WES Response to reviewers

Eamonn Organ (on behalf of co authors)

August 2026

We thank both reviewers for their comments. Here we discuss what we have updated to reflect the reviewer comments, as well as planned changes.

1 *Reviewer 1 Comments

From what I can tell, you are training with 10 m observations using the year 2023 (Line 112) and NEWA using the year 2018 (Line 136) to predict hub height winds at the wind farms over an undefined (as far as I could tell?) period of time for validation. A year or multiple years? Or less? All I can find are several figures showing 5 days, but I'm assuming the metrics in the results tables reflect longer periods of time than that? What is the data recovery and how does it vary from wind farm to wind farm? Would you expect the performance to improve if you used the consistent time periods across all datasets? Were 2018 and 2023 average, high, or low wind resource years in Ireland?

We have clarified the time periods used for the study. The height extrapolation model is trained on historical reanalysis model (2018 in the first draft) where modelled hub height wind speeds exist. We extend the training period to multiple years, to reduce potential biases and overfitting to patterns in one specific year.

Once the model is trained and model parameters are learned, the model can be used in any time period to predict hub height wind speeds where the input variables are available (essentially inputting real time meteorological data into a trained model). While one of the planned use cases is real time operational use, the testing period is after the training period to simulate real time scenarios (2023 in the original submission). We extend to multiple years which are available in the wind farm data. Testing over multiple years allows for more robust analysis on seasonal accuracy.

A brief image and analysis of the wind speeds in the training and validation years is also added.

Models are fitted for each calendar month, but there is no discussion in the results on how the proposed technique performs seasonally (or diurnally). As you mention that grid operators could be interested in this technique, such analyses are critical.

An additional table with monthly RMSE, correlation and bias is included and discussed in the results section.

Validation using the fastest wind speed in the farm is a simplistic approach but could be justified with more information. Assuming the farms are large(?), what are the impacts on the results given that your downscaled model is 250 m resolution? What location in the farm are you validating against? The center, despite the implication that the fastest wind speeds will occur at turbines at the edges of the farms typically?l.

The precise location used for prediction may influence results through two mechanisms. First, the GWA-derived climatological covariate which defines the long term mean varies spatially across the wind farm footprint. Second, the spatial GP prediction which veries over time depends on the target location. On the second point, wind farms are small relative to the estimated spatial correlation ranges of the GP, and wind speeds at turbine height are substantially smoother than at 10 m. Consequently, sensitivity to the exact prediction location is expected to be modest, with most location dependence arising from variation in the GWA climatological covariate rather than the GP interpolation itself. The average taken from the GWA will affect the mean of the predictions. This may be sensitive to the exact location used. As discussed in AC1, the location used for the prediction was the Lat and Long provided by the farm owner. Upon inspection this appears to be at a single turbine. we will perform sensitivity analysis on using different choices such as the approximate centre of the farm. As the GWA is defined on a 250m grid, an inverse distance weighting interpolation is used to the exact location of interest. Also we note that while the exact location can vary, as seen in Figure 9, the lack of terrain roughness at hub heights means the variation in mean wind speed is substantially lower than ground level.

On a similar note, how close are the wind farms in your sample to other wind farms? Figure 7 implies that the fastest wind speed in your farms could still be waked by other farms if close enough.

A table is added to the data section which includes information on neighbouring stations, per the last year in the sample (as Ireland's wind capacity expanded during the course of the testing period). We also acknowledge the limitation and include references on wind theft betweens farms. Further information on the number of turbines and the hub heights is also added to the table.

I think the wake analysis using the average wind farm speed should be removed entirely. Calculating error metrics to define the performance of your technique based on multiple unknowns is concerning. Line 428, in particular, hints of cherry picking when you don't establish an observation-based wake loss: "the 15% wake-loss adjustment yields the lowest RMSE and a mean residual close to zero, indicating stronger agreement with observed farm average wind speeds." This is acknowledged as a limitation in the conclusions section, which is appreciated, but doesn't warrant discussion in the results section without a more robust wake assessment. Why isn't ERA5 treated to the same wake loss assessment?

We acknowledge the comparison to the average wind speeds was a simplistic post-hoc analysis. Instead of including a table, with accuracy at different wake adjustments, we include a subsection with raw results and comparisons, and then some empirical results and summary statistics on the wake loss at various farms, as well as the dependence on farm size and wind speed, with reference to further papers on how wake loss could be applied.

Given the uninspiring validation of the technique (Line 417: "The proposed model achieves lower RMSE than ERA5 at three out of seven wind farms") and the very short results section (particularly if the wake assessment is removed as suggested), I don't think this manuscript is ready for publication in its current form. It's fine to try a technique and end up with unsatisfactory or inconsistent performance. But in order to be of value to the wind energy community, there needs to be a robust investigation into why the performance is what it is that is lacking in this manuscript. Is the technique struggling with faster/slower wind speeds? What is the performance like as a function of hub height or proximity to the coast or 10 m observation density or wind farm density? Does it perform better in winter versus summer, or night versus day?

We acknowledge that although direct comparison with ERA5 is important. The primary objective of the framework is not necessarily to outperform ERA5 under all circumstances, as our tool is available in potentially different use cases, but rather to provide hub-height wind estimates of comparable accuracy using observation streams available in real time with near instantaneous predictions. Nevertheless, we explored physically-motivated extensions incorporating real time variables related to atmospheric stability (e.g. wind/solar/wind gust), which yielded modest improvements in predictive accuracy.

The updated results section will also includes tables and discussion with different granular breakdowns, such as by and time of the day and year.

Please define bias somewhere. I think you are using observations – model, but other studies use model – observations to make it easy to show that the model is overperforming (positive bias) or underperforming (negative bias).

Bias has been clearly defined, as well as changed from the original paper to reflection convention

Line 41: "In the absence of wind measurements at turbine hub height, a common alternative is to vertically extrapolate wind speeds from observations at 10 m or other near-surface levels." Please include some references to support "it is common" and to identify which audiences do this. Large-scale wind farm folks will utilize tall towers and lidars. Many small project developers rely on simulated wind resource datasets.

The worded has been adjusted to more carefully reflect the situations where height extrapolation models are typically used, as well as links to a series of studies where a specific height extrapolation was required.

Section 2.1: I recommend removing the 5 observations that only have hourly resolution instead of introducing more error by linear interpolation. Given the fluctuations that occur in wind speed measurements over finer resolutions, these hourly observations are not comparable with other 18.

While we acknowledge hourly measurements lack high resolution variations, we refer to previous studies which find that due to the temporal persistence of wind speeds, even limited observations can lead to improved spatial interpolation. (e.g. https://doi.org/10.2307/2347679)

That being said to investigate the benefit, Sensitivity analysis will be added using 3 different approaches to deal with the hourly stations. Firstly these are removed and results with just 18 minute level stations is included. Secondly wind speeds are linearly interpolated, similar to the previous paper. Thirdly, a joint spatio-temporal interpolation strategy is used, which interpolates within the hour considering both observations on the hour, but also the spatial trend of nearby stations.

To investigate if these approaches have any impact on model accuracy, a table with accuracy for each of the ten minute blocks is also added to the results.

Line 127: What do you mean by “most reliable?” Smallest magnitude bias? Best correlation to capture fluctuations in the wind?

The wording in this line is sharpened with the exact results described. Specifically it was highest correlation and lowest RMSE when compared to official meteorological stations.

Line 127: What do you mean by “most reliable?” Smallest magnitude bias? Best correlation to capture fluctuations in the wind?

2 *Reviewer 2 Comments

Handling local topography and coastal heterogeneity in the Spatial GP Model: Ireland features varied terrain and distinct coastal-to-inland transitions. Standard Gaussian Process Models risk over-smoothing local topographic acceleration or abrupt changes in boundary layer dynamics near coastlines. Could the authors clarify how terrain elevation and distance to coast are accounted for within the spatial GP covariance structure or feature set? If these are not explicitly included, please discuss how topographic smoothing impacts performance at elevated or coastal wind farm sites relative to inland stations.

We account for the effect of local topography or geographical features by utilising the Global Wind Atlas average as a covariate to predict the mean structure. We expand discussion around this in the Spatial modelling section. While features such as distance to the sea and elevation have an effect on the mean structure, in analysis we found the effect tends to be quite non linear, as well as dependent on nearby geographic features rather than the site specific values. Consequently we found it to be more informative to use the predicted mean from the global wind atlas (regressed against observed speeds), as it implicitly uses physical equations to model the effects of terrain and coastline of the winds path. We also add prediction maps at different heights to give an intuitive sense of the spatial variability captured by the Global Wind Atlas.

Sub-Hourly temporal variability & Downscaling Mechanics: 10-meter meteorological observations are typically recorded as hourly or 10-minute spot/averaged measurements, while reanalysis datasets (like NEWA) are provided at coarse hourly intervals. In Section 2.1, the authors perform linear interpolation on hourly station observations to bring them to sub-hourly/10-minute resolutions. Linear interpolation of wind speed smooths out turbulent fluctuations and sub-hourly wind variability, which artificially suppresses the true variance of hub-height wind estimates. Could the authors address how linear interpolation impacts sub-hourly variance and error metrics compared to sub-sampling at top-of-hour intervals? Furthermore, how does the framework handle non-stationary variance during rapidly changing weather conditions such as during the passage of a cold front?

We expand our results analysis to study the impact of downscaling hourly wind speed measurements, in particular by showing the results by time period. As discussed in the replies to reviewer one, we compare three separate approaches

- We utilise only the 18 stations that are available at the minute level.
- We repeat the existing approach where hourly stations are linearly interpolated to each 10 minute window, and then all 23 weather stations are used across all time points.
- We interpolate the hourly stations across all time points, however we instead use a spatio-temporal interpolation strategy. Instead of only interpolating between the top of the hour measurement, the spatio-temporal predictions provides the maximum likelihood prediction, considering both the hourly observations, but also the changes neighbouring stations.

With regards the use of non-stationary covariance functions, we expand discussion on this and acknowledge the limitations of using stationary covariance functions. In particular, the sparseness of the available observations, compared to the generally high paramaterisation needed for non-stationary structures, particularly if they are temporal varying, would mean inference and identifiability. Furthermore we cite previous studies that find non-stationary data often shows sufficient predictive performance with stationary covariance structures.

While the authors explained in AC1 why 2018 NEWA data was chosen to mirror operational lag, please briefly comment in the text on whether model sensitivity was tested across multiple reanalysis years to ensure the GAM parameters aren't overly tuned to 2018 atmospheric patterns.

Additional years are included in both training and testing to provide more robust models.