



Validation of RU-WRF, the Custom Atmospheric Mesoscale Model of the Rutgers Center for Ocean Observing Leadership

Mike Optis, Andrew Kumler, George Scott, Mithu Debnath, and Pat Moriarty

National Renewable Energy Laboratory

Produced under direction of the New Jersey Board of Public Utilities by the National Renewable Energy Laboratory (NREL) under FIA-18-01872.

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Strategic Partnership Project Report

NREL/TP-5000-75209

February 2020



Validation of RU-WRF, the Custom Atmospheric Mesoscale Model of the Rutgers Center for Ocean Observing Leadership

Mike Optis, Andrew Kumler, George Scott, Mithu Debnath, and Pat Moriarty

National Renewable Energy Laboratory

Suggested Citation

Optis, Mike, Andrew Kumler, George Scott, Mithu Debnath, and Pat Moriarty. 2020. *Validation of RU-WRF, the Custom Atmospheric Mesoscale Model of the Rutgers Center for Ocean Observing Leadership*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5000-75209. <https://www.nrel.gov/docs/fy20osti/75209.pdf>.

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Strategic Partnership Project Report
NREL/TP-5000-75209
February 2020

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Support for the work was also provided by the New Jersey Board of Public Utilities under FIA-18-01872. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via www.osti.gov.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Acknowledgments

An external review committee was formed at the start of this work to review the quality and robustness of the analysis and validation presented in this report. This committee spanned wind energy experts in national laboratories, universities, and industry. The National Renewable Energy Laboratory (NREL) would like to thank the committee for its review of this analysis and report, which helped considerably in strengthening the quality of this work.

We acknowledge the U.S. Department of Energy and the Pacific Northwest National Laboratory (PNNL) for the offshore lidar data collected off the coast of New Jersey and disseminated publicly through PNNL's Data Access Portal. These data were used extensively in the validation of the Rutgers University Mesoscale Model.

We finally wish to acknowledge the New Jersey Board of Public Utilities (NJBPU) and the Rutgers University Center for Ocean Observing Leadership (RU-COOL) for the opportunity to perform this analysis. We have believed for some time that a regionally customized, ensemble-based mesoscale modeled wind resource product is the next generation for wind resource assessment. The insights gained from this work have been invaluable, and we credit NJBPU and RU-COOL for enabling a modern, state-of-the-art approach to wind resource assessment.

Executive Summary

The offshore wind industry in the United States is poised to invest billions of dollars in wind farms off the coast of New Jersey and neighboring states. An accurate characterization of the wind resource in this area is critical to ensure confidence in the analyses that require these data, such as energy prediction, grid integration, extreme events, and capacity expansion applications. Given the sparsity of offshore observations in U.S. coastal waters, especially at hub height, data from mesoscale models (i.e., atmospheric models run on scales of 1 km to about 100 km) are usually the principal wind resource data for offshore analyses. It is therefore vital that mesoscale models are accurately representing the U.S. North Atlantic wind resource.

The Rutgers University Center for Ocean Observing Leadership (RU-COOL) contracted the National Renewable Energy Laboratory (NREL) to evaluate RU-COOL's mesoscale model as well as its observation and validation capabilities for New Jersey offshore wind resource characterization. The core of this evaluation focuses on RU-COOL's custom setup of the public, open-source Weather Research and Forecasting (WRF) mesoscale model. This custom setup, named RU-WRF, is run daily as a forecast product over a domain centered on the New Jersey offshore wind lease area. This area spans from southern Massachusetts to North Carolina and covers three electricity independent system operators (Pennsylvania-New Jersey-Maryland, New York, New England). Simulation results are disseminated publicly on a daily basis and used by a range of stakeholders, including the wind energy industry for offshore wind resource assessment and forecasting. The most unique feature of RU-WRF is the use of a custom "coldest-pixel" sea surface temperature (SST) product developed by RU-COOL and designed to better capture the frequent occurrence of cold coastal upwelling of water from the sub-surface Mid-Atlantic cold pool. The cold pool is unique to the Mid-Atlantic Bight. It is caused by a strong surface temperature gradient that forms from surface heating during the summer, and it can lead to strong coastal upwelling under certain atmospheric conditions.

NREL was tasked to assess and recommend improvements to the RU-WRF model as well as the observational network and procedures used by RU-COOL to validate RU-WRF. To robustly assess the performance of RU-WRF, which is itself one particular setup of the WRF model, we considered the relative performance of a range of valid model setups. This "ensemble" was constructed based on the key differences between RU-WRF's and NREL's own custom publicly available implementation of WRF: the Wind Integration National Dataset (WIND) Toolkit. Specifically, variations in atmospheric forcing, SST forcing, WRF version, and WRF namelist (i.e., internal model setup) were considered when building the ensemble. The relative performance of ensemble members was assessed both when implemented as a forecast product (i.e., driven by a large-scale forecast product; current implementation of RU-WRF) and as a hindcast product (i.e., driven by a large-scale reanalysis product; not currently implemented).

This analysis yielded several key results, categorized here in terms of improvements to the RU-WRF setup, improvements to the coldest-pixel SST product, and improvements to the atmospheric observations and methods used to validate the RU-WRF.

Regarding the RU-WRF setup, the following results were found:

1. The RU-WRF model run in forecast mode was most accurate in the time period immediately after initialization. However, this time period is currently considered by RU-COOL as "spin-up" time for RU-WRF and data from this period are not disseminated publicly.
2. The Global Forecasting System (GFS) large-scale atmospheric forecast product, currently used to force the RU-WRF simulations, performed slightly better than the North American Mesoscale forecast model.
3. The current vertical resolution of RU-WRF is too coarse to accurately model vertical wind profiles. Increasing the vertical resolution of RU-WRF near the surface resulted in a much better representation of the wind profile.
4. The RU-WRF model, currently implemented using WRF version 3.9, demonstrated moderate to large negative biases at all the observation locations used for validation. These biases were substantially reduced when RU-WRF was implemented using WRF version 4.0. By contrast, the unbiased root-mean-squared-error (RMSE) was reduced only slightly when upgrading from WRF version 3.9 to 4.0.

5. The WIND Toolkit setup showed negligible performance changes when upgrading from WRF Version 3.9 to 4.0. The WIND Toolkit showed a larger magnitude of bias compared to RU-WRF implemented with WRF version 4.0, but lower magnitude of bias compared to RU-WRF implemented with WRF Version 3.9. The unbiased RMSE was larger in the WIND Toolkit compared to RU-WRF.
6. Differences in WIND Toolkit and RU-WRF model performance are attributed to the different planetary boundary layer parameterization schemes used in each. The scheme implemented in RU-WRF was found more accurate for offshore wind resource modeling. Increased performance in RU-WRF from WRF version 3.9 to 4.0 is believed to be related to improvements made in this scheme.
7. Most WRF models considered demonstrated large positive bias in the summer months and lower magnitude of bias in the remaining months. The cause for this seasonal trend would be a valuable subject for further research.

Regarding the coldest-pixel SST product, the following results were found:

1. Overall, the RU-COOL coldest-pixel SST product was found to be modestly more accurate compared to a widely used SST product from the National Center for Environmental Prediction (NCEP).
2. Relative performance of the two SST products depended significantly on location. Of the four buoy locations considered, the coldest-pixel product was more accurate at two locations and the NCEP product was more accurate at the other two. Causes for the relative geospatial performance of these two products would be a valuable subject for future research.
3. When used as a boundary forcing to RU-WRF, the coldest-pixel SST product provided minor improvements in wind speed modeling accuracy relative to the NCEP SST product.
4. SST measurements at the buoy locations demonstrated considerable diurnal cycles in the spring and summer months. Such cycles are not resolved in daily resolution SST products and may be significant in influencing the offshore wind resource.

Regarding the observational network and methods used by RU-COOL to validate RU-WRF, the following results were found:

1. RU-COOL has recently adopted NREL's mesoscale model validation methods used for NREL's WIND Toolkit. These methods are robust for validation of modern mesoscale models. However, these methods are limited to validating wind speed and wind direction only and not other atmospheric variables known to influence wind power production, such as atmospheric stability and turbulence.
2. It is not possible to confidently validate RU-WRF using coastal observation stations because of the coarse horizontal spatial resolution of RU-WRF (3 km) and the large coastal wind speed gradients.

Based on this analysis, we have provided several recommendations for improving RU-WRF, the coldest-pixel SST product, as well as the observational network and validation procedures used by RU-COOL. These improvements are summarized in Tables A through C.

Table A. Recommendations for Improving the RU-WRF Model

Code	Recommendation	Purpose	Effort	Value
A1	Reduce the spin-up time for the forecast model from 6 hours to 1 hour.	RU-WRF showed the lowest RMSE during the 1- to 5-hour period after model initialization. These forecast periods are certainly relevant for wind energy forecasting purposes and would correspond to 5:00 to 9:00 local time when electricity demand is ramping up. Switching to a 1-hour spin-up would be consistent with current practices used in state-of-the-art mesoscale forecast models.	Low	High
A2	Upgrade the WRF version used by RU-WRF from 3.9 to 4.0.	A considerable improvement in RU-WRF accuracy was found by upgrading the WRF version from the currently implemented 3.9 version to the recently released 4.0 version. Specifically, the negative bias found in WRF 3.9 in the New Jersey offshore area is largely corrected in WRF 4.0, which we speculate is because of improvements made to the Mellor, Yamada, Nakanishi, and Niino planetary boundary layer scheme.	Low	High
A3	Increase the vertical resolution of RU-WRF by at least a factor of two.	The current vertical resolution in RU-WRF is insufficient to capture the wind speed profile at heights relevant to offshore wind turbines. Increasing the resolution to at least that used by NREL's Wind Toolkit will make RU-WRF more valuable for wind energy purposes.	Low	High
A4	For resource assessment applications, develop a hindcast RU-WRF product that is forced by a large-scale reanalysis product.	A reanalysis-driven WRF simulation is generally more accurate than a forecast product-driven WRF simulation, particularly for large forecast lead times (e.g., greater than 12 hours). Apart from wind power forecasting, a reanalysis-driven hindcast product would provide more value for all wind energy applications (e.g., wind resource assessment, annual energy production, grid integration).	High	High

Table B. Recommendations for Improving the Modeling of SST

Code	Recommendation	Purpose	Effort	Value
B1	Investigate the causes for high negative bias of the coldest-pixel SST product at Buoy 44065.	Such an investigation may reveal inadequacies in the coldest-pixel SST product that can be improved.	Mid	Mid
B2	Investigate high RMSE during the spring/summer months at all buoys and the low relative bias of the NCEP SST product during these periods.	Such an investigation may reveal inadequacies in the coldest-pixel SST product that can be improved.	Mid	Mid
B3	Consider a blended SST product that combined the best attributes of the RU-COOL coldest-pixel SST product and the NCEP SST product.	Although the Rutgers coldest-pixel product generally outperformed the NCEP SST product, there were cases where the NCEP SST product performed better (e.g., Buoys 44009 and 44065). A blended SST product that captures the positive attributes of both SST data products would likely improve wind resource modeling in RU-WRF.	High	Mid
B4	Consider using the NCEP SST product as a backup to the coldest-pixel product.	The National Aeronautics Space Administration SST product is used by RU-COOL when coldest-pixel observations are not available (e.g., nighttime). It is possible that switching to the NCEP RTG product as a backup would provide more accurate SST and wind resource modeling.	High	Low
B5	Consider the development of an hourly SST product.	SST at offshore buoys shows a strong diurnal cycle in the spring and summer months. Exploring available data sources or interpolation methods to produce an hourly SST product that accounts for this diurnal cycle would likely improve wind resource characterization and in particular the sea breeze.	High	Mid

Table C. Recommendations for Improving RU-COOL's Observational Network and Validation Procedures

Code	Recommendation	Purpose	Effort	Value
C1	Incorporate publicly available floating lidar data	The Pacific Northwest National Laboratory's Wind-Tracer lidar off the coast of New Jersey, as well as possible future deployments by the U.S. Department of Energy (DOE), provides a valuable validation data set for modeled wind speeds, wind profiles, wind veer, and wind shear.	Low	High
C2	Consider our recommended configuration for Rutgers' WindTracer lidar.	This configuration should allow for reliable wind profile measurements at frequent intervals out to 8 km offshore, providing a valuable data set for model validation and wind resource characterization.	High	High
C3	Leverage existing or develop new partnerships with offshore wind energy developers	Lidars are increasingly being deployed in New Jersey offshore wind energy lease areas. Accessing these data through existing or new partnerships could provide the highest-quality data available to validate RU-WRF.	High	High
C4	Do not use the Rutgers Coastal Metocean Monitoring station or Oyster Creek met tower as principal validation stations	Because of strong coastal wind speed gradients and the coarse horizontal spatial resolution of the RU-WRF model, the interpolation of modeled wind speeds to these stations is too uncertain to allow for a meaningful validation of model performance.	Low	Mid
C5	Enhance validation methods to include additional variables beyond wind speed.	RU-COOL is currently following validation procedures outlined in an NREL technical report describing the WIND Toolkit (Draxl et al. 2015). We recommend continued use of these methods but also to expand the validation to atmospheric variables beyond wind speed that are known to influence the wind resource, most notably measures of atmospheric turbulence and stability.	Mid	Mid
C6	Include a RU-WRF and NOAA HRRR Performance Comparison in Validation Methods	The National Oceanographic and Atmospheric Administration produces an hourly updated, WRF-based mesoscale high-resolution rapid refresh (HRRR) forecast product for the continental United States at the same spatial resolution as RU-WRF. We recommend adding a comparison of RU-WRF and HRRR to the existing validation methods to potentially highlight the value of customized components of RU-WRF.	High	High

Acronym List

AVHRR	Advanced Very High Resolution Radiometer
DOE	U.S. Department of Energy
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA-Interim	ECMWF Reanalysis Interim
FNL	Final Operational Global Analysis
GFS	Global Forecasting System
HRRR	High Resolution Rapid Refresh
ISO	independent system operator
MAB	Mid-Atlantic Bight
MAE	mean absolute error
MYNN	Mellor Yamada Nakanishi and Niino
NAM	North American Mesoscale Model
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Center for Environmental Prediction
NDBC	National Data Buoy Center
NJBPU	New Jersey Board of Public Utilities
NOAA	National Oceanic and Atmospheric Administration
NREL	National Renewable Energy Laboratory
NYSERDA	New York State Energy Research and Development Authority
PBL	planetary boundary layer
PNNL	Pacific Northwest National Laboratory
PPI	plane position indicator
RHI	range height indicator
RMSE	root-mean-squared-error
RTG	real-time global
RU-COOL	Rutgers University Center for Ocean Observation Leadership
RU-WRF	Rutgers University custom Weather Research and Forecasting model
RUCMM	Rutgers Coastal Metocean Monitoring
RUOYC	Rutgers Oyster Creek monitoring station
SODAR	Sound Detection and Ranging
SPoRT	Short-term Prediction Research and Transition Center
SST	sea surface temperature
UTC	Universal Time Coordinated
WIND	Wind Integration National Dataset
WRF	Weather Research and Forecasting Model
WTK	Wind Integration National Dataset Toolkit
YSU	Yonsei University

Table of Contents

1	Introduction	1
1.1	Scope of Work	1
1.2	Background on RU-WRF Mesoscale Model	1

1.3	Background on RU-COOL Observational Network	3
1.4	Approach To Validating the RU-WRF Model	4
1.4.1	Time Period	4
1.4.2	Hindcast- and Forecast-Driven Models	4
1.4.3	An Ensemble Approach to Validation	5
1.5	Outline of Report	6
2	Review of RU-COOL Validation Procedures for RU-WRF	7
2.1	Observational Network	7
2.1.1	Lack of Offshore Measurements Near Hub Height	7
2.1.2	Validating Using Coastal Stations	8
2.2	Validation Methods	8
2.2.1	The Pielke Criteria	8
2.2.2	WIND Toolkit Validation Methods	10
2.3	Validating Against Other Mesoscale Forecast Products	11
2.4	Summary of Findings and Recommendations	12
3	Coldest-Pixel SST Product Evaluation	14
3.1	Analysis of Daily SST Data	14
3.2	Potential Value of Hourly SST Data	14
3.3	Summary of Findings and Recommendations	17
4	Hindcast Validation of RU-WRF	21
4.1	Ensemble Members for Hindcast Validation	21
4.2	WRF Simulation Setup	21
4.3	Data Setup and Post-Processing	22
4.4	Validation Results	23
4.4.1	Overall Performance	24
4.4.2	Seasonal Performance	26
4.4.3	Diurnal Performance	26
4.4.4	Wind Profile Comparison	28
4.4.5	Effect of Increased Vertical Resolution	30
4.4.6	Wind Direction Comparison	30
4.5	Summary of Findings and Recommendations	30
5	Forecast Validation of RU-WRF	33
5.1	Ensemble Members for Forecast Validation	33
5.2	WRF Simulation Setup	33
5.3	Validation Results	33
5.4	Summary of Findings and Recommendations	34
6	Recommendations for Improvement and Future Work	38
7	Appendix	42
7.1	Recommended WindTracer Lidar Configuration	42
7.2	Supplemental Figures	42

List of Figures

Figure 1.	The nested 9-km and 3-km WRF domain used in RU-WRF. Figure provided by RU-COOL. Magenta indicates current wind energy lease areas. Green indicates planned lease areas as of February 2019. The color shading over states indicates the Independent System Operator (ISO) for that region: blue is PJM Interconnection, green is New York ISO, and yellow is ISO New England.	2
Figure 2.	Atmospheric observation stations available to RU-COOL for RU-WRF validation.	4
Figure 3.	Mean daily wind speeds modeled by RU-WRF on October 3, 2015, in the Delaware Bay region. NDBC observing stations are shown in red.	9
Figure 4.	Same as Figure 3 but centered near the RUCMM station. The RUCMM, Oyster Creek met tower (RUOYC) and the PNNL floating LiDAR (VIND1) are shown in red.	10
Figure 5.	Observed standard deviation of hourly measurements from observing stations in the New Jersey offshore area and RMSE of RU-WRF relative to those observations. Results are based on one year of simulations from June 1, 2015 to May 31, 2016.	11
Figure 6.	Observed and modeled daily SST at four buoy locations off the coast of New Jersey.	15
Figure 7.	Performance metrics for the modeled daily SST products averaged over the full validation period	16
Figure 8.	Mean bias in the daily modeled SST products separated by calendar month.	17
Figure 9.	Unbiased RMSE for the daily modeled SST products separated by calendar month.	19
Figure 10.	Deviation from monthly average SST for the hourly SST measurements at Buoy 44065.	20
Figure 11.	Mean bias for each ensemble member at each validation station. The “Average” column on the far right is the mean of the previous five columns.	24
Figure 12.	Unbiased RMSE for each ensemble member at each validation station. The “Average” column on the far right is the mean of the previous five columns.	25
Figure 13.	Monthly performance metrics averaged across all validation stations for select WRF ensemble members. The leftmost panels consider bias and unbiased RMSE for different combinations of WRF versions and namelists while keeping the atmospheric and SST forcings constant. The rightmost panels consider bias and unbiased RMSE for different combinations of atmospheric and SST forcings while keeping the WRF version and namelist constant.	27
Figure 14.	Mean diurnal evolution of the modeled and observed 5-m wind speed at Buoy 44025. The top figure considers model variations in the namelist and WRF version. The bottom figure considers variations in atmospheric and SST forcings.	28
Figure 15.	Same as Figure 14 but for the 90-m wind speeds at the VIND1 lidar.	29
Figure 16.	Mean modeled and observed wind profiles at the VIND1 lidar considering (a) variations in model namelist and WRF version and (b) variations in atmospheric and SST forcings.	30
Figure 17.	Mean modeled and observed wind profiles at the VIND1 lidar for two vertical resolutions of the RU-WRF model.	31
Figure 18.	Modeled and observed 5-m wind direction distributions at Buoy 44025. Both variations in namelist and WRF version (a) as well as atmospheric and SST forcing (b) are considered.	31
Figure 19.	Performance metrics over the full validation period for each forecast model setup. Results are shown for each validation stations.	34
Figure 20.	Forecast model performance metrics at Buoy 44025 over the 01:00 UTC to 06:00 UTC time window.	35

Figure 21. Forecast model performance metrics at the VIND1 Lidar over the 01:00 UTC to 06:00 UTC time window.	35
Figure 22. Performance metrics at the VIND1 Lidar for different calendar months.	36
Figure 23. Performance metrics at Buoy 44025 for different calendar months.	36
Figure 24. Lidar scanning geometry for Scan 1.	43
Figure 25. Lidar scanning geometry for Scan 2.	44
Figure 26. Mean diurnal evolution of the modeled and observed 5-m wind speed at Buoy 44009. The top figure considers model variations in namelist and WRF version. The bottom figure considers variations in atmospheric and SST forcings.	45
Figure 27. Same as Figure 26 but for Buoy 44017.	46
Figure 28. Same as Figure 26 but for Buoy 44065.	47

List of Tables

Table A. Recommendations for Improving the RU-WRF Model	vi
Table B. Recommendations for Improving the Modeling of SST	vii
Table C. Recommendations for Improving RU-COOL’s Observational Network and Validation Procedures	viii
Table 1. Key Characteristics of the RU-WRF Model.	2
Table 2. Atmospheric Observation Stations Available to RU-COOL for RU-WRF Validation.	5
Table 3. Recommended Metrics, Timescales and Variables for Use in Validating Modern Mesoscale Models.	12
Table 4. A Summary of the WRF Model Components Used to Construct the 24-Member Hindcast Ensemble in this Validation Study.	22
Table 5. Comparison of Key Attributes Between RU-WRF and the WIND Toolkit.	22
Table 6. Observation Stations and Validation Heights Used to Validate RU-WRF in this Study.	23
Table 7. Key Performance Metrics for All Ensemble Members. The first section “Namelist – WRF version” consists of the mean results across all 24 ensemble members. The second section “Atmospheric – SST Forcing” consists of the results considering only the ensemble members that use the RU-WRF namelist and WRF version 4.0.	26
Table 8. Atmospheric Forcing Products and WRF Model Spin-Up Times Considered When Assessing the RU-WRF Forecast Product.	33
Table 9. Recommendations for improving the RU-WRF Model	39
Table 10. Recommendations for Improving the Modeling of SST	40
Table 11. Recommendations for Improving RU-COOL’s Observational Network and Validation Procedures	41
Table 12. Lidar Configuration for Scan 1.	42
Table 13. Lidar Configuration for Scan 2 - PPI.	42
Table 14. Lidar Configuration for Scan 2 - RHI.	42

1 Introduction

1.1 Scope of Work

The Rutgers University Center for Ocean Observing Leadership (RU-COOL) contracted the National Renewable Energy Laboratory (NREL) to evaluate RU-COOL's atmospheric observation and modeling capabilities for characterizing the New Jersey offshore wind resource. The observational network used by RU-COOL consists of mostly public but some private coastal and offshore buoy-based stations. The core wind resource modeling capability of RU-COOL is a custom setup of the Weather Research and Forecasting (WRF) mesoscale model, referred to in this report as RU-WRF. The most unique feature of RU-WRF not found in other WRF model setups is the use of custom sea surface temperature (SST) products generated by RU-COOL. This custom product was designed to better capture the unique coastal upwelling and strong storm mixing in the Mid-Atlantic Bight (MAB) that other typical SST products are not designed to capture. Funding for the development, maintenance, and use of RU-WRF by RU-COOL is provided by the New Jersey Board of Public Utilities (NJBPU), who also funded the validation work presented in this report.

In this validation study, NREL was specifically tasked to:

1. Assess the observational network used by RU-COOL to validate RU-WRF and make recommendations for improvement
2. Assess methods used by RU-COOL to validate RU-WRF and make recommendations for improvement.
3. Examine the inputs to and setup within RU-WRF, compare against available NREL data sets, and make recommendations for improvement.

1.2 Background on RU-WRF Mesoscale Model

The RU-WRF model has been operating since 2011 to provide daily forecasts of wind speed, wind direction, and other atmospheric variables off the coast of New Jersey. The RU-WRF simulations have been primarily used to forecast and characterize the New Jersey offshore wind resource. The development of RU-WRF in 2011 was motivated by the designation of New Jersey offshore lease areas and the acknowledgement from RU-COOL and NJBPU that a custom regional mesoscale model was required to account for the unique meteorological ocean (metocean) conditions of the MAB. In this region, complex metocean dynamics (e.g., coastal upwelling, air-sea interactions, low-level jets, sea breezes) have been demonstrated to influence wind conditions at heights relevant to wind power. RU-COOL believed that a custom WRF model tuned to regional atmospheric dynamics should, in principle, more accurately capture the characteristics of the wind resource than a standard WRF model setup.

The main characteristics of RU-WRF are summarized in Table 1. The model is run in forecast mode on a daily basis forced by the Global Forecasting System (GFS) $0.25^\circ \times 0.25^\circ$ forecast data product¹ initialized at 00:00 Universal Time Coordinated (UTC). The RU-WRF model is two-way nested with an outer domain at 9-kilometer (km) horizontal resolution and the inner domain at 3-km horizontal resolution. The 3-km domain spans the coastal region from southern Massachusetts to North Carolina whereas the larger 9-km domain spans from southern Maine to the northern portion of South Carolina (see Figure 1). A 6-hour spin-up time from 00:00 UTC to 06:00 UTC is used in order for the WRF model to be well developed and dynamically realistic. Forecast output from RU-WRF at 3 km resolution are provided publicly² from 6 hours to 48 hours after initialization. Forecast data from the 9-km domain are provided from 6 hours to 120 hours after initialization.

The key feature of RU-WRF is its use of a custom “coldest-pixel” SST product produced by RU-COOL. The need for a custom SST product is motivated by the unique metocean characteristics of the MAB. The region experiences strong coastal upwelling and storm mixing, which in the summer has the effect of mixing deeper, colder water to

¹The GFS forecasting product is produced by the National Center for Environmental Prediction
<https://www.nco.ncep.noaa.gov/pmb/products/gfs>

²<https://rucool.marine.rutgers.edu/data/meteorological-modeling/ruwrf-mesoscale-meteorological-model-forecast/>

Table 1. Key Characteristics of the RU-WRF Model.

Model Feature	Value
Domain	9 km – 3 km two-way nesting
WRF version	3.9 (upgraded from 3.6 in 2018)
Atmospheric forcing	Global Forecasting System $0.25^\circ \times 0.25^\circ$
Sea surface temperature forcing	RU-COOL coldest pixel product
Planetary boundary layer scheme	Mellor, Yamada, Nakanishi, and Niino
Average height of lowest model levels	53.7 m, 131.0 m, 232.5 m
Initialization time	00Z
Model spin up	6 hours
Run frequency	Daily
3-km forecast horizon	6–30 hours ahead
9-km forecast horizon	6–120 hours ahead

the surface from the persistent well-studied subsurface cold-water feature known as the Mid-Atlantic Cold Pool. This upwelling affects local and regional horizontal temperature gradients and influences the offshore wind resource, particularly sea breezes (Seroka et al. 2018).

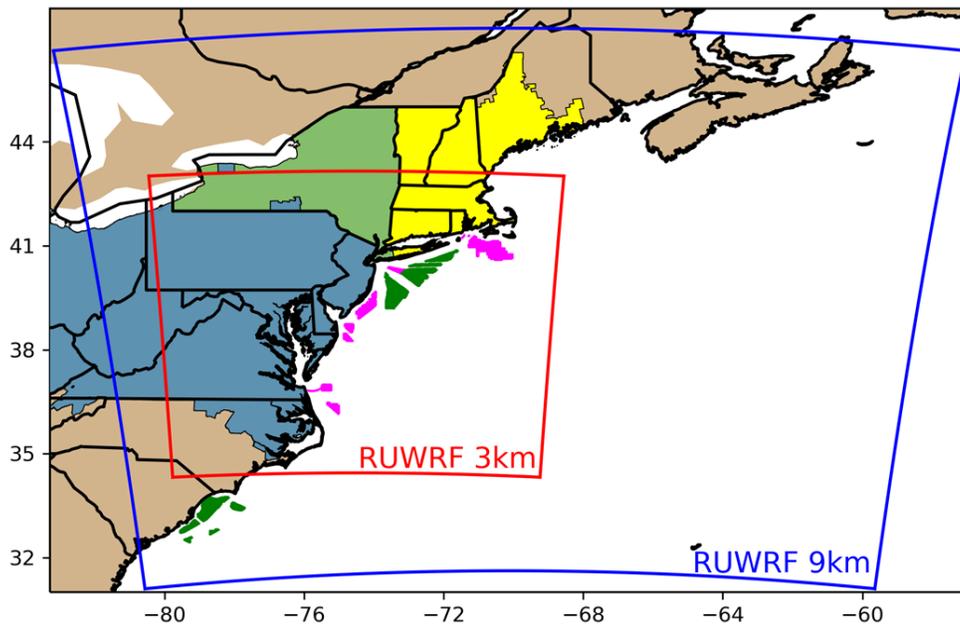


Figure 1. The nested 9-km and 3-km WRF domain used in RU-WRF. Figure provided by RU-COOL. Magenta indicates current wind energy lease areas. Green indicates planned lease areas as of February 2019. The color shading over states indicates the Independent System Operator (ISO) for that region: blue is PJM Interconnection, green is New York ISO, and yellow is ISO New England.

The coldest-pixel product uses satellite observations from the Advanced Very High Resolution Radiometer (AVHRR), as do most SST products. Data specifically from the NOAA-18 and NOAA-19 satellites are collected directly by RU-COOL using an on-site satellite dish.

The coldest pixel product differs from other SST products in the method of compositing (i.e., determining the most representative SST from a series of satellite-based observations for a given pixel location) and the declouding al-

gorithm (i.e., method for categorizing satellite observations as either a cloud or a legitimate SST measurement). A standard SST product generally takes the warmest of multiple measurements at a given pixel in order to remove any residual cloud contamination. In the MAB, this “warmest pixel” approach can have the effect of discarding cold upwelling events. By contrast, RU-COOL implements a “coldest pixel” method in which the coldest measurement from the composite is kept, therefore capturing more cold upwelling events than the standard warmest pixel approach.

The coldest pixel method, however, can be more likely to identify clouds as valid SST measurements. To adjust for this method, and to better identify cold pools, RU-COOL has implemented a custom declouding algorithm with the following characteristics:

1. The temperature threshold is changed twice a year to account for the strong seasonal cycle in the MAB,
2. The horizontal temperature gradient threshold is set to 1°C, much higher than standard SST products that are usually around 0.3°C. This higher threshold prevents removal of the strong gradients typically seen around features like upwelling fronts and the Gulf Stream’s north wall and eddy boundaries.
3. A threshold is applied to the near-infrared data channel that removes pixels with high albedo and only allows clear, low reflectance “dark” pixels through.
4. There is no comparison to an SST climatology, which typically removes important MAB features including anomalously cold summer SST caused by coastal upwelling fronts or strong storm mixing.

In cases where AVHRR data are not available, RU-COOL relies on the National Aeronautics and Space Administration (NASA) Short-term Prediction Research and Transition Center (SPoRT) SST data product³ to fill data gaps. The SPoRT data are provided daily at 06:00 UTC.

1.3 Background on RU-COOL Observational Network

This section provides a summary of the network of observational stations currently providing measurement data and therefore available to RU-COOL for validating RU-WRF. These stations are shown in Figure 2 and a summary of each station provided in Table 2.

Most observational data in the New Jersey coastline area are public data sets that are managed and disseminated by the National Data Buoy Center (NDBC), which is operated by the National Oceanic and Atmospheric Administration (NOAA). These stations include floating buoys, offshore fixed platforms, and land-based stations. Buoy measurements are at or below 5 meters (m) above water level, although the fixed platform and onshore station measurement heights range from 6 to 21 m above water and ground level, respectively.

RU-COOL privately operates the Rutgers Coastal Metocean Monitoring (RUCMM) station in the Great Bay Boulevard Wildlife management area (see Figure 2). The station has a 12-m meteorological (met) mast that measures standard meteorological variables. Data from this met mast were considered in this validation study. The station also has a Triton sodar to measure wind speed profiles up to 200 m and a recently-installed Lockheed Martin WindTracer scanning wind lidar which can measure wind vectors out to a range of 10 km. These sodar and lidar data sources were not available for the validation period used in this study and therefore are not considered in this report.

RU-COOL has access to data from the Oyster Creek met mast, located at the Oyster Creek Nuclear Generating Station (labeled “RUOYC” in Figure 2). The met mast is 120 m high and instrumented at 10 m, 45 m, and 115 m. These data were considered in this validation study.

As noted in Table 2, RU-COOL has provided information on the principal stations used for RU-WRF validation. The sites are used by RU-COOL based on proximity to New Jersey offshore wind lease areas, measurement heights, and location suitability for capturing phenomena such as the sea breeze. The table shows three of the four buoys principally used as well as the RUCMM ground station and RUOYC met tower. The remaining stations are sometimes considered by RU-COOL but are not principal validation stations.

³<https://weather.msfc.nasa.gov/sport/sst/descriptions.html#sportsstcomp>

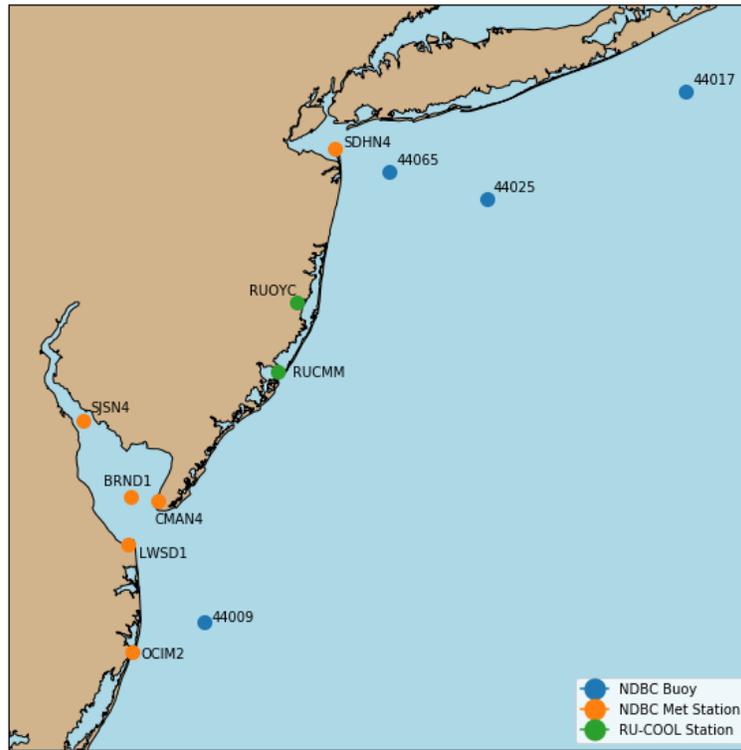


Figure 2. Atmospheric observation stations available to RU-COOL for RU-WRF validation.

1.4 Approach To Validating the RU-WRF Model

1.4.1 Time Period

In collaboration with RU-COOL, NREL defined a 1-year time period from June 1, 2015 to May 30, 2016, in which to validate RU-WRF model performance against observations. This time period was selected because of near-complete coverage of the RU-COOL coldest-pixel SST product. There were only a few days where coldest-pixel data were not available (either no daylight satellite passes were made that day or the RU-COOL satellite dish that receives AVHRR data was not operational). Furthermore, this time period provided complete coverage of SST data from the four NDBC buoys, allowing a robust validation of modeled SST against observations.

1.4.2 Hindcast- and Forecast-Driven Models

A mesoscale model such as WRF is used for wind resource modeling in two main ways:

1. **Hindcast Model:** A mesoscale model is driven by historical data from a reanalysis product, which represents the most representative state of the atmosphere.⁴
2. **Forecast Model:** A mesoscale model is driven by a large-scale forecast product, which is initialized using available observations at the initialization time and then is propagated forward in time without observations and based only on the dynamics of an atmospheric model. The RU-WRF model is currently used only as a forecast model.

Both hindcast and forecast mesoscale models are used extensively in the wind energy industry. Hindcast models are

⁴Reanalysis products combine a global network of observations using data assimilation to drive a numerical weather model and create a gridded “best estimate” of the global atmosphere.

Table 2. Atmospheric Observation Stations Available to RU-COOL for RU-WRF Validation.

Station Name	Code	Type	Meas. Height (meters)	Principal Site for RU-COOL
Buoy 44009	44009	Buoy	5.0	Yes
Buoy 44017	44017	Buoy	5.0	No
Buoy 44025	44025	Buoy	5.0	Yes
Buoy 44065	44065	Buoy	4.0	Yes
Sandy Hook	SDHN4	Ground Station	5.5	No
Cape May	CMAN4	Ground Station	9.7	No
Brandywine Shoal	BRND1	Offshore Platform	21.1	No
Lewes	LWSD1	Ground Station	9.9	No
Ship John Shoal	SJSN4	Offshore Platform	14.8	No
Ocean City Inlet	OCIM2	Ground Station	8.5	No
Rutgers Coastal Metocean Monitoring	RUCMM	Ground Station	12.0	Yes
Oyster Creek	RUOYC	Met Tower	10.0, 45.0, 110.0	Yes

used to produce historical wind resource time series and maps with applications in site assessment, grid integration analysis, and annual energy production estimates. Forecast models are used to predict and plan for future wind energy generation at plant, utility, regional, and national scales, and on timescales of an hour to days in advance.

Given the importance of both types of models, the performance of RU-WRF is assessed both in hindcast and forecast setups (a detailed description of these setups is described in Section 4). This validation study focuses predominantly on the hindcast setup. Hindcasts are generally more accurate than forecast models because historical observational data are assimilated for the entire model run. By contrast, a forecast model only has observations available at the time of model initialization and relies entirely on model dynamics after initialization. Consequently, forecast model accuracy generally decreases the longer the model is run. Focusing this validation study predominantly on the hindcast setup eliminates the influence of this forecast model “drift” and allows for a more robust performance assessment of the RU-WRF model setup. That said, a performance assessment of RU-WRF in forecast mode (as is the current implementation) is also provided.

1.4.3 An Ensemble Approach to Validation

The WRF model is modular in design and allows for a high degree of customization in both the input data to and the parameterizations and settings within the model. In this regard, the RU-WRF setup is simply one possible setup of WRF (i.e., one member of a larger ensemble of possible setups). To robustly evaluate and recommend improvements to RU-WRF, it is important to assess RU-WRF performance relative to an ensemble of reasonable WRF model setups.

This ensemble approach to validating RU-WRF was implemented in this work. In designing the hindcast ensemble, NREL contrasted the different setups in RU-WRF and NREL’s Wind Integration National Dataset (WIND) Toolkit (Draxl et al. 2015). The WIND Toolkit is a WRF-based hindcast timeseries product provided at 2-km horizontal resolution across the continental United States over the years 2007 to 2013 and is the largest grid integration wind data set publicly available to date. Based on the key differences between the inputs to and setup within the WIND Toolkit and RU-WRF, an ensemble of WRF model setups was constructed and used to validate the specific setup used in RU-WRF. A detailed description of these ensembles is provided in Section 4.

1.5 Outline of Report

Section 2 describes the observational network used by RU-COOL for validation of RU-WRF as well as the validation procedures currently used. An assessment is then provided by NREL and recommendations for improvement are made.

Section 3 evaluates the performance of the RU-COOL coldest-pixel SST product relative to both observations and a widely used SST product from the National Center for Environmental Prediction (NCEP) that is used in the WIND Toolkit.

Section 4 evaluates the performance of the RU-WRF model. First, the construction of the WRF ensembles for both the hindcast- and forecast-based evaluations of RU-WRF are described. Next, the design of the model simulations is explained. The selection and post-processing of observational data used to validate the WRF ensemble are then described. Finally, results from the hindcast and forecast validations are provided.

Section 6 summarizes the key findings and recommendations for improvement found in this study. Each recommendation is categorized based on the degree of improvement possible and the ease of implementation.

2 Review of RU-COOL Validation Procedures for RU-WRF

This section provides a review of both the observational network used by RU-COOL and the methods used to validate RU-WRF. Recommendations for improvement in both areas are provided.

2.1 Observational Network

As summarized in Section 1.3, RU-COOL makes use of a network of mostly public but some private observation stations for use in validating RU-WRF. The stations in the New Jersey offshore area primarily used by RU-COOL for validating RU-WRF include Buoys 44009, 44025, and 44065, as well as RU-COOL's RUCMM station and the Oyster Creek met tower.

2.1.1 Lack of Offshore Measurements Near Hub Height

The lack of U.S. East Coast offshore observations near hub height is a well-recognized problem in the wind industry and a key barrier in accurate characterization of the wind resource and reducing risk for offshore project development. Similar to many institutions, RU-COOL does not own nor have access to observational equipment deployed near offshore wind energy lease areas capable of measuring wind speeds at heights relevant to wind energy (i.e., up to 200 m). Although the Oyster Creek met tower and the RUCMM sodar provide wind speed profiles up to 115 m and 200 m, respectively, they are providing coastal measurements only and are located about 60 km and 30 km, respectively, from the centroid of the nearest wind energy lease area. Given this distance and the differences in the coastal and offshore wind resource regimes, the value of these observation stations for offshore wind resource characterization is limited.

By contrast, the WindTracer lidar located at the RUCMM station, which has a 10 km horizontal range, is of considerably higher value for offshore wind resource characterization. Given RU-COOL's ownership of the lidar, there is an opportunity to optimize the configuration of the lidar to characterize the wind resource at a location much closer to wind energy lease areas. To this end, we have provided recommendations in this report for lidar configuration to achieve two measurement objectives: (1) accurate measurement of 80-m wind speeds 8 km offshore at high time resolution (every 40 seconds), and (b) measurement of wind profiles in 100-m increments up to 8 km offshore at moderate time resolution (every two minutes). A detailed description of the lidar configuration required to meet these measurement objectives is provided in the Appendix.

In addition to this RU-COOL operated lidar, there are opportunities to leverage public data from previous or planned lidar deployments from governmental institutions. Funded by the U.S. Department of Energy (DOE), the Pacific Northwest National Laboratory (PNNL) deployed a lidar about 5 km off the coast of Atlantic City, New Jersey over the period period December 2015 through February 2017 (labeled "VIND1" in Figure 2). These data are freely available through PNNL's Data Access Portal¹. This lidar is still 25–30 km away from the centroid of the nearest offshore lease area and is no longer collecting data, and there are measurement bias issues at several heights (Newsom 2016); consequently, this data set is certainly not ideal for offshore wind resource characterization. However, to our knowledge, it is the highest-quality data set currently available for this purpose. We recommend that RU-COOL perform validation of earlier RU-WRF simulations using this lidar data, specifically using only the 90 m measurements which are known to have very low bias (Newsom 2016).

There are also recent lidar deployments by the New York State Energy Research and Development Authority (NY-SERDA)² and potentially future deployments by the DOE. We recommend that RU-COOL leverage these future lidar deployments and add them to the observational network used to validate RU-WRF.

Finally, wind energy developers continue to deploy floating lidar systems in their offshore wind energy lease areas, including those in New Jersey. Data from these deployments provide the highest quality observations currently avail-

¹<https://a2e.energy.gov/data>

²<https://www.nyserda.ny.gov/About/Newsroom/2019-Announcements/2019-01-31-NYSERDA-Announces-Contracts-for-Collecting-Environmental-and-Metocean-Data-in-Support-of-Offshore-Wind-Energy-Development>

able for offshore wind resource characterization and mesoscale model validation. These data are highly proprietary; however, it is possible for RU-COOL to leverage existing or develop new partnerships with offshore wind developers to use these data for research and validation purposes. Such partnerships would ultimately benefit wind energy developers who currently use or might use RU-WRF in the future, given the potential improvements to RU-WRF that result from validation against such lidars. We recommend that RU-COOL pursue these partnerships and add these private lidars to their validation network to the extent possible.

2.1.2 Validating Using Coastal Stations

As discussed in the previous section, the use of coastal stations for offshore wind resource characterization and model validation is not ideal. Coastal stations are located far from current offshore wind energy lease areas and can have very different wind resource characteristics compared to those farther offshore. More importantly, large wind speed gradients at the coastline generally prohibit a meaningful validation of mesoscale-modeled wind speeds when these models are run at coarse resolution (e.g., 3-km RU-WRF resolution). Under these conditions, modeled wind speed from one model grid box to the next can change significantly, and the interpolation of modeled wind speeds to the observation station for purposes of validation is highly uncertain.

Figures 3 and 4 illustrate this coastal gradient problem in the Delaware Bay and Atlantic City regions, respectively. Each figure shows a 3 km × 3 km grid of mean RU-WRF modeled daily wind speeds on October 3, 2015 (which was selected because of high wind conditions). In Figure 3, the SJSN4 and CMAN4 stations are both located in areas of high coastal gradients in which wind speeds in neighboring grid boxes can vary as much as 5 ms⁻¹. In Figure 3, the RUCMM and RUOYC stations (which are principal validation stations for RU-WRF) are located in areas where neighboring wind speeds can differ by 2 ms⁻¹. Comparatively, the BRND1 station (Figure 3) and the PNNL lidar (Figure 4) are located farther offshore and in areas with significantly less wind speed gradient.

For many analyses (e.g., sea breeze events, comparing various spatial resolutions for a mesoscale model), these coastal stations would be valuable reference sites. However, when validating a 3-km spatial resolution mesoscale model such as RU-WRF, the uncertainty associated with interpolating modeled grid box average wind speeds to a coastal observation station is prohibitively high. Therefore, we recommend that coastal stations, specifically RUOYC and RUCMM, are used with caution when validating RU-WRF. Specifically, when assessing performance improvements under different model setups (other than varying horizontal spatial resolution), these coastal stations should not be used in validation.

2.2 Validation Methods

This section evaluates the methods used by RU-COOL to validate RU-WRF. For the most part, these methods were not explicitly provided by RU-COOL to NREL. Rather, current validation methods used by RU-COOL have been adopted from previous reports, which RU-COOL provided as background documents for this study. These reports are evaluated in this section.

2.2.1 The Pielke Criteria

A 2013 NJBPU/RU-COOL report on RU-WRF (Glenn and Dunk 2013) described a list of three criteria used by RU-COOL to validate RU-WRF. These criteria were also described in a summary presentation of RU-WRF provided by RU-COOL to NREL in May 2018. These criteria were adopted from a 2013 study by Stanford University examining the US East Coast offshore wind energy resources (Dvorak et al. 2013), which in turn adopted the criteria from a 2002 textbook on mesoscale meteorological modeling (Pielke 2002). These criteria, all of which are based on relative comparisons of modeled wind speeds against the standard deviation of observed wind speeds, are as follows:

1. $\sigma_{obs} \approx \sigma_{mod}$
2. $RMSE < \sigma_{obs}$
3. Unbiased $RMSE < \sigma_{obs}$

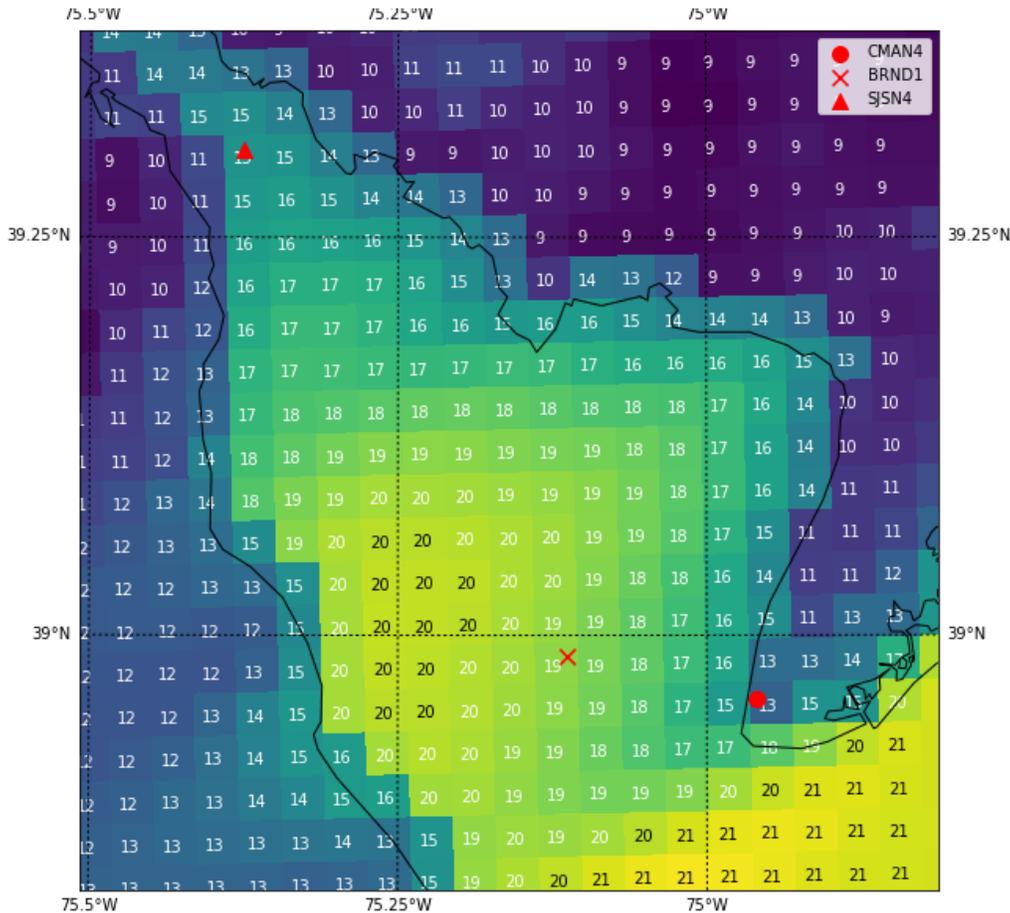


Figure 3. Mean daily wind speeds modeled by RU-WRF on October 3, 2015, in the Delaware Bay region. NDBC observing stations are shown in red.

It is our opinion that some aspects are useful in a first-order validation of a mesoscale model, although other aspects are outdated. Specifically, comparing the standard deviation of modeled and observed wind speeds is always useful in assessing how well a model captures wind speed variability. The remaining criteria, which require the root-mean-squared error (RMSE) and unbiased RMSE to be less than the observed standard deviation of wind speeds, are largely outdated. These criteria were likely more useful when the Pielke (2002) textbook was written given that mesoscale models were still in early stages of development (e.g., the WRF model was first released in 2000). For modern mesoscale models, which have undergone extensive improvements since the early 2000s, these criteria should generally be easily met.

Figure 5 compares the standard deviation and RMSE of RU-WRF modeled hourly wind speeds at the observation stations listed in Table 2. Results are based on a 1-year simulation from June 1, 2015 to May 31, 2016. As seen in the figure, comparing the variability of modeled and observed wind speeds is still a useful check. Apart from the Oyster Creek station, RU-WRF tends to underestimate variability in the New Jersey coastal area. By contrast, the figure shows that the remaining Pielke (2002) criteria are met. Specifically, the RMSE of the modeled wind speeds is always less than the standard deviation of the observations, especially at the offshore locations [i.e., the NDBC buoys, the offshore PNNL lidar (VIND1)] and the offshore fixed platform stations (BRND1 and SJSN4).

Based on this analysis, we recommend that the first Pielke (2002) criteria (i.e., $\sigma_{obs} \approx \sigma_{mod}$) continue to be used as an initial check on RU-WRF model performance. Furthermore, we recommend that the remaining criteria (i.e., RMSE

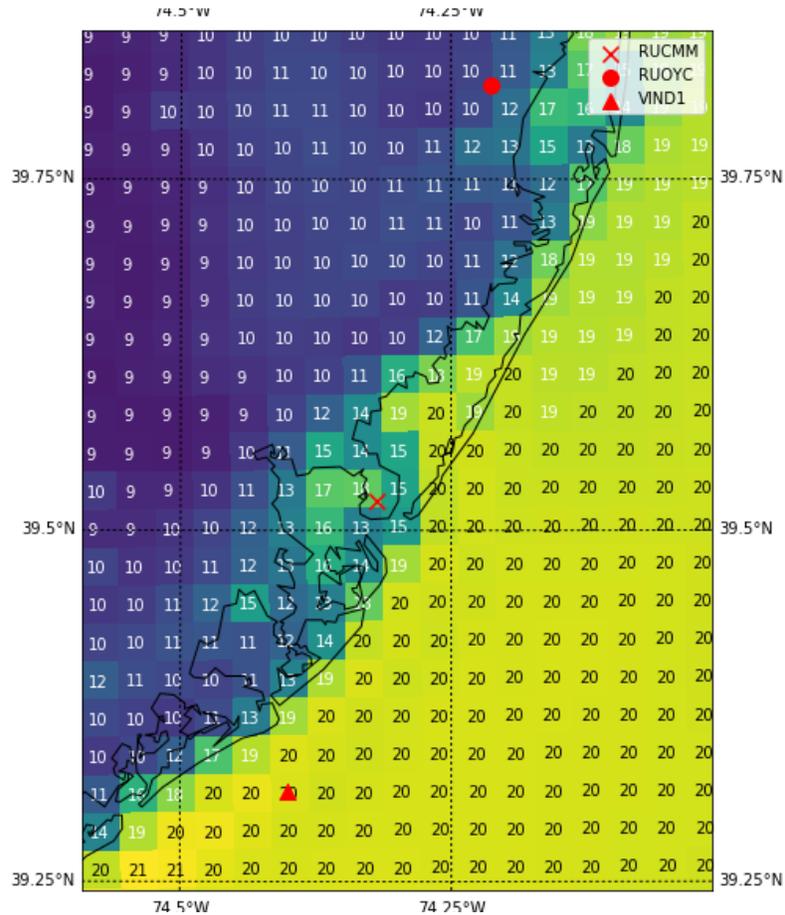


Figure 4. Same as Figure 3 but centered near the RUCMM station. The RUCMM, Oyster Creek met tower (RUOYC) and the PNNL floating LiDAR (VIND1) are shown in red.

and unbiased $RMSE < \sigma_{obs}$) are not used to validate RU-WRF.

2.2.2 WIND Toolkit Validation Methods

The summary presentation provided to NREL also cites an NREL technical report describing the validation of the WIND Toolkit (Draxl et al. 2015). This report presents a range of metrics used to validate the WIND Toolkit. It is our understanding based on this citation and subsequent conversations with RU-COOL, that the validation methods presented in Draxl et al. (2015) are currently used to validate RU-WRF.

The Draxl et al. (2015) report uses a set of performance metrics and visualizations to assess the extent to which model results resemble the observations. No specific criteria required to “validate” a model are provided. The performance metrics include RMSE, bias, unbiased RMSE, the coefficient of determination (R^2), mean absolute error (MAE), and percentage error. Bias and RMSE specifically are provided on diurnal, monthly, and annual scales.

This type of validation method is well suited for modern mesoscale model performance assessments. The range of error metrics allow for flexibility in how model performance is assessed (e.g., RMSE may be favored over MAE to more heavily penalize outliers). The unbiased RMSE is also useful to separate out the bias contribution to the overall error. The breakdown of model performance on an annual scale is useful for assessing model ability to estimate annual energy production from a wind farm. Breakdowns on monthly and diurnal scales are useful in assessing model ability to meet seasonal and diurnal energy demands and to plan for wind plant maintenance.

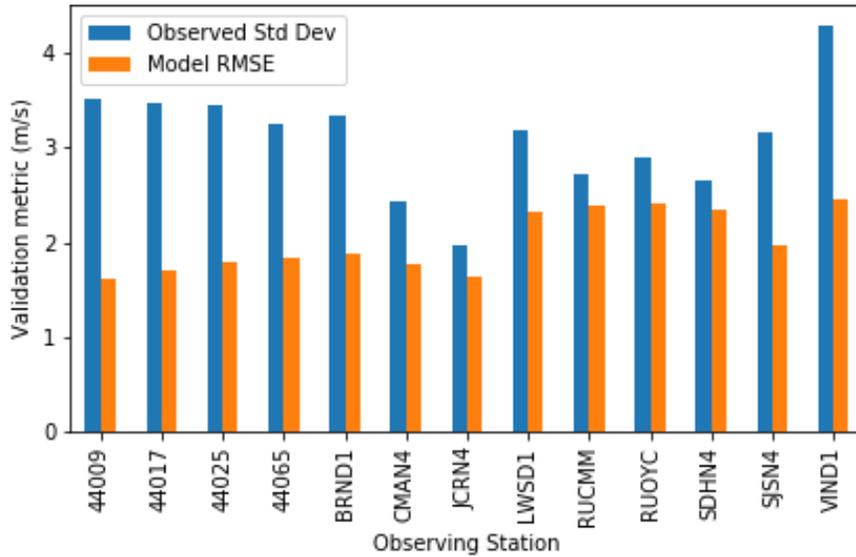


Figure 5. Observed standard deviation of hourly measurements from observing stations in the New Jersey offshore area and RMSE of RU-WRF relative to those observations. Results are based on one year of simulations from June 1, 2015 to May 31, 2016.

The methods in Draxl et al. (2015), however, do not consider atmospheric variables other than wind speed and wind direction. Several other atmospheric variables are known to significantly affect wind power generation, including air density, temperature, wind shear, wind veer, and variables quantifying atmospheric turbulence or stability (e.g., turbulence intensity, bulk Richardson number). Provided these variables are measured at observation stations (which is often not the case), we recommend that these variables be considered when validating RU-WRF.

A summary of our recommended validation metrics, timescales, and variables for validating RU-WRF is provided in Table 3.

2.3 Validating Against Other Mesoscale Forecast Products

As part of its validation procedures, RU-COOL currently compares RU-WRF output to that from the GFS forecast product and another product from NCEP: the North American Mesoscale (NAM) forecast model³. The NAM model features greater horizontal spatial resolution (12 km × 12 km) compared to the GFS (roughly 28 km × 20 km).

In general, larger-scale models such as GFS and NAM do not accurately simulate, nor are they designed to simulate, winds in the lower boundary layer of the atmosphere. Rather, these larger-scale models are designed to accurately simulate synoptic weather conditions (e.g., cold fronts, pressure systems). Therefore, it is our opinion that such models may not be useful for validating RU-WRF and that RU-WRF model performance should almost always exceed that of these larger-scale products.

Instead, we recommend that RU-COOL compare RU-WRF performance against similar mesoscale forecast products. Most notably, the NOAA High-Resolution Rapid Refresh (HRRR) model is a real-time WRF-based forecast product for the continental United States that is updated every hour.⁴ Similar to RU-WRF, the HRRR is run at 3-km × 3-km spatial resolution and data are made publicly available.⁵

³<https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesoscale-forecast-system-nam>

⁴<https://rapidrefresh.noaa.gov/hrrr/>

⁵https://rapidrefresh.noaa.gov/hrrr/HRRR/Welcome.cgi?dsKey=hrrr_jet

Table 3. Recommended Metrics, Timescales and Variables for Use in Validating Modern Mesoscale Models.

Feature	Component
Metrics	Bias Root-Mean-Squared-Error or Mean Absolute Error Unbiased Root-Mean-Squared-Error or Unbiased Mean Absolute Error Coefficient of Determination
Timescales	Annual Monthly Diurnal
Variables	Wind Speed Wind Direction Wind Shear Wind Veer Potential Temperature Vertical Gradient Bulk Richardson Number Turbulent Kinetic Energy Turbulence Intensity Obukhov Length

The HRRR model is widely considered the state-of-the-art for national mesoscale forecast products. Therefore, a comparison of HRRR and RU-WRF performance in the New Jersey MAB would be a useful exercise and could highlight the value of the custom RU-WRF components such as the coldest-pixel SST product. We recommend that such a comparison be incorporated into the validation methods of RU-WRF.

2.4 Summary of Findings and Recommendations

A summary of the key results found in this section are as follows:

1. RU-COOL has recently adopted NREL’s mesoscale model validation methods used for NREL’s WIND Toolkit. These methods are robust for validation of modern mesoscale models. However, these methods are limited to validating wind speed and wind direction only, and not other atmospheric variables known to influence wind power production, such as atmospheric stability and turbulence.
2. It is not possible to confidently validate RU-WRF using coastal observation stations because of the coarse spatial resolution of RU-WRF (3 km) and the large coastal wind speed gradients.

Based on the analysis presented in this section, we recommend the following improvements to RU-COOL’s observational network and validation methods:

1. **Incorporate publicly available floating lidar data.** PNNL’s WindTracer lidar off the coast of New Jersey, as well as possible future deployments by DOE, provide a valuable validation dataset for modeled wind speeds, wind profiles, wind veer, and wind shear.
2. **Consider our recommended configuration for RU-COOL’s WindTracer lidar.** This configuration should allow for reliable wind profile measurements at frequent intervals out to 8 km offshore, providing a valuable data set for model validation and wind resource characterization.
3. **Leverage existing or develop new partnerships with offshore wind energy developers.** Lidars are increasingly being deployed in New Jersey offshore wind energy lease areas. Accessing these data through existing or new partnerships could provide the highest-quality data available to validate RU-WRF.

4. **Do not use the RUCMM station and Oyster Creek met tower as principal validation stations.** Because of strong coastal wind speed gradients and the coarse horizontal spatial resolution of the RU-WRF model, the interpolation of modeled wind speeds to these stations is too uncertain to allow for a meaningful validation of model performance.
5. **Enhance validation methods to include additional variables beyond wind speed.** RU-COOL is currently following validation procedures outlined in an NREL technical report describing the WIND Toolkit (Draxl et al. 2015). We recommend continued use of these methods but also to expand the validation to atmospheric variables beyond wind speed that are known to influence the wind resource, most notably measures of atmospheric turbulence and stability.
6. **Include an RU-WRF and NOAA HRRR performance comparison in validation methods.** NOAA produces an hourly-updated, WRF-based mesoscale HRRR forecast product for the continental United States at the same spatial resolution as RU-WRF. Output is also publicly available. We recommend adding a comparison of RU-WRF and HRRR in the existing validation methods to help highlight the value of customized components of the RU-WRF model.

3 Coldest-Pixel SST Product Evaluation

This section provides an evaluation of the RU-COOL coldest-pixel SST product in the New Jersey offshore region. Performance is compared relative to the NCEP 1/12-degree resolution Real-Time Global (RTG) SST product.¹ This SST product is used extensively as input to WRF models across a wide range of atmospheric research institutes, including in NREL's WIND Toolkit hindcast model and NOAA's HRRR forecast model. The NCEP RTG SST product provides daily averages of SST and is derived predominately from AVHRR measurements (similar to the coldest-pixel product) but also incorporates in-situ observations from buoys and ships where and when available.

3.1 Analysis of Daily SST Data

The two SST products are compared against SST observations from four NDBC buoys — 44009, 44017, 44025, and 44065 – all of which had full SST data coverage over the validation time period. Because of the coastal gradient issue described in Section 2.1.2, no coastal stations reporting SST were used in this analysis. A validation period of June 1, 2015 to May 31, 2016, was used given the near-complete coverage of the RU-COOL coldest-pixel SST product during this period.

Figure 6 shows observed and modeled daily SST at each of the buoys over the validation period. In general, both SST products show strong agreement with observations throughout the year (R^2 correlations exceeding 0.99 in all cases). There are, however, several cases where modeled SST deviate from the observations. The NCEP SST product underestimates SST at most buoys during the winter and often overestimates SST in the spring. The RU-COOL SST product generally seems more accurate than the NCEP SST product but there are several isolated cases where modeled SST deviates considerably from the observed (e.g., Buoy 44017 around May 2016, Buoy 44065 around February 2016). Furthermore, at Buoys 44009 and 44017, the RU-COOL SST product clearly lags the observations by several days. A detailed investigation of these discrepancies is beyond the scope of this validation work.

Key performance metrics (i.e., RMSE, bias, and unbiased RMSE) for the daily modeled SST products are presented in Figure 7. The NCEP SST product has lower RMSE and unbiased RMSE for Buoys 44009 and 44065, whereas the RU-COOL SST product has lower RMSE and unbiased RMSE at Buoys 44017 and 44025. The NCEP SST product has considerably higher magnitudes of bias at Buoys 44009 (negative bias) and 44025 (positive bias) whereas the RU-COOL SST product has a higher magnitude of bias at Buoy 44065. Both products have low bias at Buoy 44017.

The bias and unbiased RMSE performance metrics on the daily modeled data are separated by calendar month in Figures 8 and 9, respectively. Figure 8 reveals that bias in the NCEP SST product can be strongly seasonal (e.g., Buoys 44017 and 44025) with a tendency to overestimate SST in spring/summer and underestimate SST in fall/winter. The RU-COOL SST product demonstrates the opposite seasonal trend (i.e., underestimating SST in spring/summer, overestimating SST in fall/winter) but at considerably less magnitude of bias relative to the NCEP SST product.

Figure 8 reveals that the RU-COOL coldest-pixel SST product tends to show a seasonal cycle in RMSE which peaks in the spring and summer months. During these periods, RMSE in the coldest-pixel SST product is considerably higher than that of the NCEP RTG SST product.

3.2 Potential Value of Hourly SST Data

Both the RU-COOL and NCEP SST products are daily or multi-day composites. When these products are used as boundary conditions to a mesoscale model, the daily SST data are interpolated for each mesoscale model time step, usually linearly. These daily SST averages do not account for diurnal variations in SST, which can be significant in the MAB. In Figure 10, the mean deviation from monthly average temperature is shown for each hour of the day and each month of the year based on in-situ observations at Buoy 44065. In the summer months, a change of more than 1°C in SST over the course of an average day is observed.

¹https://polar.ncep.noaa.gov/sst/rtg_high_res/description.shtml

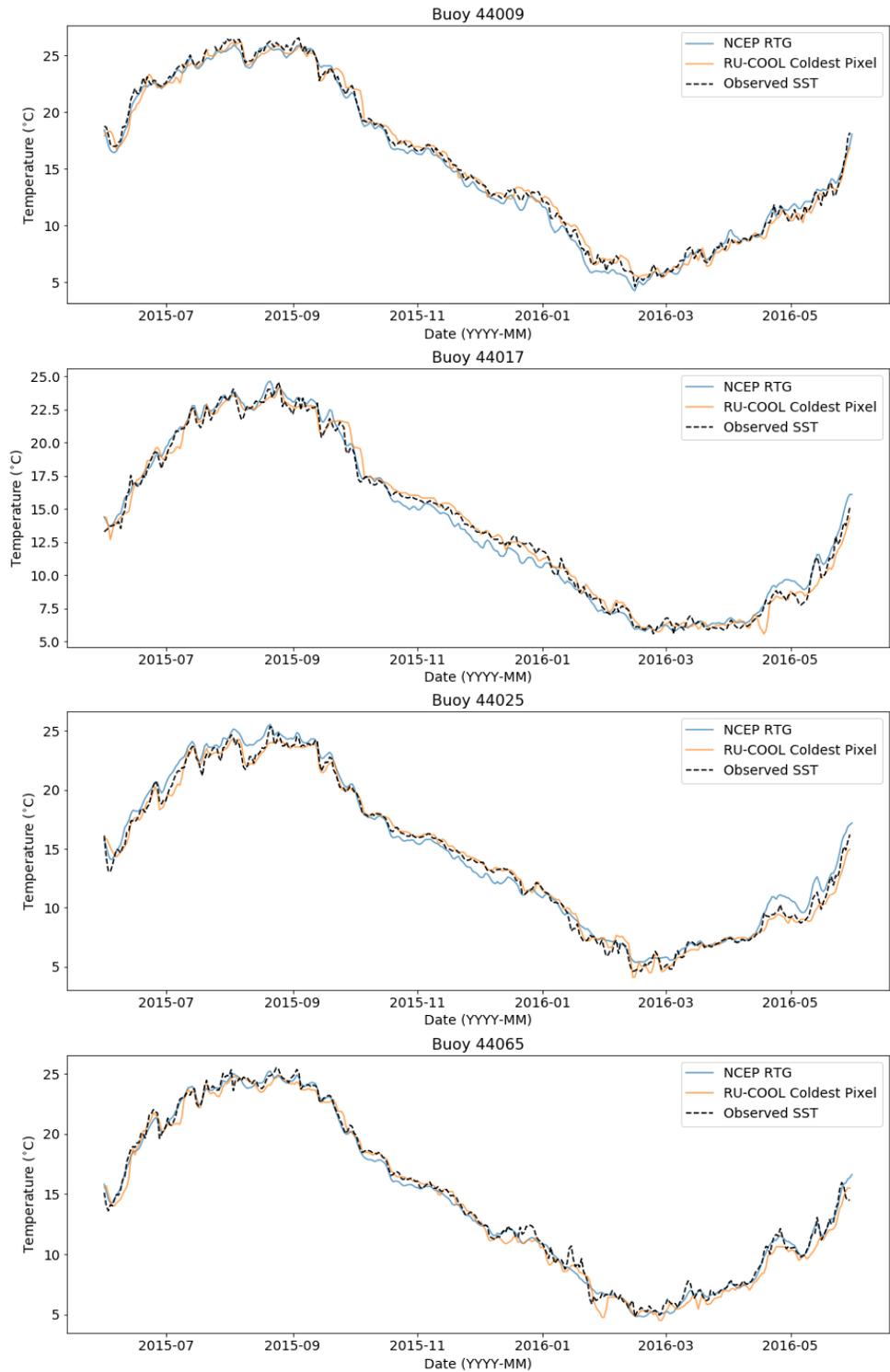


Figure 6. Observed and modeled daily SST at four buoy locations off the coast of New Jersey.

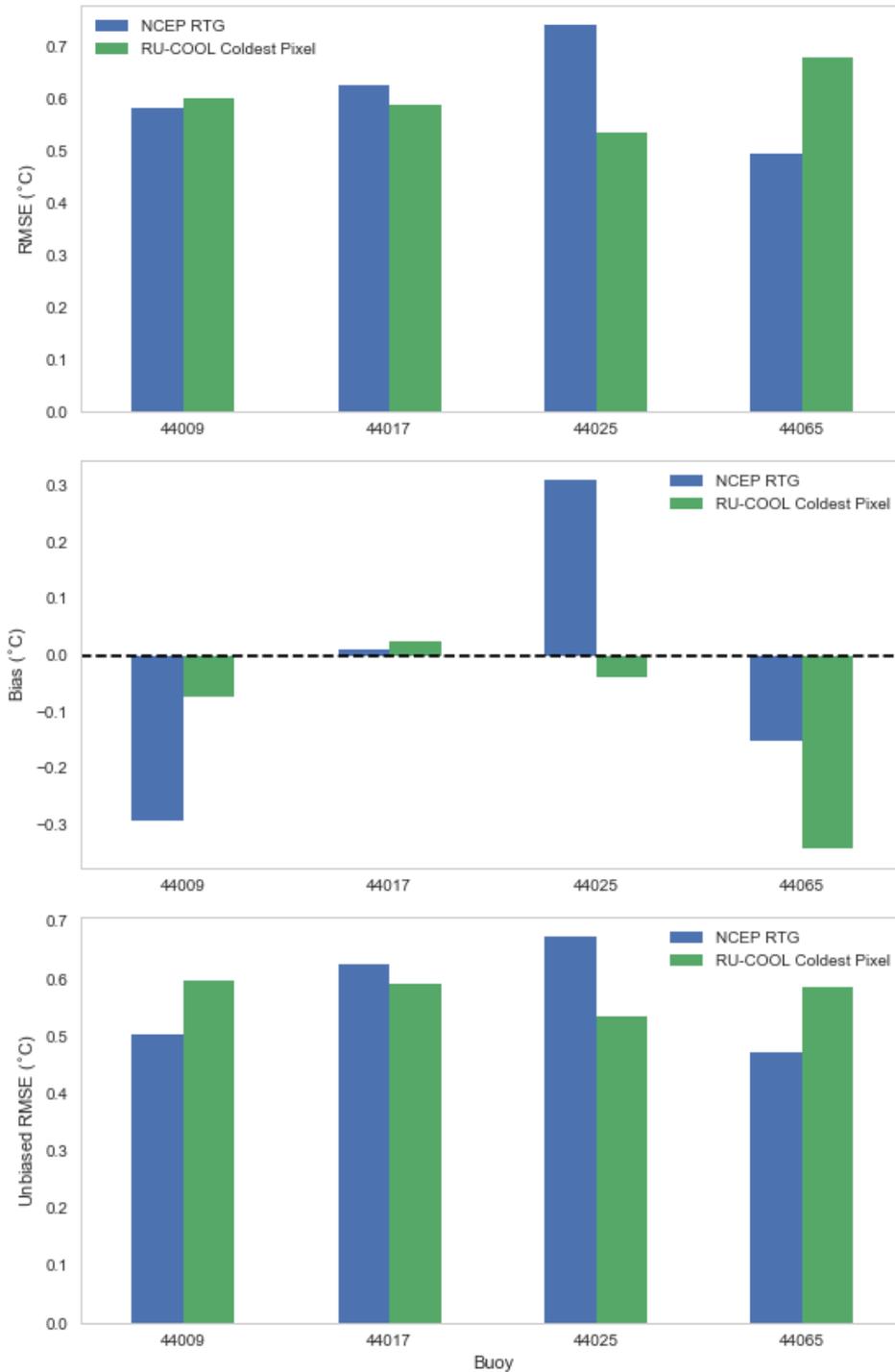


Figure 7. Performance metrics for the modeled daily SST products averaged over the full validation period

Given the role of the land-sea temperature gradient in influencing sea breeze events in the MAB, it is possible that

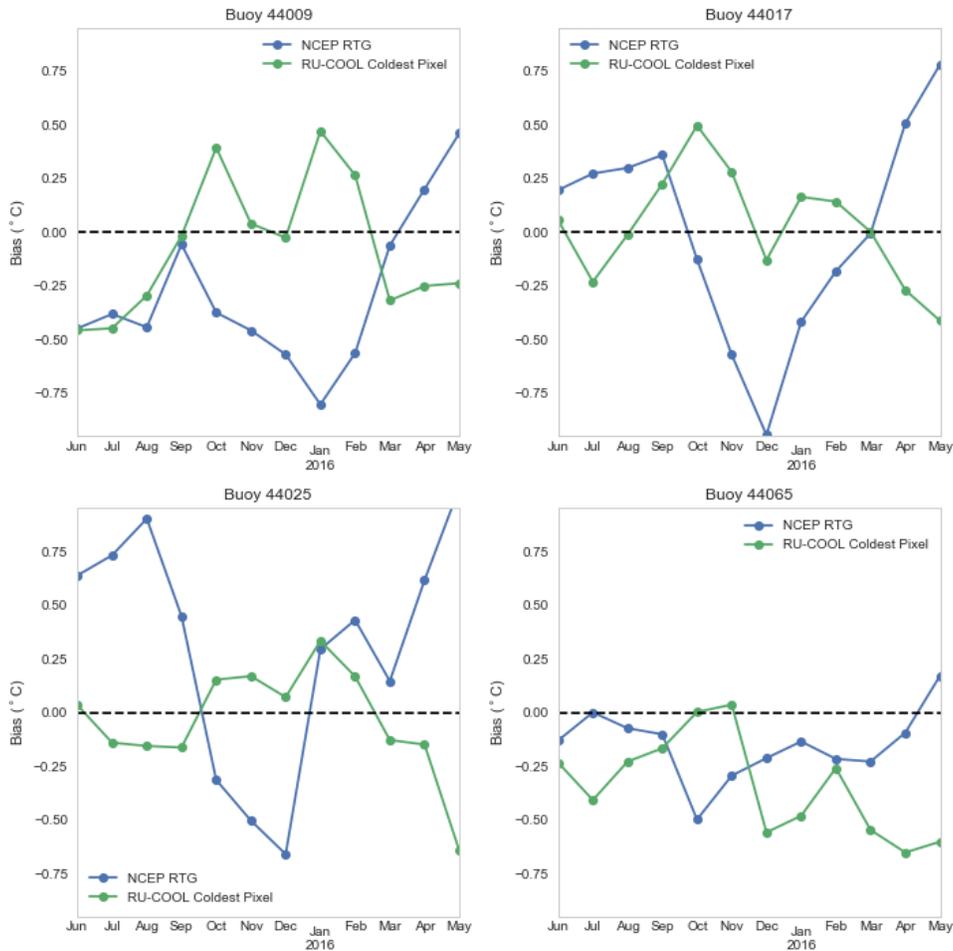


Figure 8. Mean bias in the daily modeled SST products separated by calendar month.

accounting for this diurnal variation of SST within a mesoscale model may improve simulation accuracy compared to use of typical daily SST data. Therefore, we recommend that RU-COOL explore possible data sets or interpolation methods to produce an hourly SST data set that is representative of the diurnal evolution of SST in the MAB.

3.3 Summary of Findings and Recommendations

A summary of the key results found in this section are as follows:

1. Overall, the RU-COOL coldest pixel SST product was found to be modestly more accurate compared to a widely-used SST product from the NCEP.
2. Relative performance of the two SST products depended significantly on location. Of the four buoy locations considered, the coldest-pixel product was more accurate at two locations and the NCEP product was more accurate at the other two. Causes for the relative geospatial performance of these two products would be a valuable subject for future research.
3. When used as a boundary forcing to RU-WRF, the coldest-pixel SST product provided minor improvements in wind speed modeling accuracy relative to the NCEP SST product.

4. SST measurements at the buoy locations demonstrated considerable diurnal cycles in the spring and summer months. Such cycles are not resolved in daily resolution SST products and may be significant in influencing the offshore wind resource.

Based on the analysis presented in this section, and recognizing that a detailed assessment of the coldest-pixel SST product was not within the scope of this study, we make the following recommendations for improvement of the coldest-pixel SST product:

1. Consider using the NCEP SST product as a backup to the coldest-pixel product rather than the current NASA SPoRT backup product.
2. Consider a blended SST product that combines the best attributes of the RU-COOL coldest-pixel SST product and the NCEP SST product.
3. Explore available data sources or derive methods to produce an hourly SST product that accounts for the diurnal evolution of SST during the spring and summer.
4. Investigate the causes for high negative bias at Buoy 44065.
5. Investigate high RMSE during the spring/summer months at all buoys and the low relative bias of the NCEP RTG product during these periods.

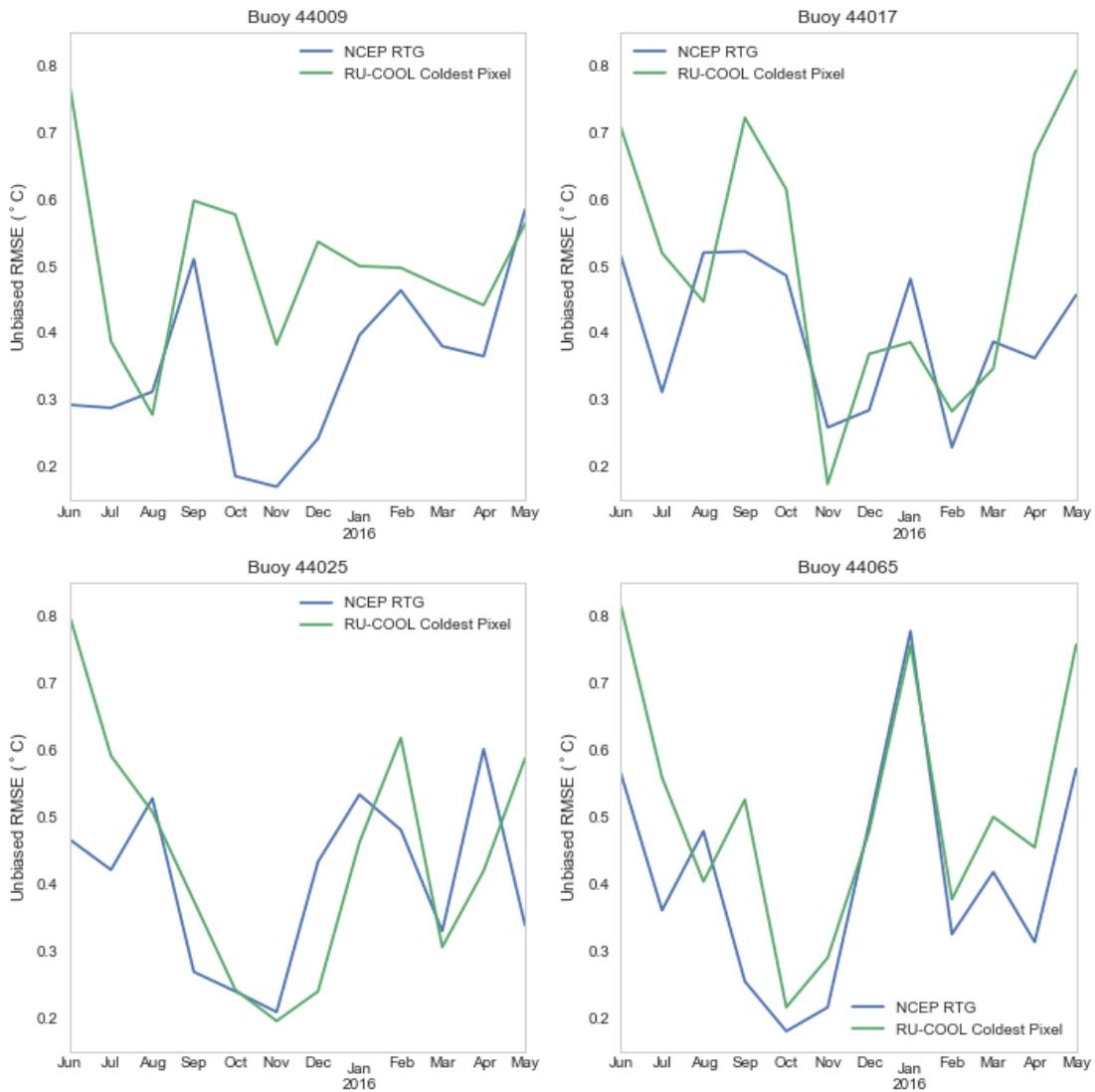


Figure 9. Unbiased RMSE for the daily modeled SST products separated by calendar month.

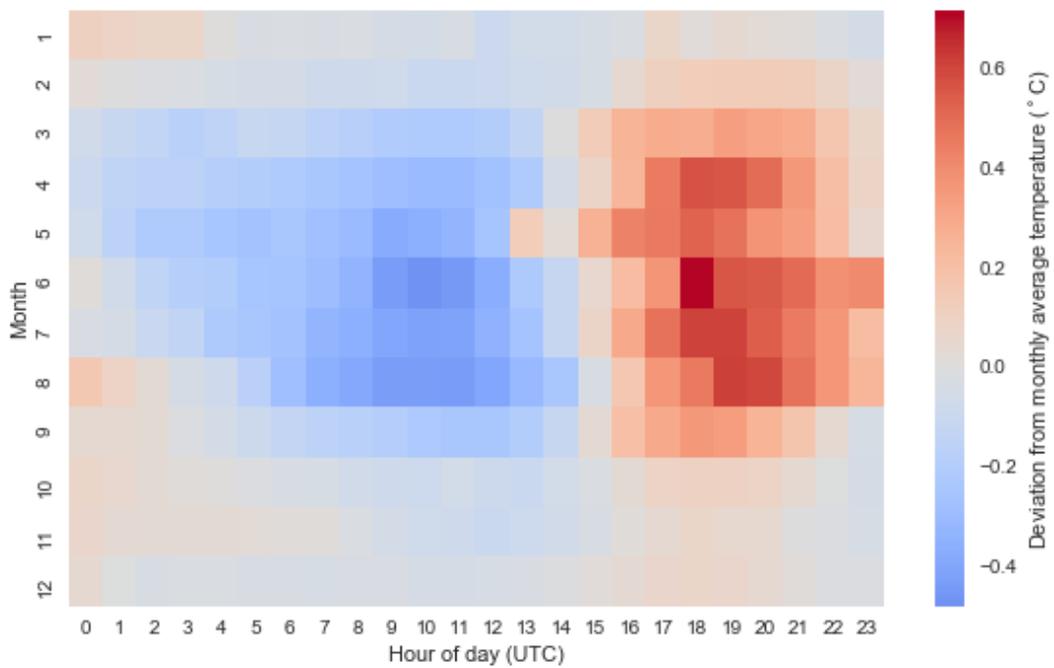


Figure 10. Deviation from monthly average SST for the hourly SST measurements at Buoy 44065.

4 Hindcast Validation of RU-WRF

This section provides an evaluation of the RU-WRF model run in hindcast mode (i.e., forced by reanalysis data). Performance is compared relative to an ensemble of WRF setups based on the key differences between RU-WRF and NREL's WIND Toolkit.

4.1 Ensemble Members for Hindcast Validation

The following four categories were considered when building the hindcast ensemble and are summarized in Table 4:

1. **Atmospheric forcing product:** WRF simulations depend significantly on the large-scale atmospheric product used as boundary forcing for the WRF simulation. In this study, we consider the use of the ERA-Interim reanalysis product¹ used in WIND Toolkit and the NCEP $1^\circ \times 1^\circ$ Final Operational Global Analysis (FNL)². Although not itself a reanalysis product, the FNL was the closest hindcast equivalent to the GFS forecast product used in the forecast mode of RU-WRF. We constructed a hindcast product from NCEP FNL by only extracting model output at initialization (every 6 hours) and concatenating these outputs together to produce a continuous 6-hour time series over the validation period. Furthermore, the $1^\circ \times 1^\circ$ resolution NCEP FNL product was used in this study and not the $0.25^\circ \times 0.25^\circ$ product because the latter was not made available until July 8, 2015, or 38 days into the validation period. This $1^\circ \times 1^\circ$ spatial resolution translates roughly to 120-km latitude \times 90-km longitude.
2. **SST input** The RU-COOL coldest-pixel and NCEP RTG SST products are both considered as possible SST inputs for the ensemble. To assess the value of these high-resolution SST data products, we also considered the default SST data included in both the ERA-Interim and NCEP FNL large-scale forcing products as an additional ensemble member.
3. **WRF version** The RU-WRF model is currently based on WRF version 3.9, whereas the WIND Toolkit, released in 2015, was based on WRF version 3.4. A comparison between WRF versions 3.4 and 3.9 was not of practical interest in this study. Given the continued advancement of WRF, RU-COOL expressed considerable interest in improving RU-WRF using new WRF versions. For this reason, we considered WRF version 3.9 as well as the recent WRF version 4.0 (released June 2018) in the ensemble. Version 4.0 of WRF feature numerous updates³ including improvements to the Mellor, Yamada, Nakanishi, and Niino (MYNN) PBL scheme used in RU-WRF.
4. **WRF namelists** The namelist.wps and namelist.input files are text files read in by WRF in which the key properties of the model are specified (e.g., land surface schemes, terrain data, parameterization schemes, spatial resolution, model dynamics). Given the range of differences between the WIND Toolkit and RU-WRF namelists (summarized in Table 5), it was not practical in this study to consider each difference when constructing the ensemble. Rather, only the WIND Toolkit and RU-WRF namelists as a whole were considered when building the ensemble.

In total, 24 ensemble members were constructed considering all possible combinations of model inputs and setups described above and summarized in Table 4.

4.2 WRF Simulation Setup

For the hindcast validation over the 1-year validation period, we ran the WRF model in separate 3-day simulation periods each offset by 2 days. The first 24 hours in the simulation were used only for WRF model spin-up and were

¹The European Centre for Medium-Range Weather Forecasts (ECMWF) released the ECMWF Reanalysis Interim (ERA-Interim) product in 2006 (<https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim>). The product has a horizontal spatial resolution of 80 km and has 60 vertical levels from the surface to 0.1 hPA.

²The NCEP Final Operational Global Analysis is based on the same model as the GFS but is prepared about an hour after the GFS is initialized in order for more observational data to be incorporated (<https://rda.ucar.edu/datasets/ds083.3/>).

³<http://www2.mmm.ucar.edu/wrf/users/wrfv4.0/updates-4.0.html>

Table 4. A Summary of the WRF Model Components Used to Construct the 24-Member Hindcast Ensemble in this Validation Study.

Category	Values Considered for Hindcast Ensemble
Atmospheric forcing	NCEP FNL $1^\circ \times 1^\circ$ ERA-Interim
SST input	RU-COOL coldest-pixel NCEP RTG Default from atmospheric forcing
WRF version	3.9 4.0
WRF namelist	RU-WRF WIND Toolkit

Table 5. Comparison of Key Attributes Between RU-WRF and the WIND Toolkit.

Feature	RU-WRF	WIND Toolkit
Atmospheric forcing	GFS $0.25^\circ \times 0.25^\circ$ forecast	ERA-Interim reanalysis
Planetary boundary layer scheme	MYNN 2.5	Yonsei University
Surface layer scheme	Monin-Obukhov with Janjic Eta	Monin-Obukhov
Microphysics	Thompson	Ferrier
Longwave radiation	RRTMG	RRTM
Shortwave radiation	Goddard	Dudhia
Topographic data	USGS GTOPO30	USGS GTOPO30
Land-use data	NLCD 3-second	NLCD 3-second
Cumulus parameterization	Off	Kain-Fritsch
Vertical levels	40	41
Near-surface level heights (meters)	53.7, 131.0, 232.5	33.9, 67.9, 102.1, 136.4, 170.8

not considered further. Therefore, the the 1-year timeseries was constructed from 2-day simulation periods concatenated together. This approach was implemented because atmospheric nudging was not used in the WRF setup (i.e., the large-scale atmospheric forcing was only applied at the boundary of the WRF domain and not applied within the domain to “nudge” the WRF simulations towards the large-scale forcing). Without atmospheric nudging, the WRF simulations can drift away from the large-scale forcing over long simulation periods. Limiting the simulation period to 3 days therefore reduces the risk of this potential drift.

4.3 Data Setup and Post-Processing

Given the uncertainty associated with using coastal stations to validate a model at relatively low spatial resolution (Section 2.1.2), the four offshore NDBC buoys and the PNNL lidar were used as validation stations in this analysis. The 90-m wind speeds measured at the VIND1 lidar are mostly used in this validation, although wind profiles are also considered. Given the known negative biases of this data set, particularly above 90 m (Newsom 2016), results from the wind profile analysis are interpreted cautiously.

Given the limited data provided by these buoys and lidar, no analysis of atmospheric stability or turbulence was possible. However, the Oyster Creek met tower, although located in the coastal zone, provides measurement data to

support such an analysis. Given the coastal gradient issue (Section 2.1.2 and the limited value of coastal observations for offshore wind energy lease areas, these data were not considered for assessing model performance in characterizing atmospheric stability. Table 6 summarizes the observation stations and validation heights used in this validation study.

Table 6. Observation Stations and Validation Heights Used to Validate RU-WRF in this Study.

Station	Validation Height (m)
44009	5.0
44017	5.0
44025	5.0
44065	4.0
VIND1	55.0, 70.0, 90.0

As shown in Table 6, validation was performed at the measurement heights of each observation station rather than at a fixed height relevant to wind energy (e.g., 100 m). For the buoys in particular, which have measurement heights at either 4.0 m or 5.0 m, we made the judgment that the extrapolation of these wind speed measurements to a typical hub height, regardless of extrapolation method, would be associated with a prohibitively large uncertainty and would prohibit a confident validation of different mesoscale model setups. To interpolate the modeled 10 m winds (diagnostic outputs from WRF) to the buoy heights, the Monin-Obukhov-derived logarithmic wind speed profile is used:

$$U_{z_2} = U_{z_1} \frac{\ln\left(\frac{z_2}{z_0}\right) - \psi\left(\frac{z_2}{L}, \frac{z_0}{L}\right)}{\ln\left(\frac{z_1}{z_0}\right) - \psi\left(\frac{z_1}{L}, \frac{z_0}{L}\right)} \quad (4.1)$$

where U is wind speed, z_2 and z_1 are heights above the surface, z_0 is the roughness length, ψ is the stability function, and L is the Obukhov length. For the stability function, the Beljaars and Holtslag (1991) formulation is used.

To interpolate model results to the 90-m validation height for the PNNL VIND1 lidar, a basic linear interpolation is performed for the u - and v - components separately. Wind speed and direction are then computed from these interpolated wind components. The linear interpolation for the Wind Toolkit (WTK) setup is a reasonable approximation given the high vertical resolution (i.e., interpolation between on average 67.9 m and 102.1 m model level heights). However, this approach is less reasonable for the RU-WRF setup given the low vertical resolution (i.e., interpolation between on average 53.7-m and 131-m model levels). To assess this reasonableness, we tested two identical RU-WRF setups, one at the default vertical resolution and the other at the finer WTK vertical resolution. Very small differences were found between the interpolated 90-m wind speeds at VIND1 and negligible differences were found in the relative validation metrics. Therefore, we believe linear interpolation to 90-m for VIND1 allows for a robust evaluation of the WRF ensemble.

4.4 Validation Results

In this section we present the results of the RU-WRF hindcast model validation. Throughout this section, the different ensemble members are labeled using the following nomenclature: '<Namelist>_<WRF version>_<Atmospheric Forcing>_<SST Forcing>'. Options for each heading name are as follows:

- <Namelist>: "rutgers" refers to the RU-WRF namelist, "wtk" refers to the WTK namelist
- <WRF version>: "3_9" refers to WRF version 3.9, "4_0" refers to WRF version 4.0
- <Atmospheric Forcing>: "eraI" refers to the ERA-I reanalysis, "fnl" refers to the FNL analysis

- <SST Forcing>: “ncep” refers to the NCEP SST product, “cp” to the RU-COOL coldest-pixel product, and “no_sst” to the default SST included in the atmospheric forcing product.

4.4.1 Overall Performance

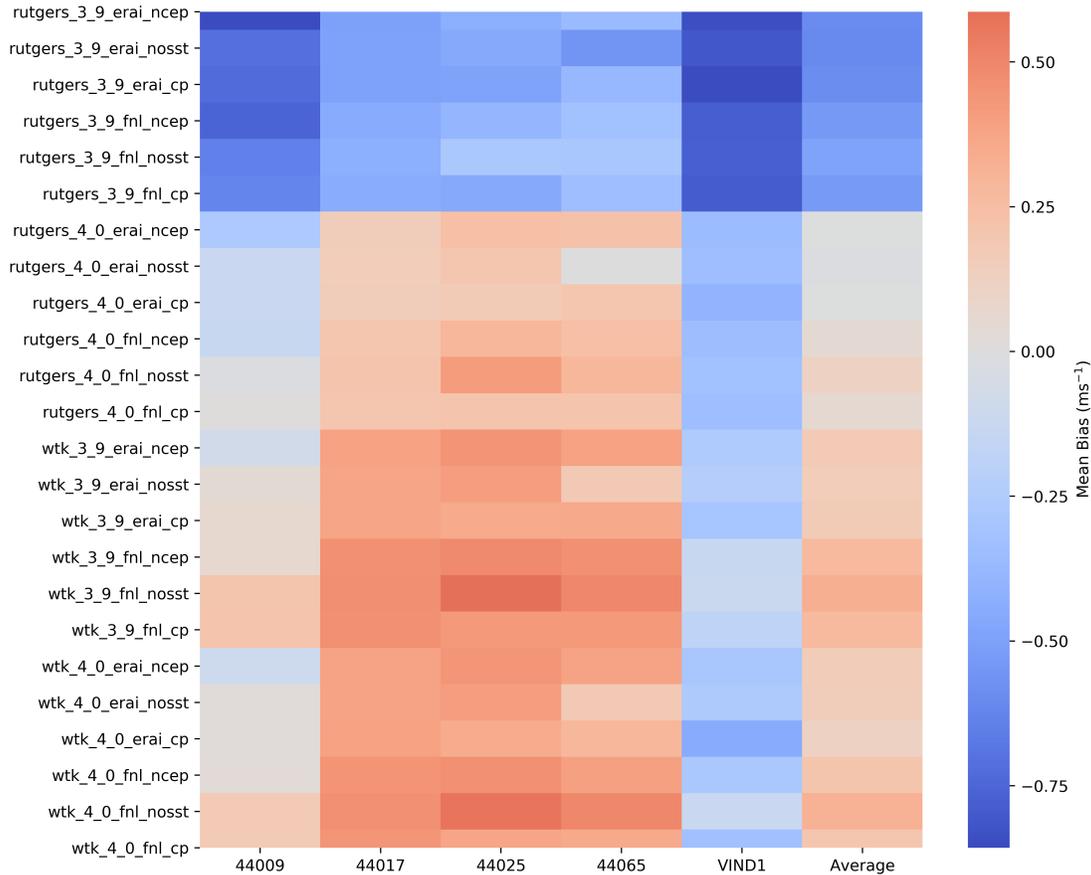


Figure 11. Mean bias for each ensemble member at each validation station. The “Average” column on the far right is the mean of the previous five columns.

Heatmaps of the mean bias and unbiased RMSE for each ensemble member at each validation station are shown in Figures 11 and 12, respectively. All ensemble members show a negative bias at the VIND1 lidar station, especially those that use the RU-WRF namelist and WRF version 3.9. The bias is reduced considerably when the RU-WRF namelist is used with WRF version 4.0. In fact, all ensemble members using the RU-WRF namelist and WRF version 3.9 setup have a mean negative bias across all validation stations. By contrast, the remaining ensemble members generally show a mean positive bias at the buoy stations. On average across the stations, ensemble members using the RU-WRF namelist and WRF version 4.0 have the lowest magnitude of bias, whereas ensemble members using the RU-WRF namelist and WRF version 3.9 have the highest magnitude of bias.

In Figure 12, the unbiased RMSE is highest at the VIND1 station whereas values are considerably lower at the buoy stations. This result is as expected given the larger wind speeds at 90 m relative to 5 m. In general, the ensemble members using the RU-WRF namelist have lower unbiased RMSE than those using the WTK namelist. Furthermore, the RU-WRF namelist ensemble members using WRF version 4.0 appear to have slightly less error than those using WRF version 3.9.

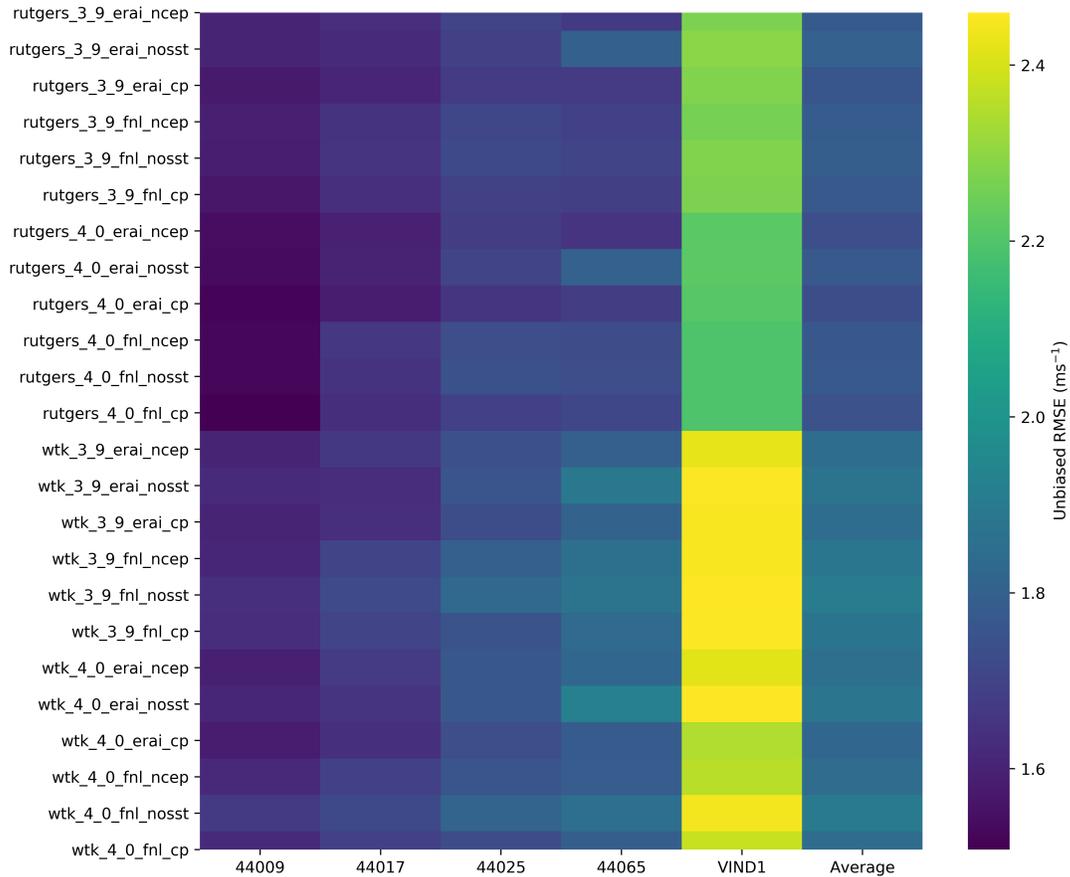


Figure 12. Unbiased RMSE for each ensemble member at each validation station. The “Average” column on the far right is the mean of the previous five columns.

A summary of key performance metrics for all ensemble members is shown in Table 7. The first section — “Namelist–WRF version” — consists of the mean results across all 24 ensemble members. The second section — “Atmospheric–SST Forcing” — consists of the results considering only the ensemble members that use the RU-WRF namelist and WRF version 4.0. Results show that the RU-WRF/WRF 4.0 models have the lowest error and highest correlation compared to the other namelist/WRF version setups, although in many cases the difference is minor. The improved performance of RU-WRF/WRF 4.0 compared to RU-WRF/WRF 3.9 is predominantly attributed to a considerably lower magnitude of bias. By contrast, the WTK namelist is not significantly affected by the change of WRF version. A detailed investigation into what specific changes in WRF versions 3.9 to 4.0 are leading to this improvement with the RU-WRF namelist is beyond the scope of this analysis; however, it is likely related to improvements made to the MYNN PBL closure scheme in WRF version 4.0, which is used in the RU-WRF namelist.

Table 7 shows that using an RU-COOL coldest-pixel SST product results in slightly less error than the NCEP RTG product, both of which provide more accurate results than when the default SST values in the atmospheric forcing product is used. The use of the ERAI reanalysis results in slightly higher magnitude of bias than the FNL analysis but lower RMSE and unbiased RMSE.

Table 7. Key Performance Metrics for All Ensemble Members. The first section “Namelist – WRF version” consists of the mean results across all 24 ensemble members. The second section “Atmospheric – SST Forcing” consists of the results considering only the ensemble members that use the RU-WRF namelist and WRF version 4.0.

	RMSE (ms ⁻¹)	Bias (ms ⁻¹)	Unbiased RMSE (ms ⁻¹)	R ²
Namelist – WRF version				
Rutgers – 3.9	1.87	-0.56	1.78	0.76
Rutgers – 4.0	1.77	+0.03	1.75	0.76
WTK – 3.9	1.90	+0.22	1.87	0.74
WTK – 4.0	1.89	+0.19	1.86	0.74
Atmospheric – SST Forcing				
ERA1 – CP	1.74	-0.04	1.73	0.77
ERA1 – NCEP	1.76	-0.05	1.74	0.77
ERA1 – Default SST	1.78	-0.02	1.77	0.76
FNL – CP	1.76	+0.06	1.75	0.77
FNL – NCEP	1.79	+0.05	1.77	0.76
FNL – Default SST	1.79	+0.12	1.77	0.76

4.4.2 Seasonal Performance

The seasonal variations of bias and unbiased RMSE are presented in Figure 13. The top panels show the mean bias and the bottom panels show the unbiased RMSE. The leftmost panels consider variations in the namelist and WRF version while keeping the atmospheric and SST inputs constant (set to ERA1 and the coldest-pixel SST products, respectively). The rightmost panels consider variations in the atmospheric and SST forcings while keeping the namelist and WRF versions constant (set to the RU-WRF namelist and WRF version 4.0, respectively).

Figure 13(a) shows that the RU-WRF/WRF 3.9 setup demonstrates considerable negative bias for most of the year but low bias from June through August. The remaining model setups demonstrate the opposite trend (i.e., low bias for most of the year but high positive bias from June through August).

Figure 13(c) shows that all models generally have relatively low unbiased RMSE from July through November and relatively higher RMSE in the other months. In general, the RU-WRF namelist setups have lower unbiased RMSE than the WTK namelist setups. The RU-WRF / WRF 4.0 setup has lower unbiased RMSE than the RU-WRF / WRF 3.9 setup and no clear seasonal pattern is evident in this relative performance.

Figure 13(b) shows that relative bias from models using different atmospheric and SST forcings is most significant in the spring and summer months (where large positive biases are observed) and much less significant in the fall and winter months (when biases are on average low). Relative biases are higher for models using different atmospheric products than those using different SST products.

Figure 13(d) shows that the ERA1 forcing generally results in lower unbiased RMSE compared to the FNL analysis. Generally, negligible seasonal differences are observed in RMSE for models using different SST products.

4.4.3 Diurnal Performance

The mean diurnal evolution of the modeled and observed wind speed is shown in Figures 14 and 15 at Buoy 44025 and the VIND1 lidar, respectively (noting the only 7-month coverage for VIND1). Diurnal plots for the remaining validation stations are provided in the Appendix. The top panel in each figure considers variations in the namelist and WRF version while keeping the atmospheric and SST inputs constant (set to ERA1 and the coldest-pixel SST products, respectively). The lower panel in each plot considers variations in the atmospheric and SST forcings while keeping the namelist and WRF version constant (set to the RU-WRF namelist and WRF version 4.0, respectively).

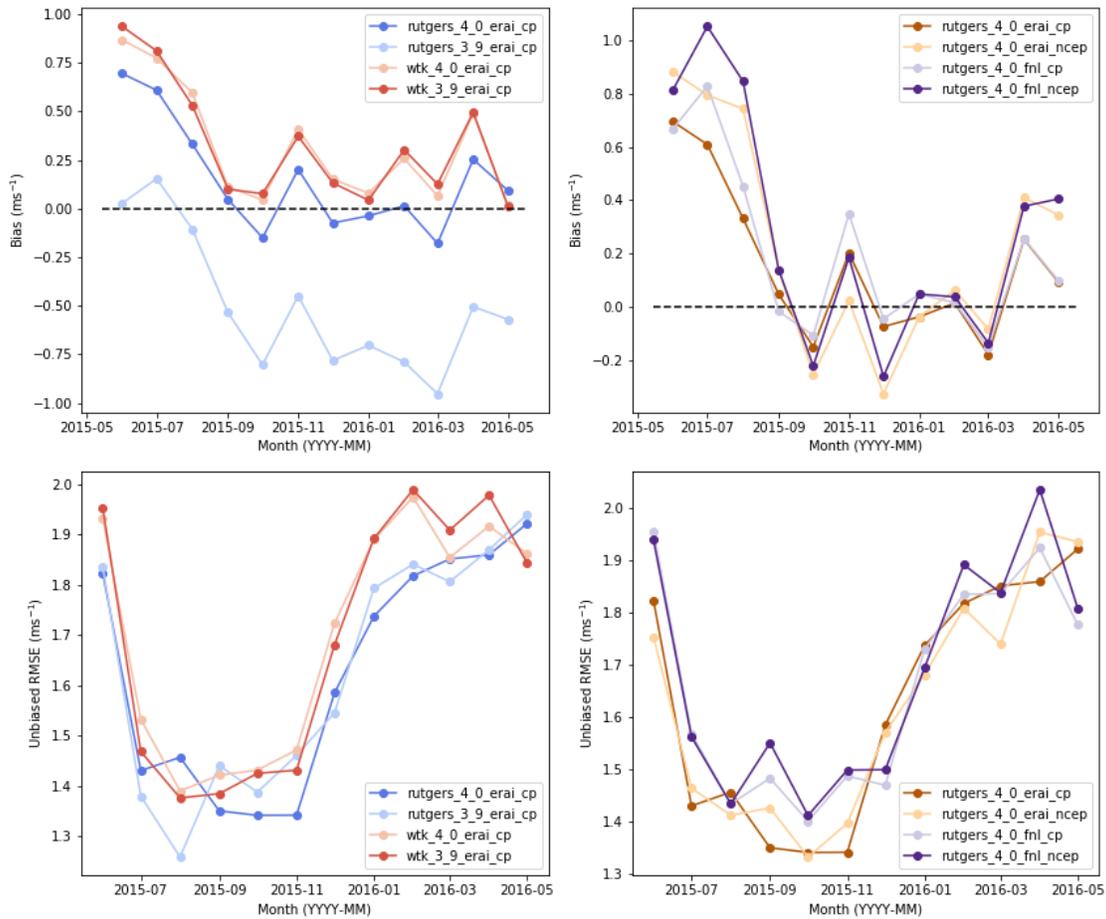


Figure 13. Monthly performance metrics averaged across all validation stations for select WRF ensemble members. The leftmost panels consider bias and unbiased RMSE for different combinations of WRF versions and namelists while keeping the atmospheric and SST forcings constant. The rightmost panels consider bias and unbiased RMSE for different combinations of atmospheric and SST forcings while keeping the WRF version and namelist constant.

At Buoy 44025, the RU-WRF / WRF 4.0 model best captures the diurnal wind speed evolution at 5 m above sea level (Figure 14[a]). It is most accurate from about 07:00 UTC through 17:00 UTC and tends to overestimate wind speeds in the remaining hours. The RU-WRF / WRF 3.9 model shows a similar trend but has a large negative bias. The WTK namelist models are approximately equivalent, follow a similar diurnal evolution as the RU-WRF namelist models, and have large positive bias at most hours with lower magnitude positive bias from about 10:00 UTC to about 17:00 UTC.

In Figure 14[b], all modeled profiles show a similar diurnal trend as the observations. Models tend to be in strong agreement from 18:00 UTC to 24:00 UTC and diverge more from 00:00 UTC to 16:00 UTC. Different atmospheric forcing products tend to produce most of this variability, although variability caused by different SST products is non-negligible.

The diurnal cycle of 90-m wind speeds at VIND1 (Figure 15) is noticeably different than the 5-m cycle at Buoy 44025. The magnitude of the diurnal cycle is less, the peak in wind speed is reached later in the day, and no distinct minimum in the cycle is found. These differences can be attributed to a shorter validation period (November to May

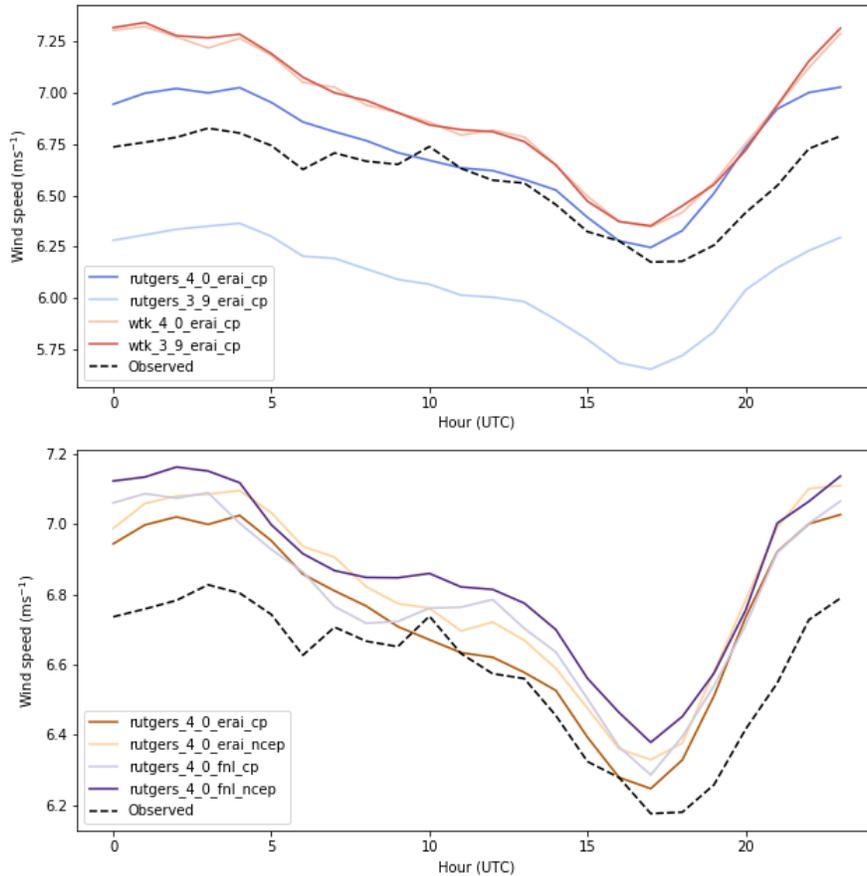


Figure 14. Mean diurnal evolution of the modeled and observed 5-m wind speed at Buoy 44025. The top figure considers model variations in the namelist and WRF version. The bottom figure considers variations in atmospheric and SST forcings.

and missing the summer season where diurnal cycles are strongest), the coastal location, and the different wind regime at 90 m.

The models tend to represent the diurnal cycle reasonably well, though all are negatively biased. The RU-WRF/WRF 4.0 model is generally consistent with the WTK namelist models. Models using different atmospheric and SST inputs are generally consistent from 00:00 UTC to 09:00 UTC but diverge in the remaining hours. Most of this divergence is attributable to different atmospheric products. In particular, the ERAI-based models show a larger magnitude diurnal cycle with lower minimum wind speeds than the FNL-based models. All models tend to capture the diurnal cycle but underestimate wind speeds, particularly during the 08:00 UTC – 18:00 UTC period.

4.4.4 Wind Profile Comparison

Observed and modeled mean wind speed profiles at VIND1 over the November 2015 – May 2016 period are shown in Figure 16. As in Section 4.4.3, a subset of models are considered to capture variations in WRF namelist and WRF version (Figure 16[a]) and atmospheric and SST forcings (Figure 16[b]). The VIND1 lidar provides wind speed measurements up to 160 m. However, because of known negative biases caused by generally low signal-to-noise ratios (Newsom 2016), especially at higher altitudes, only the wind speed profile up to 111 m is considered. The measurements up to 111 m used in this analysis are interpreted cautiously.

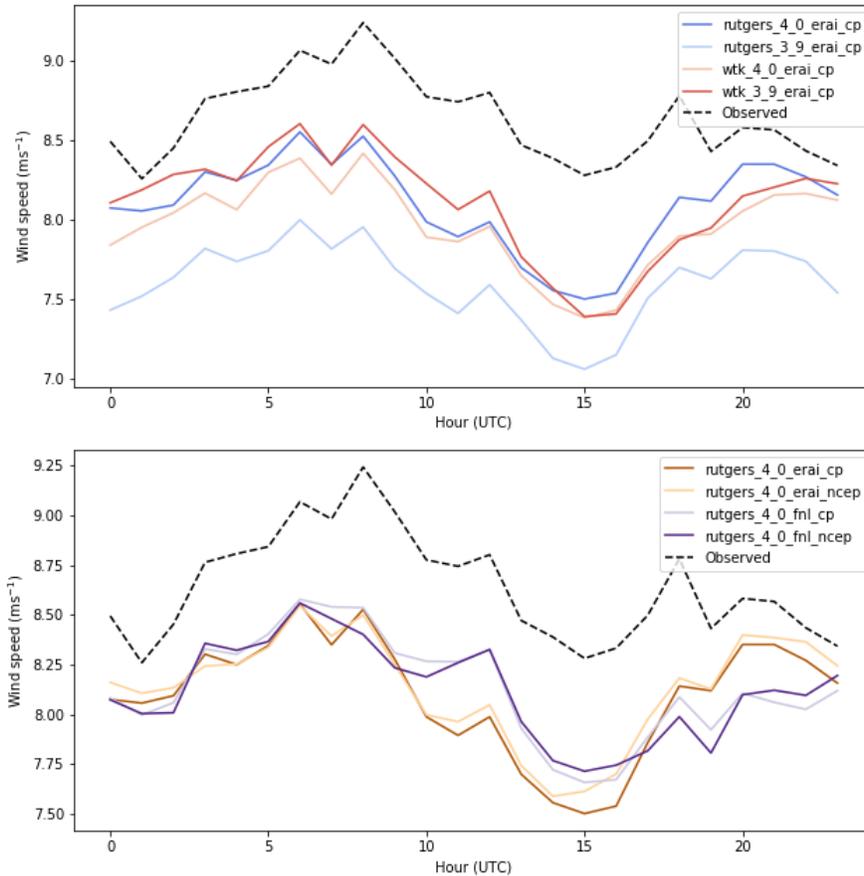


Figure 15. Same as Figure 14 but for the 90-m wind speeds at the VIND1 lidar.

The mean observed wind speeds increase from 55 m to 111 m with larger wind shear at lower heights. There is high sensitivity in the modeled wind profile to different namelists and WRF versions, particularly for RU-WRF namelist setups. Wind speed gradients from the WTK namelist models are generally in better agreement with observations than the Rutgers-namelist-based gradients. This fact is likely attributed to the low vertical resolution of the RU-WRF model. As seen in the figure, only two vertical levels between about 50 m and 130 m are used in RU-WRF, which is an insufficient number of points to adequately resolve the wind profile at heights relevant to wind power. By contrast, the WTK namelist-based wind speed profiles are relatively well-resolved. The performance of RU-WRF with higher vertical resolution is provided in Section 4.4.5.

Interestingly, the average height of the RU-WRF lowest model levels change between WRF versions 3.9 and 4.0. Average heights in WRF 3.9 are about 55 m and 130 m whereas average heights in WRF 4.0 are about 50 m and 115 m. The RU-WRF namelist does not explicitly define the vertical levels to use but rather specifies the total number of vertical model levels. Therefore, some change in WRF 3.9 to 4.0 in how model levels are assigned must have occurred. By contrast, for the WTK namelist, in which the exact model levels to use are specified, the model heights are consistent between WRF versions 3.9 and 4.0.

There is moderate to low sensitivity of the modeled wind profiles to different atmospheric and SST forcings (Figure 16[b]). The shape of the modeled wind profiles is generally consistent with small differences in wind speed magnitude.

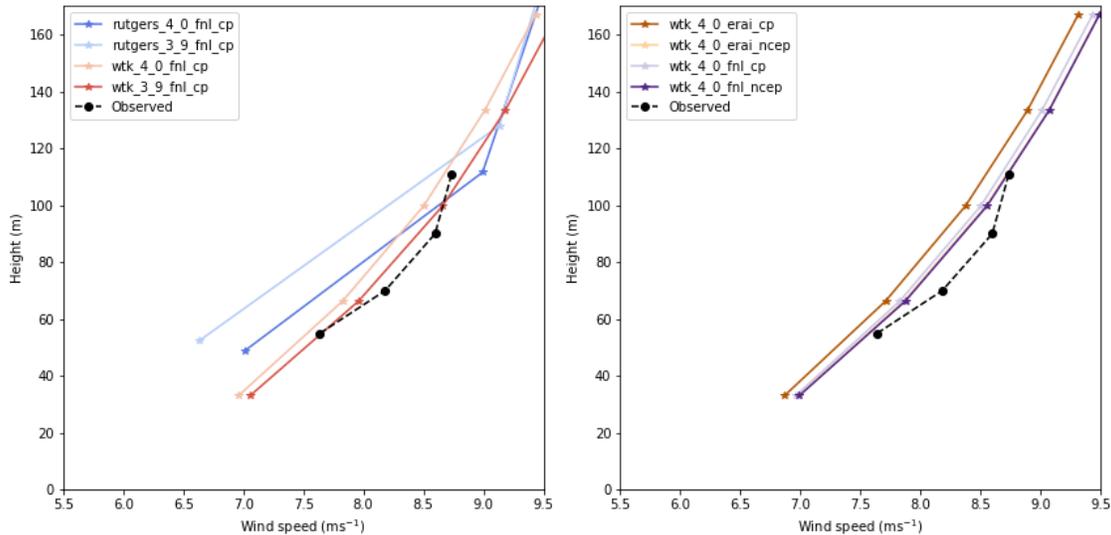


Figure 16. Mean modeled and observed wind profiles at the VIND1 lidar considering (a) variations in model namelist and WRF version and (b) variations in atmospheric and SST forcings.

4.4.5 Effect of Increased Vertical Resolution

As noted in Table 5, RU-WRF is currently setup with only two vertical levels below 150 m. To assess the value of increased vertical resolution in RU-WRF, the model setup using the Rutgers namelist, WRF version 4.0, FNL atmospheric forcing, and coldest-pixel SST forcing was run for two cases: (1) default vertical resolution provided in RU-WRF, and (2) vertical resolution provided in WTK.

Figure 17 shows the mean wind profiles from the two model setups. Increased vertical resolution in RU-WRF has considerable impact on the mean wind profile. The modeled wind profile at higher vertical resolution more closely resembles the observed wind profile. Large differences at lower heights are observed where the high-resolution profile shows about a 1 ms^{-1} greater wind speed than the low-resolution profile at about 60 m height.

4.4.6 Wind Direction Comparison

Modeled and observed distributions of 5-m wind direction at Buoy 44025 over the validation period are considered in Figure 18. Buoy 44025 was selected because it is located directly in a proposed wind energy lease area. Therefore, an accurate characterization of wind direction is especially important here given the relationship between wind direction and inter- and intra-plant wake losses.

The observed distribution shows a predominant southwest wind direction that is well represented by the models. All models show an offset around the west and northwest sector. Models using the RU-WRF namelist better capture the distribution in the 0° to 110° range, whereas the WTK namelist is generally more accurate in the other sectors. The ERAI-forced simulations generally have more accurate distributions than FNL-forced simulations.

4.5 Summary of Findings and Recommendations

A summary of the key results found in this section are as follows:

1. The current vertical resolution of RU-WRF is too coarse to accurately model vertical wind profiles. Increasing the vertical resolution of RU-WRF near the surface resulted in a much better representation of the wind profile.
2. The RU-WRF model, currently implemented using WRF version 3.9, demonstrated moderate to large negative biases at all the observation locations used for validation. These biases were substantially reduced when RU-

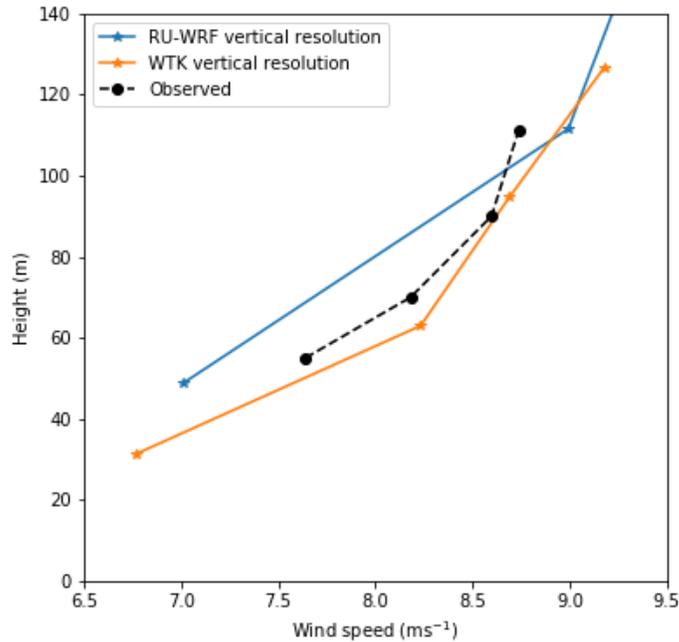


Figure 17. Mean modeled and observed wind profiles at the VIND1 lidar for two vertical resolutions of the RU-WRF model.

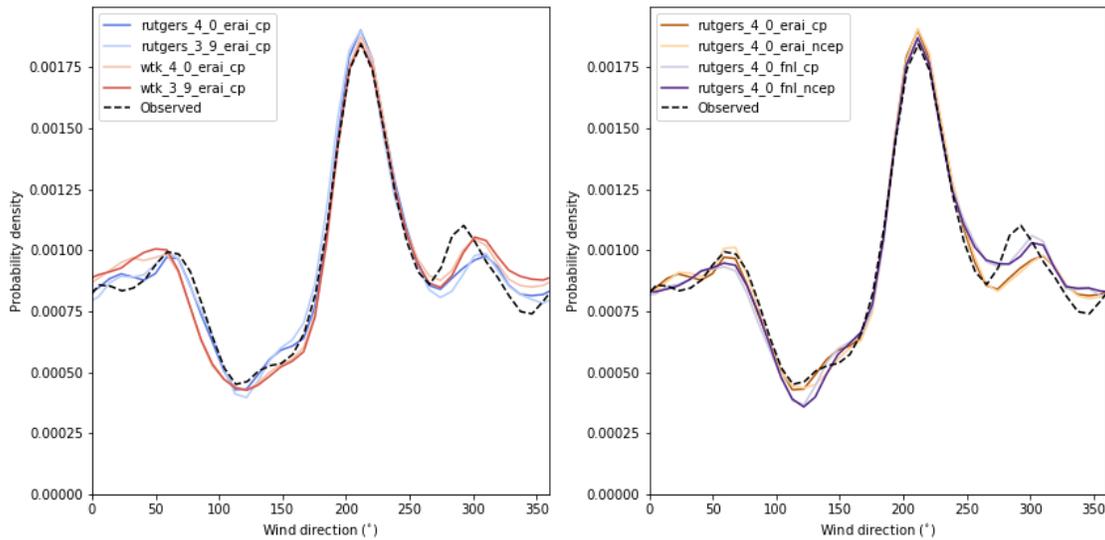


Figure 18. Modeled and observed 5-m wind direction distributions at Buoy 44025. Both variations in namelist and WRF version (a) as well as atmospheric and SST forcing (b) are considered.

WRF was implemented using WRF version 4.0. By contrast, the unbiased RMSE was reduced only slightly

when upgrading from WRF version 3.9 to 4.0.

3. The WIND Toolkit setup showed negligible performance changes when upgrading from WRF version 3.9 to 4.0. The WIND Toolkit showed a larger magnitude of bias compared to RU-WRF implemented with WRF version 4.0, but lower magnitude of bias compared to RU-WRF implemented with WRF version 3.9. The unbiased RMSE was larger in the WIND Toolkit compared to RU-WRF.
4. Differences in WIND Toolkit and RU-WRF model performance are attributed to the different planetary boundary layer parameterization schemes used in each. The scheme implemented in RU-WRF was found more accurate for offshore wind resource modeling. Increased performance in RU-WRF from WRF version 3.9 to 4.0 is believed to be related to improvements made in this scheme.
5. Most WRF models considered demonstrated large positive bias in the summer months and lower magnitude of bias in the remaining months. The cause for this seasonal trend would be a valuable subject for further research.

Based on the analysis presented in this section, we make the following recommendations for improvement to the RU-WRF model:

1. Upgrade the WRF version from the currently implemented version 3.9 to the recently released version 4.0
2. Increase the vertical resolution of RU-WRF by at least a factor of two to provide more accurate wind modeling at heights swept out by a wind turbine blade.
3. For resource assessment applications, develop a hindcast RU-WRF product that is forced by a large-scale reanalysis product.

5 Forecast Validation of RU-WRF

This section provides an evaluation of the forecast implementation of RU-WRF. Relative performance is assessed for RU-WRF when using two different large-scale forecast products to drive the simulations and the use of different forecast model spin-up times. Models in this analysis use the RU-WRF namelist, WRF version 4.0, and the coldest-pixel SST data.

5.1 Ensemble Members for Forecast Validation

For the large-scale forcing product, both the currently implemented GFS $0.25^\circ \times 0.25^\circ$ forecast product was compared against the NCEP North American Mesoscale (NAM) forecast model¹. Although both products are produced by NCEP, the NAM model features greater horizontal spatial resolution than GFS ($12 \text{ km} \times 12 \text{ km}$ compared to roughly $28 \text{ km} \times 20 \text{ km}$) and is run with a non-hydrostatic version of the WRF model. Given its availability for real-time forecasts on the U.S. East Coast, the performance of NAM relative to GFS in improving RU-WRF forecast predictions was of interest in this study.

For model spin-up times, the currently implemented 6-hour spin-up is compared against a 1-hour spin-up. The latter is consistent with current practices used for the NOAA HRRR model. For both spin-up times, only a 24 hour forecast simulation is performed. For a model initialization at 00:00 UTC, the 6-hour spin-up includes forecast data 6–30 hours after initialization, and the 1-hour spin-up includes forecast data 1–25 hours after initialization. Therefore, the difference between the two setups is the forecast horizon for the 01:00 UTC to 06:00 UTC window (i.e., much shorter for the 1-hour spin-up, much longer for the 6-hour spin-up). The key question here is if the forecast data 1 to 6 hours after initialization are inaccurate enough to justify the 6-hour spin-up, or if these data should be preserved.

Table 8. Atmospheric Forcing Products and WRF Model Spin-Up Times Considered When Assessing the RU-WRF Forecast Product.

Category	Setup Considered for RU-WRF Forecast Validation
Atmospheric forcing	NCEP GFS $0.25^\circ \times 0.25^\circ$ NCEP NAM $12\text{km} \times 12\text{km}$
Model spin-up time	6 hours 1 hour

5.2 WRF Simulation Setup

For the forecast validation over the 1-year validation period, daily WRF runs were performed with model initialization at 00:00 UTC (consistent with current RU-WRF implementation). The 1-hour spin-up model is run from 00:00 UTC to 01:00 UTC the next day with the first hour discarded. The 6-hour spin-up is initialized at the same time but run out to 06:00 the next day with the first 6 hours discarded. For both models, a total of 24 simulation hours are preserved. The full 1-year simulated time series are constructed by concatenating the individual 24-hour simulations.

5.3 Validation Results

In this section, the following labels are used to describe the different models considered:

- gfs: refers to the default RU-WRF setup that uses the GFS large-scale forcing product and the 6-hour spin-up period.
- nam: refers to the WRF setup that uses the NAM large-scale forcing product and the 6-hour spin-up period.

¹<https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesoscale-forecast-system-nam>

- gfs_mod: refers to the WRF setup that uses the GFS large-scale forcing product and the 1-hour spin-up period.

Performance metrics for each forecast model setup for each of the validation stations are shown in Figure 19. In general, the use of the GFS product results in slightly higher model performance compared to the use of the NAM product. The use of a 1-hour compared to a 6-hour spin-up time results in a considerable increase in model performance.

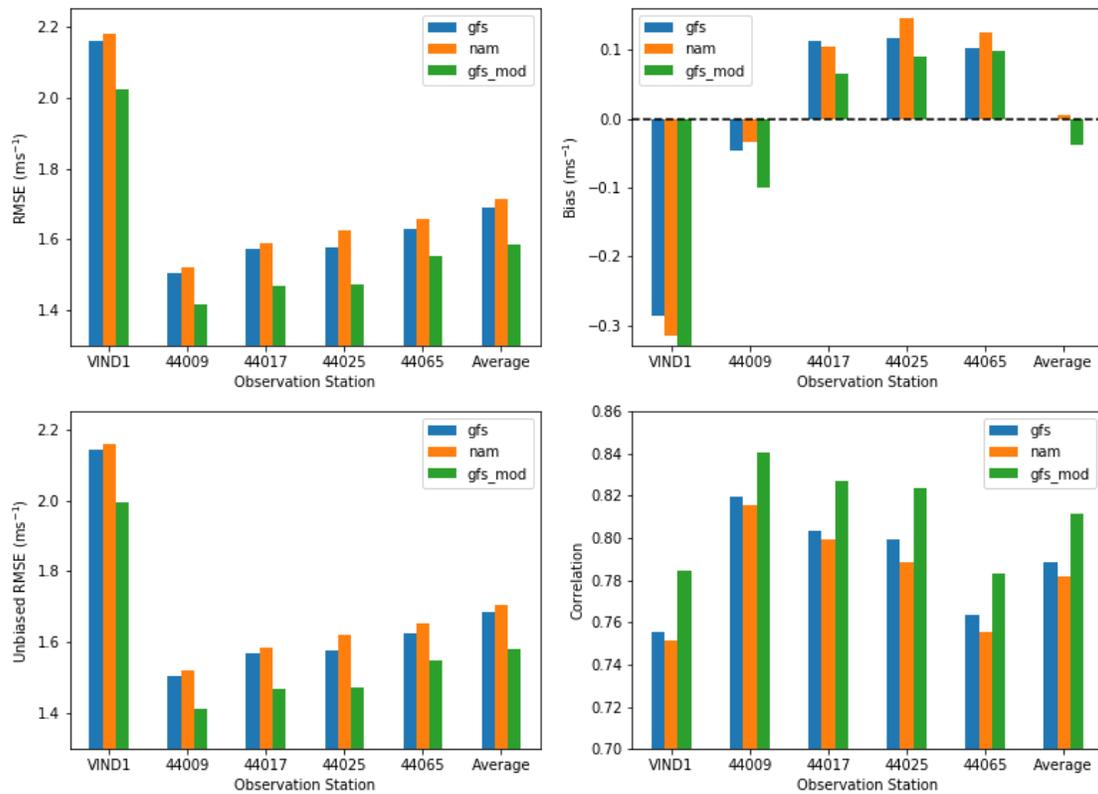


Figure 19. Performance metrics over the full validation period for each forecast model setup. Results are shown for each validation stations.

Figures 20 and 21 focus in on model performance in the 01:00 UTC to 06:00 UTC time window at the VIND1 lidar and Buoy 44025 stations, respectively. Significant improvements in model performance are found by using a 1-hour spin-up period. Reductions in RMSE are on the order of 20% to 33%, while correlation increases from between 0.70 and 0.75 to 0.84 and 0.87. A negative bias is observed at the 0100 UTC forecast time; however, it is of a magnitude similar to the positive bias seen for the six-hour spin up.

Finally, monthly trends in the performance metrics are shown in Figure 22 and 23 at the VIND1 and Buoy 44025 stations, respectively. Similar to the hindcast model performance, bias is generally higher in the summer and unbiased RMSE is lowest in the summer and fall. The GFS-based forecasts generally perform marginally better than the NAM-based forecasts across the calendar months, although NAM sometimes slightly outperforms GFS (e.g., February at Buoy 44025).

5.4 Summary of Findings and Recommendations

A summary of the key results found in this section are as follows:

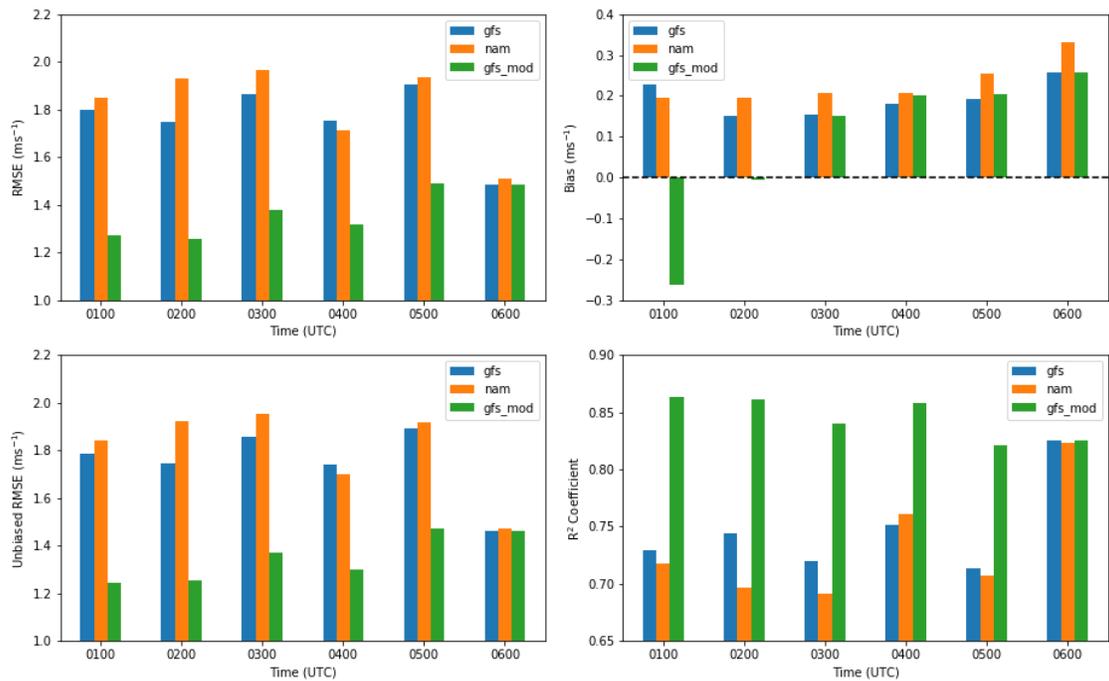


Figure 20. Forecast model performance metrics at Buoy 44025 over the 01:00 UTC to 06:00 UTC time window.

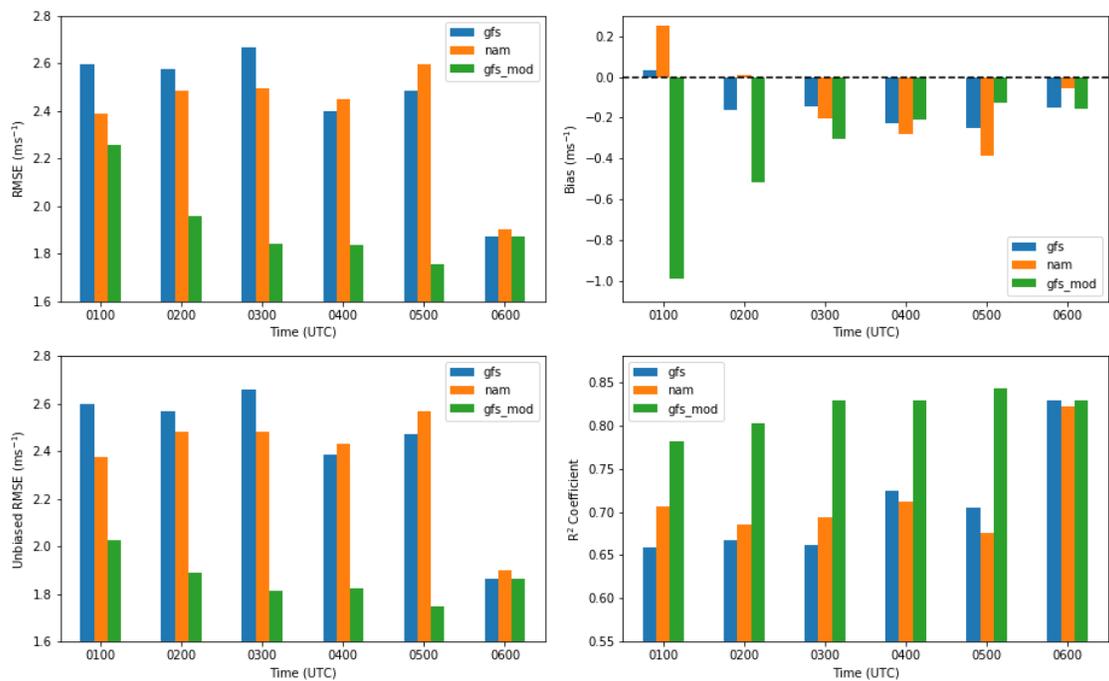


Figure 21. Forecast model performance metrics at the VIND1 Lidar over the 01:00 UTC to 06:00 UTC time window.

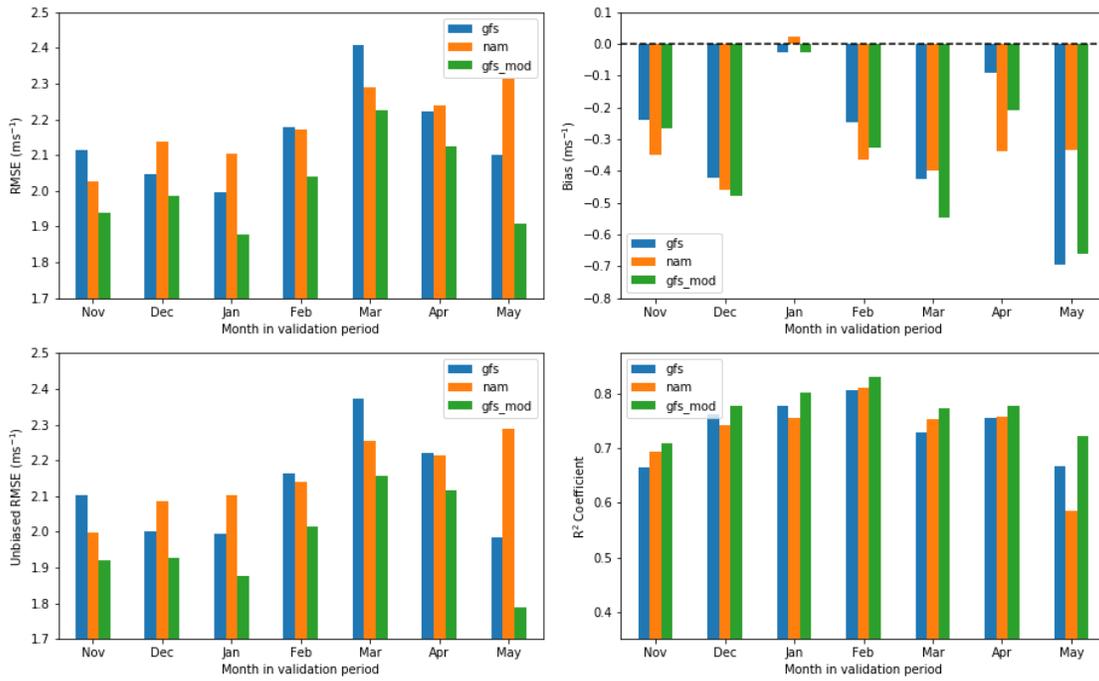


Figure 22. Performance metrics at the VIND1 Lidar for different calendar months.

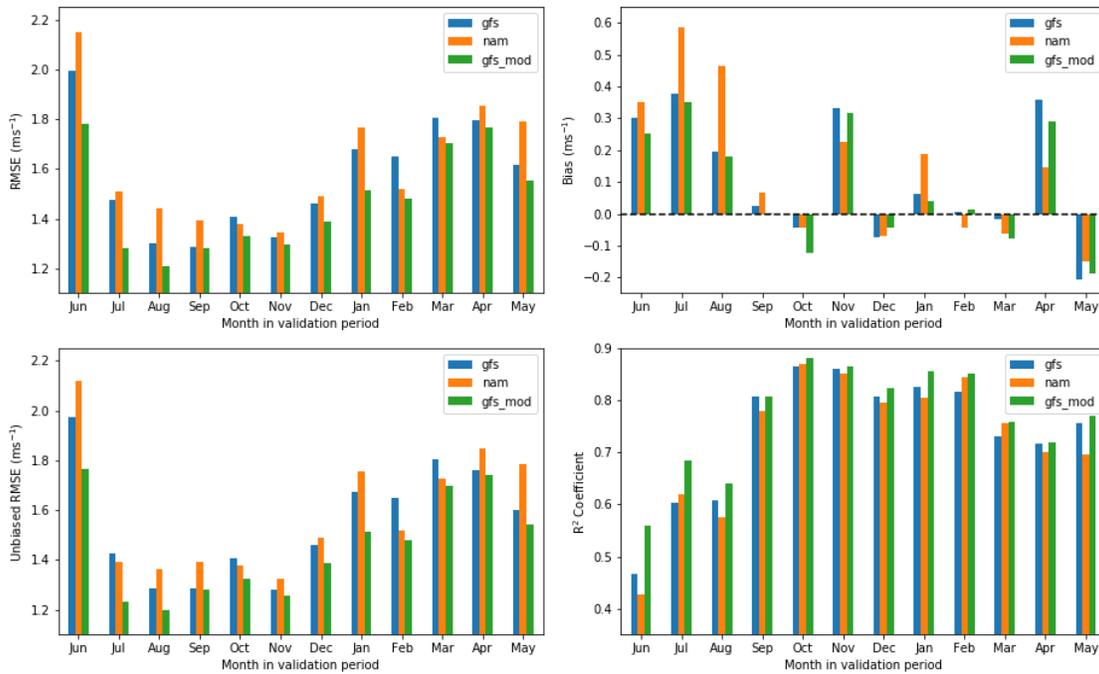


Figure 23. Performance metrics at Buoy 44025 for different calendar months.

1. The RU-WRF model run in forecast model was most accurate in the time period immediately after initialization. However, this time period is currently considered by RU-COOL as “spin-up” time for RU-WRF and data from this period are not disseminated publicly.
2. The GFS forecast product used in RU-WRF performed slightly better than the NAM product.

Based on the analysis presented in this section, we make the following recommendations for improvement to the RU-WRF model:

1. Switch to 1-hour spin-up time for RU-WRF to preserve the accurate forecast data in the 1–5 hour period after model initialization.
2. Continue using the GFS forecast product.

6 Recommendations for Improvement and Future Work

This section summarizes the key findings from this validation study. Recommendations are categorized based on those applying to RU-WRF, the coldest-pixel SST product, and the validation network and methods used by RU-COOL. Indications on ease of implementation and the magnitude of improvement are provided for each recommendation.

Regarding the RU-WRF setup, the following results were found:

1. The RU-WRF model run in forecast mode was most accurate in the time period immediately after initialization. However, this time period is currently considered by RU-COOL as “spin-up” time for RU-WRF and data from this period are not disseminated publicly.
2. The GFS forecast product used in RU-WRF performed slightly better than the NAM product.
3. The current vertical resolution of RU-WRF is too coarse to accurately model vertical wind profiles. Increasing the vertical resolution of RU-WRF near the surface resulted in a much better representation of the wind profile.
4. The RU-WRF model, currently implemented using WRF version 3.9, demonstrated moderate to large negative biases at all the observation locations used for validation. These biases were substantially reduced when RU-WRF was implemented using WRF version 4.0. By contrast, the unbiased RMSE was reduced only slightly when upgrading from WRF version 3.9 to 4.0.
5. The WIND Toolkit setup showed negligible performance changes when upgrading from WRF version 3.9 to 4.0. The WIND Toolkit showed a larger magnitude of bias compared to RU-WRF implemented with WRF version 4.0, but lower magnitude of bias compared to RU-WRF implemented with WRF version 3.9. The unbiased RMSE was larger in the WIND Toolkit compared to RU-WRF.
6. Differences in WIND Toolkit and RU-WRF model performance are attributed to the different planetary boundary layer parameterization schemes used in each. The scheme implemented in RU-WRF was found more accurate for offshore wind resource modeling. Increased performance in RU-WRF from WRF version 3.9 to 4.0 is believed to be related to improvements made in this scheme.
7. Most WRF models considered demonstrated large positive bias in the summer months and lower magnitude of bias in the remaining months. The cause for this seasonal trend would be a valuable subject for further research.

Regarding the coldest-pixel SST product, the following results were found:

1. Overall, the RU-COOL coldest pixel SST product was found to be modestly more accurate compared to a widely-used SST product from NCEP.
2. Relative performance of the two SST products depended significantly on location. Of the four buoy locations considered, the coldest-pixel product was more accurate at two locations and the NCEP product was more accurate at the other two. Causes for the relative geospatial performance of these two products would be a valuable subject for future research.
3. When used as a boundary forcing to RU-WRF, the coldest-pixel SST product provided minor improvements in wind speed modeling accuracy relative to the NCEP SST product.
4. SST measurements at the buoy locations demonstrated considerable diurnal cycles in the spring and summer months. Such cycles are not resolved in daily resolution SST products and may be significant in influencing the offshore wind resource.

Regarding the observational network and methods used by RU-COOL to validate RU-WRF, the following results were found:

1. RU-COOL has recently adopted NREL’s mesoscale model validation methods used for NREL’s WIND Toolkit. These methods are robust for validation of modern mesoscale models. However, these methods are limited to validating wind speed and wind direction only, and not other atmospheric variables known to influence wind power production, such as atmospheric stability and turbulence.
2. It is not possible to confidently validate RU-WRF using coastal observation stations because of the coarse horizontal spatial resolution of RU-WRF (3 km) and the large coastal wind speed gradients.

Based on this analysis, we have provided several recommendations for improving RU-WRF, the coldest-pixel SST product, as well as the observational network and validation procedures used by RU-COOL. These improvements are summarized in Tables 9 through 11.

Table 9. Recommendations for improving the RU-WRF Model

Code	Recommendation	Purpose	Effort	Value
A1	Reduce the spin-up time for the forecast model from 6 hours to 1 hour.	RU-WRF showed the lowest RMSE during the 1- to 5-hour period after model initialization. These forecast periods are certainly relevant for wind energy forecasting purposes and would correspond to 05:00 to 09:00 local time when electricity demand is ramping up. Switching to a 1-hour spin-up would be consistent with current practices used in state-of-the-art mesoscale forecast models.	Low	High
A2	Upgrade the WRF version used by RU-WRF from 3.9 to 4.0.	A considerable improvement in RU-WRF accuracy was found by upgrading the WRF version from the currently implemented 3.9 version to the recently released 4.0 version. Specifically, the negative bias found in WRF 3.9 in the New Jersey offshore area is largely corrected in WRF 4.0, which we speculate is because of improvements made to the Mellor, Yamada, Nakanishi and Niino planetary boundary layer scheme.	Low	High
A3	Increase the vertical resolution of RU-WRF by at least a factor of two.	The current vertical resolution in RU-WRF is insufficient to capture the wind speed profile at heights relevant to offshore wind turbines. Increasing the resolution to at least that used by NREL’s WTK will make RU-WRF more valuable for wind energy purposes.	Low	High
A4	For resource assessment applications, develop a hindcast RU-WRF product that is forced by a large-scale reanalysis product.	A reanalysis-driven WRF simulation is generally more accurate than a forecast product-driven WRF simulation, particularly for large forecast lead times (e.g., greater than 12 hours). Apart from wind power forecasting, a reanalysis-driven hindcast product would provide more value for all wind energy applications (e.g., wind resource assessment, annual energy production, grid integration).	High	High

Table 10. Recommendations for Improving the Modeling of SST

Code	Recommendation	Purpose	Effort	Value
B1	Investigate the causes for high negative bias of the coldest-pixel SST product at Buoy 44065.	Such an investigation may reveal inadequacies in the coldest-pixel SST product that can be improved.	Mid	Mid
B2	Investigate high RMSE during the spring/summer months at all buoys and the low relative bias of the NCEP RTG product during these periods.	Such an investigation may reveal inadequacies in the coldest-pixel SST product that can be improved.	Mid	Mid
B3	Consider a blended SST product that combined the best attributes of the RU-COOL coldest-pixel SST product and the NCEP SST product.	Although the Rutgers coldest-pixel product generally outperformed the NCEP SST product, there were cases where the NCEP SST product performed better (e.g, Buoys 44009 and 44065). A blended SST product that captures the positive attributes of both SST data products would likely improve wind resource modeling in RU-WRF.	High	Mid
B4	Consider using the NCEP SST product as a backup to the coldest-pixel product.	The NASA SPoRT SST product is used by RU-COOL when coldest-pixel observations are not available (e.g., nighttime). It is possible that switching to the NCEP RTG product as a backup would provide more accurate SST and wind resource modeling.	High	Low
B5	Consider development of an hourly SST product.	SST at offshore buoys shows a strong diurnal cycle in the spring and summer months. Exploring available data sources or interpolation methods to produce an hourly SST product that accounts for this diurnal cycle would likely improve wind resource characterization and in particular the sea breeze.	High	Mid

Table 11. Recommendations for Improving RU-COOL’s Observational Network and Validation Procedures

Code	Recommendation	Purpose	Effort	Value
C1	Incorporate publicly available floating Light Detection and Ranging (lidar) data	PNNL’s WindTracer lidar off the coast of New Jersey, as well as possible future deployments by DOE, provide a valuable validation dataset for modeled wind speeds, wind profiles, wind veer, and wind shear.	Low	High
C2	Consider our recommended configuration for Rutgers’ WindTracer lidar.	This configuration should allow for reliable wind profile measurements at frequent intervals out to 8 km offshore, providing a valuable data set for model validation and wind resource characterization.	High	High
C3	Leverage existing or develop new partnerships with offshore wind energy developers	Lidars are increasingly being deployed in New Jersey offshore wind energy lease areas. Accessing these data through existing or new partnerships could provide the highest-quality data available to validate RU-WRF.	High	High
C4	Do not use the RUCMM station and Oyster Creek met tower as principle validation stations	Because of strong coastal wind speed gradients and the coarse horizontal spatial resolution of the RU-WRF model, the interpolation of modeled wind speeds to these stations is too uncertain to allow for a meaningful validation of model performance.	Low	Mid
C5	Enhance validation methods to include additional variables beyond wind speed.	RU-COOL is currently following validation procedures outlined in an NREL technical report describing the WIND Toolkit (Draxl et al. 2015). We recommend continued use of these methods but also to expand the validation to atmospheric variables beyond wind speed that are known to influence the wind resource, most notably measures of atmospheric turbulence and stability.	Mid	Mid
C6	Include a RU-WRF and NOAA HRRR Performance Comparison in Validation Methods	The National Oceanographic and Atmospheric Administration (NOAA) produces an hourly-updated, WRF-based mesoscale High-Resolution Rapid Refresh (HRRR) forecast product for the continental U.S. at the same spatial resolution as RU-WRF. Output is also publicly available. We recommend adding a comparison of RU-WRF and HRRR to the existing validation methods to potentially highlight the value of customized components of RU-WRF.	High	High

7 Appendix

7.1 Recommended WindTracer Lidar Configuration

This section provides our preliminary recommendations for WindTracer lidar configuration at the RUCMM observation station to maximize the value of the lidar device for wind energy applications. The recommended approach is to incorporate two different scans for the lidar. The first is designed to accurately measure 80-m wind speed at a location 8 km offshore (Scan 1). This scan is based on a plane position indicator (PPI) setup (i.e., constant elevation angle, varying azimuthal angle). The second scan is designed to measure wind profiles at 100-m increments out to 8 km offshore (Scan 2). This scan is based on both a PPI and a range height indicator (RHI) setup (i.e., constant azimuthal angle, varying elevation angle)

Figures 24 and 25 show the configurations for Scans 1 and 2, respectively. Key parameters for Scan 1 are provided in Table 12. Key parameters for Scan 2 are provided in Tables 13 and 14 for the PPI and RHI scans, respectively.

Table 12. Lidar Configuration for Scan 1.

Parameters	Values
Elevation (°)	0.573
Azimuthal Range(°)	110
Angular resolution (°)	1.0
Time requirements for a single PPI (seconds)	36

Table 13. Lidar Configuration for Scan 2 - PPI.

Parameters	Values
Elevation (°)	0.5, 1.0, 1.5
Azimuthal Range (°)	Sector of 110
Angular resolution (°)	2.0
Time requirements for a single PPI (Sec)	18.0

Table 14. Lidar Configuration for Scan 2 - RHI.

Parameters	Values
Elevation Range (°)	0.5-1.5
Azimuthal Angle (°)	Target location
Angular resolution (°)	2.0
Time requirements for a single RHI (Sec)	5.0

7.2 Supplemental Figures

Figures 26 to 28 show the mean diurnal evolution of wind speed for Buoys 44009 (5 m measurement height), 44017 (5-m measurement height), and 44065 (4-m measurement height).

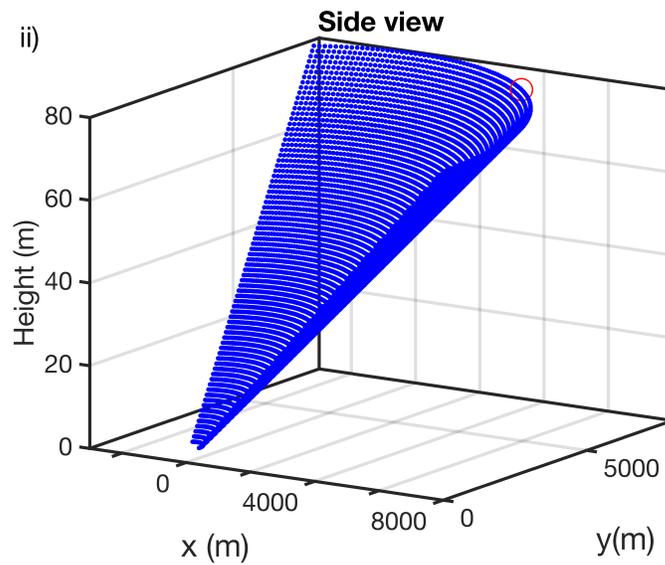
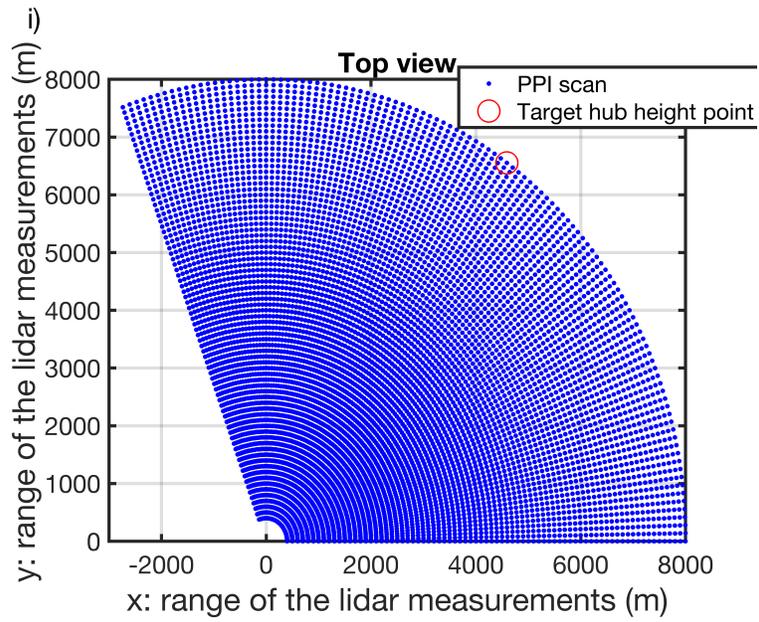


Figure 24. Lidar scanning geometry for Scan 1.

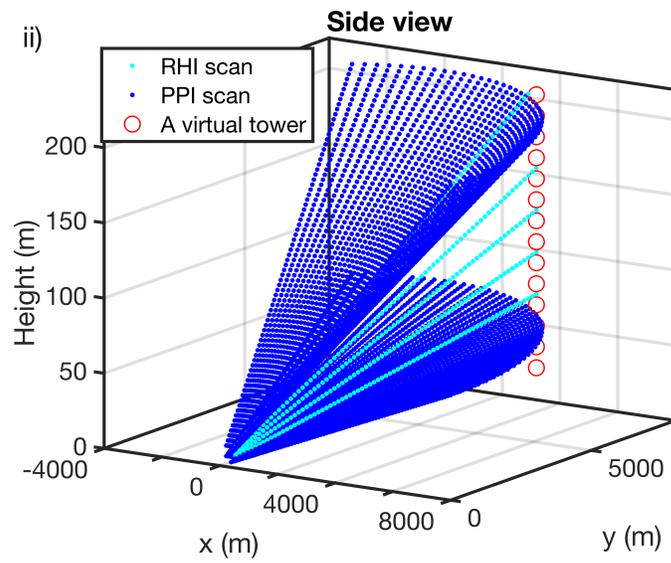
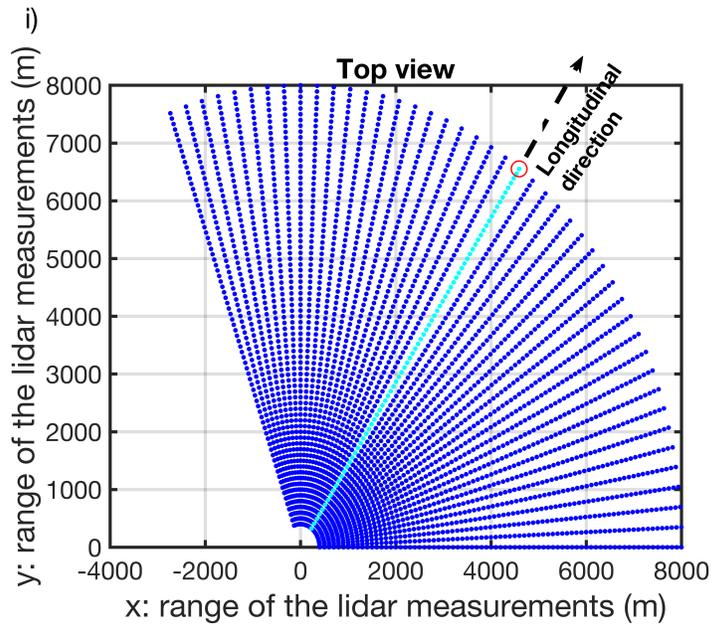


Figure 25. Lidar scanning geometry for Scan 2.

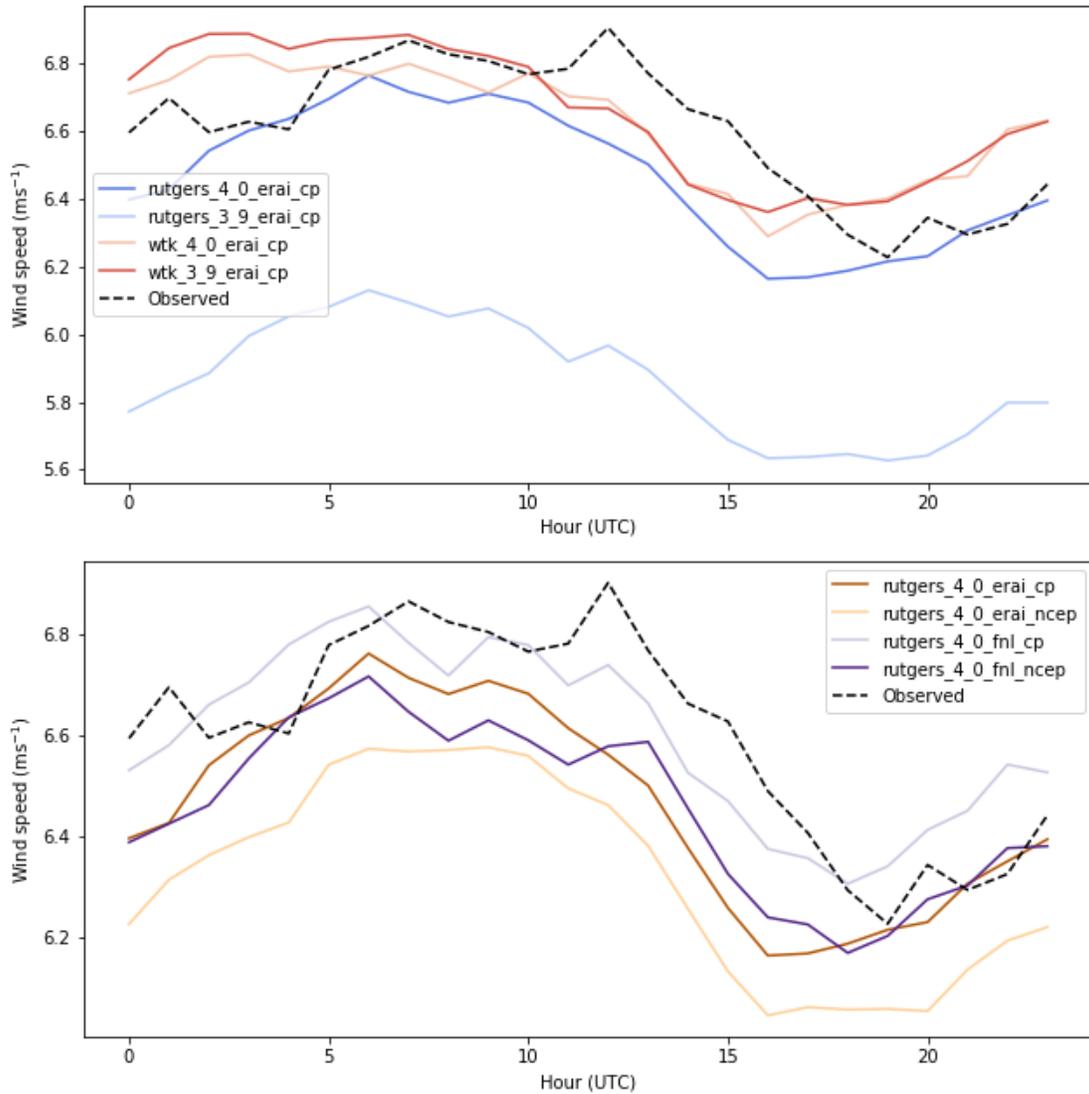


Figure 26. Mean diurnal evolution of the modeled and observed 5-m wind speed at Buoy 44009. The top figure considers model variations in namelist and WRF version. The bottom figure considers variations in atmospheric and SST forcings.

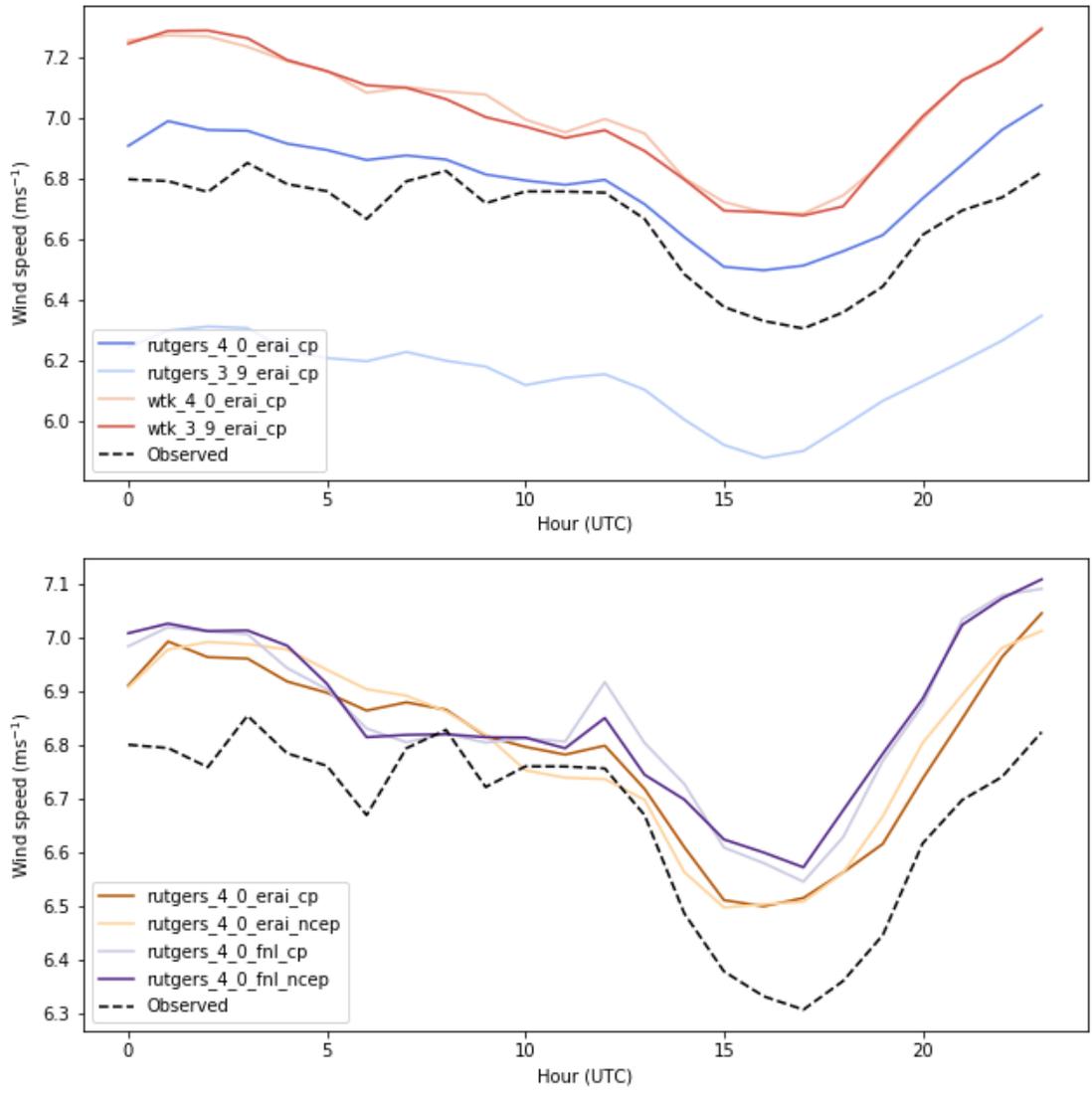


Figure 27. Same as Figure 26 but for Buoy 44017.

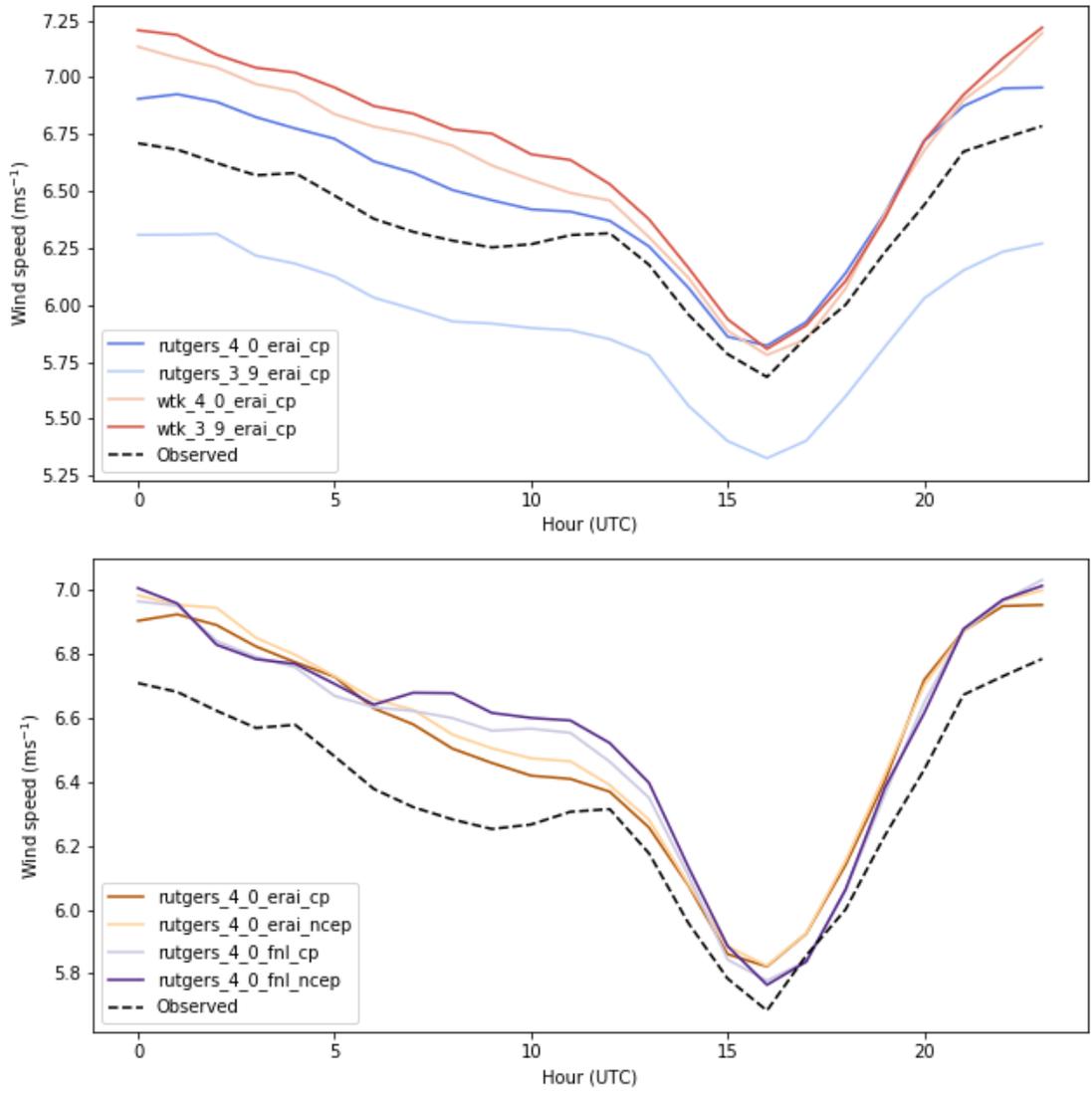


Figure 28. Same as Figure 26 but for Buoy 44065.

Bibliography

- Beljaars, A. C. M., and A. A. M. Holtslag. 1991. “Flux Parameterization over Land Surfaces for Atmospheric Models”. *Journal of Applied Meteorology* 30 (3): 327–341. doi:10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2.
- Draxl, Caroline, et al. 2015. “The Wind Integration National Dataset (WIND) Toolkit”. *Applied Energy* 151:355–366. ISSN: 0306-2619. doi:<https://doi.org/10.1016/j.apenergy.2015.03.121>. <http://www.sciencedirect.com/science/article/pii/S0306261915004237>.
- Dvorak, Michael J., et al. 2013. “US East Coast offshore wind energy resources and their relationship to peak-time electricity demand”. *Wind Energy* 16 (7): 977–997. doi:10.1002/we.1524. <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1524>.
- Glenn, S., and R. Dunk. 2013. *An Advanced Atmospheric/Ocean Assessment Program Designed to Reduce the Risks Associated with Offshore Wind Energy Development Defined by the NJ Energy Master Plan and the NJ Offshore Wind Energy Economic Development Act*. Tech. rep. New Jersey Board of Public Utilities, Rutgers Institute of Marine, and Coastal Sciences.
- Newsom, RK. 2016. *Optimizing Lidar Wind Measurements from the DOE WindSentinel Buoys*. Tech. rep. Pacific Northwest National Laboratory. https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-25512.pdf.
- Pielke, SR. RA. 2002. *Mesoscale Meteorological Modeling*. Academic Press.
- Seroka, Greg, et al. 2018. “Sea Breeze Sensitivity to Coastal Upwelling and Synoptic Flow Using Lagrangian Methods”. *Journal of Geophysical Research: Atmospheres* 123 (17): 9443–9461. doi:10.1029/2018JD028940. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD028940>.