Authors’ Response to Reviewer 2

***Reviewer’s Comment #1*:** In my view, the paper should be subject to major revision.

The study sets out an objective of very high practical importance for wind resource analysis: investigate whether short-term wind speed data from WMO weather stations realistically represent the wind speed statistics. Observations frequently cover a limited time span, which may introduce substantial errors in the estimation of several important parameters of the wind distribution, like its time variability. Therefore, guidance for selecting adequate datasets of adequate time spans is highly relevant.

***Response to Comment #1*:** We appreciate the recognition of the practical significance of our study. We sincerely thank the reviewer for the thoughtful and constructive evaluation. We acknowledge the recommendation for a major revision, and in response, we have conducted a comprehensive revision of the manuscript. Specifically, we have addressed all specific reviewer comments in detail; improved the clarity and transparency of our methodology and assumptions; added new comparative analyses (e.g., with continuous sampling and Wind Atlas data) to strengthen the generality and relevance of our findings. We hope these revisions address the reviewer’s concerns and contribute to a significantly improved manuscript.

***Reviewer’s Comment #2*:** In my view, the applied method is not adequate for this aim (“provide guidance for selecting adequate datasets of adequate time spans”). From the introduction and the description of the objectives of the study, it seems that datasets with different lenghts (720 hours to 6 years) would correspond to series of consecutive hours, so that a 720 hours dataset would correspond to 1 month of consecutive hourly data. In this way, actual observations of different lenghts would be replicated.

But from the description of the method (random sampling) it seems that the authors randomly select 720 separate hours within the full 16-year time series. Random sampling over a long time span of 16 years has much less practical relevance, due to the fact that actual station observations will be based on continuous (or nearly continuous) datasets, not on random measurements with random and large gaps between them. The small data gaps typically found in useful observations do not change this, as the full time span of the observations will still be limited.

For example, a continuous (or nearly continuous) dataset of a few months will not cover adequately the seasonal variability of the wind, but if the data are randomly selected over a large period the seasonal variability may be well represented, as indicated by the results of the study. From a practical point of view, a 720 hourly data dataset randomly selected from a multiyear full time series is not equivalent to a continuous (or nearly continuous) 720-hour dataset covering one month in total.

The result of the paper indicating that mean, standard deviation, and Weibull parameters, stabilize with relatively short records (~1 month of randomly selected hourly data) may not hold for 1-month continuous data.

The study builds on the work of Barthelmie and Pryor (2003), but that paper has a fundamental difference, as their sampling is not random. They use conditional samples to replicate data that could be obtained from remote sensing tools, which gives that paper a high practical relevance.

Response to Comment #2**:** We thank the reviewer for this insightful comment, which has helped clarify the interpretation of our methodology and its relation to our stated objectives. Our response is structured in two parts. First, we address the reviewer's remark that "Barthelmie and Pryor (2003) has a fundamental difference, as their sampling is not random." We would like to clarify that our random sampling method follows the same approach as in Barthelmie and Pryor (2003). The only technical difference lies in the choice of random number generator—the Park and Miller "Minimal Standard" in their study, versus a Permuted Congruential Generator in ours. As stated in their paper:

"*To examine the dependence of the distribution parameters on dataset density (i.e., number of observations in the time series) the dataset from Vindeby SMW <u>was randomly and multiply resampled</u> for a range of number of observations from n = 21 (assumed to be the lower bound on the dataset likely to be obtained using remote sensing) to n ~ 0.1 of the actual number of observations available from Vindeby for the entire data collection period (i.e., n = 10 000). <u>The resampling was undertaken with sample replacement (so the same observation could be selected multiple times</u> within one resampling group and/or in two or more of the 1000 resampling groups for each n) <u>using the Park and Miller "Minimal Standard" random number generator</u>. The results are presented in Fig. 2 for the four moments of the distribution and the Weibull parameters for 1000 resampling iterations for each n.*"

They further noted that their method did not retain seasonality or diurnal cycles, which motivated us to develop additional sampling strategies (i.e., diurnal-cycle-retained and seasonal-cycle-retained sampling) to investigate the effects of such temporal structures. They also acknowledged this limitation: "*It should be noted that a critical aspect of the applicability of the results (and uncertainty bounds) presented here is that <u>the datasets are randomly drawn</u> from the time series <u>with respect to seasonality</u>.*".

Second, regarding the reviewer's concern that random sampling does not reflect the practical reality of continuous datasets used in operational or observational settings, we fully agree. In response, and as elaborated in Comment #3, we incorporated additional analyses based on continuous sampling, using two stations with 46 years of uninterrupted hourly wind data. This allows for a direct comparison between random and continuous sampling, addressing the core concern raised here. We believe this addition substantially improves the practical relevance of the study and provides the guidance originally intended in our objectives.

Reviewer's Comment #3**:** If there is some misunderstanding of the method on my side, please clarify the description of the method. If I have correctly understood the method, I would suggest two possibilities:

1. - Change the objectives of the study to adapt them to the method. The practical relevance of the study would be substantially smaller, but it is in any case an interesting investigation about the characteristics of wind distributions. The comparison of random sampling to diurnal-cycle-retained and seasonality-retained data is really interesting in itself. Several aspects of the study offer new insights on wind distributions, like the different error margins analysed, and the comparative analysis of ERA data.
2. - Change the method to adapt it to the objectives. This could be done by selecting for example many different continuous datasets of the same length (e.g. 720 hour) within the full 16-year time series. The random sampling calculations already done could be retained for comparison purposes. A possibility to take advantage of the work already done, which is technically well performed, would be to use the results from the random sampling calculations to select the lenghts of the continuous datasets to analyse. That is, not all lengths until 6 years would be analysed with continuous datasets, only lengths from 720 hours to a threshold determined from the errors obtained in the random sampling calculations.

Response to Comment #3**:** We sincerely thank the reviewer for the thoughtful and constructive suggestions. In response, we followed the second suggestion proposed by the reviewer, namely, adapting the method to better align with the original objectives, by incorporating additional analyses based on continuous sampling.

To implement this, we used two additional stations located in Denmark and Scotland, each with 46 years of hourly wind speed observations, which allowed for sufficiently long and uninterrupted time series required for continuous sampling. The five Norwegian stations originally used in the study contain substantial data gaps and were therefore unsuitable for this type of analysis. Another reason for selecting these two additional stations is their geographical relevance: both are situated along the coast in the North Sea region, which is consistent with the coastal characteristics of the Norwegian stations.

The results of this additional analysis indicate that continuous sampling generally requires significantly longer time periods to achieve the same level of uncertainty in estimated distribution parameters, compared to random sampling. These findings reinforce the practical utility of random sampling when long, uninterrupted records are not available. Details of the comparison between random and continuous sampling have been added to the Discussion section of the revised manuscript. We also show the complementary contents below:

"*4.1 Sensitivity to sampling strategy and climatic non-stationarity*

In wind energy assessments, continuous sampling is more commonly used than random sampling because it preserves temporal structure and seasonal variability in wind speed time series, and most importantly, only long-term data are not available. However, continuous sampling may also introduce systematic bias, particularly over short durations, due to temporal autocorrelation and underlying climatic non-stationarity. To investigate the extent of this effect and assess the generalizability of random sampling, we conducted a sensitivity analysis using 46 years (1979–2024) of hourly wind speed data from two coastal meteorological stations: Copenhagen Airport (061800-99999, Denmark) and Leuchars (031710-99999, Scotland). These sites were chosen for their long-term records and meteorological similarity to the five Norwegian locations analyzed earlier. Copenhagen station exhibits a long-term decreasing wind speed trend (Fig. S1), consistent with broader global observations (Zeng et al., 2019).

Our results show that continuous sampling generally requires significantly longer periods to achieve the same level of uncertainty in estimated distribution parameters compared to random sampling (Fig. 7). This discrepancy arises because random sampling draws from multiple years, thereby capturing a wider range of interannual variability and reducing exposure to temporal clustering. Consequently, the 90% confidence intervals (CIs) under random sampling are symmetric for all parameters, while under continuous sampling, only the CIs for mean wind speed, Weibull scale parameter, and power density are symmetric. Shape-sensitive parameters, including standard deviation, skewness, kurtosis, and especially the Weibull shape parameter, exhibit pronounced asymmetries under continuous sampling, particularly at short durations (<2 years). This suggests that the presence of systematic climatic anomalies in continuous subsets may bias shape estimation.

These findings support earlier recommendations by Murthy et al. (2017), who advocate using at least four to ten years of data for reliable wind energy assessments. Our results suggest that when using continuous sampling, at least five years of data may be required to achieve ±10% relative uncertainty in power density estimates, although this threshold is site-specific (e.g., Copenhagen station requires more than 10 years). We further recommend that random sampling be considered as a complementary tool to identify potential biases in short-term continuous assessments." (*Lines 355-385 in the clean version of the revised manuscript*)

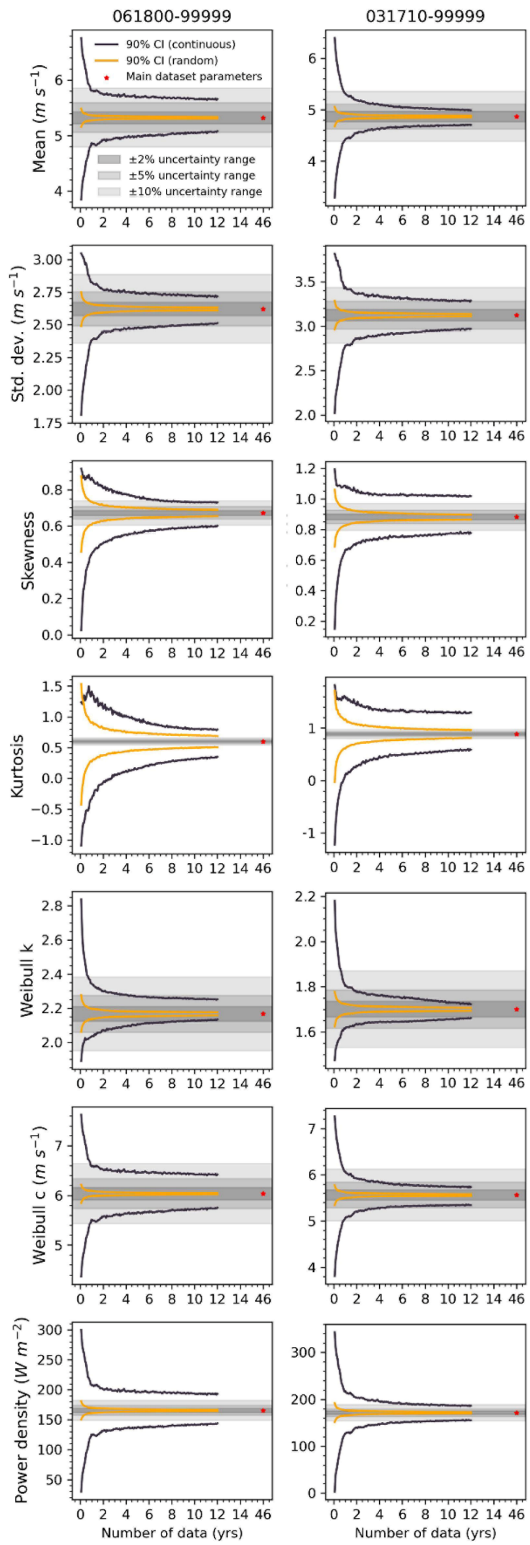

Figure 7: distribution parameters and Weibull power density derived from random sampling (orange lines) and continuous sampling (black lines), based on in-situ measurements from weather stations. Asterisks indicate values computed from the full 46-year dataset. Values for sample lengths between 14 and 46 years are omitted for visual clarity. Details of the experimental setup and sampling procedures are provided in the Methods section.

***Reviewer's Comment #4*:** A minor point that could be improved is the presentation of the results are the figures. There are too many frames in every figure that make it difficult to see the most relevant results. A selection of a couple of meteorological stations, that are representative of the whole set of stations, could perhaps be done. The full graphs could be moved to the supplementary information section. Also, the 3 curves limiting the 90% confidence intervals are difficult to distinguish, as they largely overlap.

***Response to Comment #4*:** Thank you for this helpful suggestion. This issue was also raised by Reviewer #1, and we have made corresponding revisions in the manuscript. Specifically, we now present only three key variables—mean wind speed, Weibull scale parameter (*c*), and power density. The remaining variables have been moved to the Supplementary Materials to reduce visual clutter and enhance readability. We also revised the color scheme and line styles to improve the visual distinction of the 90% confidence interval curves, which previously overlapped and were difficult to differentiate. An example of the revised figure is provided below, as it appears in the revised manuscript.

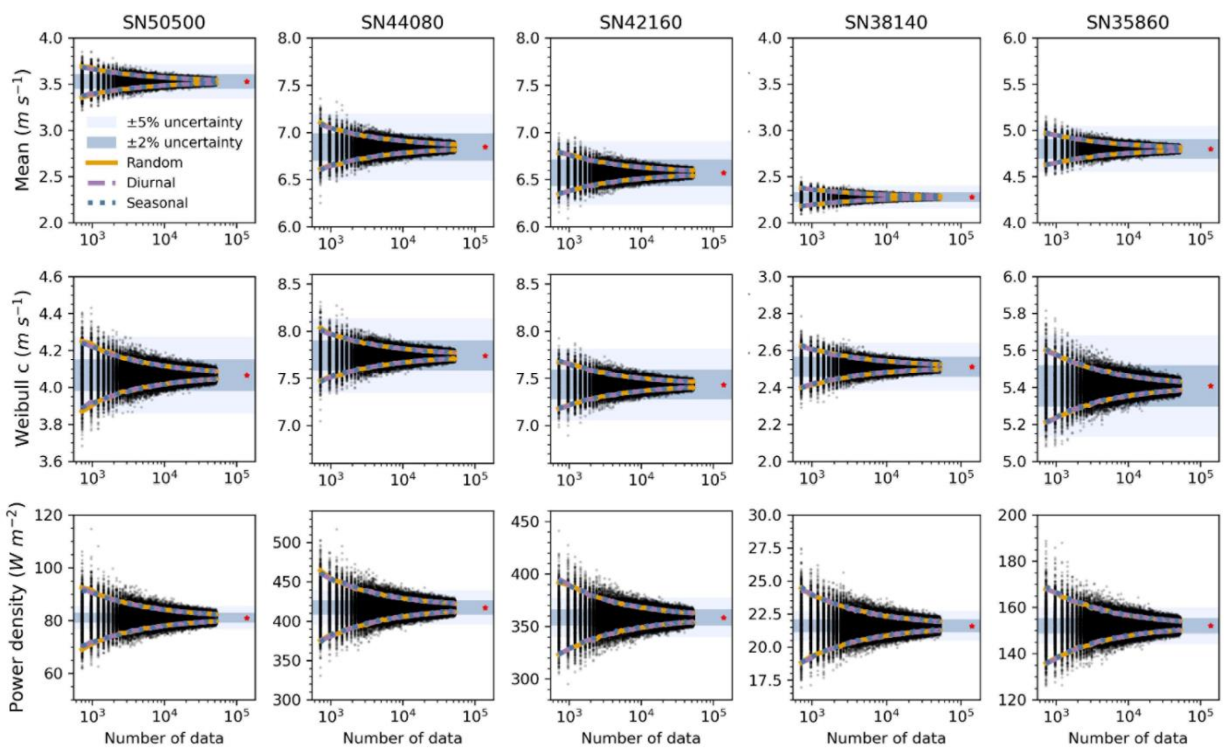

Figure 2: Estimates of mean wind speed, Weibull scale parameter, and power density from three sampling strategies, based on in-situ observations from five Norwegian stations. The 90% confidence intervals (CIs) are shown for each sampling method: random (orange), diurnal-cycle-retained (purple dashed), and seasonality-retained (blue dotted). Each black dot represents a parameter estimate from a single sampling realization of random sampling; corresponding realizations for the other two methods are not shown. Sample sizes range from 720 to 52,560 (30 days to 6 years), increasing in 240-hour (10-day) increments, with 1,000 realizations per size. Red asterisks indicate the

reference values from the full 16-year hourly dataset (see Table 2). Shaded areas represent ±2% (dark blue) and ±5% (light blue) deviation ranges from full-series values. (*Lines 220-228 in the clean version of the manuscript*)