

Review of

A comprehensive evaluation and uncertainty quantification of offshore turbulence intensity from WRF mesoscale simulations

Ref.: WES-2026-103

Recommendation: Major Revision

The authors perform a series of mesoscale simulations using WRF with different implementations of PBL schemes and horizontal grid spacings in order to validate the model in terms of wind and turbulence (TI) for a full year over the North Sea. They show that, while wind speeds are generally well captured in all simulations (differing PBL schemes, differing grid spacing), turbulence (TI) is only relatively well captured when using the SGS model TKE output and converting to TI using a *specific* formulation for TI, when compared to on-site observations. This could be a useful study that offers guidance on which model formulations provide useful information over an important region for wind energy development. This manuscript is relatively well written and the conclusions are clear, however the Figures and captions should be improved to explain acronyms and better highlight results. In general, however, I question the methodology of splitting the TKE into resolved and sub-grid scale components, considering the majority of the simulations analysed have hardly any resolved turbulence. The results that follow require scrutiny due to the splitting and seemingly cannot be generalized to other situations or to simulations with higher resolution. Overall, I recommend that the paper be resubmitted subject to major revision.

Comments:

- The main conclusion is that TI derived from the parameterized SGS TKE is the best option, and that TI_{2/3 k} gives the lowest errors. But the SGS part of the turbulence is a function of the grid spacing. As the grid is refined, more of the turbulence is resolved and the parameterized TKE decreases. The authors show this in Table 7. Therefore, the optimal TI conversion factor is not constant. TI_{2/3 k} is best at 2-3 km, but TI_{2 k} is best at 111 m. The “best formulation” is therefore an empirical correction for a specific grid spacing value, not a property of the model or the physics. The results and the PBL scheme ranking are thus valid only really for the 2 km setup. The conclusions as stated in the abstract (“independent of ...”) cannot be supported, since the TKE-TI conversion must be re-tuned at every resolution/grid spacing. I recommend that the authors evaluate total (resolved plus subgrid) TI for each grid spacing, and report whether it is approximately resolution-independent. If it is not, the conclusions on the optimal TI formulation should be restricted to the 2 km configuration, and the claims of generality across resolution, model and PBL scheme should be removed from the abstract and conclusions. Furthermore, the section on the AROME comparison is carried out by comparing SGS TKE/TI calculated at 2 km grid spacing in WRF, but at 1.3 km grid spacing in AROME, which again is inconsistent.
- The comparison “SGS TI is significantly closer to observations than model-resolved TI” is not a like-for-like comparison. At 2 km, essentially all of the turbulence is in the parameterized part by construction. A large part of the paper highlights that the standard deviation of wind speed from a 2 km WRF simulation is about 90% lower than observed, and that TI computed from it therefore has a large MAPE. This is not really a finding. At grid spacings of 2 km the energy containing eddies in the boundary layer are not resolved, so the standard deviation of the model output cannot be a realistic (or even close to a realistic) turbulence estimate. The variability it contains is mesoscale variability, plus possibly numerical noise. The authors acknowledge this themselves (lines 43-44, 282-283), so reporting the deficit as a result of the evaluation is unnecessary (and completely expected).

- To my understanding, the prognostic variable in MYNN is QKE, which is $2 \cdot TKE$. Have the authors used QKE instead of TKE for the MYNN analysis? It may explain the larger TI results in Table 5 for the MYNN simulation.
- The authors note that lidar TI increases with height (and is much larger than mast TI), which is unphysical, and that lidar and mast TI differ systematically. That undermines using lidar to claim TI_{2k} (or any formulation) better matches larger heights. Either remove those conclusions/analysis or characterise lidar TI bias, maybe? Providing some more detail on the lidar scanning pattern and how TI was calculated from the lidar measurements would help to contextualize the bias relative to the mast TI.

Minor comments:

Figure 1: What is shown here in shading? Wind speed in 2013? Averaged over the whole year? It's not clear.

Figure 2,3: What are the acronyms in the label bars? Explain in caption. Choices of colours for the contours could also be improved, as they tend to blend together when the document is printed out.

Line 223-224: "four modelled ones", and in other sections. Please be clearer throughout which simulations are being described. Presumably the 2 km grid spacing simulations.

Eq. 3: Is there a 0.5 factor missing here? Please check that the results used the correct form.

Table 6: Title, RMSE in m/s for TI? (see also Table 8)

Table 8: Caption says 333 m, but table shows 111 m values.

Figure 10c: obslso?

Typos: "weather" -> "whether" (line 288)

"proofed" (line 288),

"Juli" (line 264),

"ways to low" (line 369),

"in generall" (line 385),

"is not fully captured by none of the three PBL schemes" (double negative, line 276).