

Best Practices for the Validation of U.S. Offshore Wind Resource Models

Mike Optis, Nicola Bodini, Mithu Debnath, and Paula Doubrawa

National Renewable Energy Laboratory

Produced under direction of the Bureau of Ocean Energy Management (BOEM) by the National Renewable Energy Laboratory (NREL) under Interagency Agreement IAG-19-2122.

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.

Strategic Partnership Project Report NREL/TP-5000-78375 December 2020

Contract No. DE-AC36-08GO28308



Best Practices for the Validation of U.S. Offshore Wind Resource Models

Mike Optis, Nicola Bodini, Mithu Debnath, and Paula Doubrawa

National Renewable Energy Laboratory

Suggested Citation

Optis, Mike, Nicola Bodini, Mithu Debnath, and Paula Doubrawa. 2020. *Best Practices for the Validation of U.S. Offshore Wind Resource Models*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5000-78375. <u>https://www.nrel.gov/docs/fy21osti/78375.pdf</u>.

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-5000-78375 December 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 U.S. Department of Interior Bureau of Ocean Energy Management (BOEM) under Agreement No. IAG-19-2122. The views expressed herein 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 <u>www.nrel.gov/publications</u>.

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

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

This work was supported and funded by the Bureau of Ocean Energy Management. We would like to thank Angel McCoy specifically for her support and guidance throughout this work. We also thank our colleagues Galen Maclaurin, Caroline Draxl, and Melinda Marquis for their peer review of the report. Technical editing and support were provided by Sheri Anstedt and Tiffany Byrne, respectively.

Acronym List

BOEM	Bureau of Ocean Energy Management
cRMSE	centered or unbiased root-mean-squared-error
DTU	Technical University of Denmark
ECMWF	European Centre for Medium-Range Weather Forecasts
EMD	Earth-mover's distance
ERA5	ECMWF Reanalysis Version 5
GW	gigawatt
MAE	mean absolute error
MW	megawatt
NDBC	National Data Buoy Center
NOAA	National Oceanic and Atmospheric Administration
NREL	National Renewable Energy Laboratory
NYSERDA	New York State Energy Research and Development Authority
REWS	rotor-equivalent wind speed
RMSE	root-mean-squared error
SST	sea-surface temperature
WIND	Wind Integration National Dataset
WRF	Weather Research and Forecasting

Executive Summary

This report presents a comprehensive set of best practices for working with both modeled and measured U.S. offshore wind resource data sets. We target two key questions in this report. First, what are the best data sources and methods for validating modeled wind resource estimates? Second, what are the best methods for vertically extrapolating near-surface wind speed measurements to heights that span the rotor-swept area of modern offshore wind turbines?

These questions are motivated by state-of-the-art advancements in offshore wind measurements. Just a few years ago, measurements were limited to those from sparsely scattered coastal weather stations and near-surface buoys. Today, an increasing number of buoy-mounted lidars—generally referred to as floating lidars or "flidars"—are being deployed in U.S. offshore wind energy lease and Call Areas. These lidars are providing wind speed and direction measurement data up to 250 m (meters) that are increasingly being made public. In addition, satellite-based measurements of near-surface U.S. offshore wind speeds are emerging as a reliable and useful product, providing wind resource measurements across the entire U.S. offshore area with decent temporal coverage.

On the modeling side, advancements in numerical weather prediction models are providing increasingly accurate estimates of the wind resource. Notably, the National Renewable Energy Laboratory (NREL) is currently producing 20-year, ensemble-based offshore wind resource data sets using the Weather Research and Forecasting (WRF) model. NREL's previous WRF-based data set—the Wind Integration National Data (WIND) Toolkit—was produced before the emergence of floating lidar devices and was minimally validated with near-surface buoy data. The opportunity now to produce novel modeled data sets, and for the first time properly validate them in U.S. offshore waters, necessitates a modern review and recommendation of best practices.

The first section of this report develops a set of best practices for the validation of modeled offshore wind resource data. Several key conclusions are made here and are summarized as follows:

- 1. The following four key performance metrics at a minimum should be calculated: the bias (capturing the difference in the modeled and observed means), the unbiased or centered root-mean-squared-error (which isolates the random error from the model), the Earth mover's distance (which quantifies the difference between two distributions), and the correlation coefficient (which measures the correspondence or "pattern" between two data sets).
- 2. Models should be validated at measurement height and not at an extrapolated height above the measurements (e.g., hub height). Generally, there is too much error in even the best extrapolation methods to meaningfully validate models against extrapolated measurements.
- 3. Near-surface measurements from buoys or satellite products are not well-suited for wind resource validation for two reasons. First, winds near the surface can behave very differently than winds across the rotor-swept heights, particularly during the diurnal cycle. Second, model performance metrics can be different near the surface than at the rotor-swept heights, and validating near the surface may incorrectly characterize model performance at the more relevant rotor-swept heights.
- 4. Land-based coastal measurements should generally not be used for offshore model validation for two main reasons. First, the coastal wind resource behaves very differently from the offshore wind resource, and validating along the coast may incorrectly characterize model performance in the farther offshore wind lease and Call Areas. Second, the land-sea boundary produces strong horizontal gradients in wind speed in this area, and validating 2-kilometer or greater resolution models becomes highly uncertain given these gradients.
- 5. Given the previously mentioned considerations and the lack of offshore meteorological towers in the United States, floating lidars remain the only robust data source to perform comprehensive validation of wind speeds across rotor-swept heights.
- 6. Given the size of expected offshore wind turbine rotor diameters (e.g., 175 m and 240 m for 8-megawatt [MW] and 15-MW designs, respectively), the rotor-equivalent wind speed should be calculated and used in place

of the hub-height wind speed when performing model validation at floating lidars. Although the differences in overall model performance are small between the two, these differences become more substantial under conditions of high shear.

- 7. Numerical weather prediction models like WRF perform differently depending on atmospheric stability conditions; therefore, validating in unstable and stable conditions separately can help isolate causes for model underperformance. Generally, the bulk Richardson number (Ri_B) is the best stability metric to separate unstable and stable conditions, as it can characterize stability across the rotor-swept heights. However, Ri_B requires temperature measurements at two heights that are not available from lidar measurements. Estimates of Ri_B from the a model such as WRF, however, can provide a reasonable substitute.
- 8. Evaluating model performance by season and time of day can provide further insight into areas of model strength and deficiency. An effective way to disseminate both the seasonal and diurnal cycles is through a "12-by-24" performance matrix or heat map.
- 9. Finally, evaluating model performance during extreme events (high-shear events, extreme cold, storms, hurricanes, and so on) provides a gauge in how well a forecast model can predict anomalous events that can negatively impact an offshore wind farm.

The second section of the report considers and validates several methods for the vertical extrapolation of wind speeds, including the following:

- 1. The conventional logarithmic wind profile
- 2. A modified logarithmic wind profile accounting for long-term atmospheric stability and developed by the Technical University of Denmark (DTU)
- 3. A single-column model developed by NREL that reduces a three-dimensional model such as WRF into a single vertical dimension that can be forced at the lower boundary by measurements
- 4. A random-forest, machine-learning model developed by NREL that is trained to predict the wind profile from near-surface measurements.

We validate the models using two floating lidars deployed in offshore New Jersey by the New York State Research and Development Authority. A comparison of mean observed and modeled wind profiles is shown in Figure A. Performance metrics for the time series-based models are shown in Figure B. Overall, the random-forest model is substantially more accurate than the other extrapolation models across all performance metrics, stability classes, seasons, and time of day. Furthermore, the machine-learning model substantially outperforms the WRF model, highlighting the benefit of local observations in producing an accurate wind profile.

The strong performance of the random-forest model has important implications for future offshore wind resource characterization. Critically, in this analysis the model was trained at one floating lidar and then evaluated at the other floating lidar 83 kilometers away with minimal degradation of model performance over that distance. Given this outcome, the machine-learning approach shows great promise for spatially complete extrapolation of near-surface measurements (i.e., from satellites) using a sparse network of floating lidars for training. Future work should leverage additional publicly available floating lidar data (Atlantic Shores Lease Area in New Jersey, Mayflower Lease Area in Massachusetts, Humboldt and Morro Bay Call Areas in California) to better understand how model performance deteriorates over distance and how well the model can generalize across a range of atmospheric conditions and geographies. Next, applying the machine-learning approach to extrapolate near-surface satellite winds could be a potential game-changer in producing accurate wind profile estimates across all U.S. offshore waters.

It is also likely that the single-column and DTU models—which showed relatively poor performance in this study could be greatly improved by using better boundary conditions. Offshore atmospheric stability is strongly driven by the air-sea temperature difference but this was not accounted for in the single-column model. Rather, the model was forced at the lower boundary only by wind speed and air temperature. Accounting for the air-sea temperature difference would likely result in considerable model improvement and will be the focus of future work. The DTU



Figure A. Mean observed and modeled wind profiles at the two New York State Energy Research and Development Authority (NYSERDA) lidars E05 (left) and E06 (right). The black dotted line shows the observed profile and the different solid colors show profiles from the different vertical extrapolation models.

method was designed to use satellite-based measurements of near-surface winds extrapolated to 10 m above sea level. In this study, we applied the method to 2-m measurements, which resulted in reduced model accuracy. By contrast, using 20-m winds from the floating lidar instead produced an accurate wind profile up to 200 m. Therefore, the DTU method should be considered again for the U.S. offshore wind industry when applied exclusively to 10-m wind speeds output from satellite models. We note, however, that the model is only capable of producing a long-term wind profile, thereby putting it at a disadvantage compared to both the machine-learning and single-column models, which are time series-based.

It is likely that an update to model validation best practices will be required a few years from now. By then, two key advancements in characterizing offshore atmospheric conditions are likely to be made. The first is the use of buoy-mounted radiometers that will measure vertical profiles of temperature and humidity profiles. These devices will allow for unprecedented characterization of offshore atmospheric stability and will provide the opportunity for more comprehensive validation of model performance. The second development is the measurement and modeling of offshore turbulent intensity, the key atmospheric variable used by the wind industry to quantify atmospheric turbulence. Turbulence intensity is now being measured by floating lidars and will soon be output from the WRF model. Validating modeled estimates of turbulent intensity will be crucial to aid the emerging offshore wind industry in accurately characterizing turbine and wind farm design conditions.



Figure B. Rotor-equivalent wind speed performance metrics for the different vertical extrapolation models. Here, cRMSE denotes the unbiased root-mean-squared error, EMD denotes the Earth mover's distance, and R^2 the squared of the correlation coefficient.

Table of Contents

1	Intr	oduction
	1.1	Measuring and Modeling the Winds Offshore 1
		1.1.1 The Need for Best Practices 2
2	Vali	dating A Modeled Wind Resource
	2.1	Data
	2.2	Kev Metrics for Validation
	2.3	Best Measurement Data Source for Validation
	-10	2.3.1 Limits of Near-Surface Measurements
		2.3.2 Limits of Onshore Coastal Measurements 7
		2.3.2 Elimits of Ohishore Coustin Measurements 10.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.
	24	Best Practices for Floating Lidar Validation 11
	2.7	2.4.1 Wind Profiles and Potor Equivalent Wind Speed
		2.4.1 While Fromes and Rotor-Equivalent while speed
		2.4.2 Role of Autospheric Stability
		2.4.5 Seasonal and Diumai variations
		2.4.4 Extreme Events
3	Nov	el Extrapolation Methods for US Offshore Wind
	3.1	Extrapolation Methods
		3.1.1 Logarithmic profile
		3.1.2 Long-Term Stability Correction (DTU)
		3.1.3 Machine Learning
		3.1.4 Single-Column Model
	3.2	Model Comparison
		3.2.1 Data and Methods
		3.2.2 Results 27
	33	Explaining DTU Model Performance 29
	34	Implications for Validating at Extrapolated Height
	5.1	
4	Sum	mary and Future Work
	4.1	Summary
		4.1.1 Best Practices for Validation
		4.1.2 Extrapolation Models
	4.2	Future Work
		4.2.1 Machine-Learning and U.S. Offshore Wind Atlases
		4.2.2 Improving Single-Column-Model Performance
		4.2.3 Measuring Offshore Atmospheric Turbulence and Stability
_		
5	Арр	θεπαιχ

List of Figures

Figure A. velopm served	Mean observed and modeled wind profiles at the two New York State Energy Research and De- nent Authority (NYSERDA) lidars E05 (left) and E06 (right). The black dotted line shows the ob- profile and the different solid colors show profiles from the different vertical extrapolation models.	vii
Figure B. Here, c and R ²	Rotor-equivalent wind speed performance metrics for the different vertical extrapolation models. RMSE denotes the unbiased root-mean-squared error, EMD denotes the Earth mover's distance, the squared of the correlation coefficient.	viii
Figure 1. ber 202	Active floating lidar deployments in the North Atlantic with publicly available data, as of Novem- 20. Wind energy lease areas are shown in white and Call Areas in grey.	4
Figure 2. two dis	A comparison of two sample wind speed distributions, taken from Hahmann et al. (2020). The stributions have the same mean but different shapes, which the EMD metric is able to quantify	5
Figure 3. wind re	Observation stations used in Optis et al. (2020) to validate the new 20-year offshore California esource data set	7
Figure 4. metrics	Mean WRF-modeled and observed wind speed profiles at NYSERDA Buoy E05. Bias and cRMSE s for the 10-minute data at 20 m and 200 m are also shown.	8
Figure 5. (right)	Mean normalized WRF-modeled and observed diurnal wind speeds at 20 m (left) and 200 m at NYSERDA Buoy E05	8
Figure 6. colored	Mean daily wind speeds modeled by WRF on October 3, 2015, shown in 3-by-3-km grid cells 1 and labeled by wind speed. Offshore and coastal wind measurement stations are shown in red	9
Figure 7. ber 202	Active floating lidar deployments offshore California with publicly available data, as of Novem- 20. Wind energy Call Areas are shown in gray.	11
Figure 8. period	Mean modeled and observed wind profiles at NYSERDA Buoy E05 when considering the full of record (black) and the high-shear cases identified in Debnath et al. (2020)	13
Figure 9. speed (WRF model performance metrics at NYSERDA Buoy E05 when validating hub-height wind (blue) and REWS (orange)	14
Figure 10. ble con	Mean modeled and observed wind profiles at NYSERDA Buoy E05, shown seperately for unsta- nations (left) and stable conditions (right)	15
Figure 11.	Performance metrics for the WRF-modeled REWS separated by stability regime	16
Figure 12.	Heat map showing the 12-by-24 frequency of unstable conditions at NYSERDA Buoy E05	17
Figure 13. betwee ditions. timesca green c negativ	Validation of WRF-modeled air-sea temperature difference. The left panel shows the relationship on the frequencies of negative air-sea temperature differences and the frequency of unstable con- . The right panel compares observed and modeled air-sea temperature differences on 10-minute ales. Actual data are in faded blue and the 2-D probability distribution of the data is shown in contours. The dotted black line denotes the 1:1 line and the dotted red lines mark the shift from the to positive values.	18
Figure 14	Mean diurnal cycles of the modeled and observed REWS at NYSERDA Buov E05	18
Figure 15.	Heat maps showing the 12-by-24 performance of WRF-modeled REWS	19
2		

Figure 16.	Schematic of quantities and calculations involved in the DTU model considered herein	22
Figure 17.	Empirical vs. theoretical distribution of atmospheric stability for the two buoy sites	23
Figure 18.	Empirical vs. theoretical distribution of atmospheric stability for the two buoy sites	23
Figure 19. denote	Mean modeled and observed wind profiles at NYSERDA Buoys E05 and E06. The dotted line s the observed profile and solid colors denote the different extrapolation models.	27
Figure 20.	REWS performance metrics for the different vertical extrapolation models	28
Figure 21. (right)	Mean modeled and observed wind profiles at NYSERDA Buoy E05 in unstable (left) and stable atmospheric conditions	29
Figure 22. E05 fo	REWS performance metrics for the different vertical extrapolation models at NYSERDA Buoy r unstable and stable conditions	30
Figure 23. E06	Mean modeled and observed diurnal cycles of normalized REWS at NYSERDA Buoys E05 and	31
Figure 24. models	Heat maps (12 by 24) of REWS bias at NYSERDA Buoy E05 for the different extrapolation	32
Figure 25. based	Mean observed and modeled wind profiles at NYSERDA Buoy E05 when using the DTU method on 2-m and 20-m measurements	33
Figure 26. models	Heat maps (12 by 24) of REWS RMSE at NYSERDA Buoy E05 for the different extrapolation	40
Figure 27. models	Heat maps (12 by 24) of REWS EMD at NYSERDA Buoy E05 for the different extrapolation	41
Figure 28.	Heat maps (12 by 24) of REWS R ² at NYSERDA Buoy E05 for the different extrapolation models	42

List of Tables

Table 1.	Key Attributes of the WRF Model Used in This Study	3
Table 2.	Key Atmospheric Variables from NYSERDA Buoy/Lidars Used in This Analysis	4
Table 3.	Recommended Performance Metrics for Modeled Wind Resource Validation	6
Table 4. Data,	Active Floating Lidar Deployments in U.S. Offshore Wind Energy Areas with Publicly Available as of November 2020	10
Table 5. et al. 2	Floating Offshore Wind Turbine Technology Assumptions for the United States by Year (Beiter 2020)	12
Table 6.	Input Features Used for the Random-Forest Model	24
Table 7.	Algorithm Hyperparameters Sampled in the Random-Forest Cross Validation	24

1 Introduction

1.1 Measuring and Modeling the Winds Offshore

This report presents a comprehensive set of best practices for working with both modeled and measured U.S. offshore wind resource data sets. We target two key questions in this report. First, what are the best data sources and methods for validating modeled wind resource estimates? Second, what are the best methods for vertically extrapolating near-surface wind speed measurements to heights that span the rotor-swept area of modern offshore wind turbines?

There are several motivations for this report. The first is the rapid development of the U.S. offshore wind industry. Currently, there is a total of 28.5 gigawatts (GW) of planned capacity in U.S. offshore waters, including 30 megawatts (MW) operating capacity (Block Island Wind Farm), 6.4 GW of assigned lease areas, 17,440 GW planned with developer site control, and an additional 4,600 GW in early planning phases (Musial et al. 2020). These areas span multiple bodies of water including the Atlantic and Pacific coasts, Hawaii, and the Great Lakes.

Accurate characterization of the offshore wind resource is vital for the success of this industry. Generally, characterization involves a combination of wind measurements providing accurate but single-point localized resource characterization and modeled wind estimates that characterize an entire region but require validation against observations to understand model strengths and limitations. Numerical weather prediction (NWP) models have historically provided the best modeled estimates of the U.S. offshore wind resource. Specifically, the Wind Integration National Dataset (WIND) Toolkit, produced by the National Renewable Energy Laboratory (NREL), has provided the best publicly available NWP-modeled data set since 2013 (Draxl 2015). Covering the entire continental United States and out to 200 nautical miles offshore, the WIND Toolkit provides 2-kilometer (km) and 5-minute time resolution data up to 200 meters (m) over the 7-year period from 2007 to 2013, and was produced using the state-of-the-art Weather Research and Forecasting (WRF) numerical weather prediction model.

The WIND Toolkit was minimally validated offshore, however, because of the lack of high-quality observations available at the time. Historically, offshore wind observations in the United States have been largely limited to near-surface, buoy-based measurements from the National Data Buoy Center, managed by the National Oceanic and Atmospheric Administration (National Data Buoy Center 1971). Further, wind profile measurements covering the expected rotor-swept heights of modern wind turbines—soon expected to extend up to nearly 300 m above sea level—have been critically lacking. Because of this, offshore validation of the WIND Toolkit was limited to a handful of buoys with measurements below 5 m above sea level. (Draxl 2015).

Moving forward to the present day, there have been substantial advancements in U.S. offshore wind measurement capabilities. Most notable is the emergence of lidar-based technologies, which allow measurements of winds up to about 250 m (Carbon Trust 2018). These lidars may be attached to low-height offshore meteorological towers (Woods Hole Oceanographic Institute 2020) but more commonly are mounted on buoys and referred to as floating lidars or "flidars." Validated extensively against meteorological tower measurements, floating lidar devices are transforming how the offshore wind resource is measured. It is expected there are over 20 floating lidars currently deployed across U.S. offshore wind energy lease and Call Areas. Although most data are owned by wind energy developers and kept proprietary, an increasing number are being made freely available to the public (DNV-GL 2020; Atlantic Shores Offshore Wind 2020; Mayflower Offshore Wind 2020; Pacific Northwest National Laboratory 2020). These floating lidars offer an unprecedented opportunity to robustly validate NWP estimates of the U.S. offshore wind resource.

Another emerging offshore technology is the use of satellites to measure near-surface wind speeds. Here, the backscatter of pulsed laser signals sent from the satellite reflect off the ocean surface. The more calm the water (low wind), the more the signal is received back by the satellite. The more rippled the water (higher winds), the more scattered the returning signal and the less is received back by the satellite. Based on geophysical transfer functions, those signals are then used to produce estimates of the 10-m above-sea-level wind speeds. This technology is increasingly being used to characterize the wind resource for wind energy purposes (e.g., Badger et al. 2015) and was

recently used by NREL and colleagues to characterize the 10-m wind resource in the U.S. Atlantic offshore region (Ahsbahs et al. 2020).

On the modeling front, a series of NWP-related advancements since 2013 have motivated an update to U.S. modeled offshore wind resource data sets. As described in Optis et al. (2020), the WRF model has been extensively updated since 2013, with targeted improvements for wind energy applications in particular. There have also been extensive advancements in atmospheric and oceanic products used as boundary conditions to the WRF model, including the new ERA5 reanalysis product produced by the European Centre for Medium-Range Weather Forecasts (Hersbach et al. 2020) and sea-surface temperature (SST) products (Kaiser-Weiss and Minnett 2012). Because of these advancements, NREL is currently producing updated 20-year WRF-based data sets for all U.S. offshore waters and is leveraging the advancement in offshore wind measurements to perform more comprehensive validation than was possible with the WIND Toolkit back in 2013.

1.1.1 The Need for Best Practices

These advancements in offshore wind measurements and modeling, as ground-breaking as they are, come with an important caveat: how do we validate NWP-modeled wind estimates in this new technology environment? More specifically, the following questions present themselves:

- What is the relative value of using different measurement data sources for offshore wind validation, including floating lidar, satellite-based winds, buoy-measured winds, and even land-based coastal weather stations? What are the strengths and limitations of each, and which should be used and why?
- What are the most important performance metrics to use when validating a modeled wind estimate?
- Is it sufficient to just validate the hub-height wind speeds, given the increasing diameter of offshore wind turbines?
- How should data be filtered to perform more robust validation? For example, is mean wind sufficient, or should other considerations such as time of day, season, or specific atmospheric conditions be considered?
- Should near-surface measurements be extrapolated vertically (e.g., to hub height) and then validated at this more relevant height? If so, what is the best extrapolation method for this application and could any novel approaches be leveraged?

Given the importance of these questions, it is clear that a set of best practices is required to address recent measurement and modeling advances. The aim of this report is to provide these best practices. We begin in Section 2 by exploring and recommending general best practices for validation. Here, we use recent WRF simulations in the U.S. Atlantic offshore area to justify and demonstrate these best practices. In Section 3, we consider a range of conventional and novel methods to vertically extrapolate near-surface winds to heights more relevant to wind energy, such as hub height. Based on model performance, we determine whether validation is best performed at the measurement location or whether validation at an extrapolated height is acceptable.

2 Validating A Modeled Wind Resource

This section provides a set of recommended practices for the validation of modeled offshore wind speeds. We first discuss the best metrics for validation and summarize four key metrics. Next, we present the limits of using nearsurface, land-based coastal wind measurements for validation and conclude that buoy-based floating lidar measurements provide the only robust and comprehensive data source for model validation. We then present arguments as to why the rotor-equivalent wind speed (REWS) is a better variable than the hub-height wind for validation. Finally, we discuss the importance of validating on diurnal and monthly timescales, validating separately based on atmospheric stability conditions, and validating specific extreme events.

2.1 Data

Throughout this section, we present best practices by validating recent WRF simulations performed by NREL in the U.S. Atlantic offshore region. These simulations cover the period September 2019 through August 2020. The key WRF model attributes are summarized in Table 1.

Feature	Specification
WRF version	4.2.1
Nesting	6 km, 2 km
Output time resolution	5 minutes
Vertical levels	61
Near-surface-level heights (m)	12, 34, 52, 69, 86, 107, 134, 165, 200
Atmospheric nudging	Spectral nudging on 6-km domain, applied every
	6 hours
Planetary boundary layer scheme	Mellor-Yamada-Nakanishi-Niino
Microphysics	Ferrier
Longwave radiation	Rapid radiative transfer model
Shortwave radiation	Rapid radiative transfer model
Topographic database	Global multiresolution terrain elevation data
	from the U.S. Geological Survey and National
	Geospatial-Intelligence Agency
Land-use data	Moderate Resolution Imaging Spectroradiometer
	30s
Cumulus parameterization	Kain-Fritsch

Table 1. Key Attributes of the WRF Model Used in This Study

For this validation exercise, we use the data from the New York State Energy Research and Development Authority (NYSERDA) E05 and E06 floating lidars as the basis for our measurements. In addition to lidar-based wind measurements, the buoys also measure key near-surface atmospheric variables. A summary of the data fields available are provided in Table 2. The location of the buoys is shown in Figure 1.

2.2 Key Metrics for Validation

When properly quantifying the error in a modeled wind resource estimate it is important to understand how and why a model might be underperforming. For time series analysis, model error can generally be decomposed into **bias** and **random error**. The bias quantifies how different the mean modeled result is from the observed result, whereas the random error quantifies the variations of the model around the mean.

Metrics such as the root-mean-squared-error (RMSE) and mean absolute error (MAE) include both the bias and random components of error. To attribute model error to bias and random error separately, RMSE or MAE can be

Table 2. Key Atmospheric Variables from NYSERDA Buoy/Lidars Used in This Analysis

Variable	Vertical Coverage
Wind speed	2 m and then 20 m to 200 m in
	20-m increments
Wind direction	2 m and then 20 m to 200 m in
	20-m increments
Air temperature	2 m
Air pressure	2 m
SST	Ocean surface



Figure 1. Active floating lidar deployments in the North Atlantic with publicly available data, as of November 2020. Wind energy lease areas are shown in white and Call Areas in grey.

decomposed into a bias component and what is called the "centered" or "unbiased" component of RMSE or MAE. Taylor 2001 demonstrated this separation for RMSE and bias by showing the relation between their squares:

$$RMSE^2 = Bias^2 + cRMSE^2 \tag{2.1}$$

where cRMSE is the unbiased RMSE. The bias component is calculated as:

$$Bias = \overline{p} - \overline{o} \tag{2.2}$$

where p are the modeled estimates and o are the observations and the bar denotes the mean. The RMSE component is calculated as:

$$RMSE = \left[\frac{1}{N}\sum_{n=1}^{N} (p_n - o_n)^2\right]^{1/2}$$
(2.3)

where N is the number of data points. It can be shown then that the cRMSE is calculated as

$$cRMSE = \left[\frac{1}{N}\sum_{n=1}^{N}\left[(p_n - \overline{p}) - (o_n - \overline{o})\right]^2\right]^{1/2}$$
(2.4)

Another key performance metric is the correlation coefficient, R, which measures the "correspondence" or "pattern" between two variables. It is defined as:

$$R = \frac{\frac{1}{N} \sum_{n=1}^{N} (p_n - \overline{p}) - (o_n - \overline{o})}{\sigma_n \sigma_o}$$
(2.5)

where σ_p and σ_o are the standard deviations of the predictions and observations, respectively. The correlation coefficient, R, reaches a maximum value of 1, indicative of perfect correlation, and a minimum value of -1, indicative of perfect negative correlation. Usually the square of the correlation coefficient is used, R^2 , which has maximum and minimum values of 1 and 0, respectively, which becomes more useful for characterizing relationships when there are multiple independent variables.

A fourth metric that is very useful is the Earth-mover's distance (EMD), recently popularized for wind energy in the making of the New European Wind Atlas (Hahmann et al. 2020). The EMD (also called the Wasserstein distance) is a measure of the difference between two distributions. Specifically, the metric is equal to the area between two cumulative distribution functions and can be interpreted as the amount of "dirt" needed to move from one probability distribution (or pile) to another to make them equal. The key advantage of the EMD metric is that it accounts for cases where two distributions may have the same bias but have different shapes, as shown in Figure 2.



Figure 2. A comparison of two sample wind speed distributions, taken from Hahmann et al. (2020). The two distributions have the same mean but different shapes, which the EMD metric is able to quantify.

Based on the arguments in this section, we recommend that the four metrics listed in Table 3 should, at a minimum, be used to validate modeled estimates of the wind resource. We note that MAE is not included in this list. It always

has less magnitude than RMSE given that it penalizes outliers less. However, both metrics trend similarly and including both generally does not reveal new information. We also note that bias and cRMSE can be converted to percentage units but we do not made explicit recommendations here for that conversion. There are many ways to normalize those errors to get percentages—including the mean, range, and standard deviation of the observations with no clear guidance on which normalization metric is most appropriate. Therefore, no recommendations on percentage-based performance metrics are provided in this report.

Name	Abbreviation	Description
Bias	Bias	Difference between the mean modeled and observed result
Unbiased RMSE	cRMSE	The random error component after bias is re- moved, describing the differences in model variations around the mean
Square of correlation coefficient	R ²	The correspondence or pattern between the modeled and observed variable
Earth-mover's distance	EMD	Difference between the probability distributions between the modeled and observed variable

Table 3. Recommended Performance Metrics for Modeled Wind Resource Validation

2.3 Best Measurement Data Source for Validation

In this section, we discuss the benefits and drawbacks of different wind speed measurement data sources for validation, including onshore coastal measurements, near-surface buoy measurements, satellite-based products, and floating lidar.

2.3.1 Limits of Near-Surface Measurements

Near-surface buoy-based measurements of wind speed are common in U.S. waters, mostly maintained and accessed through the National Data Buoy Center (National Data Buoy Center 1971). For example, the Optis et al. (2020) report on the updated California wind resource data set considered model performance at a range of buoy locations (shown in Figure 3). These sites are appealing, given the data is public, often provide multiple years of 10-minute or hourly-averaged data, and are spread across a domain and not isolated to specific regions. Finally, the data are quality-controlled by the National Oceanic and Atmospheric Administration and are therefore reliable.

Satellite-based scatterometers retrieve wind speeds at 10 m above sea level and have been used to create a satellitebased offshore wind atlas for the U.S. Atlantic offshore area (Ahsbahs et al. 2020). This study demonstrated that satellite-based winds can have very low bias overall relative to buoy-based anemometers, but fairly high RMSE ranging from $1.3-1.5 \text{ ms}^{-1}$, depending on the satellite data source. Therefore, in principle, the Ahsbahs et al. (2020) wind atlas could be used to validate 10-m modeled wind bias across the U.S. Atlantic but not the other recommended performance metrics in Table 3.

The key question, though, is whether validation at 10 m or lower is meaningful and representative for wind energy applications. More specifically, are model performance metrics at 10 m or lower representative of model performance at hub height or at other rotor-swept heights? We explore this question in Figure 4, wherein we compare mean wind profiles from the WRF simulations in the North Atlantic against the NYSERDA Buoy E05 floating lidar measurements. As shown in the figure, the WRF-modeled profile is accurate on average below about 30 m, above which the WRF-modeled profile increasingly underestimates wind speeds as height increases. Because of this trend, bias at 20 m is slightly positive (0.2 ms^{-1}) and bias at 200 m is moderately negative (-0.4 ms^{-1}) . Based on validation



Figure 3. Observation stations used in Optis et al. (2020) to validate the new 20-year offshore California wind resource data set

only at lower heights (e.g., using only buoy measurements), one would conclude that WRF is positively biased at this location, despite actually being negatively biased at rotor-swept heights. Furthermore, we see that the cRMSE is slightly higher at 200 m (2.0 ms^{-1}) than 20 m (1.8 ms^{-1}) . So again, validating only at lower heights would provide an inaccurate characterization of model performance at rotor-swept heights.

We continue the 20 m and 200 m comparison in Figure 5, in which the mean modeled and observed diurnal wind speed cycles are shown. Looking first at the observed profiles, we see that at 20 m the diurnal cycle does not have a clear trend. By contrast, the 200-m cycle is much more pronounced with a minimum around 15:00 UTC and a maximum around 00:00 UTC. This comparison alone reveals the limitations of surface-only analyses, which in this case would miss the diurnal variations in the actual wind resource. In terms of validation, we see that the modeled diurnal cycle shows some amplitude at 20 m, which does not match the observations (RMSE = 0.035 ms^{-1}). By contrast, the modeled cycle is in reasonable agreement with the observed cycle (RMSE = 0.028 ms^{-1} , or a 20% improvement). Therefore, validating at 20 m would underestimate model performance at rotor-swept heights and would mischaracterize the ability of WRF to capture the diurnal cycle at rotor-swept heights.

Based on the results in this section, we conclude that validating at lower heights is inappropriate wind energy applications. Not only can key performance metrics change substantially with height, but so can the characteristics of the wind resource.

2.3.2 Limits of Onshore Coastal Measurements

Another common data source for offshore wind resource validation is land-based weather stations located along the coastline. These stations generally have measurement heights above those of buoy-based measurements, which could provide more meaningful validation at heights more relevant to wind power.



Figure 4. Mean WRF-modeled and observed wind speed profiles at NYSERDA Buoy E05. Bias and cRMSE metrics for the 10-minute data at 20 m and 200 m are also shown.



Figure 5. Mean normalized WRF-modeled and observed diurnal wind speeds at 20 m (left) and 200 m (right) at NYSERDA Buoy E05

However, the land-sea interface at the coastline provides several challenges to conduct meaningful validation of a modeled wind resource. First, coastal stations are generally located far from current offshore wind energy lease and Call Areas and can have very different wind resource characteristics than those located farther offshore (e.g., more

pronounced sea breezes).

More importantly, large wind speed gradients at the coastline generally prohibit a meaningful validation of modeled wind speeds when these models are run at a coarse resolution (e.g., 2-km WRF resolution in the Optis et al. [2020] data set). 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. This phenomenon was documented in Optis et al. (2020) and shown again in Figure 6. The figure focuses on three measurement stations along the New Jersey coastline near Atlantic City: the Rutgers University Oyster Creek meterological tower (RUOYC) and Coastal Metocean Monitoring station (RUCMM) and a Vindicator floating lidar (VIND1) deployed by the Pacific Northwest National Laboratory (PNNL) from 2015 to 2017. The colored and labeled boxes show the mean wind speeds in each 3 km grid cell, as modeled by WRF, during a high wind event on October 3, 2015. We see that the coastal RUCMM and RUOYC stations are located in areas where wind speeds in adjacent WRF cells can differ by 2 ms⁻¹. Comparatively, the VIND1 lidar located farther offshore consistently has wind speeds of 20 ms⁻¹ in neighboring cells.



Figure 6. Mean daily wind speeds modeled by WRF on October 3, 2015, shown in 3-by-3-km grid cells colored and labeled by wind speed. Offshore and coastal wind measurement stations are shown in red.

For many analyses (e.g., sea breeze events, comparing various spatial resolutions for an NWP model), these coastal stations would be valuable reference sites. However, when validating a 2-km spatial resolution NWP model such as WRF, the uncertainty associated with interpolating modeled grid-box average wind speeds to a coastal observation station is prohibitively high. Therefore, we advise against the use of coastal stations for validation, especially when comparing the performance of different models.

2.3.3 Floating Lidar Measurements

Given the limitations of near-surface and coastal measurements, as well as the absence of tall meteorological towers in U.S. offshore waters, the remaining data sources for robust model validation are buoy-based floating lidars. As of November 2020, there are six active floating lidar deployments in U.S. offshore wind energy areas whose data are publicly available. These lidars are listed in Table 4 and shown in Figures 1 and 7 for the Atlantic and California offshore areas, respectively.

Location	Organization	Time Resolu- tion	Start Date for Public Data	Maximum Measurement Height	Data Access
Hudson South Call Area, New Jersey	NYSERDA	10 minute	2019-09-04	200 m	DNV-GL (2020)
Hudson North Call Area, New Jersey	NYSERDA	10 minute	2019-08-12	200 m	DNV-GL (2020)
Atlantic Shores, New Jersey	Shell/EDF	10 minute	2020-02-26	250 m	Atlantic Shores Offshore Wind (2020)
Mayflower, MS	Shell/EDPR	Daily	2020-04-13	250 m	Mayflower Offshore Wind 2020
Humboldt, Califor- nia	PNNL/Bureau of Ocean Energy Management (BOEM)	1 second	2020-10-XX	250 m	Pacific Northwest National Laboratory (2020)
Morro Bay, Califor- nia	PNNL/BOEM	1 second	2020-10-XX	250 m	Pacific Northwest National Laboratory (2020)

Table 4. Active Floating Lidar Deployments in U.S. Offshore Wind Energy Areas with Publicly Available Data, as of November 2020

Most of these public floating lidar data are at a 10-minute resolution or better, apart from the Mayflower data that is provided daily. The recently deployed floating lidars in California by PNNL and BOEM are providing data at a 1-second and 10-minute resolution. Most data sources provide measurements up to 250 m, except for the NYSERDA lidars, which only measure up to 200 m.

These data sources provide the best means for robust validation of the offshore wind resource, with the ability to assess wind shear and wind veer (i.e., change in wind direction with height) across the entire rotor-swept area of expected 8- to 12-MW offshore turbines and the majority of the 15-MW wind turbines (Beiter et al. (2020); Table 5). Furthermore, a growing body of validation exercises are showing strong agreement between lidar-measured and anemometer-measured wind speeds, giving confidence to the exclusive use of floating lidar for offshore wind profile measurements (Carbon Trust 2018).

The key limitation to these floating lidars is the sparsity of measurements. A total of six data sources for all U.S. offshore waters is relatively small compared to land-based validations stations (e.g., Draxl 2015). There are many more floating lidars deployed privately by wind developers in current Atlantic lease areas; however, these data are



Figure 7. Active floating lidar deployments offshore California with publicly available data, as of November 2020. Wind energy Call Areas are shown in gray.

considered highly proprietary and not shared publicly.

2.4 Best Practices for Floating Lidar Validation

Having established floating lidars as the best data source for validation of modeled wind resource data sets, we now turn to best practices for how to use these data for model validation. We focus first on the advantages of REWS over hub-height wind speeds, and then discuss the role of atmospheric stability and how validation should account for this. We then argue for both seasonal- and diurnal-based validation and finally discuss the importance of validating specific extreme events.

2.4.1 Wind Profiles and Rotor-Equivalent Wind Speed

Offshore wind turbines will be considerably larger than their land-based counterparts. Recently, Beiter et al. (2020) compared four future build-out scenarios for offshore California with floating wind turbine capacities ranging from 8 MW to 15 MW (Table 5). Rotor diameters are expected to reach nearly 200 m by 2022 and 240 m by 2032, corresponding to swept height ranges from 30 m to 206 m and 30 m to 270 m, respectively.

2019	2022	2027	2032
8	10	12	15
175	196	215	240
118	128	138	150
30 m–206 m	30 m–226 m	30 m–246 m	30 m–270 m
	2019 8 175 118 30 m-206 m	2019 2022 8 10 175 196 118 128 30 m-206 m 30 m-226 m	2019202220278101217519621511812813830 m-206 m30 m-226 m30 m-246 m

Table 5. Floating Offshore Wind Turbine Technology Assumptions for the United States by Year (Beiter et al. 2020)

Large wind turbine performance is influenced by the high vertical wind shear across the rotor (Barthelmie, Shepherd, and Pryor 2020) and research has shown that weighting the wind speed over the rotor swept area provides a more accurate estimate of kinetic energy passing through the rotor than the simple consideration of hub-height wind speeds (Wagner et al. 2014; Redfern et al. 2019). This weighted wind speed is referred to as the rotor-equivalent wind speed and is defined as follows:

$$REWS = \left(\sum_{h=1}^{N_h} v_h^3 \frac{A_i}{A}\right)^{1/3}$$
(2.6)

where n_h is the number of available measurement or model heights within the rotor layer, v_i is the wind speed at height h, A is the total area swept by the rotor, and A_h is the area of the hth segment, calculated according to:

$$A_{h} = \int_{z_{h}}^{z_{h+1}} c(z)dz$$
 (2.7)

where z_h is the height of the hth segment separation line and:

$$c(z) = 2\sqrt{R^2 - (z - H)^2}$$
(2.8)

where R is the rotor diameter and H the hub height (Wagner et al. 2014).

Given that REWS is a better predictor of turbine energy than hub-height wind speed, should modeled REWS be validated instead of hub-height wind speed? Are there meaningful performance metric differences between the two to justify that change? To explore these questions, we contrast the use of validating hub-height wind speeds and REWS using the assumption of a 10-MW wind turbine shown in Table 5. We contrast two cases where validation is performed on the full data set and when validation is performed only on a subset of high-wind-shear cases observed at the NYSERDA E05 and E06 floating lidars, as documented in Debnath et al. (2020). Comparative mean profiles for each case are shown in Figure 8. As shown in the figure, the high-shear events are associated with high wind speeds with a mean difference ranging from 40 m to 200 m of about 6 ms⁻¹. By contrast, the mean profile for the entire period of record shows considerably less shear and mean winds.

Figure 9 contrasts the four key validation metrics for hub-height wind speed and REWS for both the full data and high-shear subsets. The results generally show that validating using hub-height wind speed vs. REWS does not have considerable impact when considering the full period of record at the lidars, with only slight improvements in model performance when validating against REWS. By contrast, for the high-shear cases, there are more meaningful model performance improvements using REWS than hub-height wind speeds, especially for the bias and EMD metrics. We



Figure 8. Mean modeled and observed wind profiles at NYSERDA Buoy E05 when considering the full period of record (black) and the high-shear cases identified in Debnath et al. (2020)

note that model performance metrics all increase when validating REWS rather than hub height. We do not expect this to be a general principle but is more likely specific to the WRF model setup and wind resource characteristics of the region.

Based on the analysis in this section, we encourage the use of REWS over hub-height wind speeds when validating offshore model performance. Such an approach will give a more accurate characterization of model performance as it pertains to wind energy applications.

2.4.2 Role of Atmospheric Stability

A decade of research has established the role that atmospheric stability plays in observed and modeled wind profiles (Optis, Monahan, and Bosveld 2014, 2016). In general, atmospheric stability can be separated into unstable and stable conditions. Unstable conditions usually occur when warm air sits underneath colder air (e.g., during strong land surface heating in summer). The lower density of the warmer air causes it to rise above the cold air, inducing strong vertical mixing of momentum and leading to wind profiles that are relatively constant with height. By contrast, stable conditions occur when colder air sits underneath warmer air (e.g., during land surface cooling at night). The higher density of the underlying colder air suppresses turbulent vertical mixing and produces wind profiles with strong wind shear.

Because of ocean mixing and the high heat capacity of water, the sea surface is generally not subject to the same rapid changes in temperature as land and magnitudes of stability are generally less. However, on longer seasonal timescales, the SST can change considerably. Furthermore, air coming from land can be advected offshore and lead to considerable differences in land-sea temperature. Most interestingly, stable stratification can be induced when warmer air flows over a colder sea and lead to phenomena such as high shear and low-level jets (Dörenkämper et al. 2015; Debnath et al. 2020).

Planetary boundary layer schemes in NWP models like WRF can account for unstable and stable conditions very



Figure 9. WRF model performance metrics at NYSERDA Buoy E05 when validating hub-height wind speed (blue) and REWS (orange)

differently. Therefore, when analyzing model performance offshore, it is important to consider both unstable and stable conditions. In many cases, and provided there is sufficient data, these broad stability regimes can be further broken down into extreme, moderate, and weakly unstable and stable conditions (Optis, Monahan, and Bosveld 2014; Optis et al. 2020a). Such breakdown provides more insight into model performance across a range of stability regimes.

There are several metrics used to quantify atmospheric stability and selecting the right one is important. A common metric is the Obukhov length, which quantifies the ratio of turbulent momentum and temperature fluxes at a given height, usually the surface. The flux Richardson number is similar but also considers the wind and temperature gradients. A more simplified and representative metric is the bulk Richardson number, Ri_B , which quantifies the atmospheric stability in a vertical layer based on wind and temperature differences between the top and bottom of that layer:

$$Ri_B = \frac{g}{\theta_v} \frac{\Delta z \Delta \theta_v}{\left(\Delta U\right)^2 + \left(\Delta V\right)^2}$$
(2.9)

where g is gravitational acceleration, θ_v is absolute virtual potential temperature, $\Delta \theta_v$ is the virtual potential temperature difference across a layer of thickness, Δ_z , and ΔU and ΔV are the changes in horizontal wind components across that same layer. Values of $Ri_B < 0$ represent thermodynamically unstable conditions and $Ri_B > 0$ represent stable conditions, with instability or stability increasing monotonically with the magnitude of Ri_B . For neutral conditions,

$Ri_B \approx 0.$

The Ri_B metric has several advantages over other stability metrics. First, it is able to quantify stability over a specific height range (e.g., rotor-swept heights of a wind turbine), whereas other metrics are restricted to a single height. Therefore, it can provide a more accurate characterization of stability in the context of wind turbine operation. Second, it requires only measurements of wind and temperature, which are often measured at meteorological towers and are typical output from atmospheric models such as WRF. The flux-based metrics, by contrast, require more robust sonic anemometer measurements that are not as widely available.

The main limitation of Ri_B is that it does not provide a direct measure of turbulence. Rather, it characterizes the "likely" turbulence given the wind and temperature gradient in a vertical layer. By contrast, flux-based measurements do provide direct measures of turbulence and, if measured at hub height, may be more useful metrics for characterizing atmospheric stability experienced by a wind turbine.

In the U.S. wind offshore environment, however, measurements generally do not exist to calculate any stability metric. The lack of temperature measurements at different heights preclude the use of Ri_B while the lack of flux measurements at buoys preclude the use of other metrics. To validate separately by stability regime requires that the stability metric be calculated from the model being validated. In this case, we use the WRF model and select Ri_B as the metric to distinguish the stability regime.

Figure 10 shows the mean modeled and observed wind profiles under unstable and stable conditions. The observed profile shows low shear in unstable conditions and high shear in stable conditions, as we would expect. We further see differing model performance in each stability regime: in unstable conditions, WRF tends to underestimate wind speeds and wind shear, whereas in stable conditions WRF overestimates wind speeds but also underestimates wind shear. These differences highlight the benefit of validating by stability regime and can lead to more detailed investigations on how and why the WRF model may underperform in certain conditions.



Figure 10. Mean modeled and observed wind profiles at NYSERDA Buoy E05, shown seperately for unstable conditions (left) and stable conditions (right)

Figure 10 also reveals that unstable conditions are predominant at Buoy E05, occurring 70.5% of the time. We note that this frequency is specific to the Atlantic region does not represent other offshore areas. For example, Optis et

al. (2020) found that very stable conditions were dominant in offshore California wind energy Call Areas, mainly because of strong coastal upwelling of deeper colder water and the resulting induction of stable stratification.

We further break down model performance by stability regime in Figure 11 using the four key performance metrics and validating on REWS. Here, we see that cRMSE is considerably lower in unstable conditions, whereas the bias and EMD are lower in stable conditions. The change in the R^2 coefficient is generally small, with slight improvement in unstable conditions. The figure further illustrates the value in validating by stability regime. In this case, for example, an effort to improve WRF model bias in the U.S. Atlantic offshore region should focus on unstable rather than stable conditions.



Figure 11. Performance metrics for the WRF-modeled REWS separated by stability regime

Atmospheric stability is largely driven by vertical temperature differences and therefore can vary significantly by season and time of day. Examining these trends can be important in understanding at what times of the day or year a model may underperform. A useful tool for such investigation is a 12-by-24 matrix that plots the mean of some variable over each month and hour of the year (i.e., 12 months \times 24 hours). This matrix can be nicely illustrated by a heat map, which we use in Figure 12 to characterize the frequency of unstable conditions at Buoy E05 throughout the year. The figure shows a strong seasonal trend in atmospheric stability, with 80% to 100% of the data representing unstable conditions from August through December, but only 50% to 85% from January through July. Therefore, we would expect WRF model bias (which was highest in unstable conditions) to be highest during the latter half of the year. We also observe diurnal trends in stability although no clear consistent pattern emerges across the different months.

So far this section has illustrated how modeled wind speeds can perform differently under different stability conditions. The final step for robust model validation is to quantify how well the model characterizes atmospheric stability. For example, if a model is underestimating wind shear in stable conditions, it may be because it is underestimating the magnitude of stability. Unfortunately, the lack of observations in U.S. offshore wind areas that are needed to characterize stability make such a validation challenging. The best available approach is to validate near-surface stability by leveraging wind and air temperature measurements from buoys in addition to SST. These variables can yield calculations of the air-sea temperature difference (itself a useful stability metric) or the Ri_B between the buoy measurement height and the surface. These near-surface metrics, although not ideal, can provide some indication of atmospheric stability in the air aloft. For example, Debnath et al. (2020) found that the air-sea temperature difference



Figure 12. Heat map showing the 12-by-24 frequency of unstable conditions at NYSERDA Buoy E05

at Buoys E05 and E06 was strongly associated with the occurrence of high shear and low-level-jet events.

In Figure 13, we validate WRF in its ability to model the air-sea temperature difference. In the left panel, we first show that the frequency of unstable conditions modeled at Buoy E05 (as measured by Ri_B between 200 m and the surface) is moderately correlated to the frequency of negative air-sea temperature differences modeled by WRF ($R^2 = 0.448$). So it is clear that the air-sea temperature difference plays a role in determining atmospheric stability between 200 m and the surface, but also highlights the limit of near-surface measurements in fully characterizing atmospheric conditions relevant to wind power. The right-most figure compares the modeled and observed air-sea temperature difference. In general, the correlation is very strong ($R^2 = 0.961$). We do see that in the region in which most of the data are contained (darker green), the WRF model tends to slightly overestimate the air-sea temperature difference (total bias of 0.23° C). Overall, we can conclude that the WRF model is representing the air-sea temperature difference well.

2.4.3 Seasonal and Diurnal Variations

As demonstrated in the previous section, U.S. offshore atmospheric stability varies with season and time of day. Given the different relative performance of the WRF model in unstable and stable conditions, it is important to break down that performance by season and time of day as well. Such analysis provides useful information on when a model would be most accurate and when model performance decreases.

A diurnal cycle plot of modeled and observed REWS, as shown in Figure 14, can provide an overall indication of how well a model is capturing the diurnal cycle. Such a comparison is especially important in the context of offshore wind energy meeting diurnally varying electricity loads and understanding how well a model captures that diurnal cycle in wind speeds. In Figure 14, the observed REWS shows a moderate diurnal cycle, peaking around 10.5 ms⁻¹ and reaching a minimum of around 9.7 ms⁻¹. By contrast, the WRF model predicts a larger amplitude diurnal cycle that reaches a minimum of almost 9.0 ms⁻¹. Indeed, much of the negative bias in WRF-modeled REWS (Figure 9) occurs in the 10:00 UTC to 19:00 UTC, which would correspond to 05:00 to 14:00 local time at Buoy E05.



Figure 13. Validation of WRF-modeled air-sea temperature difference. The left panel shows the relationship between the frequencies of negative air-sea temperature differences and the frequency of unstable conditions. The right panel compares observed and modeled air-sea temperature differences on 10-minute timescales. Actual data are in faded blue and the 2-D probability distribution of the data is shown in green contours. The dotted black line denotes the 1:1 line and the dotted red lines mark the shift from negative to positive values.



Figure 14. Mean diurnal cycles of the modeled and observed REWS at NYSERDA Buoy E05

To assess diurnal performance by season, we show the 12-by-24 heat maps in Figure 15 for all four performance metrics. The figure reveals key times when the model performs well and when it does not. For example, the large negative bias between 10:00 to 19:00 UTC in Figure 14 is especially strong in December and January. Furthermore, the small positive bias we saw from 00:00 to 03:00 UTC is largely caused by large positive bias in the summer months (especially July), whereas the remaining months trend toward negative bias during this time. Other anomalies are revealed as well, such as the tendency to have very high cRMSE during the later hours in December and the corresponding drop in correlation.



Figure 15. Heat maps showing the 12-by-24 performance of WRF-modeled REWS

2.4.4 Extreme Events

Finally, we discuss the value in evaluating model performance during targeted extreme events. An exhaustive list of extreme events is not considered here but could include storm or hurricane events, extremely cold temperatures, very high turbulence, and so on. Here, we focus only on the extreme high-shear events presented earlier in Section 2.4.1 when discussing the value of REWS. As Debnath et al. (2020) demonstrated, Buoys E05 and E06 experienced nearly 100 independent high-shear events over a 1-year period of record. These events can considerably impact turbine operation through increased loading, fatigue, and so on; therefore, the ability to predict and characterize such events through a model such as WRF is important.

WRF performance during these extreme shear events was presented earlier in this section (Figure 8 showing the mean profile and Figure 9 showing the performance metrics). The fact that WRF underestimates wind shear during these events is an important result and could motivate future work to better understand how and why WRF underestimates the magnitude of these events.

3 Novel Extrapolation Methods for US Offshore Wind

In this section, we evaluate a range of methods for the vertical extrapolation of near-surface wind speed measurements to model wind profiles across the rotor area of expected offshore wind turbines. This study is motivated by two questions: can vertical extrapolation error be made sufficiently low to validate modeled REWS or hub-height winds using extrapolated near-surface wind speed measurements? If so, can this extrapolation technique be applied to satellite-based near-surface wind speed measurements to produce an observation-based wind atlas for all of U.S. offshore waters?

To address the first question, we evaluate a range of extrapolation methods and validate them using the floating lidar data at NYSERDA buoys E05 and E06. We consider both the conventional logarithmic wind profile and a recent modified formulation to the logarithmic profile with a long-term stability correction developed by researchers at the Technical University of Denmark (DTU) (Badger et al. 2015). We also propose and evaluate more novel methods including a single-column model and a machine-learning approach. These four models are validated against offshore floating lidar observations using the recommended set of best practices described in the previous section.

We note that the conventional power law profile is not considered in this analysis. The power law profile requires measurements at two heights to calculate the power law coefficient, α . This study is focused on extrapolating near-surface winds to rotor-swept heights using only near-surface measurements. Therefore, a second height is not available to calculate α . In fact, this two-height requirement makes the power law profile largely impractical for offshore wind energy purposes, given the prevalence of single-height measurements from buoys and satellites.

All time-series-based analysis in this section uses the 10-minute-averaged wind speeds.

3.1 Extrapolation Methods

3.1.1 Logarithmic profile

The logarithmic wind profile is defined as:

$$U(z) = \frac{u_*}{\kappa} \left[\ln\left(\frac{z}{z_0}\right) - \psi_m\left(\frac{z}{L}, \frac{z_0}{L}\right) \right]$$
(3.1)

where U is the wind speed, κ is the von Kármán constant (normally taken to be 0.4), z is the height above the surface, u_* is the friction velocity, z_0 is the roughness length, ψ_m is the stability function for momentum that adjusts the wind profile depending on atmospheric stability, and L is the Monin-Obukhov length that characterizes atmospheric stability. The friction velocity, u_* , requires high-frequency sonic anemometer measurements that are not available at the NYSERDA buoys. Instead, we reformulate Eq. 3.1 to use the 2-m wind speeds as a reference measurement, allowing the wind profile to be calculated according to:

$$U(z) = U_{2m} \left[\frac{\ln(z/z_0) - \psi_m(z/L, z_0/L)}{\ln(z_{ref}/z_0) - \psi_m(z_{2m}/L, z_0/L)} \right].$$
(3.2)

Here, we implement the ψ_m formulations from Jiménez et al. 2012, which have become standard correction functions and are currently used in the WRF mesoscale model surface layer parameterization.

The calculation of *L* typically requires measurements of the momentum and turbulent temperature fluxes, which are not available from buoy measurements but require more sophisticated sonic anemometers and thermometers measuring at high frequency. Instead, we can calculate a "bulk" *L* based on the bulk Richardson number, Ri_B :

$$Ri_B = \frac{g}{\theta_{avg}} \frac{z_{2m}(\theta_{z_2} - \theta_{surf})}{U_{z_{2m}}^2}$$
(3.3)

where z_{2m} is the height 2 m above the surface, g is the acceleration as a result of gravity, $\theta_{z_{2m}}$ is the potential temperature at 2 m, θ_{surf} is the potential temperature at the surface, and U_{2m} is the 2-m wind speed. Combining the Equations 3.2 and 3.3 yields a relationship between L and Ri_B :

$$Ri_{B} = \frac{z_{2m}}{L} \frac{ln\left(\frac{z_{2m}}{z_{0}}\right) - \psi_{h}\left(\frac{z_{2m}}{L}, \frac{z_{0}}{L}\right)}{\left[ln\left(\frac{z_{2m}}{z_{0}}\right) - \psi_{m}\left(\frac{z_{2m}}{L}, \frac{z_{0}}{L}\right)\right]^{2}}$$
(3.4)

where ψ_h is the stability function for temperature, also taken from Jiménez et al. 2012.

Using Eq. 3.4, we iteratively solve for L given Ri_B , which combined with Eq. 3.2 allows the calculation of the vertical wind profile.

3.1.2 Long-Term Stability Correction (DTU)

The second model we considered is a modification of the logarithmic profile that intends to provide a more accurate long-term stability correction. This model has been developed over the last decade by DTU (Kelly and Gryning 2010; Badger et al. 2015) and is herein referred to as the DTU method. The motivation for this model is the fact that the logarithmic profile is frequently violated under conditions of stable stratification (Optis, Monahan, and Bosveld 2014, 2016) and that the logarithmic profile would be inaccurate when applied on a time-series basis (e.g., to 10-minute-averaged atmospheric conditions). Instead, the DTU method aims to calculate a single long-term wind profile that accounts for the distribution of L values throughout the year to perform a single, long-term stability correction. As such, the DTU method is only suitable for long-term wind resource assessment, as it requires at least 1 year of data and ideally many years (Kelly and Gryning 2010).

The stability correction applied to the log extrapolation is height-dependent and computed based on empirical constants and atmospheric conditions at the site: the percentage of stable vs. unstable conditions; the quadratic mean of the kinematic heat flux; the mean, near-surface air temperature; and the time-averaged friction velocity. These input parameters are taken from the WRF simulations (described in Section 2.1) and are combined with correction functions, ψ_m , based on similarity theory to compute a vertical profile of the correction function (Figure 16). This correction is then added to the log extrapolation to yield a wind speed profile, as in Eq. 3.1, where u_* is usually obtained from the local measurements and z_0 is computed using the Charnock relationship, $z_0 = \alpha u_*^2/g$, with g being the acceleration caused by gravity and $\alpha = 0.0144$.



Figure 16. Schematic of quantities and calculations involved in the DTU model considered herein

The first step in the DTU method is to verify that the probability distribution functions for atmospheric stability are a good fit to the empirical distributions. This comparison is given in Figure 17. The theoretical functions shown in this figure take into account the percentage of stable vs. unstable conditions at the NYSERDA buoy sites (n_{stable} and $n_{unstable}$), scales of variation for L^{-1} (σ_{stable} and $\sigma_{unstable}$), and empirical constants ($C_{stable} = 5$ and $C_{unstable} = 12$). Note that previous work focusing on other data sets used different values for the $C\pm$ constants (e.g., both were set to 3.0 to extrapolate satellite-derived wind speed measurements in Badger et al. 2015).



Figure 17. Empirical vs. theoretical distribution of atmospheric stability for the two buoy sites

The vertical profile of the stability correction function, ψ , for both buoys is shown in Figure 18. The correction is unstable (i.e., positive) below 80 m and stable (i.e., negative) above it.



Figure 18. Empirical vs. theoretical distribution of atmospheric stability for the two buoy sites

3.1.3 Machine Learning

The third model considered is a machine-learning-based method. Here, we consider a relative simple ensemblebased regression tree method, known as a random-forest model, which has shown strong predictive power in previous wind speed extrapolation work (Bodini and Optis 2020b, 2020a), and in relating wind farm energy production to on-site atmospheric variables (Optis and Perr-Sauer 2019). Here, we use the RandomForestRegressor module

in Python's Scikit-learn (Pedregosa et al. 2011). We consider a range of input features available from the NYSERDA buoys: 2-m wind speed, wind direction, pressure, and air temperature; the SST and air-sea temperature difference, as well as the time of day and month of year. Wind direction, time of day, and month of year are all decomposed into their sine and cosine components to preserve circularity (i.e., 0° and 360° directions are equivalent, as are 00:00 and 24:00).¹ A summary of these variables is listed in Table 6.

Input Feature	Acronym	Measurement Height (m AGL)
2-m wind speed	WS 2 m	2
Sine of 2-m wind direction Cosine of 2-m wind direction	WD	2
2-m air temperature	Т	2
Sea-surface temperature	SST	0
Air-sea temperature difference	T - SST	-
2-m pressure	р	2
Sine of time of the day Cosine of time of the day	time	-
Sine of month Cosine of month	month	-

Table 6. Input Features Used for the Random-Forest Model

To make sure the observation sets over which the random forest is trained and tested cover as much of the seasonal variability as possible, we build the testing set using a consecutive 20% of the observations from each month in the period of record. We evaluate different combinations of the hyperparameters with a five-fold cross validation, and randomly sample 20 sets. The hyperparameters considered in the cross validation and their sampled ranges are shown in Table 7. We evaluate the performance of the learning algorithm based on the root-mean-squared error between measured and predicted wind speed at extrapolation height: the set of hyperparameters that leads to the lowest root-mean-squared error is selected and used to assess the final performance of the learning algorithm.

As described in detail in Bodini and Optis (2020), it is both impractical and unfair to evaluate a machine-learning model at the same site where it is trained. Critically, the model requires observations of the lidar-measured wind speeds up to 200 m in order to be trained. Evaluating model performance at the training site is impractical because the wind profiles are already known and unfair as the other extrapolation methods do not have such knowledge of lidar-measured wind profiles. Instead, model performance must be assessed through a "round-robin" approach, in which the model is evaluated at a site not used to train the model. Specifically in this study, the random forest model is trained on data at NYSERDA buoy E05 and then evaluated against other extrapolation models at NYSERDA buoy E06 located 83 km away, and then vice-versa. This round-robin approach ensures a fair comparison of the different extrapolation methods and that no model has prior knowledge of lidar-measured wind profiles at the site where it is evaluated.

 Table 7. Algorithm Hyperparameters Sampled in the Random-Forest Cross Validation

Hyperparameter	Possible Values
Number of estimators	10-800
Maximum depth	4-40
Maximum number of features	1-11
Minimum number of samples to split	2-11
Minimum number of samples for a leaf	1–15

¹Both are needed because each value of sine only (or cosine only) is linked to two different values of the cyclical feature.

3.1.4 Single-Column Model

The fourth model considered is a single-column model. Essentially, it is a stripped-down version of a three-dimensional (3D) model, such as WRF, in which only vertical exchanges are considered and horizontal homogeneity is assumed. This greatly simplifies the governing equations of a 3D model and reduces the SCM to a one-dimensional model in the vertical only. By further assuming no moisture processed or cloud radiation, the equations of motion simplify further and depend only on the horizontal pressure gradients, the Coriolis force, and the vertical turbulent flux of momentum and temperature:

$$\frac{\partial u}{\partial t} = f(v - v_G) - \frac{\partial (\overline{u'w'})}{\partial z}$$

$$\frac{\partial v}{\partial t} = f(u - u_G) - \frac{\partial (\overline{v'w'})}{\partial z}$$

$$\frac{\partial \theta}{\partial t} = \frac{\partial (\overline{\theta'w'})}{\partial z}$$
(3.5)

where u, v, and w are the three vector wind components, t is time, z is the height above the surface, θ is potential temperature, and u_G and v_G are the u- and v- components of the geostrophic wind. The $\overline{u'w'}$, $\overline{v'w'}$ terms represent the u- and v- components of the vertical turbulent momentum flux and $\overline{\theta'w'}$ represents the vertical turbulent temperature flux.

The momentum and temperature fluxes are not solved directly but rather parameterized based on well-established eddy-diffusivity relationships:

$$\overline{u'w'} = -K_{\rm m} \frac{\partial u}{\partial z}$$

$$\overline{v'w'} = -K_{\rm m} \frac{\partial v}{\partial z}$$

$$\overline{\theta'w'} = -K_{\rm h} \frac{\partial \theta}{\partial z}$$
(3.6)

where K_m and K_h are the eddy diffusivities for momentum and temperature, respectively. These terms are themselves parameterized with a range of possible options in the literature (Optis and Monahan 2016, 2017). We adopt a relatively simple first-order closure model that includes eddy diffusivities that are related to the wind speed gradient and a stability function that depends on the Richardson number:

$$K_{m} = l_{m}^{2} \frac{\partial U}{\partial z} f_{m}(R_{i})$$

$$K_{h} = l_{m} l_{h} \frac{\partial U}{\partial z} f_{h}(R_{i})$$
(3.7)

where l_m and l_h are the mixing lengths for momentum and heat, respectively, and f_m and f_h are the stability functions for momentum and temperature, respectively. There are a range of proposed formulations for the mixing lengths and stability functions. Here, we use the one developed by Smith (1990), which showed strong results when used in a single-column model in previou studies (Optis and Monahan 2016, 2017). A detailed explanation and the equations of the stability functions and mixing lengths can be found in (Smith 1990; Cuxart et al. 2006; Optis and Monahan 2017).

The single-column-model equations are solved on a logarithmically stretched grid from a 2-m to 2,000-m height with 200 grid levels to provide higher resolution near the surface. The lower boundary conditions at 2 m are the measured wind speed components and temperature from the NYSERDA buoys. The upper boundary conditions are the 800 hectopascal pressure level data provided by the ERA5 reanalysis. A zero temperature gradient boundary condition is also applied at the top of the domain.

Recognizing that the geostrophic wind can change with height in conditions of horizontal temperature gradients, we calculate a geostrophic wind profile at each time step to force the simulations. This is done by first assuming that the 800 hPa winds from ERA5 are geostrophic, which is a reasonable assumption at 2000 m, where surface friction effects should be negligible. Next, we calculate the geostrophic wind at the surface using surface pressure and air temperature data from the ERA5 reanalysis product:

$$u_{G} = -\frac{1}{f\rho} \frac{\partial P}{\partial y}$$

$$v_{G} = \frac{1}{f\rho} \frac{\partial P}{\partial x}$$
(3.8)

where ρ is air density and *P* is pressure. The horizontal pressure gradient terms are calculated by taking a planar best fit of the closest 9 ERA5 grid points that surround the buoy locations. Equation 3.8 is used to calculate the geostrophic wind components at 2 m, and finally the geostrophic wind profile is found by linearly interpolating the 2-m and 800-hPa values to the different single-column-model heights.

To initialize the simulation, we start by solving for an initial neutral profile and impose an equilibrium condition (i.e., $\partial u/\partial t = 0$; $\partial v/\partial t = 0$; $\partial (\overline{\theta'w'})/\partial z = 0$) and solving for the u and v profiles. Then the simulation moves forward as a time-marching algorithm using the complete set of equations provided in this section. A continuous simulation is launched for the whole year time period without interruption.

3.2 Model Comparison

3.2.1 Data and Methods

The four vertical extrapolation models presented in the previous section are all validated against lidar data from NYSERDA Buoys E05 and E06 over the full period of record. For each lidar, we consider only the time periods where wind speeds are reported at every height from 20 m to 200 m. Based on recommendations from Section 2, we validate using the REWS calculated from each extrapolation model based on the 10-MW offshore reference wind turbine (Beiter et al. 2020).

For comparative purposes, we also include WRF performance alongside the vertical extrapolation model performance. This inclusion allows us to compare the performance between comprehensive, 3D models that lack the knowledge of local observations (i.e., WRF) and highly idealized models that do benefit from knowledge of those observations (i.e., the extrapolation models).

We note that the DTU method is only capable of modeling the mean wind profile. Therefore, performance metric analysis throughout this section excludes the DTU method.

3.2.2 Results

We begin with a comparison of the mean wind profile in Figure 19, showing results at both NYSERDA buoys E05 and E06. The observed wind profile shows moderate shear, increasing from about 8.5 ms^{-1} to 10.5 ms^{-1} at E05, and 8.0 ms^{-1} to 10.3 ms^{-1} at E06. We see that the random-forest machine-learning model provides excellent agreement with the mean profile, whereas the other models are deficient in some respects. The single-column model underestimates wind speeds at E05 but is very close to the observed profile at E06. The logarithmic profile captures the upper winds relatively well with a slight positive bias, but has increasingly higher bias at lower heights. The DTU method significantly overestimates wind speeds, especially at the upper heights, with nearly a 1.5-ms^{-1} bias at 200 m. Finally, we see that the WRF model tends to underestimate the wind profile (as shown in the previous section).



Figure 19. Mean modeled and observed wind profiles at NYSERDA Buoys E05 and E06. The dotted line denotes the observed profile and solid colors denote the different extrapolation models.

REWS-based performance metrics for the different models are shown in Figure 20. Again the strong performance of the machine-learning model is apparent, with considerably lower error metrics and higher correlation to observations relative to the other models. The bias is notably negligible at Buoy E05 and slightly negative at E06 By contrast, the single-column model has the weakest performance across all metrics at E05 and all but the bias at E06. The logarithmic profile performance falls in between the machine-learning and the single-column models and is the only model with a positive bias at both buoys. Finally, the WRF model tends to perform similarly to the logarithmic model with slightly lower unbiased RMSE and higher correlation, but higher magnitude of bias and EMD.

We next consider the role of atmospheric stability in relative model performance, first by comparing the mean wind profiles by stability regime in Figure 21. Stability regimes are based on the WRF-modeled bulk Richardson number between 200 m and the surface. Here, we focus only on Buoy E05 and note that relative performance is similar at both buoys. The machine-learning model shows similar performance in unstable and stable conditions, accurately capturing the unstable profile and slightly underestimating the stable profile. The single-column model performs reasonably well in unstable conditions but is unable to capture the high shear in the stable regime and significantly underestimates wind speeds. The log profile similarly underestimates wind speeds in stable conditions but overestimates in unstable conditions. Finally, the WRF model underestimates the wind profile in unstable conditions while accurately capturing winds above 100 m in stable conditions, but overestimating them below 100 m. Overall, we see that all models apart from the random forest struggle with consistent accuracy across stability regimes.



Figure 20. REWS performance metrics for the different vertical extrapolation models

This relative consistency is further illustrated in Figure 22, where we show the REWS performance metrics by stability regime. Again, we focus on Buoy E05 and note the similar relative performance between models at Buoy E06. We also see the random forest with the strongest performance metrics, apart from slightly higher magnitude bias and higher EMD in stable conditions relative to the WRF model. The single-column model shows lower magnitude bias and EMD in unstable relative to stable conditions but high unbiased RMSE and correlation across both regimes. The log profile performs better in unstable conditions than unstable conditions for all performance metrics, whereas the WRF model cRMSE and R^2 are lower in unstable conditions, but bias and EMD are higher relative to stable conditions.

We next compare normalized diurnal profiles for each model in Figure 23. The observed profiles show moderate diurnal trends with peak wind speeds from 00:00 to 05:00 UTC at 5% above the mean, and minimum wind speeds around 16:00 UTC that are 5% to 7% below the mean. We also note the larger diurnal amplitude at Buoy E06 relative to Buoy E05. The random-forest model shows excellent agreement on average throughout the diurnal cycle, with RMSE in the normalized wind speeds of only 0.25% and 0.8% at E05 and E06, respectively. The single-column model performs reasonably well at capturing the diurnal cycle (0.82% and 1.0%) while both the WRF model and logarithmic profile have the largest error (RMSE greater than 2% at both buoys). Specifically, the WRF model tends to overestimate the amplitude of the diurnal cycle, whereas the logarithmic profile tends to underestimate it.

Finally, we present 12-by-24 heat maps to show the combined diurnal and monthly trends of model performance. We show only the bias heat maps in Figure 24 while the remaining performance metric heat maps are provided in the Appendix. We see that the machine-learning model has consistently low magnitude bias throughout the diurnal and monthly cycles with no clear diurnal trends but a tendency to overestimate wind speeds in the fall. The single-



Figure 21. Mean modeled and observed wind profiles at NYSERDA Buoy E05 in unstable (left) and stable (right) atmospheric conditions

column model shows considerable negative bias throughout the year with a tendency to overestimate wind speeds in November. Interestingly, the bias in December is positive from 01:00 to 12:00 and negative form 13:00 to 00:00. The WRF model shows some trends with positive bias in spring in the early hours with negative bias in the midhours. Finally, the logarithmic profile shows substantial trends with strong overestimation of winds through most of the year and underestimates in spring with the largest magnitude of those underestimates in the early hours.

3.3 Explaining DTU Model Performance

Figure 19 showed that the DTU method significantly overestimated wind speeds. This is a surprising result given its strong performance in Badger et al. {**badger_extrapolating_2015**} in which 10-m satellite-measured winds were extrapolated. To explore this, we compare DTU model performance using both 2-m and 20-m measurements as the basis for extrapolation. The results are shown in Figure 25. We see that the extrapolation from the 2-m measurements does not match the measured wind speed profile. This is likely because the measurement height is too low and located within the viscous sublayer where log-law approximations are not valid. When the same method is used to extrapolate from the 20-m lidar measurements, we see a good match between extrapolated and measured values. This analysis reveals that this DTU method is not suitable for extrapolation based on buoy wind speed measurements, which are often made with propeller or cup anemometers between 2 m and 5 m above the sea surface. Instead, this method should be applied to short offshore meteorological masts and satellite-derived wind speed estimates.

3.4 Implications for Validating at Extrapolated Height

This section has established that the random-forest machine-learning model significantly outperforms the other extrapolation models in its ability to predict wind profiles based on surface measurements. Here, we consider whether machine-learning model performance is sufficiently high such wind speed validation could be performed at hub height, based on extrapolation near-surface measurements to hub height first. In principle, this approach would overcome the limitations of validating near the surface, as discussed in the previous section.



Figure 22. REWS performance metrics for the different vertical extrapolation models at NYSERDA Buoy E05 for unstable and stable conditions

First, we acknowledge that the machine-learning model was only found to perform well across the two NYSERDA buoy locations, and its ability to generalize across the U.S. Atlantic wind energy areas (or even the Pacific, Hawaii, and so on) is unclear and requires further research.

Assuming the machine-learning model would perform similarly well across the U.S. Atlantic, for example, then the low bias found in this study would support the use of machine learning to at least validate mean wind profiles. However, the cRMSE for modeling the REWS was nearly 1 ms^{-1} , considerably less than the other models, but still a moderate error. Therefore, the extrapolation of near-surface wind speeds for the purpose of validating at hub height or REWS would not be appropriate on a time series basis and would be associated with prohibitive uncertainty to allow for a meaningful validation. Therefore, as recommended in the previous section, validating modeled wind speeds using direct offshore lidar measurements remains the key recommendation of this report.



Figure 23. Mean modeled and observed diurnal cycles of normalized REWS at NYSERDA Buoys E05 and E06.



Figure 24. Heat maps (12 by 24) of REWS bias at NYSERDA Buoy E05 for the different extrapolation models



Figure 25. Mean observed and modeled wind profiles at NYSERDA Buoy E05 when using the DTU method based on 2-m and 20-m measurements

4 Summary and Future Work

4.1 Summary

Two key questions were answered in this report. First, what are the best data sources and methods for validating modeled U.S. offshore wind resource data sets? And second, what are the best methods for vertically extrapolating near-surface wind speed measurements to heights that span the rotor-swept area of modern offshore wind turbines?

4.1.1 Best Practices for Validation

Based on the analysis in this report, we make the following recommendations for validating U.S. offshore wind resource models:

- 1. The following four key performance metrics at a minimum should be calculated: the bias (capturing the difference in the modeled and observed means), the unbiased or centered root-mean-squared-error (which isolates the random error from the model), the Earth mover's distance (which quantifies the difference between two distributions), and the correlation coefficient (which measures the correspondence or "pattern" between two data sets).
- 2. Models should be validated at measurement height and not at an extrapolated height above the measurements (e.g., hub height). Generally, there is too much error in even the best extrapolation methods to meaningfully validate models against extrapolated measurements.
- 3. Near-surface measurements from buoys or satellite products are not well-suited for wind resource validation for two reasons. First, winds near the surface can behave very differently than winds across the rotor-swept heights, particularly during the diurnal cycle. Second, model performance metrics can be different near the surface than at the rotor-swept heights, and validating near the surface may incorrectly characterize model performance at the more relevant rotor-swept heights.
- 4. Land-based coastal measurements should generally not be used for offshore model validation for two main reasons. First, the coastal wind resource behaves very differently from the offshore wind resource, and validating along the coast may incorrectly characterize model performance in the farther offshore wind lease and Call Areas. Second, the land-sea boundary produces strong horizontal gradients in wind speed in this area, and validating 2-kilometer or greater resolution models becomes highly uncertain given these gradients.
- 5. Given the previously mentioned considerations and the lack of offshore meteorological towers in the United States, floating lidars remain the only robust data source to perform comprehensive validation of wind speeds across rotor-swept heights.
- 6. Given the size of expected offshore wind turbine rotor diameters (e.g., 175 m and 240 m for 8-megawatt [MW] and 15-MW designs, respectively), the rotor-equivalent wind speed should be calculated and used in place of the hub-height wind speed when performing model validation at floating lidars. Although the differences in overall model performance are small between the two, these differences become more substantial under conditions of high shear.
- 7. Numerical weather prediction models like WRF perform differently depending on atmospheric stability conditions; therefore, validating in unstable and stable conditions separately can help isolate causes for model underperformance. Generally, the bulk Richardson number (Ri_B) is the best stability metric to separate unstable and stable conditions, as it can characterize stability across the rotor-swept heights. However, Ri_B requires temperature measurements at two heights that are not available from lidar measurements. Estimates of Ri_B from the a model such as WRF, however, can provide a reasonable substitute.
- 8. Evaluating model performance by season and time of day can provide further insight into areas of model strength and deficiency. An effective way to disseminate both the seasonal and diurnal cycles is through a "12-by-24" performance matrix or heat map.

9. Finally, evaluating model performance during extreme events (high-shear events, extreme cold, storms, hurricanes, and so on) provides a gauge in how well a forecast model can predict anomalous events that can negatively impact an offshore wind farm.

4.1.2 Extrapolation Models

Of the four wind speed vertical extrapolation models considered, the random- forest machine-learning model significantly outperformed the other models and was able to accurately represent winds across the vertical profile in different seasons and times of day, and in different stability regimes. Furthermore, the machine-learning model substantially outperforms the WRF model, highlighting the benefit of local observations in generating wind profiles.

We emphasize that machine-learning model performance was based on training the model at one floating lidar (e.g., Buoy E05) and then evaluating its performance along with the other models at the other lidar (e.g., Buoy E06). In fact, very little performance degradation was found compared to testing the machine-learning model directly at the site it was trained, likely a result of the horizontally homogeneous conditions in the U.S. Atlantic offshore wind areas. This result has important implications for future use in creating observation-based wind atlases (see next section).

4.2 Future Work

4.2.1 Machine-Learning and U.S. Offshore Wind Atlases

The machine-learning approach to vertically extrapolate near-surface wind speeds shows considerable promise in the ability to produce observation-based wind atlases for U.S. offshore waters. Currently, the best observationbased wind atlas is the satellite-based atlas developed in Ahsbahs et al. (2020), a collaboration between NREL and DTU. This atlas uses only the 10-m wind speeds using the scatterometer-measured near-surface wind speeds that are extrapolated to 10 m using a geophysical transfer function. Training a machine-learning model on these data to predict offshore wind profiles at lidar sites should be a priority for future research and the first step toward an observation-based U.S. offshore wind atlas that can provide extrapolated measurements across the rotor-swept heights.

Ultimately, a machine-learning approach must rely on floating lidar data to train the model. Given the sparsity of these data sources, a future U.S. offshore wind atlas built from this approach will be relying on a relatively small number of training sites. Therefore, it is crucial to understand how machine-learning prediction accuracy decreases with distance offshore and how accurate, for example, we can expect a machine-learning-based wind atlas to be when several hundred kilometers from a floating lidar training site. Only two floating lidars 83 km apart were considered in this preliminary analysis, and the machine-learning model showed very little degradation across that distance. Immediate future work should consider all six currently deployed floating lidar data and apply a "round-robin" approach, in which a model trained at one site is evaluated at all remaining sites (as in Bodini and Optis 2020a). Such a study will provide critical information on how machine-learning extrapolation models perform not only with distance, but even across different oceans (i.e., a model trained in the Atlantic and applied in the Pacific).

This proposed scope of future research will be aided by continued efforts to make floating lidar data public. Most deployed lidars are currently owned by wind energy developers and publicly available. Public access to these data would greatly improve our understanding of the U.S. offshore wind resource and help produce more accurate hubheight observation-based offshore wind atlases.

4.2.2 Improving Single-Column-Model Performance

The single-column-model performance offshore could be improved considerably through better accounting of nearsurface stability. The model was forced at its lower boundary only by the 2-m wind speed and temperature and critically did not consider the role of sea surface temperature and related heat flux. Therefore, the single-column model really had no way to account for or characterize the role of atmospheric stability that was demonstrated in this study to be an important driver of the wind profile. By contrast, the WRF model is able to capture these effects and the machine-learning model used the air-sea temperature difference as an input variable which considerably improved model results. Improving the single-column-model design to account for atmospheric stability (e.g., by substituting the temperature lower-boundary condition with a flux-based measurement) should be an area of future work.

4.2.3 Measuring Offshore Atmospheric Turbulence and Stability

The ability to measure and characterize atmospheric stability in U.S. offshore waters is greatly hindered by the lack of observational data. In this study, the best metric possible was the air-sea temperature difference that describes only near-surface atmospheric stability and has limited predictive accuracy on stability aloft.

The emergence of microwave radiometry has the potential to rapidly improve this characterization. Radiometers are capable of measuring temperature and humidity profiles and were demonstrated in the Wind Forecasting Improvement Project 2 (WFIP2) in complex terrain (Shaw et al. 2019a) and are being researched for offshore applications (Shaw et al. 2019b). Recently, a U.S. Department of Energy small business funding award is exploring the use of buoy-mounted radiometers to measure temperature and humidity profiles offshore.

Should such data become available soon, an immediate effort to characterize atmospheric stability and validate the WRF model's ability to predict stability should be undertaken. It is currently unknown how well it characterizes atmospheric stability in U.S. offshore wind areas and, given the importance of stability in governing wind profiles, it is vital to understand this to improve WRF in offshore areas and more accurately characterize the U.S. offshore wind resource.

Finally, validating WRF model ability to estimate turbulence intensity should be a near-term priority. Turbulence intensity is the key atmospheric variable used by the wind industry to quantify atmospheric turbulence, and is now being measured by offshore floating lidars. Turbulence intensity is not a current output from WRF but is expected to be within a year or two. Validating modeled estimates of offshore turbulent intensity will be crucial to aid the emerging offshore wind industry in accurately characterizing turbine and wind farm design conditions.

References

- Ahsbahs, Tobias, et al. 2020. "US East Coast synthetic aperture radar wind atlas for offshore wind energy". Publisher: Copernicus GmbH, *Wind Energy Science* 5, no. 3 (): 1191–1210. ISSN: 2366-7443, visited on 11/14/2020. doi:https://doi.org/10.5194/wes-5-1191-2020.https://wes.copernicus.org/ articles/5/1191/2020/.
- Atlantic Shores Offshore Wind. 2020. "Atlantic Shores Floating LiDAR Buoy Data". https://erddap.maracoos.org/erddap/tabledap/AtlanticShores_833d_ebdb_69d0.graph?&time%3E= 2020-05-30T00%3A00%3A00Z&time%3C=2020-06-06T00%3A00%3A00Z.
- Badger, Merete, et al. 2015. "Extrapolating Satellite Winds to Turbine Operating Heights". *Journal of Applied Meteorology and Climatology* 55 (4): 975–991. ISSN: 1558-8424, visited on 01/13/2020. doi:10.1175/JAMC-D-15-0197.1. https://journals.ametsoc.org/doi/full/10.1175/JAMC-D-15-0197.1.
- Barthelmie, R. J., T. J. Shepherd, and S. C. Pryor. 2020. "Increasing turbine dimensions: impact on shear and power". Publisher: IOP Publishing, *Journal of Physics: Conference Series* 1618 (): 062024. ISSN: 1742-6596, visited on 11/11/2020. doi:10.1088/1742-6596/1618/6/062024. https://doi.org/10.1088% 2F1742-6596%2F1618%2F6%2F062024.
- Beiter, Philipp, et al. 2020. "The Cost of Floating Offshore Wind Energy in California Between 2019 and 2032": NREL/TP-5000-77384; BOEM-2020-48. doi:10.2172/1710181. https://www.nrel.gov/docs/fy21osti/77384.pdf.
- Bodini, Nicola, and Mike Optis. 2020a. "How accurate is a machine learning-based wind speed extrapolation under a round-robin approach?" Publisher: IOP Publishing, *Journal of Physics: Conference Series* 1618 (): 062037. ISSN: 1742-6596, visited on 11/14/2020. doi:10.1088/1742-6596/1618/6/062037. https: //doi.org/10.1088%2F1742-6596%2F1618%2F6%2F062037.
- . 2020b. "The importance of round-robin validation when assessing machine-learning-based vertical extrapolation of wind speeds". Publisher: Copernicus GmbH, *Wind Energy Science* 5, no. 2 (): 489–501. ISSN: 2366-7443, visited on 11/14/2020. doi:https://doi.org/10.5194/wes-5-489-2020.https://wes. copernicus.org/articles/5/489/2020/.
- Carbon Trust. 2018. Carbon Trust Offshore Wind Accelerator Roadmap. Tech. rep. Visited on 11/12/2020. https: //prod-drupal-files.storage.googleapis.com/documents/resource/public/owa-wuflr-updated-fl-roadmap_18102018.pdf.
- Cuxart, J., et al. 2006. "Single-Column Model Intercomparison for a Stably Stratified Atmospheric Boundary Layer". *Boundary-Layer Meteorology* 118, no. 2 (): 273–303. ISSN: 1573-1472, visited on 11/14/2020. doi:10.1007/s10546-005-3780-1. https://doi.org/10.1007/s10546-005-3780-1.
- Debnath, Mithu, et al. 2020. "Extreme Wind Shear Events in US Offshore Wind Energy Areas and the Role of Induced Stratification". Publisher: Copernicus GmbH, *Wind Energy Science Discussions* (): 1–22. ISSN: 2366-7443, visited on 11/14/2020. doi:https://doi.org/10.5194/wes-2020-103.https://wes.copernicus.org/preprints/wes-2020-103/.
- DNV-GL. 2020. "NYSERDA Floating LiDAR Buoy Data". https://oswbuoysny.resourcepanorama. dnvgl.com/.
- Draxl, C. 2015. "Overview and Meteorological Validation of the Wind Integration National Dataset Toolkit". *Renewable Energy*: 87.
- Dörenkämper, Martin, et al. 2015. "On the Offshore Advection of Boundary-Layer Structures and the Influence on Offshore Wind Conditions". *Boundary-Layer Meteorology* 155, no. 3 (): 459–482. ISSN: 1573-1472, visited on 11/14/2020. doi:10.1007/s10546-015-0008-x. https://doi.org/10.1007/s10546-015-0008-x.
- Hahmann, Andrea N., et al. 2020. "The making of the New European Wind Atlas Part 1: Model sensitivity". Publisher: Copernicus GmbH, *Geoscientific Model Development* 13, no. 10 (): 5053–5078. ISSN: 1991-959X, visited on 11/14/2020. doi:https://doi.org/10.5194/gmd-13-5053-2020.https://gmd.copernicus.org/articles/13/5053/2020/.

- Hersbach, Hans, et al. 2020. "The ERA5 global reanalysis". *Quarterly Journal of the Royal Meteorological Society* (): qj.3803. ISSN: 0035-9009, 1477-870X, visited on 06/21/2020. doi:10.1002/qj.3803. https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803.
- Jiménez, Pedro A., et al. 2012. "A Revised Scheme for the WRF Surface Layer Formulation". *Monthly Weather Review* 140 (3): 898–918. ISSN: 0027-0644. doi:10.1175/MWR-D-11-00056.1. https://doi.org/10. 1175/MWR-D-11-00056.1.
- Kaiser-Weiss, Andrea, and Peter Minnett. 2012. *GHRSST Strategy and Implementation Plan*. Tech. rep. Group for High Resolution Sea Surface Temperature. Visited on 11/17/2020. https://www.ghrsst.org/wp-content/uploads/2016/10/GHRSST-GDIP-2012-2022.pdf.
- Kelly, Mark C., and Sven-Erik Gryning. 2010. "Long-Term Mean Wind Profiles Based on Similarity Theory". Publisher: Springer Netherlands, *Boundary-Layer Meteorology* 136 (3): 377–390. ISSN: 0006-8314, visited on 11/14/2020. doi:10.1007/s10546-010-9509-9. https://orbit.dtu.dk/en/publications/ long-term-mean-wind-profiles-based-on-similarity-theory.
- Mayflower Offshore Wind. 2020. "Mayflower Floating LiDAR Buoy Data". http://www.neracoos.org/ erddap/tabledap/SHELL_MAYFLOWER_winds_csv_all.html.
- Musial, Walter, et al. 2020. 2019 Offshore Wind Technology Data Update. Tech. rep. NREL/TP-5000-77411, 1677477, MainId:26357. Visited on 11/12/2020. doi:10.2172/1677477.https://www.osti.gov/servlets/purl/1677477/.
- National Data Buoy Center. 1971. "Meteorological and oceanographic data collected from the National Data Buoy Center Coastal-Marine Automated Network (C-MAN) and moored (weather) buoys". https://www.ndbc.noaa.gov/.
- Optis, Michael, and Adam Monahan. 2016. "The Extrapolation of Near-Surface Wind Speeds under Stable Stratification Using an Equilibrium-Based Single-Column Model Approach". Publisher: American Meteorological Society, *Journal of Applied Meteorology and Climatology* 55, no. 4 (): 923–943. ISSN: 1558-8424, visited on 11/14/2020. doi:10.1175/JAMC-D-15-0075.1.https://journals.ametsoc.org/jamc/article/55/4/ 923/342359/The-Extrapolation-of-Near-Surface-Wind-Speeds.
- . 2017. "A Comparison of Equilibrium and Time-Evolving Approaches to Modeling the Wind Profile under Stable Stratification". Publisher: American Meteorological Society, *Journal of Applied Meteorology and Climatology* 56, no. 5 (): 1365–1382. ISSN: 1558-8424, visited on 11/14/2020. doi:10.1175/JAMC-D-16-0324.1. https://journals.ametsoc.org/jamc/article/56/5/1365/23433/A-Comparison-of-Equilibrium-and-Time-Evolving.
- Optis, Michael, Adam Monahan, and Fred C. Bosveld. 2014. "Moving Beyond Monin–Obukhov Similarity Theory in Modelling Wind-Speed Profiles in the Lower Atmospheric Boundary Layer under Stable Stratification". *Boundary-Layer Meteorology* 153 (3): 497–514. ISSN: 1573-1472, visited on 11/14/2020. doi:10.1007/ s10546-014-9953-z. https://doi.org/10.1007/s10546-014-9953-z.
- . 2016. "Limitations and breakdown of Monin-Obukhov similarity theory for wind profile extrapolation under stable stratification". Wind Energy 19 (6): 1053-1072. doi:https://doi.org/10.1002/we.1883. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/we.1883.https: //onlinelibrary.wiley.com/doi/abs/10.1002/we.1883.
- Optis, Michael, et al. 2020a. 2020 Offshore Wind Resource Assessment for the California Pacific Outer Continental Shelf. Tech. rep. NREL/TP-5000-77642. National Renewable Energy Laboratory (NREL), Golden, CO (United States). Visited on 11/14/2020. doi:10.2172/1677466. https://www.osti.gov/biblio/1677466.
- Optis, Mike, and Jordan Perr-Sauer. 2019. "The importance of atmospheric turbulence and stability in machinelearning models of wind farm power production". *Renewable and Sustainable Energy Reviews* 112 (): 27–41. ISSN: 1364-0321, visited on 11/14/2020. doi:10.1016/j.rser.2019.05.031.http://www. sciencedirect.com/science/article/pii/S1364032119303442.

Optis, Mike, et al. 2020b. "Validation of RU-WRF, the Custom Atmospheric Mesoscale Model of the Rutgers Center for Ocean Observing Leadership". *Renewable Energy*: 61.

Pacific Northwest National Laboratory. 2020. "Buoy Lidar Data, California". https://a2e.energy.gov/ data?ProjectFilter=%5B%22buoy%22%5D.

Pedregosa, F., et al. 2011. "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research* 12:2825–2830.

- Redfern, Stephanie, et al. 2019. "Incorporation of the Rotor-Equivalent Wind Speed into the Weather Research and Forecasting Model's Wind Farm Parameterization". Publisher: American Meteorological Society, *Monthly Weather Review* 147, no. 3 (): 1029–1046. ISSN: 0027-0644, visited on 10/20/2020. doi:10.1175/MWR-D-18-0194.1.https://journals.ametsoc.org/mwr/article/147/3/1029/107084/ Incorporation-of-the-Rotor-Equivalent-Wind-Speed.
- Shaw, William J., et al. 2019a. "The Second Wind Forecast Improvement Project (WFIP2): General Overview". _eprint: https://journals.ametsoc.org/bams/article-pdf/100/9/1687/4871688/bams-d-18-0036_1.pdf, *Bulletin of the American Meteorological Society* 100 (9): 1687–1699. ISSN: 0003-0007. doi:10.1175/BAMS-D-18-0036.1. https://doi.org/10.1175/BAMS-D-18-0036.1.
- Shaw, William J., et al. 2019b. "Workshop on Research Needs for Offshore Wind Resource Characterization" (). doi:10.2172/1572142.
- Smith, R. N. B. 1990. "A scheme for predicting layer clouds and their water content in a general circulation model". *Quarterly Journal of the Royal Meteorological Society* 116 (492): 435–460. doi:https://doi.org/10. 1002/qj.49711649210.eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10. 1002/qj.49711649210.https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/ qj.49711649210.
- Taylor, Karl E. 2001. "Summarizing multiple aspects of model performance in a single diagram". *Journal of Geophysical Research: Atmospheres* 106 (D7): 7183–7192. ISSN: 2156-2202, visited on 11/06/2020. doi:10.1029/2000JD900719. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900719.
- Wagner, R., et al. 2014. "Rotor equivalent wind speed for power curve measurement comparative exercise for IEA Wind Annex 32". Publisher: IOP Publishing, *Journal of Physics: Conference Series* 524 (): 012108. ISSN: 1742-6596, visited on 10/20/2020. doi:10.1088/1742-6596/524/1/012108. https://doi.org/10.1088%2F1742-6596%2F524%2F1%2F012108.
- Woods Hole Oceanographic Institute. 2020. "Martha's Vineyard Coastal Observatory". https://mvco.whoi.edu/.

5 Appendix



Figure 26. Heat maps (12 by 24) of REWS RMSE at NYSERDA Buoy E05 for the different extrapolation models



Figure 27. Heat maps (12 by 24) of REWS EMD at NYSERDA Buoy E05 for the different extrapolation models



Figure 28. Heat maps (12 by 24) of REWS R² at NYSERDA Buoy E05 for the different extrapolation models