Journal of Physics:
Conference Series

PURPOSE-LED PUBLISHING™

PAPER • OPEN ACCESS

Benchmarking Atmospheric Model Data for Wind Resource Assessment: A Comparative Study Across 170 Sites

To cite this article: Sam Williams *et al* 2026 *J. Phys.: Conf. Ser.* **3224** 022040

View the article online for updates and enhancements.

You may also like

- Avian whiffling-inspired gaps provide an alternative method for roll control
Piper Sigrest and Daniel J Inman
- The use of physics in the design and construction of roads
A C Whiffin
- Electron nuclear double resonance at 35000 Mc/s
R J Cook

SmarAct | group

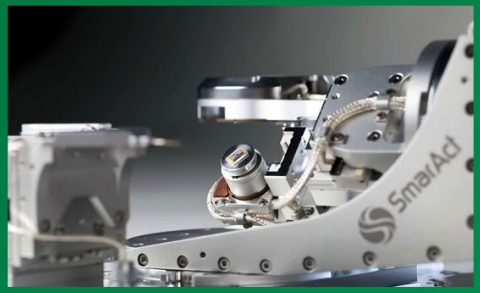
webinar

More details and registration

One sample, many dimensions: prospects for multimodal nanoscale characterization

Explore how precise sample manipulation enables complementary nanoscale techniques to work together on one sample

Live at 2 p.m. GMT/3 p.m. CET,
12 November 2026



This content was downloaded from IP address 31.151.17.97 on 13/11/2026 at 19:31

Journal of Physics: Conference Series **3224** (2026) 022040 doi:10.1088/1742-6596/3224/2/022040

Benchmarking Atmospheric Model Data for Wind Resource Assessment: A Comparative Study Across 170 Sites.

Sam Williams[1], Pim van Dorp[2], Gibson Kersting[3] and Sumitra Sharma[1].

[1]RWE Offshore Wind GmbH, Hamburg, Germany
[2]Whiffle BV, Delft, Netherlands
[3]RWE CleanEnergy LLC, Austin, USA

E-mail: sam.williams@rwe.com

Abstract. Wind resource assessments rely increasingly on numerically modelled atmospheric data, yet the accuracy of various models (from global reanalyses to high-resolution simulations) remains uncertain. This study benchmarks three fidelities of model data: i) global reanalyses (ERA5, MERRA-2), ii) a 2-km mesoscale model (Whiffle Meso), and iii) a nested mesoscale–LES approach (Whiffle Meso-LES), against 194 measurements at 170 wind farm sites worldwide. Wind speed, power, and wind direction metrics are used to determine a model's suitability for wind resource assessment, MCP, and AEP. The results find that reanalysis data have the largest spread and bias, with ERA5 consistently outperforming MERRA-2. Across all metrics, the largest step change in model performance occurs when downscaling from reanalysis to mesoscale, while additional improvements from mesoscale-to-LES downscaling are smaller and site-dependent. Terrain plays a dominant role: offshore and flat-terrain sites are already well characterized by large-scale models, while moderate and complex terrains see substantial error reductions from further LES downscaling, indicating that wind flow complexity drives model skill differences.

1. Introduction

Characterization of the wind resource is essential for the design and operation of wind farms. However, deploying measurement equipment can be difficult, expensive, and time-consuming. Instead, the wind energy industry relies on numerically modelled historical atmospheric datasets for i) initial prospecting of new wind farm locations, ii) correction of short-term measurement data to account for long-term interannual variability and iii) production planning of future power generation. There are many commercial and public datasets available, and it is essential that the optimum dataset is selected for each use case. However, the wind energy industry has a poor understanding of how these different datasets compare, and their respective strengths and weaknesses.

Previous studies have compared modelled data sets to wind measurements [1–8]. However, these previous studies have faced two limitations: i) A small sample size of the measurements reduces the statistical significance of the findings. The intense local and temporal variations in wind speed can lead to large errors between model and measurement. For a small number of

Content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

sites, the inherent uncertainty of model performance may lead to incorrect conclusions. ii) the small geographic distribution of measurements. Global and regional mesoscale models use parameterisations to represent sub-grid scale atmospheric processes. An important parameterisation for resolving the wind at wind turbine height is the model's boundary layer scheme. Boundary layer schemes are known to differ in their performance in different atmospheric regimes. These limitations make it difficult for a wind energy developer to understand how the uncertainty or bias in numerically modeled data propagates into AEP assessments.

The above limitations are addressed by conducting an extensive benchmarking of model data to wind measurements. This study leverages the benefits of a multi-year big-data project to digitalise, automatically clean and harmonise every wind measurement deployed by RWE globally. This database contains 170 sites with over 194 independent wind measurements. Each site contains at least one wind measurement with over 1 year of recorded data. Sites are globally distributed in all of the key wind energy markets as shown in Figure 1.

2. Methodology

2.1. Measurement Data

This study uses data from 170 site locations distributed in key areas of wind energy development (Fig. 1). Each site location has a minimum of one measurement device within a 30 km radius. Each measurement is treated independently in this study. Atmospheric parameters relevant to resource assessment are captured using three types of measurement devices: (i) Met Masts, (ii) LiDARs, and (iii) floating LiDARs (FLiDARs) (Fig. 2 (right)). Each measurement device has been deployed for a minimum of 12 months, measuring at least wind speed and wind direction. Measurement data are collected and cleaned according to industry best practice. Time stamps with duplicated entries, erroneous data, low availability, or low carrier-to-noise ratios are removed. The majority of the analysis was conducted at 100 m height ($\pm$5 m) for both measurement and model (except MERRA-2 which is at 50 m). A number of met masts were shorter than 100 m and analysis was conducted on the highest matching height for model and measurement. The effects of specific weather events (such as icing) are flagged, and distortion from Met Masts corrected for. Data is collected as 10 minute average values. A timestamp of 12:10, contains data from 12:00 to 12:10. 10 minute values are then averaged to hourly values. For an hourly average value to be valid, all six 10 minute timestamps need to also be valid for that hour. Topographic complexity for each location was calculated using the standard deviation of the terrain height for the LES domain. The terrain complexity and tree cover at each site is shown in Fig. 2 (left).

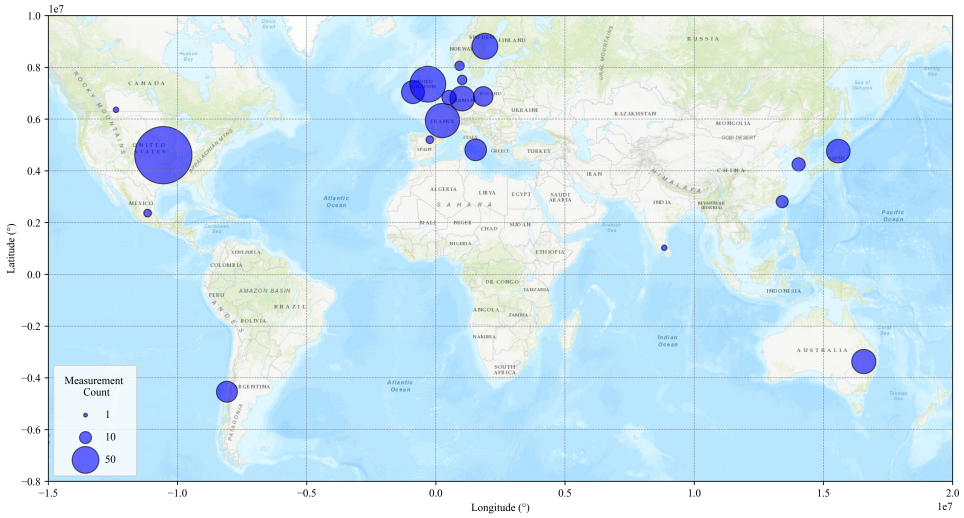


Figure 1. Global distribution of 170 measurement locations. Data are aggregated at the country level, with site coordinates averaged. Bubble sizes represent the number of measurement counts per country. The basemap is provided by Esri WorldTopoMap.

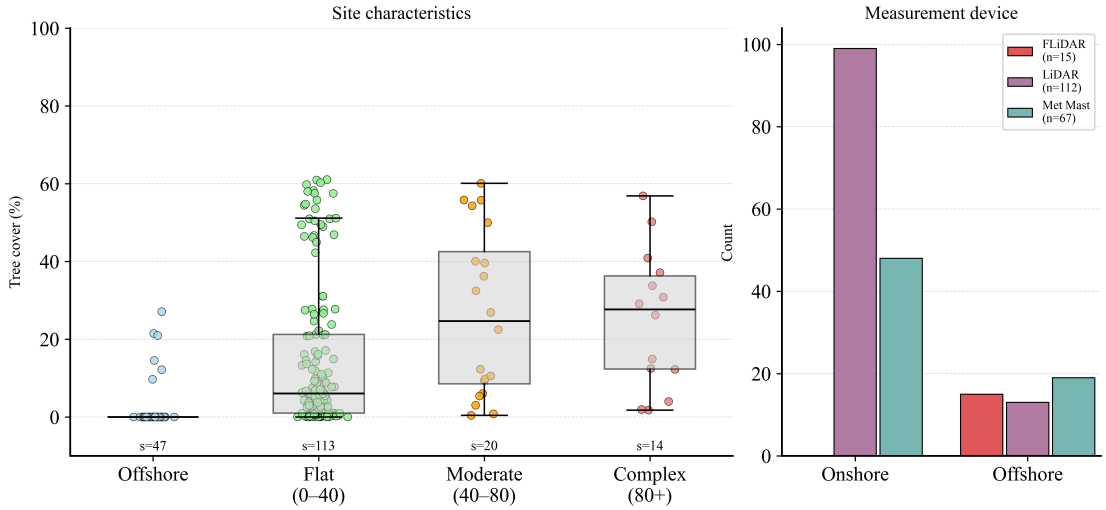


Figure 2. Overview of site characteristics and measurement device. **Left:** Distribution of site-level tree cover percentage by terrain class. The tree cover percentage was determined using [9]. Sites are categorised into terrain complexity based on the standard deviation of terrain height (m) of the LES domain. The number of sites is s. **Right:** Counts of measurement devices across all sites by type.

2.2. Numerical Modelling

The reanalysis datasets are ERA5 [10] and MERRA-2 [11]. The atmospheric model used is a coupled mesoscale-LES code called *ASPIRE*, developed by *Whiffle BV*. In this study, *ASPIRE* is run in two modes: mesoscale-only and meso(scale)-LES. The mesoscale-only simulations, also known as the *Whiffle Wind Atlas*, use ERA5 reanalysis data as boundary conditions, and

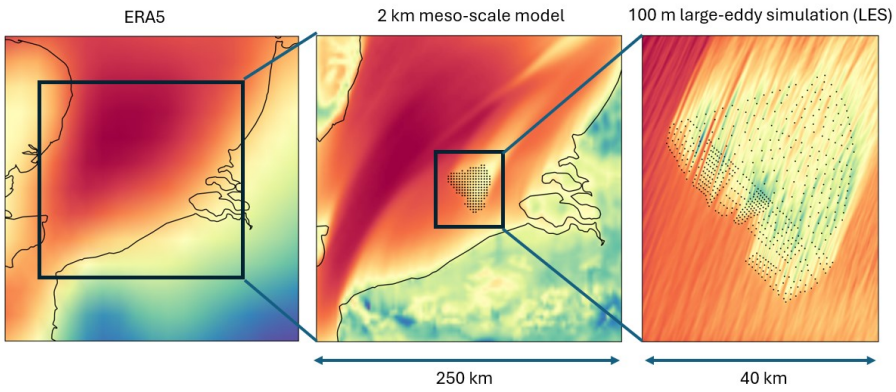


Figure 3. Nesting of *ASPIRE* grids across scales: ERA5 reanalysis (left) provides boundary conditions for a 2 km mesoscale simulation (center), which in turn drives a 100 m large-eddy simulation (LES; right). The domains progressively zoom in from regional (250 km) to local (40 km) scales, enabling detailed resolution of fine-scale atmospheric processes.

have been produced for all areas of wind energy interest for 2005–2025 for hourly atmospheric variables at a horizontal resolution of 2 km. In the mesoscale-only simulations turbulence is completely parameterized. The mesoscale-LES simulations add a nested LES domain at 100 m resolution into the mesoscale domain. The resolution of the LES is selected based on a cost-accuracy trade-off. Although 100 m resolution will be too coarse to resolve all turbulence down to the inertial range in all atmospheric conditions, grid sensitivity studies have shown the performance decrease with respect to higher resolutions is limited in case of real weather LES [12, 13]. In ASPIRE, case and time specific inflow turbulence is generated by concurrently running a number of periodic LES's around the boundary. The turbulent fluctuations generated by these concurrent simulations are prescribed as inflow turbulence for the main simulation. The concurrent simulations are driven by the mean profiles of the main LES simulation, ensuring that the resulting fluctuations are representative for the local atmospheric conditions. The mesoscale domain serves to bridge the gap between the coarse ERA5 horizontal grid ($\sim$ 30 km) and the fine LES grid (100 m) [1]. A typical nested model set-up is shown in Fig. 3. See [14] for a more detailed model description. For each site location, both a mesoscale-only and mesoscale-LES simulation were run for one-year. For each site location, a single year was determined based on the highest availability of valid measurement device timestamps. Each site location was used to generate a model domain that contained all of the measurement devices for that location with a minimum buffer of 5 km from the domain boundary. The minimum domain size was 250x250 km and 25x25 km for the mesoscale-only and mesoscale-LES respectively. The nearest grid point for each simulation is used to compare to the measurement device. No interpolation or special selection is used.

3. Analysis Metrics

Model evaluation metrics were selected to reflect the distinct ways in which modelled data is used across the wind energy industry, including resource assessment, energy yield assessment (EYA), turbine and wind farm design, and operational forecasting. No single metric fully characterises model performance for all applications; instead, each metric emphasises different aspects of model fidelity that are relevant to specific wind energy use cases.

Metrics based on mean values, such as the mean bias error (MBE) in wind speed and the power

mean bias error (PMBE), primarily quantify systematic bias. These metrics are particularly relevant for long-term EYA, financial modelling, and bankability studies, where persistent over- or underestimation of wind speed or power directly affects predicted AEP and associated revenue estimates. Error-based metrics such as the Root Mean Square Error (RMSE) capture both bias and variability in instantaneous wind speed differences. RMSE is therefore more relevant for applications that depend on short-term accuracy, including operational forecasting, load analysis, and control strategy development. Elevated RMSE values may indicate deficiencies in representing transient phenomena such as gusts, ramps, or diurnal variability. Correlation-based metrics (r), assess the temporal correspondence between modelled and observed variability independent of absolute bias. While the Kolmogorov–Smirnov (KS) statistic evaluates the ability of models to reproduce the full statistical structure of wind speed variability. Both r and the KS statistic may be used to determine which datasets are best for long-term correction or Measure-Correlate-Predict of short-term measurement datasets. Power-based metrics translate wind speed errors into their nonlinear impact on turbine output via the power curve. Mean power metrics directly reflect implications for energy production, while analyses based on mean-scaled power more closely reflect errors that propagate through a long-term correction or MCP methodology. Directional metrics (Wind Rose Difference, WRD) address model performance in representing wind direction statistics, which are critical for wind farm layout design, wake interaction analysis, and sector-based EYAs. Taken together, the selected metrics provide a multidimensional assessment of model performance that aligns with the diverse requirements of wind energy analysis, from long-term climatological accuracy to short-term operational relevance.

3.1. Wind Speed Metrics

Let $\{u_i^{\mathrm{obs}}\}_{i=1}^n$ denote the measured (observed) wind speed time series and $\{u_i^{\mathrm{mod}}\}_{i=1}^n$ the corresponding modelled wind speed series at matched time steps. Overbars denote temporal means (e.g. $\bar{u}^{\mathrm{obs}} = \frac{1}{n}\sum_{i=1}^n u_i^{\mathrm{obs}}$).

Mean Bias Error (MBE). The bias in modelled wind speed was quantified as the difference between the mean modelled and mean observed wind speeds:

$$\mathrm{MBE} = \overline{u}^{\mathrm{mod}} - \overline{u}^{\mathrm{obs}}. \tag{1}$$

An ideal value is $\mathrm{MBE} = 0\ \mathrm{m\,s^{-1}}$.

Root Mean Square Error (RMSE). Overall agreement in wind speed magnitude was assessed using the root mean square error:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(u_i^{\mathrm{mod}} - u_i^{\mathrm{obs}}\right)^2}. \tag{2}$$

Lower RMSE values indicate improved agreement between modelled and measured wind speeds.

Pearson correlation coefficient (r). Temporal correspondence between measured and modelled wind speed variability was quantified using the Pearson correlation coefficient r:

$$r = \frac{\sum_{i=1}^n \left(u_i^{\mathrm{obs}} - \overline{u}^{\mathrm{obs}}\right)\left(u_i^{\mathrm{mod}} - \overline{u}^{\mathrm{mod}}\right)}{\sqrt{\sum_{i=1}^n \left(u_i^{\mathrm{obs}} - \overline{u}^{\mathrm{obs}}\right)^2 \sum_{i=1}^n \left(u_i^{\mathrm{mod}} - \overline{u}^{\mathrm{mod}}\right)^2}}. \tag{3}$$

An ideal value of $r = 1$ indicates perfect temporal agreement.

Kolmogorov–Smirnov (KS) statistic. Differences between the distributions of measured and modelled wind speeds were assessed using the two-sample Kolmogorov–Smirnov statistic:

$$D_{\mathrm{KS}} = \sup_x \left| F_{\mathrm{obs}}(x) - F_{\mathrm{mod}}(x) \right|, \tag{4}$$

where F_{obs} and F_{mod} are the empirical cumulative distribution functions of u^{obs} and u^{mod}, respectively. A value of $D_{\mathrm{KS}} = 0$ indicates identical distributions.

3.2. Power Metrics

Measured and modelled wind speeds were converted to power time series using a turbine power curve $P(\cdot)$:

$$p_i^{\mathrm{obs}} = P\left(u_i^{\mathrm{obs}}\right), \qquad p_i^{\mathrm{mod}} = P\left(u_i^{\mathrm{mod}}\right). \tag{5}$$

The bias in modelled power was quantified as the difference between the mean modelled and mean observed power: $\mathrm{PMBE} = \overline{p}^{\mathrm{mod}} - \overline{p}^{\mathrm{obs}}$. An ideal value is $\mathrm{PMBE} = 0$ W.

Scaled Power. To isolate temporal shape differences independent of mean bias, modelled power was mean-scaled to the measured power:

$$\tilde{p}_i^{\mathrm{mod}} = p_i^{\mathrm{mod}} \frac{\overline{p}^{\mathrm{obs}}}{\overline{p}^{\mathrm{mod}}}. \tag{6}$$

3.3. Wind Direction Metrics

Wind Rose Difference (WRD). Directional frequency distributions were computed by binning wind directions into K equal-width sectors (here $K = 12$) and normalising. Let f_k^{obs} and f_k^{mod} denote the normalised directional frequencies in sector k for measured and modelled wind directions, respectively. The wind rose difference was computed as:

$$\mathrm{WRD} = \frac{1}{2} \sum_{k=1}^{K} \left| f_k^{\mathrm{obs}} - f_k^{\mathrm{mod}} \right| \times 100\ \%. \tag{7}$$

A WRD value of zero indicates identical directional frequency distributions.

4. Results

Across all sites, the highest-resolution models most often achieved the smallest mean bias error (MBE) (Fig. 4). Ranking the best-performing source by smallest absolute MBE across 194 measurements, Mesoscale-LES most frequently achieved the lowest absolute MBE (91 measurements, 46.91%), followed by Mesoscale (63, 32.47%) and ERA5 (40, 20.62%). Violin and notched box plots of mean difference (MBE), RMSE and r show reanalysis with the largest bias and dispersion; ERA5 consistently outperforms MERRA-2 with narrower RMSE and higher correlation. Mesoscale downscaling recentres the MBE near zero and tightens the interquartile spread, while LES yields the smallest errors and highest correlations, indicating the best reproduction of temporal variability. KS statistics are consistent with these patterns, with progressively smaller distributional gaps relative to observations as resolution increases.

Splitting the results by Offshore and Onshore terrain complexity (see Fig. 4) reveals that the greatest bias reduction and decrease in model spread arises from mesoscale downscaling. Offshore sites exhibit relatively small MBE and limited spread across all modelling approaches, indicating that large-scale atmospheric processes dominate the wind regime and are already well represented in reanalysis and mesoscale simulations. In simple onshore terrain, the

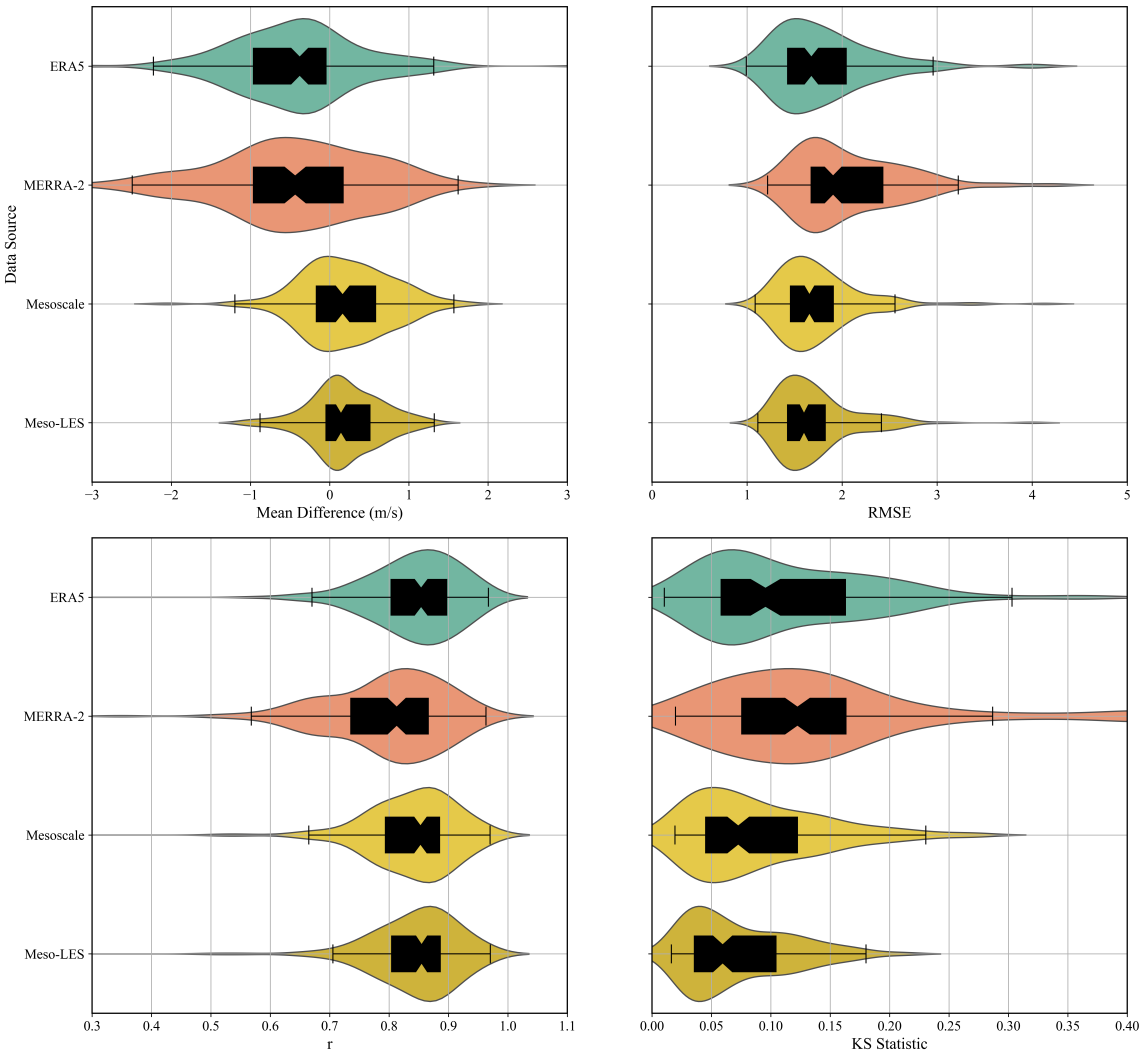


Figure 4. Comparison of wind metrics across model sources. Violin plots show the distributions of MBE ($\mathrm{m\,s^{-1}}$), RMSE, r, and KS statistic for ERA5, MERRA-2, Mesoscale, and Meso-LES models. Black boxplots overlaid on each violin indicate the median and interquartile range. These distributions summarise model performance across all 170 evaluated sites.

mesoscale–LES configuration adds little additional value beyond the mesoscale model. This suggests that wind conditions at these sites are primarily governed by synoptic and mesoscale flow features that are adequately captured at kilometre-scale resolution.

In contrast, moderate and complex terrain show a clearer benefit from progressive downscaling. Reanalysis consistently underestimates wind speeds, while mesoscale modelling mitigates but does not fully remove this bias. The mesoscale–LES configuration further reduces the bias and narrows the distribution, indicating improved representation of site-specific flow features. The wind regime in complex terrain is more strongly influenced by small-scale orographic effects, which are partially resolved in the mesoscale–LES but remain unresolved in the coarser models.

Extending the analysis to the power metrics delivers results consistent with the wind metrics. Downscaling from global reanalysis to mesoscale to LES yields monotonic improvements in accuracy for all metrics, with the greatest gains evident in more complex terrain (Fig. 6).

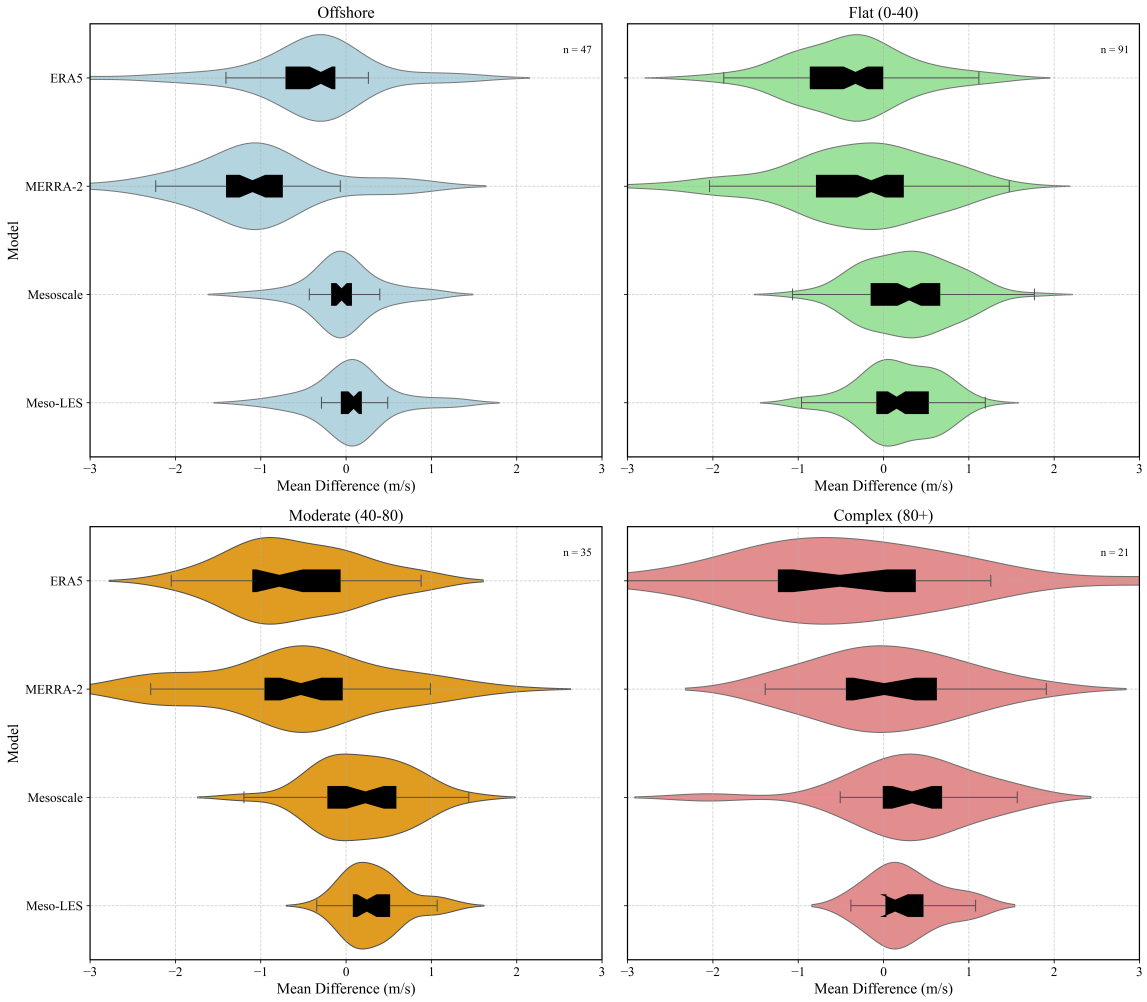


Figure 5. Distribution of wind speed MBE ($\mathrm{m\,s^{-1}}$). Top left shows all Offshore sites and then subsequently Onshore sites grouped into Flat (0–40), Moderate (40–80), and Complex (80+) terrain classes based on terrain height standard deviation. Violin plots represent the full distribution across sites, with overlaid boxplots indicating the median and interquartile range. Results are shown separately for ERA5, Mesoscale, and Meso-LES models. The number of measurements for each plot is denoted by n.

PMBE is predictably similar to the MBE results, with the largest reduction in model error coming from the downscaling step of reanalysis to mesoscale.

Many wind energy applications do not rely on using modelled data for absolute wind speed values, instead modelled wind speed and direction data is scaled by measurement data for long-terming or MCP analysis. The scaled power RMSE (Fig. 6(b)) highlights how well each model replicates the fluctuations in power (e.g. diurnal cycles, synoptic variations) independent of any mean bias offset. The same hierarchy of performance is evident. ERA5 and MERRA-2 have the highest scaled RMSE values on average, indicating that even after bias correction, there are significant errors in the pattern of power production compared to measurements. Again, the greatest reduction in error is found at the mesoscale downscaling. This suggests that at 2-km resolution, many fine-scale temporal variations in the wind are captured more accurately, reducing the mismatch in power output variability. The spread among sites also decreases,

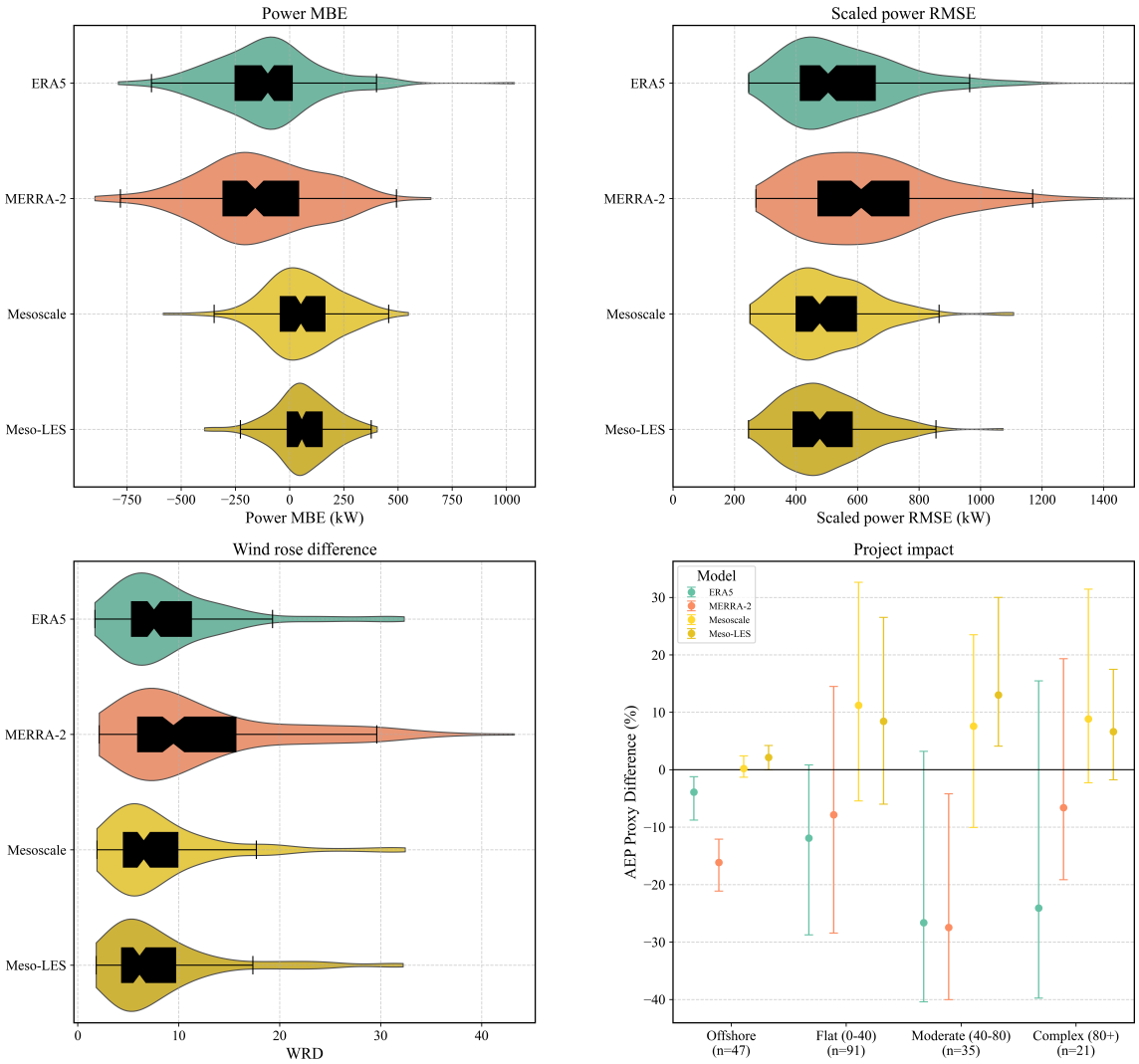


Figure 6. Top left, top right and bottom left show distributions of PMBE, scaled power RMSE, WRD across all sites. Bottom right summarises the median and interquartile range of an AEP proxy (percentage power sum difference relative to the reference) for offshore, flat, moderate, and complex terrain.

implying more consistent performance.

It is harder to determine the impact of wind direction uncertainty on wake modelling, EYAs and wind farm design [15, 16]. However, modelled wind direction is used to either scale or sometimes even replace a measured wind rose. The WRD between model and measurement (Fig. 6) quantifies the risk analysts take with this approach. The monotonic increase in model improvement extends from wind speed to wind direction, with the biggest reduction in error being from downscaling to mesoscale.

The previous analysis has consistently demonstrated the benefits of using higher fidelity model data for wind energy analysis. Analysis of 170 diverse sites worldwide allows for median errors on AEP for different site classifications (Fig. 6). The sum of the gross power is used as a proxy AEP (no dependence on wake model or wind direction), for analysts to determine the impact on typical projects. Notably, for Offshore sites little bias in AEP is shown for Mesoscale

and Meso-LES datasets. For all other applications, the importance of site calibration is evident with errors of $\sim \pm 10\%$ at a minimum.

Across all metrics considered, a clear pattern emerges: the dominant performance gain occurs when moving from reanalysis to mesoscale, whereas the additional benefit of LES downscaling (at 100 m) is smaller and strongly site-dependent. This suggests diminishing returns from further refinement beyond mesoscale for many locations, and implies that future improvements may increasingly depend on higher-quality driving reanalysis products (e.g. next-generation datasets such as ERA6).

5. Conclusions

The accuracy of 2 global reanalysis datasets, a 2-km mesoscale downscaling and a 100 m nested meso-LES downscaling were assessed against measurements across 170 sites worldwide. The systematic reduction of bias, variability error, and wind direction misrepresentation from reanalysis to mesoscale to Meso-LES has direct consequences for wind resource assessment and EYAs. In practical terms, every application using numerically modelled data would be improved by using mesoscale data over reanalysis. The mesoscale downscaling significantly reduces wind speed and power bias and improves representation of the wind speed distribution with respect to ERA5. Additional downscaling from the mesoscale with LES further improves error metrics across the board, especially for micro-scale dominated complex sites.

The scale and geographical distribution of the sites in this benchmark study allows for clear recommendations for the different numerically modelled datasets for wind energy applications.

- ERA5 outperforms MERRA-2 for offshore and flat sites. However, both global reanalyses are too coarse to accurately capture mean wind speed or wind speed distribution at moderate and complex onshore sites.
- Mesoscale downscaling is generally sufficient to capture mean wind speed within ~ 1 m/s, except for moderate and complex sites.
- Additional LES downscaling should be applied for moderate and complex sites to retain acceptable levels of uncertainty in modelled wind speed.
- For wake modelling and wind farm design, using mesoscale data significantly reduces the wind direction uncertainty.
- Downscaling shows clear benefits in capturing wind speed distribution and, related, mean power error. This translates into lower AEP (proxy) differences with respect to the reanalysis datasets.
- Projects that only utilize modelled data in their design will see errors of $\sim \pm 10\%$ (using Mesoscale or Meso-LES) with the exception of Offshore sites which show minimal AEP difference.

The current study focuses on prediction of wind speed, wind direction, and gross power. These metrics are most relevant for wind resource assessment and MCP. In future work, this study will be extended towards a full EYA benchmark and uncertainty quantification. As a next step, the multi-measurement sites within this benchmark study will be used to evaluate cross-prediction accuracy in a similar, statistically rigorous way. It is expected to see increased value of LES downscaling when considering relative predictions across sites since the synoptic and mesoscale modelling errors cancel out. As a final step to EYA, SCADA data from many wind farms will be involved to assess power and wake modelling accuracy. Finally, future improvements in atmospheric modelling skill are likely to depend increasingly on higher-quality large-scale boundary conditions (i.e. next-generation reanalyses, e.g. ERA6), as downscaling performance is ultimately constrained by the quality of the driving reanalysis data.

References

[1] Storey R and Rauffus R 2024 Mesoscale-coupled large eddy simulation for wind resource assessment *Journal of Physics: Conference Series* **2767** 052040

[2] Witha B, Hahmann A, Sīle T, Dörenkämper M, Ezber Y, García-Bustamante E, González Rouco J F, Leroy G and Navarro J 2019 Report on wrf model sensitivity studies and specifications for the mesoscale wind atlas production runs (newa deliverable 4.3)

[3] Kantharaju J *et al.* 2023 *Journal of Physics: Conference Series* **2507** 012015

[4] Gutiérrez R, Browning K A, Parés-Sierra A, Vogel F, Puertas-Castillo J, Gallardo-Altamirano E and Vásquez C 2022 Evaluation of era5 and merra-2 reanalyses for wind energy applications in mexico *Wind Energy Science* **7** 659–681

[5] Murcia J P, Koivisto M J, Luzia G, Olsen B T, Hahmann A N, Sørensen P E and Als M 2022 Validation of european-scale simulated wind speed and wind generation time series *Applied Energy* **305** 117794

[6] Gruber K, Regner P, Wehrle S, Zeyringer M and Schmidt J 2022 Towards global validation of wind power simulations: A multi-country assessment of wind power simulation from merra-2 and era-5 reanalyses bias-corrected with the global wind atlas *Energy* **238** 121520

[7] Eneau C 2023 Evaluation of modeled data for wind resource assessment (po170) Poster presented at the WindEurope Annual 2023 Conference, Copenhagen, Denmark URL `https://windeurope.org/annual2023/conference/ri-activities/poster-pitches/`

[8] Bruneau M 2020 *Accuracy of EMD-WRF reanalysis for wind power estimations, using WindPRO* Master's thesis Lund University Sweden URL `https://www.iea.lth.se/publications/MS-Theses/Full%20document/5436_full_document.pdf`

[9] NASA MEaSUREs Program 2015 Global forest canopy cover (gfcc), 30 m, version 3 accessed January 10, 2025 URL `https://developers.google.com/earth-engine/datasets/catalog/NASA_MEASURES_GFCC_TC_v3`

[10] Hersbach H, Bell B, Berrisford P *et al.* 2020 The era5 global reanalysis *Quarterly Journal of the Royal Meteorological Society* **146** 1999–2049

[11] Gelaro R, McCarty W, Suárez M J *et al.* 2017 The modern-era retrospective analysis for research and applications, version 2 (merra-2) *Journal of Climate* **30** 5419–5454

[12] Postema B, Verzijlbergh R A, van Dorp P, Baas P and Jonker H J J 2025 Estimating long-term annual energy production from shorter-time-series data: methods and verification with a 10-year large-eddy simulation of a large offshore wind farm *Wind Energy Science* **10** 1471–1484, supp. URL `https://wes.copernicus.org/articles/10/1471/2025/wes-10-1471-2025-supplement.pdf`

[13] Baas P, Folkersma M and van Dorp P 2026 On the validity and value of real-weather, hectometric-scale large-eddy simulation for wind energy applications *Submitted for the Torque 2026 conference, IOP Publishing*

[14] Baas P, Verzijlbergh R, van Dorp P and Jonker H 2023 Investigating energy production and wake losses of multi-gigawatt offshore wind farms with atmospheric large-eddy simulation *Wind Energy Science* **8** 787–805 URL `https://wes.copernicus.org/articles/8/787/2023/`

[15] Gaumond M, Réthoré P E, Ott S, Peña A, Bechmann A and Hansen K S 2014 Evaluation of the wind direction uncertainty and its impact on wake modeling at the horns rev offshore wind farm *Wind Energy* **17** 1169–1178

[16] Chang L H 2024 *Propagation of Wind Measurement Uncertainty in Wake Modelling* Master's thesis Delft University of Technology Delft, The Netherlands