The authors introduce a Graph Neural Operator architecture for wind farm wake flow prediction, where turbine-turbine interactions are learned through message-passing on a graph and flow velocities can be queried at arbitrary probe locations. Trained on ~21,000 PyWake simulations spanning diverse procedurally-generated layouts, the model generalizes across farm configurations and seems to achieve low RMSE in the predictions while being approximately 10x faster than the engineering baseline (PyWake).

In general, I find that the manuscript is well written and comprehensive: all components of the framework, from procedural dataset generation to GNO architecture and training, are described in sufficient detail to enable independent implementation. The breakdown of the results is also very detailed, and I appreciate the thorough reporting on computational aspects. I recommend minor revisions, with specific comments below.

Specific comments:

- Section 2
  - A lot of detail is provided for the models that are used by PyWake to generate the training data. I think some of this can be moved to the appendix.
  - The part about the additional coordinate system is a bit confusing and could do with some extra clarification.
  - Why did you choose to use the coordinates of the sender nodes as an edge feature, instead of the relative coordinates (vector coordinates) between the the two connected nodes, as is typically done in other GNN work?
  - I'm not sure I see the advantage of having the decoding stage setup as it is. I get that you want to first freeze the turbine node latents, so that you have the option of adding any amount of probe nodes after, but why not have undirected probe edges between a probe and other probes, as well as between a probe and the turbine nodes. I guess the current setup makes sense when only using a single probe message-passing step, but why not have multiple steps during this phase? It would increase performance (albeit with extra computational costs). Please consider clarifying this.
  - Could you have not resumed training after 72 hours? Or were the models sufficiently trained by this time and it was not needed?
  - I'm not sure the MAPE metric is correct, as typically we would divide the error by the target value. In some sense, what you show here is more of a MAE that is normalized by the inflow velocity, which may still be a useful metric.
  - You mention the cardinality of the graph, is this actually used somewhere?
- Section 3
  - You could consider moving some of the parameter tables for the hyperparameter study to the appendix.

    - No ablation on the number of message-passing steps (M=3 throughout) is performed. This seems like an important hyperparameter to study, given that message-passing propagates information only to immediate neighbors. Would increasing the number of steps increase performance for cluster farm setups, where wake superposition seems to be the biggest hurdle?
    - The training procedure regarding probe node sampling deserves clarification. How are the probe locations selected? Random uniform sampling across the domain, or weighted toward regions with stronger wake effects? This choice likely influences model performance in different flow regions.
    - I would have really liked to see a complete flow map comparison between the GNO's output vs PyWake (along with a difference plot), especially in the near-wake region. It would be extremely useful to understand where the model falls short. You could do this for the IEA Wind 740-10-MW reference reference farms for instance.
    - This point is optional, but I was wondering what the learned RBF kernels look like. Showing the learned RBF distributions after training would provide insight into what spatial scales the model finds important.
    - Figure 9. I think that you can plot the farm in a nicer way, it is hard to actually see the positions of the turbines.
    - Figure 12. In general I find this plot a bit hard to read, the color scheme can be improved and the bars made a bit wider. For subplot (a): the KDE extends into negative RMSE values, which is not physically meaningful. Consider using a boundary-corrected KDE method, or simply truncating/reflecting at zero.
- Section 4
    - I would suggest more measured language that emphasizes the nice methodological contribution without overclaiming impact ('breakthrough').

Minor technical comments:

- Line 120, line 283: alignment issues
- Line 320: text overflow
- "Windfarm" vs "Wind farm" used interchangeably, try to stay consistent. Same for "flowmap" vs "flow map".
- Caption of Table 2: typo in caption ('b' instead of 'bold'?)