Assessment of NWP Forecast Models in Simulating Offshore Winds through the Lower Boundary Layer by Measurements from a Ship-Based Scanning Doppler Lidar

Yelena L. Pichugina,^{a,b} Robert M. Banta,^b Joseph B. Olson,^{a,b} Jacob R. Carley,^c Melinda C. Marquis,^b W. Alan Brewer,^b James M. Wilczak,^b Irina Djalalova,^{a,b} Laura Bianco,^{a,b} Eric P. James,^{a,b} Stanley G. Benjamin,^b and Joel Cline^d

^a CIRES, Boulder, Colorado

^b NOAA/Earth System Research Laboratory, Boulder, Colorado ^c I. M. Systems Group, Inc., and NOAA/NWS/Environmental Modeling Center, College Park, Maryland ^d Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, Washington, D.C.

(Manuscript received 30 November 2016, in final form 14 July 2017)

ABSTRACT

Evaluation of model skill in predicting winds over the ocean was performed by comparing retrospective runs of numerical weather prediction (NWP) forecast models to shipborne Doppler lidar measurements in the Gulf of Maine, a potential region for U.S. coastal wind farm development. Deployed on board the NOAA R/V Ronald H. Brown during a 2004 field campaign, the high-resolution Doppler lidar (HRDL) provided accurate motion-compensated wind measurements from the water surface up through several hundred meters of the marine atmospheric boundary layer (MABL). The quality and resolution of the HRDL data allow detailed analysis of wind flow at heights within the rotor layer of modern wind turbines and data on other critical variables to be obtained, such as wind speed and direction shear, turbulence, low-level jet properties, ramp events, and many other wind-energy-relevant aspects of the flow. This study will focus on the quantitative validation of NWP models' wind forecasts within the lower MABL by comparison with HRDL measurements. Validation of two modeling systems rerun in special configurations for these 2004 cases—the hourly updated Rapid Refresh (RAP) system and a special hourly updated version of the North American Mesoscale Forecast System [NAM Rapid Refresh (NAMRR)]-are presented. These models were run at both normal-resolution (RAP, 13 km; NAMRR, 12 km) and high-resolution versions: the NAMRR-CONUS-nest (4 km) and the High-Resolution Rapid Refresh (HRRR, 3 km). Each model was run twice: with (experimental runs) and without (control runs) assimilation of data from 11 wind profiling radars located along the U.S. East Coast. The impact of the additional assimilation of the 11 profilers was estimated by comparing HRDL data to modeled winds from both runs. The results obtained demonstrate the importance of high-resolution lidar measurements to validate NWP models and to better understand what atmospheric conditions may impact the accuracy of wind forecasts in the marine atmospheric boundary layer. Results of this research will also provide a first guess as to the uncertainties of wind resource assessment using NWP models in one of the U.S. offshore areas projected for wind plant development.

1. Introduction

Assessment and improvement of numerical weather prediction (NWP) model skill require accurate profile measurements of meteorological quantities, including wind. Here we use high-precision, high-resolution wind profiles measured by shipborne Doppler lidar during a monthlong research cruise to evaluate the performance of two modeling systems in the marine atmosphere over the Gulf of Maine, an especially difficult environment in which to obtain such measurements. The focus of this research is on using these measurements and models to support the emerging offshore wind energy industry, which has an urgent need for wind data in the turbinerotor layer of the marine atmospheric boundary layer (MABL). Offshore locations off the coast of the United States show great potential for generating electrical power from the wind (Schwartz et al. 2010; Musial and Ram 2010). Direct measurements of wind properties

DOI: 10.1175/MWR-D-16-0442.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Corresponding author: Yelena Pichugina, yelena.pichugina@ noaa.gov



FIG. 1. Google Earth image of the northeastern United States, showing ship tracks for the entire NEAQS-04 campaign, from 9 Jul to 12 Aug 2004 (gray circles). Locations of inland wind profiling radars (white pins).

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 (Elliott et al. 2012; Drechsel et al. 2012).

Because of the lack of appropriate measurements offshore, current estimates of this resource have been obtained from NWP model output (e.g., Schwartz et al. 2010; Musial and Ram 2010; James et al. 2017, manuscript submitted to *Wind Energy*), by extrapolation of shoreline measurements outward over the ocean or other water body (such as the Great Lakes), or by vertical extrapolation from measurements near the water surface. However, the lack of quality offshore profile measurements means that NWP output has not been well validated there and that the extrapolations may lead to significant errors in estimating winds at hub height (Nunalee and Basu 2014; Pichugina et al. 2017).

The high potential of the area off the U.S. East Coast for wind energy (WE) development (Musial and Butterfield 2004) requires measurement campaigns to obtain information on turbine-level winds. Such measurement campaigns offshore are expected to be expensive; therefore, an effort to use existing wind profile datasets was made by Pichugina et al. (2012) to characterize the rotor-layer winds, including their spatial and temporal variability and vertical structure. Results from this research led to the DOE–NOAA collaborative project the Positioning of Offshore Wind Energy Resources (POWER) project (Banta et al. 2014). The present paper describes results from that project. One of the key objectives of the POWER study was to verify hub-height winds predicted by two different NOAA NWP forecast models, to address the need to verify NWP models over the ocean for wind energy and other applications.

The present study uses high-resolution wind profile measurements from NOAA's scanning high-resolution Doppler lidar (HRDL) mounted on a research ship that sailed in the Gulf of Maine during the New England Air Quality Study (NEAQS-04) in the summer of 2004 (Pichugina et al. 2012). The lidar dataset is used to evaluate NWP model skill in simulating offshore winds through and above the turbine-rotor layer. The two modeling systems used for validation were a special hourly updated version of NOAA's North American Mesoscale Forecast System (NAM), known as NAM Rapid Refresh (NAMRR), and the hourly updated Rapid Refresh (RAP) system. The two modeling systems used were run at both normal-resolution ("parent" domain model) and high-resolution versions: the NAMRR-CONUS-nest and the High-Resolution Rapid Refresh (HRRR), respectively.

Hourly measurements of wind speed and direction from an array of 11 coastal and inland 915-MHz wind profiling radars (profilers) were also available as part of the NEAQS-04 experiment, located along the coastal area (Fig. 1). In the POWER experiment, data from these profilers were assimilated into *experimental* model runs to quantify the impact of the additional data on model skill. Runs using all four models, with and without assimilation of data from these coastal and inland profilers, were verified against shipborne lidar measurements.

The paper is organized as follows: Section 2 discusses the NEAQS-04 research cruise, the measurement and modeling systems used in the project, as well as the approaches and metrics used for the model validation. It also discusses the effects of the vertical interpolation and provides examples of two approaches: first, when measurements are interpolated to model output levels (lidar to model); and second, when modeled variables are interpolated to the heights of lidar measurements (model to lidar). Section 3 presents an overview of wind speed and wind direction during two periods selected for model validation. The HRDL data analyses and lidarmodel comparisons are given in section 4. This section also presents results from comparisons of measured and modeled variables. Section 5 provides quantitative results of model validation using HRDL data, and presents statistical metrics between modeled and measured winds along with the estimated impact of assimilation of additional data from inland profilers on model accuracy. Section 5 also discusses meteorological conditions associated with large deviations between measured and modeled variables. Conclusions and recommendations are given in section 6.

2. Measurement and modeling systems

Two air quality field studies, both aimed at characterizing local pollution sources in the New England region, were conducted in the early 2000s (White et al. 2007). The first study was the NEAQS-02 campaign in summer 2002 (Angevine et al. 2006; Darby et al. 2007) and the second study was the NEAQS-04 field campaign (Wolfe et al. 2007; Fairall et al. 2006), the dataset used in the present study. 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 Transformation-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 key processes producing these transports. Land-based, airborne, and shipborne instrumentation contributed to the dataset, as described by Fehsenfeld et al. (2006).

The major offshore measurement platform was the 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. Tracks where the ship traveled during NEAQS-04 are shown in Fig. 1. The ship's remote sensing instrumentation included NOAA's HRDL and a NOAA

915-MHz profiler. Wind profile measurements were also available from radiosondes launched from the deck of the ship every 6 h.

a. HRDL

The 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). HRDL provides precise rangeresolved measurements of the line-of-sight or radial wind, that is, the component of the velocity parallel to the beam, and aerosol backscatter, at a range resolution of 30 m. Deployed on board the RHB, HRDL was operated over the Gulf of Maine 24 h per day during the NEAQS-04 field campaign (Wolfe et al. 2007; Pichugina et al. 2012), providing accurate profiles of wind speed and direction from the deck of the RHB every 15 min. Details of HRDL's adaptations for marine use, such as its motion-compensation system, HRDL technical specifications, accuracy of lidar measurements offshore, and scanning procedures, are described by Pichugina et al. (2012). The HRDL system includes full scanning capability in azimuth (conical), elevation (vertical slice), and staring modes (Banta et al. 2002). Profiles of wind speed and direction used in this study are calculated from azimuth scans at constant elevation (conical scans) using a velocity-azimuth display (VAD) procedure (Browning and Wexler 1968; Banta et al. 2002, 2015). RMS instrument-noise uncertainty for these mean profiles has been determined to be less than $0.1 \,\mathrm{m\,s^{-1}}$ (Grund et al. 2001; Pichugina et al. 2008). Wind flow properties during the cruise, including statistics and distributions of wind speed and wind direction, frequency of low-level jet (LLJ) occurrence, and wind shear across the turbine-rotor layer, obtained from all available Doppler lidar measurements during the NEAQS-04 experiment, are given in Pichugina et al. (2012, 2017) as examples of longer-term averages.

An analysis of coincident HRDL and radiosonde data from the entire experiment shows good agreement between rawinsondes and HRDL horizontal wind components, with correlation coefficients of 0.97 and 0.98 for the two components, for all heights above 100 m (Wolfe et al. 2007). High correlation ($R^2 > 0.98$) between the two instruments was also shown in Pichugina et al. (2017) for each individual height of measurement from 10 up to 2000 m above the water surface. The correlation is reduced below 100 m as a result of the influence of the ship's atmospheric wake on the rawinsonde measurements at these levels, whereas HRDL profiles are obtained from conical scan data (as just described) that sample well outside this wake. Because of the low frequency (every 6h) and uncertainty of measurements within the turbine-rotor layer of the atmosphere (taken here to be roughly 50–150 m), rawinsonde measurements are not used in the present study for model evaluation.

b. Profilers: Wind profiling radars

During NEAQS-04, hourly profiles of wind speed and direction were also obtained from the ship-based 915-MHz profiler (33-cm wavelength: Carter el al. 1995; Strauch et al. 1984; Wilczak et al. 1996), which could be operated in two modes, high resolution and low resolution. In the high-resolution mode, the first available measurement height was 216m with a vertical measurement step of 58 m; in the low-resolution mode, the lowest measurement height was 310 m, with vertical spacing of 101 m (White et al. 2007). Detailed analysis and comparison between ship-based HRDL and profiler measurements show a reasonable agreement between the two instruments, with slightly increased scatter and reduced height coverage for the high-resolution mode "as might be expected because of lower transmitted power and therefore lower return signal in this mode" (Wolfe et al. 2007, section 3).

Eleven land-based profilers were deployed along the U.S. East Coast and other locations in the northeastern United States during the summer of 2004 for the NEAQS-04 experiment (Fig. 1). These profilers were also 915-MHz radars, having vertical resolutions of $\sim 60 \,\mathrm{m}$ (high-resolution mode) and $\sim 100 \,\mathrm{m}$ (low-resolution mode). The maximum height with detectable signal varied with atmospheric conditions (a stronger backscatter signal occurs in a moister, more turbulent atmosphere), but the coverage typically ranged from the lowest level up to around 1.5 km above ground level (AGL) for the highresolution mode and up to around 4 km AGL for the lowresolution mode. As with the RHB profiler, the lowest height of measurement was different for high- and lowresolution modes and varied among the profilers used: 70-190 m for the high-resolution mode and 100-300 m for the low-resolution mode.¹ A listing of the first available height and vertical resolution of all 11 land-based and shipborne profiler measurements used during NEAQS-04 is provided in Banta et al. (2014).

c. NOAA model forecast systems used in the study

In this study we used two analog versions of NWP models run at the NOAA/National Centers for Environmental Prediction (NCEP) and the Earth System Research Laboratory (ESRL) Global Systems Division (GSD). These are 1) an experimental hourly updated version of the North American Mesoscale Forecast System (NAMRR; Carley et al. 2015; Rogers et al. 2009) model and its finer-resolution nest, the NAMRR-CONUS-nest; and 2) the hourly updated RAP model (Benjamin et al. 2004, 2016) and its embedded HRRR model. Operational renditions of these models at NCEP provide foundational meteorological predictions at time and space scales useful to the wind energy industry. The two versions are based on different model frameworks: the NAMRR models were derived from the Nonhydrostatic Multiscale Model on the B grid (NMMB; e.g., Janjić 2003; Janjić and Gall 2012) and the RAP-HRRR model came from the Advanced Research version of the Weather Research and Forecasting (WRF-ARW) Model.

Maps showing the model domains used in this study are given in Fig. 2. The NAMRR and the RAP were run at horizontal grid intervals of 12 and 13 km, respectively, and for the finer-mesh sizes, the NAMRR-CONUS-nest had a grid interval of 4 km and the HRRR had 3-km grid spacing (Banta et al. 2014). Tables 1 and 2 give an overview of the model configurations and physical parameterizations used in these 2012 versions.

For initialization, the NAMRR parent/nest and the RAP are updated hourly using the 3DVAR algorithm of the Gridpoint Statistical Interpolation analysis system (GSI) (Wu et al. 2002). Both NWP systems employ a partial-cycling procedure (Rogers et al. 2009) in POWER that fully cycles the land states and includes a regular reinitialization and spinup of the atmospheric state from a global model [e.g., the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) for POWER]. Specific information on the data assimilation (DA) configurations for RAP, HRRR, and NAMRR parent/nest for the POWER experiments are described in detail by Banta et al. (2014). Further information for the cycling DA system used for the operational RAP and HRRR is described in Benjamin et al. (2016). All NWP systems in POWER produce hourly forecasts out to 18h.

A truncated HRRR domain was used for this POWER intercomparison centered over the northeastern United States (Fig. 2; Banta et al. 2014, 2017, manuscript submitted to *Bull. Amer. Meteor. Soc.*, hereafter B17). The model physics and data assimilation used in these RAP and HRRR versions for this study (Table 2) corresponded to the 2012 versions of the RAP version 1 model/assimilation options (see Benjamin et al. 2016). For example, current features, such as the hybrid ensemble– variational data assimilation, were not available for

¹ More information on the technical parameters of the NOAA profilers can be found at http://www.esrl.noaa.gov/psd/data/obs/ instruments/WindProfilerDescription.html.



FIG. 2. Google maps of model domains: (a) NAMRR (orange) and NAMRR-CONUS-nest (cyan), and (b) RAP (orange) and HRRR (cyan).

RAP/HRRR or NAMRR/NAMRR-CONUS-nest, nor was any assimilation of radar reflectivity data. These trial versions used for the POWER project are referred to as the RAP2012-P and HRRR2012-P, respectively, in B17, but we will refer to them as RAP and HRRR, respectively, for simplicity.

In this study model values were stored hourly on the hour. We used hourly averaged HRDL wind profile data to validate output wind profiles from retrospective runs. A second objective was to assess the impact of assimilating the hourly profiler data on model skill offshore. This was done by comparing the differences in model errors for the retrospective runs, using HRDL profile data as a reference, for output model wind values without (control) versus with (experimental) assimilation of the data from the 11 coastal and inland profilers. The selected periods for these retrospective models runs are described in section 3. For comparisons of modeled versus HRDL-measured winds, the gridded model wind values were either extracted for the nearest model grid point to the ship location (NAMRR and nest) or interpolated horizontally to the ship position using a parabolic interpolation scheme (RAP-HRRR). The former is the same procedure used operationally to generate model soundings (i.e., "BUFR soundings") at a desired location. These values were then linearly interpolated vertically to the lidar measurement heights. It is worth noting here that the ship-location profile data

TABLE 1. The 12-km NAMRR and 4-km NAMRR-CONUS-nest of	domain configurations (Banta et al. 2014); CONUS	, conterminous
United	States.		

12-km NAMRR parent description	Configuration			
Points in x, y, z directions	954, 835, 60			
Time step (s)	$26.66\overline{6}$			
Microphysics parameterization	Ferrier et al. (2002, 2011)			
Boundary layer parameterization	Janjić (2001)			
Convective parameterization	Janjić (1994)			
Longwave/shortwave radiation parameterization	Iacono et al. (2008)/Mlawer et al. (1997)			
Land surface model	Ek et al. (2003)			
Gravity wave drag parameterization	Alpert (2004)			
4-km CONUS nest description	Configuration			
Points in x, y, z directions	1371, 1100, 60			
Time step (s)	$8.88\overline{8}$			
Convective parameterization	Janjić (1994): Modified to be less active for higher resolution			
Gravity wave drag parameterization	None			

13-km RAP description, CONUS domain reduced from whole domain	Configuration		
Points in x, y, z directions	758, 567, 51		
Time step (s)	60		
Cloud microphysics parameterization	Thompson et al. (2008)		
Boundary layer parameterization	Janjić (2001)		
Convective parameterization	Grell 3D/Grell shallow cumulus		
Longwave/shortwave radiation parameterization	Chou and Suarez (1994)/Mlawer et al. (1997)		
Land surface model	Smirnova et al. (1997, 2000)		
3-km HRRR description reduced from the whole domain	Configuration		
Points in x, y, z directions	520, 450, 51		
Time step (s)	20		
Convective parameterization	Turned off		

TABLE 2. The 13-km RAP and 3-km HRRR domain configurations for POWER (from Banta et al. 2014, B17).

were extracted before they were provided to us, so we had no control over how they were handled. We have done sensitivity tests with another subsequent dataset over land and found that using different techniques in this way produced only small differences between the methods.

1) EFFECT OF VERTICAL INTERPOLATION

To evaluate model performance using lidar data and to obtain comparison statistics, modeled and observed wind flow variables need to be interpolated to the same heights.

For this analysis, measurement data could be interpolated to model grid heights, or model values to measurement heights. If measured data were at coarser resolution relative to the model's resolution, such as the wind profiling radar (WPR) and NAMRR in Fig. 3, then either technique may yield similar results. But if the measurements are at much finer resolution than the model's resolution, such as the lidar data in Fig. 3, then this finescale information should be used in calculating model errors through a layer. Such "model to lidar"level interpolation evaluates the model against the detailed atmospheric structure observed and provides more data points within a layer for better error statistics.

Examples of these interpolation approaches are illustrated in Figs. 4 and 5 using HRRR model output for illustration (NAMRR showed the same behavior). Time-height cross sections of wind speed in Fig. 4 are shown for lidar data (Figs. 4a and 4d) and for HRRR values (Figs. 4b and 4e). The left panels are all at lidar resolution, and the right panels show the coarser model resolution. Figure 4a illustrates the detailed atmospheric structure measured by HRDL, and Fig. 4e shows the HRRR model initial (hour 0) simulated winds for the same period. When the coarse model output is interpolated to the lidar levels (Fig. 4b), the plot looks smoother but no new information on atmospheric vertical structure is added. But when the finescale lidar data are averaged to the coarse resolution of the model (Fig. 4d), considerable vertical structure is lost. Figures 4c and 4f show the model-minus-lidar differences (biases), or errors, calculated at the finescale lidar levels (Fig. 4c) and at the coarser-scale model grid heights (Fig. 4f). Careful inspection of these two panels reveals that the vertical structure and magnitudes of these errors differ between the two plots. We conclude that employing the high-resolution data of the left panels in calculating model error is more appropriate than using the coarse data from the right panels, as this



FIG. 3. Heights used for the mean wind profile from (a) shipborne measurements in the first 600 m MSL, and (b)–(e) vertical levels of wind output from NWP models used in the study. Rotor layer of 50-150 m (gray bar). Heights of shipborne profiler measurements for (f) high-resolution mode (solid line) and (g) low-resolution mode (dashed line).



FIG. 4. Time-height cross sections of wind speeds and model bias ($m s^{-1}$, color scales at top) and direction (arrows) on 11 Aug, shown up to 1.5 km MSL. Values displayed at (left) finescale HRDL resolution and (right) coarser-resolution model heights. (a) HRDL data shown at actual finescale lidar height levels. (b) Model initial values interpolated to lidar levels. (c) Difference (bias) between interpolated model values and HRDL-measured wind speeds (middle panel minus upper panel) portrayed at finer lidar resolution. (d) HRDL speeds averaged to model grid intervals. (e) HRRR-modeled wind speed and direction on HRRR's vertical grid. (f) Difference between model and lidar on HRRR grid.



FIG. 5. Examples of bias ($WS_{MODEL} - WS_{lidar}$) in wind speed profiles measured by lidar and modeled by HRRR for 1000–1500 UTC 11 Aug 2004. Modeled profiles are shown for the initial time (forecast 0). Bias in observed and modeled winds when the lidar-to-model interpolation approach is used (red lines); heights of model output (red circles). Bias in observed and modeled winds when the model-to-lidar interpolation approach is used (blue lines); heights of lidar measurements (blue diamonds). The turbine-rotor layer of 50–150 m is indicated (horizontal dotted lines).

allows finescale structure of the lidar data, with such features as thin shear layers and LLJ noses, to be accounted for in the error computations. These plots also suggest that increasing the vertical resolution of the simulations should have a greater impact on improving model skill than increasing horizontal resolution, especially over the ocean, where horizontal gradients are gentler than over many land surfaces.

To further illustrate details of the interpolation, Fig. 5 shows profiles of the difference (bias) between measured and HRRR-modeled wind speed profiles in the lower 300 m AGL for 6 h (1000–1500 UTC) during the same day as in Fig. 4. In general the "lidar to model" lines (red) fall near the model-to-lidar points (blue), but discrepancies of up to 0.5 m s^{-1} can be seen. These differences are not taken into consideration when only model-level data are used in the lidar-to-model case, such as when calculating mean statistics over a vertical layer. Similar analyses for other days (and for the NAMRR) show that in most cases models

underestimate wind speed (negative bias) for both approaches, but occasionally a positive bias was also observed (such as for hour 11; Fig. 5, above 100 m). In general—and not surprisingly—models with larger vertical grid spacing exhibit larger discrepancies between the two methods.

Hereafter, model output values are interpolated to lidar heights, to allow us to compare errors for different models at a single height (such as turbine hub height) or through layers of interest and to capture the true, measured variability of the winds in finer detail. The increased number of points within the averaging layer also provides more robust validation statistics, as mentioned.

2) EFFECTS OF AVERAGING DEPTH

Averaging over deeper layers should produce better model-measurement agreement, but what is the magnitude of this improvement? The POWER dataset is well suited to quantify this effect, consisting (as it does)



FIG. 6. Error statistics between lidar-measured and modeled wind speed are shown for the August study period as a function of forecast lead time. (from left to right) Coefficient of determination R^2 , RMSE, and model RMSE improvement as a result of the assimilation of coastal profiler data. Winds from NAMRR (red) and NAMRR-CONUS-nest (blue) for experimental (solid) and control (dashed) runs. Statistics computed at a height of (a) 100 m, (b) through the rotor layer of 50–150 m, (c) through the lowest 0.3 km MSL, and (d) through the lowest 1 km MSL.

of both detailed profile measurements and highresolution model output. Coefficients of determination R^2 and root-mean-square error (RMSE) as a function of forecast lead time for the NAMRR-CONUS-nest and NAMRR models during the August period are compared in Fig. 6 for winds at hub height (here, taken to be 100 m) and for mean values over three distinct vertical-layer depths: the turbine-rotor layer (here, 50–150 m), the layer from 10 to 300 m, and the layer from 10 m to 1 km MSL. The number of points involved in the RMSE calculations was 168 at hub height, 1176 in the rotor layer, 2688 in the lowest 300 m, and 6048 in the lowest 1000 m. These numbers are for lead hour 0, and they slightly decrease by lead hour 9 as a result of occasional gaps in the data. RAP/ HRRR output could just as well have been used for this illustration without affecting the conclusions.

The R^2 (left) and RMSE (middle) columns of Fig. 6 demonstrate better agreement between measured and



FIG. 7. RMSE between measured and NAMRR and NAMRR-CONUS-nest modeled winds are shown as a function of forecast hour out to 9 h. Results in all panels are shown for hub height and for mean values over several vertical layers according to the legend in the bottom-right plot. (top) Control runs and (bottom) experimental runs of both models.

modeled winds when averaged over a deeper layer (bottom panels) than over shallower layers. Deeperlayer data indicate smaller RMSEs by $\sim 0.5 \,\mathrm{m\,s^{-1}}$ ($\sim 20\%$) and larger coefficients of determination (0.9 vs 0.8) compared with these statistics found for the turbinerotor layer or for the hub height.

The rightmost column in Fig. 6 shows the improvement of the model forecast as a function of forecast lead time as a result of assimilation of additional data from coastal profilers, which will be further investigated in section 5. The improvement was computed as a difference in RMSE between measured winds and those modeled by the experimental and control runs (RMSE_{EXP} and RMSE_{CNTR}, respectively), normalized by RMSE_{CNTR}:

$$Improvement(\%) = 100 \times (RMSE_{CNTR} - RMSE_{EXP})/RMSE_{CNTR}.$$
 (1)

Similar to the other statistics, larger improvement was found for averages over deeper layers. For example, up to 10% improvement was found early in the simulations and positive values for up to seven forecast hours for the 1-km MSL layer (bottom panel) compared to <4%improvement found for as few as 2h (NAMRR-CONUS-nest) for the hub-height winds (top panel).

To better visualize the differences in statistics found for different layers, including hub height, these statistics are plotted together in Figs. 7 and 8. Figure 7 shows that the most exaggerated differences between RMSE statistics for the different layers occurred at the initial time of the forecast, and these differences decreased with forecast lead time. A clear stratification is also shown in Fig. 8 for R^2 , when the R^2 is calculated for each layer. Deeper layers showed larger correlations (R^2), but unlike RMSE, this trend tended to persist through the 9-h forecast period shown.

Assessment of the model accuracy later in section 5 will be performed for the 10–500-m MSL layer (red lines in Figs. 7 and 8), which is important for wind turbine operations. It was shown (Pichugina et al. 2017) that the majority of LLJs during NEAQS-04 occurred there, producing shear through the turbine-rotor layer. Shear and shear-generated turbulence in the rotor layer can be strong enough to increase turbine loads and adversely affect hardware and operations (Banta et al. 2006; Kelley et al. 2004).

3. Periods selected for model validation

Two study periods, limited to one week each by computer-resource constraints, were selected from the NEAQS-04 dataset for model validation. The selection of these periods was primarily based on the availability of HRDL measurements for several consecutive days,



preferably for a week. The first study period selected was 6-12 August, corresponding to the longest lull between frontal passages according to White et al. (2007). The second study period (10–17 July) includes a day of rain on 14 July; however, model runs were conducted over the 8-day range to avoid having to restart the run on 15 July. Although the intervening rainy day was not originally intended for analysis, we have included this period in this study to assess the ability of the models to predict atmospheric events associated with a transient mesoscale cyclonic storm system. In addition to the continuity of available measurements, other factors considered for the selection of study periods included a variety of wind flow conditions, the presence of LLJs, and the ship-track pattern, as a mixture of tracks close to the shore and farther out to sea was desirable. Two days of measurements (13 and 16 July), when the ship was stationary for several hours, were included to obtain comparison statistics free of spatial variability. Ship tracks for the two study periods are shown in Fig. 9.

Overview of wind speed and wind direction during selected periods

Sequential 15-min HRDL-measured wind profiles were combined into time-height plots to provide an overview of the range and diurnal variability of wind flow conditions for all days in the selected periods. Figure 10 shows a time-height cross section of the lowest 500 m of profile data for 7 days each of the July (top) and August (bottom) study periods. Wind speed values are color coded and plotted as a function of time (UTC) for each day, and wind direction is shown by arrows. These plots illustrate temporal and vertical variability of the wind flow, which result from time-dependent changes in the flows, from either the spatial variability of the flows, or from a combination of both. Stronger winds and LLJ events within turbine-rotor heights and above are evident on 16–17 July and 10–12 August.

Profiles of wind speed and direction in these plots were computed from the conical 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 are associated with times when lidar measurements are unavailable because of thick fog and precipitation, such as can be seen in Fig. 10a between 1700 UTC 13 July and 1400 UTC 15 July. Many episodes of high wind speed shear across the presumed turbine-rotor layer (between the horizontal parallel black lines) can be seen during both periods. Because of the variable vertical structure of winds, estimating wind resources based on nearsurface measurements obtained from buoys or occasional ships can lead to significant errors (Pichugina et al. 2012, 2017).

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 (Pichugina et al. 2012). The variability of hub-height (100 m) wind speed along



FIG. 9. Ship tracks during the two study periods: (a) 10-17 Jul and (b) 6-12 Aug. Tracks for each day are shown by the color according to the legend in the upper-left corner of each plot. The white rectangles represent an area of $241 \text{ km} \times 250 \text{ km}$ with the following coordinates at the corners: NW (71°N, 44°W), SW (68°N, 44°W), NE (71°N, 41.5°W), and SE (68°N, 41.5°W). Each circle represents the location of a lidar-measured profile of wind speed and direction along ship tracks.

ship tracks during each selected period is shown in Fig. 11. Plots like these for the entire cruise are presented by Pichugina et al. (2017). Figures 10 and 11 illustrate the forecasting challenge posed by the observed temporal and spatial variability of winds, and show the need for offshore profile measurements through the MABL to validate models.

Distributions of wind speed and direction in the rotor layer are shown in Fig. 12 for all hours (gray bars), for daytime/evening transition hours (1500-2400 UTC, red), and for nighttime/morning transition hours (0000-1500 UTC, blue). Wind speeds during both periods varied from 0 to $15 \,\mathrm{m\,s^{-1}}$, with distributions slightly shifted toward lower wind speeds for daytime hours and toward stronger winds at night. Wind directions also varied widely, with distinct westerly and southwesterly modes. Westerly and southwesterly winds in the Gulf of Maine were common in summer 2004 (Angevine et al. 2006; Fairall et al. 2006; Pichugina et al. 2012). However, some periods of northerly and northwesterly winds were also observed (see Fig. 10 for 11 July and 9 August), producing a second mode in wind direction histograms (Fig. 12d).

Mean statistics of these distributions (Table 3) show stronger nighttime winds during both periods compared to daytime hours. Since the HRDL measurements during NEAQS-04 are the only available high-resolution measurements offshore, we added the three bottom rows to Table 3 to show percentages of wind speeds in the rotor layer of a 1.5-MW wind turbine below the nominal cut-in threshold ($0-4 \text{ m s}^{-1}$), between the cut-in and rated winds (4–12 m s⁻¹), and greater than rated speeds (12–25 m s⁻¹). The results indicate wind conditions that would be favorable for turbine operations in \sim 74%–85% of the time for both periods. Obviously



FIG. 10. Time-height cross sections of lidar-measured wind speed (color bar, scaled from 0 to 16 m s^{-1}) and direction (arrows: up = north), computed from HRDL measurements for the (top) July and (bottom) August 2004 study periods. The vertical axis is height above sea level (m), and the horizontal axis show days from each selected period. The presumed turbine-rotor layer between 50 and 150 m is indicated (horizontal dashed lines).



FIG. 11. Lidar-measured wind speed at 100 m along ship tracks during (a) 10–17 Jul and (b) 6–12 Aug. Winds are color coded from 0 to 15 m s^{-1} according to the color bar at the top of the figure.

longer-term measurements are needed to provide stable statistics for periods when winds are less than 4 m s^{-1} and turbines will not operate, or when winds exceed cutoff speeds larger than 25 m s^{-1} (we did not observe such strong winds during the entire cruise).

4. Direct comparisons of measured and modeled winds

Figures 13-15 show examples of direct comparisons of observed and modeled winds without vertical interpolation using the NAMRR models, noting that corresponding RAP-HRRR examples behave similarly. These comparisons are shown for the initial conditions (lead hour 0) to investigate how well models agree with the measurements. Sample profiles of measured and NAMRR-modeled wind profiles are shown in Fig. 13, where black curves represent lidar data, and NAMRR-CONUS-nest and NAMRR model values are shown by red and blue colors, respectively. Solid red and blue lines in this figure represent model experimental (profiler assimilation) runs, and dashed lines show winds from the control runs. In these cases one sees large lidar-model discrepancies associated with LLJ "noses" or maxima, which were more prevalent during nighttime hours (Pichugina et al. 2017), producing large wind speed errors in the turbine-rotor layer at speeds that would produce even larger relative errors in predicted power production, as also found by B17. Analysis of all hourly averaged profiles from both periods, similar to those shown in Fig. 13, shows better for periods of weak or moderate wind speeds without LLJ structure (Banta et al. 2014).

In many cases the models underestimate the observed wind speeds, as in the top panels of Fig. 13 below 500 m. In other cases the models overestimate the wind speeds (e.g., Fig. 13, top panels, above 500 m). Averaged over longer-term periods (Fig. 14), positive and negative deviations can compensate, leading to better agreement between observed and modeled mean wind speed profiles; for example, Fig. 14 shows the means for the August period. Standard deviation bars indicate variability for the weeklong sample, reflecting the diversity of wind conditions encountered during this study period.

The larger deviations below 200 m during nighttime hours (Fig. 14, middle panels) are most likely due to LLJs, which produce wind maxima in this layer (see Fig. 13; Pichugina et al. 2017) but which get smoothed out of the mean profiles. As shown in Fig. 14 (middle panels), winds from the experimental runs at night agree better with measurements than winds from the control runs for those hours, illustrating the effectiveness of assimilating the landbased profiler data. Quantitative assessment of the impact of profiler data assimilation is further investigated in section 5.

Time series of modeled winds at the fourth level of all models and those measured by HRDL at the closest height to this level are shown in Fig. 15 for all days in the



FIG. 12. Distribution of rotor layer (50-150 m) wind (a),(c) speed and (b),(d) direction from lidar data during two periods selected for the study: (top) 10–17 Jul and (bottom) 6–12 Aug. All data (gray), and data selected for daytime/evening transition (1500–2400 UTC, red) and nighttime/morning transition (0000–1500 UTC, blue) hours. The total number of occurrences in each bin is indicated along the left vertical axis of the histogram, and the percentage of these occurrences is shown along the right vertical axis. Mean and median values of each distribution are shown in Table 3.

August period. These examples show reasonable agreement in wind speed except for some episodes of stronger winds, such as in 10–11 August (see also Fig. 10). A small contribution to the differences in

deviations of modeled winds during episodes of larger vertical wind shear may be due to height differences of model output used for these plots: RAP at 165.8 m, HRRR at 164.3 m, and lidar at 164.2 m; and NAMRR at

TABLE 3. Mean values of turbine-rotor-layer wind speed and wind direction distributions from lidar.

Variables	Mean values	10–17 Jul		6–12 Aug			
		All hours	Daytime	Nighttime	All hours	Daytime	Nighttime
Wind speed $(m s^{-1})$	Mean	7.42	5.99	8.98	7.31	6.77	7.83
	Median	7.28	6.27	8.88	7.43	6.16	7.97
	Std dev	3.32	2.87	3.07	3.19	3.31	2.98
Wind direction (° from N)	Mean	205.5	189.7	223.0	249.3	247.8	250.8
	Median	211.0	196.7	222.9	256.5	252.6	257.0
	Std dev	70.2	67.2	69.6	46.4	47.5	45.3
Speed range (m s ^{-1})		Wind speeds in the rotor layer (%)					
<4		15.6	23.7	6.3	12.5	15.2	9.8
4–12		73.2	69.8	77.1	78.2	76.8	79.6
12–25	11.2	6.5	16.6	9.3	8.0	10.5	



FIG. 13. Examples of hourly averaged lidar-measured and modeled wind profiles on (top) 16 Jul and (bottom) 9 Aug for lidar profiles (black), and NAMRR-CONUS-nest (red) and NAMRR (blue). Experimental runs (solid) and control runs (dashed). Symbols indicate 35 heights of lidar measurements and 17 heights of model outputs in the first 1 km AGL. Time (UTC) is shown in the upper-left corner of each graph.

143.6 m, NAMRR-CONUS-nest at 145.1 m, and lidar at 141.3 m.

5. Results: Validation of models by lidar data

a. Time-height comparisons

Comparisons of time-height cross sections of measured versus modeled wind flow provide a semiquantitative picture of the model performance. Figure 16 shows examples of time-height cross sections of the HRDLmeasured wind speed data (top panels) plotted against modeled winds from control and experimental runs of the NAMRR and NAMRR-CONUS-nest models for forecast lead hour 0. Examples are given for 2 days of the project: The left five panels are for 17 July and the right five panels are for 9 August, each characterized by different wind regimes. The results for RAP-HRRR (not shown here) are similar.

During these days, measurements, as well as extracted modeled winds, were taken along ship tracks at different distances from the shore (Fig. 11) and water depth. On 17 July, the ship cruised over deep (140-180 m) waters of the open ocean 60-160 km from the coast. On 9 August, the ship sailed northeast out of Boston, Massachusetts, along the coast over shallow (10-25 m) waters at 5–7 km from shoreline toward Cape Ann. After about 30 km, the ship turned southeast toward open ocean, with the farthest point about 60 km from the coast.



FIG. 14. Period-mean wind speed profiles for 6–12 Aug 2004 are shown for (top) RAP (blue) and HRRR (red) models and (bottom) NAMRR (blue) and NAMRR-CONUS-nest (red) models. Profiles are shown as means for (left) diurnal period, (middle) nighttime (0300–1200 UTC) hours, and (right) daytime (1500–2300 UTC) hours. Output is shown from the experimental model runs (solid) and control runs (dashed); Lidar data (black solid). Symbols indicate the heights of measurements (black) and model output (red, blue). Plus/ minus standard deviation of lidar data (black horizontal) and models' experimental runs (red and blue horizontal) for lead hour 0.

The time-height comparisons show that the overall features and trends of the wind flow measured at various locations and distances from the shore are represented in the models. All models captured the wind direction and wind speed pattern, and they show stronger winds during nighttime/morning transition hours (0000-1000 UTC) as well as episodes of LLJs in the lowest 300 m on 17 July. The models also simulated the 9 August trend of wind flow that included episodes of weak surface winds at 0200–0600 UTC and stronger nighttime winds above 200 m MSL. Quantifiable differences in flow

strength and timing can also be seen among the models, and between models and HRDL-measured values. For example, the LLJ episode (Fig. 16, left plots) is modeled but misplaced in time and height, particularly in the lowresolution model (NAMRR). Low wind conditions (Fig. 16, right plots) are simulated by all models, but the depth of this weak flow layer was overestimated.

b. Statistical analysis

The R^2 , bias, and RMSE between lidar-measured and modeled winds were computed for the July and August



FIG. 15. Observed and modeled (top) wind speed and (bottom) wind direction for 6–12 Aug 2004. Legend indicates output from model experimental runs: (left) NAMRR and NAMRR-CONUS-nest and (right) RAP and HRRR. Data are shown for initial conditions at the fourth level of all models. Lidar data (black) representing wind speed and direction at the closest (less than 4 m) height for each model.

periods for both the NAMRR and RAP modeling systems. These metrics were plotted as profiles for each forecast hour, and as a function of forecast-hour lead time averaged for the 10–500-m MSL layer.

Vertical profiles of period-mean statistical metrics for lead hour 0 are shown in Fig. 17 for RAP and HRRR (top row) and NAMRR and NAMRR-CONUS-nest (bottom row) control and experimental runs during the 6–12 August study period. All models produced similar trends in period-mean statistics, showing the same or improved metrics for experimental runs (solid lines with dots) compared to control runs (dashed lines), and the improvement is generally more pronounced at higher altitudes. As was the case for individual profiles (Fig. 13) or mean wind speed profiles (Fig. 14), profiles of the error statistics below 150 m show larger deviations of modeled winds versus lidar measurements than higher in the MABL, as also found by B17.

Figure 18 illustrates the performance of all four models for the August study period showing the RMSE as a function of forecast lead hour for both scalar wind (simple wind speed) and vector wind (which includes directional deviations). During this period, the RMSEs for the higher-resolution models in blue (HRRR and NAMRR-CONUS-nest) overlap or fall below those of the lower-resolution or "parent" models (red, RAP and NAMRR) for a short time after initialization. It is of interest to note that after that initial hour or so, the coarser-resolution models exhibit lower RMSE error statistics than the fine-mesh nests. This kind of degradation of model skill with increasing grid resolution has been noted before (e.g., Mass 2002; Mittermaier 2014). It has been attributed to the fact that strong horizontal variations exist on the smaller scales, and even if the models were to capture this well, small displacements of model horizontal structure with respect to the atmosphere's would produce model error versus a measurement at a given location.

The overall impact of assimilation of the land-based profiler data is summarized in Fig. 19, illustrating the percent improvement in scalar and vector winds from each model, averaged over the lowest 500 m. In general, during the August study period profiler assimilation produced an improved initialization by $\sim 0.2 \,\mathrm{m \, s^{-1}}$ and forecast improvement by as much as 5%-10% early (generally in the first 2h), with positive improvement indicated out to 3-4 h. Using profiler data averaged over the deeper 100-2000-m layer for 12 experiment days during July and August, Djalalova et al. (2016) similarly found forecast improvement early in the simulation for cases with assimilation of the profiler data. But with respect to parent versus nest grid interval over this deeper layer, they did not find that the finer-mesh runs exhibited larger errors after the initial hour.

For the July study period as a whole, in contrast to the August case, the assimilation of profiler data produced a drop in model performance (Fig. 20). After a brief interval of forecast improvement (~1 h mostly), the RMS error of the experimental profiler-assimilation runs exceeded that of the control runs. In other words, the assimilation of profile data from the profiler coastal array made the model forecast worse at the location of the *RHB*, consistent with results for the 100- to 2000-m layer.



FIG. 16. Time-height cross sections of (top) lidar-measured and (four bottom panels) modeled wind speed and direction for (left) 17 Jul and (right) 9 Aug 2004. The modeled winds used for this comparison are from control and experimental runs of NAMRR and NAMRR-CONUS-nest as indicated at the top of each panel. Winds ($m s^{-1}$) are scaled according to the color scale, vertical axes are height (m MSL), and horizontal axes are time (UTC). Modeled winds are shown for lead hour 0 (forecast 0).

Meteorological conditions during this period were dominated by a mesoscale cyclonic storm system that passed through the southern portions of the Gulf of Maine, producing easterly flow, rain, and fog over the study region. Surface meteorological data recorded on Appledore Island (White et al. 2007) on 13–14 July show an eastward-propagating cold frontal passage at surface levels. The frontal passage and accompanying wind speed vortices were also evident on Doppler weather radar reflectivity maps and satellite images (neither shown).

As shown previously (see Fig. 10, top), lidar measurements on 13 July were limited to 17 h (0000–1700 UTC). HRDL data were then unavailable until 1400 UTC 15 July as a result of the heavy rain and fog conditions also observed from the deck of the ship (Pichugina et al. 2014). On the first day that the storm affected the study area (13 July), data were taken from a stationary position when the ship was located ~15 km from shore (ocean depth of ~80 m). Changes in wind direction from southwesterly to easterly, seen in the HRDL time-height cross section (Fig. 21), were due to the cyclone vortex advancing into the area. On 17 July (Fig. 10, top), the day after the storm system moved off to the east, a time-height cross section of HRDL winds shows southwesterly to westerly flow weakening through the day.

For the easterly-flow period of 13–15 July, the region being sampled by the *RHB* was thus *upstream* of the coastal profiler array, so any influences of the profiler wind data on the model forecast would have to be transferred upwind by the model. During most of the rest of the project, in contrast, the flow had a westerly component. Thus, the Gulf of Maine during much of the project was downwind of the profiler coastal array so that assimilation of the profiler data would have a more direct downstream impact on the model solutions over the study area.

To further investigate the error behavior using HRDL data during the July period, Fig. 22 shows RMSE as a function of forecast lead time for 13 and 17 July

500

400

300

200

100

z, m

RAP EXP RAP CNTR HRRR EXP

HRRR CNTR





FIG. 17. Profiles of period-mean statistics between measured and modeled wind speeds in 6–12 Aug 2004 at forecast hour 0. (top) RAP and HRRR, and (bottom) NAMRR and NAMRR-CONUS-nest models. The legends indicate the experimental runs (solid lines) and the control runs (dashed lines). Symbols indicate heights of interpolations to lidar measurements. Rotor layer of 50–150 m (horizontal dotted lines).

individually. On 13 July, the RMSEs for the control and experimental assimilation runs are about equal at initialization, but then the RMSEs from the experimental runs rapidly increase with lead time relative to the control runs. During the westerly component wind conditions on 17 July, both assimilation model runs show improvement out to a 12-h lead time as a result of the assimilation of inland profiler data. This discrepancy between days suggests that the anomalous behavior of the models for the July period could be limited to the stormy easterly precipitation period and that this period dominates the July verification statistics. To verify this hypothesis, the HRDL data for 13-15 July were excluded from the sample, and verification statistics were calculated only for the combined 10-12 and 16-17 July datasets. Figure 23 shows RMSE (top) and improvement as a function of the forecast lead hour (bottom) for the scalar wind (speed) from the models (RAP/HRRR, left; NAMRR, right) for these 5 days. It shows smaller RMSE for the assimilation runs than for the control runs out to a lead time of several hours, in accord with the results found for the August period.

These results are consistent with those of Djalalova et al. (2016; refer to it for details) using the profiler wind dataset for this July period (which was available for all days, including 13–15 July) for all profilers in the coastal array—the ones being assimilated into the models for the experimental runs. They determined the RMSE of model winds versus profiler winds, averaged over the lowest 2000 m MSL, and similarly found that after excluding certain easterly-flow profiler data from the verification dataset for those three July days, assimilation of the coastal-profiler data improved model performance. Including these data for those three days in the verification dataset produced a significant degradation in model performance.



FIG. 18. RMSEs between observed and modeled (top) scalar and (bottom) vector winds averaged over 50-m layer MSL are shown as a function of forecast lead time for the August 2004 period. (left) RAP and HRRR models, and (right) NAMRR and NAMRR-CONUS-nest models. RMSE from experimental (solid lines) and control (dashed lines) runs according to the legend in each panel.

The July decreases in model skill by the assimilation of profiler data thus are most likely a consequence of a 3-day period of easterly winds over the Gulf of Maine study region. In attempting to explain this behavior, the most obvious issue is that the coastal profiler array is poorly placed-downwind of the verification regionfor NWP model assimilation under easterly-flow conditions. Any impact that the assimilation of these data would have in the study region would have to be transferred upstream, and model processes-either numerical, model physics, or model initialization related-that would do this are more likely to degrade than to increase the accuracy of the forecast. In general, such decreases in model accuracy have been noted before. For example, Morss and Emanuel (2002) and Semple et al. (2012) found that assimilated measurements can at times degrade a model's analysis or prediction. The impact of assimilation data on a model forecast may depend on the quality of the measurements, the nature of the forecast model, the assimilation methodology, or how well the atmospheric regime in question is represented by the measured dataset (e.g., Morss and Emanuel 2002).

In our case, Banta et al. (2014) and Djalalova et al. (2016) have opined that a lack of flow dependence in the background error term used by the model's 3DVAR data assimilation algorithm is a likely contributor to this problem. Subsequent to the model runs described in this study, hybrid ensemble-variational assimilation was implemented in RAP version 2 in 2014 at NCEP (Benjamin et al. 2016). The present example illustrates how assimilation of profile data in a model having static background error covariance can, at least in some cases, degrade the forecast in regions of the simulation domain that are upstream, or in otherwise unfavorable locations, relative to the sites of the assimilated measurements.

6. Summary and conclusions

This paper has presented an evaluation of model skill in simulating and predicting winds aloft over the ocean by comparing retrospective runs of two NWP forecast models to shipborne Doppler lidar wind measurements over the Gulf of Maine. Deployed on board the R/V Ronald H. Brown during the 2004 NEAQS field campaign, the NOAA high-resolution Doppler lidar provided accurate motion-compensated measurements from the water surface up to several hundred meters above mean sea level. High-precision and high-vertical-resolution lidar data provide an important capability for investigating and understanding wind flow conditions that influence model accuracy at turbine-rotor-layer heights and above, and for evaluating model skill through at least the lowest several hundred meters of the atmosphere. These measurements have given insight into boundary



FIG. 19. Improvement (%) of model wind forecast as a result of assimilation of the coastal profilers for the August 2004 study period is shown as a function of forecast lead time. (left) RAP (red) and HRRR (blue) models. (right) NAMRR (red) and NAMRR-CONUS-nest (blue) models. Improvements are shown for mean RMSE from experimental and control runs in the 10–500-m layer MSL (see Fig. 18).

layer behavior, including nocturnal stable and LLJ conditions, which are among the most difficult atmospheric conditions to characterize, understand, and model (Banta et al. 2006, 2013; Pichugina et al. 2008; Pichugina and Banta 2010), and the ability of forecast models to simulate them.

The study presents validation of two different modeling systems: a new hourly updated version of the NAM



FIG. 20. As in Fig. 18, but for the 10-17 Jul 2004 study period.



FIG. 21. Time-height cross sections of wind speed and direction on 13 Jul from hourly averaged lidar measurements and experimental and control runs of NAMRR-CONUS-nest and NAMRR models. Modeled winds are shown for lead hour 0.

forecast system, the NAMRR; and a 2012 version of the hourly updated RAP system and associated embedded finescale models, the NAMRR-CONUS-nest and the HRRR, respectively. Two issues pertinent to the analysis procedures were addressed. First, model values were interpolated to the heights of the lidar data (rather than vice versa) to preserve the finescale structure and to increase the sample size for statistics. Second, the effect of averaging over layers of different depth, where deeper layers produce better error statistics, were quantified, revealing, for example, that RMSEs calculated over a 1000-m layer were 20% smaller than those for an individual level.

Lidar-model comparisons showed that all models captured major trends in the wind field relatively well but that larger quantitative discrepancies between modeled and observed winds were evident below 100– 200 m (e.g., Fig. 13) and during nighttime LLJs, illustrating a need to improve boundary layer-to-surface exchange physics.

A finding of the POWER study was that the finerresolution domain-embedded models produced worse wind speed predictions than the coarser parent domains for forecast lead times greater than about 3h, although for shorter lead times the finer-resolution models performed better. The underperformance of finerresolution models at these scales has been previously pointed out (e.g., Mass 2002; Mittermaier 2014), and these results show that similar behavior can be found in accurately measured hourly winds aloft over the ocean.

Lidar data also were used to estimate the impact of additional data assimilation from 11 profilers located along the U.S. East Coast by comparing these data to model runs with and without the profiler data assimilated. Assimilation of profiler data led to a few percent (5%–10%) improvement in all models for the first 2–4 forecast hours. After that time, and even more dramatically for a 3-day period in July, the experimental simulations (with profiler assimilation) performed worse than the control runs. The significant degradation of skill during July as a result of the assimilation was associated



FIG. 22. RMSE between observed and modeled winds in the first 500 m MSL for (a) 13 Jul and (b) 17 Jul 2004. RMSE for experimental (solid) and control runs (dashed) for NAMRR (red) and NAMRR-CONUS-nest (blue).



FIG. 23. (top) RMSEs between observed and modeled scalar winds averaged over 500-m layer MSL are shown as a function of forecast lead time for 5 days in 10–12 and 16–17 Jul 2004. (left) RAP and HRRR models, and (right) NAMRR and NAMRR-CONUS-nest models. Solid (dashed) lines represent RMSE from experimental (control) runs according to the legend in each panel. (bottom) Improvement (%) of models as a result of the additional assimilation of inland profiler data.

with a mesoscale cyclonic storm system that passed through the study region that generating a period of easterly flow, which put the profiler-assimilation array downstream of the region. Why the effects on model skill upstream of the assimilation data are negative instead of neutral is an important question that is not answered here. The assimilation scheme and its background error covariance have been suggested as possible contributors, but more in-depth measurement-modeling studies will be required to address this issue.

It is noteworthy that two models having different origins, numerics, microphysical schemes, land surface models, and other attributes produced many similar error properties, including often similar error magnitudes, similar fine-mesh versus coarse-mesh behavior, and a similar response to assimilation of profiler data. This is reminiscent of the study of Zhong and Fast (2003, p. 1301), who evaluated three significantly different mesoscale models against a comprehensive measurement dataset and found that the "types of forecast errors were surprisingly similar" and exhibited similar sensitivities for all three models. Such similarities indicate that some basic aspects of contemporary NWP, such as the finitedifference approach at currently achievable grid resolutions, may impose fundamental limitations on model error reduction and model skill (B17).

Acknowledgments. The present study, the Positioning of Offshore Wind Energy Resources (POWER) project, was sponsored by the Wind and Water Power Program of the U.S. Department of Energy's Energy Efficiency and Renewable Energy Office, by the NOAA Atmospheric Science for Renewable Energy Program, and by the NOAA/ESRL Air Quality Program. We thank our colleagues Raul Alvarez, Dan Law, Janet Machol, Richard Marchbanks, Brandi McCarty, Rob K. Newsom, Sara Tucker, Scott Sandberg, and Ann Weickmann for their long hours collecting HRDL data on the ship during the NEAQS 2004 campaign, and Lisa Darby for her review of this manuscript.

REFERENCES

- Alpert, J., 2004: Sub-grid scale mountain blocking at NCEP. 20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction, Seattle, WA, Amer. Meteor. Soc., P2.4, https://ams.confex.com/ams/84Annual/techprogram/ paper_71011.htm.
- 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., R. K. Newsom, J. K. Lundquist, Y. L. Pichugina, R. L. Coulter, and L. Mahrt, 2002: Nocturnal low-level jet

characteristics over Kansas during CASES-99. *Bound.-Layer Meteor.*, **105**, 221–252, doi:10.1023/A:1019992330866.

- —, Y. L. Pichugina, and W. A. Brewer, 2006: Turbulent velocityvariance profiles in the stable boundary layer generated by a nocturnal low-level jet. *J. Atmos. Sci.*, **63**, 2700–2719, doi:10.1175/JAS3776.1.
- —, —, N. D. Kelley, R. M. Hardesty, and W. A. Brewer, 2013: Wind energy meteorology: Insight into wind properties in the turbine-rotor layer of the atmosphere from high-resolution Doppler lidar. *Bull. Amer. Meteor. Soc.*, **94**, 883–902, doi:10.1175/BAMS-D-11-00057.1.
- Banta, R., and Coauthors, 2014: NOAA study to inform meteorological observation for offshore wind: Positioning of Offshore Wind Energy Resources (POWER). NOAA Final Tech. Rep. to DOE, 145 pp., http://www.esrl.noaa.gov/gsd/renewable/ AMR_DOE-FinalReport-POWERproject-1.pdf.
- —, and Coauthors, 2015: 3D volumetric analysis of wind-turbine wake properties in the atmosphere using high-resolution Doppler lidar. J. Atmos. Oceanic Technol., 32, 904–914, doi:10.1175/JTECH-D-14-00078.1.
- Benjamin, S. G., and Coauthors, 2004: An hourly assimilation– forecast cycle: The RUC. Mon. Wea. Rev., 132, 495–518, doi:10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2.
- —, and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, 144, 1669–1694, doi:10.1175/MWR-D-15-0242.1.
- Browning, K. A., and R. Wexler, 1968: The determination of kinematic properties of a wind field using Doppler radar. J. Appl. Meteor., 7, 105–113, doi:10.1175/ 1520-0450(1968)007<0105:TDOKPO>2.0.CO;2.
- Carley, J. R., and Coauthors, 2015: Ongoing development of the hourly-updated version of the NAM forecast system. 27th Conf. on Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction, Chicago, IL, Amer. Meteor. Soc., 2A.1, https://ams.confex.com/ams/27WAF23NWP/ webprogram/Paper273567.html.
- Carter, D., K. S. Gage, W. L. Ecklund, W. M. Angevine, P. E. Johnston, A. C. Riddle, J. Wilson, and R. William, 1995: Developments in UHF lower tropospheric wind profiling at NOAA's Aeronomy Laboratory. *Radio Sci.*, **30**, 977–1001, doi:10.1029/95RS00649.
- Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. Technical Report Series on Global Modeling and Data Assimilation, Vol. 3, NASA Tech. Memo. 104606, 85 pp., https:// ia600502.us.archive.org/23/items/nasa_techdoc_19950009331/ 19950009331.pdf.
- Darby, L. S., and Coauthors, 2007: Ozone differences between near-coastal and offshore sites in New England: Role of meteorology. J. Geophys. Res., 112, D16S91, doi:10.1029/ 2007JD008446.
- Djalalova, I. V., and Coauthors, 2016: The POWER Experiment: Impact of assimilation of a network of coastal wind profiling radars on simulating offshore winds in and above the wind turbine layer. *Wea. Forecasting*, **31**, 1071–1091, doi:10.1175/ WAF-D-15-0104.1.
- Drechsel, S., G. J. Mayr, J. W. Messner, and R. Stauffer, 2012: Wind speeds at heights crucial for wind energy: Measurements and verification of forecasts. *J. Appl. Meteor. Climatol.*, 51, 1602–1617, doi:10.1175/JAMC-D-11-0247.1.
- 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, 8851, doi:10.1029/2002JD003296.

- Elliott, D., and Coauthors, 2012: Offshore resource assessment and design conditions: A data requirements and gaps analysis for offshore renewable energy systems. DOE Tech. Rep. DOE/ EE-0696, 66 pp., doi:10.2172/1219742.
- Fairall, C. W., and Coauthors, 2006: Turbulent bulk transfer coefficients and ozone deposition velocity in the International Consortium for Atmospheric Research into Transport and Transformation. J. Geophys. Res., 111, D23S20, doi:1029/ 2006JD007597.
- Fehsenfeld, F. C., and Coauthors, 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, 19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction, San Antonio, TX, Amer. Meteor. Soc., 10.1, https://ams.confex.com/ams/SLS_WAF_ NWP/techprogram/paper_47241.htm.
- —, 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., 14B.1, https:// ams.confex.com/ams/91Annual/webprogram/Paper179488. html.
- Grund, C. J., R. M. Banta, J. L. George, J. N. Howell, M. J. Post, R. A. Richter, and A. M. Weickmann, 2001: High-resolution Doppler lidar for boundary layer and cloud research. *J. Atmos. Oceanic Technol.*, **18**, 376–393, doi:10.1175/ 1520-0426(2001)018<0376:HRDLFB>2.0.CO;2.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by longlived greenhouse gases: Calculations with the AER radiative transfer models. J. Geophys. Res., 113, D13103, doi:10.1029/ 2008JD009944.
- Janjić, 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, doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.
- —, 2001: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp., http://www.emc.ncep.noaa.gov/officenotes/ newernotes/on437.pdf.
- —, 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, 82, 271–285, doi:10.1007/s00703-001-0587-6.
- —, 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, 75 pp., doi:10.5065/D6WH2MZX.
- Kelley, N., M. Shirazi, D. Jager, S. Wilde, J. Adams, M. Buhl, P. Sullivan, and E. Patton, 2004: Lamar low-level jet project interim report. National Renewable Energy Laboratory Tech. Rep. NREL/TP-500-34593, 216 pp., http://www.nrel.gov/docs/ fy04osti/34593.pdf.
- Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. *Bull. Amer. Meteor.*

Soc., **83**, 407–430, doi:10.1175/1520-0477(2002)083<0407: DIHRPM>2.3.CO;2.

- Mittermaier, M. P., 2014: A strategy for verifying near-convectionresolving model forecasts at observing sites. *Wea. Forecasting*, 29, 185–204, doi:10.1175/WAF-D-12-00075.1.
- 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, 16663–16682, doi:10.1029/ 97JD00237.
- 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, doi:10.1256/00359000260498897.
- Musial, W., and S. Butterfield, 2004: Future for offshore wind energy in the United States. National Renewable Energy Laboratory Conf. Paper NREL/CP-500-36313, 14 pp.
- —, and B. Ram, 2010: Large-scale offshore wind power in the United States: Assessment of opportunities and barriers. National Renewable Energy Laboratory Tech. Rep. NREL/TP-500-40745, 222 pp., https://www.nrel.gov/docs/fy10osti/40745.pdf.
- Nunalee, C. G., and S. Basu, 2014: Mesoscale modeling of coastal low-level jets: Implications for offshore wind resource estimation. *Wind Energy*, **17**, 1199–1216, doi:10.1002/we.1628.
- Pichugina, Y. L., and R. M. Banta, 2010: Stable boundary-layer depth from high-resolution measurements of the mean wind profile. J. Appl. Meteor. Climatol., 49, 20–35, doi:10.1175/ 2009JAMC2168.1.
- —, —, N. D. Kelley, B. Jonkman, R. K. Newsom, S. C. Tucker, and W. A. Brewer, 2008: Horizontal-velocity and variance measurements in the stable boundary layer using Doppler lidar: Sensitivity to averaging procedures. *J. Atmos. Oceanic Technol.*, 25, 1307–1327, doi:10.1175/2008JTECHA988.1.
- —, —, W. A. Brewer, S. P. Sandberg, and R. M. Hardesty, 2012: Doppler lidar–based wind-profile measurement system for offshore wind-energy and other marine boundary layer applications. J. Appl. Meteor. Climatol., 51, 327–349, doi:10.1175/ JAMC-D-11-040.1.
- —, and Coauthors, 2014: Assessment of NWP forecast model skill to represent marine boundary layer features. 21st Symp. on Boundary Layers and Turbulence, Leeds, United Kingdom, Amer. Meteor. Soc., 14A.4, https://ams.confex.com/ams/ 21BLT/webprogram/Paper248663.html.
- —, and Coauthors, 2017: Properties of the offshore low level jet and rotor layer wind shear as measured by scanning Doppler lidar. *Wind Energy*, **20**, 987–1002, doi:10.1002/we.2075.
- Rogers, E., and Coauthors, 2009: The NCEP North American Mesoscale modeling system: Recent changes and future plans. 23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction, Omaha, NE, Amer. Meteor. Soc., 2A.4, https://ams.confex.com/ams/23WAF19NWP/ techprogram/paper_154114.htm.

- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc., 91, 1015–1057, doi:10.1175/ 2010BAMS3001.1.
- Schwartz, M., D. Heimiller, S. Haymes, and W. Musial, 2010: Assessment of offshore wind energy resources for the United States. National Renewable Energy Laboratory Tech. Rep. NREL/TP-500-45889, 97 pp., http://www.nrel.gov/docs/fy10osti/ 45889.pdf.
- Semple, A., M. Thurlow, and 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, doi:10.1175/MWR-D-11-00273.1.
- Smirnova, T. G., J. M. Brown, and S. G. Benjamin, 1997: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. *Mon. Wea. Rev.*, **125**, 1870–1884, doi:10.1175/1520-0493(1997)125<1870: PODSMC>2.0.CO;2.
- —, S. G. Benjamin, J. M. Brown, B. Schwartz, and D. Kim, 2000: Validation of long-term precipitation and evolved soil moisture and temperature fields in MAPS. Preprints, *15th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 1.15, https://ams.confex.com/ams/annual2000/techprogram/paper_ 6846.htm.
- 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. Oceanic Technol., 1, 37–49, doi:10.1175/ 1520-0426(1984)001.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, doi:10.1175/2008MWR2387.1.
- White, A. B., and Coauthors, 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. *Bound.-Layer Meteor.*, 78, 321–349, doi:10.1007/BF00120940.
- Wolfe, D. E., and Coauthors, 2007: Shipboard multisensor merged wind profiles from the New England Air Quality Study 2004. J. Geophys. Res., 112, D10S15, doi:10.1029/2006JD007344.
- Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, doi:10.1175/ 1520-0493(2002)130<2905:TDVAWS>2.0.CO;2.
- Zhong, S., and J. D. Fast, 2003: An evaluation of the MM5, RAMS, and Meso-Eta models at subkilometer resolution using field campaign data in the Salt Lake Valley. *Mon. Wea. Rev.*, **131**, 1301–1322, doi:10.1175/1520-0493(2003)131<1301: AEOTMR>2.0.CO;2.