Reviewer comments on the paper

Assessing the Accuracy of a Three-Year High-Resolution Mesoscale Wind Farm Wake Simulation with Lidar and Satellite Radar Data

General comments

This paper evaluates the long-term capability of the WRF–Fitch WFP modeling framework representing wake effects from an offshore wind farm cluster in the Belgium–Dutch cluster. About two years of profiling lidar observations at several coordinates and SAR data from four selected dates are used for comparison.

There is indeed a literature gap on the evaluation of the cluster wakes simulated by WRF-WFP over more realistic meteorological conditions using observations. This is no easy task, as it is difficult to cover large areas for a long period of time with measurement devices. Thus, the research problem and gap are well stated. The text is very well written, and the figures are well made; some of them convey information very nicely and look great (Figs. 14 and 15). The methodological approach makes sense to me, with some reservations. First, evaluate the error metrics in detail to support the subsequent analyses of the wake effects.

However, my main concerns are the superficial interpretation of results, the limited engagement with the literature, and consequently, very mild conclusions for a study with such strong potential. The paper often remains at a descriptive level (“the wake extends”, “bias improves”) without explaining why certain behaviors occur or how they compare to previous WRF–WFP, LES, or observational studies. If the authors strengthen the physical interpretation of their results and relate them more thoroughly to existing literature, the manuscript could make a significant contribution.

Specific comments

1) Please provide details on the turbine types used, including hub height, rotor diameter, and ideally thrust/power coefficient curves. These curves can be placed in the Appendix. If some data are confidential, clarify which components are public and which are not.

2) A more detailed characterization of the sources of model error needs to be made. The paper clearly shows that the simulations with the Fitch WFP lead to improved metrics of wind speed and wind direction based on the lidar observations. In Section 3.2.1, wind speed metrics for several lidar locations and models are shown. However, these error metrics are aggregated for all wind directions, and so they represent simultaneously the effect of the WF simulation representing the undisturbed wind and

wake effects. This is a problem because wake effects can be misrepresented, both because of errors in the representation of the free wind and specificities to the WFP. This information is crucial to establishing what makes the model more or less accurate (inflow error, wake modeling error, spatial gradients errors). As we know, errors in the undisturbed wind speed propagate into errors in the wake effects. This piece of evidence will strengthen the discussions when a few selected cases are presented in the SAR evaluation (Fig. 15).

3) Many more discussions are needed. For instance, Section 3.3 only describes the results without a single reference to the literature. That is unfortunate, because Figures 14 and 15 are the apex of the paper in my opinion. It is a missed opportunity to compare your evidence to the literature and describe the contrasts. Please cite papers that did WRF-WFP vs. lidar and/or SAR observations. Resort to the good references you already have and consider including Platis et al. (2018) and Cañadillas et al. (2020), which are great papers on the same scope. You cited Ali et al. (2023) in the Introduction, but their work is never mentioned again. The reader wants to know how the representation of the wake effects compares with other studies that use idealized LES or mesoscale simulations with observations near offshore wind farms.

4) Following on the aforementioned point, Figures 10 and 11 are not being discussed as well, which is another missed opportunity to communicate whether your results make sense with the literature or not.

5) Figure 11: You state that "Notably, intra-cluster wake behavior differs under NE winds, with a continuous reduction in wind speed observed throughout the wind farm", and no explanation as to why this behavior takes place is provided. Why is it that the wind speed deficit is more flattened for the SW direction? As readers, we want these insights from you. Sometimes the most interesting aspects of the research appear in these details. What is the role of the different transects relative to the wake center? What is the role of the heterogeneous layout and density of the wind turbines?

6) L467–470: "Systematic biases in both wind speed and direction are evident, particularly in the wind speed metrics. However, it should be highlighted that the model is driven by ERA5 hourly data, which consistently introduce a negative bias, affecting the boundary conditions in WRF and propagating wind speed underestimations to the analyzed locations." And what is the effect of this modeling bias on the representation of the wake effects? This discussion is fundamental because sometimes you have a bias in the free wind speed that can actually improve, by accident, the validation metrics. This must be explained.

7) Many figures address modeling error metrics (Figs. 5–8), but none for the actual wake effects. I think we can increase the value of the paper if you use Figure 13 as a template and plot a wake-related metric (e.g. wind speed deficit). That template is great because it simultaneously separates the effects of wind speed classes and stability.

8) On the marine ABL being mostly unstable: “Similar conditions have been mentioned in the work of Rosencrans et al. (2024) analyzing the ABL stratification on the Rhode Island–Massachusetts area in a WRF execution from September 2019 to September 2020.” Actually, Rosencrans et al. show that unstable vs. stable are almost 50-50, with a few occurrences of neutral cases. I think this statement needs an adjustment. Pointing out that your case is more unstable is actually interesting and adds to the discussion.

9) You mention the Vollmer et al. correction in the Introduction, but it was not used in the paper. Can you comment on what effect the Vollmer et al. correction would have on your simulations?

10) The SAR analysis looks great. But why have only 4 dates been selected over 2 years?

11) Are the biases of SAR observations fixed or vary in space and time? Can you estimate the SAR bias at 10 m for your case? The agreement does look very good in the wake region (errors of about 0.5 m/s or less at times), but more information on the reliability of these observations, preferably citing the literature, can strengthen your claims that the model works well.

12) In L367–368 you state that “greater wind speed magnitudes are associated with more pronounced downstream wind speed deficits”. Why does the wind speed deficit increase with the wind speed in Figures 10–12?

13) The Conclusions need an upgrade after the description and discussion of the results. As of now, it states that “Our findings demonstrate that the use of the Fitch WFP, in combination with this particular WRF configuration, is largely able to capture wake effects and improve the general accuracy over a simulation without wind farms included.” Then, you proceed to describe that “The WF simulations accurately capture the wake structure when wind speeds range between 5 to 11 ms^{-1}.” As you can see, in the Conclusions, you merely restate the results. In my view, this restatement of results occurs because of the insufficient description and discussion of the results, as I said before.

14) In Table 2, you use a stability classification that includes classes such as “stable–neutral”. Whereas this is valid, most commonly you will find a five-class system with “neutral”, “weakly stable”, and “very stable”. Perhaps using the five-class system will make it easier to discuss your work with the literature and for other researchers to interpret your results. This is not criticism, just something to be aware of. Your work, your choice.

Technical corrections

1) Figure 1 needs domain dimensions in degrees or km.

2) The L parameter is just Obukhov length, not Monin-Obukhov length, even though it belongs to the Monin-Obukhov Similarity Theory. This requires changes in Tables and in the text e.g. when you write “MO length”.

3) Please include a wind direction arrow in Fig. 15, as you did for Fig. 11.

Suggested references

- Cañadillas, B., Foreman, R., Barth, V., Siedersleben, S., Lampert, A., Platis, A., Djath, B., Schulz-Stellenfleth, J., Bange, J., Emeis, S., & Neumann, T. (2020). Offshore wind farm wake recovery: Airborne measurements and its representation in engineering models. *Wind Energy*, *23*(5), 1249–1265. https://doi.org/10.1002/we.2484

- Platis, A., Siedersleben, S. K., Bange, J., Lampert, A., Bärfuss, K., Hankers, R., Cañadillas, B., Foreman, R., Schulz-Stellenfleth, J., Djath, B., Neumann, T., & Emeis, S. (2018). First in situ evidence of wakes in the far field behind offshore wind farms. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-20389-y

- Siedersleben, S. K., Platis, A., Lundquist, J. K., Djath, B., Lampert, A., Bärfuss, K., Cañadillas, B., Schulz-Stellenfleth, J., Bange, J., Neumann, T., & Emeis, S. (2020). Turbulent kinetic energy over large offshore wind farms observed and simulated by the mesoscale model WRF (3.8.1). *Geoscientific Model Development*, *13*(1), 249–268. https://doi.org/10.5194/gmd-13-249-2020