Original comments; Author responses

The manuscript "Model sensitivity across scales: a case study of simulating an offshore low-level jet" by the authors P. Hawbecker, W. Lassman, T. W. Juliano, B. Kosovic and S. E. Haupt deals with the investigation of the sensitivity of mesoscale-to-microscale simulations of a specific case of an offshore low-level jet (LLJ) to varying sea surface temperature (SST) datasets. The authors aim to understand how the representation of physical characteristics of the observed LLJ changes between mesoscale and microscale simulations. Moreover, the authors investigate the question whether a better mesoscale setup leads also an improved result of the microscale simulation coupled to this mesoscale simulation.
The study shows for the specific case investigated that low-level shear and jet nose height are better represented in microscale simulations compared to mesoscale simulations. Specifically, low-level shear improves by reducing near-surface wind speeds and lowering the jet nose height to be closer to observations. The study highlights the challenges in predicting the mesoscale setup that will result in the best performance of the microscale model. While the best performing mesoscale setup (CMC SST data) ends up being the best performing microscale setup, the second-best performing mesoscale setup becomes the worst performing microscale setup. This suggests that the differences between mesoscale and microscale numerical methods and model setup are large enough that one of the best performers on the mesoscale may lead to one of the worst performances of the microscale model.

While the paper is overall well written and addresses with mesoscale-microscale coupling a hot topic in the field of wind energy research, my main concern refers to the question how generalizable the results presented by the authors actually are. I think it would strengthen the manuscript a lot if the authors could show that their conclusions from the single low-level jet case at one site can be transferred to other low-level jet cases at the same site and a next step also to low-level jet cases at another site. If no additional sites or measurements are included in the manuscript it should be mentioned more clearly that the reader should be careful to transfer the findings of this paper to other sites.

We would like to thank the reviewer for their comments and suggestions for this paper. We appreciate the need to further address the generalizability of these findings and have done so through adding the following text to the manuscript:
Line 69 of original manuscript:
*We note that the study considers a single case study for a specific topic; thus, it is unclear whether the resulting findings generalize to other cases and atmospheric phenomena. However, we explore fundamental differences in mesoscale and microscale simulation techniques that are generally applicable for other atmospheric studies.*
Line 375 of original manuscript:
*These fundamental differences emphasize that studies should generally use caution when assuming that mesoscale sensitivities will directly translate to the microscale when simulating atmospheric phenomena (such as low-level jets) that have known dependencies on model grid spacing and turbulence closure techniques.*

Original comments; Author responses

With that said, it is not feasible at this time to perform this analysis for another case. We have included this idea in the summary as future work. The following text has been added to the manuscript:
Line 383 of the original manuscript:
*Additionally, similar studies for additional cases, different atmospheric phenomena, and different parametric sensitivities will be required in future studies to determine when and where mesoscale sensitivity can directly translate to microscale sensitivity.*

We agree that it is possible that these findings are not completely generalizable to all situations. However, this is meant as a caution to people who are making the assumption that their mesoscale sensitivity study will in some way ensure high levels of performance on the microscale. It is possible, and is shown here, that depending on the metrics of interest and what is deemed important for a study, the best mesoscale simulation setup may not lead to the best microscale setup due to the differences in mesoscale and microscale numerics and turbulence closure.

I have the following specific comments on the paper:

1. Line 64: Please change “shear over the rotor swept” to “shear over the rotor swept area”
   Thank you for catching this mistake. It has been fixed in the revised manuscript.
2. Line 91: Please change “Initial and boundary conditions for model” to “Initial and boundary conditions for the model”
   Thank you for catching this mistake. It has been fixed in the revised manuscript.
3. Line 91: I ask the authors to provide a motivation for using the MERRA-2 data for providing the initial and boundary conditions. E.g., it would be an important information if previous papers have shown that this reanalysis dataset performs best in the region under investigation by the authors.
   Thank you for mentioning this. The following text has been added to the manuscript in original line 138 to provide reasoning to the choice of MERRA-2 as the boundary condition:
   *The European Centre for Medium-Range Weather Forecasts (ECMWF) v5 reanalysis (Hersbach et al., 2020, ERA5) dataset was also tested and performance with MERRA-2 was found to be slightly better for the specific case day considered here (not shown).*
4. Line 115: I ask the authors to add additional information on the correction of measurement data for tidal variation.
   We’d like to thank the reviewer for pointing this out. This statement is erroneous and needs to be fixed. The lidar data are not corrected for waves/tides as they are averaged over 10 minutes. When averaging over 10 minutes, it has been shown that correcting for waves is unnecessary for these purposes (see Figure 14c in the reference below).
   Krishnamurthy, R., García Medina, G., Gaudet, B., Gustafson Jr, W. I., Kassianov, E. I., Liu, J., ... & Mahon, A. M. (2023). Year-long buoy-based observations of the air–sea transition zone off the US west coast. Earth System Science Data Discussions, 2023, 1-53.
   The original sentence in question has been removed and the following text has been added to the manuscript on Line 115 of the original manuscript:
   *The data are available at 10-minute averaged output and are not corrected for wave or tidal*

*variation. This correction has been shown to have a negligible effect in offshore floating lidar measurements when averaged at 10 minute timescales (Krishnamurthy et al. 2023).*

5. Figure 2: I found it slightly difficult to determine the height of the maximum wind speed from this plot (is it really always at about 120 m as stated by the authors in the text?). I think it would be helpful for the reader if the authors added markers showing the position of the maximum at each time to this figure.
   This is a great idea and markers have been added to the figure to denote jet height. Additional text to explain this has also been added to the caption.
6. Line 130: Is a resolution of 10 m for the simulation of a stably stratified marine atmospheric boundary layer that shows an LLJ sufficient for an LES? I think it would be good to slightly lower expectations already at this point.
   This is a fair point and we agree that expectations should be properly set for the ability of the LES to fully resolve turbulence in a stable boundary layer with 10 m grid spacing. The following text has been added to the manuscript on Line 124 of the original manuscript:
   *Note that even with a $\Delta x$ of 10 m, we are likely not fully resolving the inertial subrange of the stable boundary layer and associated low-level jet (Beare and Macvean, 2004; Beare et al., 2006). Simulations at higher resolutions (down to sub-meter grid spacing) have not shown clear convergence in simulating the very stable boundary layer (Sullivan et al., 2016) and a small ensemble of cases at these resolutions is out of the scope of this current project. We do not anticipate the overall findings from this study to be impacted by the use of 10 m horizontal grid spacing as all simulations are equally impacted.*
7. Line 151: I ask the authors to please clarify whether the output that is produced every 10 minutes is instantaneous or time-averaged data.
   Thank you for bringing up this oversight. Yes, the model data is also averaged in time. The following line has been added to line 165 in the original text:
   *Thus, the simulation datasets are also averaged at 10-minute intervals for comparison with observations.*
8. Line 163: Please replace "closer" by "closure"
   Thank you for catching this spelling mistake. It has been fixed in the revised manuscript.
9. Line 178/179: The authors mention several times that it is computationally expensive to run the LES simulations. My suggestion is to provide information on the resources that were actually consumed for the simulations by the authors. How many core-h on what type of HPC infrastructure have been consumed?
   This is a very good point. The following has been added to the manuscript to provide evidence of the computational expense:
   *The simulations were run on the NSF National Center for Atmospheric Research (NCAR) Cheyenne supercomputer. To provide evidence of the computational expense of LES for real-data cases, the following information is provided. Each LES run required 1,296~cores running for between 11 and 12 hours of wall-clock time in order to produce 10~minutes of simulation. Thus, to run for the full 6 hours, 35 restarts were required to fit within the Cheyenne 12-hour job limit. In total, this results in between 513,000~to~560,000 core hours per LES simulation -- 3.6-3.9~million~core hours in total.*
10. Line 188/189: In recent years a number of criteria for detecting low-level jets have been suggested in literature. I think the authors should refer to these criteria.

We appreciate the reviewer's suggestion and have added the following text to the manuscript:
On line 114: "While there exist many techniques to detect low-level jets from wind profiles (see, for example, Piety (2005) and Hallgren et al. (2023)), the observed low-level jet was detected using the technique described in Debnath et al. (2021)."
On line 188: "The height of maximum wind speed is used to define the low-level jet height."

11. Line 199: From my point of view the observation of a too strong shear in the microscale simulations of the authors was to be expected. The chosen resolution is too coarse to really be an LES of a stable atmospheric boundary layer. I expect that this leads to a too low turbulence in the model. Thus, the shear becomes too large.
This is a good perspective to mention and we will include this as a possibility for the overestimation of low-level shear within the discussion. The portion of the paragraph starting on line 352 of the original manuscript now reads:
*This finding leads us to believe that one or multiple of the following scenarios are occurring:*
*– The mesoscale MYNN 2.5 PBL parameterization may overly mix the stable boundary layer*
*– The 1.5 order TKE sub-grid turbulence scheme on the LES domains misrepresents surface drag over the ocean*
*– The grid spacing on the LES domains is too large to resolve turbulence within the stable boundary layer, resulting in an overproduction of shear*
12. Figure 5: Please add markers that provide information on the core height also in this figure.
This is a great idea and markers have been added to the figure to denote jet height. Additional text to explain this has also been added to the caption.
13. Line 257/258: "The mesoscale domains befit from slightly under-predicting wind speeds below hub height and over-predicting wind speeds above" Doesn't this sentence contradict the result that the shear in the mesoscale model is low? (see e.g. figure 9) In Line 260 the authors state: "Analysing the time series of the ensemble mean of bias in low-level shear the mesoscale domains underpredict low-level shear while LES over-predict." This sounds contradictory to the sentence in line 257/258 to me.
This is admittedly a confusing sentence and the word "slightly" was meant to apply to both the underprediction and overprediction of wind speeds below and above the jet nose, respectively. The sentence has been reworded to be:
*While agreement between the observations and simulations is decent below 100~m, the mesoscale domains do not overpredict the wind speeds as much as the LES domains do above the jet height (Figure 8d), which reduces error in REWS.*
14. Line 268: "It is interesting to note that the mesoscale domains produce larger negative values of SHFX than the LES domains, which indicates more stable conditions" Couldn't this be checked by checking the profile of potential temperature? Another possible explanation is that the eddy viscosity could be overestimated by the mesoscale model. This would fit to the wind profiles having less shear.
In terms of the stability, we appreciate the suggestion to check the potential temperature profiles and we had done so in our analysis, but did not include the plots in the manuscript. Although it may be helpful to add to the manuscript,we  do not believe that adding an additional figure is necessary to make this point. The following sentence has been added in Line 262 of the original manuscript:

Original comments; Author responses

*This is reinforced when checking the potential temperature profiles in which the lapse rate near the surface for the mesoscale domains is stronger than that of the LES domains (not shown).*

With respect to eddy viscosity, this is indeed a possibility and we suggest this in the conclusions without explicitly mentioning eddy viscosity (Line 353 of the original manuscript). We agree that it is important to mention the mechanism that may be the cause of the over-mixing. The following sentence has been added to Line 353 of the original manuscript: *[T]he mesoscale MYNN 2.5 PBL parameterization may overly mix the stable boundary layer potentially as a result of over-predicting eddy viscosity…*