The HRRR Meteorology, Energy, and Transmission (MET) Toolkit: Advancing high-resolution atmospheric data for contiguous U.S. energy applications

Nicola Bodini[1*], Emina Maric[1*], Ulrike Egerer[1*], Grant Buster[1], Luke Lavin[1], Pavlo Pinchuk[1], Brandon Benton[1], and David D. Turner[2]

[1]National Laboratory of the Rockies, Golden, CO, USA
[2]NOAA Global Systems Laboratory, Boulder, CO, USA

Correspondence: Nicola Bodini (nicola.bodini@nlr.gov)

* N.B., E.M., and U.E. contributed equally to this work.

Abstract. High-quality, multiyear atmospheric data are foundational for power system planning and grid integration. While the legacy Wind Integration National Dataset (WIND) Toolkit has long served as the industry standard, its historical record ends in 2013, leaving a critical gap in current modeling capabilities. Modern alternatives, such as the WIND Toolkit Long-term Ensemble Dataset (WTK-LED) and its Climate variant, offer extended coverage but exhibit higher wind speed biases and are computationally intensive to produce. This study introduces the High-Resolution Rapid Refresh Meteorology, Energy, and Transmission (HRRR MET) Toolkit, a repackaged version of the National Oceanic and Atmospheric Administration's native HRRR data. The HRRR MET Toolkit is designed to overcome the significant technical barriers associated with accessing native HRRR formats by providing a streamlined, user-friendly dataset with high vertical resolution at power generation-relevant heights. To ensure seamless continuity for long-term studies, the HRRR MET Toolkit is provided on the same uniform 2 km horizontal grid as the legacy WIND Toolkit, offering both modern accessibility and spatial consistency with the established historical record. To evaluate potential performance gains, we also assessed an experimental bias-corrected version using quantile mapping against the WIND Toolkit as a climatological reference. We provide a comprehensive validation of both HRRR variants alongside the WTK-LED, its Climate variant, and the 2023 National Offshore Wind (NOW-23) dataset against long-term observations across the contiguous United States. Results indicate that the HRRR MET Toolkit significantly outperforms the WTK-LED suite; for instance, it reduces hub-height average wind speed bias to 0.10 m s^{-1} (compared to 0.82 m s^{-1} for the WTK-LED) and achieves an hourly wind speed correlation of 0.82. Critically, the comparison between the native and bias-corrected HRRR variants reveals that the statistical correction offers marginal benefit and in some cases exacerbates positive wind speed biases in complex terrain. We conclude that the native HRRR physics are sufficiently robust for energy applications and therefore recommend the HRRR MET Toolkit as a highly accessible, accurate, and less complex standard for modern power system studies in the United States.

Copyright statement. This work was authored in part by the National Laboratory of the Rockies for the U.S. Department of Energy (DOE), operated under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Critical Minerals and Energy Innovation Wind Energy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE, NOAA, 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.

1 Introduction

High-quality land-based and offshore wind and other atmospheric data covering multiple years of weather patterns are essential for planning and operating power systems with increasing shares of wind energy generation. With wind energy generation exceeding 10 % of annual electricity generation in the United States in 2022 (U.S. Energy Information Administration, 2023), accurate and temporally extensive wind resource data can improve efficiency of the North American power system. The energy community increasingly relies on large-scale modeled datasets to support long-term planning, grid operations, and resource assessment.

Historically, the National Laboratory of the Rockies' (NLR's) Wind Integration National Dataset (WIND) Toolkit (Draxl et al., 2015) has served as the gold standard for such analyses in North America, providing high-resolution meteorological data for the years 2007–2013 with a relatively low wind speed bias. More recently, NLR released the WIND Toolkit Long-term Ensemble Dataset (WTK-LED) (Draxl et al., 2024) and its Climate variant ~~to~~ (a 20 year, 4 km resolution continental simulation developed to provide a climatological record) to extend this record. However, these newer datasets present some trade-offs: they are computationally expensive to produce, and recent evaluations – derived from a comparison of legacy and modern validation analyses (Draxl et al., 2015, 2024) as well as the analysis presented herein – show they exhibit larger biases against meteorological observations than the legacy WIND Toolkit. For offshore wind analysis, NLR released the 2023 National Offshore Wind (NOW-23) Data Set (Bodini et al., 2024), which spans over 20 years but currently lacks continuous updates and has not been validated across all modeled regions. Consequently, these datasets do not fully meet all the best-practice criteria outlined by the Energy Systems Integration Group for modern power systems planning (Energy Systems Integration Group, 2023), particularly regarding multi-decadal coverage and a mechanism for continuous annual extension.

To address the need for a continuously updatable, high-fidelity dataset, the community has increasingly looked toward archived data from operational weather forecast models. The National Oceanic and Atmospheric Administration's (NOAA's) High-Resolution Rapid Refresh (HRRR) model (Dowell et al., 2022) provides an hourly updating operational weather forecast capability across the contiguous United States (CONUS). The potential of the HRRR archive for renewable energy applications was recognized early on: James et al. (2017) developed a unified high-resolution wind and solar energy dataset from the 2013–2015 HRRR record, and James et al. (2018) extended this effort to offshore wind speed estimates. Both studies included comparisons against the legacy WIND Toolkit using 80 m winds, thereby anticipating the transition to the HRRR-based energy dataset described herein. However, three significant barriers have historically prevented the widespread use of archived native HRRR data in power system analyses. First, the native data are stored in GRIB formats on hybrid levels that are difficult for

energy modelers to utilize. Second, the model provides limited vertical resolution at heights relevant to modern wind turbines. Third, while HRRR shows strong weather forecast skill, previous studies have documented some wind speed biases (Buster et al., 2024b), particularly in regions with complex terrain (Pichugina et al., 2019; Collins et al., 2024) and for offshore ramp events (Yin and Peña, 2024).

To close this gap, NLR developed the HRRR Meteorology, Energy, and Transmission (MET) Toolkit. This initiative provides a streamlined, energy-ready version of the HRRR (2015–2025) by performing spatial regridding, vertical interpolation, and temporal gap-filling. Critically, to maintain continuity with a decade of industry-standard modeling, the HRRR MET Toolkit is provided on the same horizontal grid as the legacy WIND Toolkit, ensuring the data are both accessible and spatially consistent with the historical record. Furthermore, motivated by the high accuracy of and desire of consistency with the legacy WIND Toolkit, NLR also developed an experimental Bias-Corrected version (BC-HRRR) (Buster et al., 2024b, c). This version applies a quantile mapping methodology to anchor the HRRR distribution to the legacy WIND Toolkit's climatological baseline. This study presents both the native HRRR MET Toolkit and the BC-HRRR variant to determine whether statistical post-processing adds significant value over a robustly repackaged operational dataset.

Validation of wind resource datasets is fundamental for their acceptance in stakeholder applications; while modeled data provide consistent spatiotemporal coverage, they remain prone to systematic errors and biases when compared to in situ measurements. The Energy Systems Integration Group (Energy Systems Integration Group, 2023) explicitly identifies validation "against real conditions" as a primary requirement for high-quality resource datasets intended for power systems analyses. Ideally, models are validated against direct wind speed measurements from meteorological towers or remote sensing devices taken at commercial hub heights. Significant validation efforts have been undertaken internationally to meet this standard, most notably for the New European Wind Atlas. This initiative utilized an extensive network of masts and large-scale lidar campaigns to validate wind climates across Europe (Dörenkämper et al., 2020; Murcia et al., 2022; Cheynet et al., 2025). However, replicating this level of validation across North America remains challenging. The public availability of multi-decadal, (sub-)hourly time series at nominal turbine hub heights (approximately 80–140 m) is limited, creating a validation gap for large-scale spatiotemporal datasets in the United States.

While valuable validation studies exist within the North American domain, they typically face at least one of three critical limitations: low measurement height, restricted geographic scope, or reliance on vertical extrapolation. First, many studies rely on observations well below modern hub heights (Sheridan et al., 2025c, a) as they often depend on towers ranging from 10 m to 60 m. Even the recent comprehensive validation of the WTK-LED Climate dataset by Peco et al. (2025) utilized 26 towers where the vast majority of sensors were located well below 100 m, failing to fully capture the wind shear profile relevant to modern rotors. Second, to overcome height limitations, researchers sometimes extrapolate near-surface measurements to hub height using power law, logarithmic profiles, or machine learning (Bodini et al., 2023), an approach that introduces additional uncertainty to the validation. Third, studies that do utilize high-quality hub-height data are often geographically restricted or dataset-specific. This includes validations focused exclusively on offshore environments (Optis et al., 2020b; Sheridan et al., 2025b), a single site in Oklahoma (Pronk et al., 2022) or in the southeastern United States (Liu et al., 2025), or several sites all located within the Columbia River Gorge (Bianco et al., 2019). A recent and notable exception by Wilczak et al. (2024)

provided a continent-wide validation of hub-height winds across the contiguous United States, however their work was focused on evaluating a global reanalysis product (ERA5). Consequently, a comprehensive, continent-wide validation at hub height that inter-compares the full suite of modern, high-resolution national datasets remains outstanding in published literature.

In this paper, we perform a comprehensive comparison and validation of an entire suite of wind resource datasets for CONUS: the legacy WIND Toolkit, WTK-LED, WTK-LED Climate, NOW-23, the HRRR MET Toolkit, and BC-HRRR. We specifically investigate the value of bias-correcting an existing operational meteorological dataset (HRRR) to a wind-energy-specific baseline (WIND Toolkit) and whether these datasets can effectively replace the computationally intensive simulation of new datasets. To answer these questions, we inter-compare the individual models and subsequently validate them against observations across CONUS using multiyear measurement periods at hourly resolution. Section 2 introduces the modeled datasets used in our work, which are inter-compared in Sect. 3. Section 4 presents the observational datasets and methods used for model validation, the results of which are described in Sect. 5. We conclude and suggest future directions in Sect. 6.

2 Modeled datasets

A variety of wind resource datasets have been developed to support the U.S. power system sector, ranging from historical reanalysis-type products to operational forecast-based models. While the legacy WIND Toolkit (Draxl et al., 2015) served as the industry standard for a decade, its temporal coverage ended in 2013, necessitating the development of the next-generation products described in this section. A table comparing the main technical specifications of the various datasets is included in the Appendix (Table A1).

2.1 WTK-LED and its Climate variant

The WTK-LED represents a significant update to the legacy WIND Toolkit, transitioning from a deterministic framework to a dataset that provides high-resolution time series alongside some model uncertainty estimates (Draxl et al., 2024). The dataset was generated using the Weather Research and Forecasting (WRF) model (v4.1.3) with boundary conditions provided by the ERA5 reanalysis (Hersbach et al., 2020). The core WTK-LED CONUS product provides data at a 2 km horizontal resolution and a 5 min temporal resolution. While originally released for the 2018–2020 period, additional years (2014–2015) have since been added and made publicly available. To characterize the wind resource across the entire rotor layer of modern turbines, the vertical grid includes nine heights between 10 and 200 m (10, 40, 60, 80, 100, 120, 140, 160, and 200 m).

To complement the high-resolution short-term simulations, the WTK-LED Climate variant was developed by Argonne National Laboratory. This climate-scale simulation covers the North American continent at a 4 km spatial resolution and 1 h temporal resolution, spanning a 20-year period from 2001 to 2020. While WTK-LED Climate offers a more robust climatological record, initial validations across 13 CONUS sites have shown a generally positive wind speed bias ranging from 0.5 to 2 m s^{-1}, with performance metrics varying significantly by site and terrain complexity (Draxl et al., 2024).

2.2 NOW-23

The NOW-23 dataset is an offshore-specific, WRF-based product that supersedes the offshore components of the original WIND Toolkit (Bodini et al., 2024). NOW-23 provides a state-of-the-art wind resource atlas for all U.S. offshore waters (excluding Alaska), covering the Atlantic, Pacific, Gulf of Mexico, Hawaii, and Great Lakes regions. Unlike the uniform national approach of WTK-LED, NOW-23 employs a regional tuning strategy. For each offshore region, the WRF model (v4.2.1) setup was selected based on validation against available observations (e.g., buoys and lidars) to account for unique regional meteorological phenomena. NOW-23 offers a consistent 20-year record (starting in 2000 and extending to 2019–2022, depending on the region) with 2 km horizontal resolution and 5 min temporal resolution. The dataset is particularly optimized for energy applications, featuring 12 vertical levels in the lowest 300 m of the atmosphere to accurately resolve the wind profile across the rotor disk. Validation efforts for NOW-23 (Bodini et al., 2024) have demonstrated its superior accuracy compared to legacy products.

2.3 HRRR MET Toolkit

The HRRR is an operational, hourly updating, convection-allowing atmospheric model developed by NOAA. Built on the Advanced Research Weather Research and Forecasting core, the HRRR covers CONUS with a 3 km horizontal resolution (Dowell et al., 2022). The model is initialized hourly using a hybrid ensemble-variational data assimilation system that integrates radar reflectivity and other conventional observations to accurately capture mesoscale weather phenomena. Since its initial operational implementation in September 2014, the HRRR has undergone continuous development, resulting in four distinct versions (HRRRv1 through HRRRv4) that introduced improvements in physics and data assimilation (Table 1). The HRRR is known to be a highly accurate operational weather forecast model in the United States (James et al., 2022). Previous validation efforts show native HRRR data better match available wind generation observations than common alternatives like ERA5 (Hersbach et al., 2020) and MERRA-2 (Gelaro et al., 2017) at a hub height of 80 m (Millstein et al., 2023), although with a tendency to underestimate winds in selected U.S. regions and for some wind speed ranges (Buster et al., 2024b; Pichugina et al., 2019; Collins et al., 2024; Yin and Peña, 2024).

Table 1. HRRR model versions and operational start dates.

Version	Start
HRRRv1	30 September 2014
HRRRv2	23 August 2016
HRRRv3	12 July 2018
HRRRv4	02 December 2020

Unlike reanalysis datasets (e.g., ERA5 or MERRA-2) that are generated retrospectively with a frozen model configuration, the HRRR archive consists of real-time operational forecasts. While the HRRR provides a highly accurate record, the nature

of the operational archive presents significant challenges for power system applications: the data are distributed as individual hourly GRIB files containing instantaneous fields with forecast lead times ranging from 0 hours (analysis) to up to 48 hours, stored on irregular hybrid vertical levels that offer limited resolution at modern turbine heights, and containing various temporal gaps (prior to 2019) in the public archive.

To resolve these barriers, NLR developed the HRRR MET (Meteorology, Energy, and Transmission) Toolkit. This toolkit "repackages" the native HRRR archive into a continuous, energy-ready dataset that is both accessible and spatially consistent with legacy NLR datasets. The creation of the HRRR MET Toolkit involves a multistage processing pipeline:

- Data acquisition and formatting. We acquire open-source HRRR data from NOAA in GRIB format. To ensure the most realistic representation of atmospheric conditions, we utilize the f02 forecast hour (two hours after initialization) rather than the analysis hour (f00). This choice allows for sufficient model spin-up, balancing the ingested observational data with the model's physical consistency~~(Vibrant Clean Energy, 2024).~~ . James et al. (2017) discuss the rationale for preferring short-range HRRR forecasts over the f00 analysis in detail, having adopted the same approach for their HRRR-based renewable energy dataset (see also Vibrant Clean Energy, 2024). We compile these hourly forecasts to create a continuous historical record.

- Spatial regridding. To ensure compatibility with the widely used WIND Toolkit, we re-grid the native 3 km HRRR data onto the 2 km WIND Toolkit spatial grid. This is achieved using inverse distance-weighted (IDW) interpolation with the four nearest neighbors. Each HRRR grid cell is matched to the four nearest WIND Toolkit grid cells, and the resulting value is calculated by weighting the neighbors inversely by distance, preserving spatial variability while aligning the projections. However, users should be aware that interpolating from a coarser grid to a finer grid inherently smooths the data without generating new physical information. While the four-point IDW interpolation introduces modest additional spatial smoothing, the primary constraint is dynamical: established spectral analyses indicate that internal numerical diffusion limits the effective resolution of WRF-based models to approximately $7\Delta x$ (Skamarock, 2004). Consequently, while the HRRR MET Toolkit is provided on a 2 km grid, the native 3 km HRRR – and by extension, this regridded dataset – resolves true physical features at scales of roughly 20 km.

- Vertical interpolation. The HRRR model natively stores data on irregular hybrid sigma-pressure levels, which vary spatially and temporally. Common turbine hub heights (e.g., 60–140 m) typically fall between the lower model levels (e.g., 8, 23, 53, 106, and 198 m). We calculate the HRRR native level heights above ground as the geopotential height minus the model's orography. We linearly interpolate wind speeds from these time-varying native levels to fixed heights above ground level. The resulting HRRR MET Toolkit dataset provides wind resource variables at the same vertical levels as the WIND Toolkit: 10, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 400, and 500 m. This method is computationally efficient and preserves negative wind shear profiles critical for wind power analysis. It makes the data immediately usable for industry-standard power conversion models.

– Temporal gap filling. The public HRRR f02 native files contain missing data, specifically 1,789 missing hours between 2015 and 2018 (approximately 5.1 % of the period). To provide a complete hourly record, we employ a hierarchical gap-filling strategy:

1. Alternative forecast horizons: Missing f02 time steps are populated using forecasts from previous initialization hours, utilizing horizons up to f06 (e.g., filling a missing 06:00 UTC f02 using the f03 forecast initialized at 03:00 UTC).
2. Pressure-level data: If native-level data are unavailable through f06, we utilize f02 pressure-level data, which have significantly fewer gaps (only 47 missing hours in the 2015–2017 period).
3. Linear interpolation: For the remaining 25 hours of missing data (all in 2016, including one full missing day), values are linearly interpolated between the closest available valid time steps.

The HRRR MET Toolkit currently covers 2015–2025 and is intended to be extended annually. The following variables are available as 60 min instantaneous data: boundary layer height, friction velocity, latent heat flux, precipitation rate, pressure, relative humidity, roughness length, sensible heat flux, skin temperature, specific humidity, temperature, wind direction, and wind speed. Note that boundary layer height, friction velocity, latent heat flux, roughness length, and sensible heat flux are not available for the HRRRv1 period.

Radiation and solar irradiance variables were not included in the current release. For high-fidelity solar resource data in the United States, users typically rely on the National Solar Radiation Database (NSRDB, Sengupta et al. (2018)). Users specifically requiring solar data that is physically consistent with the HRRR MET Toolkit meteorology can extract shortwave and longwave radiation fields directly from the native NOAA HRRR archive. However, systematically evaluating the accuracy of these native HRRR radiation fields against the NSRDB remains an important area for future research, particularly to support combined wind-solar grid integration studies. Should solar radiation variables be incorporated into a future release of the HRRR MET Toolkit, we note that – in contrast to the wind fields, for which the f02 forecast is preferred to allow model spin-up – the HRRR downward shortwave radiation is most accurate at the 0-h analysis time (f00), as the HRRR exhibits a positive shortwave radiation bias that grows with forecast horizon (Benjamin et al., 2025).

2.3.1 Characterization of wind speed in the HRRR MET Toolkit

Since it is a newly released dataset, here we provide a climatological characterization of the hub height wind conditions as modeled in the HRRR MET Toolkit dataset. Fig. 1 shows the mean wind speeds for the 2015–2025 period at 100 m (a proxy for land-based hub height). We note that the HRRR dataset spans four model versions (HRRR v1–v4; Table 1), each introducing improvements in model physics and data assimilation. The impact of these version transitions on temporal consistency of the dataset is analyzed in Section 5 (Figure 18).

The annual cycle of 100 m wind speed, expressed as the difference of each month's mean from the overall annual mean, is shown in Fig. 2. Overall, the wind resource is larger in the winter and early spring months across the majority of CONUS, with

the lowest wind speeds occurring in the summer. The Atlantic coast exhibits the largest magnitude in the annual cycle, whereas the magnitude is more limited in the central U.S. wind belt where the majority of wind turbines are currently deployed.

The diurnal cycle, expressed as the difference of each UTC hour's mean from the overall mean, is presented in Fig. 3. A stronger diurnal cycle is observed in the mountainous areas of the western U.S. and in the southern portion of the central U.S. (around Texas), characterized by stronger winds in the local evening (03:00 UTC) and lowest winds in the local morning (16 UTC). The timing of the cycle varies geographically: mountain regions show an earlier cycle where stronger winds begin after local noon and end around midnight. In contrast, the central Plains exhibit a delayed cycle with stronger winds starting in the late evening and lasting through dawn, a phenomenon potentially linked to the frequent nocturnal low-level jets in those regions.

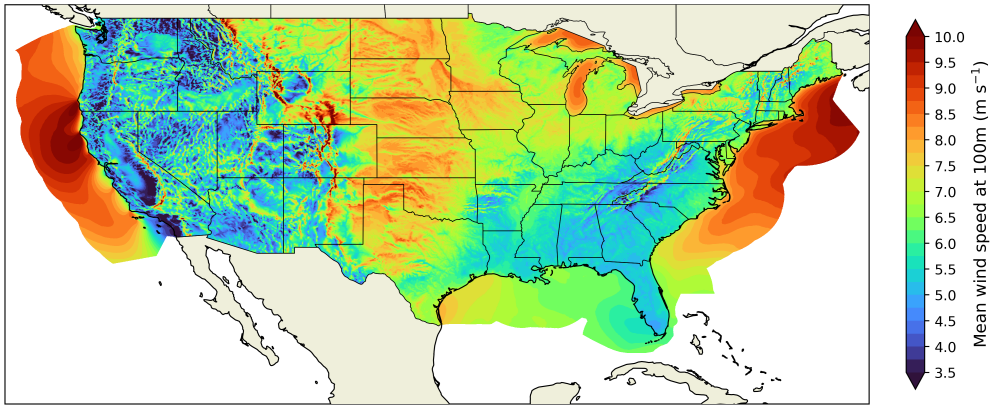


Figure 1. Mean 100 m wind speed from the HRRR MET Toolkit (2015–2025).

2.4 BC-HRRR

The BC-HRRR dataset is an experimental variant (2015–2023) derived from the HRRR MET Toolkit~~to test~~ , developed to consider whether applying a statistical bias correction (BC) against a wind-energy-specific baseline could ~~further~~ improve the accuracy of the operational data~~and ensure a more seamless transition between legacy and modern records.~~ ; the effectiveness of this BC approach is evaluated in Sect. 5 and discussed in Sect. 6. The "BC" in BC-HRRR refers to a quantile mapping bias correction (Maraun, 2016) applied to the interpolated wind speeds to anchor the HRRR distribution to the WIND Toolkit baseline. The corrected wind speed x_{BCHRRR} is calculated as:

$$x_{\mathrm{BCHRRR}} = CDF^{-1}_{\mathrm{WIND\ Toolkit}}(CDF_{\mathrm{HRRR}}(x_{\mathrm{HRRR}})) \tag{1}$$

where CDF_{HRRR} is the cumulative distribution function of the 2015–2023 HRRR MET Toolkit data, and $CDF^{-1}_{\mathrm{WIND\ Toolkit}}$ is the inverse CDF of the 2007–2013 WIND Toolkit data. This process is performed with CDFs unique to each spatial grid point and height level using 100 quantiles sampled with an inverse log relationship to better resolve the high-wind-speed tail of the

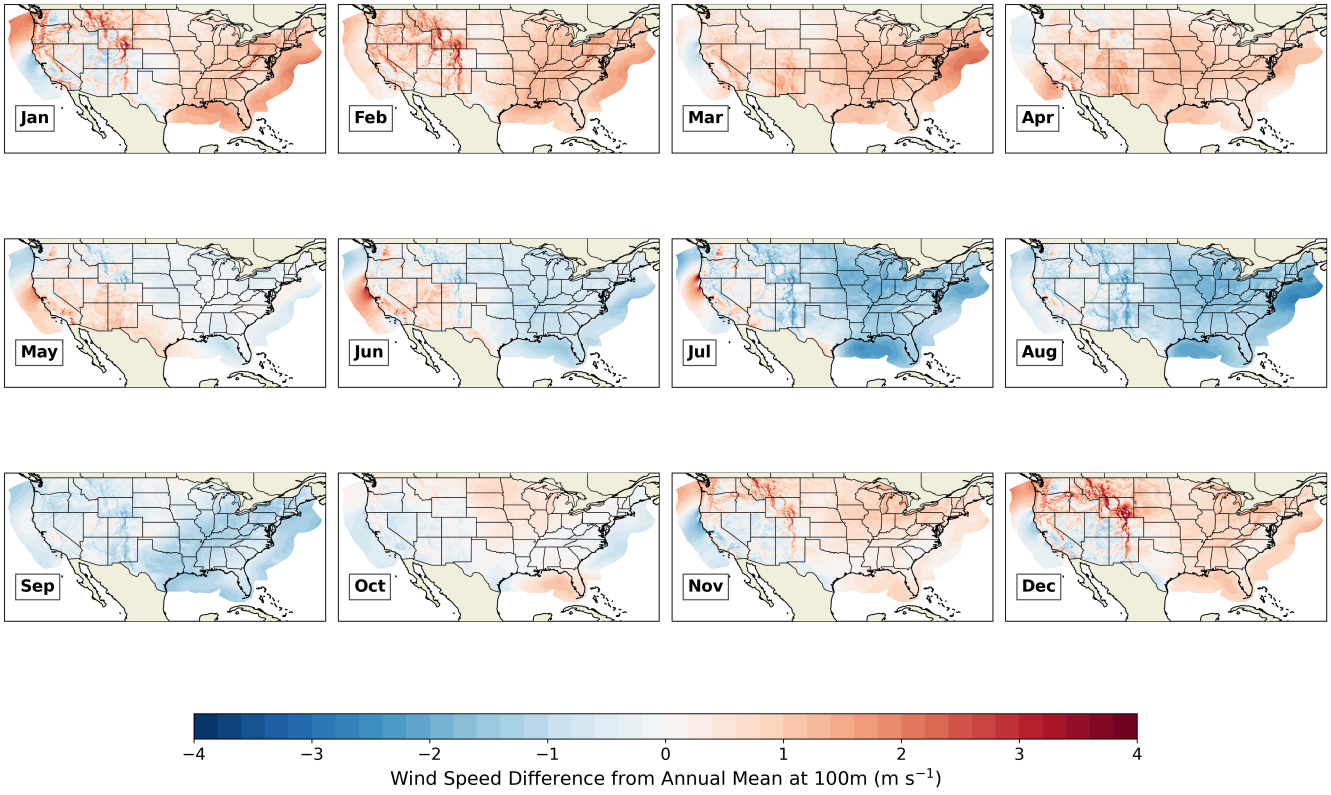


Figure 2. Maps showing the annual cycle in 100 m wind speed in the HRRR MET Toolkit, expressed in terms of the difference for each month's mean wind speed from the overall mean.

distribution. By construction, the quantile mapping procedure eliminates mean distributional differences between the HRRR and the WIND Toolkit at each grid point and height level. Due to standard numerical interpolation between discrete quantiles, negligible residual differences remain (± 0.03 m s^{-1}), meaning the long-term mean wind speed bias is effectively zero (see Appendix C). We note that wind direction is not affected by this correction. Also, future work may explore the use of quantile delta mapping (Cannon et al., 2015) as a method to allow for long-term trends and annual deviations in wind resource from the WIND Toolkit baseline. Appendix B analyzes ~~ERA-5~~ ERA5 long-term data to assess the extent to which the WIND Toolkit period represents long-term conditions. ~~Appendix C compares the mean wind speed from the BC-HRRR and the WIND Toolkit to check that the bias correction approach worked as intended.~~ We utilize ERA5 reanalysis data for this assessment because it provides a consistent, multi-decadal dataset necessary for evaluating relative temporal variability. We acknowledge that ERA5 is not bias-free; for instance, Wilczak et al. (2024) and Pronk et al. (2022) have documented systematic wind speed biases in ERA5 over the contiguous United States. However, ERA5 is used here exclusively to evaluate internal temporal shifts rather than as a validation reference. Because our analysis focuses strictly on relative temporal anomalies – comparing specific ERA5 epochs against the ERA5 long-term mean – these systematic biases are effectively differenced out.

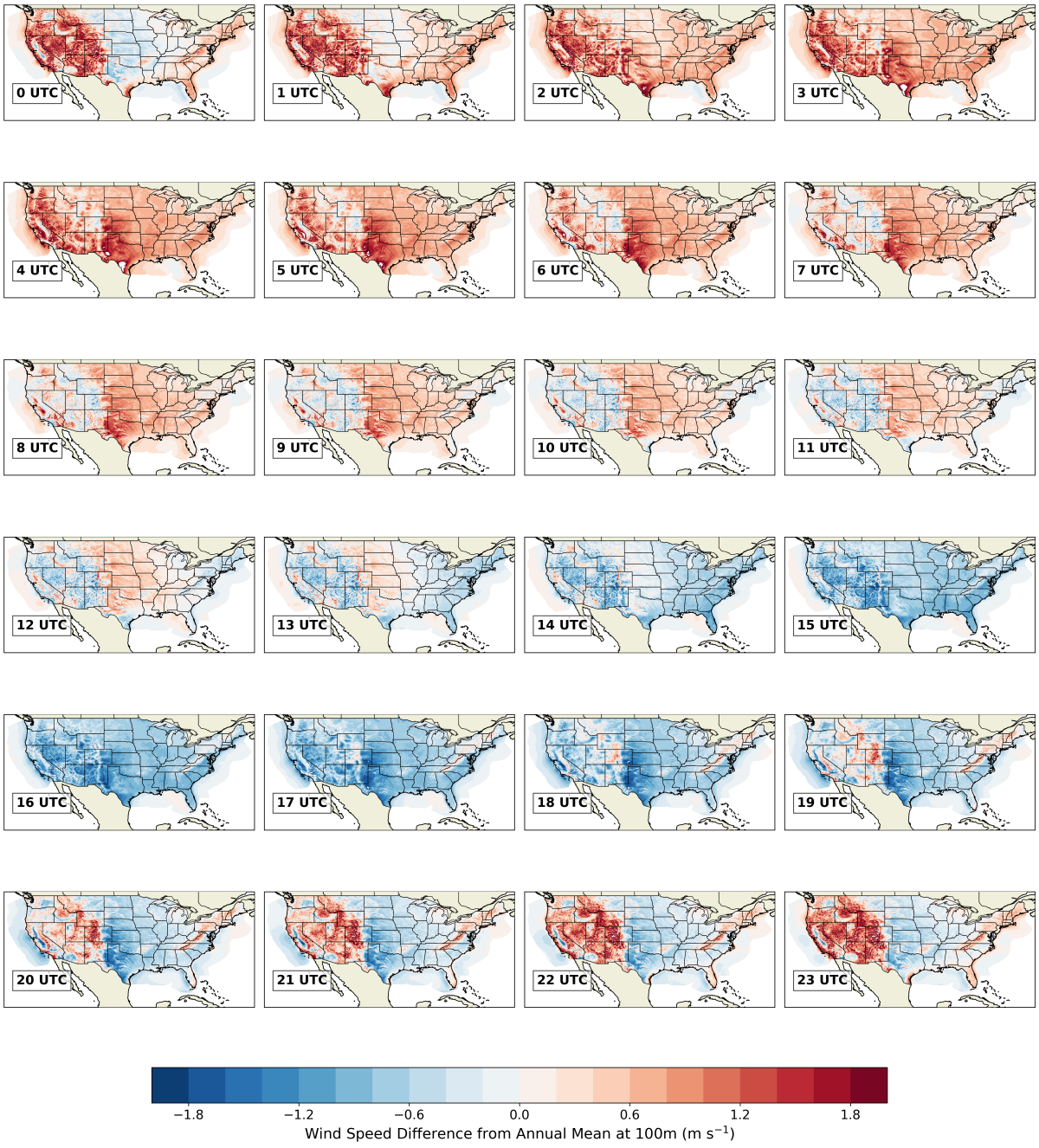


Figure 3. Maps showing the diurnal cycle in 100 m wind speed in the HRRR MET Toolkit, expressed in terms of the difference for each (UTC) hour's mean wind speed from the overall mean.

The following variables are included as 60 min instantaneous data in BC-HRRR: wind speed and direction at 10, 20, 40, 60, 80, 100, 120, 140, 160, 180, and 200 m; temperature at 2, 10, 20, 40, 60, 80, 100, 120, 140, 160, 180, and 200 m; pressure at 0 m; cumulative precipitation; and specific and relative humidity at 2 m.

3 Inter-comparison of modeled datasets

It is critical to validate synthetic wind resource datasets to ensure their suitability for power systems analysis (Energy Systems Integration Group, 2023). Direct validation against in situ measurements is the gold standard and will be the subject of Sect. 4 and Sect. 5. However, continental-scale validation at modern turbine hub heights is constrained by the scarcity of publicly available, long-term observational campaigns. Thus, it is also important to characterize the relationship between the different wind resource datasets across the United States. Understanding these differences is critical for users intending to concatenate multiple datasets to create multi-decadal records or those transitioning established workflows from legacy products to newer ones.

We present a comprehensive inter-comparison of all overlapping dataset pairs in Fig. 4. To ensure ~~methodological consistency,~~ a fair, like-for-like comparison that isolates model differences from inter-annual climate variability, each pairwise difference (Row minus Column) in this matrix is computed exclusively using the overlapping temporal periods ~~available~~ for that specific ~~pair.~~ dataset pair. We note that this strict temporal matching differs from the experimental bias-correction methodology used for BC-HRRR (Section 2.4). In that experimental approach, the WIND Toolkit (2007–2013) was deliberately used as a climatological target for the HRRR (2015–2023) despite the absence of temporal overlap – a constraint necessitated by dataset availability, but justified by the representativeness of the WIND Toolkit epoch for long-term CONUS wind conditions (see Appendix B). For the direct model inter-comparisons presented here, however, exact temporal overlap is strictly required.

The WIND Toolkit is excluded from this matrix because it shares minimal temporal overlap with the other datasets, intersecting only with NOW-23 and WTK-LED Climate. Readers interested in a comparison between the WIND Toolkit and NOW-23 are referred to Bodini et al. (2024).

The inter-comparison reveals a systematic divergence between the WTK-LED family and the HRRR-based products. As illustrated in the upper-right quadrant of Fig. 4, both WTK-LED and WTK-LED Climate exhibit consistently higher mean wind speeds than the HRRR products across the majority of the CONUS domain. Subsequent validation (Sect. 4 and Sect. 5) confirms that the lower wind speeds in HRRR MET Toolkit and BC-HRRR represent a significant gain in accuracy, as the legacy WTK-LED products tend to overestimate the resource. Notably, the two WTK-LED variants show relatively strong internal consistency, with the original WTK-LED modeling slightly stronger winds (mean differences <0.5 m s^{-1}) than its Climate counterpart (except for some larger deviations in the eastern United States).

For the offshore resource, comparisons involving the regionally segmented NOW-23 dataset reveal that it generally predicts slightly higher wind speeds than BC-HRRR and HRRR MET Toolkit. The closest agreement occurs off the southern Pacific coast, Gulf of America, and off the southern Atlantic coast. This alignment is likely driven by methodological similarities, as these regions utilized a planetary boundary layer scheme (see details in Table A1) for NOW-23 that is consistent with the

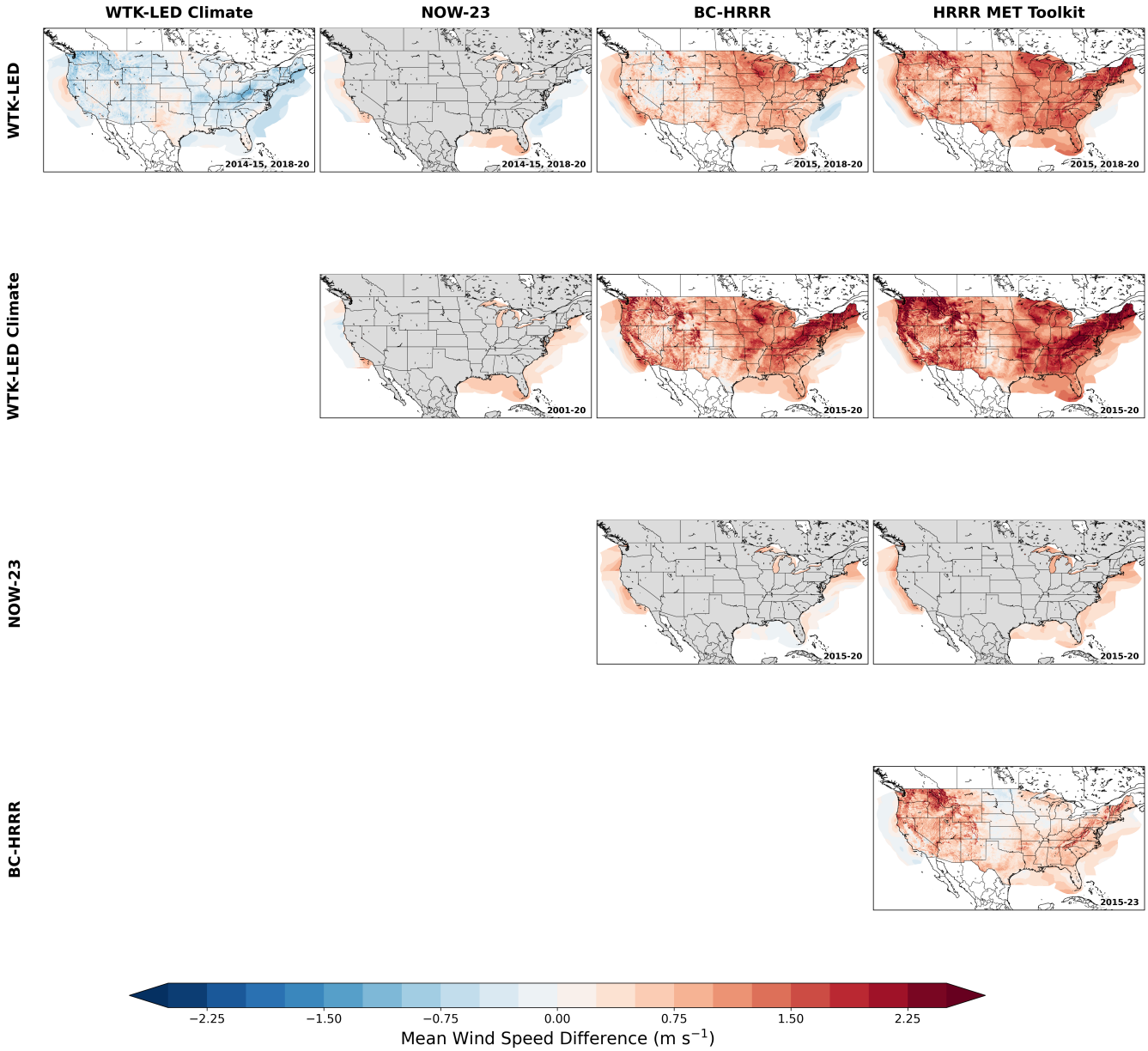


Figure 4. Matrix of pairwise mean 100 m wind speed differences between the considered modeled datasets. Each panel displays the mean difference calculated as the dataset on each row minus the dataset on each column. Differences are computed using only the overlapping time periods available for each specific pair, as indicated by the year labels in the bottom-right corner of each panel. Onshore regions are masked in gray for comparisons involving NOW-23, which is exclusively an offshore dataset.

WIND Toolkit (which serves as the bias-correction proxy for BC-HRRR). In contrast, BC-HRRR models lower mean wind

speeds than NOW-23 in other coastal regions, though these differences typically remain below 1 m s^{-1}. Overall, dataset spread is notably more muted offshore than onshore, suggesting higher modeling consensus in marine environments.

Finally, comparing BC-HRRR against the HRRR MET Toolkit (processed with identical regridding and interpolation but without bias correction) isolates the specific impact of the quantile mapping. The most pronounced adjustments occur in the complex terrain of the western United States and the Appalachian Mountains, where BC-HRRR yields higher mean 100 m wind speeds. Conversely, in the "wind belt" in the central United States – the region currently hosting the highest density of land-based wind capacity – the differences are minimal.

4 Data and methods for model validation

Observational datasets for model validation were selected to span a range of geographies, climates, and land–sea interfaces across CONUS to ensure that model strengths and weaknesses are captured across multiple environments. We prioritize datasets that are publicly accessible, consistently archived, and with near-hub-height measurements (e.g., tall towers and lidars). This allows a more direct and physically meaningful comparison to model wind speeds at heights relevant for the wind energy industry without relying on vertical extrapolation methods. We further target long, continuous time series, prioritizing datasets with multiyear records and at least hourly resolution to capture diurnal cycles. The observational data selected for this study are summarized in Table 2, and Fig. 5 shows a terrain map with the observational stations.

The five Gulf of Mexico stations (42361–42364 and 42394) consist of anemometers mounted on structures atop large offshore oil and gas platforms. While sensor heights range from 100 to 122 m above mean sea level, the host platforms have topside decks sitting roughly 12–34 m above the waterline, and their structures can induce localized flow distortion and wakes. That said, the anemometers are positioned approximately 65–110 m above the main deck level, placing them above the zone of most severe structural blockage. Because platform-specific structural metadata are absent from the public database, no external flow distortion corrections were applied beyond the standard quality control filtering used across the whole Tall Tower Dataset (Ramon et al., 2020), as described below (see Step 2b).

We validate model performance against wind speed observations. The steps for preparing the observations and model data for comparison include:

1. Download observations and standardize them into a common format across all sites. For lidar datasets, if not already provided, calculate wind speed and wind direction from velocity-azimuth-display scans (Frehlich et al., 2006).

2. Apply data quality control (QC) for wind speed observations:

   (a) Apply dataset-specific QC flags that are recommended in the source datasets. For the Tall Tower Dataset (Ramon et al., 2020), we apply the Quality Control Software Suite for Tall Towers (QCSS4TT) quality flags and retain only observations with QC flag = 1 (passing all applicable checks). The QCSS4TT includes a dedicated tower shadow test that removes wake-contaminated observations at stations with more than one anemometer at a given height (the specific stations are not identified in the public metadata). For the Iowa Mesonet ETTI4 tower, which has dual anemometers

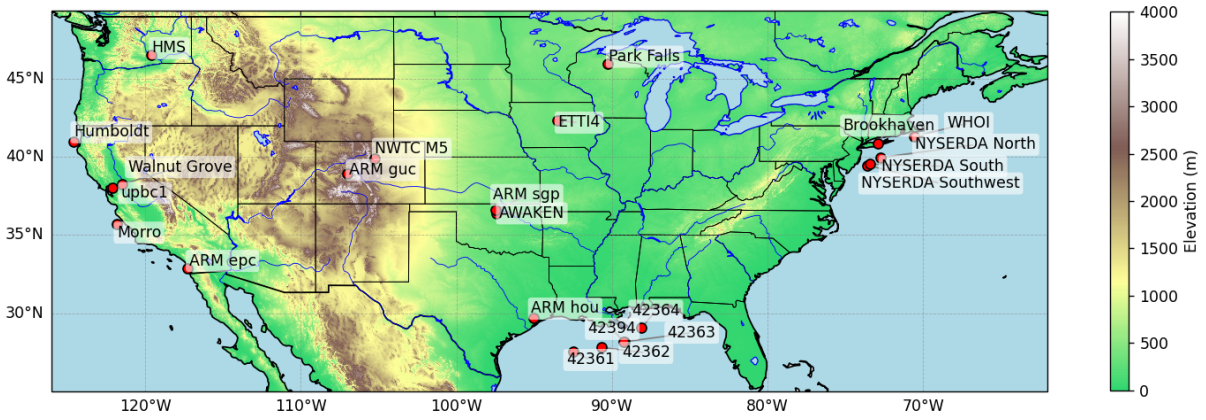


Figure 5. Map of the observational stations used for model validation. Terrain height data are derived from https://apps.nationalmap.gov/downloader and down-sampled to 1 km resolution.

on opposing booms at each height, we select the upwind sensor at each time step based on concurrent measurements, explicitly excluding tower-wake-affected observations.

(b) Filter outliers: values outside of 2 times the standard deviation within a rolling 6 h window, and wind speeds over certain station-specific thresholds, defined based on maximum wind speed occurrences at each site.

(c) For the raw lidar datasets (ARM SGP, GUC, and EPC): apply signal-to-noise ratio filter with a site-specific signal-to-noise ratio threshold (-19 dB for ARM SGP, -20 dB for GUC and EPC) that effectively filters out noise.

(d) Exclude suspicious data (e.g., negative wind speeds, repeated values, long near-zero runs).

(e) Fix time stamp, time zone and unit errors.

3. Extract model data (WTK-LED, WTK-LED Climate, NOW-23, HRRR MET Toolkit, and BC-HRRR) at the nearest grid point to each observational site.

4. Resample all data. We evaluate the WRF-based models using hourly instantaneous values, as some modeled datasets are restricted to this resolution. For the observations, however, we calculate hourly averages (centered on the model output time) to ensure consistency across sites, as certain datasets are provided solely as hourly means and preclude the derivation of instantaneous values.

5. Linearly interpolate model data at the heights of each observational site.

6. Keep all the time stamps at which the observations overlap with at least one modeled dataset.

Optis et al. (2020a) provide recommendations for metrics when comparing model vs. observational datasets. Following these recommendations, we calculate the modeled wind speed (U) bias at each site (and height) as

$$\text{bias} = \overline{U}_{\mathbf{model}} - \overline{U}_{\mathbf{obs}},$$

the centered root-mean-square error (cRMSE) as

$$\text{cRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(U_{\mathbf{model,i}} - \overline{U}_{\mathbf{model}}) - (U_{\mathbf{obs,i}} - \overline{U}_{\mathbf{obs}})\right]^2},$$

and the Pearson correlation coefficient r. To compare probability density functions (PDFs) of wind speed, we use the Wasserstein distance (or Earth mover's distance, EMD, Lupu et al. (2017) and Hahmann et al. (2020)). The Wasserstein distance provides a physically interpretable measure of dissimilarity between two probability distributions by quantifying the minimum "effort" required to transform one distribution into the other. It captures differences in PDF shape, spread, and tail behavior, making it well suited for evaluating model–observation discrepancies. In one dimension, the first-order Wasserstein distance between two PDFs $F(U)$ is defined as

$$W(F_1, F_2) = \int_0^1 \left|F_1^{-1}(U) - F_2^{-1}(U)\right| dU.$$

The Wasserstein distance is expressed in the same physical units as the distribution (e.g., $\mathrm{m\,s^{-1}}$ for wind speed), and a value of 0 indicates identical distributions.

5 Results of model validation

We first validate model performance at heights relevant to commercial wind turbines. For each station, we select the measurement height closest to 100 m (land-based) or 140 m (offshore) to serve as a proxy for hub height. Figure 6 summarizes validation metrics across all sites, and Table 3 provides the aggregate means. A clear performance hierarchy emerges: the HRRR-based products (HRRR MET Toolkit and BC-HRRR) and the offshore-specific NOW-23 consistently outperform the WTK-LED family. Averaged across all stations, HRRR MET Toolkit and BC-HRRR achieve Pearson correlation coefficients (r) of 0.82, significantly surpassing WTK-LED (0.67) and WTK-LED Climate (0.39). Similarly, the cRMSE is substantially lower for the HRRR family ($\approx$2.1 $\mathrm{m\,s^{-1}}$) compared to WTK-LED (3.3 $\mathrm{m\,s^{-1}}$) and WTK-LED Climate (4.4 $\mathrm{m\,s^{-1}}$). For offshore stations, NOW-23 demonstrates superior performance, achieving the lowest mean bias (0.02 $\mathrm{m\,s^{-1}}$) and highest correlation (0.87) of any dataset, validating its regional tuning strategy. The HRRR-based products also perform robustly in offshore environments, closely trailing NOW-23, whereas the WTK-LED variants show significant degradation. While correlation and cRMSE favor the HRRR family and NOW-23, wind speed bias exhibits more complex behavior. All models generally display a positive domain-averaged bias, but magnitudes vary. WTK-LED Climate exhibits the largest average positive bias (1.05 $\mathrm{m\,s^{-1}}$), followed by WTK-LED (0.82 $\mathrm{m\,s^{-1}}$). Notably, the bias correction applied to BC-HRRR shifts the results relative to the HRRR MET Toolkit. In the aggregate, HRRR MET Toolkit has a lower mean bias (0.10 $\mathrm{m\,s^{-1}}$) than BC-HRRR (0.36 $\mathrm{m\,s^{-1}}$); however, these global averages mask important regional distinctions detailed below. Also, we note that residual tower shadowing may systematically depress observed wind speeds, contributing to the apparent positive model biases documented at sites where the tower wake effect could not be accounted for. This effect would not be present at the lidar-based validation sites, which are not subject to tower wake contamination.

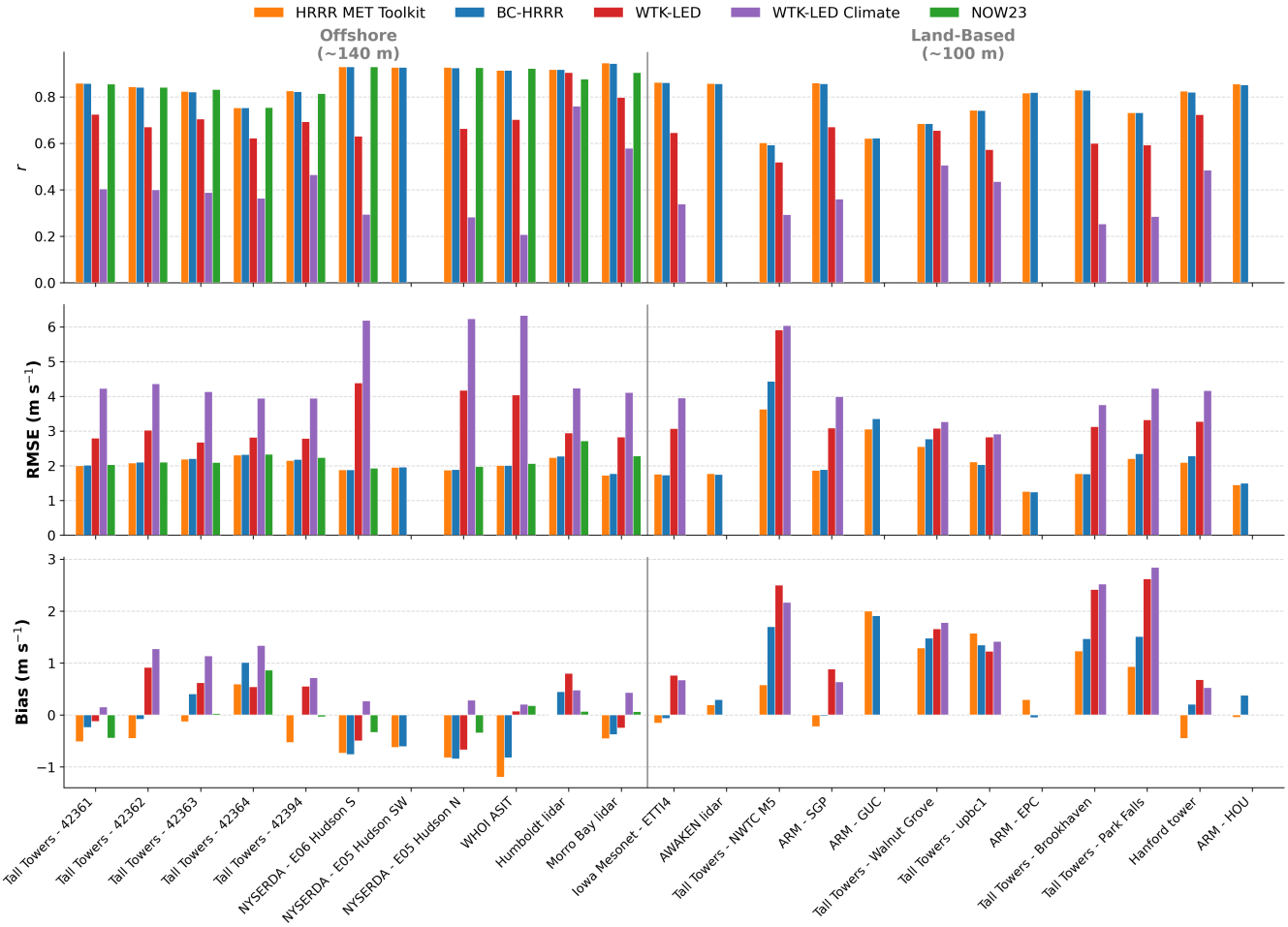


Figure 6. Hub-height validation metrics (bias, cRMSE, and r) for each station and model. Metrics are computed over the station-specific periods that overlap each model's availability within 2015–2025. The vertical dashed line separates offshore sites (left) from land-based stations (right).

To ensure a strict inter-comparison, we restrict the subsequent station-level analysis to overlapping temporal periods where all modeled datasets are available (2015, 2018, 2019, and 2020). Figures 7 and 8 show the PDFs of wind speeds. Onshore, HRRR MET Toolkit and BC-HRRR distributions align closely with observations, whereas WTK-LED products exhibit broader distributions and overpredict the frequency of high wind speeds. This distributional offset drives the higher positive biases in WTK-LED. Agreement is highest at sites with simple terrain (e.g., ARM SGP, Iowa Mesonet), whereas all models struggle to capture distributions in complex terrain (e.g., NWTC M5), highlighting the limitations of mesoscale modeling in resolving localized flow. Offshore (Fig. 8), all datasets better match observed distributions, with NOW-23 providing the best fit (in terms of bias and correlation), followed closely by the HRRR-based products (with HRRR MET Toolkit showing the low-

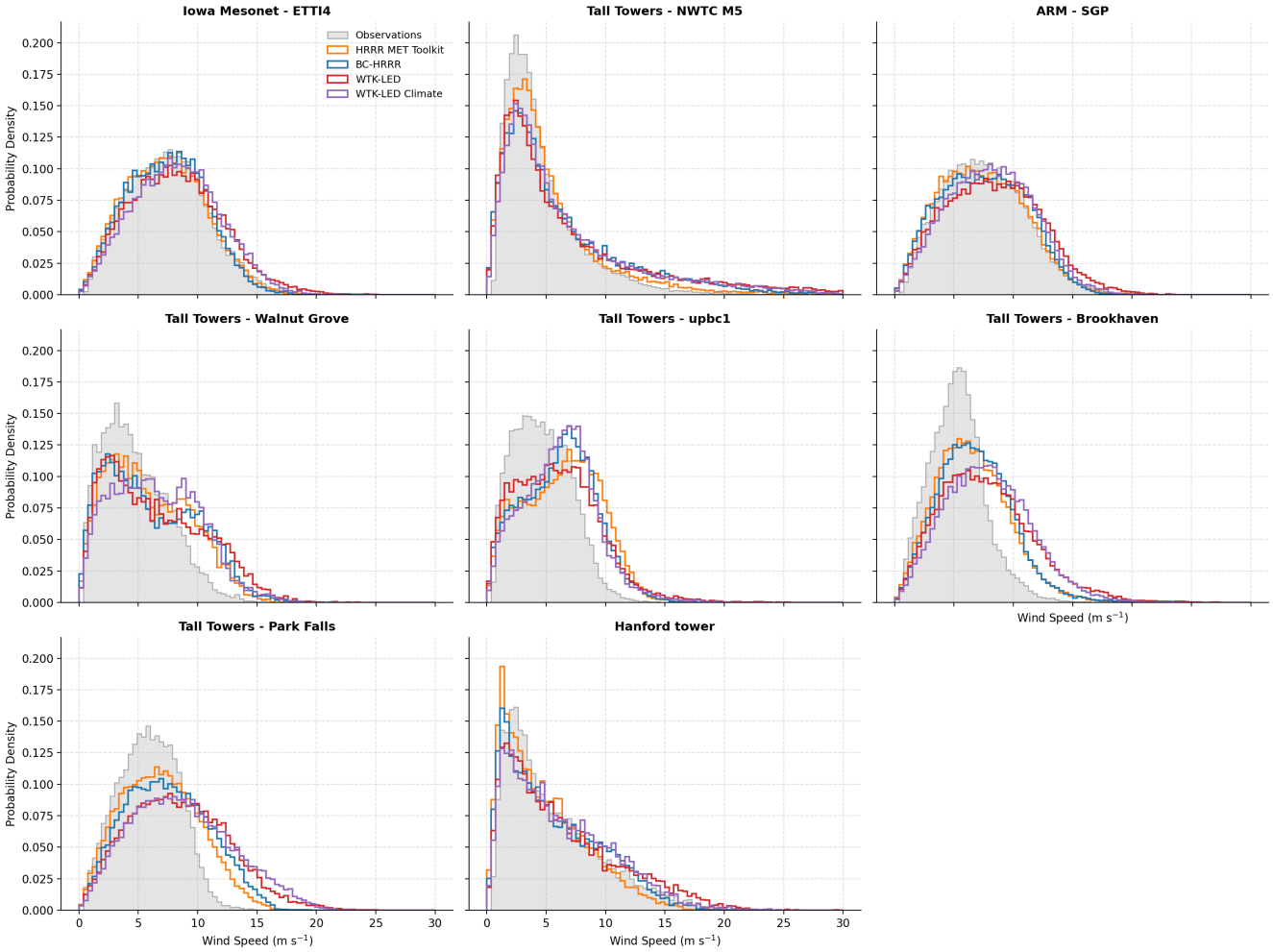


Figure 7. PDFs for land-based sites comparing hub-height (closest height to 100 m with observations) observed and modeled wind speeds over 2015, 2018, 2019, and 2020.

est cRMSE). Furthermore, the comparison reveals a generally better model–lidar agreement than model–tower agreement. This discrepancy is consistent with the presence of residual tower shadowing effects at certain tower locations. Aerodynamic wake contamination from the tower geometry can artificially decelerate the incoming flow, shifting true high-wind-speed observations into lower velocity bins. This structural artifact inflates the apparent low-wind frequency and truncates the high-wind tail in the tower-derived PDFs – a signature likely affecting stations such as upbc1, Brookhaven, and Park Falls. These distributional characteristics reinforce our earlier interpretation: a portion of the apparent positive model bias at tall tower locations likely reflects a localized measurement artifact rather than an unmitigated model deficiency alone.

Figure 9 presents a "scorecard" heatmap of validation metrics for the overlapping period, clarifying the regional impact of the bias correction in BC-HRRR. For offshore sites, the HRRR MET Toolkit exhibits an average negative bias (mean

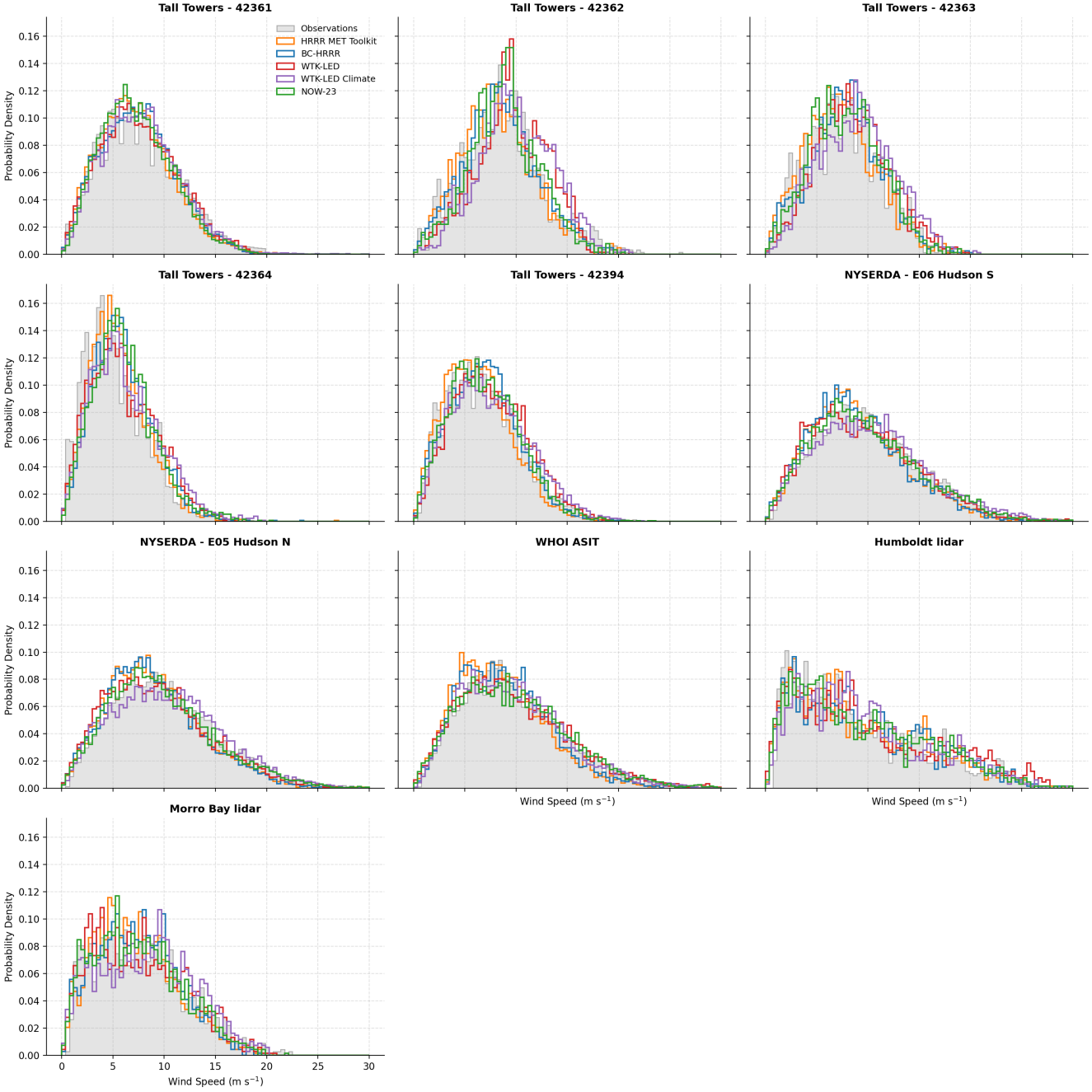


Figure 8. PDFs for offshore sites comparing hub-height (closest height to 140 m with observations) observed and modeled wind speeds over 2015, 2018, 2019, and 2020.

	HRRR MET Toolkit				BC-HRRR				WTK-LED				WTK-LED Climate				NOW-23			
	EMD	Bias	cRMSE	r	EMD	Bias	cRMSE	r	EMD	Bias	cRMSE	r	EMD	Bias	cRMSE	r	EMD	Bias	cRMSE	r
Tall Towers - 42361	0.45	-0.45	1.97	0.86	0.25	-0.18	1.98	0.86	0.12	-0.09	2.76	0.73	0.23	0.09	4.20	0.38	0.39	-0.33	2.04	0.85
Tall Towers - 42362	0.55	-0.54	2.38	0.80	0.23	-0.16	2.40	0.80	1.11	0.92	3.02	0.67	1.35	1.26	4.37	0.32	0.48	0.13	2.31	0.81
Tall Towers - 42363	0.33	-0.31	2.10	0.82	0.26	0.23	2.10	0.82	0.79	0.79	2.51	0.74	1.28	1.28	3.89	0.41	0.23	0.20	1.88	0.85
Tall Towers - 42364	0.34	0.26	2.08	0.75	0.70	0.67	2.11	0.75	0.74	0.71	2.90	0.56	0.97	0.96	3.75	0.31	0.75	0.74	2.20	0.73
Tall Towers - 42394	0.55	-0.55	2.13	0.81	0.25	-0.01	2.16	0.81	0.58	0.56	2.73	0.70	0.64	0.63	3.74	0.48	0.18	0.06	2.20	0.80
NYSERDA - E06 Hudson S	0.98	-0.98	1.87	0.93	1.01	-1.01	1.87	0.93	0.52	-0.50	4.38	0.63	0.37	0.27	6.19	0.29	0.39	-0.34	1.93	0.93
NYSERDA - E05 Hudson N	1.00	-0.99	1.88	0.93	1.03	-1.00	1.90	0.93	0.70	-0.67	4.17	0.66	0.42	0.29	6.23	0.28	0.41	-0.34	1.97	0.93
WHOI ASIT	1.16	-1.16	1.98	0.92	0.79	-0.79	1.98	0.92	0.35	0.07	4.03	0.70	0.15	-0.09	5.95	0.28	0.39	0.16	2.01	0.93
Humboldt lidar	0.22	0.05	2.32	0.93	0.57	0.46	2.37	0.93	0.91	0.80	2.94	0.90	0.52	0.48	4.23	0.76	0.46	0.36	2.84	0.90
Morro Bay lidar	0.69	-0.64	2.05	0.89	0.55	-0.47	2.08	0.88	0.29	-0.25	2.82	0.80	0.55	0.43	4.11	0.58	0.44	-0.44	2.51	0.83
Iowa Mesonet - ETTI4	0.13	-0.13	1.73	0.87	0.11	-0.04	1.71	0.87	0.77	0.76	3.07	0.65	0.79	0.79	3.98	0.34				
Tall Towers - NWTC M5	0.64	0.56	3.61	0.58	1.73	1.65	4.40	0.57	2.43	2.35	5.78	0.49	2.06	2.02	5.84	0.28				
ARM - SGP	0.23	-0.23	1.85	0.86	0.24	-0.02	1.87	0.86	0.82	0.79	3.08	0.66	0.56	0.53	3.98	0.36				
Tall Towers - Walnut Grove	1.18	1.17	2.49	0.70	1.35	1.34	2.70	0.70	1.68	1.68	3.18	0.62	1.79	1.78	3.21	0.51				
Tall Towers - upbc1	1.66	1.64	2.09	0.75	1.43	1.41	2.00	0.75	1.29	1.25	2.83	0.58	1.48	1.47	2.88	0.44				
Tall Towers - Brookhaven	1.28	1.28	1.79	0.82	1.52	1.51	1.79	0.82	2.42	2.42	3.11	0.60	2.57	2.57	3.73	0.26				
Tall Towers - Park Falls	1.00	1.00	2.27	0.72	1.58	1.58	2.40	0.72	2.55	2.54	3.32	0.59	2.83	2.82	4.23	0.29				
Hanford tower	0.77	-0.77	2.41	0.80	0.32	-0.05	2.52	0.80	0.76	0.68	3.27	0.72	0.59	0.53	4.16	0.48				
Mean (Offshore)	**0.63**	**-0.53**	**2.08**	**0.86**	**0.56**	**-0.23**	**2.09**	**0.86**	**0.61**	**0.23**	**3.23**	**0.71**	**0.65**	**0.56**	**4.67**	**0.41**	**0.41**	**0.02**	**2.19**	**0.86**
Mean (Land-based)	**0.86**	**0.57**	**2.28**	**0.76**	**1.03**	**0.92**	**2.43**	**0.76**	**1.59**	**1.56**	**3.46**	**0.61**	**1.58**	**1.56**	**4.00**	**0.37**				
Mean (All)	**0.73**	**-0.04**	**2.17**	**0.82**	**0.77**	**0.28**	**2.24**	**0.82**	**1.05**	**0.82**	**3.33**	**0.67**	**1.06**	**1.01**	**4.37**	**0.39**	**0.41**	**0.02**	**2.19**	**0.86**



Figure 9. Scorecard for metrics per station and average metrics (across 2015, 2018, 2019, and 2020).

375 $-0.53\ \mathrm{m\,s^{-1}}$). The bias correction successfully mitigates this underspeeding, improving the BC-HRRR offshore mean bias to $-0.23\ \mathrm{m\,s^{-1}}$ (closer to the NOW-23 benchmark of $0.02\ \mathrm{m\,s^{-1}}$). Conversely, at land-based sites where the HRRR MET Toolkit already possesses a positive mean bias ($0.57\ \mathrm{m\,s^{-1}}$), the bias correction – which anchors the distribution to the legacy WIND Toolkit – increases the positive bias to $0.92\ \mathrm{m\,s^{-1}}$. This suggests that ~~while quantile mapping effectively corrects offshore underspeeding, it may exacerbate positive biases onshore where local terrain effects dominate~~the onshore positive bias in
380 BC-HRRR relative to HRRR MET Toolkit reflects the spatial heterogeneity of WIND Toolkit biases: the quantile mapping anchors the HRRR distribution to a reference that is itself positively biased relative to observations in complex terrain. In offshore environments, where the WIND Toolkit is better calibrated, the same correction reduces the negative bias present in the native HRRR. Despite this shift, BC-HRRR maintains the high correlation and low cRMSE of the native HRRR MET Toolkit, offering a significant improvement over the WTK-LED family (mean land-based cRMSE: $3.46\ \mathrm{m\,s^{-1}}$).

The spatial distribution of validation metrics (Figs. 10–12) reveals coherent regional structures aligned with terrain complexity. Bias maps (Fig. 10) indicate that offshore and flat inland sites generally exhibit near-zero bias across most models, reflecting reduced terrain forcing and homogeneous flow. In contrast, complex terrain and coastal transition zones pose challenges, particularly for WTK-LED and WTK-LED Climate, which display large positive biases consistent with unresolved subgrid variability. The HRRR-based models demonstrate smaller biases overall, though the land-sea contrast remains evident. Correlation (Fig. 11) and cRMSE (Fig. 12) degrade markedly in regions of complex terrain, particularly throughout the western United States. This pattern is consistent across all models, suggesting that unresolved terrain-driven variability may impose a fundamental limit on the agreement achievable between mesoscale models and point observations.

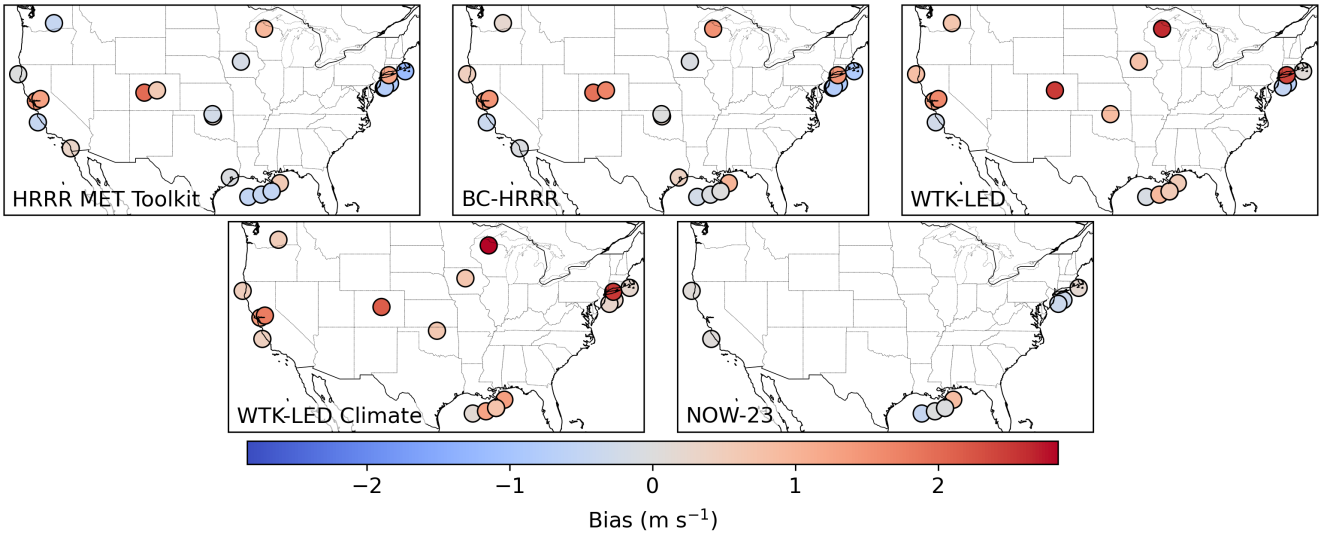


Figure 10. Maps showing hub-height wind speed bias over the full period of record of each modeled dataset.

The dependence of model performance on terrain complexity is more systematically quantified in Fig. 13, which plots metrics against the standard deviation of terrain height within a 20 km radius of each station. We observe a log-linear degradation in performance for all datasets as terrain complexity increases: correlation decreases and cRMSE increases, consistent with known limitations of WRF in resolving complex flow (e.g., Carvalho et al. (2012)). While all models suffer in rugged environments, the HRRR-based products maintain a performance advantage over WTK-LED across the complexity spectrum. Notably, the bias for the WTK-LED family remains high regardless of terrain, whereas the HRRR family shows lower biases that scale with complexity.

A regime-dependent evaluation of model behavior, binned by observed wind speed (Fig. 14), reveals a pronounced dependence of bias on wind speed magnitude. Across all models, positive biases are largest at low wind speeds and generally decrease as wind speeds increase. Several datasets transition toward negative bias in the highest bins, indicating a tendency to overpredict light winds and underpredict strong winds. The WTK-LED Climate dataset exhibits the largest biases throughout the distribution and the most distinct shift toward negative bias at high wind speeds, whereas the HRRR-based products and NOW-23

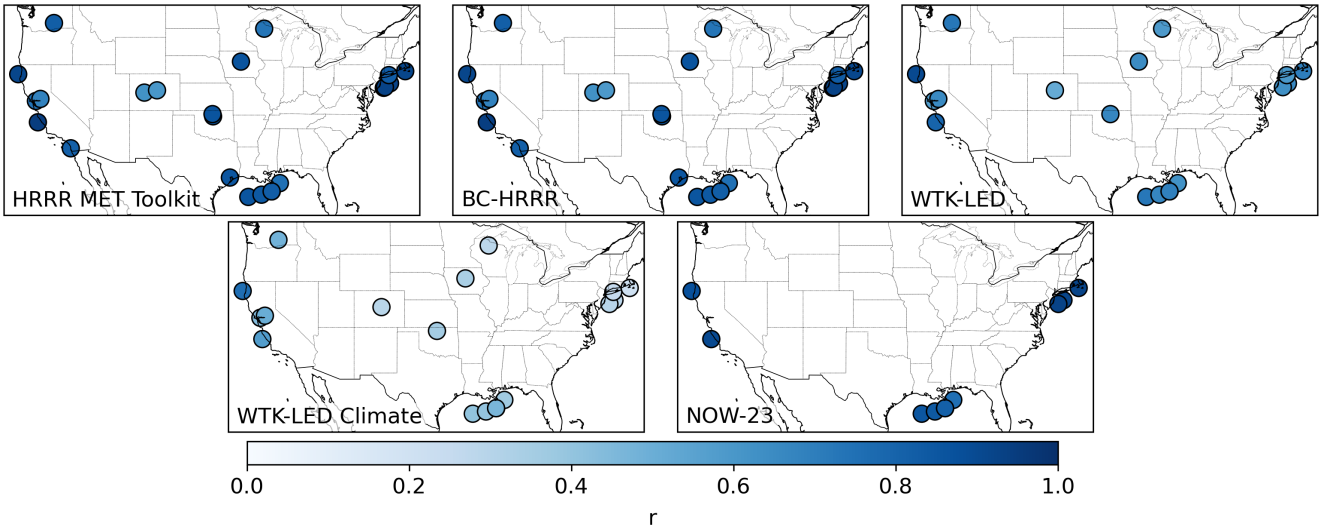


Figure 11. Maps showing hub-height wind speed correlation coefficient over the full period of record of each modeled dataset.

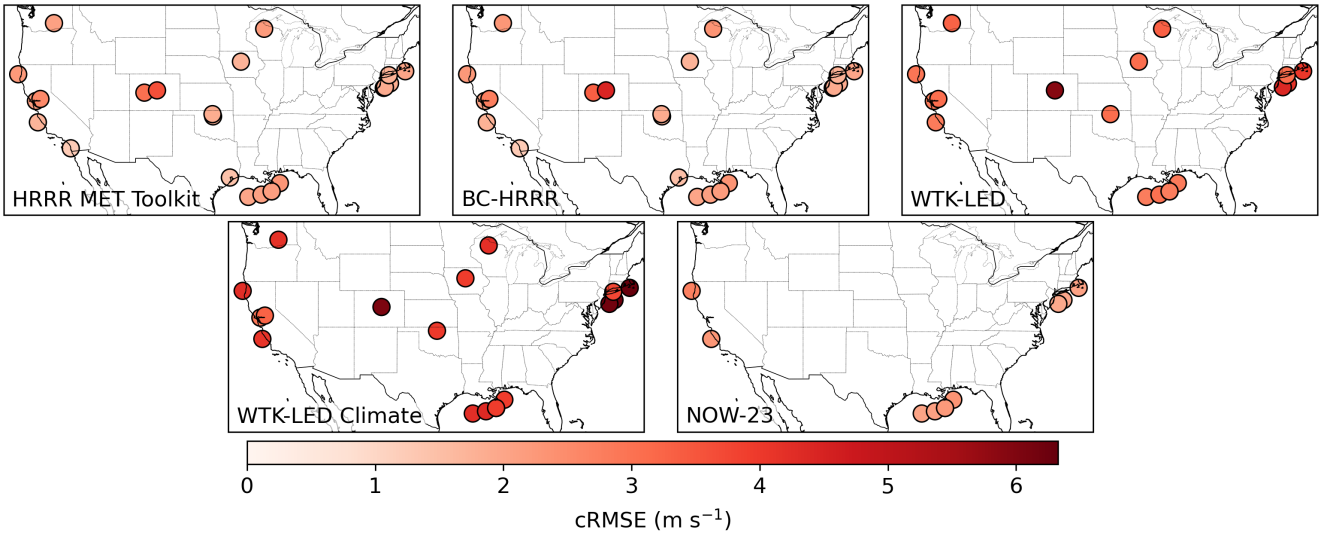


Figure 12. Maps showing hub-height wind speed cRMSE over the full period of record of each modeled dataset.

maintain smaller biases across most regimes. Notably, BC-HRRR does not show larger biases than HRRR MET Toolkit for either very low or very high wind speeds, when the effects of the bias correction might mute extremes. Additionally, cRMSE increases monotonically with wind speed for all datasets, reflecting the expected growth in absolute error magnitude as wind-speed variability increases. We omit Pearson's correlation coefficients from this binned analysis, as restricting observations to narrow ranges artificially reduces variance, rendering r a measure of scatter rather than true model skill.

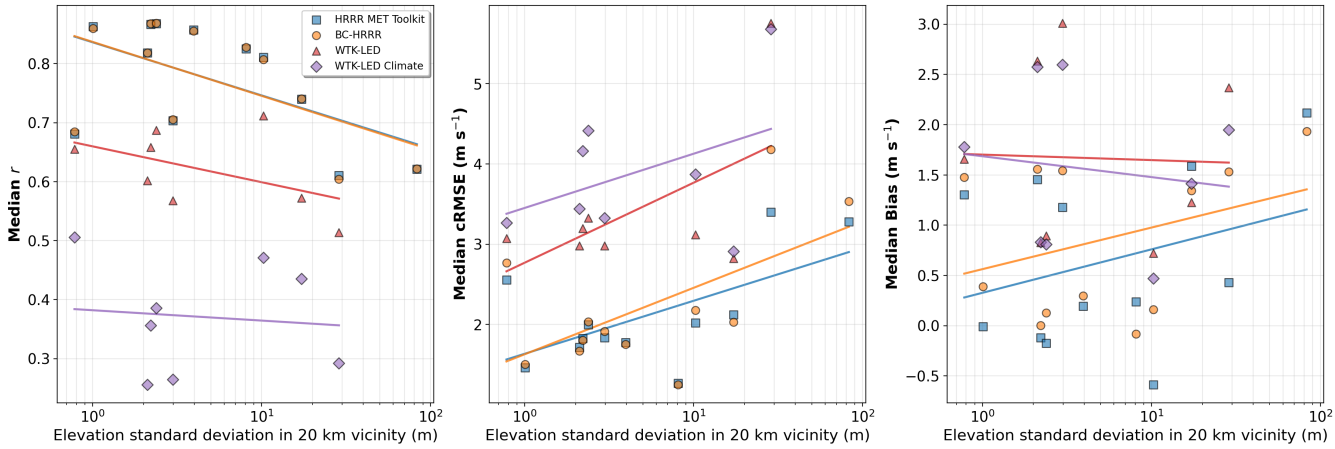


Figure 13. Correlation r (left), cRMSE (center), and bias (right) depending on terrain complexity (standard deviation of terrain height within a 20 km radius). Metric values are medians across all heights per measurement station. NOW-23 data and offshore stations are not shown because terrain complexity does not impact offshore sites. Note that the correlation statistics between the HRRR MET Toolkit and BC-HRRR in the left panel are nearly identical and thus the data overlap.

Figure 15 illustrates the diurnal variation of hub-height validation metrics. The left column displays metrics averaged across all observational stations, with shaded regions denoting the interquartile range (IQR; 25th–75th percentile) at each local hour. The IQR characterizes site-to-site variability robustly, minimizing sensitivity to outliers – a critical feature given the diversity of terrains and flow regimes in the CONUS-wide station network. Across the diurnal cycle, the relative performance hierarchy remains consistent with the aggregated results. WTK-LED Climate exhibits the largest biases and cRMSE and the lowest correlations at all hours. In contrast, the HRRR-based products and NOW-23 show substantially improved performance, with NOW-23 behaving comparably to the HRRR family, reflecting its strong offshore validation. The most distinct diurnal structure appears in the bias. For all datasets, positive biases peak during the night and early morning, reaching a minimum between late morning and afternoon (approximately 09:00–16:00 local time). This behavior aligns with the daily cycle of boundary layer stability. During daytime convective conditions, enhanced turbulent mixing deepens the boundary layer and reduces vertical wind shear, improving the coupling between hub-height winds, surface forcing, and the larger-scale flow. Under these conditions, mesoscale models generally represent mean hub-height winds more accurately. Conversely, stable nighttime conditions suppress turbulence and increase sensitivity to surface-layer and planetary boundary layer parameterizations. This regime often promotes complex phenomena – such as nocturnal low-level jets, drainage flows, and terrain-induced decoupling – that are known to exacerbate model errors (Deppe et al., 2013; Bianco et al., 2019; Olson et al., 2019). The diurnal trend in cRMSE, which shows higher values at night and a minimum during the day, is consistent with this stability-driven contrast in predictability. To determine whether this aggregate diurnal signal is uniform across the network or driven by specific locations, the right column of Fig. 15 presents station-resolved diurnal cycles for the HRRR MET Toolkit. Stations are ranked by diurnal amplitude (defined as the peak-to-trough range of the metric over 24 hours). The eight stations with the largest amplitudes are

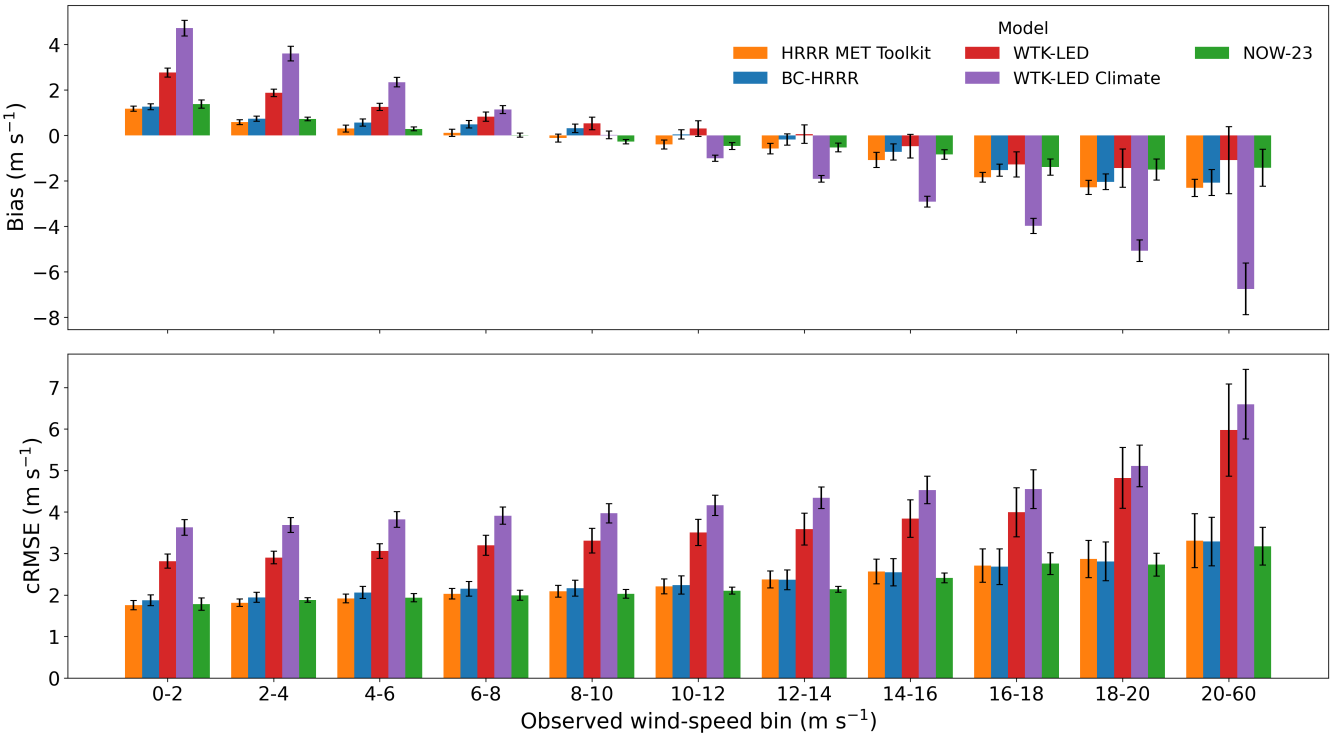


Figure 14. Hub-height validation metrics as a function of observed wind speed, averaged across all stations: (top) bias and (bottom) cRMSE.

highlighted in color, while the remaining stations appear in gray. The multi-station median (black line) shows a much flatter profile than the high-amplitude stations, indicating that pronounced diurnal variability is not universal but rather concentrated at a subset of sites. These high-amplitude stations likely correspond to environments with strong stability transitions or locally forced circulations – such as complex terrain or coastal zones – consistent with the terrain-dependent degradation identified in Fig. 13. These results confirm that while diurnal performance variability is a robust feature at the continental scale, its magnitude is strongly site-dependent and linked to local flow complexity.

Seasonally (Fig. 16), the relative performance hierarchy remains consistent: HRRR-based products and NOW-23 outperform the WTK-LED family across all months. Among metrics, correlation displays the clearest seasonal dependence, with lower values in summer and peaks in winter/spring. This likely reflects enhanced summer convective activity and atmospheric instability, which drive small-scale intermittency that mesoscale models struggle to resolve at the point scale (Bianco et al., 2019; Liu et al., 2025). Consequently, temporal coherence degrades during the warm season even where mean errors remain modest. In contrast, bias and cRMSE show weaker seasonal structure when averaged across the domain. The substantial width of the IQR bands suggests that systematic offsets are driven less by seasonality and more by site-specific factors – such as terrain complexity and surface roughness – consistent with prior findings that terrain exerts a stronger control on hub-height performance than season alone (Carvalho et al., 2012).

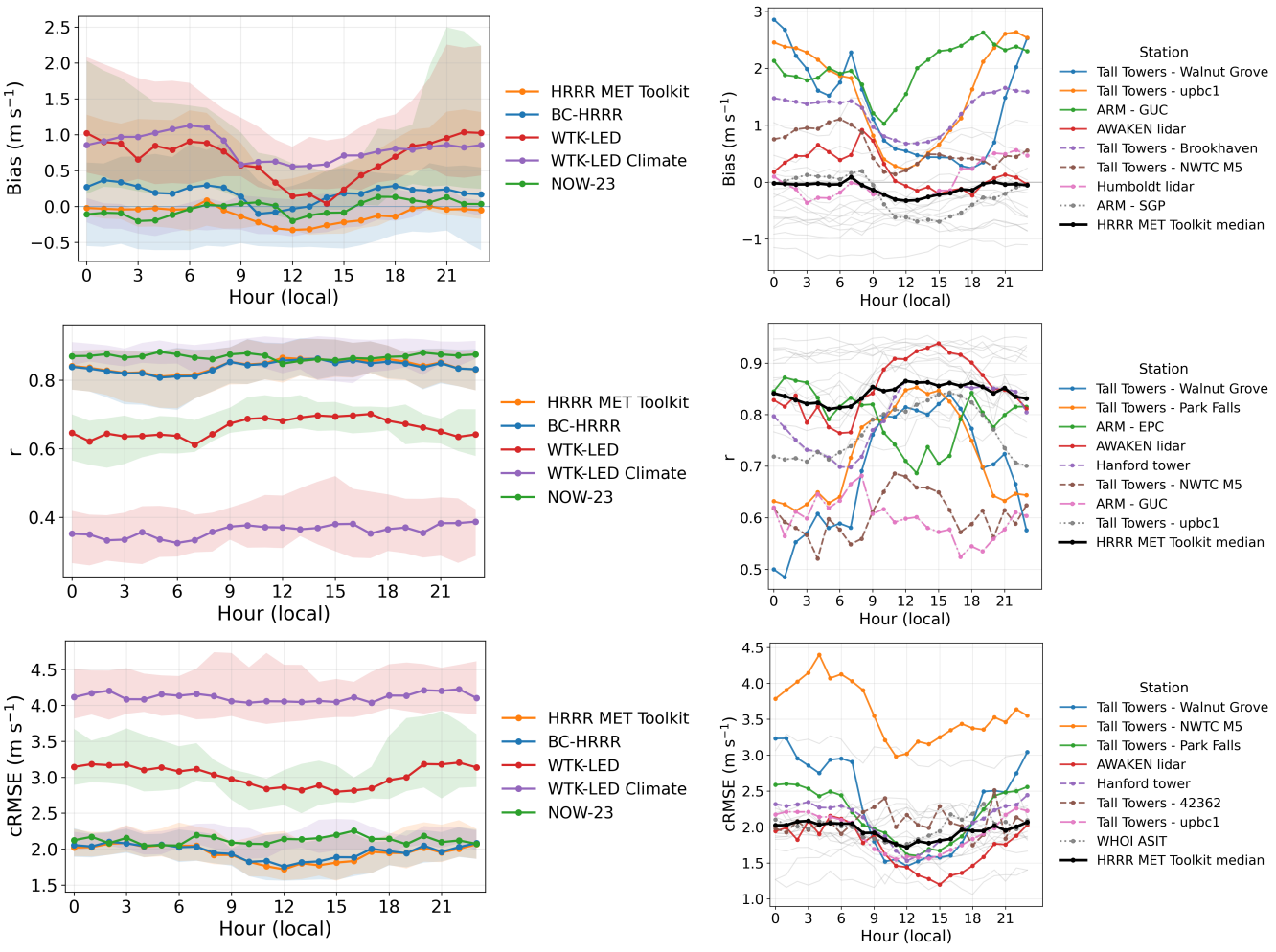


Figure 15. Diurnal variation of hub-height bias, r, and cRMSE. The left column shows station-averaged metrics with interquartile range (IQR) shading. The right column details station-level diurnal cycles for the HRRR MET Toolkit, highlighting the eight stations with the largest diurnal amplitude (colored lines) against the remaining stations (gray lines) and the network median (black line).

Figure 17 illustrates validation metrics calculated against observation sites that provide multilevel measurements. A consistent trend is observed where cRMSE increases with height for all models. This degradation is likely driven by the higher mean wind speeds experienced at greater altitudes, where absolute errors naturally scale with the magnitude of the wind vector. Correlation and bias show more variability between stations, without a singular dominant vertical trend. However, the general coherence of HRRR MET Toolkit with observations across these levels confirms that the vertical interpolation strategy effectively preserves the shear signal inherent in the native HRRR model data, making it suitable for analyzing wind resource across the entire rotor-swept area.

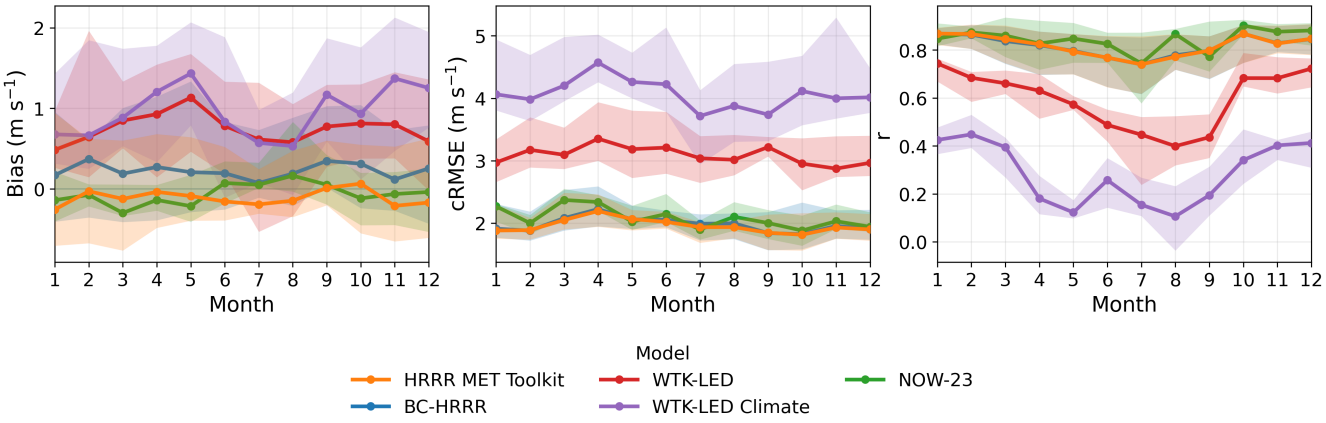


Figure 16. Monthly variation of hub-height validation metrics. Lines show metrics averaged across all observational stations, while shaded regions indicate the interquartile range (25th–75th percentile) across stations for each month.

Finally, as an operational weather model, the HRRR has undergone several major upgrades since 2014 (Table 1). A critical question for long-term resource assessment is whether these version changes introduce artificial steps or trends in the dataset's accuracy. Figure 18 displays the temporal evolution of weekly validation metrics for the HRRR MET Toolkit, averaged across the 16 stations that span multiple HRRR versions. The analysis reveals a positive trend in model performance over time, corresponding to improvements in the underlying HRRR model physics and data assimilation systems. From HRRRv1 to HRRRv4, the station-averaged correlation improves from approximately 0.67 to 0.8, cRMSE decreases from 2.2 m s^{-1} to 1.8 m s^{-1}, and mean bias drops significantly from 0.7 m s^{-1} to 0.1 m s^{-1}. These results align with findings from Bianco et al. (2025), who reported improved ramp forecasting skill and reduced error in HRRRv4 compared to v3, as well as the results shown in James et al. (2022) who performed a similar analysis using wind observations at 10 m. To ensure these trends reflect model improvements rather than changes in the observational network quality, we performed an identical analysis on the WTK-LED dataset (which uses a frozen model configuration) and observed no such temporal trends. This confirms that HRRR MET Toolkit (and BC-HRRR) benefits from the continuous evolution of the operational HRRR system, providing a dataset that becomes increasingly accurate in the recent record. It is worth noting that at least some of these improvements – particularly the transition to HRRRv3 and v4 – were facilitated by the collaborative DOE-NOAA-private sector field campaigns under the Wind Forecast Improvement Projects (WFIP and WFIP2), which identified and helped correct systematic HRRR model errors (Bianco et al., 2019; Olson et al., 2019). This demonstrates the value of sustained investment in targeted field campaigns for operational model improvement. A related open question concerns the degree to which the HRRR's performance advantage stems from its advanced data assimilation system versus improvements in model physics. One approach to disentangle these contributions would be to evaluate HRRR skill at longer forecast horizons (e.g., hour 18 or 48), where the memory of assimilated observations decays and performance differences would reflect physical parameterization quality more directly. This is an important avenue for future research.

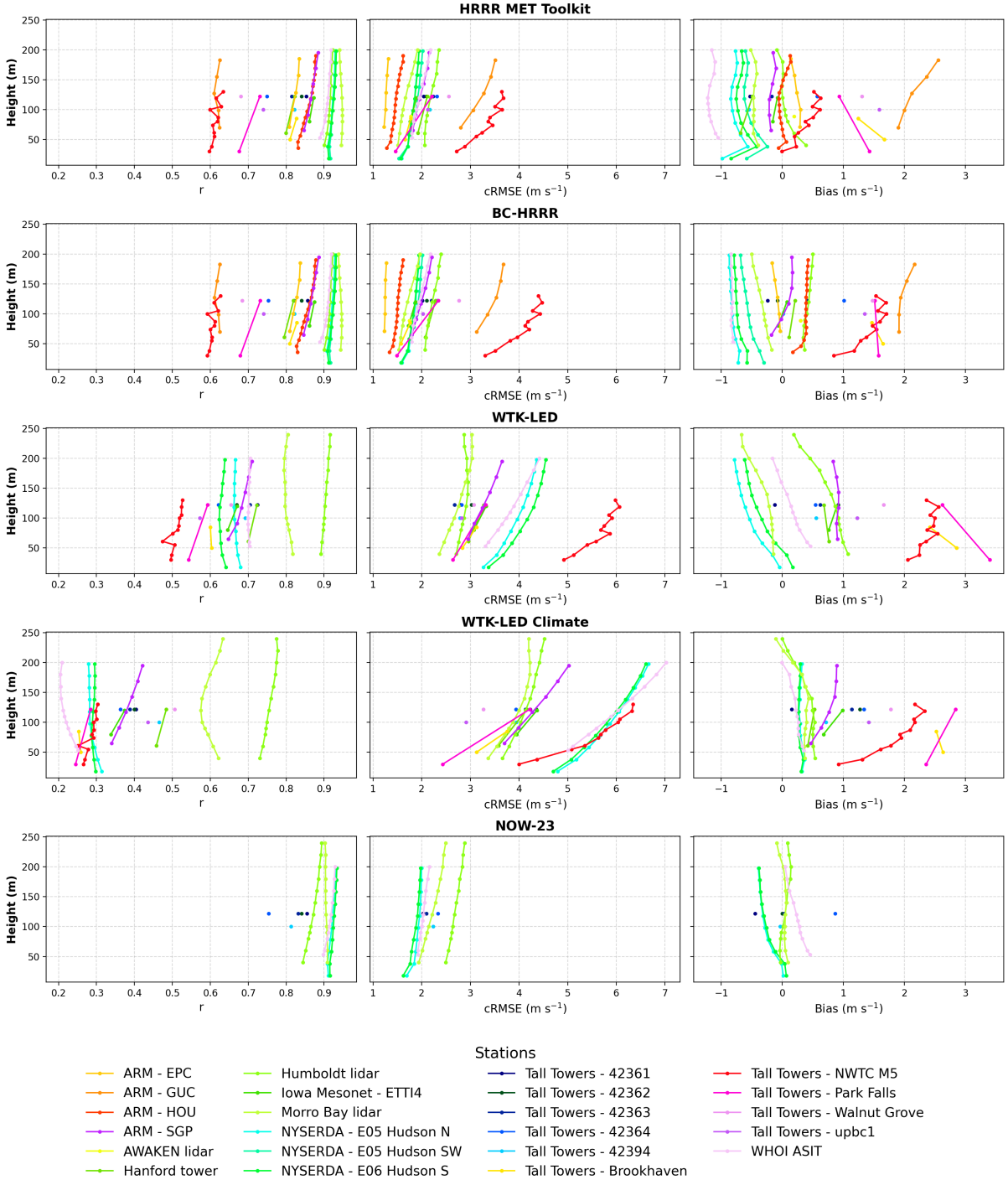


Figure 17. Validation metrics as a function of observational heights for all models and observation stations.

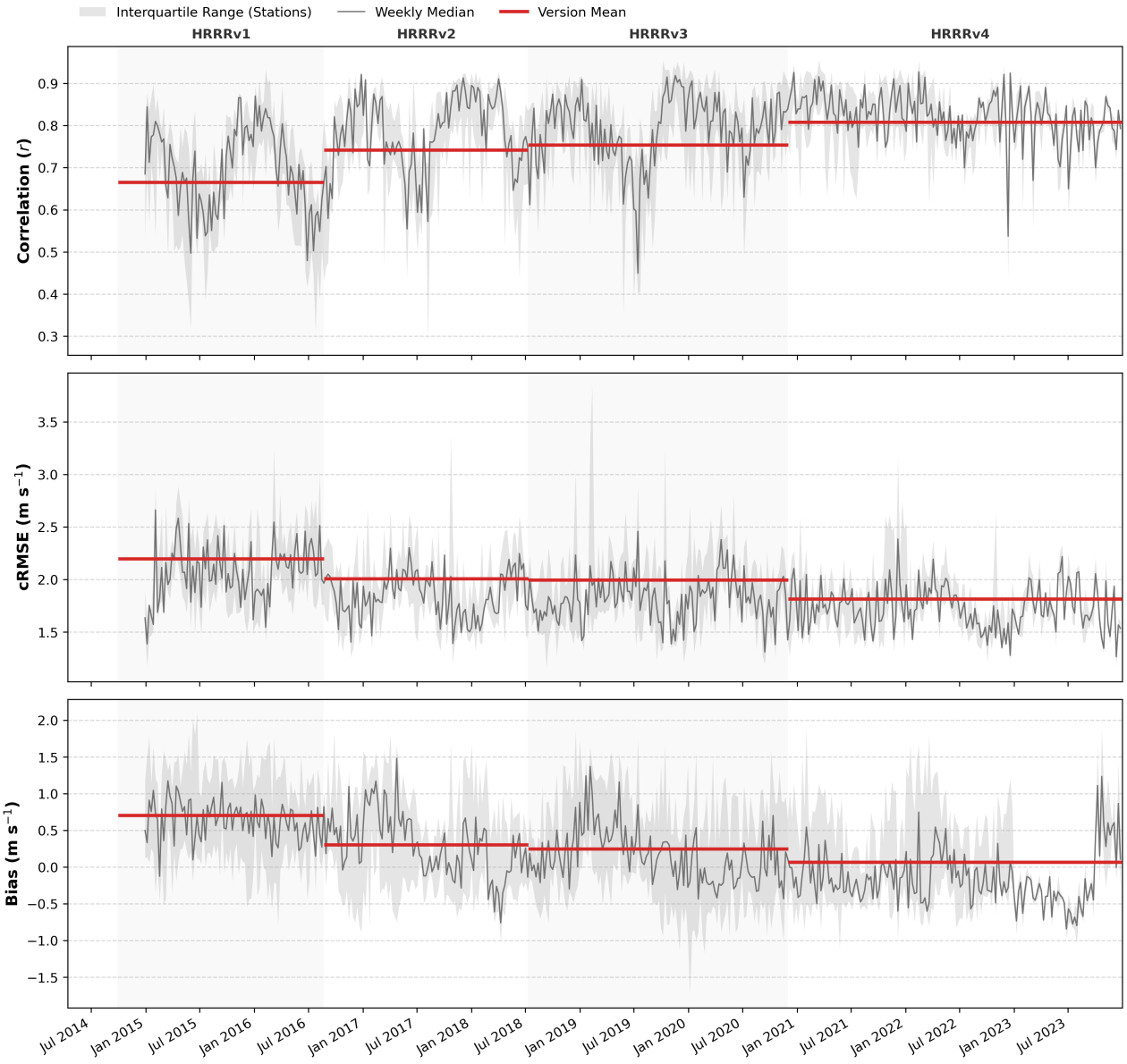


Figure 18. Temporal evolution of HRRR MET Toolkit validation metrics from 2015 to 2023 (as only one observational site has observations after early 2024), categorized by the underlying operational HRRR model version. The panels display the weekly (a) Pearson's correlation coefficient, (b) cRMSE, and (c) bias. The solid dark gray line represents the weekly median across all validation stations, while the light gray shading denotes the interquartile range (25th–75th percentile), illustrating variability across sites. Thick red horizontal lines indicate the mean performance for each distinct HRRR version period (delineated by alternating background shading).

6 Discussion and Conclusions

In this study, we introduced the HRRR MET (Meteorology, Energy, and Transmission) Toolkit and performed the first comprehensive, multi-model validation of wind resource datasets across CONUS at utility-scale hub heights. By repackaging the operational HRRR archive into a user-friendly format and anchoring it to the same uniform 2 km horizontal grid as the legacy WIND Toolkit, the HRRR MET Toolkit provides a seamless, high-resolution wind resource record spanning 2015–2025. This bridges the critical gap between historical industry-standard products and modern operational forecasts.

Our validation against a diverse set of tall meteorological towers and lidars yields several key findings. First, the HRRR-based datasets (the HRRR MET Toolkit and its experimental BC-HRRR variant) consistently outperform the computationally intensive WTK-LED and WTK-LED Climate datasets. Across all stations, the HRRR MET Toolkit achieves a significantly higher correlation (0.82 vs. 0.67) and lower average hub-height bias (0.10 vs. 0.82 m s^{-1}) than WTK-LED. This suggests that leveraging continuously updated operational weather forecast models, which benefit from advanced data assimilation, is a more effective and efficient strategy for maintaining current historical records than running static WRF hindcasts.

Second, our results demonstrate that the experimental bias correction applied to BC-HRRR does not universally improve the accuracy of the native HRRR model. While the correction effectively mitigates some negative biases in offshore environments, it provides marginal benefit in land-based regions and can actually exacerbate positive biases in mountainous or topographically complex terrain. These findings indicate that the native HRRR physics are already robust for energy applications without the added complexity of statistical post-processing. We conclude that a uniform, country-wide bias correction against legacy datasets is not a universal solution, especially when it introduces non-stationarity risks in extreme event analysis.

It is also important to acknowledge several potential sources of uncertainty inherent to the validation process itself. First, the observational "ground truth" is subject to measurement errors specific to the instrumentation and inconsistencies in quality control processes and site standards across the diverse observational networks used. Second, methodological uncertainties arise from comparing grid-based models to point measurements; this includes spatial representation errors inherent in mapping a 2 km grid cell to a specific instrumentation location, as well as vertical interpolation uncertainties stemming from the linear interpolation of coarser model levels to specific sensor heights. Third, despite rigorous filtering, some validation sites may still be impacted by microscale surroundings or complex local flow features that are unsuitable for mesoscale model validation. Finally, the results are conditioned by the geographic representativeness of the available validation locations and the partly varying validation time periods across datasets. While a different constellation of validation sites might yield slightly different error statistics, the broad spatial distribution of the current sites suggests that the overall conclusions regarding the relative performance of the datasets remain robust.

Ultimately, the HRRR MET Toolkit represents a marked improvement in data accessibility and usability over the native HRRR archive. The native data are distributed in complex GRIB formats on irregular hybrid vertical levels that vary in height over time and space, creating a significant barrier to entry for energy analysts. The HRRR MET Toolkit removes these barriers by providing a regularized, user-friendly dataset interpolated to standard wind-energy-specific heights. Furthermore, the toolkit's accuracy naturally improves over time as it tracks updates to the underlying NOAA operational model. The HRRR

MET Toolkit is publicly available through the NLR Open Energy Data Initiative (OEDI) and the Wind Resource Database (WRDB). Users should remain aware of the specific regional nuances and limitations summarized in Table 4.

Data availability. The post-processed modeled data distributed by NLR (WIND Toolkit, WTK-LED, WTK-LED Climate, NOW-23, HRRR MET Toolkit, and BC-HRRR) are stored in Hierarchical Data Format version 5 (HDF5) on NLR's Kestrel high-performance computing system. To support broad dissemination, the datasets are also hosted on the Amazon Web Services (AWS) Simple Storage Service (S3) as part of the Open Energy Data Initiative (OEDI). The HDF5 files contain self-describing metadata – including variable names, units, coordinate systems, and spatial/temporal resolutions – ensuring the data are FAIR (Findable, Accessible, Interoperable, and Reusable). Users can access the data through two primary channels:

1. Wind Resource Database (WRDB): For general users, the datasets are integrated into the NLR Wind Resource Database (National Laboratory of the Rockies, 2024). This web-based platform allows users to visualize layers, select specific locations or regions (e.g., U.S. counties), filter by year and variable, and download the resulting subsets in common formats (currently .csv).
2. Direct Cloud/API Access: For bulk access or programmatic analysis, the data are served via the HDF Group's Highly Scalable Data System (HSDS). This architecture allows users to slice and extract specific subsets of the multi-terabyte dataset without downloading full files. A repository of example Python scripts is provided to demonstrate how to connect to the HSDS service, query specific spatiotemporal subsets, and process the data. Detailed documentation and these resources can be accessed at https://developer.nlr.gov/docs/wind/wind-toolkit/.

For further assistance, a dedicated user support team is available at wrdb@nlr.gov.

Native HRRR data can be downloaded from NOAA's webpage at https://rapidrefresh.noaa.gov/hrrr/.

Observations used for model validation are stored (in a common .nc format) at https://doi.org/10.5281/zenodo.18177810 (Bodini et al., 2026).

Appendix A: Technical specifications of modeled datasets

Technical specifications of modeled datasets are reported in Table A1.

Appendix B: Analysis of representativeness of WIND Toolkit period for bias correction

The bias correction method used to produce BC-HRRR assumes stationarity in the wind speed distribution. A consequence of this approach is that high-wind extreme events in the BC-HRRR are constrained by the WIND Toolkit record, potentially masking long-term climatic trends or extreme events unique to the recent period. Future iterations of this dataset may employ Quantile Delta Mapping (Cannon et al., 2015) to better preserve these non-stationary signals.

It is important to assess whether the WIND Toolkit period (2007–2013) is representative of the long-term conditions over CONUS and whether the BC-HRRR period and the WIND Toolkit period share a similar wind climatology, given that the WIND Toolkit is used as the basis for bias correction. We use ERA5 reanalysis data (Hersbach et al., 2020) for this assessment,

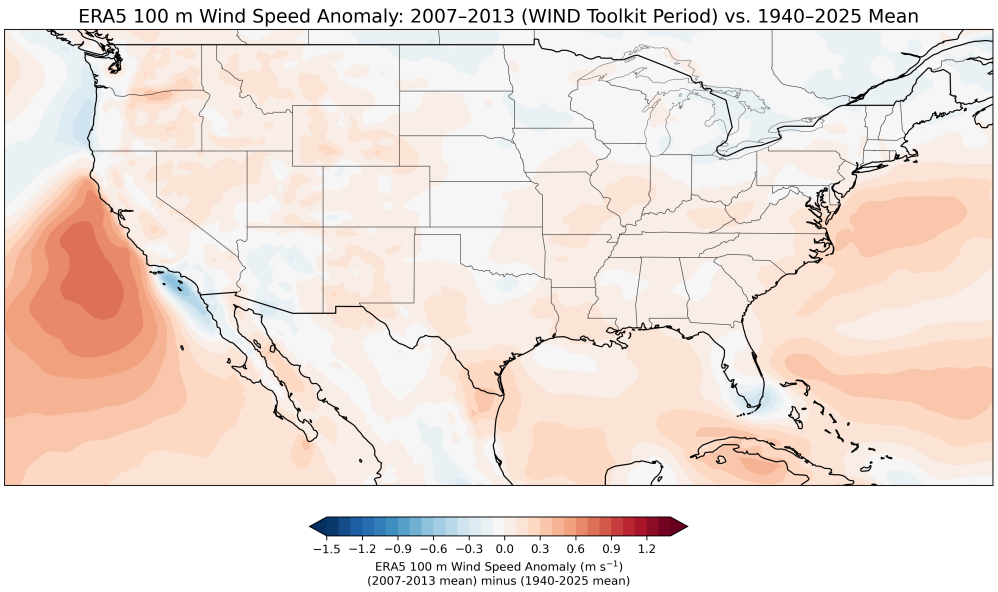
as it offers a consistent dataset for evaluation over the long term, though we note that ERA5 itself possesses a known overall negative bias across much of the contiguous United States (Wilczak et al., 2024).

Fig. B1 illustrates the spatiotemporal variability of ERA5 ~~10~~ 100 m wind speed changes over time. ~~Because monthly averaged wind speed data from ERA5 are readily available at 10 m, we utilize this height for the initial comparison.~~ Overall, the wind resource during the WIND Toolkit period is representative of long-term wind conditions (~~1991–2020~~1940–2025) over CONUS, especially onshore. ~~Variability~~ Temporal variability is limited to less than 0.1 m s^{-1} in most areas, with the exception of ~~limited offshore regions . The~~ some offshore regions (especially off the Pacific coast). The ERA5 data also indicates that the shift in wind resource magnitude between the WIND Toolkit period and the BC-HRRR period ~~in ERA5 shows~~ (2015–2023) exhibits distinct spatial variability across the contiguous United States. While the eastern and northwestern US experienced slightly lower wind speeds ~~than the WIND Toolkit period. Because the~~ during the recent BC-HRRR period, much of the central wind corridor and the southwest saw slight increases. Because our bias-correction process anchors the 2015–2023 HRRR data to the ~~slightly windier~~ 2007–2013 ~~distribution, a small positive bias in~~ WIND Toolkit distribution, this temporal mismatch means the resulting BC-HRRR ~~wind speeds relative to observations may be expected; however, differences are minimal~~ product may exhibit minor, regionally dependent biases relative to actual ground observations taken during the 2015–2023 period (e.g., a slight positive bias in the east and northwest, and a slight negative bias in the central US). However, the magnitude of these temporal differences remains generally minimal across the onshore domain (average < 0.2 m s^{-1}), with ~~some larger differences noted off the west coast~~the largest anomalies restricted primarily to offshore regions.

To further confirm this representativeness at hub height across the domain, we extracted 100 m hourly wind speed data from ERA5 at ~~selected sites~~ a geographically distributed set of locations covering the years 2000–2024. ~~As shown in Fig. B2, these locations cover all of~~ Because the observational sites used in this study are unevenly distributed, the specific locations used here were selected to ensure a robust spatial sampling across CONUS, including both onshore and offshore ~~sites.~~ location (Fig. B2). The distribution of temporal differences confirms the ~~findings at 10 m~~spatial findings: While the WIND Toolkit period is slightly above the long-term average, the magnitude of the difference is small (0.05 m s^{-1} on average). Similarly, the BC-HRRR period is slightly ~~lower~~ less windy than the WIND Toolkit period, but the differences are negligible. This confirms that the WIND Toolkit epoch is a suitable candidate for use as the basis for bias correction.

Appendix C: Comparison between BC-HRRR and WIND Toolkit

The WIND Toolkit serves as the distributional target for the quantile mapping bias correction. Comparing their climatological statistics verifies that the bias correction has successfully harmonized the datasets, enabling their combined use as a 16-year record. Figure C1 shows the difference in mean 100 m wind speed. The observed mean differences are negligible (<~~0.1~~0.03 m s^{-1}) across the majority of the domain, confirming that BC-HRRR successfully anchors the modern HRRR era to the legacy WIND Toolkit baseline.



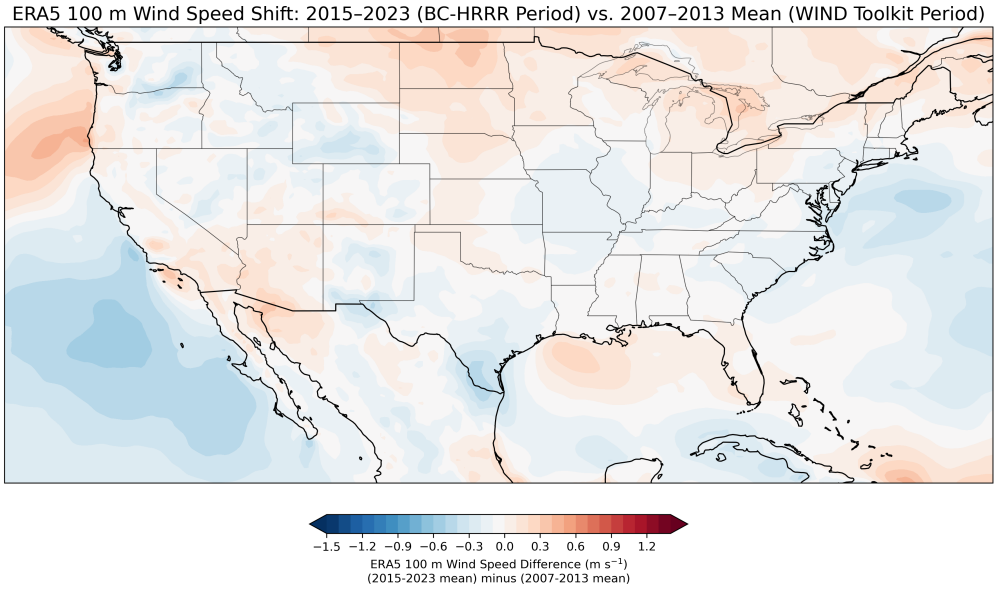


Figure B1. Spatiotemporal variability of ERA5 ~~10~~ 100 m wind speeds over the contiguous United States~~:~~. Note that both panels utilize only ERA5 data to evaluate shifts between specific epochs. (Top) The anomaly of the WIND Toolkit reference period (2007–2013) relative to the long-term climatology (~~1991~~1940–~~2020~~2025), illustrating the representativeness of the WIND Toolkit ~~period~~epoch. (Bottom) The shift in wind resource magnitude during the BC-HRRR period (2015–2023) compared to the WIND Toolkit period.

Author contributions. N.B and G.B. led the development of the HRRR MET Toolkit dataset. L.L. and G.B. led the development of the BC-HRRR dataset. P.P. and B.B. developed the software infrastructure for BC-HRRR data processing. D.D.T. provided expertise on the HRRR

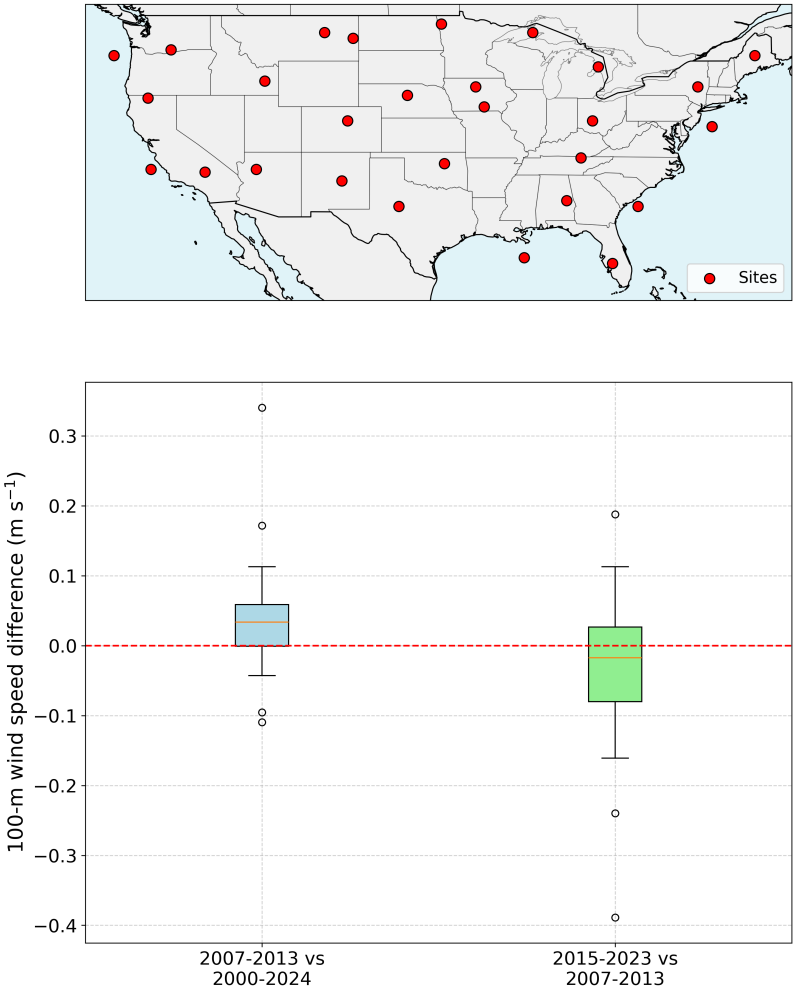


Figure B2. (Top) Map of locations at which 100 m ERA5 hourly wind speed was extracted. (Bottom) Distribution of ERA5 100 m wind speed differences across these sites. The left boxplot shows the anomaly of the WIND Toolkit reference period (2007–2013) relative to the long-term climatology (2000–2024), while the right boxplot illustrates the shift in wind resources during the BC-HRRR period (2015–2023) compared to the WIND Toolkit period.

model. N.B. coordinated the work, and led the data analysis. E.M. and U.E. performed the model validation. N.B., E.M., and U.E. led the manuscript preparation. All authors reviewed the manuscript draft.

Competing interests. The authors declare that they have no competing interests.

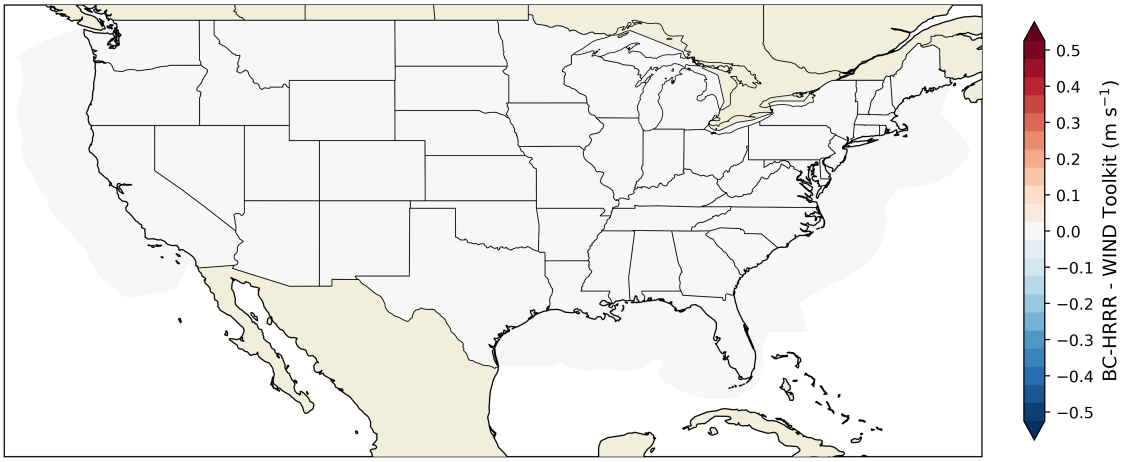


Figure C1. Difference in mean 100 m wind speed between the BC-HRRR and the WIND Toolkit. Values represent the BC-HRRR climatological mean (2015–2023) minus the WIND Toolkit climatological mean (2007–2013).

Acknowledgements. This work would not be possible without the significant efforts by NOAA to produce and maintain the HRRR. Collecting and storing atmospheric observations is a challenging task, and the authors would like to acknowledge all the people and institutions involved in the collection of observations used for model validation. Some data were obtained from the Atmospheric Radiation Measurement User Facility, a U.S. Department of Energy Office of Science user facility managed by the Biological and Environmental Research Program. Some data were obtained from the Wind Data Hub funded by U.S. Department of Energy Office of Critical Minerals and Energy Innovation and operated and maintained by Pacific Northwest National Laboratory at https://wdh.energy.gov. We acknowledge the use of data from the Woods Hole Oceanographic Institution (WHOI) MetOcean Data Initiative, collected by WHOI. We also thank the Barcelona Supercomputing Center for providing access to the Tall Tower Dataset. Data were also obtained from the Iowa Environmental Mesonet, maintained by the Department of Agronomy at Iowa State University. Finally, we acknowledge the Hanford Meteorological Station, managed by Hanford Mission Integration Solutions for the U.S. Department of Energy, for the provision of meteorological data from the Hanford Site. A portion of this research was performed using computational resources sponsored by the U.S. Department of Energy's Office of Critical Minerals and Energy Innovation and located at the National Laboratory of the Rockies.

References

Benjamin, S. G., James, E. P., Turner, D. D., Balmes, K. A., Sedlar, J., Lantz, K. O., Jensen, A. A., Riihimaki, L. D., and Augustine, J. A.: Excessive downward shortwave radiation in the HRRR and RAP weather models and testing strategies for improvements, Monthly Weather Review, 153, 2279–2293, https://doi.org/10.1175/MWR-D-25-0094.1, 2025.

Bianco, L., Djalalova, I. V., Wilczak, J. M., Olson, J. B., Kenyon, J. S., Choukulkar, A., Berg, L. K., Fernando, H. J., Grimit, E. P., Krishnamurthy, R., et al.: Impact of model improvements on 80 m wind speeds during the second Wind Forecast Improvement Project (WFIP2), Geoscientific Model Development, 12, 4803–4821, https://doi.org/10.5194/gmd-12-4803-2019, 2019.

Bianco, L., Mendeke, R., Lindblom, J., Djalalova, I. V., Turner, D. D., and Wilczak, J. M.: Evaluating the ability of the operational High Resolution Rapid Refresh model version 3 (HRRRv3) and version 4 (HRRRv4) to forecast wind ramp events in the US Great Plains, Wind Energy Science, 10, 2117–2136, https://doi.org/10.5194/wes-10-2117-2025, 2025.

Bodini, N., Castagneri, S., and Optis, M.: Long-term uncertainty quantification in WRF-modeled offshore wind resource off the US Atlantic coast, Wind Energy Science, 8, 607–620, https://doi.org/10.5194/wes-8-607-2023, 2023.

Bodini, N., Optis, M., Redfern, S., Rosencrans, D., Rybchuk, A., Lundquist, J. K., Pronk, V., Castagneri, S., Purkayastha, A., Draxl, C., Krishnamurthy, R., Young, E., Roberts, B., Rosenlieb, E., and Musial, W.: The 2023 National Offshore Wind data set (NOW-23), Earth System Science Data, 16, 1965–2006, https://doi.org/10.5194/essd-16-1965-2024, 2024.

Bodini, N., Egerer, U., and Maric, E.: Formatted hub-height wind speed observations across the Continental United States, https://doi.org/10.5281/zenodo.18177810, 2026.

Buster, G., Benton, B. N., Glaws, A., and King, R. N.: High-resolution meteorology with climate change impacts from global climate model data using generative machine learning, Nature Energy, https://doi.org/10.1038/s41560-024-01507-9, 2024a.

Buster, G., Pinchuk, P., Lavin, L., Benton, B., and Bodini, N.: Bias Corrected NOAA HRRR Wind Resource Data for Grid Integration Applications, https://doi.org/10.25984/2480349, dataset, 2024b.

Buster, G., Pinchuk, P., Lavin, L., Benton, B., and Bodini, N.: Bias Correcting NOAA's High-Resolution Rapid Refresh (HRRR) Wind Resource Data for Grid Integration Applications, Tech. Rep. NREL/PR-6A20-91749, National Renewable Energy Laboratory, https://doi.org/10.2172/2479268, 2024c.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?, Journal of Climate, 28, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.

Carvalho, D., Rocha, A., Gómez-Gesteira, M., and Santos, C.: A sensitivity study of the WRF model in wind simulation for an area of high wind energy, Environmental Modelling & Software, 33, 23–34, https://doi.org/10.1016/j.envsoft.2012.01.019, 2012.

Cheynet, E., Diezel, J. M., Haakenstad, H., Breivik, Ø., Peña, A., and Reuder, J.: Tall wind profile validation of ERA5, NORA3, and NEWA datasets using lidar observations, Wind Energy Science, 10, 733–754, https://doi.org/10.5194/wes-10-733-2025, 2025.

Collins, E., Lebo, Z. J., Cox, R., Hammer, C., Brothers, M., Geerts, B., Capella, R., and McCorkle, S.: Forecasting High Wind Events in the HRRR Model over Wyoming and Colorado. Part I: Evaluation of Wind Speeds and Gusts, Weather and Forecasting, 39, 705–723, https://doi.org/10.1175/WAF-D-23-0036.1, 2024.

Deppe, A. J., Gallus Jr, W. A., and Takle, E. S.: A WRF ensemble for improved wind speed forecasts at turbine height, Weather and Forecasting, 28, 212–228, https://doi.org/10.1175/WAF-D-11-00112.1, 2013.

Dörenkämper, M., Olsen, B. T., Witha, B., Hahmann, A. N., Davis, N. N., Barcons, J., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Sastre-Marugán, M., Sīle, T., Trei, W., Žagar, M., Badger, J., Gottschall, J., Sanz Rodrigo, J., and Mann, J.: The Making

of the New European Wind Atlas – Part 2: Production and evaluation, Geosci. Model Dev., 13, 5079–5102, https://doi.org/10.5194/gmd-13-5079-2020, 2020.

Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., Blake, B. T., Brown, J. M., Olson, J. B., Hu, M., Smirnova, T. G., Ladwig, T., Kenyon, J. S., Ahmadov, R., Turner, D. D., Duda, J. D., and Alcott, T. I.: The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description, Weather and Forecasting, 37, 1371–1395, https://doi.org/10.1175/WAF-D-21-0151.1, 2022.

Draxl, C., Clifton, A., Hodge, B.-M., and McCaa, J.: The Wind Integration National Dataset (WIND) Toolkit, Applied Energy, 151, 355–366, https://doi.org/10.1016/j.apenergy.2015.03.121, 2015.

Draxl, C., Wang, J., Sheridan, L., Jung, C., Bodini, N., Buckhold, S., Aghili, C., Peco, K., Kotamarthi, R., Kumler, A., Phillips, C., Purkayastha, A., Young, E., Rosenlieb, E., and Tinnesand, H.: WTK-LED: The WIND Toolkit Long-Term Ensemble Dataset, Tech. Rep. NREL/TP-5000-88457, National Renewable Energy Laboratory, https://doi.org/10.2172/2473210, 2024.

Energy Systems Integration Group: Weather Dataset Needs for Planning and Analyzing Modern Power Systems (Full Report), Tech. rep., Energy Systems Integration Group, Reston, VA, https://www.esig.energy/weather-data-for-power-system-planning, a Report of the Weather Datasets Project Team, 2023.

Frehlich, R., Meillier, Y., Jensen, M. L., Balsley, B., and Sharman, R.: Measurements of Boundary Layer Profiles in an Urban Environment, Journal of Applied Meteorology and Climatology, 45, 821 – 837, https://doi.org/10.1175/JAM2368.1, 2006.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Hahmann, A. N., Sīle, T., Witha, B., Davis, N. N., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Olsen, B. T., and Söderberg, S.: The making of the New European Wind Atlas – Part 1: Model sensitivity, Geoscientific Model Development, 13, 5053–5078, https://doi.org/10.5194/gmd-13-5053-2020, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

James, E. P., Benjamin, S. G., and Marquis, M.: A unified high-resolution wind and solar dataset from a rapidly updating numerical weather prediction model, Renewable Energy, 102, 390–405, https://doi.org/10.1016/j.renene.2016.10.059, 2017.

James, E. P., Benjamin, S. G., and Marquis, M.: Offshore wind speed estimates from a high-resolution rapidly updating numerical weather prediction model forecast dataset, Wind Energy, 21, 264–284, https://doi.org/10.1002/we.2161, 2018.

James, E. P., Alexander, C. R., Dowell, D. C., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., Brown, J. M., Olson, J. B., Hu, M., Smirnova, T. G., Ladwig, T., Kenyon, J. S., and Turner, D. D.: The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part II: Forecast Performance, Weather and Forecasting, 37, 1397–1417, https://doi.org/10.1175/WAF-D-21-0130.1, 2022.

Liu, Y., Feng, S., Berg, L. K., Wharton, S., Arthur, R., Turner, D. D., and Fast, J. D.: Benchmarking near surface winds in the HRRR analyses using multi-source observations over complex terrain in the southeastern United States, Journal of Applied Meteorology and Climatology, https://doi.org/10.1175/JAMC-D-24-0163.1, 2025.

Lupu, N., Selios, L., and Warner, Z.: A new measure of congruence: The Earth Mover's Distance, Political Analysis, 25, 95–113, https://doi.org/10.1017/pan.2017.2, 2017.

Maraun, D.: Bias Correcting Climate Change Simulations—A Critical Review, Current Climate Change Reports, 2, 211–220, https://doi.org/10.1007/s40641-016-0050-x, 2016.

Millstein, D., Jeong, S., Ancell, A., and Wiser, R.: A database of hourly wind speed and modeled generation for US wind plants based on three meteorological models, Scientific Data, 10, 883, https://doi.org/10.1038/s41597-023-02804-w, 2023.

Murcia, J. P., Koivisto, M. J., Luzia, G., Olsen, B. T., Hahmann, A. N., Sørensen, P. E., and Als, M.: Validation of European-scale simulated wind speed and wind generation time series, Applied Energy, 305, 117 794, https://doi.org/10.1016/j.apenergy.2021.117794, 2022.

National Laboratory of the Rockies: Wind Resource Database (WRDB), https://wrdb.nrel.gov/, 2024.

Olson, J. B., Kenyon, J. S., Djalalova, I., Bianco, L., Turner, D. D., Pichugina, Y., Choukulkar, A., Toy, M. D., Brown, J. M., Angevine, W. M., et al.: Improving wind energy forecasting through numerical weather prediction model development, Bulletin of the American Meteorological Society, 100, 2201–2220, 2019.

Optis, M., Bodini, N., Debnath, M., and Doubrawa, P.: Best Practices for the Validation of U.S. Offshore Wind Resource Models, Tech. rep., National Renewable Energy Laboratory, Golden, CO (United States), https://doi.org/10.2172/1755697, 2020a.

Optis, M., Kumler, A., Scott, G., Debnath, M., and Moriarty, P.: Validation of RU-WRF, the Custom Atmospheric Mesoscale Model of the Rutgers Center for Ocean Observing Leadership, Tech. Rep. NREL/TP-5000-75209, National Renewable Energy Laboratory, https://www.nlr.gov/docs/fy20osti/75209.pdf, 2020b.

Peco, K., Wang, J., Jung, C., Sever, G., Sheridan, L., Feinstein, J., Kotamarthi, R., Draxl, C., Young, E., Purkayastha, A., and Kumler, A.: Evaluation of a High-Resolution Regional Climate Simulation for Surface and Hub-height Wind Climatology over North America, Wind Energy Science Discussions, 2025, 1–36, https://doi.org/10.5194/wes-2025-13, 2025.

Pichugina, Y. L., Banta, R. M., Bonin, T., Brewer, W. A., Choukulkar, A., McCarty, B. J., Baidar, S., Draxl, C., Fernando, H. J. S., Kenyon, J., Krishnamurthy, R., Marquis, M., Olson, J., Sharp, J., and Stoelinga, M.: Spatial Variability of Winds and HRRR–NCEP Model Error Statistics at Three Doppler-Lidar Sites in the Wind-Energy Generation Region of the Columbia River Basin, Journal of Applied Meteorology and Climatology, 58, 1633–1656, https://doi.org/10.1175/JAMC-D-18-0244.1, 2019.

Pronk, V., Bodini, N., Optis, M., Lundquist, J. K., Moriarty, P., Draxl, C., Purkayastha, A., and Young, E.: Can reanalysis products outperform mesoscale numerical weather prediction models in modeling the wind resource in simple terrain?, Wind Energy Science, 7, 487–504, https://doi.org/10.5194/wes-7-487-2022, 2022.

Ramon, J., Lledó, L., Pérez-Zanón, N., Soret, A., and Doblas-Reyes, F. J.: The Tall Tower Dataset: a unique initiative to boost wind energy research, Earth System Science Data, 12, 429–439, https://doi.org/10.5194/essd-12-429-2020, 2020.

Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G., and Shelby, J.: The National Solar Radiation Data Base (NSRDB), Renewable and Sustainable Energy Reviews, 89, 51–60, https://doi.org/10.1016/j.rser.2018.03.003, 2018.

Sheridan, L. M., Duplyakin, D., Phillips, C., Tinnesand, H., Rai, R. K., Flaherty, J. E., and Berg, L. K.: Evaluating the potential of short-term instrument deployment to improve distributed wind resource assessment, Wind Energy Science, 10, 1451–1470, https://doi.org/10.5194/wes-10-1451-2025, 2025a.

Sheridan, L. M., Krishnamurthy, R., Nguyen, T. M., Chen, Y.-L., Gustafson Jr., W. I., Liu, Y., Hsiao, F., Newsom, R. K., Spicer, P., Kassianov, E., Pekour, M., Bodini, N., and Severy, M.: Performance of reanalysis and mesoscale models for wind resource assessment off the coast of Hawaii, Wind Energy Science Discussions, 2025, 1–43, https://doi.org/10.5194/wes-2025-167, 2025b.

Sheridan, L. M., Wang, J., Draxl, C., Bodini, N., Phillips, C., Duplyakin, D., Tinnesand, H., Rai, R. K., Flaherty, J. E., Berg, L. K., Jung, C., Young, E., and Kotamarthi, R.: Performance of wind assessment datasets in United States coastal areas, Wind Energy Science, 10, 1551–1574, https://doi.org/10.5194/wes-10-1551-2025, 2025c.

Skamarock, W. C.: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, Monthly Weather Review, 132, 3019–3032, https://doi.org/10.1175/MWR2830.1, 2004.

U.S. Energy Information Administration: Annual Energy Outlook 2023, Tech. rep., U.S. Energy Information Administration, https://www.eia.gov/outlooks/aeo/pdf/AEO2023_Narrative.pdf, 2023.

Vibrant Clean Energy: Vibrant Clean Energy Resource Adequacy Renewable Energy (RARE) Power Dataset (Version 1.0.0), https://doi.org/10.5281/ZENODO.13937522, dataset, 2024.

Wilczak, J. M., Akish, E., Capotondi, A., and Compo, G. P.: Evaluation and Bias Correction of the ERA5 Reanalysis over the United States for Wind and Solar Energy Applications, Energies, 17, 1667, https://doi.org/10.3390/en17071667, 2024.

Yin, Y. and Peña, M.: Analysis of bias correction of HRRR model outputs for offshore wind power ramp events, Renewable Energy, 228, 120 581, https://doi.org/10.1016/j.renene.2024.120581, 2024.

Table 2. Summary of observational datasets used in this study. The temporal range indicates the period used in our validation analysis. For offshore stations, the last column reports the approximate distance to the nearest shoreline.

Dataset	Heights~~(m)~~ (m)	Resolution~~(min)~~ (min)	Years	Setting	Distance to shore (km)
Tall Towers - 42361 (Auger Platform)	122	60	2014–2018	Offshore	~220
Tall Towers - 42362 (Brutus Platform)	122	15	2015–2017	Offshore	~140
Tall Towers - 42363 (Mars Platform)	122	15	2014–2016	Offshore	~85
Tall Towers - 42364 (Ram-Powell Platform)	122	15	2014–2018	Offshore	~90
Tall Towers - 42394 (Olympus Platform)	100	60	2014–2016	Offshore	~85
Tall Towers - Brookhaven	50, 85	10	2014–2022	Land	–
Tall Towers - NLR National Wind Technology Center (NWTC) M5	30–130	10	2014–2022	Land	–
Tall Towers - Park Falls	30, 122	60	2014–2022	Land	–
Tall Towers - upbc1 (Union Pacific Railroad Bridge)	100	10	2014–2021	Land (coastal)	–
Tall Towers - Walnut Grove	122	60	2014–2016	Land	–
AWAKEN event log lidar	88.5	10	2023–2024	Land	–
Hanford tower	61, 122	15	2020–2024	Land	–
Iowa Mesonet - ETTI4 tower	80, 120	5	2016–2021	Land	–
Humboldt lidar	40–240	10	2020–2021	Offshore	~35
Morro Bay lidar	40–240	10	2020–2021	Offshore	~40
ARM - Southern Great Plains (SGP) Central Facility lidar	65–195	15	2014–2023	Land	–
ARM - East Pacific (EPC) lidar	71–185	10	2023	Land (coastal)	–
ARM - Gunnison (GUC) lidar	70–183	15	2021–2023	Land	–
ARM - Houston (HOU) lidar	36–190	15	2021–2022	Land	–
NYSERDA - E05 Hudson North lidar	18–198	10	2019–2021	Offshore	~85
NYSERDA - E05 Hudson SouthWest lidar	18–198	10	2022–2023	Offshore	~50
NYSERDA - E06 Hudson South lidar	18–198	10	2019–2022	Offshore	~60
WHOI Air-Sea Interaction Tower (ASIT) lidar	53–200	10	2016–2018	Offshore	~3

Table 3. Mean hub-height performance metrics by model, averaged over all stations (corresponding to averages from Fig. 6).

Model	Mean bias ($\mathrm{m\,s^{-1}}$)	Mean cRMSE ($\mathrm{m\,s^{-1}}$)	Mean r
HRRR MET Toolkit	0.10	2.08	0.82
BC-HRRR	0.36	2.16	0.82
WTK-LED	0.82	3.34	0.67
WTK-LED Climate	1.05	4.44	0.39
NOW-23	0.01	2.17	0.87

Table 4. Limitations of the HRRR MET Toolkit dataset and potential user actions.

Limitation	**Description**	**Potential Action**
Spatial/temporal resolution	Provided at 2 km hourly resolution.	Use sub-hourly alternative datasets for ramping studies or localized downscaling for site-specific needs.
Complex terrain performance	Modeling challenges remain in complex topography.	Exercise caution in complex areas; utilize site-specific measurements to verify local performance.
Accessibility	Designed for energy modelers; excludes some raw meteorological variables found in native GRIBs.	Utilize the raw NOAA HRRR archives for specialized atmospheric research beyond power system needs.
Incomplete validation	Validation is currently limited to sites with publicly available observations and focused on hub heights (80–140 m).	Conduct additional validation if using the dataset for low-level (e.g., 10 m) or high-altitude applications.
Gaps in record	Missing year (2014) and small gaps prior to 2019 filled via interpolation/alternative forecasts.	For 2014 specifically, users should refer to WTK-LED or other reanalysis products.
Future climate	Historical records do not account for future changes in the climate signals.	Incorporate climate-adjusted datasets, such as Sup3rCC (Buster et al., 2024a), for long-term (2050+) planning.

Table A1. Comparison of technical specifications for WIND Toolkit, WTK-LED, WTK-LED Climate, NOW-23, and HRRR.

Feature	**WIND Toolkit**	**WTK-LED**	**WTK-LED Climate**	**NOW-23**	**HRRR**
WRF Version	3.4.1	4.1.3	4.1.3	4.2.1	WRF-ARW, version varies: v1: v3.5+, v2: v3.6+, v3: v3.8.1+, v4: v3.9+
Years	2007–2013	2014–2015 and 2018–2020	2000–2020	2000–2019/2023 (based on region)	Sep 2014–present
Domains	Four nested	One	One	Two nested	One
Horizontal Resolution	2 km (innermost domain; outer domains: 54 km, 18 km, 6 km)	2 km	4 km	2 km (innermost domain; outer domain: 6 km)	3 km
Model Δt	11.1 s (innermost domain; outer domains: 300 s, 100 s, 33.3 s)	5 s	Adaptive (min 5 s)	6.67 s (innermost domain; outer domain: 20 s)	20 s
Output Resolution	5 min	5 min	60 min	5 min	60 min (15 min for a subset of 2D fields)
Boundary Conditions	ERA-Interim	ERA5	ERA5	ERA5	RAP
SST Forcing	ERA-Interim (NCEP Real-Time Global and coarsened OSTIA)	ERA5 (coarsened OSTIA)	ERA5 (coarsened OSTIA)	OSTIA	GFS
Planetary Boundary Layer Scheme	YSU	MYNN	YSU	MYNN or YSU (regional tuning)	MYNN
SL Scheme	MM5	MYNN	MM5	MYNN or MM5 (regional tuning)	MYNN
LSM	Noah	Noah	Noah	Noah	RUC
Microphysics	Ferrier	Morrison	Morrison	Ferrier	Thompson-Eidhammer
Data Assimilation	No	No	No	No	Yes

Note: SST: Sea Surface Temperature; SL: Surface Layer; LSM: Land Surface Model; MYNN: Mellor–Yamada–Nakanishi–Niino; YSU: Yonsei University; MM5: Fifth-Generation Penn State/National Center for Atmospheric Research Mesoscale Model; OSTIA: Operational Sea Surface Temperature and Sea Ice Analysis; RAP: Rapid Refresh; RUC: Rapid Update Cycle; ERA5: ECMWF Reanalysis v5; GFS: Global Forecast System; NCEP: National Centers for Environmental Prediction.