Referee report: WES-2026-120

Overall assessment

This manuscript presents a hierarchical Bayesian framework for combining 10-minute SCADA calibrations of the Jensen and Gaussian TurbOPark wake models across four offshore wind farms. It distinguishes farm-level and cluster-level parameters from predictions for an additional farm, and examines how the calibrated parameters vary with atmospheric stability. The distinction between a pooled parameter and a neutral-stability parameter is useful when comparing SCADA calibrations with values from idealised simulations.

I recommend major revision. The main concerns are the interpretation of the stability trends, the representation of uncertainty, and the influence of data selection and residual-mismatch weighting. The scope of the new-farm predictions also deserves attention.

1. Stability dependence under comparable inflow conditions

The trends discussed at lines 680–684 show an association between stability and the calibrated parameters. However, Section 5 also shows that the stability proxies vary with wind speed and direction, neither of which enters the stability-dependent central-parameter model explicitly. How much of the trend persists under comparable inflow conditions? A comparison within a few well-populated wind-speed and direction ranges would help answer this question. The interpretation should also account for this covariation.

2. Prediction for another wind farm within the offshore cluster

Lines 359–366 correctly state that the new-farm prediction assumes membership of the same offshore-cluster population, including the range of layout effects represented by the observed farms. A useful test would be to fit the model to three farms and compare its prediction for the fourth with a separate estimate from that farm's data, including its uncertainty. This would strengthen the evidence for transferability. The conclusions should distinguish model-based predictive uncertainty from demonstrated prediction accuracy.

3. Data selection and temporal dependence

The exclusion of rejected or curtailed timestamps at lines 122–126 is motivated by uncertainty in turbine thrust. The remaining question is how this selection changes the data available for inference. Please report the available and retained timestamp counts for each farm and wake model, together with the wind-speed, direction, and stability coverage before and after filtering.

Successive 10-minute calibrations may also remain correlated after accounting for stability. Has this been checked in the residuals of the fitted log-scale model? If appreciable correlation remains, its effect on the reported posterior intervals should be assessed. More widely spaced observations could provide a sensitivity check, with the spacing chosen from the residual correlation and the resulting sample coverage reported.

4. Interpretation of the log-Student-t model

The finite-variance statement at lines 337–340 needs to distinguish the logarithm of the calibrated parameter from the parameter itself. For the untruncated model in Equation 5, $\nu > 2$ gives finite variance for $\log \hat{\omega}_{p,i}$, not for $\hat{\omega}_{p,i}$. Accordingly, Equation 8 defines an approximate scale conversion: $\sigma_{\mathrm{tot},p,i}$ should not be interpreted as the standard deviation of the original-scale predictive distribution.

Section 6.1 states that positivity constraints are imposed on the central parameters. Please specify how non-positive new-farm draws are handled in Equations 20–22 as well. The sampling settings and numerical convergence summaries should also be given.

5. Influence of residual-mismatch weighting

The residual-mismatch contribution at lines 302–306 gives different observations and different uncertainty scales. Table 1 already compares the scale parameters with and without this term, and the text explains why the low-mismatch predictions in Figure 6 are narrower. What is less clear is whether the weighting also shifts the central estimates or changes the stability trends. Please compare those quantities between the two formulations. Showing how the residual-mismatch indicator varies with stability would help establish which conditions are given less weight.

Minor comments

1. The reference to "yield" at lines 626 should be made consistent with the fixed-wind-distribution wake-loss sensitivity study described in Section 6.3.
2. For the stability-dependent analysis introduced at lines 621–624, specify which mast-based and ERA5-based proxy series is assigned to each farm, and how these series are matched to the 10-minute SCADA timestamps.
3. The TI fit described at lines 155–158 is shown in Figure 3, but its mathematical form and fitting procedure are not given.
4. Alongside the distributions in Figure 5, discussed at lines 435–445, tabulate the posterior medians and 5–95% intervals of the cluster-level and farm-level central parameters. Table 1 currently gives the scale and tail parameters.
5. In the proxy calculation at lines 530–545, give the measurement heights and sea-surface-temperature source, and state how low wind speeds are treated in the bulk Richardson number. For the ERA5 calculation, identify the input variables used in Equation 32.