Reviewer 1

This manuscript investigates how short measurements could be long-term corrected (LTC) with reference data from ERA5 in order to obtain accurate estimates of long-term mean wind speed and energy production. The motivation is that full-scale (1y+) measurement campaigns are costly and often unfeasible for small-scale wind power projects.

Overall, the paper is clear and the study is robust. The dataset of wind speed measurements is relatively large and representative for small-scale wind power in the US. The manuscript addresses a real problem, which is well described, and is thus relevant. A nice aspect is the focus on worst case scenarios in addition to mean statistics.

There are however some possible improvements to methodology and text. Below are comments listed approximately in order of importance.

We really appreciate you taking the time to review our work and provide great suggestions for improvement. We have addressed and incorporated your ideas and edits as follows. Again, thank you!

1. ERA5 is state-of-the-art for LTC and as input to mesoscale models. Due to e.g. its coarse resolution it is however not suitable for mean wind estimates at particular sites. Raw ERA5 is thus a poor baseline and all comparisons becomes more or less irrelevant. It would strengthen the paper greatly if e.g. GWA (global wind atlas, contains mean wind speed and parameters needed for energy calculations) was used as benchmark.

We really appreciate this recommendation for improving the baseline for comparison. To address your helpful suggestion, we have added Section 3.5 and Figure 13, which compare the performance of the MCP-based estimates with the long-term estimates from two higher-resolution wind datasets: 1) your suggestion of Global Wind Atlas (250 m resolution) and 2) the recently released wind climatology component of the WIND Toolkit Long-term Ensemble Dataset (4 km resolution).

"3.5 Performance comparison with higher-resolution wind datasets

ERA5 has shown to be a valuable reference dataset for developing long-term wind speed estimates via MCP with short-term observations thanks to its extensive temporal coverage and relative success at representing fluctuations in observed wind speeds. However, ERA5 has the limitations of coarse horizontal resolution and a tendency to exhibit a slow wind speed bias (Ramon et al., 2019; Gualtieri, 2021; Sheridan et al., 2022; Wilczak et al., 2024), which urges comparison of the long-term MCP results with long-term estimates from higher-resolution wind datasets. We explore the performance of the MCP-based long-term wind speed estimates relative to Global Wind Atlas version 3 (GWA3) and the climatology component of the WIND Toolkit Long-term Ensemble Dataset (WTK-LED Climate), for which long-term wind speed estimates are freely and easily accessed through user friendly web applications.

GWA3 is produced by the Technical University of Denmark (DTU) and the World Bank Group. The developers used the Weather Research and Forecasting (WRF) mesoscale model (Skamarock et al., 2008) in conjunction with the Rapid Radiative Transfer Model (RRTM) for the longwave and shortwave radiation schemes (Mlawer et al., 1997; Iacono et al., 2008), the Mellor-Yamada-Janjić planetary boundary layer (PBL) scheme (Janjić 1994), and ERA5 as the input and boundary conditions to produce simulated wind data at a horizontal resolution of 3 km (Davis et al., 2023). Next, microscale modelling was performed using the Wind Atlas Analysis and Application Program (WAsP) model (Troen and Petersen, 1989) with an output grid spacing of 250 m for GWA3. GWA3 provides global coverage for land-based wind estimates and offshore wind estimates within 200 km of shorelines. Long-term wind data are output at five heights between 10 m and 200 m based on the 10-year period of 2008-2017, and wind speed indices illustrate trends at annual, monthly, and diurnal temporal resolutions. Users can access GWA3 through its web application (DTU, 2024).

WTK-LED Climate was released in 2024 as the wind climatology component, developed by Argonne National Laboratory, of the WIND Toolkit Long-term Ensemble Dataset, a wind resource dataset led by the National Renewable Energy Laboratory. The climatology dataset uses an accelerated version of RRTM for general circulation models (RRTMG) for the radiation schemes, ERA5 for the initial and boundary conditions, and YSU for the PBL scheme (Draxl et al., 2024). WTK-LED Climate covers North America at a 4-km horizontal spatial resolution and 1-hr temporal resolution for the 20-year period of 2000-2020. Through the WindWatts web application (NREL, 2024), users can access WTK-LED Climate long-term average and monthly wind speeds at seven output heights between 30 m and 140 m.

Across the 66 observation sites, we extract the long-term average wind speed estimates from GWA3 and WTK-LED Climate at the surrounding output heights to the measurement heights and adjust them to the measurement height with the power law (Eq. 4 and 5). We determine the bias and relative error as in Eq. 2 and Eq. 3, keeping in mind that this comparison involves the different definitions of "long-term" that a user of the GWA3 and WindWatts web applications would experience (10 years for GWA3, 20 years for WTK-LED Climate and varying lengths for the observations as shown in Figure 2b). For our observation sites, the median bias magnitudes for the WTK-LED Climate and GWA3 long-term estimates are 0.71 m s-1 and 0.36 m s-1, respectively, and the median relative errors are 12.8% and 6.1%. On average, using even one month of wind speed observations to create long-term MCP-based wind speed estimates provides an improvement in accuracy over the long-term estimates provided by the higher-resolution WTK-LED Climate and GWA3 (Figure 13, Table 2)."

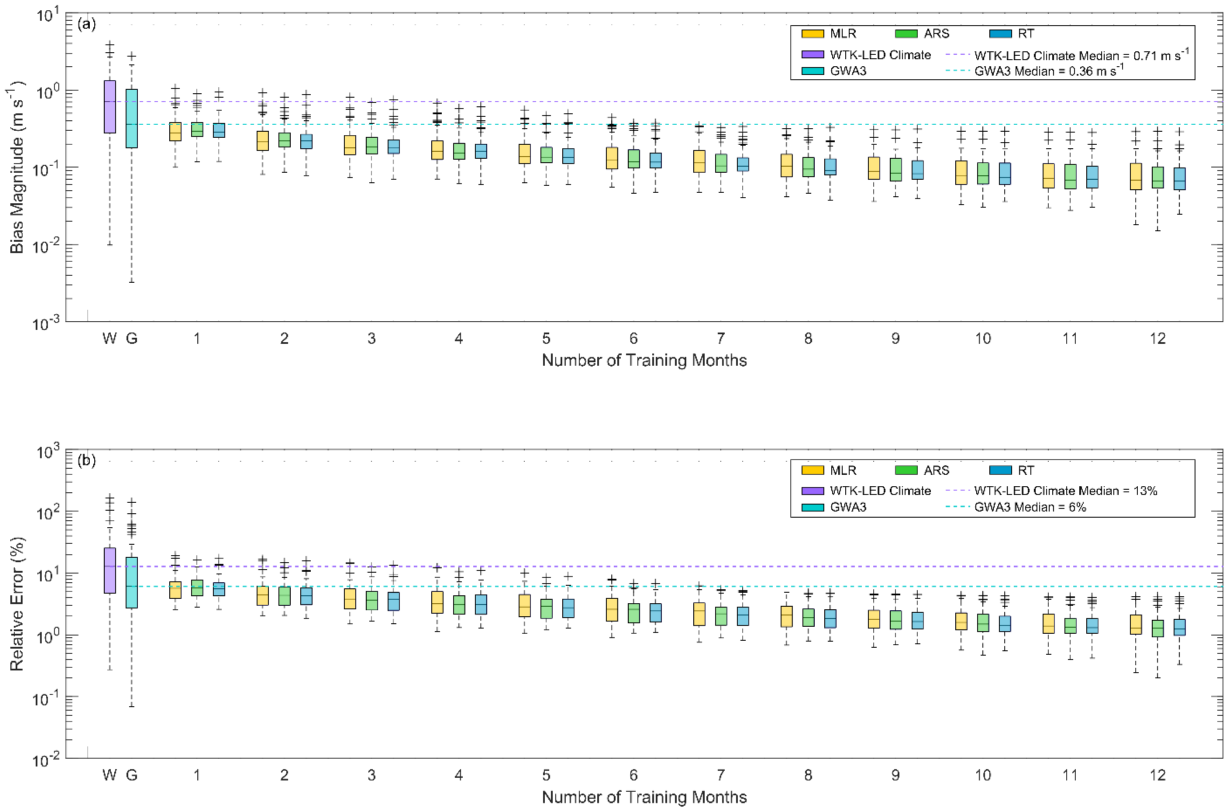

Figure 1. Average long-term wind speed (a) bias magnitude and (b) relative error across 66 sites based on the long-term estimates of GWA3 and WTK-LED Climate and the MCP-based estimates from ERA5 and short-term observations.

2. Overall, the paper is well written and methods are clear. The paragraphs on MCP and quality threshold (L197-207 + L223-235) are however difficult to follow in detail. As an example, it is not fully clear weather the authors exclude MCP-simulated wind speeds for periods with missing observations when computing means. A shorter, clearer and stepwise description would be better. From my understanding

a. The mean of the full observation period is taken as ground truth.

b. A model is trained on a short (few months) period

c. The model is used to predict the full observation period.

d. Mean statistics are computed and compared to ground truth. Samples with missing observations are excluded from the calculations.

Thank you for your help in adding clarity and reducing wordiness to the discussion of the MCP process! We have modified Lines 208-217 to read:

"As an initial test of the performance of the algorithms and reference variable combinations, we develop ensembles of MCP-based long-term wind speed estimates at each measurement site using consecutive 12-month training periods, according to the following steps:

1. Establish that 75% of the observations in each month in the training period are available after applying the quality control checks discussed in Section 2.1 (all 66 observations utilised in this work have average and median monthly data recovery and quality rates exceeding the 75% threshold).

2. A model is trained on temporally aligned observation data and reference data during the training period.

3. The model is used to predict the full observation period (Figure 2b).

4. Performance statistics are computed with respect to the observations (Table 1). Timestamps with missing observations are excluded from the statistics."

And Lines 233-236 have been reworded as such:

"The months-long analysis follows the same ensemble formula as the 12-month exercise, just with shorter consecutive training periods. Across the measurement sites, calendar months in the spring and fall had the most single instances of ≥75% data recovery and quality, followed by summer, and lastly winter (Figure 4). Median measurement data recovery and quality percentages according to calendar month ranged from 99.2% (December) to 99.7% (May) (Figure 4)."

3. One of the main reasons why MCP-methods will work well or not on short measurement series is the seasonality of the errors in the reference series. It is not a problem per se if the measurement period is not representative for the long-term as long as the reference series accurately captures the variations. As an extreme example, with perfectly correlated observation and reference series, one observation sample is enough to estimate the true long-term mean wind speed. In practice, as is well demonstrated in the manuscript, the risk of getting poor results are higher if measurements are taken during e.g. low-wind periods. But these phenomena should not be confused (as in e.g. L24-27, L440-442).

Thank you for pointing this out. Per your helpful note, we have reworded to the following:

Lines 24-27: "However, in cases when the shortest observational periods (one to two months) used for correction are not well correlated with the overlapping ERA5 reference, the resultant long-term wind speed errors are worse than those produced using ERA5 without correction."

Lines 484-486: "While even one month of onsite wind speed measurements improves long-term wind speed estimates on average, incorporating at least four months of onsite measurements is a better option to mitigate the errors that could occur if some of the measured and reference wind speeds during the measurement period are poorly correlated."

Lines 491-492: "However, MLR is the least risky algorithm given the possibility of poor correlations between the measurements and the reference data."

4. Energy results should really be in the results section (not in discussion). Since energy production is more important than mean wind speed for the intended application, it would be good to emphasize these results more (on the expense of mean wind speed), e.g. in discussion and abstract.

We appreciate this suggestion to emphasize the energy results, and Reviewer 2 also made this point. We have moved the energy results to a new results subsection: Section 3.4 "Implications for energy production estimates."

Additionally, we have added the following text to the abstract (Lines 32-36): “Translating the analysis to wind energy, median relative errors in the capacity factor are on average within 10% using one month of training. If the observation period used for correction is not well correlated with the reference data, however, misrepresentation of the observed capacity factor can be substantial. The risk associated with poor correlation between the observed and reference datasets decreases with increasing training period length. In the worst correlation scenarios, the median capacity factor relative errors from using one, three, and six months are within 47%, 26%, and 16%, respectively.”

5. It would be good to include descriptions and references of state-of-the-art MCP methods.

Thank you! We have added the following discussion and references on Lines 175-182:

“One of the advantages of utilising MCP for long-term wind resource estimation is the variety of algorithm choices, which range from simplistic linear regression to machine learning techniques that can be applied to link the short-term and long-term wind speeds. Early MCP methodologies focused on linear (as reported via Rogers et al., 2005: Derrick, 1992; Landberg and Mortenson, 1993; Woods and Watson, 1997; Vermeulen et al., 2001) and quadratic fits (as reported via Rogers et al., 2005: Joensen et al., 1999; Riedel and Strack, 2001). From there, distribution-based probabilistic techniques emerged (García-Rojo, 2004; Sheppard, 2009; Carta and Velázquez, 2011). With the onset of machine learning techniques came applications to MCP-based wind resource analysis, such as using artificial neural networks, support vector machine, and random forest to estimate long-term wind speeds (Díaz et al., 2017).”

Carta, J. A. and Velázquez, S.: A new probabilistic method to estimate the long-term wind speed characteristics at a potential wind energy conversion site, Energy, 36(5) 2671-2685, https://doi.org/10.1016/j.energy.2011.02.008, 2011.

Derrick, A.: Development of the measure-correlate-predict strategy for site assessment, Proceedings of the BWEA, 1992.

Díaz, S. Carta, J. A., and Matías, J. M.: Performance assessment of five MCP models proposed for the estimation of long-term wind turbine power outputs at a target site using three machine learning techniques, Applied Energy, 209, 455-477, https://doi.org/10.1016/j.apenergy.2017.11.007, 2018.

García-Rojo, R.: Algorithm for the estimation of the long-term wind climate at a meteorological mast using a joint probabilistic approach, Wind Engineering, 28, 213-224, https://doi.org/10.1260/0309524041211378, 2004.

Joenson, A., Landberg, L., and Madsen, H.: A new measure-correlate-predict approach for resource assessment, Proceedings of the EWEA, 1999.

Landberg, L. and Mortenson, N. G.: A comparison of physical and statistical methods for estimating the wind resource at a site, Proceedings of the BWEA, 1993.

Riedel, V. and Strack, M.: Robust approximation of functional relationships between meteorological data: alternative measure-correlate-predict algorithms, Proceedings of the EWEA, 2001.

Rogers, A. L., Rogers, J. W., and Manwell, J. F.: Comparison of the performance of four measure-correlate-predict algorithms, Journal of Wind Engineering and Industrial Aerodynamics, 93(3) 243-264, https://doi.org/10.1016/j.jweia.2004.12.002, 2005.

Sheppard, C. J. R.: Analysis of the measure-correlate-predict methodology for wind resource assessment, Thesis, California State Polytechnic University, https://scholarworks.calstate.edu/concern/theses/9593tx516 , 2009.

Vermeulen, P. E. J., Marijanyan, A., Abrahamyan, A., den Boom J. H.: Application of matrix MCP analysis in mountainous Armenia, Proceedings of the EWEA, 2001.

Woods, J. C. and Watson, S. J.: A new matrix method of predicting long-term wind roses with MCP, J. Wind Eng. Ind. Aerodyn., 66(2), 85-94, https://doi.org/10.1016/S0167-6105(97)00009-3, 1997.

6. Many different regression tree algorithms exist. State which one you use and give a reference.

We have modified Lines 194-195 to read: "In this analysis, the ensemble aggregation method used is least-squares boosting with 100 learning cycles per the Matlab algorithm fitrensemble (MathWorks, 2024)."

MathWorks: fitrensemble, https://www.mathworks.com/help/stats/fitrensemble.html, last access: 13 December 2024.

7. The quality control of measurements is performed automatically. This is fine, in particular given the large number of observations. Periods with disturbed observations might however persist after the control, an example is partial icing of the anemometers. This could maybe explain the poorer results in Alaska.

Great point. We have added the following to Lines 355-358: "A potential factor impacting the results for Alaska is the quality of the observations. While the automated quality-control techniques discussed in Section 2.1 remove periods of nonvarying wind speeds due to outages or icing, they may not capture more subtle impacts on the observations, such as partial icing of the anemometers."

8. L33-34: normally, one describes it as that reference data (e.g. ERA5) are used to correct a short measurement, not the other way around.

We have reworded the sentence as follows (Lines 38-40): "In the utility-scale wind energy industry, short-term (less than five years) wind measurements are temporally extended using long-term (decades-long) wind resource simulations to produce a long-term wind energy generation estimate at a site of development interest in an expedient manner."

9. L58-65 could be rewritten to sound less like a sales text for ArcVera.

We agree and have reduced the discussion on ArcVera's procedures to the following (Lines 64-66): "For example, ArcVera uses at least one full year of observational data to bias correct their high-resolution model output (ArcVera, 2023)."

10. Refs to fig 5c should be 5d (L357 + L428)

Thank you! We have made the changes to 5d and appreciate you catching this typo.

11. L398 "versus performing a bias correction" could be "versus estimating mean wind speed" (the term bias correction should not be reserved for mean wind speed estimates).

We have reworded per your recommendation, thank you.

Lines 400-402: “Additionally, the power production at the low wind sites will be dominated by the tail end of the wind speed distribution, leading to potential significant differences between the skill of the MCP algorithms in reproducing the highest percentiles of wind speeds versus estimating mean wind speeds, as in Sections 3.1-3.3.”

Reviewer 2

The authors have thoroughly and appropriately addressed the reviewer's comments. I therefore recommend the manuscript for publication and only suggest one smaller technical correction:

• Section 3.4: The heading currently reads "Discussion". I suggest renaming this section for two reasons. First, the title "Discussion" suggests that the discussion of the results is limited to 3.4. However, results are evaluated and discussed in other parts of section 3 as well. Second, new results are presented in subsection 3.4 (e.g. table 3 and figure 12). A new heading should refer to the topic of the sub-section – i.e. "Implications for Capacity Factors/Annual Energy" production or similar.

Thank you for this helpful suggestion and for reviewing the update manuscript! We have renamed Section 3.4 "Implications for energy production estimates." We have also moved the final three paragraphs of the manuscript under a new Section 3.6 "Recommendations and future work." Again, we appreciate your time and your thoughtful review!