forecast period than the control runs without assimilation, and that the higher-resolution configurations should in general outperform the coarser-gridded versions. For the RAP and HRRR runs the error statistics were found not to be significantly different for most of the forecast period. The NAMRR CONUSNEST did have smaller errors than its NAMRR parent for the first hour; NAMRR CONUSNEST, however, had larger errors afterwards. The larger error found in the higher-resolution NAMRR CONUSNEST, relative to its parent, is a property

which at times has been noted as problematic in verification of NWP models with finer grid-spacing.

Although finer grid spacing generally offers the ability to resolve more detail, and may often be subjectively better, these additional details may be unfairly penalized by traditional verification methods, such as RMSE (e.g.; Mittermaier 2013). Previously in the WFIP results (Wilczak et al. *BAMS* 2014-15, submitted), higher-resolution models were also found to only equal or even have slightly poorer skill for point verification, but, when using a verification procedure designed to test the timing of events, the higher resolution models demonstrated the ability to better capture wind *ramp* events. Furthermore, it should be noted that the RAP and HRRR used a constricted domain (Fig.X.2.b) for the POWER experiment with a boundary near the Gulf of Maine, smaller than the operational RAP/HRRR domains (Fig. X.2.a). This was expected to contribute to forecast errors in the POWER configuration.

The most significant result was that, **for the August period, the experimental runs with profiler data added all had lower errors than the control runs** by up to 0.2 m s^{-1} (8%) for at least the first 3-4 hr. Results for the July period, on the other hand, included a period of 3 days where the control runs consistently outperformed the experimental runs after the first hour or so, and the higher-resolution nested runs did not consistently perform better than the lower-resolution parent models. Outside of that 3-day period, the July results were consistent with the August period. It is not uncommon for data impact experiments to show brief periods of negative impact for a given observation type, even though the overall impact is clearly positive (Benjamin et al. 2010, Benjamin personal communication), as also found in the WFIP analysis. This effect is further explained below.

The RMSE vs. forecast-lead-time analysis was also performed with the *RHB* profiler winds averaged over the 0.1- to 2.0-km layer. The results for the study period (other than the days affected by the mid-July frontal passage) were generally similar to the HRDL comparisons, the RMSE of the scalar winds decreasing by 0.1-0.3 m s^{-1} for the experimental runs compared with the control runs. The **HRDL and WPR profile data from the *RHB*, which were not assimilated into the models, independently confirm the usefulness of the coastal profiling radar network**. The scalar speed RMSE improvement for the August episode and offshore flow days in the July episode (12 days total) was near 8% at forecast hour 1, decreasing with time and becoming insignificant by forecast hour 8. It is worth noting that the overall reduction in RMSE generally lasted much longer into the forecast with the RHB profiler-based verification when compared with the HRDL-based verification. A primary difference between the two instruments is the depth over which the forecast winds are verified, which is 0.1-2.0 km for the *RHB* profiler and is 20-500 m for the HRDL. Given the noted difference in the verification these results suggest that **the shallower HRDL based verification is highlighting a need for additional model devel-**

opment to improve the winds in the 20-500-m layer. Such work is under active pursuit for the modeling systems discussed in this work (e.g., Sušelj and Sood 2010).

The model forecast winds were also verified against the wind data from the profilers in the coastal and inland array that was used to assimilate into the models. These comparisons allowed us to see whether the verification results were consistent with those from the *RHB*, and also to see how long the assimilation has a positive influence on the forecasts. One expects, of course, that the experimental runs should significantly improve on the control runs for the initial fields and the very short lead times (2-3 hr), because the predicted wind profiles are strongly constrained by the measurements, and this was the case. The improvements for the initial conditions (forecast hour 0) were up to 0.6 m s^{-1} (27% error reduction -- compared with 8% (0.2 m s^{-1}) for the *RHB* winds) for the average scalar wind speed in the 0.1-2.0-km layer, and **forecast improvements were noted out to 5-hr lead-time**. Overall, for all of the August study period and the July period (excluding the Gulf of Maine WPR sites on days having onshore flow, from the mid-July cyclone and frontal passages), **the layer-average scalar wind speed RMSE of the RAP and HRRR models ranged from 2-3 ms^{-1} over the 12 forecast hours simulated**. The results for the land-based profiler arrays were thus qualitatively similar to the *RHB* lidar and WPR results, but the magnitudes of the improvements due to assimilation were larger.

The negative impact of assimilated measurements on model forecast skill, such as occurred in the middle of the July study period, has been mentioned several times throughout the report and in these conclusions. The degradation of the model forecasts during this period, in spite of assimilation of the special land-based WPR observations into the models, may stem from variety of reasons. For this study, the most likely causes, in no particular order, were: (1) the lack of data, including assimilation of offshore profiles, and the arrangement of the observational arrays, (2) the quality of the forecast model, (3) the data assimilation methodology, and (4) the overall complexity of the atmospheric regime (Morss and Emanuel 2002). In 2004 the available datasets lacked many potentially important coastal/offshore observations available today, such as radar reflectivity and radial winds, commercial aircraft data, and only limited satellite data were available. Also, more advanced versions of all the models have been developed since the start of this study. The NAMRRX, CONUSNESTX, and RAPX (-X = experimental) data assimilation systems have also since been upgraded and may have been able to improve their skill for this period with a more diverse set of observations that extend over the coast and a larger domain (Fig.2.9) for these POWER experiments. However we remain somewhat skeptical that a significantly larger domain would reduce much of degradation in the absence of additional observations given that the NAMRR system implemented for POWER used its full-sized domains for both parent and CONUS-NEST domains (Fig. 2.9-right) with similar results. Furthermore, the degradation may have been caused by the lack of flow dependence in the background error term associated with 3DVar, since this period featured a transient short-wave trough, which would not be well described by static background error covariance used in 3DVar (Lorenc 2003; Kalnay 2003). Finally, the profiler observations are primarily restricted to the coast with none residing farther offshore and none over western New England or New York State (Fig. 2.6). Therefore it is quite likely that the profiler observations would incompletely sample the

onshore flow produced by synoptic systems moving up the coast, which can heavily impact low-level wind forecasts. Future studies should use more recent periods with current observational data and more state-of-art assimilation than used for POWER (e.g., NAMv3.1 – Rogers et al., 2014; RAPv2 – http://rapidrefresh.noaa.gov). We do not rule out that new well-offshore (200-500km) profile measurements might also help forecast accuracy for turbine-height winds off the New England coast. In the recent DOE-supported Wind Forecast Improvement Project (WFIP), which included these newer data sets, it was shown that routine assimilation of widely spaced profiler observations over two regions also offered consistent improvements in wind forecasts from RAP and NAM systems. New, well-offshore (200-500-km) profile measurements may also help improve forecast accuracy for turbine-height winds off the New England coast. This previous work from WFIP, as well as the bulk of the POWER results depicted here, supports the potential high-value of assimilating profile observations on the overall improvement of wind forecasts for WE needs.

It has also been noted that offshore model speeds were biased slow in the turbine layer by up to 2 m s^{-1}, with larger biases at night, when using the MYJ PBL scheme in this project. Sušelj and Sood (2010) found improvements to the MYJ PBL scheme over water simply by replacing the original mixing length scale formulation with that from the MYNN PBL scheme and updating the flux-profile relationships at the lower boundary to be consistent with the new mixing new lengths. Recently, improvements in model wind speeds have been found in the RAP/HRRR over land by switching to the MYNN from the MYJ PBL scheme, so we expect that this bias will be lower (better) with the new MYNN BL scheme.

Finally, **a simple two-member average of each NAMRR and RAP configuration produced two-member ensemble predictions of the winds, which were also evaluated to demonstrate the effects of ensembling**. Evaluations were performed using routine NCEP verification software on the 0.1-2.0-km-layer averaged wind speeds. The verification routines were first applied to the individual model runs, which produced very similar results to the previously described verification against the land-based profiler winds, and in fact showed an even stronger improvement in the experimental runs. The **ensemble mean values showed statistically significant improvements over the runs using the individual models**, generally through the end of the 12-hr forecast period simulated.

Sampling requirements

One of the most important decisions in developing a wind plant is where to put it. Differences of 1-2 m s^{-1} in annual average wind speed can translate to $10s M (if not $100s M) over several years. Locations where the resource 'goes away' for a year or even a season can be very detrimental. Measurements with the *RHB* Doppler lidar in this report have demonstrated the variable nature of the wind resource in time and space, and that spatial variability is evident in horizontal composites of wind-speed distribution. An important question for this kind of sam-

pling is whether such spatial variability would be smoothed out of longer-term average distributions.

Longer-term mean distributions from model output for time periods ranging from a season to 2 years, also presented here (Figs. 3.1-3.2), seem to show such smoothing. Immediately evident in these wind-speed maps is the significant gradient in wind speed starting at the shoreline, as wind speeds increase dramatically in the cross-shore (or shore-perpendicular) direction from lower inland wind speeds to much stronger wind speeds well offshore—often several m s^{-1}—a prime reason for developing offshore wind resources. In the along-shore, or shore-parallel, direction the strength of the land-to-sea wind-speed gradient just offshore can vary considerably. For example, the along-shore gradient is significant for the 30-km-offshore wind speeds between the northern and southern Gulf of Maine, and again between Long Island and the Delmarva peninsula (Fig. 3.1).

Based on the HRRR 2-year wind-speed composite maps, an observing network that would be useful for wind resource assessment is shown in Fig. 4.1. For long-term average maps such as this to be used with confidence by the wind industry, verification of key meteorological features is required. Some of the principal features present in Fig. 4.1 are 1) a strong cross-shore wind speed gradient in the first 75-100 km from shore, at all locations; 2) the variability of this gradient in the cross-shore direction, with the strongest gradients closest to the coastline, decreasing in strength the farther one goes from the shore; and 3) the variability of the gradient in the along-shore direction. To verify these key features, a proposed set of buoy-based wind-profiling systems has been added to the wind resource map. This network consists of approximately four cross-shore lines of 3-5 buoy profilers with each line placed in a region of strong, moderate, or weak cross-shore gradient. The along-shore gradient is measured by the differences between the lines in the north-south direction. Another feature of the proposed network is that the spacing of the buoys be kept smaller when near shore, with larger spacing as one goes farther offshore, where the speed gradients become weaker.

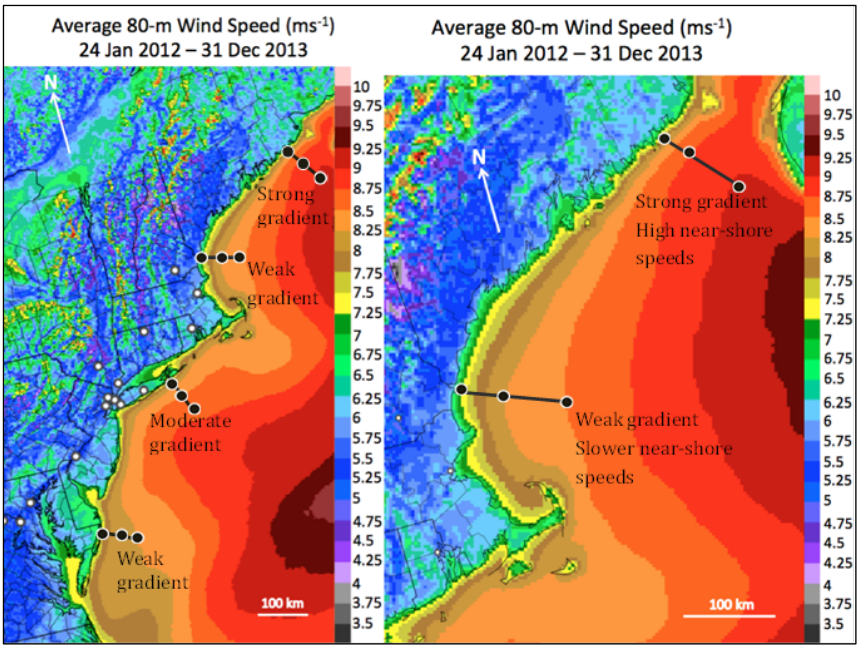


Figure 4.1. Wind-profile measurement array configuration focused on measuring cross-shore gradients of wind speed, (left) larger-scale view, and (right) Gulf of Main and New England centered view.

Figure 4.2 shows the same network, but this time on a map of the percent of time that the wind speed is greater than a speed threshold. Again, the four cross-shore lines are seen to sample the minimum and maximum gradients, and also provide information on the along-shore difference in this quantity.

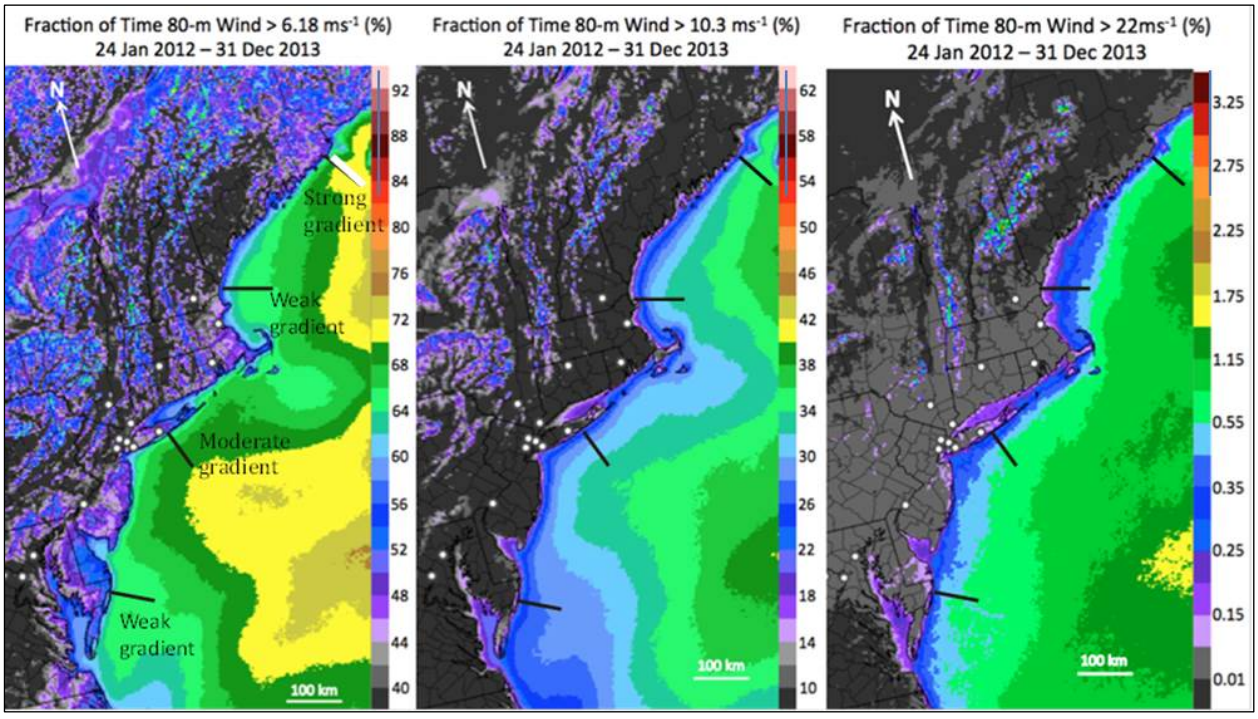


Figure 4.2. Same as Fig. 4.1, except superimposed on maps of percent of time that the wind speed exceeds a given speed threshold.

Such regional differences are important to verify observationally and to understand. An advantage of measurement arrays such as Fig.4.1 is that the cross-shore variations in wind speeds are well sampled and regional differences can be assessed over distances of 500 km or more. A disadvantage is that along-shore variability over distances on the order of 100 km or less is essentially unsampled. Such along-shore variability at smaller scales is also important for making siting decisions, for example which of two candidate wind-farm sites separated by 30 km (or 50 km, or 100 km) in the along-shore direction should be preferred for development. The larger question is, what are the important distance scales of variability in the along-shore direction and what are their amplitudes or potential impacts on WE? Along-shore flow variability may result from coastline shape, including headlands that locally block or accelerate flow parallel to the coastline (for an example, see Rothermel et al. 1998), or terrain effects, such as gaps in coastal topography that focus wind-speed maxima, which can eject out over the sea and may occur often enough to appear in long-term maps of mean wind speed. It is possible for even modest terrain relief to produce such flows, especially in stable conditions that are often present over this region.

An example of this kind of along-shore variability that does show up in the 2-yr HRRR model composites is the area of weak flow off the extreme eastern coast of Maine (black arrow in upper right corner of Fig. 4.3b), southeast of Eastport near the island of Grand Manan, where the 2-yr average wind speed is more than 1 m s^{-1} less than neighboring locations 20 or 30 km away at similar offshore distances. Having a chance at identifying high- or low-wind regions in the along-shore direction with this kind of dimension will require greater granularity in the cross-shore direction than the arrays in Fig. 4.1. An example of an array design where some profiling sites are deployed cross shore and some are deployed along shore is shown in Fig. 4.3a. A variation covering the same general area in the along-shore direction with one fewer sensor is shown in Fig. 4.3b, where the cross-shore sampling is arranged in a diamond pattern. The advantage of these kinds of arrangement is enhanced characterization of the along-shore variability in wind speed, which can be scaled down to 50- or 100-km intervals as needed. The disadvantage compared with Fig.4.1 is that the array has fewer sensors in the cross-shore lines, so that the wind-speed gradients in that direction are not as well sampled.

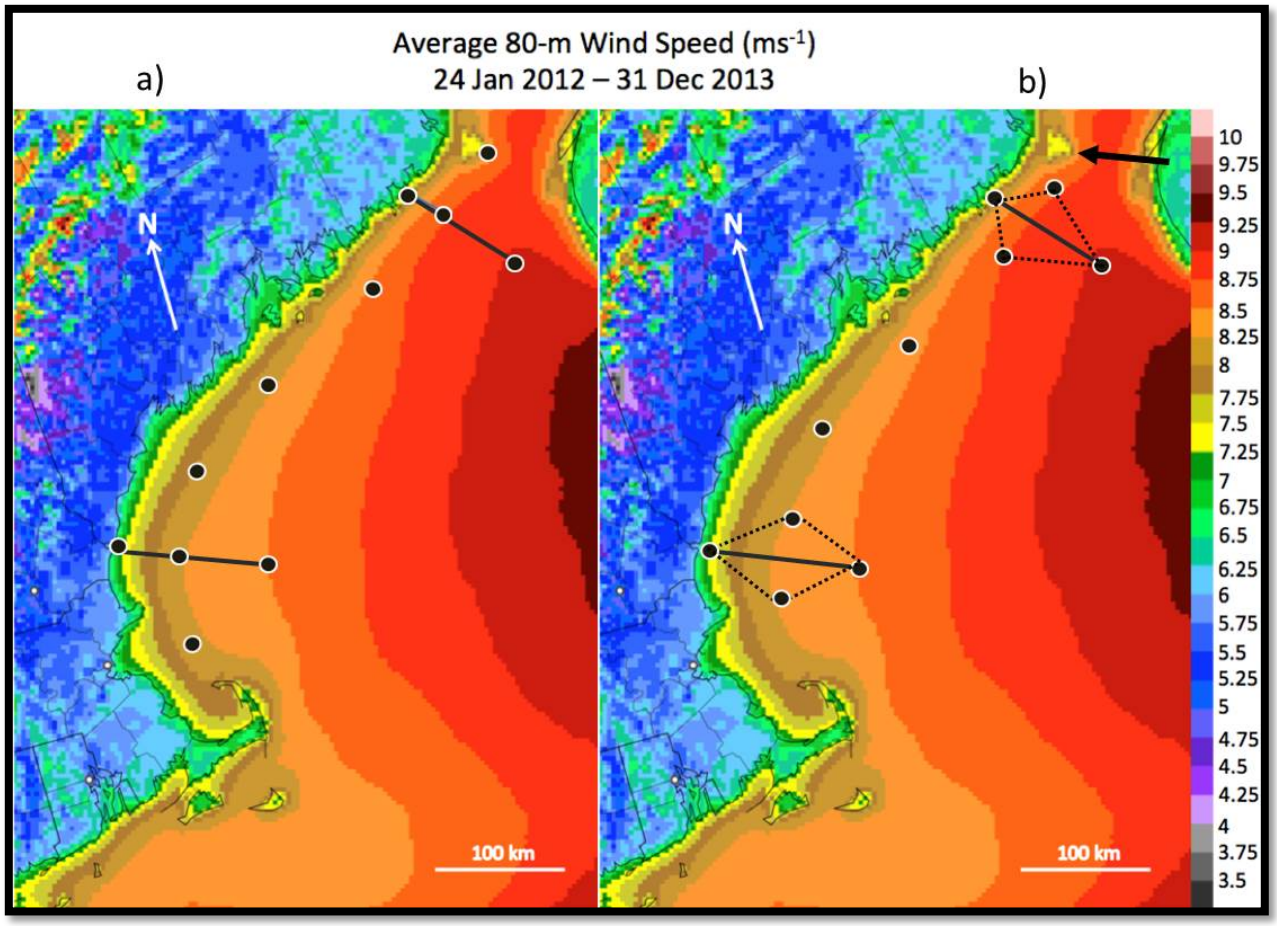


Figure 4.3. a) Example of a deployment of offshore wind profiling sites (black dots) having both along-shore and cross-shore sampling. b) Similar to a, except cross shore sampling is accomplished in a diamond pattern. Black arrow in upper right corner of panel b) indicates area of weak averaged wind speed off extreme eastern Maine.

The model-generated annual and 2-yr maps of hub-height wind speeds calculated for this study appear relatively smooth in the along-shore direction at the smaller scales. The most prominent along-shore variation is a retarding or sheltering effect of Cape Cod, with lower wind speeds indicated in its vicinity in all the long-term averages. An example of even smaller-scale is the eastern Maine area of low speeds just described. This location would seem to be the kind of area to avoid in finding optimum sites for offshore WE development, if these deficits were real and in the right place. The more general question to be addressed is, do other such areas exist, of either higher or lower mean speeds, that did *not* appear in the model composite? **The apparent smoothness of the simulated long-term mean horizontal distribution means either that smaller-scale along-shore variability (on scales of 20-100 km or so) is smoothed out of the annual or longer means, or that the models are not reproducing this variability.** Before we rely heavily on models for answers on siting decisions, it is important to find out which alternative is true.

The available measured hub-height horizontal distributions from lidar data of the NEAQS-04 cruise (Figs.3.13 and C-5) show some similarities to the model maps, including strong cross-shore wind-speed gradients, strong mean wind speeds over the Gulf of Maine, and the slowing of the flow associated with Cape Cod. Peak speeds over the Gulf of Maine are much stronger in the measurements, and the Cape Cod area speeds are much slower, indicating that the similarities are qualitative and not quantitative. It is reasonable to hypothesize that a reason for the quantitative discrepancies is that the measurements were taken over a much shorter time period and that the study region was not uniformly sampled, because of the relatively slow ship speeds. Although the hypothesis is undoubtedly true, at this time the only map of the horizontal distribution of winds measured at hub height over the ocean is the one presented here (Figs. 3.13 and C-5). Proving the hypothesis will require more such measurements over a much longer time period.

Measurements of the horizontal distribution of offshore hub-height winds are important because, **if it can be verified that the models reproduce well all aspects of the rotor-layer flow important to WE, then models can be run for resource assessment much less expensively than offshore measurement deployments that would be required**. Equally as important, **if models do *not* accurately reproduce the important aspects of the winds, measurement deployments become even more critical to document and understand offshore wind patterns and behavior**. Without question, some long-term offshore wind observations will be necessary to corroborate model resource assessments. In turn, the model assessments themselves will improve in accuracy. Both will be needed.

Although expensive, such field measurement campaigns can help to avoid even more costly misjudgments in siting or other decisions, such as appropriate hardware to deploy. In fact, even if model generated wind fields proved not to be reliable enough to make final siting decisions, such simulated fields would likely be used as a first guess as to promising sites, with measurements or measurement-based information to finalize the decisions. Besides providing the information itself, a goal of such measurement-model interactions is to diagnose what factors limit model accuracy so they can be addressed, so that overall the process becomes less reliant on large-scale instrument deployments.

Thus, several important questions require measurement campaigns for answers. What are the important scales of horizontal variability of rotor-level winds in the cross-shore and along-shore directions, is one. Others include, what is the nature of extreme events, of ramp events, and of shear, stability, and turbulence encountered by the turbines? What is the nature of meteorological phenomena that may affect wind-speed statistics seasonally or annually? For example, LLJs over the ocean occurring mostly during summer months (Freedman, p.c.), which were also noted during NEAQS-04, may vary strongly in strength in the along-shore direction (Colle and Novak 2010). Diurnal sea-breeze circulations that may form an important contribution to the offshore resource, are driven by the large diurnal changes in surface heating inland. Along-

coastal variation in coastal topography, roughness (including vegetation type), soil moisture, soil type, or land use may generate along-shore variability in the wind resource through their modification of surface heat flux over land, with a resultant change in the local intensity of the sea-breeze circulation. **Recurrence of localized phenomena** and effects such as these and others **may occur often enough to be part of the long-term average wind resource at offshore locations**. Some such local effects have been pointed out in the 2-yr HRRR-based wind composite, such as the area of weak winds off the extreme eastern coast of Maine, apparent sheltered areas from Cape Cod, Nantucket, and Martha's Vineyard, and diurnal and seasonal variations in many locations (including Long Island Sound and Buzzards Bay).

The requirement for winds aloft at least within the rotor layer, and hopefully above, dictates the use of profiling devices or other means of getting sensors into this layer, such as aircraft, or very tall towers, which in general are very expensive and cannot be moved once installed. **The need to document the horizontal variability of flows points to a need to deploy profiling devices to fixed locations in arrays**, as with buoy-mounted Doppler lidars. Flow features seen in the HRDL measurements indicate that the profiling devices should have a vertical resolution of 10 m through the rotor layer to document the vertical structure, and at time intervals of 15 min or less. These intervals would allow climatologies of quantities beyond wind speed to be compiled, such as shear magnitude, ramp properties, extreme events, LLJ properties, and others.

The issue of what horizontal density of profiling sites (as, e.g., in Figs.4.1 and 4.3) is necessary to properly measure horizontal scales of variability is difficult to address, especially given the lack of measurements over the ocean to base any estimates on. It will require one or more measurement campaigns to determine what these scales are, but if the sampling array is not of high enough density, important scales may be missed. A way to sample the spatial variability is to use mobile platforms with profiling capability, such as the *RHB* in this study, or instrumented aircraft (including UAVs). Mobile platforms have their own set of issues to address, because slow-moving platforms see a combination of spatial and temporal variability. So, for example, if wind speeds increase over an entire region in a 3-hr period, a ship-borne measurement could not discern by itself whether the changes were in space, due to ship movement, or in time. Aircraft measure one vertical level at a time and can only be out for limited time periods, making long-term data sets and statistics difficult to compile. **Studies of spatial variability will most likely require multiple approaches, e.g., a mixture of fixed-platform sensors and sensors on mobile platforms** should be considered. If the mobile platform were a ship, sailing past fixed-location profiling devices would aid in sorting out the temporal vs. spatial variability of the measured flows. **All-weather capability is needed for these observing systems since low-level fog, low cloudiness, and precipitation can be frequent occurrences** in many offshore regions, as recommended over land by Banta et al. (2013).

The recommendation for designing a program to understand offshore variability is therefore to involve different types of platform. **Fixed long-term arrays should consist of both cross-**

shore lines at intervals along the shoreline and, between the cross-shore sensors, along-shore lines of profiling devices at some appropriate distance from shore, such as 50 km. The cross-shore lines may extend out to 75-100 km and consist of 3-5 profiling systems. Mobile platforms having profiling capability, such as the *RHB*, should be used as much as they can, perhaps in IOPs, during the sampling campaign, which would hopefully be for more than one year. Then the along-shore arrays (as in Fig.4.3) may be deployable at site intervals of 50-100 km, the distance in between to be filled in with data from the moving platforms. Once the scales of variability are better understood, the array spacing and arrangement, which may vary from location to location, should be adjusted—which could be done during the campaign.

The fundamental question of what is the necessary density of sensors to address offshore WE needs is not one that can be *fully* answered using currently available resources, may take iterations of field campaigns, and may have different answers in different U.S. coastal waters. Thus, the results and recommendations we make regarding requirements for a profiling-sensor network must be considered in this light and preliminary, and require additional research. In the SOW for this project, we stated:

"Limited meteorological observations in the wind turbine layer have been collected during special field campaigns, and utilizing this data for evaluating wind resource models should be viewed as the first step in improving offshore wind resource characterization. The scope of this study (SOW) is smaller than what is needed to fully determine the requirements for an observation network to support offshore wind resource assessment. Collection of long-term data sets (multi-year) in multiple over-land coastal and offshore areas along the Atlantic coast and in the Great Lakes, possibly coupled with well-calibrated Observing System Simulation Experiments (OSSE's), are required to determine such requirements."

5. CONCLUSIONS and RECOMMENDATIONS

Results from this POWER study include:

- **An initial observed characterization of offshore wind behavior at turbine-level (based on lidar data: unprecedented)**
- **An initial error assessment of short-range weather model forecasts for offshore turbine-level winds, showing roughly similar skill to those for land-based turbine-level winds in WFIP**
- **More accurate offshore turbine-level wind forecasts, in general, by assimilation of land-based regional wind-profiling radar observations**
- **A 2-year 80-m wind-speed composite based on an hourly updated 3-km-grid model to recommend areas for long-term offshore wind observations**

Considerations:

- Any future observational study for offshore wind evaluation will benefit from recently developed improvements in data assimilation and modeling now used (as of March 2014) in real time with the RAP and HRRR and developmental NAMRR. Such a study will also benefit from improved observational datasets now available.
    - Introduction of new RAP/NAMRR data assimilation techniques (Hybrid-Ensemble-Variational) now better uses the wind profiler data according to flow dependent meteorological conditions.
    - Assimilation of more recently available offshore data, such as satellite and scanning weather radar data now improve model skill offshore.
- A coastal observing system to support offshore wind energy will likely have two distinct components: one for resource assessment, with the primary goal to validate model climatologies; and a second for wind-energy forecasting that will provide real-time information to be assimilated into forecast models. The resource assessment observations should span at least a several year time frame, whereas the forecasting observations will be ongoing.

Recommendations:

- **A useful offshore wind-energy measurement network would consist of an appropriate mixture of cross-shore transects of buoy-mounted wind-profiling instruments and along-shore lines of these sensors**. A cross-shore transect should extend through the high wind-speed gradient zone in the first 75-100 km from shore, with sufficient resolution along the transect to resolve variations in the cross-shore gradient. Each transect would require 2-5 buoys, with closer spacing near the coast. Multiple transects up and down the coastline would validate the model climatology showing the presence of high, moderate, and low gradient regions, and also validate the large-scale, along-shore variability of the wind resource (e.g. from the northern Gulf of Maine to

Maryland). Along-shore lines of sensing arrays would characterize offshore wind-speed variations in the direction parallel to the coastline.
    - For offshore wind resource assessment, as well as overall characterization and understanding of offshore wind meteorology, several sensor-distribution options have been suggested in this report, and are summarized below in Figs.5.1 and 5.2.
    - The use of multiple profiling devices at each site should be considered, to assure all-weather data availability.
- **An interior augmentation of boundary-layer profilers is recommended** beyond an array similar to that deployed for NEAQS-04, as shown in Fig. 5.3, for improved 1-12-hr wind forecasts. This deployment would be supplemented by some of the offshore wind measurement systems indicated in Figs. 5.1.or 5.2.
- **Mobile sensing platforms such as the *RHB* or aircraft will be required** along with the long-term deployment of fixed-sensor (e.g., buoy lidar) sensor arrays to characterize along-shore variability and discover recurrent areas of stronger and weaker flows. Intensive Observational Periods (IOP's) are recommended to focus resources and understand meteorological processes driving the flow variability in the along- and cross-shore directions.
- **The new HRRR-based 80-m wind composite for the U.S. including offshore areas should be extended beyond the current 2-year period**. A proper offshore wind climatology or wind-resource chart will require more than 2-years' worth of archived model runs, since significant differences have been noted between two even consecutive years (2012-1013) or two different summers (2002 and 2004). Year-to-year variations in mean atmospheric long-wave patterns produce such variations. As improved boundary-layer physics and data assimilation are improved, retrospective re-determination of ~2h HRRR forecasts to cover prior periods should be considered.
- **The initial 2-year HRRR offshore wind composite can be applied to other geographical regions** since it is available over all coastal regions, including the Great Lakes, the southeastern U.S. Atlantic coast, and the Gulf of Mexico coastline, although it will need to be extended into additional years.
- **A future offshore study with latest observations but including offshore profiling measurements (such as from the *R/V RHB*) would be helpful to extend this study.**
- Buoy-based vertical wind-profiling data that have a spatial distribution sufficient to resolve the gross characteristics of offshore mesoscale vortices, such as observed during the July 2004 study period, should be considered. We note that today improved forecasts for such periods would be likely for similar weather scenarios, given 2014-era data assimilation systems and current 2014 observation distribution.
- Additional observations from the RFORE platform should be used in the future to confirm the effectiveness of model improvements.
- Advanced versions of hourly assimilated/output RAP/HRRR and NAM model, NAMRR, specifically upgraded for the POWER experiment (and used in this study) should be used in future offshore projects.

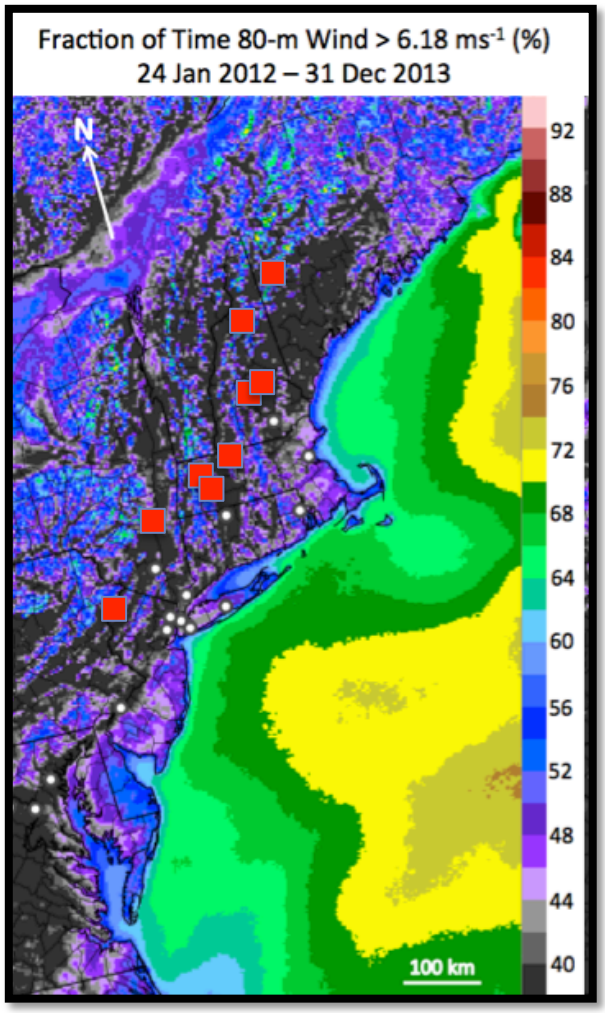


Figure 5.1: Possible 500-km-scale distribution for future offshore wind measurement systems (red squares), including 7 along-shore locations and having two locations with some cross-shore transects: potential locations for cross-shore sampling are indicated by double squares. In this study, NOAA found an important degree of large-scale, along-shore variability from Maine to Virginia within the new 2-yr composite map of 80-m wind speed from the 3-km-grid HRRR, and also variability in the cross-shore direction. Some combination of along-shore and cross-shore all-weather turbine-level wind observations is needed to make a long-term (multi-year) assessment to corroborate this initial model-based result. We assume that these observations should be approximately 10-30 km offshore, subject to further discussion with DOE.

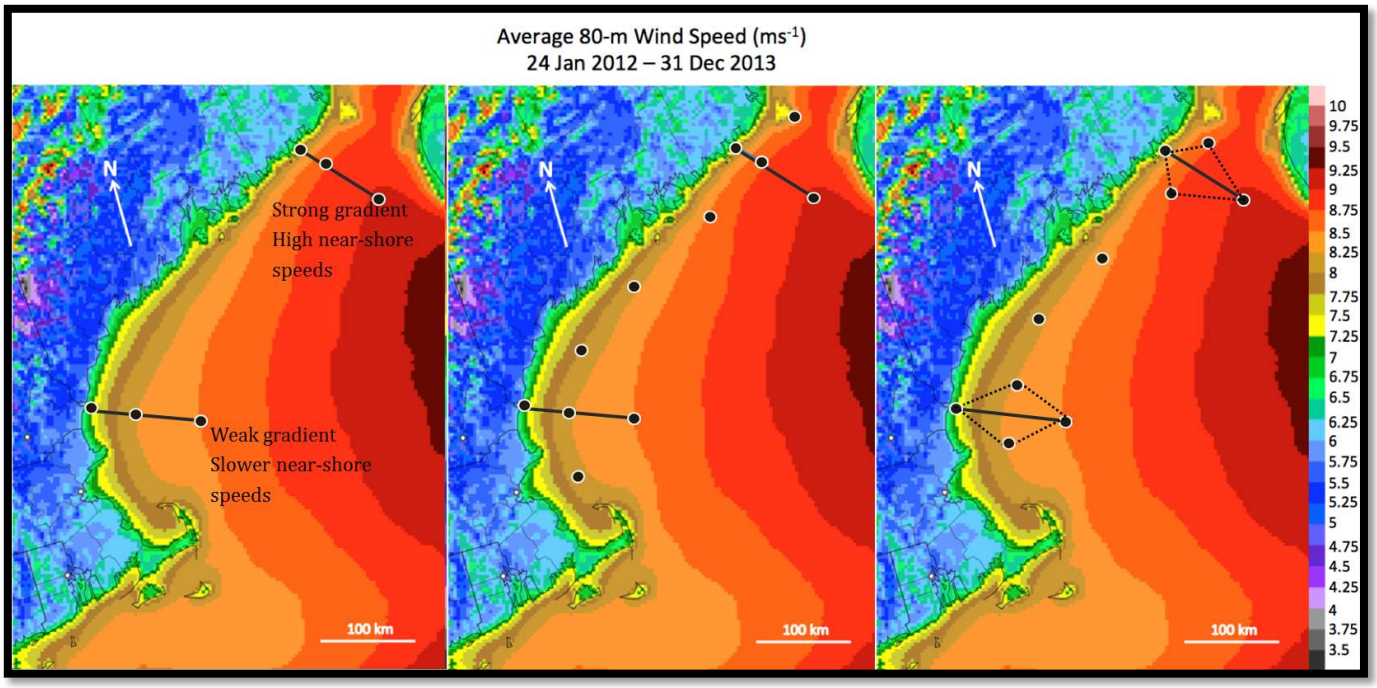


Figure 5.2. Three options for deployment of profiling sensors to characterize the wind resource for offshore wind energy.

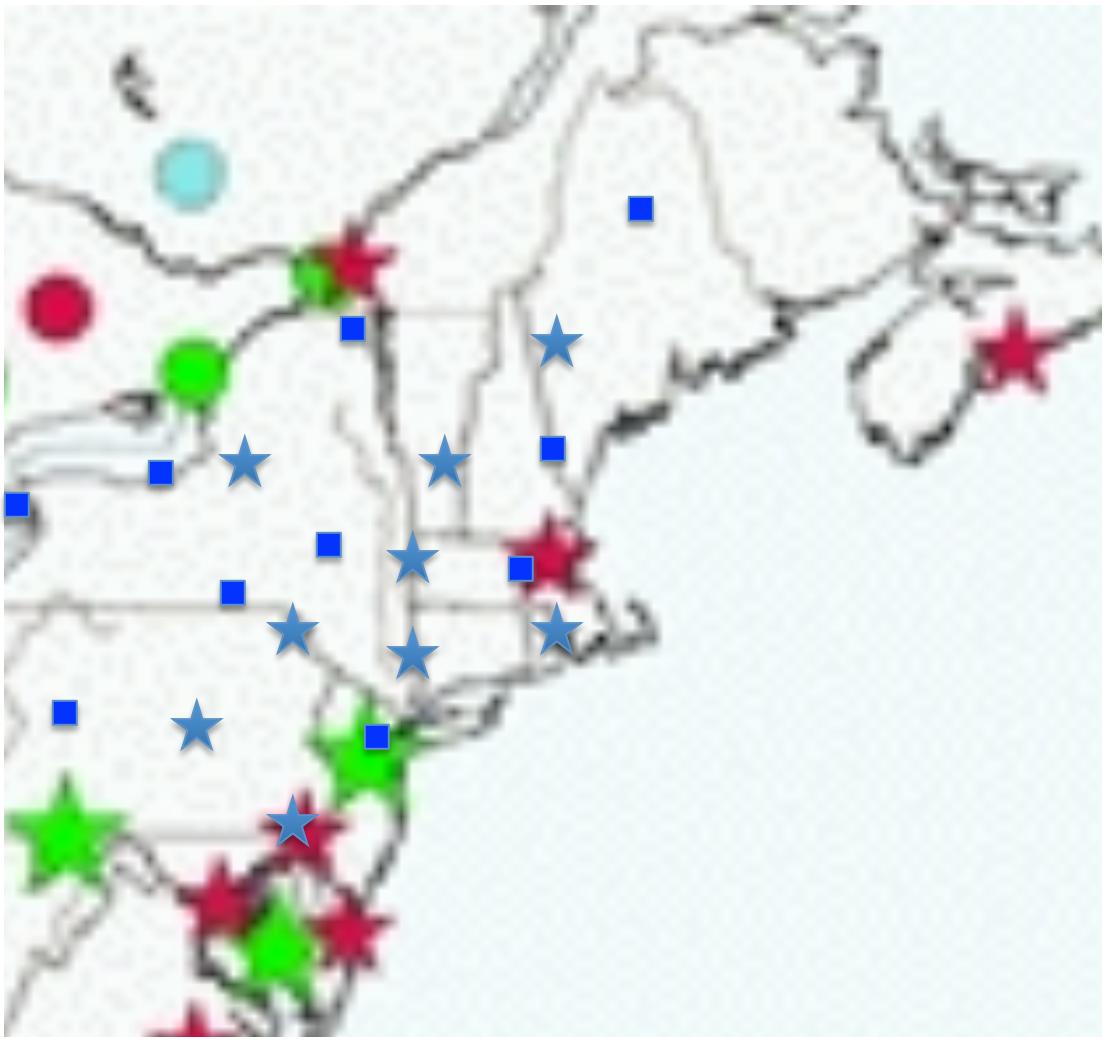


Figure 5.3. Possible distribution for future inland boundary-layer profiler wind measurement systems (blue stars) to improve accuracy of short-term wind forecasts including New England offshore possible locations. Other observing systems shown, including existing NWS WSR-88D radars producing frequent velocity-azimuth display (VAD) vertical wind profiles (blue squares), and existing other boundary-layer profilers (stars – green and red).

REFERENCES

Alpert, J., 2004: Sub-grid scale mountain blocking at NCEP. Proc. 20Th Conference on Weather on Analysis and Forecasting/17th Conference on Numerical Weather Prediction, American Meteorological Society. Seattle, WA. P2.4. [Available online at https://ams.confex.com/ams/pdfpapers/71011.pdf].

Angevine, W. M., J. E. Hare, C. W. Fairall, D. E. Wolfe, R. J. Hill, W. A. Brewer, and A. B. White, 2006: Structure and formation of the highly stable marine boundary layer over the Gulf of Maine, *J. Geophys. Res*., 111, D23S22, doi:10.1029/2006JD007465.

Banta, R.M., C.M. Shun, D.C. Law, W.O.J. Brown, R.F. Reinking, R.M. Hardesty, C.J. Senff, W.A. Brewer, M.J. Post, L.S. Darby, 2012: Observational techniques: Sampling the mountain atmosphere. Ch. 8 in ***Mountain Weather Research and Forecasting***, Eds., *F.K. Chow, S. De Wekker, and B. Snyder*, Springer, Dordrecht, the Netherlands, 409-530.

Banta, R.M., Y.L. Pichugina, N.D. Kelley, W.A. Brewer, and R.M. Hardesty, 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, S. G., and Coauthors, 2004a: An Hourly Assimilation‐Forecast Cycle: The RUC. *Mon. Wea. Rev.*, 132, 495‐518.

Benjamin, S. G., S.S. Weygandt, D. Devenyi, J.M. Brown, G.A. Manikin, T.L. Smith, and T.G. Smirnova, 2004b: Improved moisture and PBL initialization in the RUC using METAR data. Extended Abstracts. 22nd Conference on Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc. 17.3.

Benjamin, S.G., B.D. Jamison, W.R. Moninger, S.R. Sahm, B. Schwartz, T.W. Schlatter, 2010: Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR, and mesonet observations via the RUC Hourly Assimilation Cycle. *Mon. Wea. Rev.*,138, 1319-1343. PDF

Carter, D. A., K.. S. Gage, W.L.F.cklund, W.M. Angevine, P.E. Johnston, A. C. Riddle, J. Wilson, and C.R.Williams (1995), Developments in UHF lower tropospheric wind profiling at NOAA's Aeronomy Laboratory, *Radio Sci.*, 30, 997-1001.

Darby, L.S., S.A. McKeen, C.J. Senff, A.B. White, R.M. Banta, M.J. Post, W.A. Brewer, R.D. Marchbanks, R.J. Alvarez II, S.E. Peckham, H. Mao, and R. Talbot, 2007: Ozone differences between near-coastal and offshore sites in New England: Role of meteorology. *J. Geophys. Res*., 112, D16S91, doi:10.1029/2007JD008446.

Daley, R., 1991: Atmospheric Data Analysis. Cambridge University Press, 457 pp.

Desroziers, G. and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation error parameters in a variational assimilation. Quart. J. Roy. Meteor. Soc., 127, 1433-1452.

Drechsel, S., G. J. Mayr, J. W. Messner, R. Stauffer, 2012: Wind Speeds at Heights Crucial for Wind Energy: Measurements and Verification of Forecasts. J. Appl. Meteor. Climatol., 51, 1602-1617.

Drechsel, S., G.J. Mayr, J.W. Messner, and R. Stauffer, 2012: Lower boundary wind speeds: Measurements and verification of forecasts. J. Appl. Meteor. Climatol., submitted.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. J. Geophys. Res., 108 (D22), 16, doi:10.1029/2002JD003296.

Fairall, C. W., Ludovic Bariteau, A.A. Grachev, R. J. Hill, D.E. Wolfe, W. Brewer, S. Tucker, J. E. Hare, and W. Angevine, 2006: Coastal effects on turbulent bulk transfer coefficients and ozone deposition velocity in ICARTT. J. Geophys. Res., 111, D23S20, doi:1029/2006JD007597.

Fehsenfeld, F.C., et al., 2006: International Consortium for Atmospheric Research on Transport and Transformation (ICARTT): North America to Europe—Overview of the 2004 summer field study. *J. Geophys. Res.*, 111, D23S01, doi: 10.1029/2006JD007829.

Ferrier, B. S., Y. Jin, Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and precipitation scheme in the NCEP Eta model. Preprints, 15th Conf. on Numerical Weather Prediction, San Antonio, TX, Amer. Meteor. Soc., 280-283.

Ferrier, B. S., W. Wang, and E. Colon, 2011: Evaluating cloud microphysics schemes in nested NMMB forecasts. 24th Conf. on Weather Analysis and Forecasting/20th Conf. on Numerical Weather Prediction, Seattle, WA, Amer. Meteor. Soc.

Grund, C.J., R.M. Banta, J.L. George, J.N. Howell, M.J. Post, R.A. Richter, A.M. Weickmann, 2001: High-resolution Doppler lidar for boundary layer and cloud research. J. Atmos. Oceanic Tech., 18, 376-393.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. Wea. Forecasting, 14, 155-167.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. Mon. Wea. Rev., 124, 1225-1242.

Holm, E., E. Andersson, A. Beljaars, P. Lopez, J.-F. Mahfouf, A. Simmons, and J.-N. Thepaut, 2002: Assimilation and modelling of the hydrologic cycle: ECMWF's status and plans. ECMWF Tech. Memo., 55 pp. [383].

Huang, X.-Y. and P. Lynch, 1993: Diabatic Digital-Filtering Initialization: Application to the HIRLAM Model. Mon. Wea. Rev., 121, 589-603.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. J. Geophys. Res., 113, D13103, doi:10.1029/2008JD009944.

Janjic, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon. Wea. Rev., 122, 927-945.

Janjic, Z. I., 2001: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP meso model. NCEP Office Note, 91 pp. [437].

Janjic, Z. I., 2003: A nonhydrostatic model based on a new approach. Meteorology and Atmospheric Physics, 82, 271-285.

Janjic, Z. I., 2005: A unified model approach from meso to global scales. Geophysical Research Abstracts, General Assembly, Vienna, Austria, European Geosciences Union, 7, 05 582, SRef-ID: 1607-7962/gra/EGU05-A-05 582.

Janjic, Z. I. and T. L. Black, 2007: An ESMF unified model for a broad range of spatial and temporal scales. Geophysical Research Abstracts, General Assembly, Vienna, Austria, European Geosciences Union, 9, 05 025, 2007, SRef-ID: 1607-7962/gra/EGU2007-A-05 025.

Janjic, Z. I. and R. Gall, 2012: Scientific documentation of the NCEP nonhydrostatic multiscale model on the B grid (NMMB). Part I Dynamics. NCAR Tech. Note NCAR/TN- 489 +STR. 74 pp.

Kain, J. S., et al., 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. Wea. Forecasting, 23, 931-952.

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

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. Quart. J. Roy. Meteor. Soc., 112, 1177-1194.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP−a comparison with 4D-Var. Quart. J. Roy. Meteor. Soc., 129, 3183-3203.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. J. Atmos. Sci., 20, 130-141.

Lynch, P., D. Giard, and V. Ivanovici, 1997: Improving the Efficiency of a Digital Filtering Scheme for Diabatic Initialization. Mon. Wea. Rev., 125, 1976-1982.

Mittermaier, M. P., 2013: A strategy for verifying near-convection-resolving model forecasts at observing sites. Wea. Forecasting, In press.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res., 102, 16 663-16 682.

Morss, R. E. and K. A. Emanuel, 2002: Influence of added observations on analysis and forecast errors: Results from idealized systems. Quart. J. Roy. Meteor. Soc., 128, 285-321.

Musial, W., and S. Butterfield, 2004: Future for offshore wind energy in the United States, Preprint for Energy Ocean 2004, Palm Beach, Florida, June 28.29, 2004.

_____, and B. Ram, 2010: Large-scale offshore wind power in the United States: Assessment of opportunities and barriers. NREL/TP-500-40745.

Novak, D., C. Bailey, K. Brill, P. Burke, W. Hogsett, R. Rausch, and M. Schichtel, 2013: Precipitation and temperature forecast performance at the Weather Prediction Center. Wea. Forecasting. In press.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. Mon. Wea. Rev., 120, 1747-1763.

Parrish, D. F., J. C. Derber, R. J. Purser, W.-S. Wu, and Z.-X. Pu, 1997: The NCEP global analysis system: Recent improvements and future plans. J. Meteor. Soc. Japan, 75, 359-365.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. Mon. Wea. Rev., 131, 1524-1535.

Rife, D. L., C. A. Davis, Y. Liu, and T. T. Warner, 2004: Predictability of Low-Level Winds by Mesoscale Meteorological Models. Mon. Wea. Rev., 132, 2553-2569.

Rogers, E., and Coauthors, 2009: The NCEP North American Mesoscale modeling system: Recent changes and future plans. Preprints, 23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction, Omaha, NE, Amer. Meteor. Soc., 2A.4. [Available online at http://ams.confex.com/ams/pdfpapers/154114.pdf].

Rogers, E., B. Ferrier, Z. Janjic, W. S. Wu, and G. DiMego, 2014: The NCEP North American Mesoscale (NAM) Analysis and Forecast System : Near-term plans and future evolution into a high-resolution ensemble. *26th Conf. on Weather Analysis and Forecasting/22nd Conf. on Numerical Weather Prediction,* Atlanta, GA, Amer. Meteor. Soc., J1.3.

Rothermel, J., D.R.Cutten, R.M. Hardesty, R.T.Menzies, J.N. Howell, S.C. Johnson, D.M. Tratt, L.D. Olivier, and R.M. Banta, 1998: The multi-center airborne coherent atmospheric wind sensor, MACAWS. *Bull. Amer. Meteor Soc.*, **79**, 581-599.

Saha, S., and Coauthors, 2010: The NCEP climate forecast system reanalysis. Bull. Amer. Meteor. Soc., 91, 1015‑1057.

Schwartz, M., D. Heimiller, S. Haymes, and W. Musial, 2010: Assessment of offshore wind energy resources for the United States. NREL/TP-500-45889. National Renewable Energy Laboratory, Golden CO.

Semple, A., M. Thurlow, S. Milton, 2012: Experimental determination of forecast sensitivity and the degradation of forecasts through the assimilation of good quality data. Mon. Wea. Rev., 140, 2253‑2269.

Sušelj, Kay, and Abha Sood, 2010: Improving the Mellor–Yamada–Janjic parameterization for wind conditions in the marine planetary boundary layer. *Bound.-Layer Meteorol.*, **136**, 301–324, DOI 10.1007/s10546-010-9502-3

Strauch, R.G., D.A. Merritt, K.P. Moran, K.B. Earnshaw, and D. Van De Kamp, 1984: The Colorado Wind-Profiling Network, J. Atmos. Ocean. Technol., 1, 37-49. 271 272

Thiébaux, J., E. Rogers, W. Wang, and B. Katz, 2003: A new high-resolution blended real-time global sea surface temperature analysis. Bull. Amer. Meteor. Soc., 84, 645‑656.

Weber, B.L., D.B. Wuertz, D.C. Welsh, R. McPeek, 1993: Quality controls for profiler measurements of winds and RASS temperatures. J. Atmos. Oceanic Technol., 10, 452‑464. doi: http://dx.doi.org/10.1175/1520-0426(1993)010<0452:QCFPMO>2.0.CO;2

White, A.B., L.S. Darby, C.J. Senff, C.W. King, R.M. Banta, J. Koermer, J.M. Wilczak, P.J., Neiman, W.M. Angevine, and R. Talbot, 2007: Comparing the impact of meteorological variability on surface ozone during the NEAQS (2002) and ICARTT (2004) field campaigns. J. Geophys. Res., 112, D10S14, doi:10.1029/2006JD007590.

Wilczak, J.M., E.E. Gossard, W.D. Neff, and W.L. Eberhard, 1996: Ground-based remote sensing of the atmospheric boundary layer: 25 years of progress, Boundary-Layer Meteor., 78, 321-349.

Wilczak, J. M., et al., 2013: Public/Private Partnership for Improving Short Term Wind Energy Forecasts and Quantifying the Benefits of Utility Operations. , 146 pp, submitted to DOE.

Wolfe, D.E. and Coathors, 2007: Shipboard multisensory merged wind profiles from the New England Air Quality Study 2004. J. Geophys. Res., 112 , DOI: 10.1029/2006JD007344

Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variationalanalysis with spatially inhomogeneous covariances. Mon. Wea. Rev., 130, 2905–2916.

Zhu, Y., J. Derber, A. Collard, D. Dee, R. Treadon, G. Gayno, and J. A. Jung, 2013: Enhanced radiance bias correction in the National Centers for Environmental Prediction's Gridpoint Statistical Interpolation data assimilation system. Q.J.R. Meteorol. Soc. In press. doi: 10.1002/qj.2233

Appendix A: Gridpoint Statistical Interpolation (GSI) System and Data Assimilation

The implementation of the Gridpoint Statistical Interpolation (GSI) system for the POWER project was under the context of 3-dimensional variational assimilation (3DVar), which minimizes a cost function that measures the distance to the background forecast and observations (Kalnay, 2003). The model analysis is then globally adjusted to all the observations available during the assimilation period (Talagrand, 1997). For GSI 3DVar, the following incremental cost function is minimized

$$J = \frac{1}{2}[\mathbf{x}^{T}\mathbf{B}^{-1}\mathbf{x} + (\mathbf{Hx} - \mathbf{y})^{T}\mathbf{R}^{-1}(\mathbf{Hx} - \mathbf{y})]$$

Where $\mathbf{x}$ is a column vector of analysis increments, $\mathbf{x} = \mathbf{x}^{a} - \mathbf{x}^{f}$. Here superscripts 'a' and 'f' denote an analysis and forecast, respectively. The vector$\mathbf{x}$, and its constituents, has length $n$which corresponds to the total number of model gridpoints times the number of analysis variables (e.g. streamfunction, temperature, etc.). Matrix $\mathbf{B}$ is the background error covariance matrix and is of dimension n x n. $\mathbf{H}$ is the, possibly nonlinear, observation operator which maps forecast variables to observations and has dimension p x n, where p corresponds to the total number of observations to be assimilated. Finally, $\mathbf{y}$ is a column vector of observation innovations and takes the form of $\mathbf{y} = \mathbf{y}_{obs} - \mathbf{Hx}^{f}$ and has length p, where subscript 'obs' denotes the actual observations (e.g. nacelle wind speeds). To find the analysis increment which minimizes the cost function the iterative preconditioned conjugate gradient algorithm of Derber and Rosati (1989) is used.

In the practical implementation of 3DVar the background error covariance matrix, $\mathbf{B}$, must estimated *a priori.* The structure of this matrix is quite important, as it largely determines how the information from the assimilated observations is spread out into the analysis, i.e. the analysis increments (Daley, 1991; Kalnay, 2003). In the GSI $\mathbf{B}$ may be estimated using the so-called NMC method (Parrish and Derber, 1992) or the ensemble-analysis method (Houtekamer et al. 1996). Both techniques follow the methodology of using the average of many sets of forecast differences, verifying at the same time, to estimate $\mathbf{B}$. Along with the estimation of $\mathbf{B}$, the GSI includes additional steps to model the cross-variable covariances using linear balance relationships via statistical regression, which allows for a coupling of mass and wind fields in the resulting analysis (Parrish et al., 1997; Wu et al., 2002). Finally, the spatially univariate correlations are modeled using an isotropic recursive filter, which has the response of a Gaussian function and spreads the analysis increments to nearby grid points (Purser et al., 2003).

The observation error covariance matrix, $\mathbf{R}$, contains instrumentation errors, representativeness errors, and errors associated with the observation operator, $\mathbf{H}$. In practice it is assumed that all observation errors are independent and uncorrelated (Lorenc, 1986; Kalnay, 2003), thus rendering $\mathbf{R}$ a diagonal matrix of error variances. In GSI observation errors for conventional

observations, e.g. nacelle winds and radiosondes, are specified through an external error table file which stores the observation error standard deviations as a function of vertical pressure level. Observations also have a certain amount of self-descriptive meta-data associated with them which can yield additional information about observation quality. This extra data can be used to quality control observations during the analysis process by adaptively inflating observation error and/or rejecting observations completely. For example, observations may be rejected through a gross error test, which checks the observation against the background forecast, if the magnitude of the difference is too large, the observation may be rejected.

Appendix B:

We performed some additional analysis of the HRRR climatology, stratifying the results by season and by time of day. In this Appendix we present some daytime (15-00 UTC) and nighttime (03-12 UTC) plots from the summers (June – August) of 2012 and 2013. Overall, the summer of 2013 was windier than the summer of 2012. Maps of the average wind speed for these two years (Figure B-1) show that in both years the highest wind speeds within a few dozen kilometers of the coast occur just east of the southeastern tip of Cape Cod.

Some interesting differences appear when comparing the daytime and the nighttime plots. During the daytime, plots of the 6 m s^{-1} threshold show a distinct low-wind-speed region near and north of Cape Cod. In the summer of 2012, the daytime frequency of exceedances of the 6 m s^{-1} threshold is as low as 5-10% in this region (Fig B-2a,b), compared to 30% - 45% during the nighttime (Figure B-3). Another interesting feature is a daytime maximum in exceedance of both the 6- and 10-m s^{-1} (not shown) thresholds within and just to the southwest of Buzzards Bay off the southern coast of Massachusetts. This feature shows up in both 2012 and 2013.

Looking now at the average 80-m wind speeds during daytime and nighttime of the two years, we see quite similar patterns from year to year. During the day, winds are relatively slow in a broad region to the north and northeast of Cape Cod (Figs. B-4 and B-5); we hypothesize that this region of lower wind speeds is due to the sheltering effect of eastern Massachusetts and Rhode Island against the prevailing southwesterly flow. During the daytime, we also see a localized wind speed maximum just to the southwest of Buzzards Bay in both years. This wind speed maximum has speeds from 0.5 to 1.0 m s^{-1} stronger than some of the surrounding coastal areas. Both of these interesting features are only apparent during the daytime hours.

The cost of installing offshore turbines is strongly dependent upon the depth of the ocean floor. To this end, Figure B-6 shows the location of a couple of important bathymetry contours relative to the coast of New England, as well as the example wind energy regions mentioned in this report. 30-m ocean depth is an approximate break point between inexpensive and moderately expensive anchoring systems, and 80-m ocean depth is an approximate upper depth threshold for modern anchoring systems. We can see from this map that some of the southern regions are located in shallower seas, particularly the Atlantic City region. All of the sites except for the Monhegan Island/Boothbay Harbor site are located in waters less than 80 m deep. Bathymetry data will play an important role in the decision making process for placing offshore wind farms.

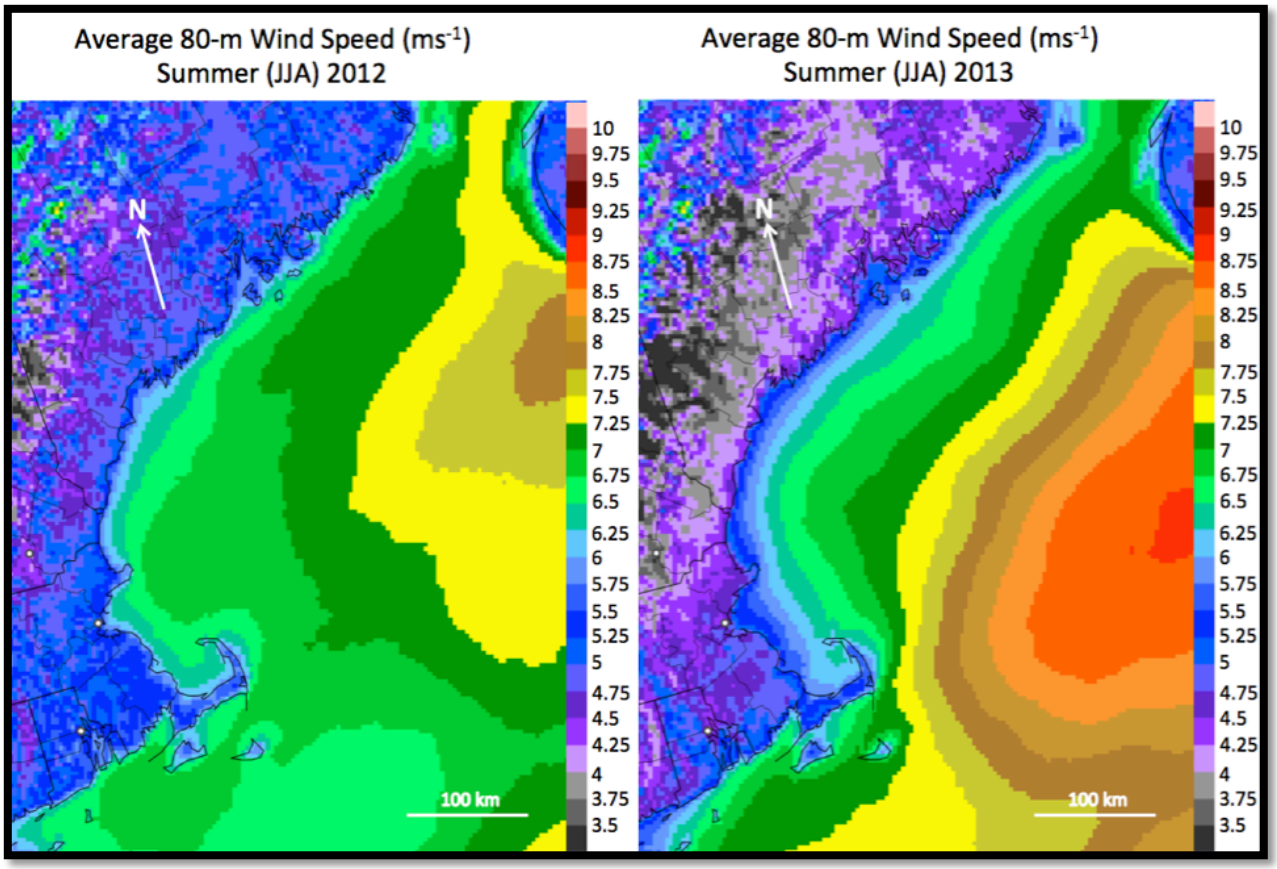


Figure B-1: Average 80-m wind speed (m s^{-1}) during (left) June-August 2012 and (right) June-August 2013 according to 2-h HRRR forecasts.

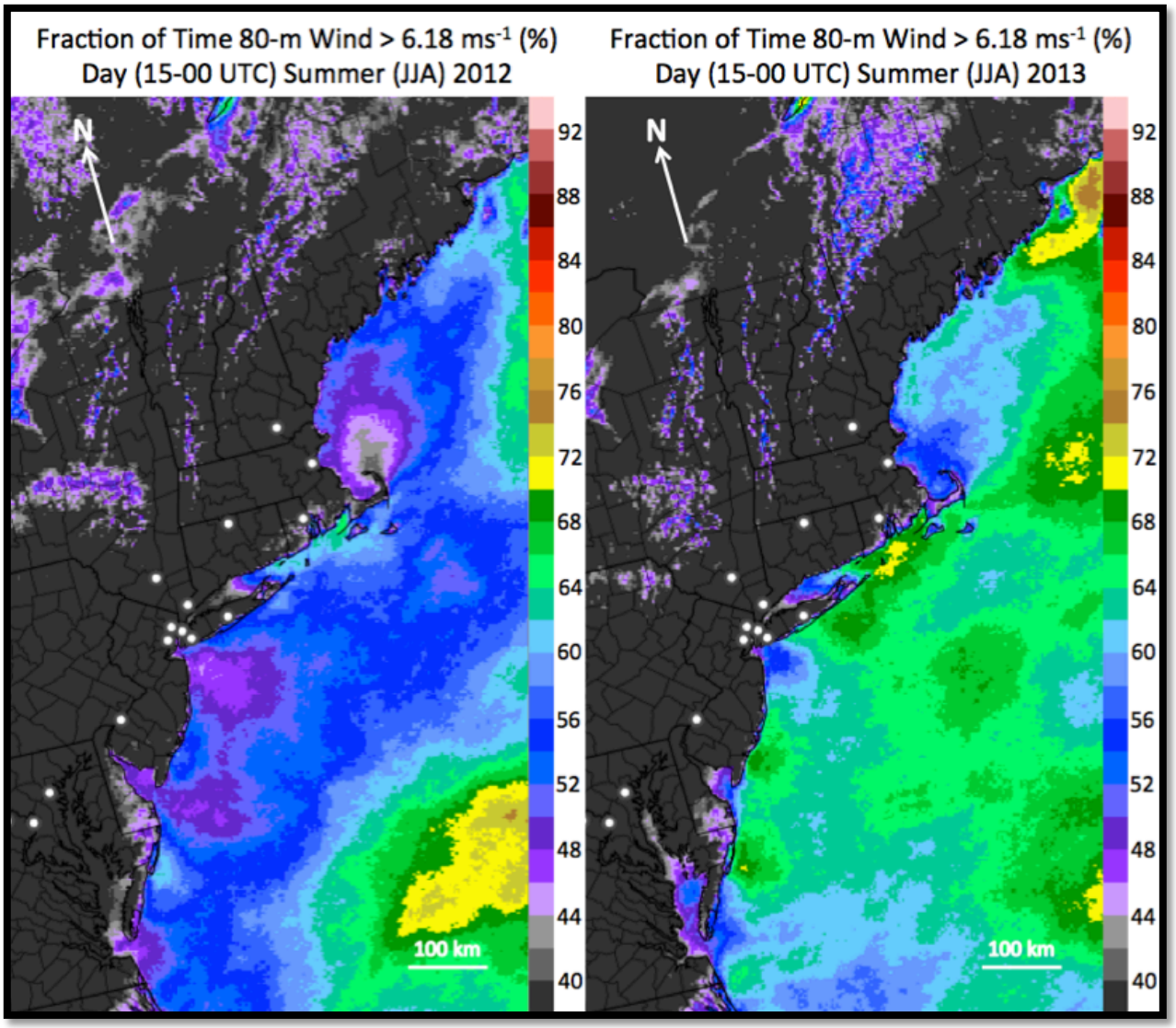


Figure B-2a: Percentage of time that 80-m wind exceeds 6.18 m s^{-1} (12 kts) during the daytime (15-00 UTC) of (left) June-August 2012 and (right) June-August 2013 according to 2-h HRRR forecasts.

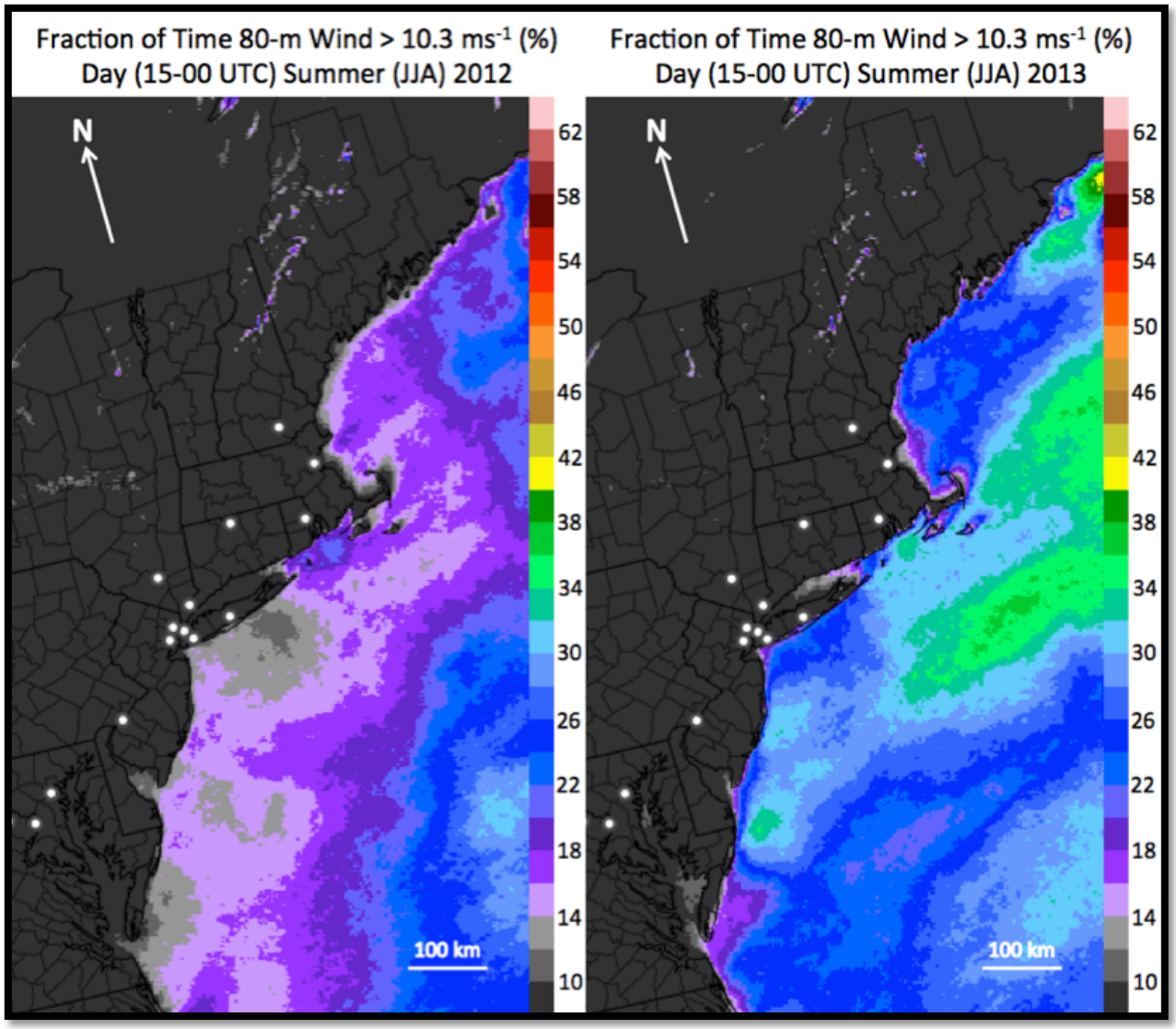


Figure B-2b: Percentage of time that 80-m wind exceeds 10.3 m s^{-1} (20 kts) during the daytime (15-00 UTC) of (left) June-August 2012 and (right) June-August 2013 according to 2-h HRRR forecasts.

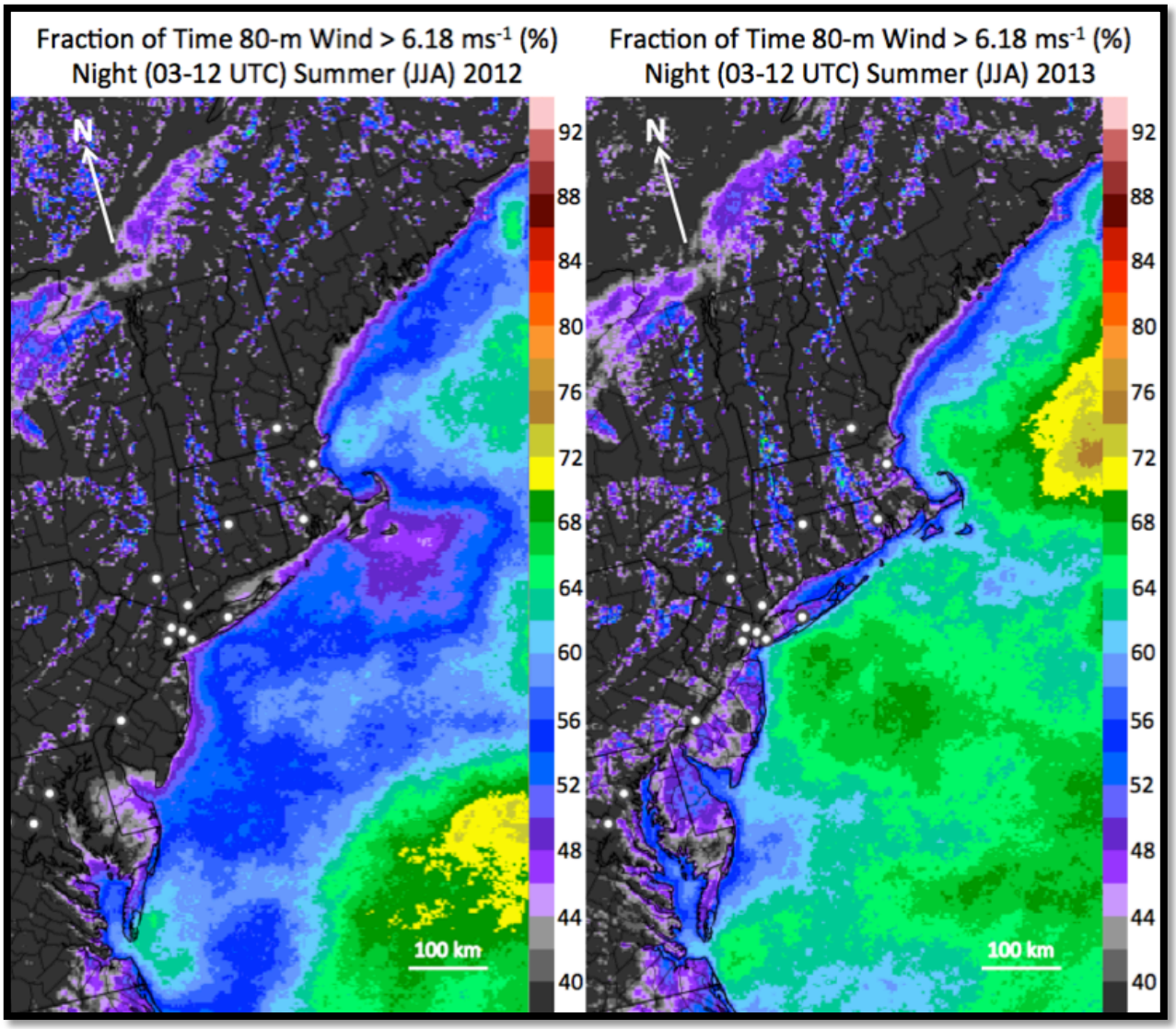


Figure B-3a: Percentage of time that 80-m wind exceeds 6.18 m s^{-1} (12 kts) during the nighttime (03-12 UTC) of (left) June-August 2012 and (right) June-August 2013 according to 2-h HRRR forecasts.

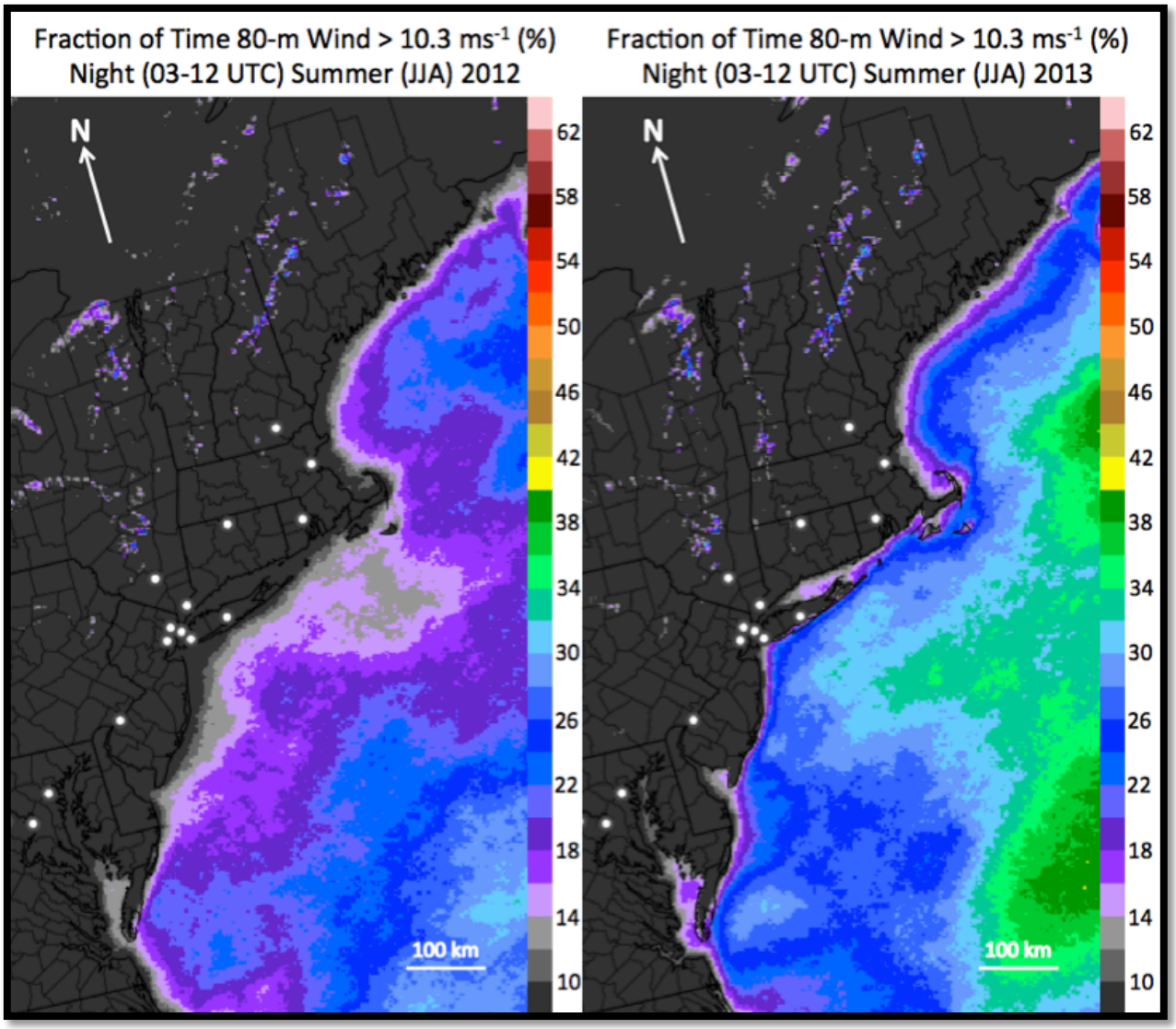


Figure B-3b: Percentage of time that 80-m wind exceeds 10.3 m s^{-1} (20 kts) during the nighttime (03-12 UTC) of (left) June-August 2012 and (right) June-August 2013 according to 2-h HRRR forecasts.

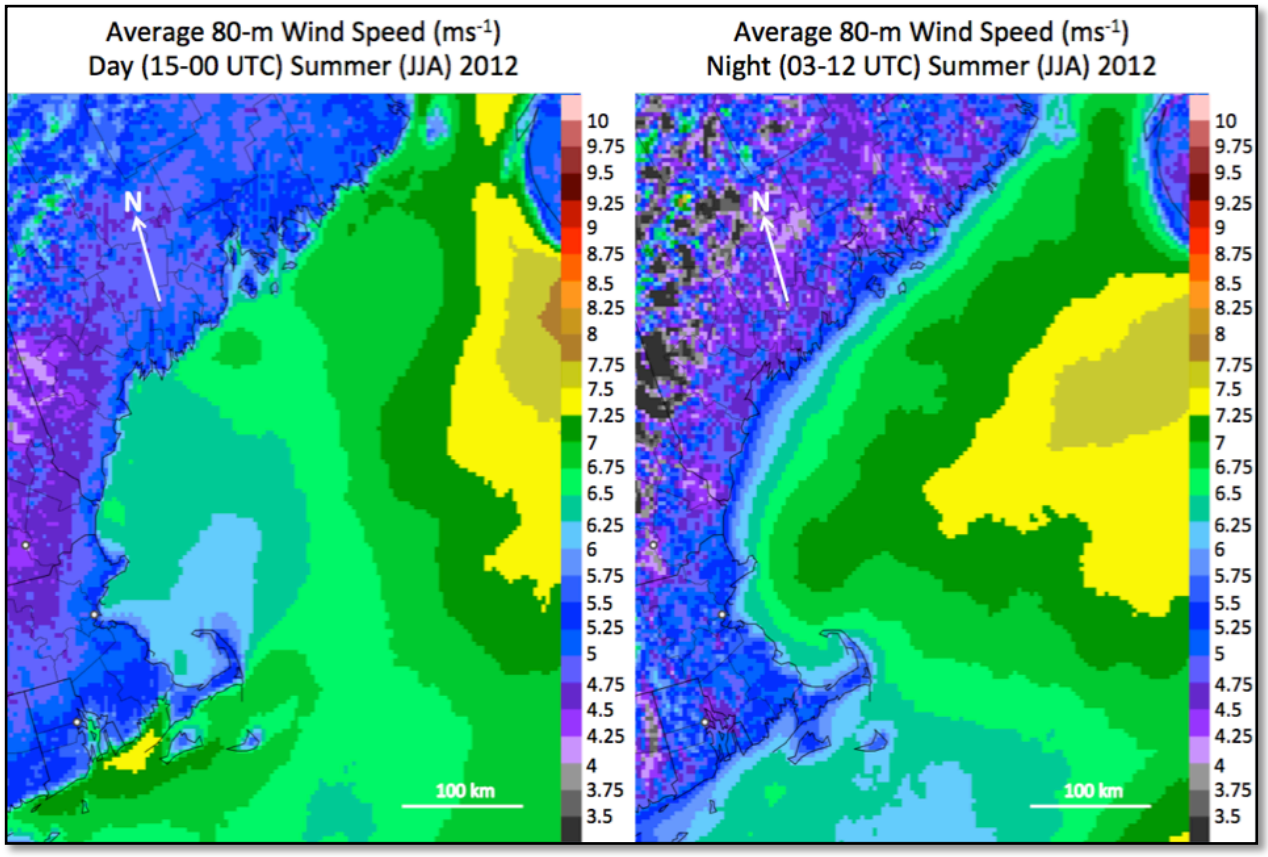


Figure B-4: Average 80-m wind speed (m s^{-1}) during the daytime (15-00 UTC: left) and nighttime (03-12 UTC: right) hours for June-August 2012 according to 2-h HRRR forecasts.

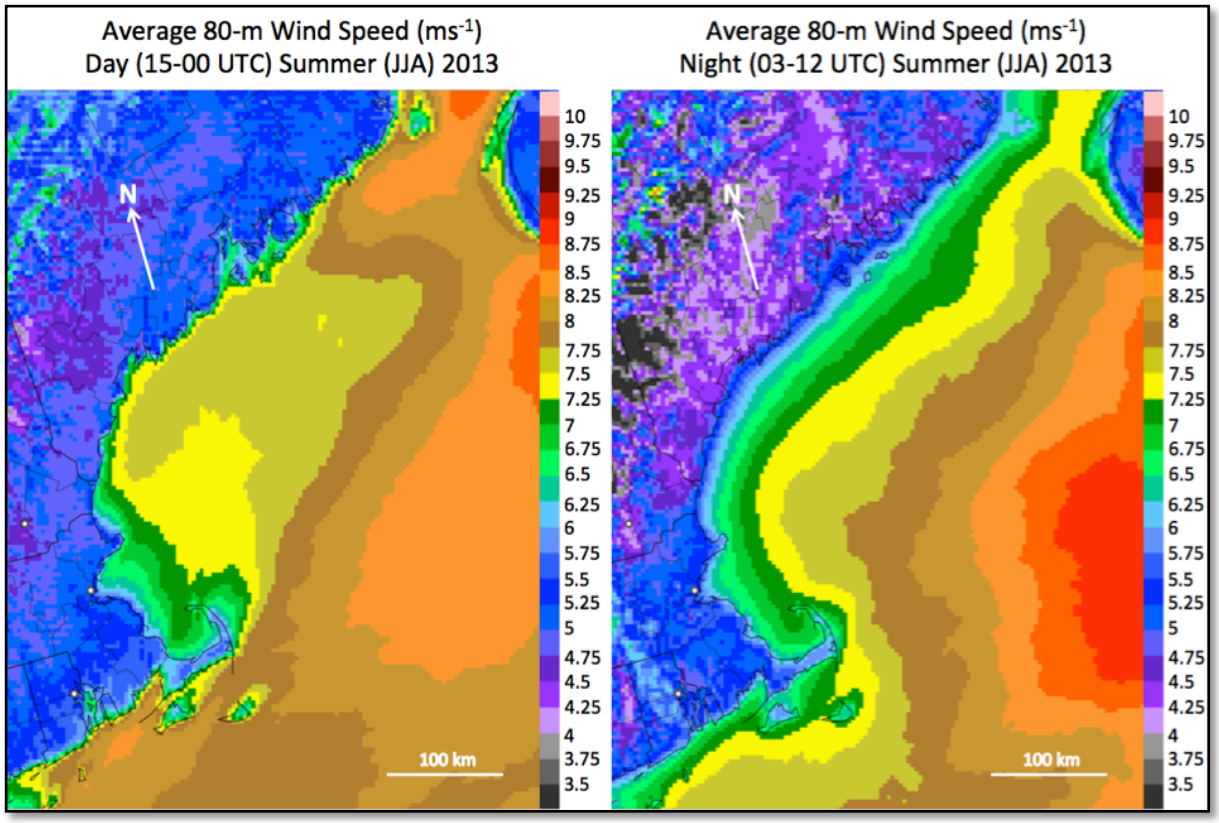


Figure B-5: Average 80-m wind speed (m s^{-1}) during the daytime (15-00 UTC: left) and nighttime (03-12 UTC: right) hours for June-August 2013 according to 2-h HRRR forecasts.

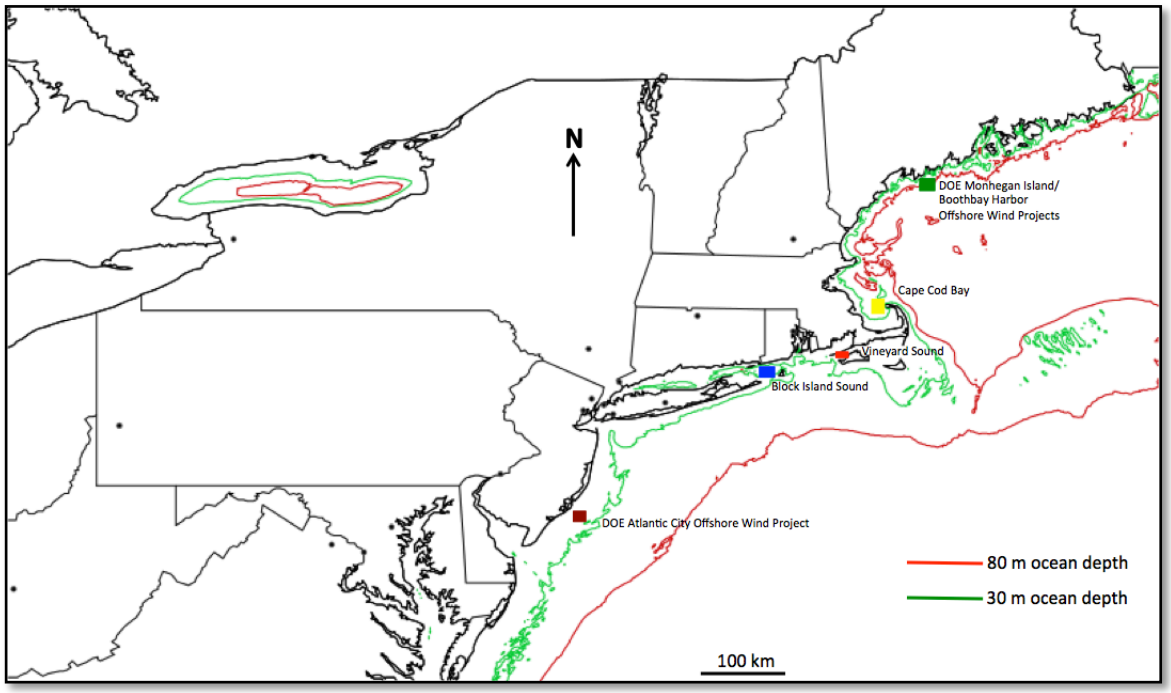


Figure B-6: Location of the 30-m and 80-m bathymetry contours relative to the coast of New England the locations of the example wind energy sites analyzed in the report.

Appendix C: Supplementary HRDL analyses

Lidar measurements

HRDL's high resolution and precision allow the lidar data to be analyzed for temporal variation in wind field and vertical differences (shear) across the blade-swept area. Examples of such analysis are shown in Figure C-1 for 16 July and 9 August, where time series show wind speed and wind direction at the top (145 m), hub (95 m) and bottom (45 m) height of a presumptive turbine rotor layer.

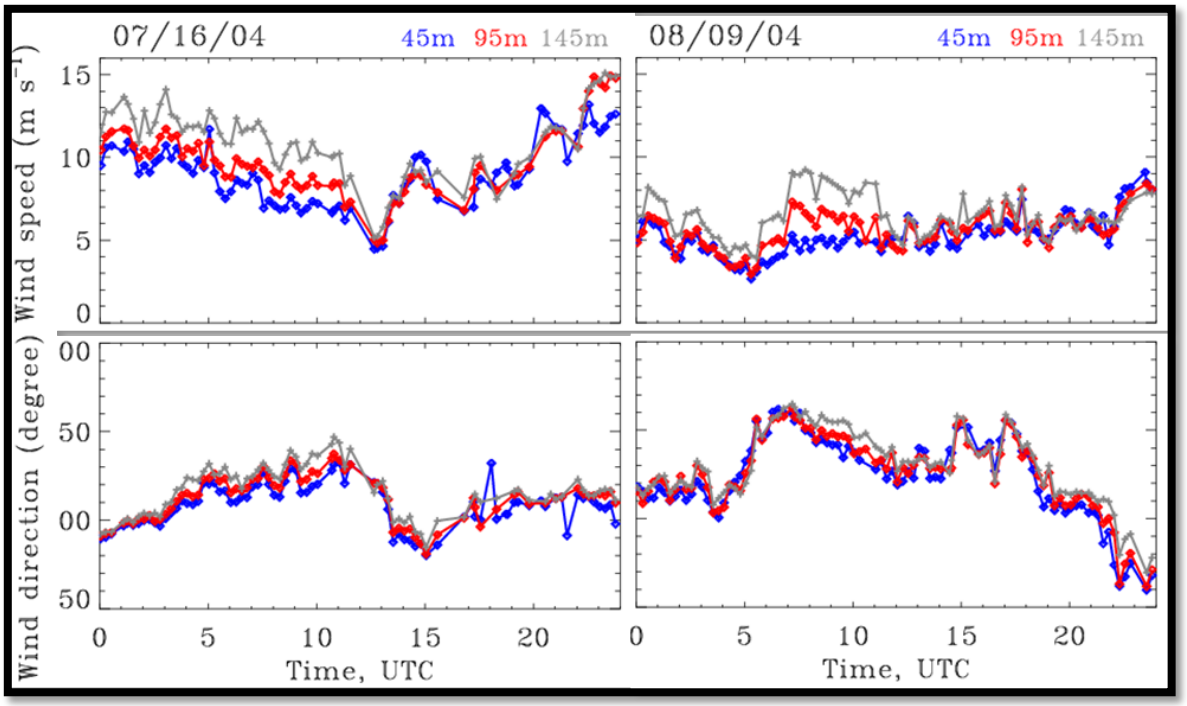


Figure C-1. Time series of wind speed (top panels, m s^{-1}) and wind direction (bottom panels, degrees) at several heights (as indicated at the top of the figure) within the turbine rotor-swept layer are shown for 16 July (left panels) and 9 August (right panels) 2004.

Ship tracks during both days are shown in Figure C-2. On 16 July nighttime lidar measurements were obtained at a fixed location, when RHB was stationed 4.5 mi off the coast east of (north of Cape Anne). During daytime hours the ship slowly moved farther to the ocean then turned to the north and continued to move towards southern Maine at a distance of ~30 km from the shore line. On this day (Figure C-1) winds at all heights show significant diurnal variation, decreasing to <5 m s^{-1} by 1300 UTC (just after sunrise) and increasing again to >15 m s^{-1}

during the daytime hours. Nighttime shears of 3-4 m s^{-1} can be seen between top and bottom heights of the rotor on 16 July. On 9 August measurements from 00 to 14 UTC were taken in Boston Harbor, then the ship sailed along the coast, turning towards the open ocean at 18 UTC. A narrower span of wind speeds from 3 to 10 m s^{-1} was observed for 9 August, but a strong shear of up to 4 m s^{-1} across the rotor could be seen at the midnight through early-morning transition hours. Figure C-1 shows that wind speeds in the rotor layer can vary over a wide range of values even on individual nights. As expected, shear magnitudes were larger on the stronger wind day. These examples demonstrate the ability of lidar measurements to provide an accurate estimate of wind properties across the rotor layer, and show that strong shear and rapid changes in wind speed and direction can occur.

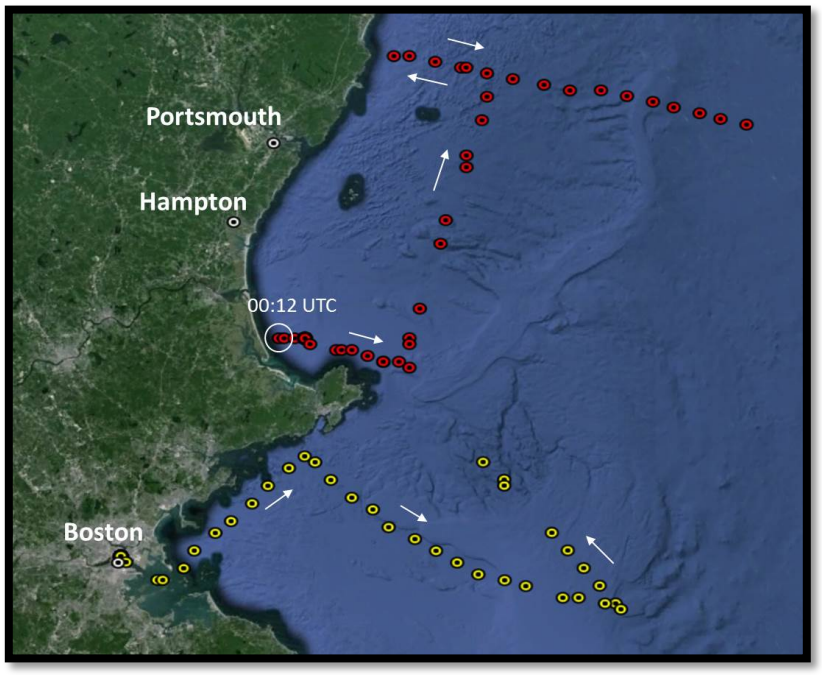


Figure C-2. Ship tracks during 16 July (red) and 9 August (yellow). Circles indicate points of lidar measurements averaged over ~15 min time interval.

Time-series of wind speed and direction for 7 days in each study period (Figure C-3) illustrate the forecasting challenge posed by strong variability in wind-flow conditions often seen from day to day and between selected periods.

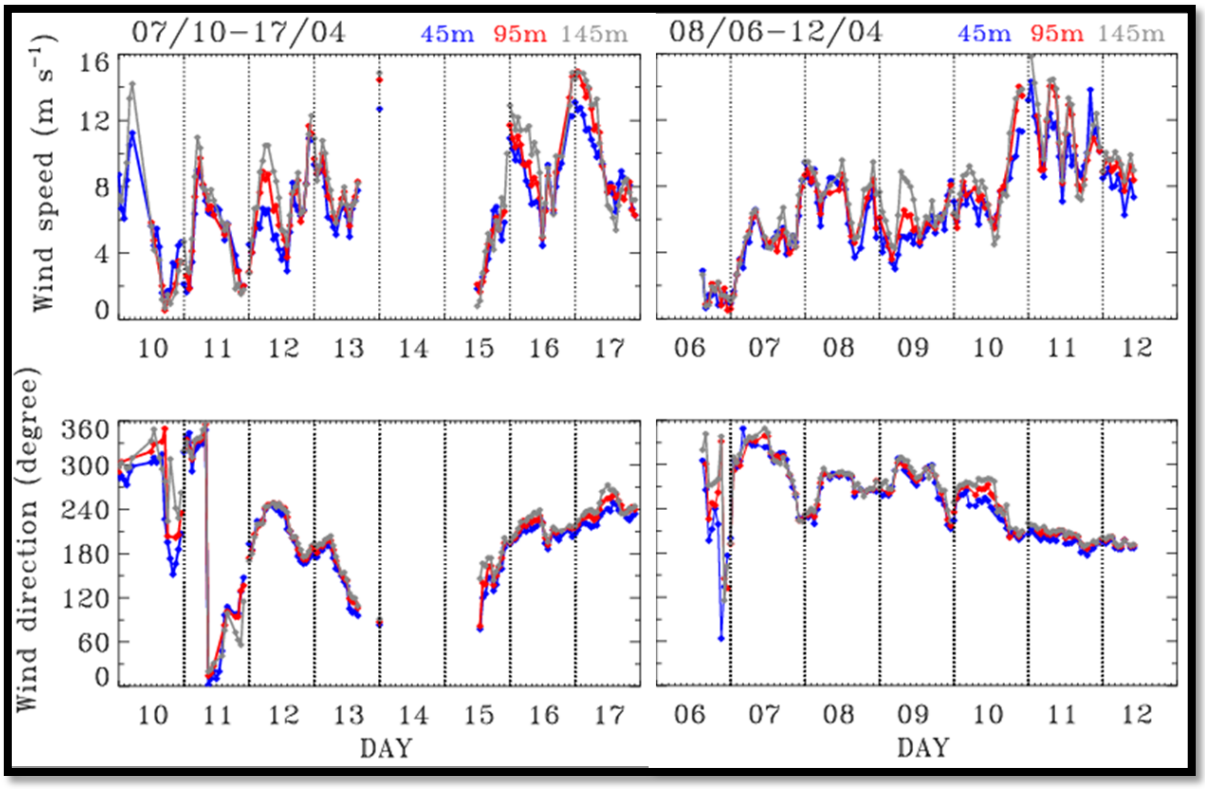


Figure C-3. Time series of wind speed (top panels, m s^{-1}) and wind direction (bottom panels, degrees) at several heights within the turbine rotor-swept layer are shown for two periods (left) July 10-17 and (right) August 6-12.

Due to differing synoptic conditions, stronger wind-direction variations were seen during the July period, due to the heavy rain and dense fog observed from the evening of 13 July until mid-afternoon of 15 July. HRDL was not operated during these hours (blank area on the left panels). Besides the day-to-day differences in meteorological conditions between the two selected periods, the differences in wind properties also reflect spatial differences in the location of the ship tracks, leading to the measurements being taken at different off-shore distances or locations along the coast line.

Rotor-level data measured by HRDL can also be used to determine wind-speed distribution histograms. Overall, meteorological conditions at heights of the presumptive turbine rotor-swept area (40-160m) for the two selected periods were dominated by westerly and northwesterly wind directions and a wide range (2-17 m s^{-1}) of winds speeds with mean values of 7-8 m s^{-1} as shown in Figure C-4.

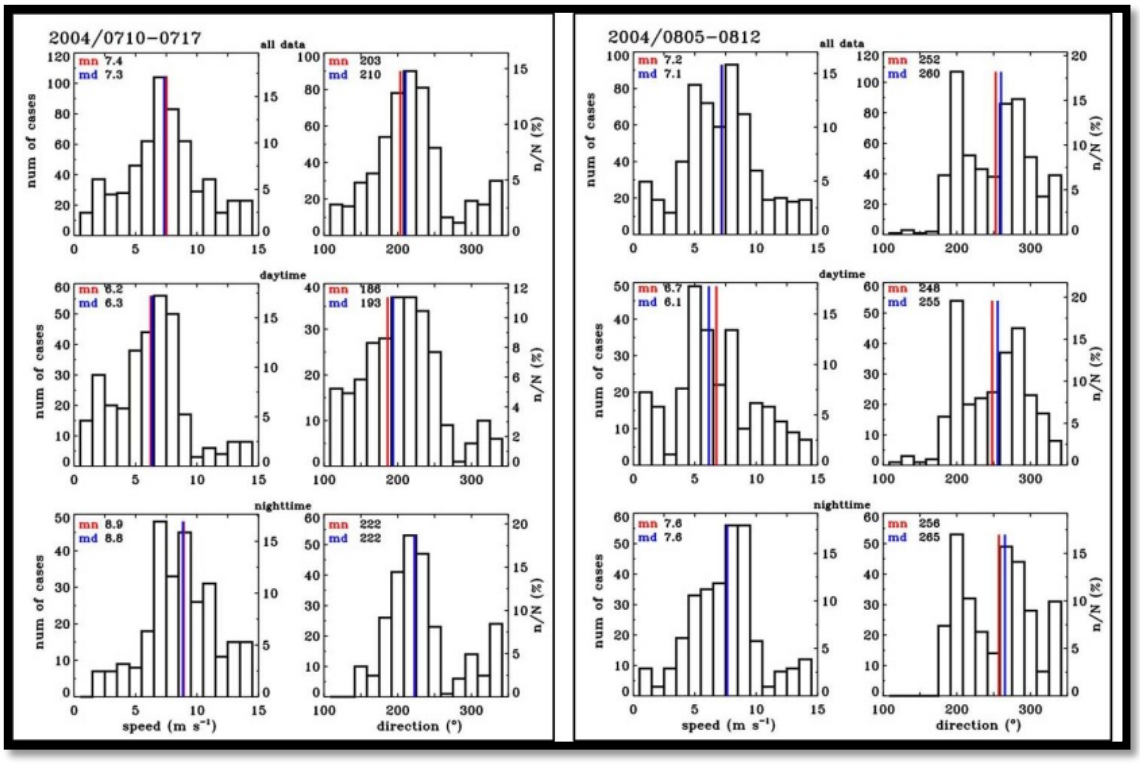


Figure C-4 Distribution of rotor layer (40-160-m) wind speed and direction during two periods selected for the study: (left) 10-17 July and (right) 5-12 August. Top panels—all measurements; middle panels—daytime measurements; bottom panels—nighttime measurements. Color lines indicate mean and median of distributions.

Percentages of wind speeds in the rotor layer were divided into 4 wind speed ranges as shown in Table C-1 for the July and August periods. The results are shown for all hours, for daytime (roughly 12-00 UTC), and for nighttime (roughly 00-12 UTC). Such information, obtained for short-term periods, illustrates the kind of information that can be obtained during long-term measurements to provide estimates on the percent of time, e.g., when winds are less than 4 m s^{-1} and turbine will not operate, or a percent of time when winds are strong enough to exceed the cutoff speed (we did not observe such cases during both study periods: line 2, Table C-1).

Table C-1. Percent of the rotor layer winds in four wi`nd speed categories

Wind speed range	July 10-17			August 6-12		
	All	Daytime	Nighttime	All	Daytime	Nighttime
ws < 4 m s^{-1}	15.6	23.69	6.34	12.48	15.21	9.82
ws > 25 m s^{-1}	0	0	0	0	0	0
4 < ws < 12 m s^{-1}	73.23	69.84	77.11	78.25	76.81	79.64
4 < ws < 25 m s^{-1}	84.4	76.31	93.66	87.52	84.78	90.18

Statistics, similar to weekly distributions presented in FigureC-4 could be obtained for the duration of the whole project (July 9-August 12) from all available Doppler-lidar profiles for mean wind speed and shear across the blade layer as examples of longer-term averages. Since few if any other long-term measurements are available aloft in the blade layer over the ocean, these measurements could be considered a first guess for summer time climatology of winds aloft in this layer over the Gulf of Maine. Obviously, longer-term deployments would be needed to calculate longer-term (e.g., annual) averages.

Comparison of near-surface and hub-height winds

The assumption of a standardized wind-speed profile shape, such as the power-law wind-profile relation: $U=U_{ref}(z/z_{ref})^{\alpha}$ is a common approach used to estimate the wind speed U at a higher level (z) using near-surface measurements. Some Illustrative examples of significant deviations of measured wind profiles from those computed by power law, using combinations of two values of the shear exponent and wind speed from two reference heights, are given in Pichugina et al. (2012). A question is, can the horizontal *distribution* of near-surface winds be used to infer the distribution of hub-height winds.

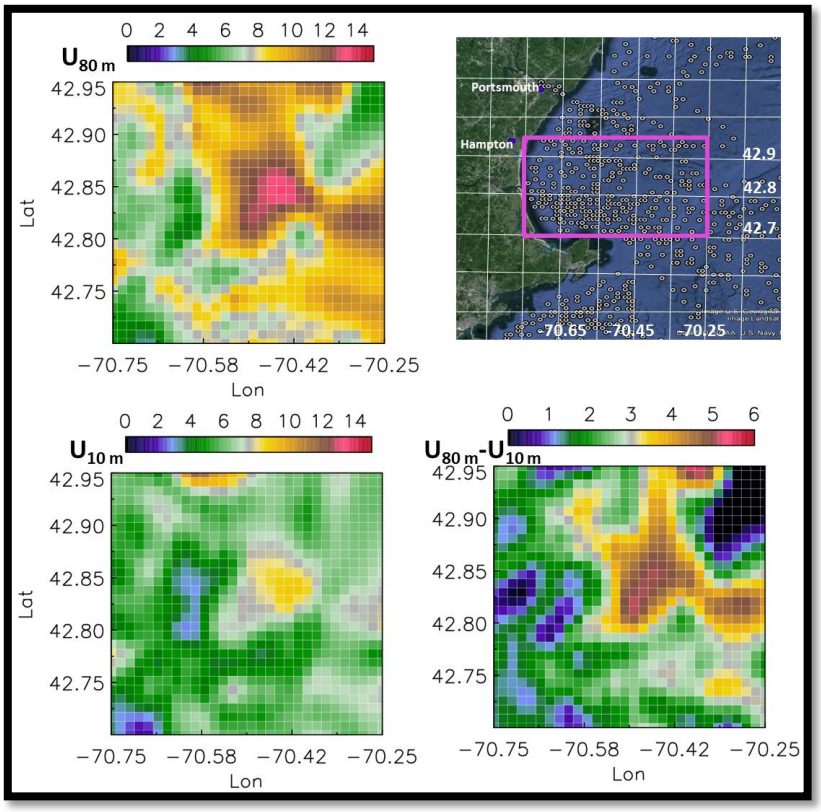


Figure C-5. Spatial distribution of winds at 80 and 10 m. Two left plots show pseudo-maps of (top) wind speed at 80 m ASL and (bottom) speed at 10 m from lidar 15-min data taken during the whole 2004 measurement campaign in the Gulf of Maine. These maps were obtained by weighted gridding of data in the small area (magenta box on the top, right) most covered by lidar data during 6 weeks of measurement. The 80m-10m difference in wind speed pattern is illustrated in the bottom, right plot.

In the text, it was shown using HRDL wind data for the entire cruise and the entire region covered, that the horizontal distribution of 10-m winds differed from the 80-m distribution (Figure _). The region considered was unevenly sampled. In Figure C-5, the spatial distributions are determined again, focusing in on the region frequently traversed by the *RHB* during the cruise over a range of weather and wind conditions. Again, the two distributions (left top and bottom) differ. Quantitatively, FigureC-5d shows that the 80-10-m speed difference goes from near 0 to more than 5 m s^{-1} depending on location, a range too large to be accommodated by standardized profile extrapolation—again, extrapolating wind speeds vertically from near the surface would lead to large errors in many parts of the domain at least.

Model validation

For comparison with modeled winds, lidar data were averaged over 1-hr intervals and interpolated to the model heights. The models use pressure as a vertical coordinate, resulting in

slight variations of model geometric heights from hour to hour and day to day. For this analysis, we used mean height values for each model in the selected period (as was shown in Figure _). Sample profiles of measured and NAM modeled wind profiles are shown in Figure C-6. Solid red and blue lines in this Figure represent model experimental runs, while dashed lines show winds from control runs. The largest discrepancies were observed for stronger winds and LLJ-like profiles which are more frequent during nighttime hours as demonstrated in the top four profiles of 16 July. The better agreement was observed for weak and moderate wind speeds as shown in all profiles for 9 August. Similar results were observed for all days in both study periods.

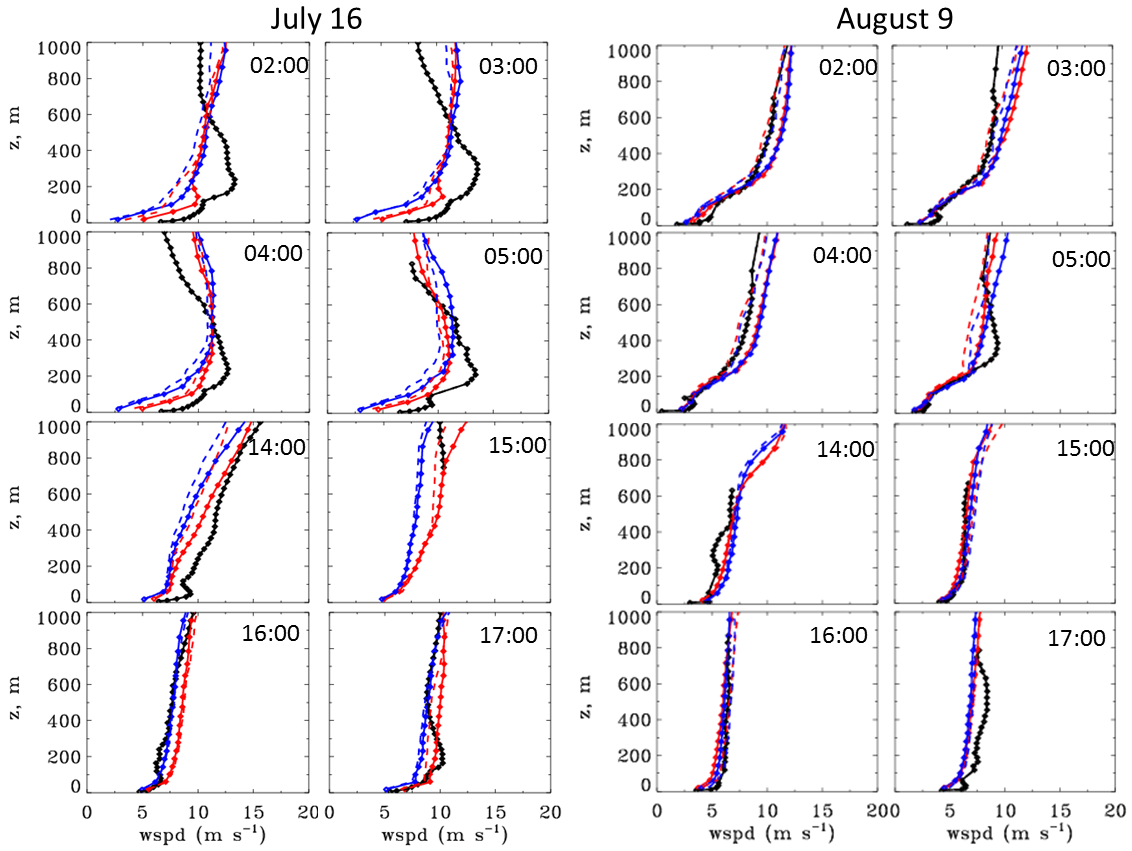


Figure C-6 -Examples of lidar-measured and modeled wind profiles (left two columns) July 16; (right two columns) August 9. Lidar profiles are shown in black. Red and blue lines are NAM CONUS and NAM PARENT correspondently. Solid color lines represent experimental runs, dashed color lines represent control lines. Symbols indicate heights of observational and modeled data.

A time-series comparison of modeled wind speed and direction with HRDL-measured winds is shown in FigureC-7 for all days in the August period. The comparisons show reasonable agreement in wind speed except for some episodes of increased winds due to LLJ occurrences when jet maxima were close to these heights. Significant variations in wind direction during 6 August have been attributed to the approach of the first hurricane of the season, Hu-

ricane Alex, that formed on 31 July and dissipated on 6 August, reaching peak wind speeds of 120 mph (54 m s^{-1}) while off the coast of New England.

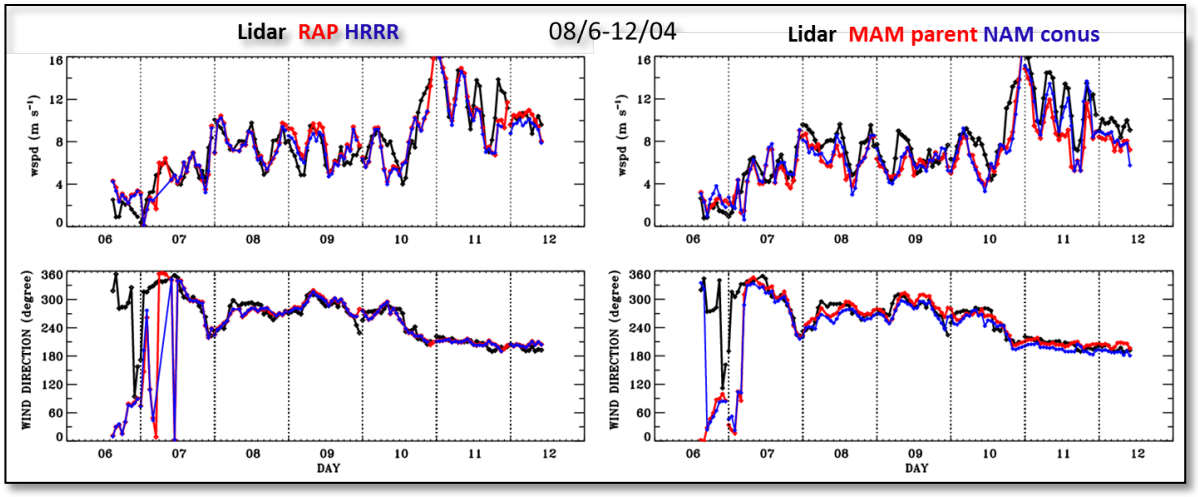
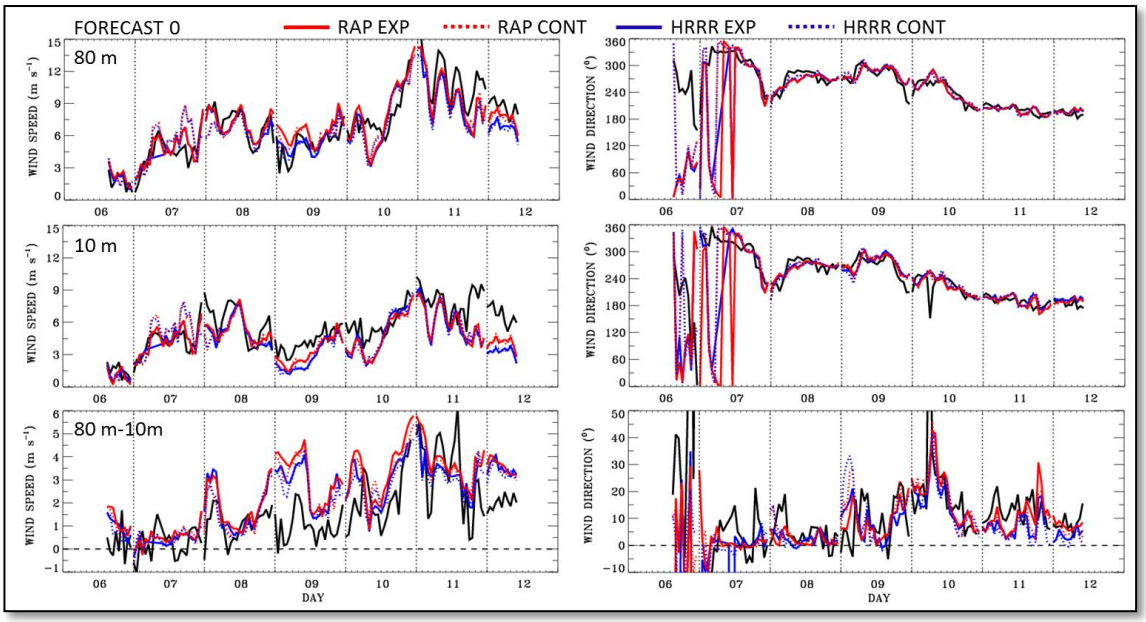


Figure C-7. Observed and modeled (top panels) wind speed and (bottom panels) wind direction for 6-12 August. Data are shown for initial conditions at the fourth level of all models (RAP and HRRR ~165m, NAM ~ 142m).

Time series of modeled and observed wind speed at two different vertical levels for all days in the August period (in Figure C-8) demonstrate some differences in (left) wind speed and (right) wind direction pattern near the water surface and at a presumptive turbine hub height.



FigureC-8. Modeled and observed (left) wind speed and (right) wind direction for 06-12 August: Top panels- at 80 m; middle panels – at 10m; bottom panels show difference in data at 80m and 10 m. Black solid line indicates lidar data; Solid and dotted red and blue lines indicate modeled data according to the legend at the top of the plot. All data are shown for the initial conditions.

Greater deviations of modeled wind speeds from measured are seen at the lower heights, confirming the previous results as seen in Figure C-5. Significant differences in the trend of winds at both heights could be seen in both lidar-measured data and model predictions (for example in 9 or 11 August). Smoother for modeled data these differences are substantial in lidar measurements for 6-12 August, similar to results in Figure C- 5 shown for the whole experiment.

Scatter plots of wind speed in the first 500 m above the water surface (FigureC-9) show nice agreement in observed and modeled winds below 10 m s^{-1}, where linear-fit lines shown in red and blue are close to the best-fit shown by the black solid line. Deviations from the best fit line increase with increasing wind speed, reaching 2-3 m s^{-1} at the peak measured speed of 18 m s^{-1}. The number of data points with strong winds is fewer compared with the number of data points close or below 10 ms^{-1}, but overall the plots show a slow bias in the modeled wind speeds, increasing with speed.

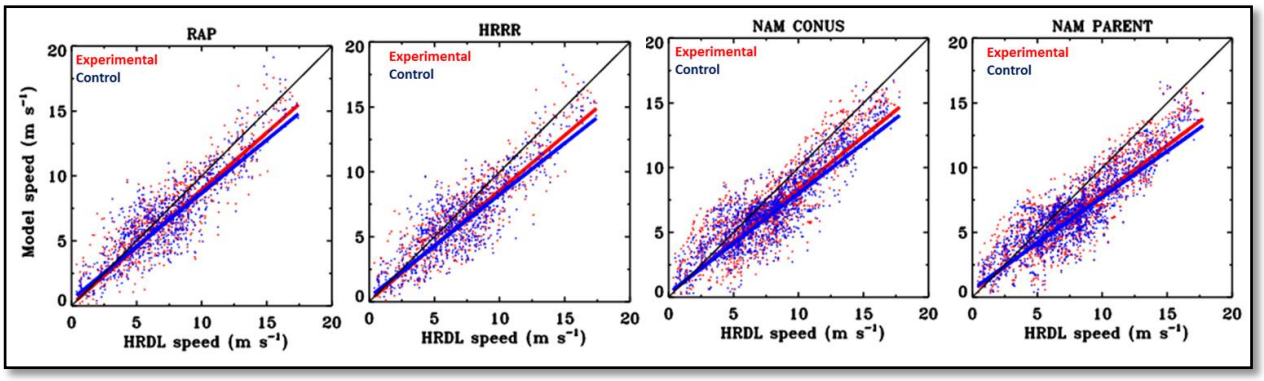


Figure C-9. Scatter plots of lidar measured and modeled wind speed in the layer of 500 m above the ground during August 6-12. All plots are shown for initial conditions. Solid color lines are the best fit, red- for experimental runs, blue -for control runs. Black line is 1:1 fit.

Correlation statistics for these plots are shown in Table C2.

Table C-2. Correlation statistics between measured and modeled wind speed in the layer of (NAM models) 20-500 m and (RAP, HRRR models) 11-500 m during August period.

Model	count	R^2	Bias	Slope
RAP exp	810	0.88	0.11 ± 0.14	0.88 ± 0.02
RAP cont	810	0.87	0.49 ± 0.16	0.82 ± 0.02
HRRR exp	750	0.88	0.09 ± 0.15	0.85 ± 0.02
HRRR cont	804	0.86	0.36 ± 0.15	0.79 ± 0.02
NAM conus exp	1479	0.88	0.25 ± 0.12	0.81 ± 0.02
NAM parent exp	1479	0.88	0.49 ± 0.12	0.74 ± 0.01

Profiles of the mean bias between modeled and observed scalar wind for various model lead times are presented in Figure C-10. The figure shows that in the first several forecast hours, models underestimate wind speed (negative bias) and after 6-7 hr lead time the bias became positive. Similar to profiles of correlation coefficient and RMSE (Figure 3.17.) and mean bias profiles shown for the lead hour 0, bias profiles for all lead hours show larger discrepancies below 100 m and these deviations are stronger near the water surface.

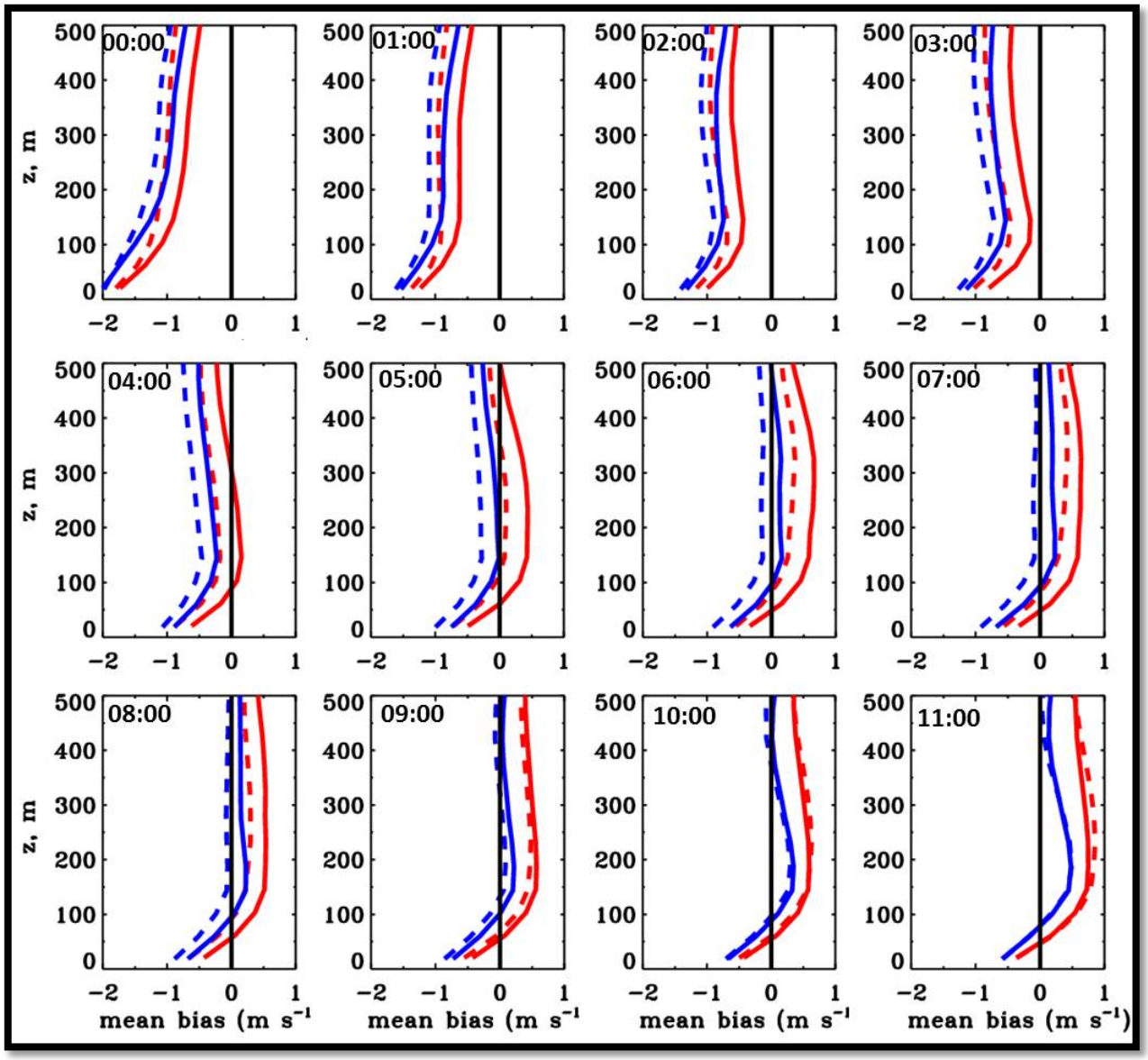


Figure C-10. Profiles of the mean bias between modeled and observed scalar wind for 11 lead times. Red and blue solid lines are for the NAM- CONUS and NAM- parent experimental runs; red and blue dashed lines are for the NAM- CONUS and NAM- parent control runs

APPENDIX D: Detailed analysis, July study period

Disturbed weather conditions characterized by large-scale vortices having complex wind-flow patterns over coastal and ocean areas can lead to large uncertainties in model forecasts. The anomalous behavior of the model error statistics during the mid-July study period was an example of such a case. To further investigate the reasons for this behavior, a detailed analysis has been performed for the July period.

Figure D-1 provides an overview of surface and cloud conditions during this period. The surface wind data superimposed on the satellite images show significant shifts in wind directions during the 13-14 July period, characterized by the mesoscale low-pressure vortex, dense fog, and heavy rain.

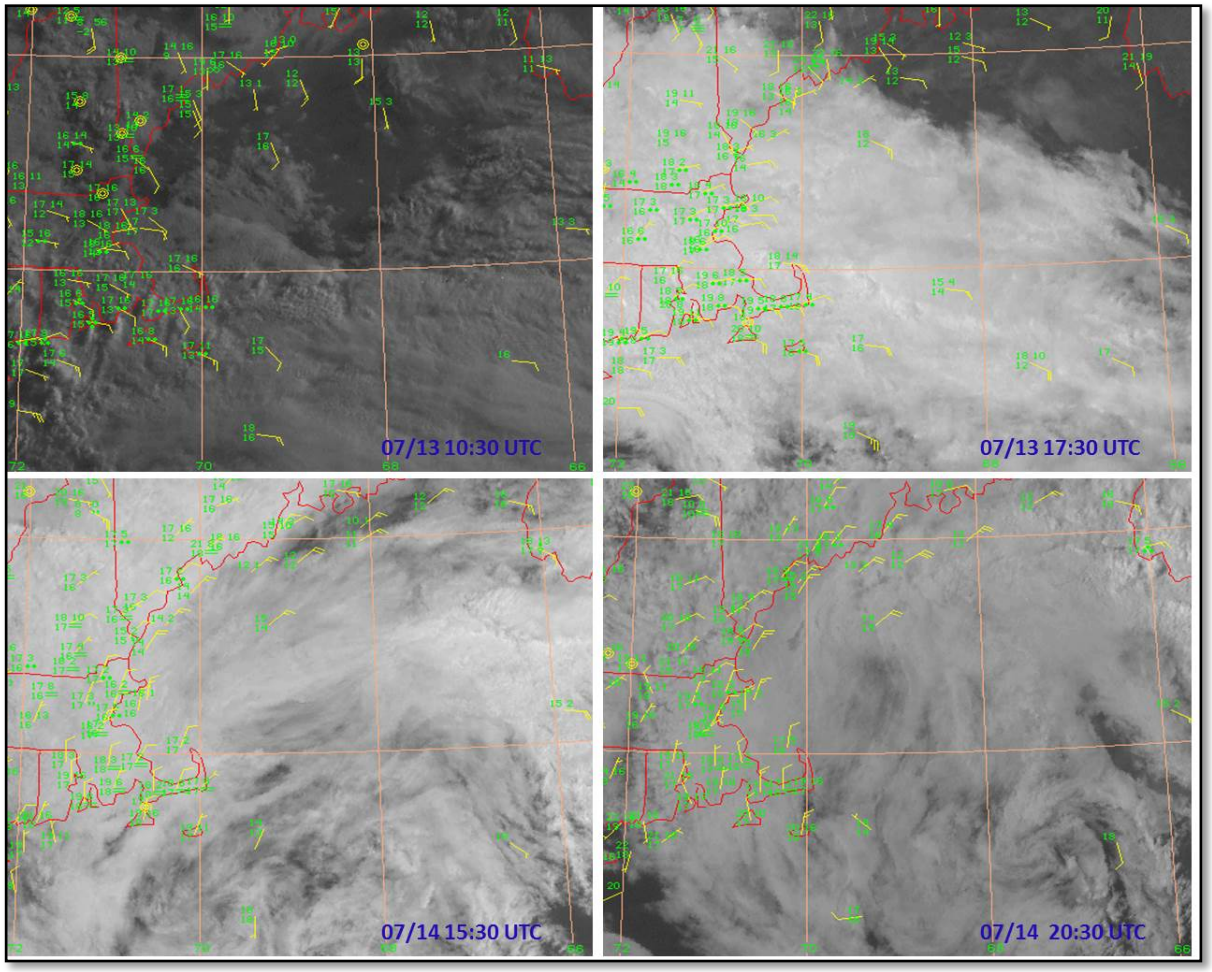


Figure D-1. Examples of the surface data combined with satellite images from GOES-EAST (1-km resolution, visible) demonstrate meteorological conditions observed during 13 -14 July.

Leading up to this event, on July 13 HRDL data were taken from a stationary position (Figure D2, left) when the ship was located ~9.5 mi from the shore about 16 hr prior to moving to another location. Time-height cross sections of all available data (FigureD2, right) show areas of stronger (10-12 m s^{-1}) winds propagating from near the surface to higher elevations from a south-westerly direction. Around 09 UTC the wind direction changed to southerly then to south-easterly at all elevations. After sunrise (~12 UTC) winds became easterly in the lowest 200 m.

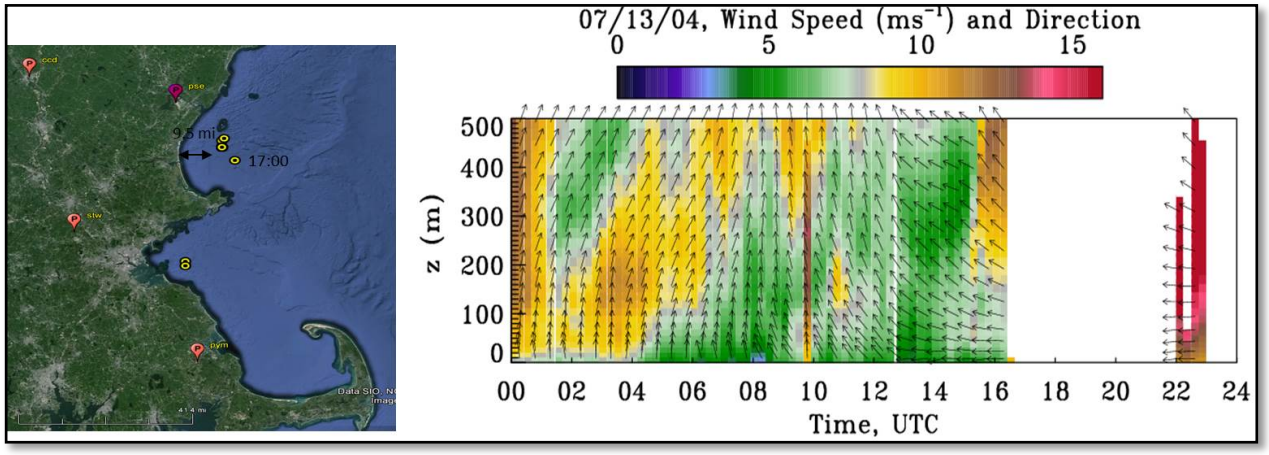


Figure D-2. Left: location of RHB during 16 hours on July 13. Right: time-height cross section of wind speed and direction in the lowest 500 m on July 13. No lidar data were obtained from 17-22 UTC due to the intense rain.

Time-series of wind speed and direction at several heights (Figure D3) show the changes in these variables in finer detail.

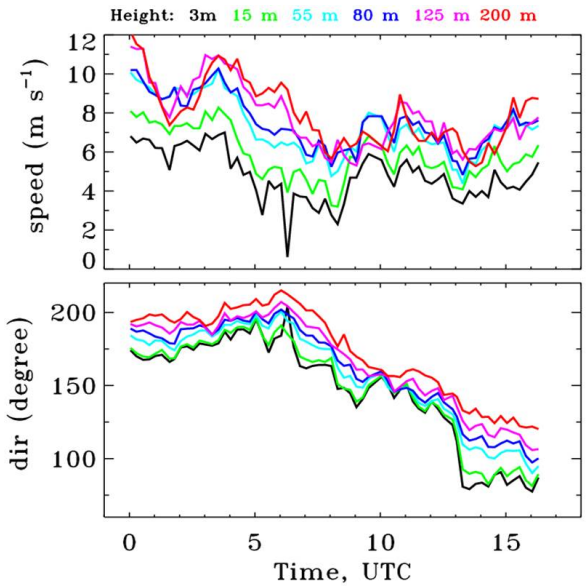


Figure D-3. Time-series of (top) wind speed and (bottom) wind direction measured by HRDL for 16 hours on July 13.

Figure D-4 demonstrates an increase in RMSE discrepancies between measured and modeled winds in the lowest 500 m as the mesoscale storm approached the study area on 13 July. As described in the main text, these errors were larger for model runs with assimilation of coastal wind profiler data, in contrast to expectation.

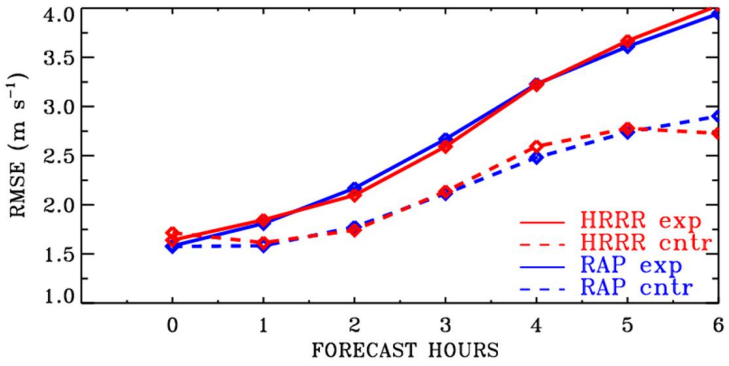


Figure D-4. RMSE between HRDL-measured and modeled winds averaged in the first 500-m layer.

Using the land-based profiler-array data as a reference, an analysis of RMSE statistics vs. forecast lead time has been presented in the main text for the July period (Figure X.9). A similar statistical analysis has been reproduced for each individual day of the period (FIGURED-5). The statistics are shown only for the low-resolution models, the RAP and NAM. It reveals a big dif-

ference in the behavior of the RMSE between the experimental and control runs for three days in particular, 13-15 July 2004, indicated by the red boxes in the middle of both plots.

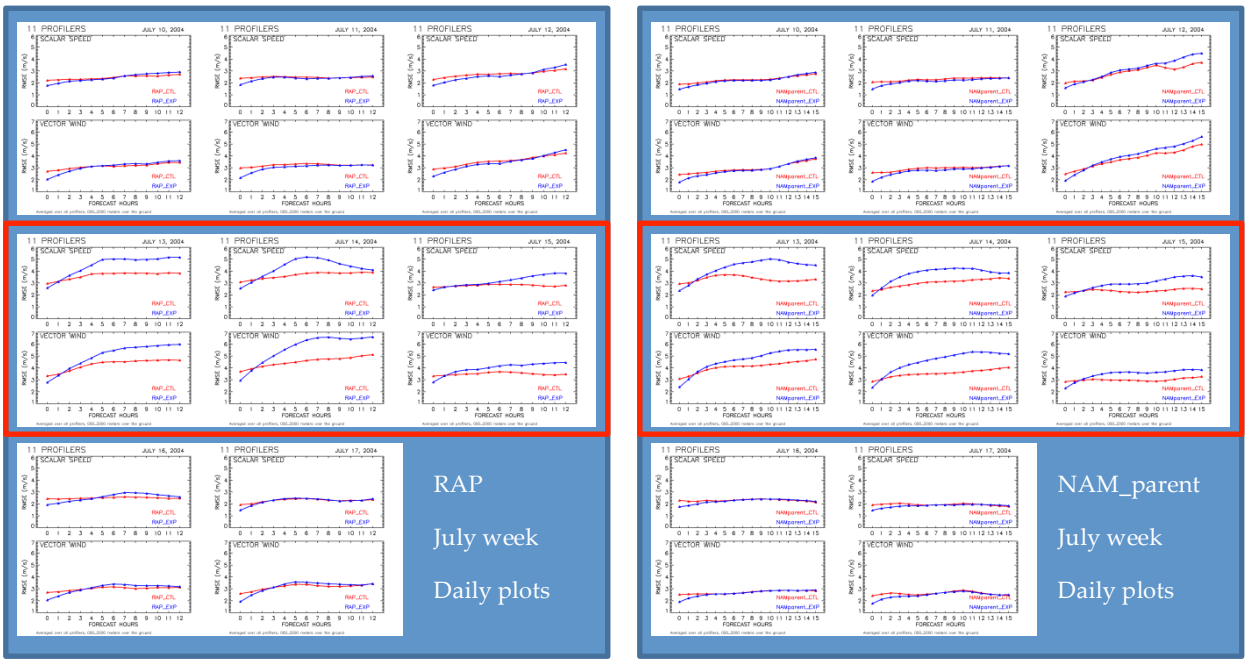


Figure D-5. RMSE daily statistics for the RAP model on the left and NAM model on the right for the scalar wind speed averaged over 2000 meters above the ground.

For additional analysis, we also used a dataset of buoy observations that provides wind speed measurements 2-5 m above the ocean surface (FigureD-6). The map with the buoy locations is in the left panel. We separated the buoys into three sets: one set includes the buoys along the coast (blue in the figure), the second set is in the central part of the Gulf of Maine (orange in the figure), and the third set includes the buoys farthest from the coast (red in the figure). We compared the buoy data to the model data at the same location, at the initialization time for the first model level, which is around 10 meters above the ocean (RAP 11 m, NAM 20 m, see Figure 6). We calculated the RMSE statistics according to the buoy sets and present it in the six right panels (the RAP model is in the middle and NAM model is on the right). The large discrepancy between the experimental and control models is highlighted in the circles. Both models show the same "problematic" days of July 13-15, 2004, when the model experimental runs have bigger RMSE compared to the control runs.

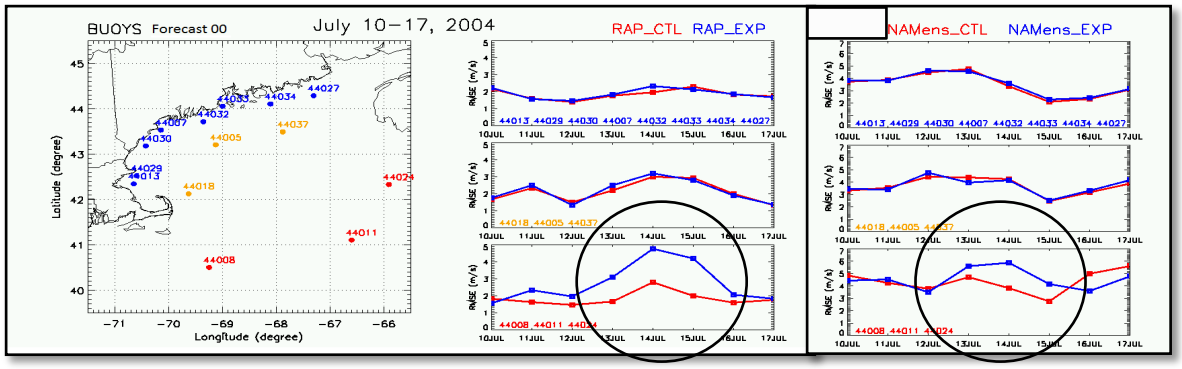


Figure D-6. Location of the buoys on the left, separated in three categories: along the coast – blue, in the center of the gulf of Maine – orange, and far in the ocean – red, with the corresponding RMSE statistics from RAP model in the middle and from NAM model on the right, over the 7 days of the July period.

The buoy data were assimilated into both the control and experimental runs of the models but the large difference between RMSE of the experimental and control runs is seen only for the set located far in the ocean. What was the cause of this discrepancy? To find out, the model outputs were analyzed on an hourly basis for the whole July period. We found that a cyclonic storm system was moving over the ocean during these "problematic" days. The model low-level wind vector field is shown in FigureD-7. This instantaneous output of the RAP model shows a mesoscale vortex over the gulf of Maine, and also reveals a large difference between the experimental and control simulation winds, especially evident in the direction of the winds. The storm vortex produced easterly-component winds on its north side as evident in the model prediction (FigureD-7), and also evident in surface-satellite data (FigureD-1) and the Appledore Island profiler data presented in the main text (Figure X.13b). Thus, during the problematic time period of 13-15 July, the wind direction changed to the south, southeast and east directions, whereas for the other days the west or north wind directions prevailed.

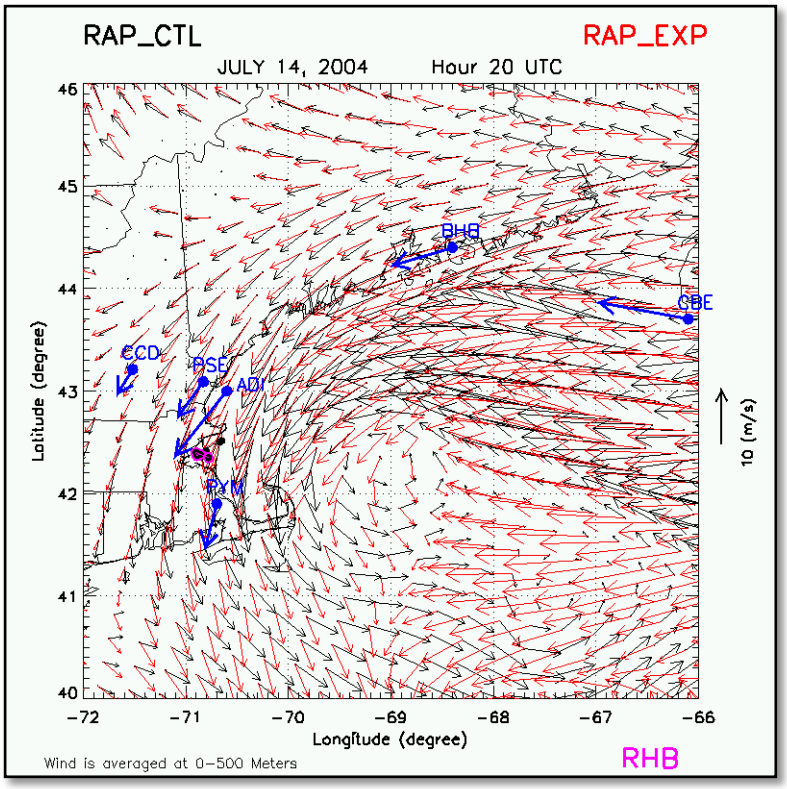


Figure D-7. One instant of RAP-model output data for the control (black) and experimental (red) runs. Winds are averaged over 500 m above the surface at hour 20 UTC of 14 July.

It is useful to show the larger domain of the models, RAP and NAMRR, (FIGURED-8), to demonstrate that the vortex is represented differently in the experimental and control runs, especially the shifted locations of the centers of the vortices.

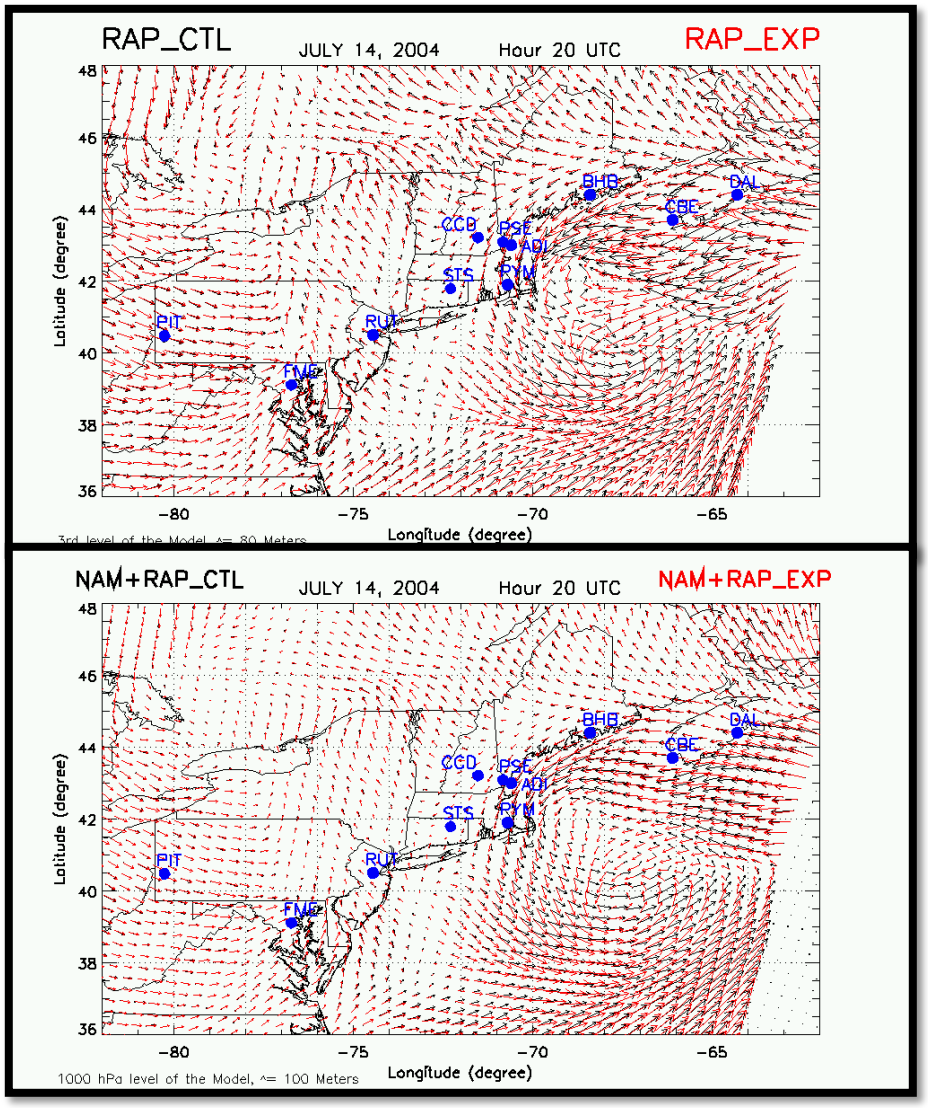


Figure D-8. RAP model third level wind output and NAMRR+RAP mean output on the bottom at hour 20 UTC of 14 July.

To find out how this discrepancy influenced the comparison of the wind profiling radar data inland, the profiler data from the lowest range gates were analyzed in FIGURED-9. Again, we separated the sites into three sets: one set includes the profilers along the coast (blue in the figure), the second set are the profilers a little further from the coast of the gulf of Maine (orange in the figure), and the third set includes the profiler farther inland (red in the figure). The RAP model data are in the upper plots, and the NAM model, in the bottom plots. The RMSE statistics for all inland profilers together for all July study days are shown under the map images on the left, for both plots. These curves show the RMSE improvement of the experimental runs over the control runs for all days of the July period at the initialization time, forecast hour 00, as veri-

fied at the profiler sites. The same RMSE improvement is shown on the three additional panels on the right side on each model image for each of the three profiler sets, where we see that the RMSE of the experimental runs is smaller compared to the RMSE of the control runs.

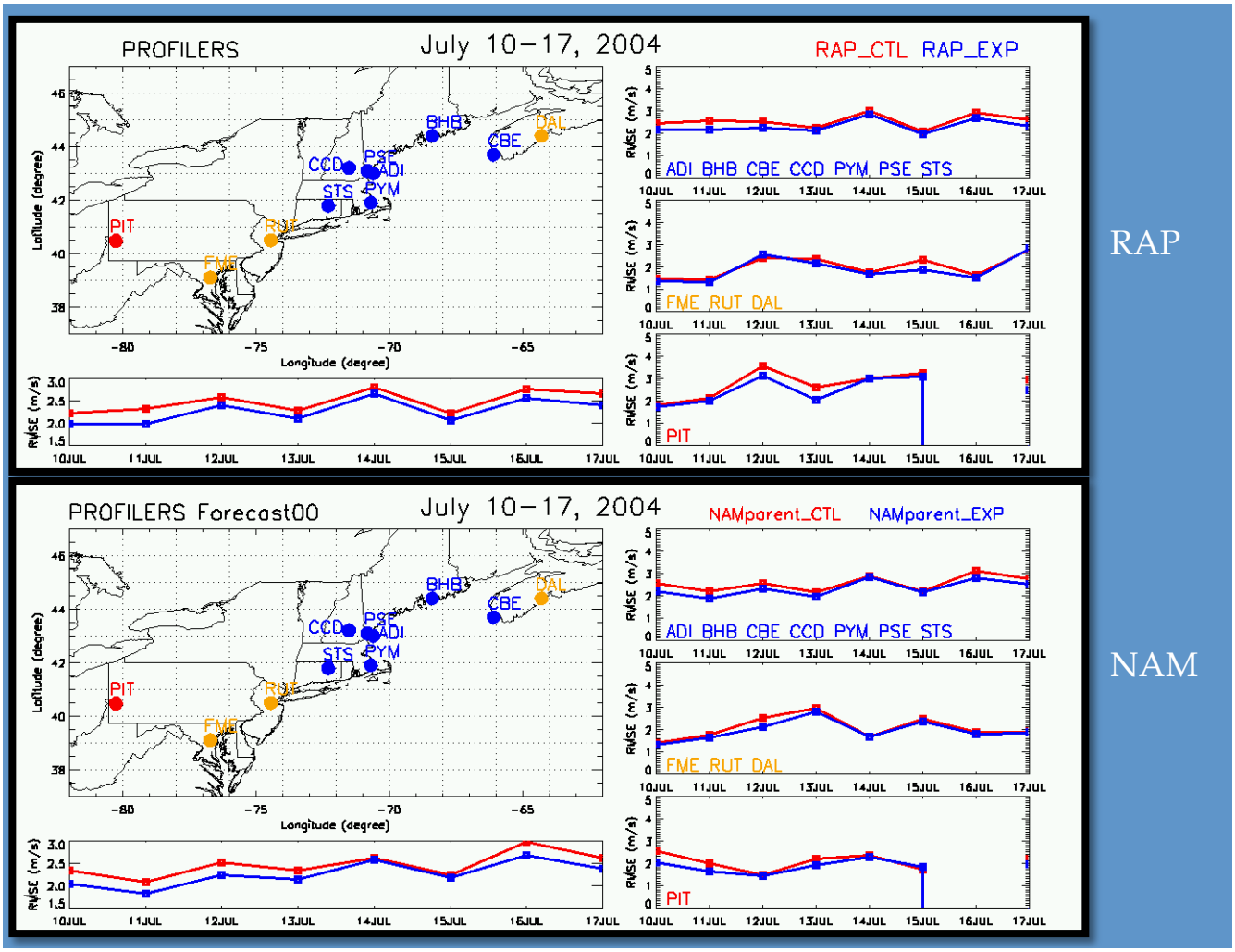


Figure D-9. Map of the profiler locations on the left, separated into three categories: around the Gulf of Maine – blue; on the shore far from the Gulf of Maine – orange; and far inland (only one profiler in Pittsburg) – red. Corresponding RMSE statistics for the RAP model are given in the upper plots, and for the NAM model, in the lower plots, for the July period. Below each profiler map is the RMSE statistics for all profilers. The model outputs are compared at forecast hour 0.

The same statistics, but for forecast hour 6, show the experimental-run statistics being worse compared to the control run ones (FigureD10). The statistics computed for all profilers (below each map image) present discrepancies between the experimental and control runs for several days of the July time period, particularly for 13-15 July. The three panels on the right show that this discrepancy mostly occurred for the profilers located along the cost of the Gulf of Maine.

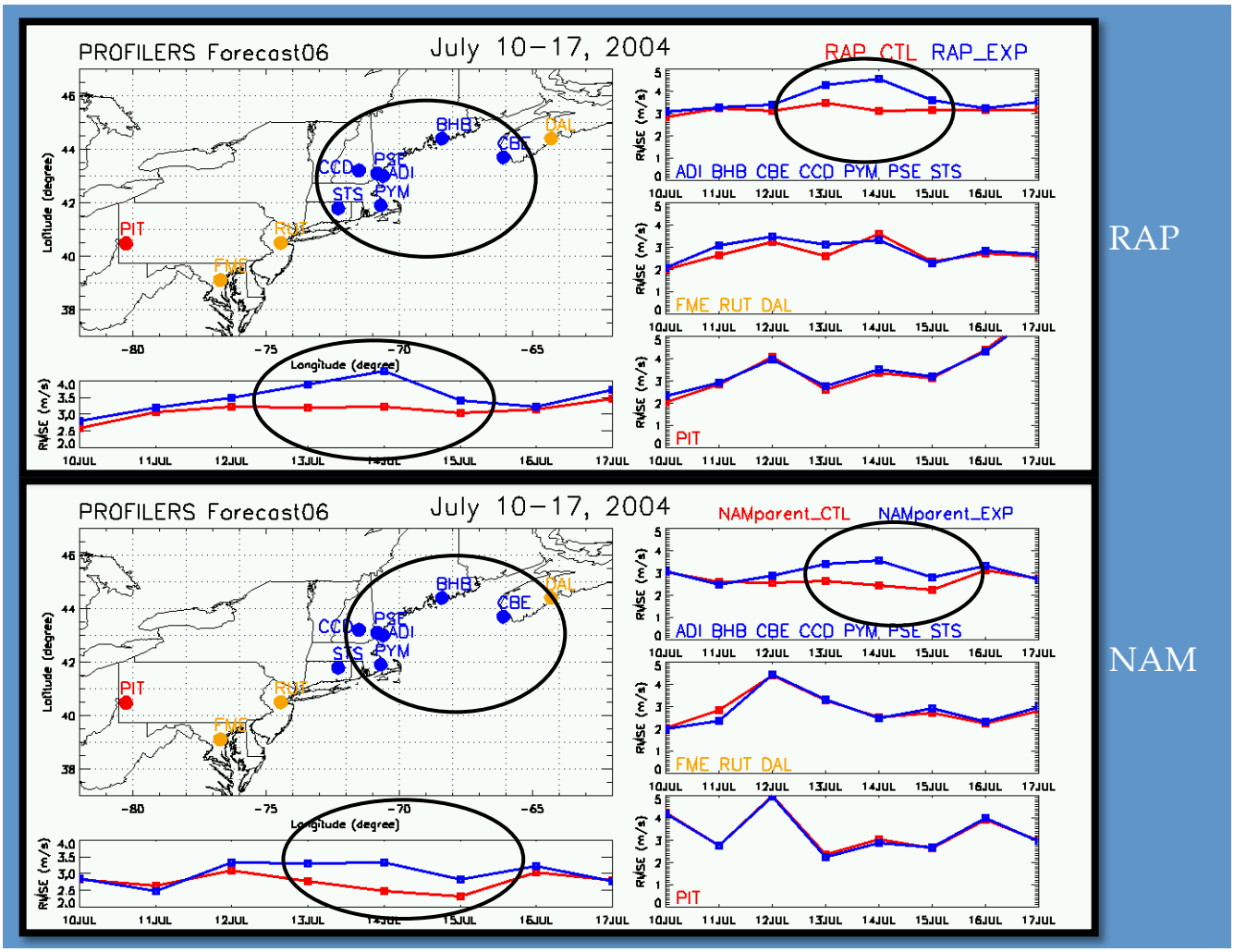


Figure D-10. Same as in FigureD9 but for forecast hour 6.

This confirms what we had found from the buoy analysis, that the degradation in the experimental run was more likely due to the initial differences over the ocean between the experimental and control runs. We note that the data assimilation technique used for this experiment was 3DVar, while more advanced assimilation techniques (such as EnKF) likely would have made better use of the assimilated data, therefore improving the results. Unfortunately those updated techniques could not be used for this experiment due to the lack of available of GDAS EnKF members from 2004, which both RAP and NAM/CONUSNEST systems now use as a part of a hybrid ensemble-3DVar analysis. In addition, for a meteorological situation like this, the availability of more offshore observations, such as buoy-mounted wind profiling systems, might have provided better constraints on the model at the initialization time, possibly yielding better forecast.

APPENDIX E: NAMRR and NAMRR-CONUSNEST - Forecast verification against surface observations

For verification against conventional surface observations of 2-m temperature and 10-m wind (e.g. from METAR), the model forecast parameters were interpolated to their observed counterparts and the differences calculated. Since the supplemental profiler observations in the POWER project were restricted to the eastern part of the United States (e.g. Figure 2), only surface verification statistics from this region are presented here (Figure E1) and are intended to yield only a bulk perspective. Note that the NAMRR referenced in this section is the same as the NAM referenced in other parts of this report.

6–12 August 2004

Figures E2 and E3 show the RMSE and bias for 2-m AGL temperature forecasts from the NAMRR system and the ENSMEAN forecasts, respectively. In Figure E2, we see an overall, very slight degradation in the temperature RMSE verification in the CONUSNESTX, and a slight degradation in the NAMRRX and CONUSNESTX bias (an increase). However, while statistically significant, these differences are all exceptionally small and are on the order of 0.05° K or less. In comparison, the ENSMEAN(X) verification shows improvement in RMSE over both NAMRR(X) and RAP(X), on the order of 0.25 ° to 0.5° K. It is also worth noting that the RAP(X) shows lower RMSEs relative to NAMRR(X) during the short forecast window; however the RMSEs for both systems tend to converge by forecast hour 12, Figure E-3. Bias from the ENSMEAN(X) experiments is relatively mixed; it shows an improvement relative to RAP(X) but degradation relative to NAMRR(X). Differences in bias relative to ENSMEAN(X), while statistically significant, are quite small and are on the order of 0.05° to 0.10° K. The bias from the NAMRR(X), compared to the RAP(X), is closer to zero throughout the forecast window.

Figure E4 shows the effect of assimilating POWER data on the 10-m wind forecast, which shows almost no impact during the August study period, both in terms of RMSE and bias for either the NAMRRX or CONUSNESTX domains.

In Figure E5 significant improvement of the ENSMEAN(X), relative to both RAP(X) and NAMRR(X), in RMSE is seen throughout the forecast period. It is also interesting to note that while the RAP(X) 10-m wind forecasts depict a slightly lower RMSE at the beginning of the period, around 0.05 m s^{-1}, the RMSEs of RAP(X) and NAMRR(X) generally converge by forecast hour 12. The ENSMEAN(X) bias shown in Figure E5 is an improvement relative to the

NAMRR(X)'s high bias, but degrades the RAP(X)'s bias, which is closer to zero by a narrow margin (~0.05 m/s).

10 – 17 July 2004

Forecast verification of surface observations for the July study period generally followed a similar trend to those seen in the August period; the differences which exist are generally quite small and the ENSMEAN(X) forecasts tend to show a reduction in RMSE relative to RAP(X) and NAMRR(X).

For example, Figure E6 shows that the effect of assimilating the extra POWER data has almost no impact on the NAMRRX 2-m temperature RMSEs and an exceptionally small degradation in the NAMRR_CONUSNESTX RMSE. Biases for both domains again show very slight degradation (warming), on the order of ~0.025° K.

Figure E7 depicts the RMSEs and bias for the ENSMEAN(X) 2-m temperature forecasts. As in the August period we note that the ENSMEAN(X) shows a statistically significant improvement in the RMSE over NAMRR(X) for the entire period and is improved over RAP(X) for forecast hour 5 and beyond. It is also noted that the RMSEs for both RAP(X) and NAMRR(X) converge by forecast hour 12.

Somewhat interestingly, the characteristics of the 2-m temperature bias for RAP(X) and NAMRR(X) are nearly opposite of what they were for the August period (Figure E3 compared to Figure E7). Here the RAP(X) shows a bias closer to zero when compared to NAMRR(X). Unsurprisingly, the ENSMEAN(X) bias behaves in a nearly identical manner to its bias shown during the August study period – the bias in July falls between the NAMRR(X) and RAP(X). Therefore it shows improvement relative to NAMRR(X) and degradation relative to RAP(X).

Figures E8 and E9 depict the RMSE and bias for 10-m wind forecasts from the NAMRR(X) system and the ENSMEAN(X) experiments respectively. In both the NAMRRX and CONUSNESTX, we see a statistically significant, yet very small, ~0.05 m/s degradation of the RMSE approximately beyond forecast hour 5. Coincidentally both NAMRRX and CONUSNESTX show very slight improvements in the 10-m wind speed bias. This improvement is more pronounced in the NAMRRX than in the CONUSNESTX.

When we switch perspectives and look at the ENSMEAN(X) 10-m wind forecasts (Figure E9), we again see consistent results with what we found during the August period, which were similar to those seen in the 2-m temperature verification (Figure E3). Significant improvement

of the ENSMEAN(X), relative to both RAP(X) and NAMRR(X) in RMSE is again seen throughout the forecast period. The RAP(X) 10 m wind forecasts again depict a slightly lower RMSE at the beginning of the period of ~0.05 m s^{-1}, but then the RMSEs of RAP(X) and NAMRR(X) converge by forecast hour 12. Here the NAMRR(X) has a high speed bias and the RAP(X) has somewhat of a low speed bias.

Overall, comparison of surface measurements of temperature and wind speed against NAMRR/NAMRRX and CONUSNEST/CONUSNESTX forecasts tended to show very slight degradation or no change in both RMSE and bias. If degradation was present, it was on the order of 0.05 m s^{-1} for winds and 0.05° K for temperature (or less). Interestingly, both July and August periods had temperature biases that increased (warmed) with the addition of the POWER observations (Figs E2 and E6). Furthermore, while at certain forecast hours these results showed *statistical* significance (e.g. Figure E6), one must question if a change in RMSE of 0.05 m s^{-1} for winds and 0.05° K is really of practical significance, since such a change is, subjectively, very minor. To evaluate this effectively, additional experiments of possibly longer duration or which span multiple seasons should be conducted.

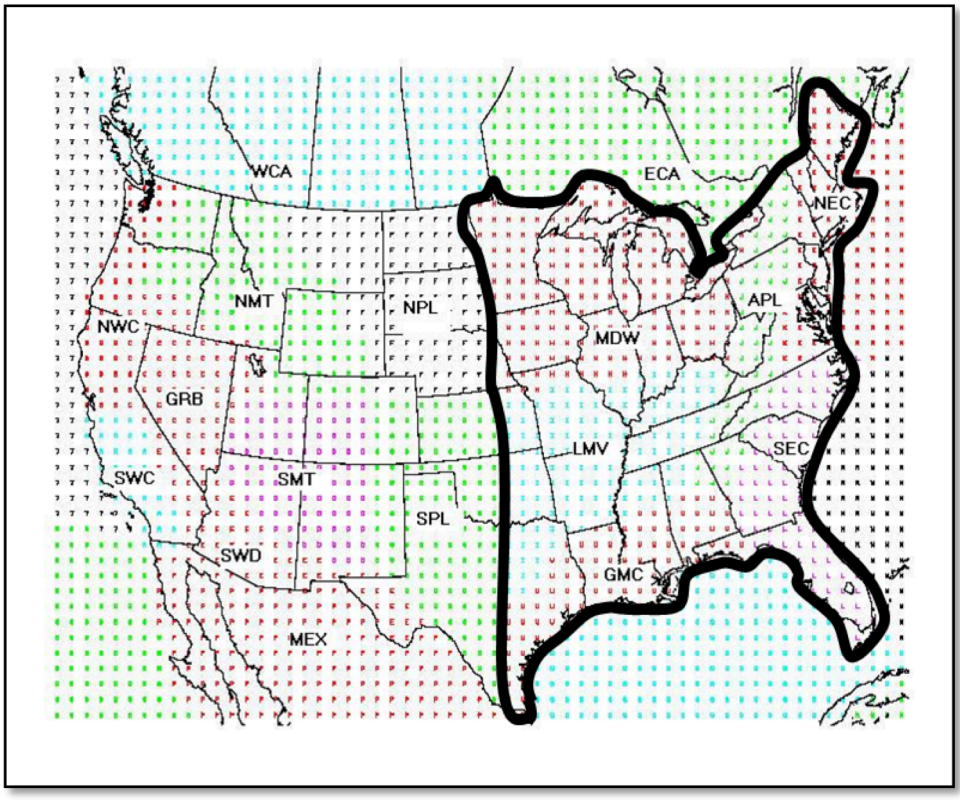


Figure E-1. Forecast verification regions in the NCEP Forecast Verification System. For conventional verification in POWER, regions MDW, LMV, GMC, SEC, APL, and NEC were combined to form the Eastern verification region (black outline).

The impact of averaging the RAP(X) and NAMRR(X) forecasts into a very simple two-member ensemble mean, known as ENSMEAN(X), had the overall effect of reducing RMSEs for surface observation based verification, regardless of the assimilation of extra profiler observations.

Brief comparisons between the RAP(X) and NAMRR(X) were also done as a part of evaluating the ENSMEAN forecasts against surface observations. Surface-based verification indicated that RAP(X) forecasts had RMSEs that tended to be lower than NAMRR(X) earlier in the forecast period, but then approached those of the NAMRR(X) toward the end of the forecast window (e.g. Figure E3). Initially this appears to imply that the NAMRR(X) system does not fit the surface observations quite as closely as the RAP(X) system during the analysis. This would not be unexpected, as this is one of the first applications of the NAMRR system, which has historically been more focused on forecasts in the 12-84 hour range rather than the first 0-6 or 0-12 hours, which is generally the primary focus of the RAP forecast window.

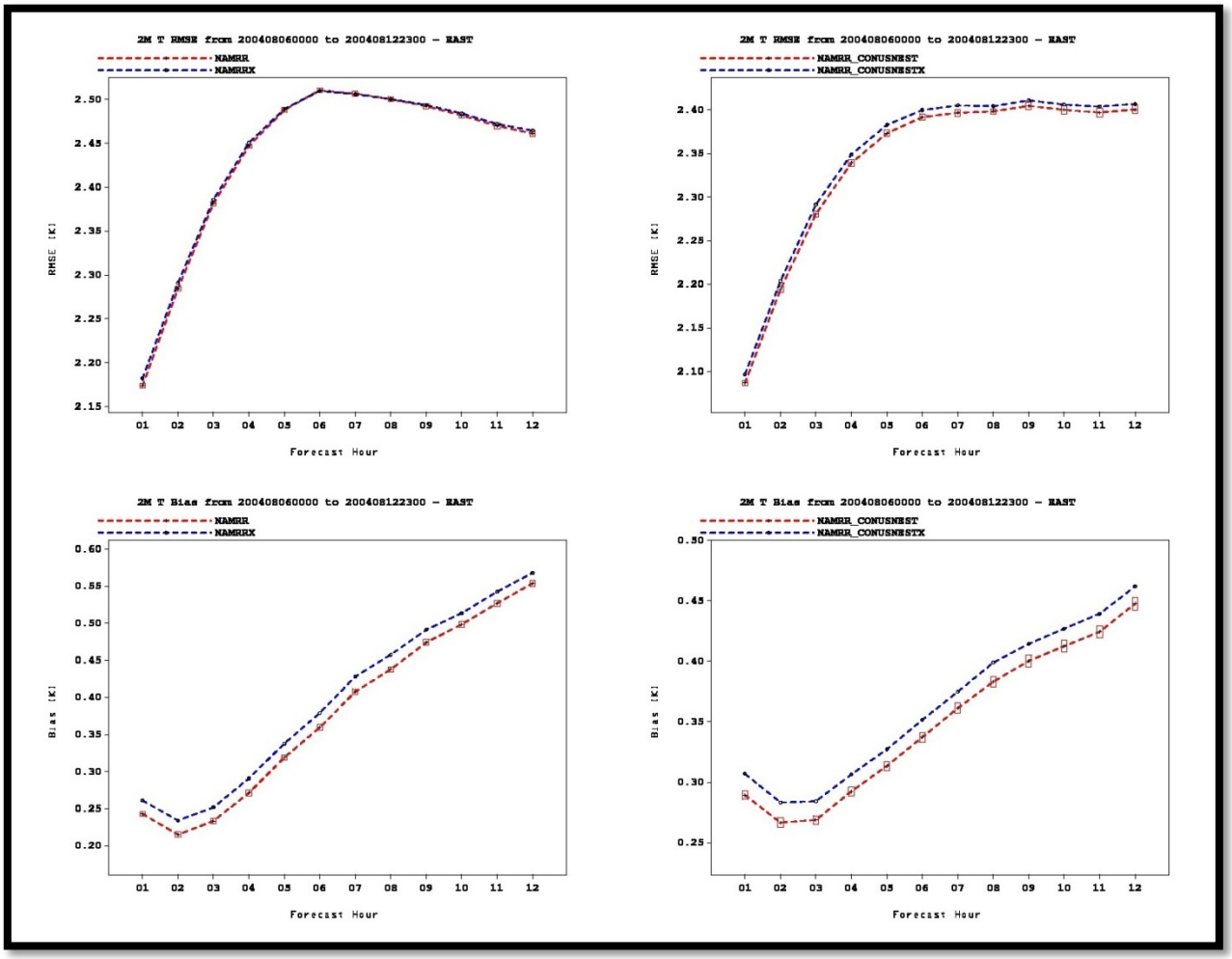


Figure E-2. 2-m temperature RMSE (top) and bias (bottom) against surface observations over the East (Figure E1) for the August period. Statistics from the 12-km parent domain occupy the left panels and forecasts from the 4-km nest domain. Red traces correspond to the control, and blue traces correspond to the experiment.

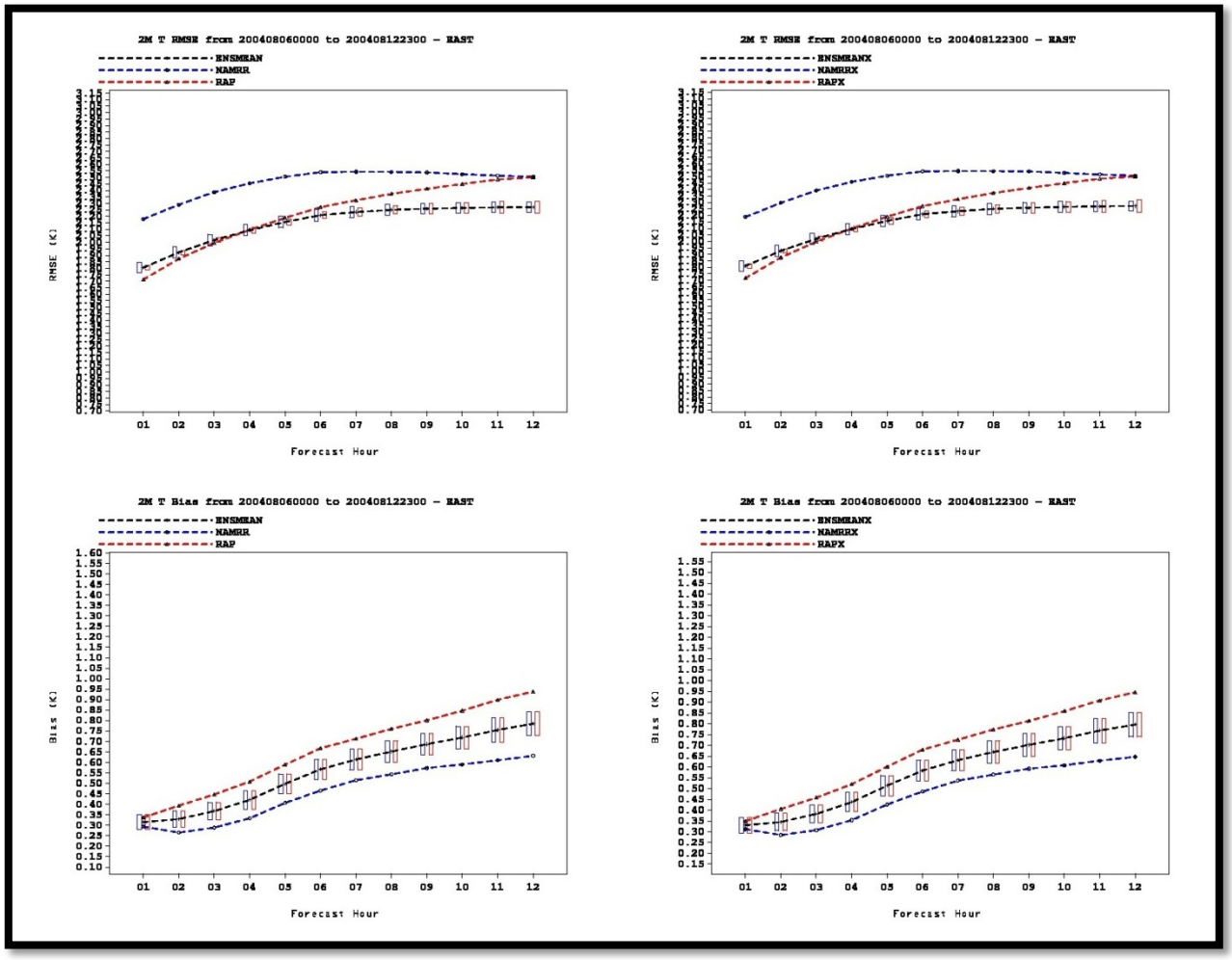


Figure E-3. 2-m temperature RMSE (top) and bias (bottom) against surface observations over the East (Figure E1) for the August period. Each panel shows statistics comparing RAP (red), NAMRR (blue), and their mean (ENSMEAN; black). The left panel corresponds to the control simulations, and right panel corresponds to the experiment simulations.

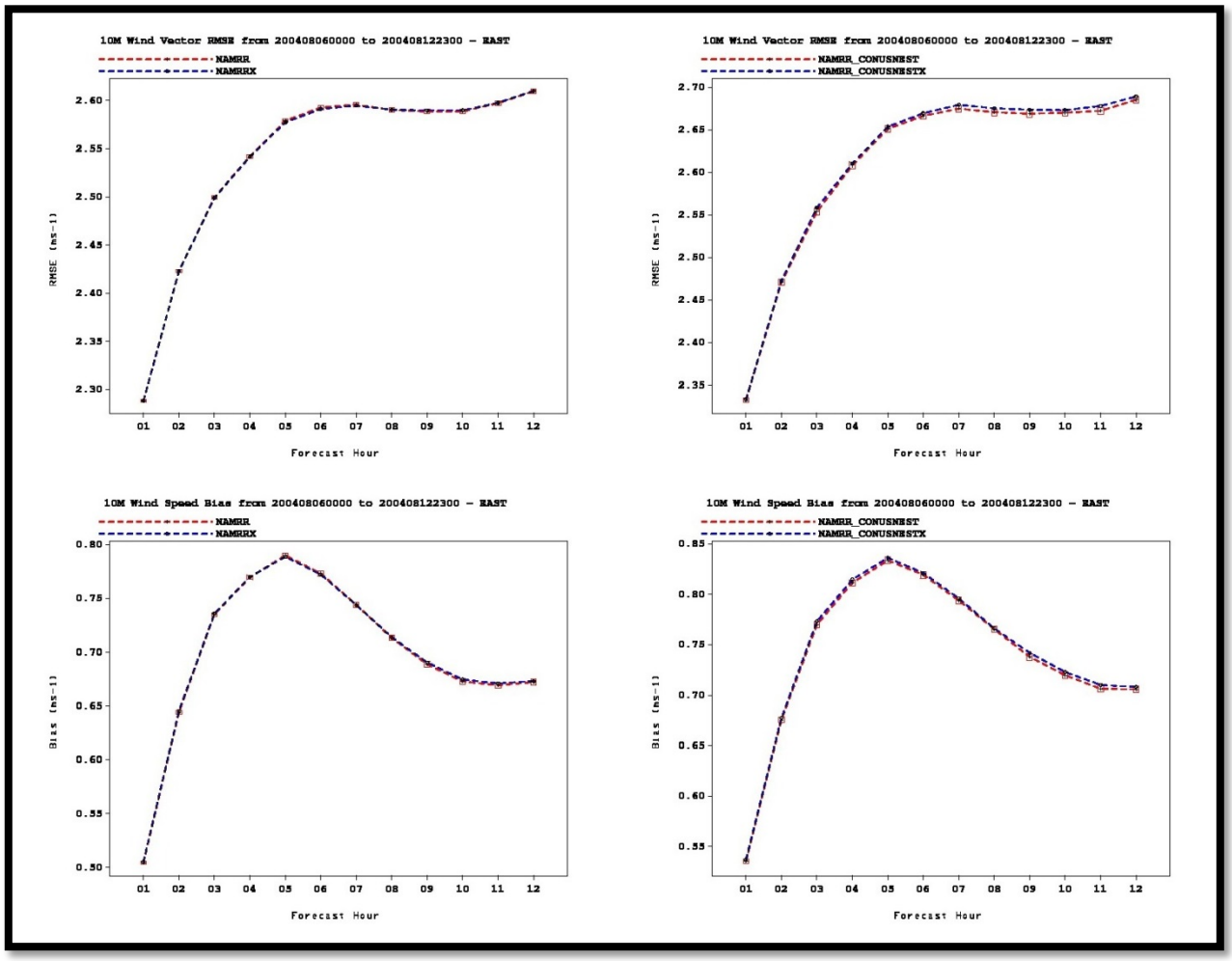


Figure E-4. As in Figure E-2, except for 10-m winds, August period. Note that RMSE is vector RMSE.

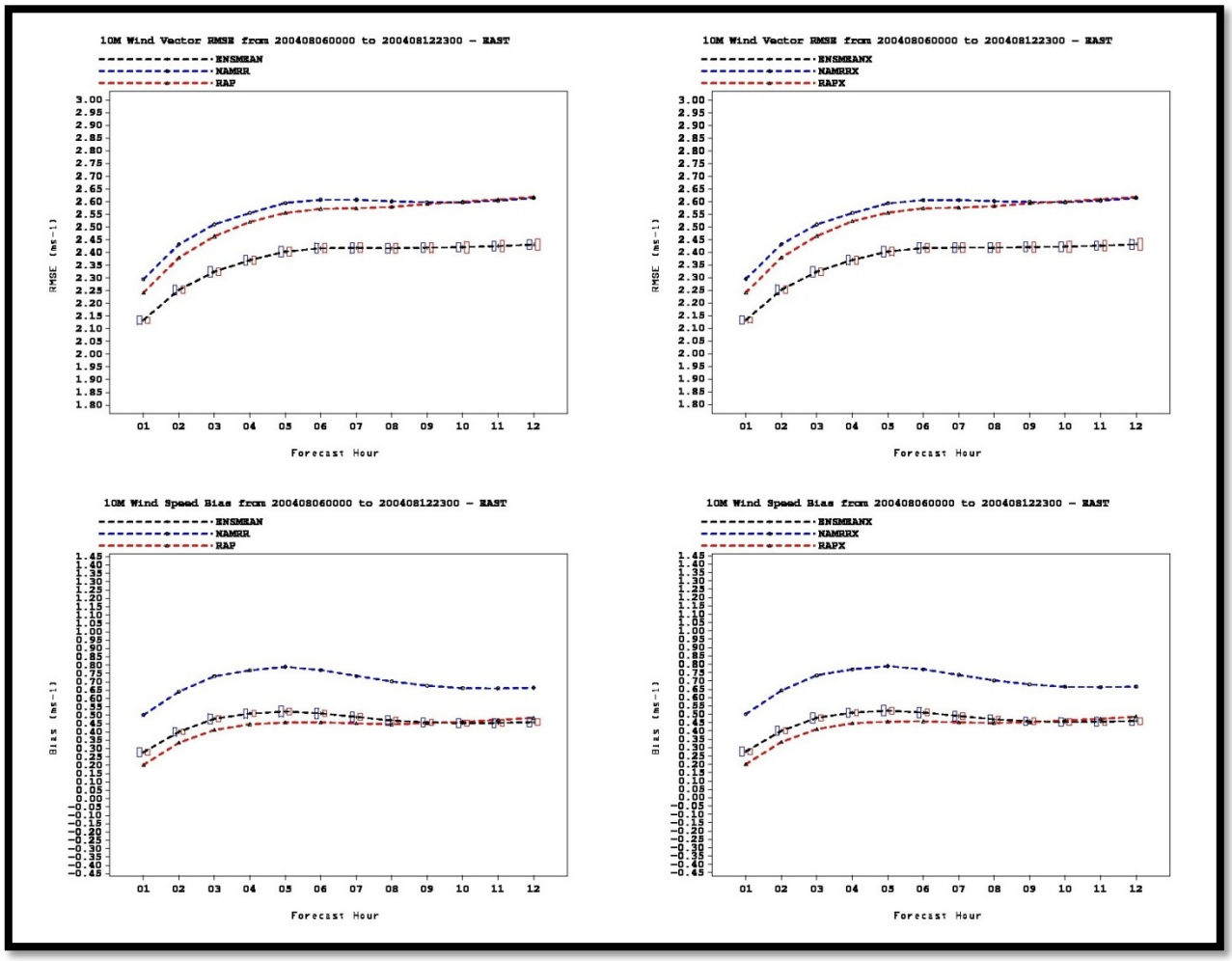


Figure E-5. As in Figure E-3 except for 10-m winds, August period. Note that RMSE is vector RMSE.

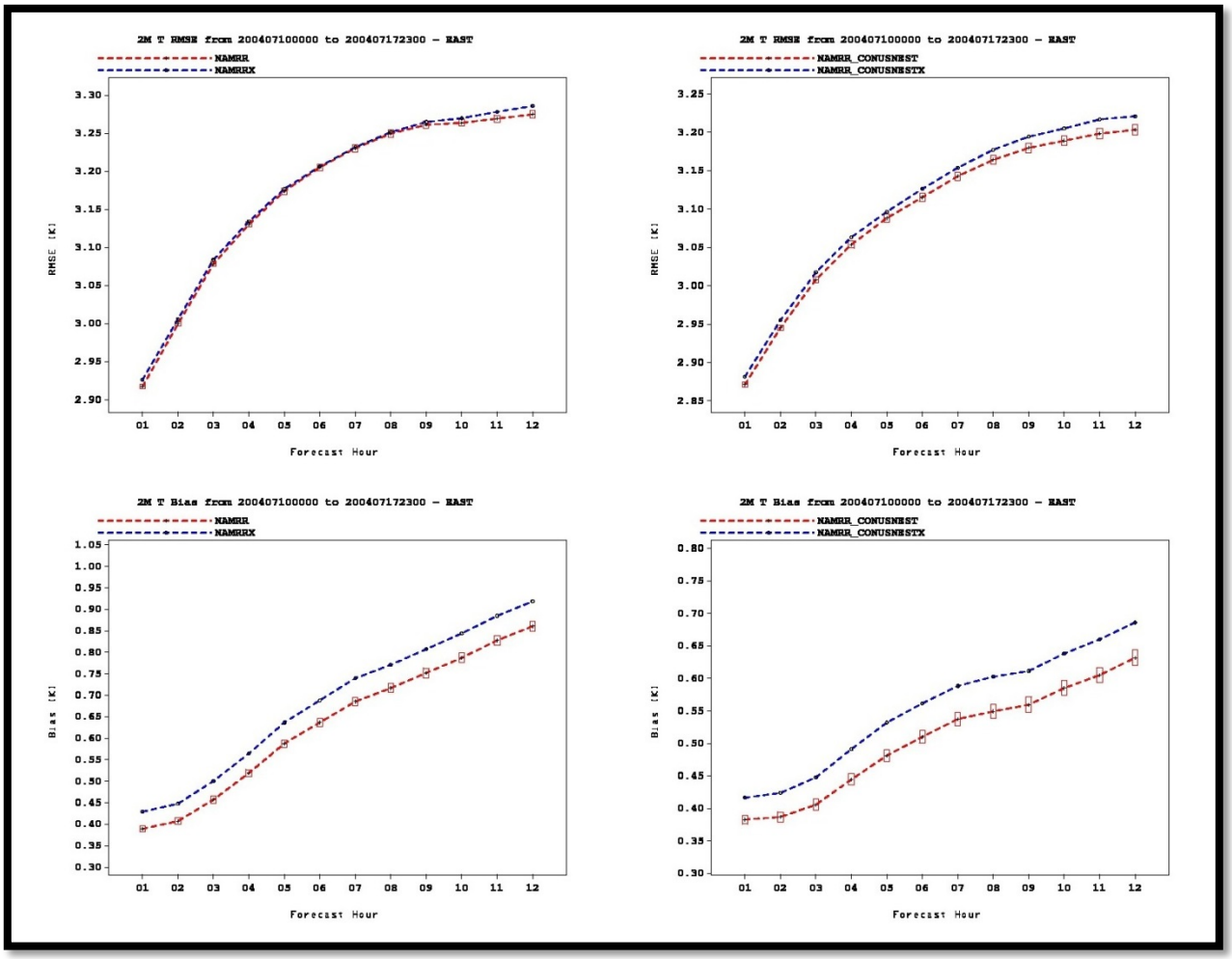


Figure E-6. As in Figure E-2 except for the July study period.

2-m temperature RMSE (top) and bias (bottom) against surface observations over the East (Figure 11) for the July period. Statistics from the 12-km parent domain occupy the left panels and forecasts from the 4-km nest domain. Red traces correspond to the control, and blue traces correspond to the experiment.

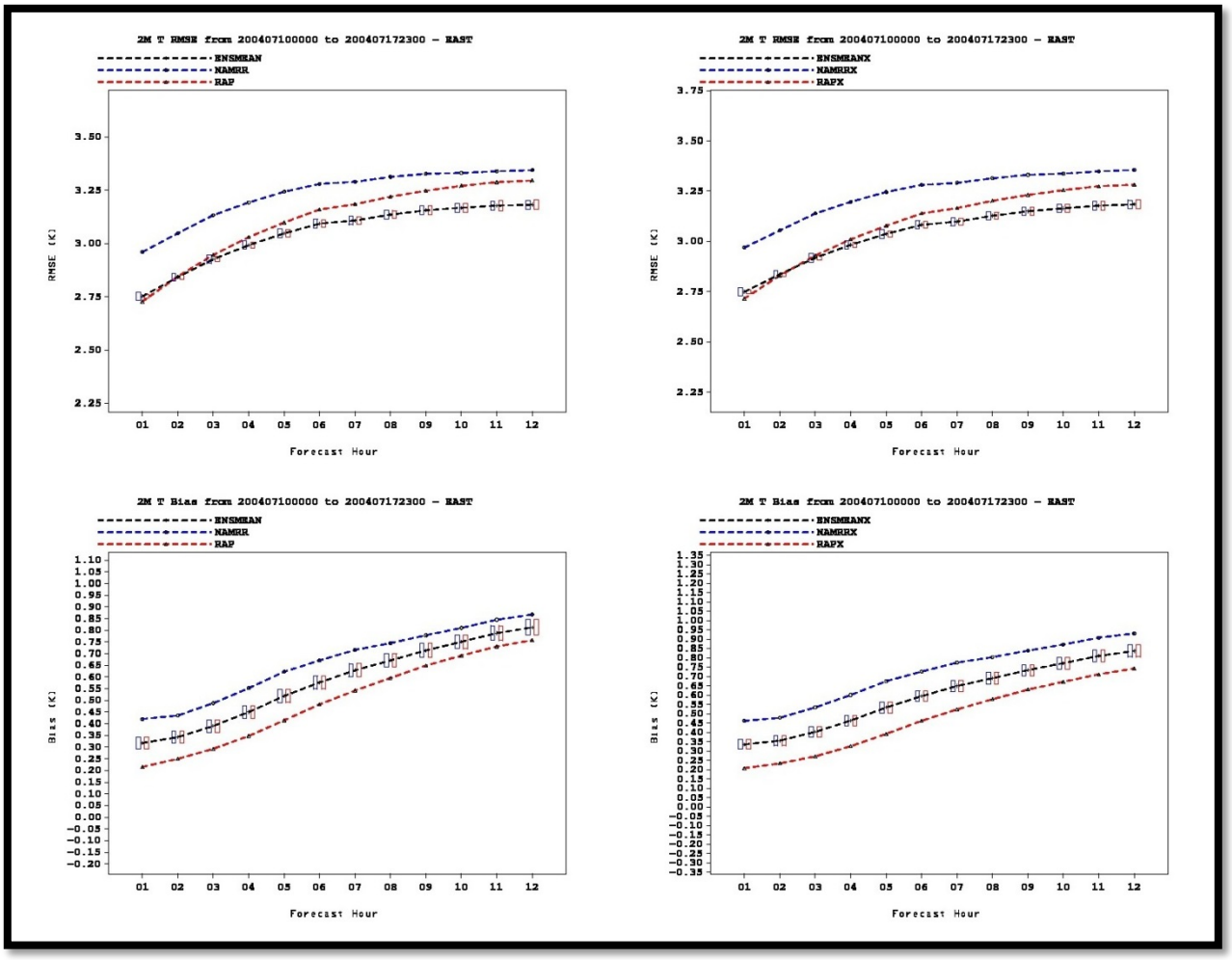


Figure E-7. As in Figure E-3 except for the July study period.

2-m temperature RMSE (top) and bias (bottom) against surface observations over the East (Figure 11) for the July period. Each panel shows statistics comparing RAP (red), NAMRR (blue), and their mean (ENSMEAN; black). The left panel corresponds to the control simulations, and right panel corresponds to the experiment simulations.

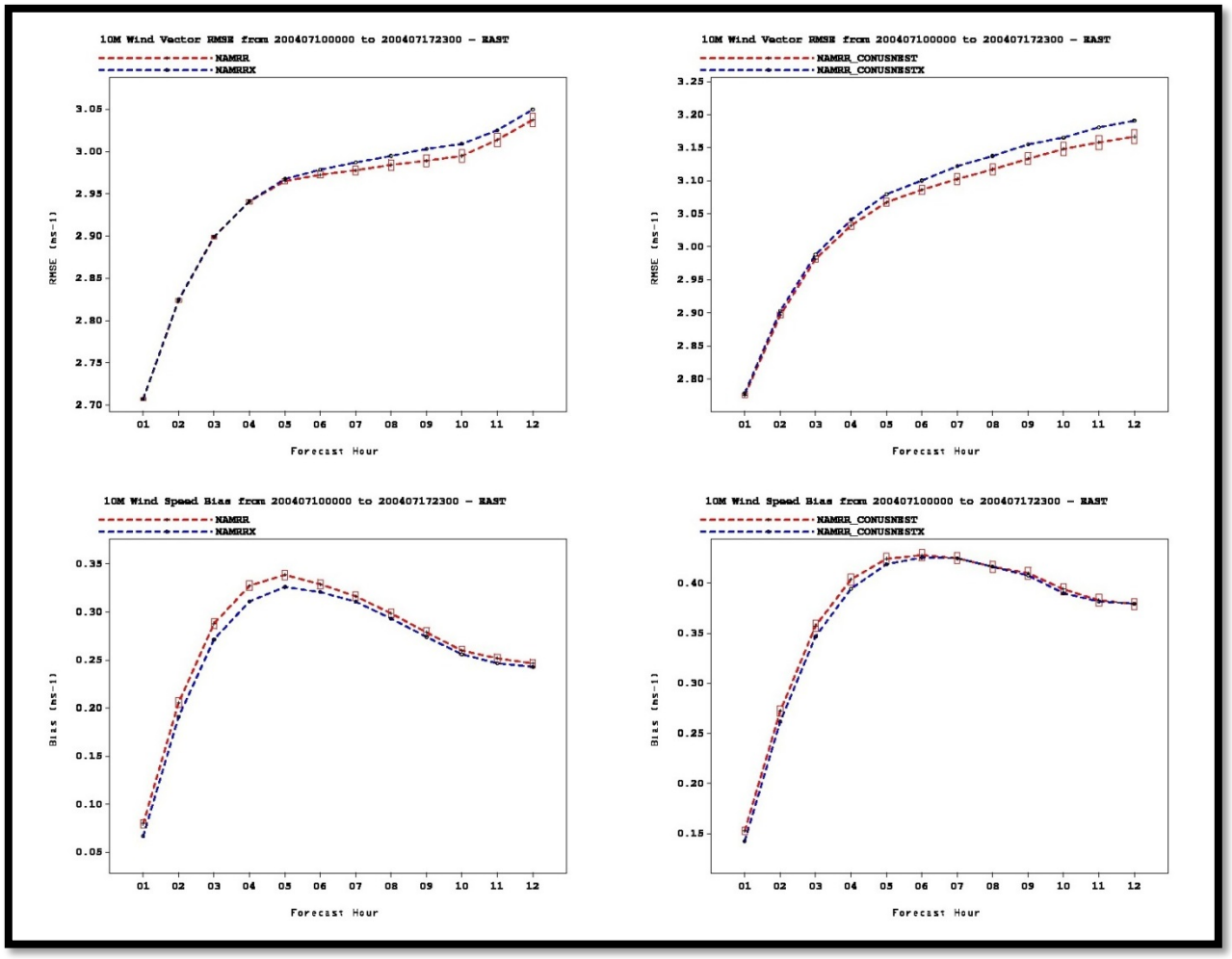


Figure E-8. As in Figure E-6 except for 10-m winds, July. Note that RMSE is vector RMSE.

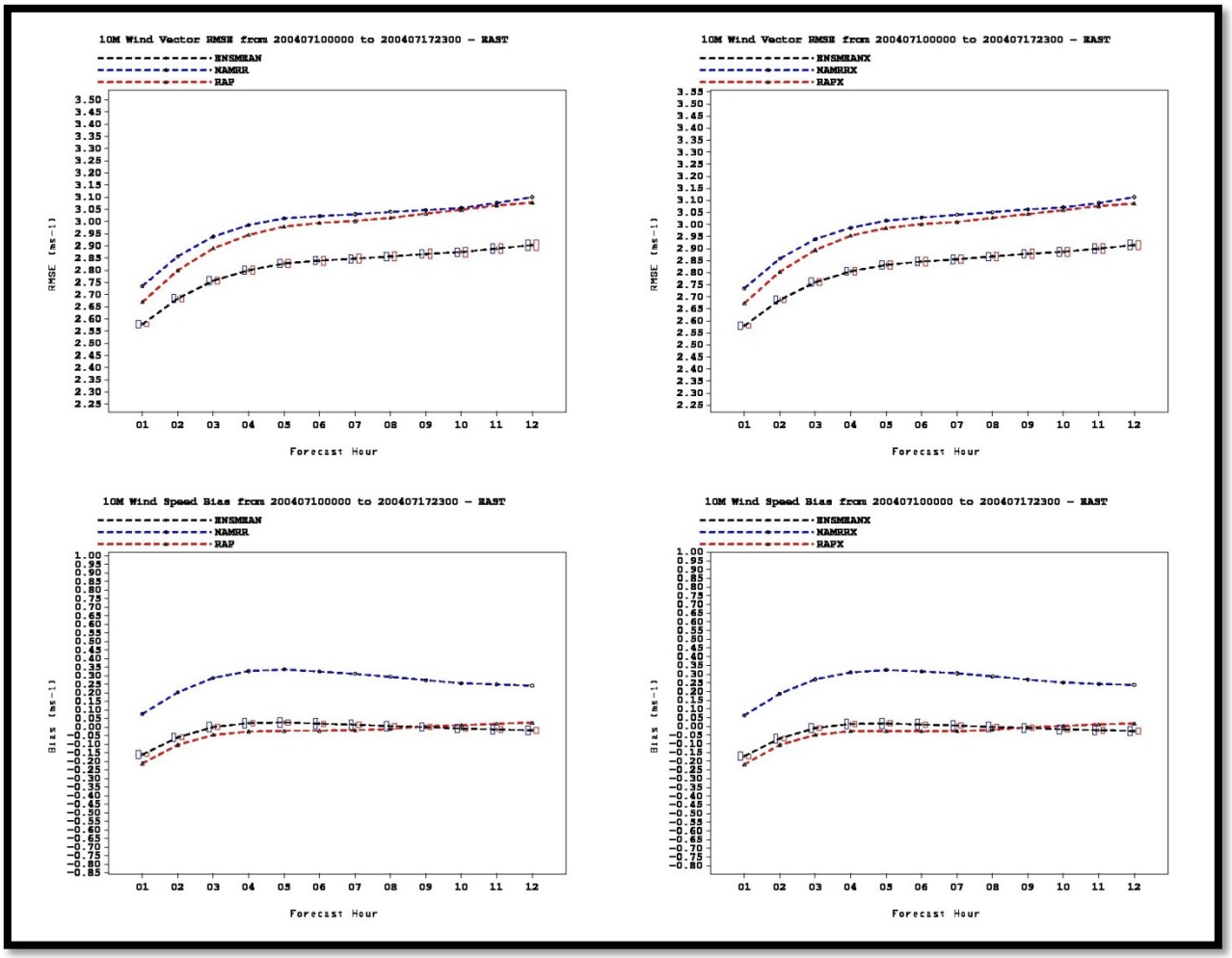


Figure E-9. As in Figure E-7 except for 10-m winds, July. Note that RMSE is vector RMSE.

List of Figures

List of Tables