Review of “The HRRR Meteorology, Energy, and Transmission (MET) Toolkit: Advancing high-resolution atmospheric data for contiguous U.S. energy applications” by Bodini et al.

General comments:

This manuscript is extremely important as it demonstrates the very significant value that the operational HRRR model provides for wind resource characterization relative to other historical data sets, and moreover provides a tool to make the HRRR data set readily available to the wind energy community. It demonstrates this value through a series of well-conceived figures, displaying a wide variety of metrics. The manuscript if well-written, and easy to follow, and was a pleasure to read. My only two significant comments, are first, that the analysis appears to assume that the ERA5 is bias-free, which it is not. Second, and more importantly, the analysis does not appear to take into account tower shadowing effects in tall-tower data. If that is true, then I believe that tower shadowing can explain some, if not most, of the large differences in the bias behavior found between onshore tall tower sites and lidar sites. Ideally, the tall-tower analysis should be repeated removing data values potentially affected by tower shadowing.

We thank the Reviewer for their thorough and constructive review. We have addressed the specific comments (which include the two significant comments summarized above) in the rest of this document.

Specific comments:

Line 38: What is the “climate variant” of the Wind-LED Toolkit?

The WTK-LED Climate is a 20-year (2001-2020), 4 km resolution climate-scale WRF simulation covering North America, developed by Argonne National Laboratory to complement the shorter, higher-resolution core WTK-LED product. It is described in detail in Draxl et al. (2024), and briefly in Section 2.1 of our paper. We have added a short sentence providing more context at line 38: "...the WIND Toolkit Long-term Ensemble Dataset (WTK-LED) (Draxl et al., 2024) and its Climate variant (a 20 year, 4 km resolution continental simulation developed to provide a climatological record)..."

Lines 38-40: Is there a reference to the analysis showing that the WIND-LED Toolkit has larger biases than does the WIND Toolkit?

This statement is a synthesis derived from comparing the baseline validation metrics reported in the original legacy WIND Toolkit documentation (Draxl et al., 2015), the WTK-LED technical report (Draxl et al., 2024), and the validation results presented later within this manuscript.
Because a single standalone paper comparing these specific historical biases does not exist, we have revised the manuscript text to point the reader to these reference documents and our own findings to substantiate the comparison:
"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, Draxl et al., 2024) as well as the analysis presented herein -- show they exhibit larger biases against meteorological observations than the legacy WIND Toolkit."

Lines 68-70: A similar evaluation analysis as that of Dorenkamper et al. 2020 was provided by Wilczak et al, 2024, https://doi.org/10.3390/en17071667 for hub-height winds across the US.

We thank the Reviewer for pointing to Wilczak et al. (2024). We have added this reference:
"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."

Lines 150-154: Can the effective spatial resolution of the re-gridded data set be estimated? Due to the interpolation from four 3-km resolution grid points, I would expect this resolution to be on the order of 4-6 km. Can it be determined more precisely? In any case, the larger effective resolution should be mentioned.

The Reviewer's intuition is correct: the four-point inverse distance-weighted (IDW) interpolation from the native 3 km HRRR grid to the 2 km output grid cannot introduce new physical information at scales finer than the source grid spacing. The geometric support of the four nearest neighbors introduces modest additional smoothing, consistent with the Reviewer's estimate of ~4-6 km. However, it is important to note that the dominant constraint on effective resolution is not the interpolation step but rather the numerical diffusion inherent to the WRF dynamical core. Following Skamarock (2004), the effective dynamical

resolution of WRF is approximately 7Δx, which for the 3-km HRRR corresponds to ~21 km.
We have added a clarifying sentence to the manuscript noting that, despite the 2 km grid spacing, the effective spatial resolution of the regridded dataset remains approximately 20 km, consistent with the spectral characteristics of the HRRR model dynamics:
"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 Δx (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."

Reference:

- ***Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. Mon. Wea. Rev., 132, 3019–3032, doi:10.1175/MWR2830.1.***

Lines 172-176: For grid integration studies, having solar variables consistent with the HRRR-MET wind variables would be useful. Is there a reason why radiation variables were not included, and is there any plan to include solar variables in the future?

We completely agree with the Reviewer that co-located wind and solar variables are critical for modern grid integration studies. The exclusion of solar variables in this release was primarily programmatic: this specific development was funded with a focus to improve wind resource characterization. Additionally, NLR already develops and maintains the National Solar Radiation Database (NSRDB), which serves as the flagship dataset for solar applications. That said, we recognize the growing need for unified datasets for hybrid plant and grid studies. A valuable next step would be to evaluate how the native HRRR radiation fields perform against the NSRDB to understand their viability for integrated studies.

We have revised the manuscript to point users to the NSRDB and the native HRRR archive, and have noted this evaluation as an important avenue for future work:
"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."

Figure 1. The color-scale here is a bit frustrating, as most mean windspeed values fall within a rather narrow range of 6-9 m/s, with very little color gradation present. As a result, all of the high plains states have an almost uniform color.

We agree that the color scale in Figure 1 compresses variability in the 6-9 m s⁻¹ range that dominates across the Great Plains. In the revised manuscript we have adjusted the colormap:

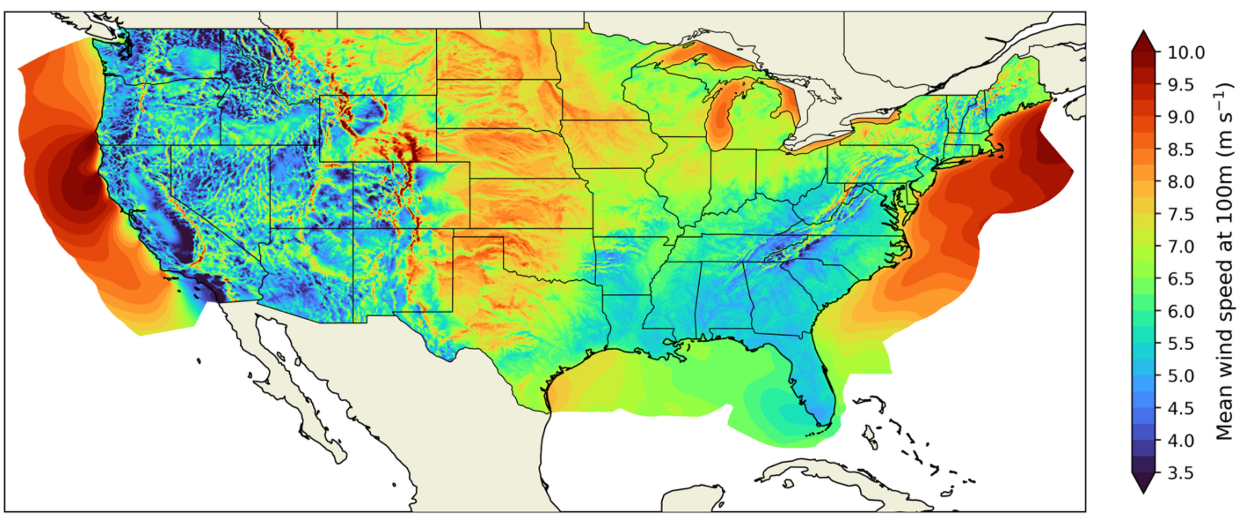


Figure 1. Is there any evidence in the multi-year data set of jumps in the CONUS-averaged mean wind speeds that occur at the transition dates between the various HRRR versions? Quantifying such jumps would be important for those concerned about the impact of the changes to the model physics between the model versions. – OK, I see that this has been addressed later on Lines 382-395. It would be good to foreshadow that later analysis here or in the Introduction.

We agree that foreshadowing the version-transition analysis earlier would improve the paper's clarity, and we have added the following text in Section 2.3.1: "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)."

Line 200. After applying the quantile bias correction to the HRRR, are the biases exactly zero at each grid point?

By construction, the quantile mapping procedure maps the HRRR cumulative distribution function (CDF) onto the WIND Toolkit CDF at each grid point and height level. In theory, this drives the bias in the long-term mean to zero. In practice, due to numerical interpolation between discrete quantile bins and the handling of extreme tail values, negligible residual differences remain (on the order of ±0.03 m s-1). Therefore, we consider the long-term mean wind speed bias between BC-HRRR and the WIND Toolkit to be effectively zero (see Appendix C and the updated figure there). We have clarified this in the text:
"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)."

Line 204-205. However, the ERA5 has been shown to have its' own biases across the CONUS (Wilczak et al, 2024; Pronk et al, 2022). Don't those biases need to be removed? See also comments on manuscript lines 463-465 below.

The Reviewer is correct that ERA5 has documented biases over CONUS. Our use of ERA5 in Appendix B is intended only as a tool for assessing the relative representativeness of different time periods (i.e., whether 2007-2013 is a climatologically typical window relative to a longer baseline), not to validate absolute wind speeds. We now clarify this distinction:
"Appendix B analyzes ERA5 long-term data to assess the extent to which the WIND Toolkit period represents long-term conditions. 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."

Line 218-220. I don't understand what is meant by "methodological consistency" in requiring that the exact same time periods are used to compare the biases between the different products. Previously, the Wind Tool Kit (2007-2013) was used to bias correct the HRRR-MET (2015-present) event though the two model data sets have no period of overlap whatsoever. Wouldn't the same methodological (in)consistency apply in that case?

We thank the Reviewer for catching this apparent contradiction. The Reviewer is entirely correct that the bias-correction process for BC-HRRR utilized non-overlapping periods, and we acknowledge that our phrasing was unclear.
The rationale for this difference stems from the distinct goals of each methodology. In Section 3 (Inter-comparison of modeled datasets), the goal is to evaluate differences in model physics, resolution, and configuration. To isolate these model differences, we must eliminate inter-annual meteorological differences by enforcing overlapping time periods.
Conversely, the BC-HRRR dataset (Section 2.4) was developed as an <u>experimental</u> product to test whether a legacy climatology (WIND Toolkit, 2007-2013) could effectively constrain a more recent dataset (HRRR, 2015-2023). The lack of overlap was a forced constraint of that experiment, justified by the fact that the WIND Toolkit epoch is broadly representative of long-term climate (as shown in Appendix B).
Ultimately, as noted in the manuscript, the pure HRRR MET Toolkit proved superior to this experimental BC-HRRR product.

We have removed the phrase "methodological consistency" and revised the manuscript text to explicitly contrast the strict temporal matching required for the inter-comparison against the experimental approach used to create BC-HRRR:
"We present a comprehensive inter-comparison of all overlapping dataset pairs in Fig. 4. To ensure 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 for that specific 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."

Lines 250-260: Tower wind speeds can be significantly impacted by tower shadowing effects when the cup anemometers or sonic anemometers are downwind of the tower structure. Was anything done to minimize these effects? If not, why not? No mention is made whether the QC flags remove all downstream values affected by tower shadowing. See also comments for Line 295 below and Figure 9 below.

We thank the Reviewer for raising this important point, and we agree the paper should make these details explicit.
For the Tall Tower Dataset (Ramon et al., 2020), the Quality Control Software Suite for Tall Towers (QCSS4TT) includes a dedicated tower shadow test among its 16 quality checks. This test flags observations contaminated by tower wake at stations where more than one anemometer is installed at a given height -- though the metadata do not specify which stations meet this criterion. Because we retained only observations with QC flag = 1 (passing all applicable checks), tower-shadowed data were removed wherever this test was applicable.
For the Iowa Mesonet ETTI4 station, dual anemometers are installed on opposite booms at each measurement height. We explicitly addressed tower wake by selecting the upwind sensor at each time step based on concurrent wind direction measurements.
We have revised the manuscript to reflect these details:
"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 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."

Reference:

- ***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 Syst. Sci. Data, 12, 429–439, https://doi.org/10.5194/essd-12-429-2020, 2020.

Also, I believe that many of the "tall towers" listed in the Gulf of Mexico are in fact short towers mounted on top of huge offshore drilling platforms, in which case flow distortion effects from the platforms can be very large, even larger than the tower shadowing effects. Was anything done to account for those flow distortion effects? At a minimum, the heights above the tops of the drilling structures should be mentioned, and the possibility for flow distortion effects should be acknowledged.

The Reviewer raises a valid and important concern. The five Gulf of Mexico stations (42361-42364 and 42394) are indeed anemometers mounted on structures atop large offshore oil and gas platforms, not conventional meteorological towers. The multi-level topsides of these platforms can introduce localized flow distortion and structural wakes that may bias wind speed measurements.
The public Tall Towers metadata do not document whether any platform-specific flow distortion corrections were applied to the archived data, so none were applied in our analysis beyond whatever corrections are embedded in the source dataset (once again, only measurements satisfying QC flag = 1 were retained). We agree this context must be transparently communicated. We have made two changes:

1. ***Table 2 now explicitly identifies these stations as platform-mounted.***
2. ***Section 4 includes the following added text:***
   "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).."

For reference, the structural context of each station, based on available documentation, is as follows:

Station	*Platform*	*Anemometer height (MSL)*	*Approx. deck height (MSL)*	*Anemometer height above deck*
42361	***Auger***	***122 m***	***22.6 m***	***~99 m***
42362	***Brutus***	***122 m***	***12.2 m***	***~110 m***
42363	***Mars***	***122 m***	***13.7 m***	***~108 m***
42364	***Ram-Powell***	***122 m***	***12.2 m***	***~110 m***
42394	***Olympus***	***100 m***	***34.1 m***	***~66 m***

Line 276: Is there a reference for the Wasserstein distance?

Thank you for pointing this out. We have added a reference for the Wasserstein distance (Lupu et al., 2017) to the revised manuscript. To provide helpful context for our readers, we also added a citation to Hahmann et al. (2020), which demonstrates a successful recent application of this metric specifically within atmospheric science.
References:

- ***Lupu, N., Selios, L., and Warner, Z.: A new measure of congruence: The Earth Mover's Distance, Polit. Anal., 25, 95–113, https://doi.org/10.1017/pan.2017.2, 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, Geosci. Model Dev., 13, 5053–5078, https://doi.org/10.5194/gmd-13-5053-2020, 2020.***

Line 295: I suspect that much of the overall positive biases found in all of the models, and especially when compared against onshore tall towers, is due to the underestimation of the observed wind speeds due to tower shadowing.

We thank the Reviewer for this insightful and important interpretation. We agree it deserves explicit acknowledgment in the paper. We have added the following sentences to Section 5 to make this caveat explicit:
"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."

We note that even at lidar sites, a positive bias remains (see Table 3 and Figure 9) for the WTK-LED and WTK-LED Climate, confirming that at least some fraction of

the positive bias reflects a genuine model tendency rather than exclusively an observational artifact. Disentangling these contributions fully would require co-located anemometer and lidar measurements at the same sites -- a valuable direction for future validation work.

Lines 312-313: I don’t understand the sentence “This suggests that while quantile mapping effectively corrects offshore under-speeding, it may exacerbate positive biases onshore where local terrain effects dominate”. This seems to put the blame on the quantile mapping procedure, when it is probably due to inadequacies in the Wind ToolKit data, specifically to differences in the onshore and offshore Wind Toolkit biases relative to the truth.

The Reviewer correctly points out that the sentence could be read as attributing the onshore positive bias exacerbation to the quantile mapping procedure per se, when the root cause is the spatial variability in the WIND Toolkit's own biases. We have revised the text for clarity:
"This suggests that the onshore positive bias in 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."

Figure 9: I note here that the models all tend to closely follow the observed PDFs when measured by lidars, whereas in Fig.7 for tall towers, larger differences exist, which are consistent with tower shadowing effects in which observed high wind speeds are shifted to lower wind speed bins,

Comparing the PDFs in Figures 7 and 8, there is indeed a pattern where models track the observed wind speed distributions more closely at lidar sites than at (most) tall tower sites. We have added a sentence to the results discussion of the PDF figures to make this interpretation explicit:
"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 18: This is a very nice and important figure showing the time evolution of the HRRR model as it has improved over time. It may be worth noting here that at least some of these improvements occurred because of the concerted, collaborative effort between DOE, NOAA, and the private sector at finding and correcting HRRR model errors through the WFIP field campaigns. This demonstrates the value of those collaborations and in the investments made in them. Also, I'm curious if the authors can speculate on why the bias increases pretty sharply towards the end of the timeseries, while the other metrics show continued improvement?

We thank the Reviewer for this important contextual suggestion. Regarding the WFIP campaigns, we have gladly added an acknowledgment:
"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."

Regarding the apparent bias increase near the end of the time series: this is an interesting feature. We speculate it may reflect the very reduced number of validation stations with observations extending into late 2023, which increases sensitivity to site-specific anomalies (in this case, in bias).

References:

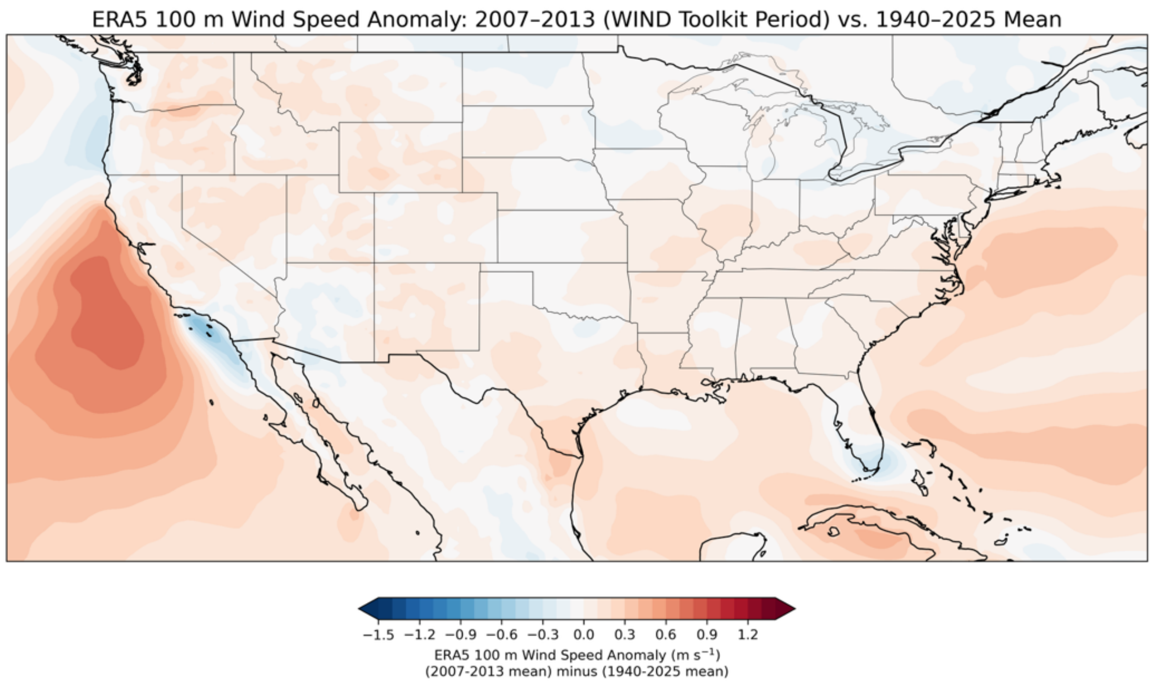
- ***Olson, J. B., Kenyon, J. S., Djalalova, I., Bianco, L., Turner, D. D., Pichugina, Y., Choukulkar, A., Toy, M. D., Brown, J. M., Angevine, W. M., et al.: Improving wind energy forecasting through numerical weather prediction model development, Bulletin of the American Meteorological Society, 100,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

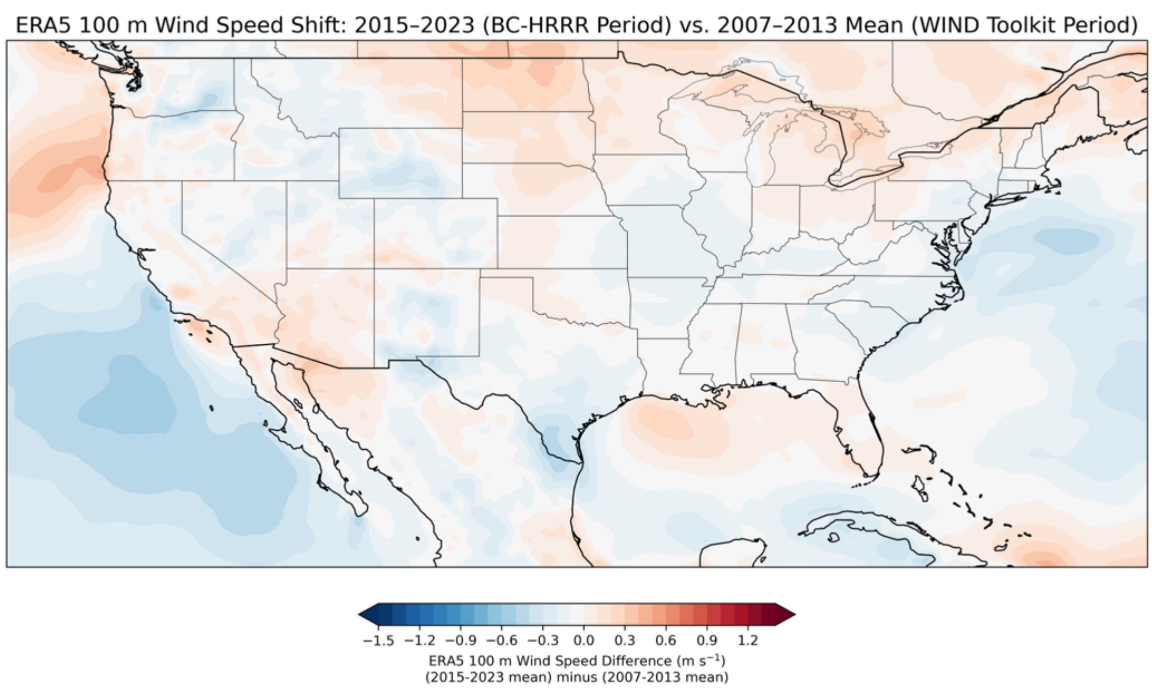
Line 405: The impact of data assimilation is an interesting and important question. To some degree, one could test whether the higher skill of the HRRR comes from DA, or from better model physics, by looking at HRRR skill at hour 18, or even hour 48 when those longer forecasts are available. At those longer forecast horizons, the value of DA will diminish, and higher skill in the HRRR relative to the other models will depend more on more physics. This question probably is beyond the scope of this paper, but it could be mentioned as an area for future research.

The Reviewer's suggestion of evaluating HRRR at longer forecast horizons (hour 18 or 48) to disentangle DA from physics contributions is very valuable. We have added this as a future research direction:
"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."

Line 459: I find it strange that 10m ERA5 data are used in this comparison, even though 100m data are readily available. The rationale for this seems to be that monthly values are pre-averaged at 10 m, but it would be trivial to average hourly values to monthly values, and would make the results much more valuable for wind energy purposes.

We entirely agree with the Reviewer. We have downloaded the 100 m ERA5 data, processed it, and updated the analysis and Figure B1 in the Appendix to reflect 100 m wind speeds, making it much more relevant for energy applications.
Updated Figure:





Line 461. If I interpret this sentence correctly, ERA5 data only from 1991-2020 are used to determine the long-term conditions, even though the ERA5 data set spans a much longer period. If so, why restrict it to only these 30 years?

Thank you for this suggestion. We have expanded our analysis to utilize the full available ERA5 record from 1940 through 2025 to determine the long-term climatology. Figure B1 and the corresponding text have been updated to reflect this extended period.

Line 464: "… relative to observations …". I'm confused here, because I thought this comparison was against the ERA5, not observations. The ERA5 values are not observations, but are highly dependent on the IFS model.

We apologize for the ambiguity. We were attempting to explain a downstream consequence of the bias correction process. Because the 10-m ERA5 data indicated the WIND Toolkit period (2007-2013) was slightly windier than the BC-HRRR period (2015-2023), anchoring (via bias correction) the BC-HRRR data to that windier historic baseline could result in the BC-HRRR product having a slight positive bias when evaluated against actual ground measurements (observations) taken during the 2015-2023 period. We have rewritten this section in the manuscript to make this logic more explicit, and to reflect the fact that, when now considering 100-m winds in ERA5, the temporal variability is more region-dependent:
"The ERA5 data also indicates that the shift in wind resource magnitude between the WIND Toolkit period and the BC-HRRR period (2015--2023) exhibits distinct spatial variability across the contiguous United States. While the eastern and northwestern US experienced slightly lower wind speeds 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 2007--2013 WIND Toolkit distribution, this temporal mismatch means the resulting BC-HRRR 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 the largest anomalies restricted primarily to offshore regions."

Lines 463-465. A significant portion of the Wind Toolkit positive bias relative to the ERA5 shown in Figure B1 is likely due to the known low-bias of the ERA5 (Wilczak et al, 2024). It may be beneficial to acknowledge this, otherwise readers will mistakenly interpret the bias in this figure as a shortcoming of the Wind Toolkit, when in fact it is more likely a shortcoming in the ERA5.

We appreciate the Reviewer highlighting the important findings of Wilczak et al. (2024) regarding the low bias of ERA5, and we have added a citation to this work in the revised Appendix to provide better context:
"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)."

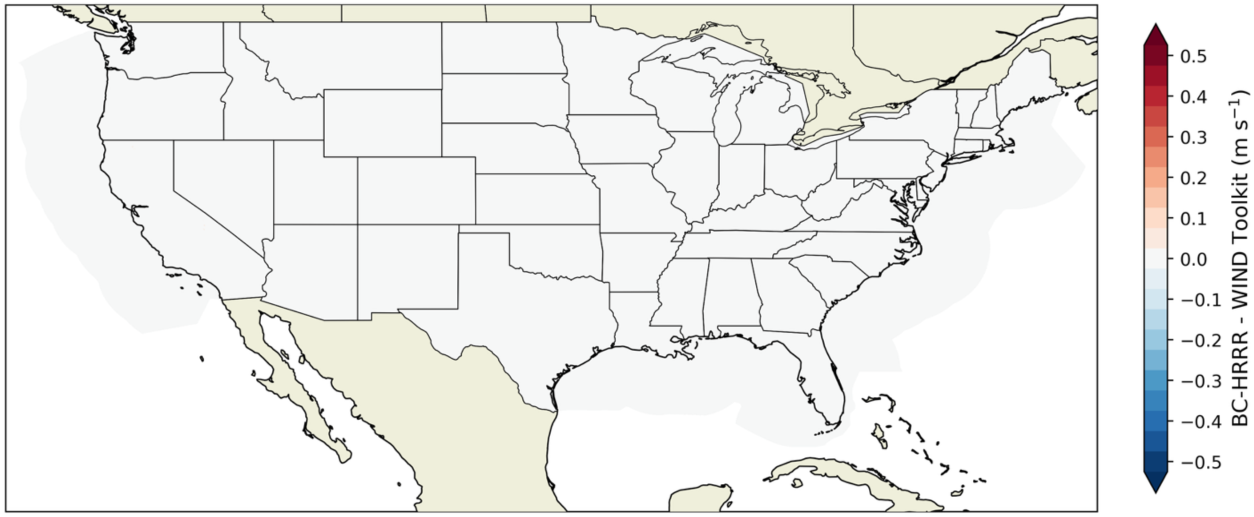
However, we would also like to respectfully clarify a potential misunderstanding regarding Figure B1: this figure does not compare the WIND Toolkit data to ERA5 data. Rather, it compares ERA5 to itself across different time periods (specifically, the ERA5 mean during the 2007–2013 WIND Toolkit epoch minus the ERA5 long-term mean). It is a purely temporal comparison to assess climate representativeness. To prevent future readers from making this same misinterpretation, we have carefully adjusted the text and captions to emphasize that this is a temporal anomaly utilizing solely ERA5 data.

Line 467: At selected observation sites?

We thank the Reviewer for pointing out this ambiguity. To clarify, these were not the observation sites used earlier in the study. Because the availability of our observational data did not span the entire CONUS domain equally, we intentionally selected a geographically distributed set of reference points to ensure a robust spatial representation across the entire domain. We have revised the manuscript to explicitly state how and why these points were chosen:
"To confirm this representativeness at hub height, we extracted 100 m hourly wind speed data from ERA5 at a geographically distributed set of locations covering the years 2000--2024. 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 locations (Fig. B2)."

Figure C1. The plots say WTK-LED but the caption says Wind Toolkit (no LED). Aren't those different models?

We had the wrong figure file in the draft – our apologies! The figure file has been updated in the revised draft, and it is included below:


0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
BC-HRRR - WIND Toolkit (m s^{-1})