Review of “Impact of physical parameterizations on wind simulation with WRF V3.9.1.1 over the coastal regions of North China at PBL gray-zone resolution” by Yu et al.

This paper examines wind forecasts during a relatively long period of stable conditions when a haze event affected China. Surface meteorological observations are used to evaluate the WRF model’s ability to predict the evolution of winds during the event. The authors conduct a number of WRF simulations (640 total), altering the PBL, radiation, and microphysics schemes to determine the sensitivity of wind speed and direction forecasts to choice of model physics. Pearson’s correlation coefficient, bias, RMSE, and Taylor skill score are utilized to perform the model evaluation. Overall, the study shows the largest spread in wind speed within the PBL schemes tested, followed by radiation, and then microphysics schemes. Delineation between coastal/inland stations as well as stations at different elevation are examined to understand any model biases specific to land type and characteristics. An important finding is that WRF predicts wind speed less accurately for coastal stations compared to inland stations, and error metrics tended to degrade with increasing elevation.

Overall, this study has interesting components and would be a nice contribution to the literature especially due to the very large ensemble that was run. However, there are several aspects that should be addressed before the paper is suitable for publication. My main concerns are related to the authors’ model setup, lack of some background information about the case, and insufficient physical explanations for some of their results. My general and specific comments are listed below.

<u>Major/general comments:</u>

1. WRF model configurations: I wonder why the authors chose to run all of the physics parameterizations as default except for YSU, which was run using a topographic correction for surface winds and the top-down mixing option. Were the impacts of these YSU options tested? I believe that YSU is not run this way by default, so it would be good to know the impact, especially since your results show that YSU is one of the best performers. For instance, MYNN has a number of namelist tuning options, so why not modify these? Also, what is the motivation for running with the top-down mixing option in YSU if this is a statically stable case? Are stratus/fog conditions expected in some of the coastal regions? Please explain.

As a separate but related issue, the authors use different surface layer schemes between the PBL schemes. This means that it is impossible to attribute all of the differences in results specifically to the PBL scheme. There is no discussion about this at all in the paper, although it is definitely important considering the station observations likely fall within the first model grid cell. Furthermore, why not use the revised MM5 scheme for all of the PBL schemes? I believe that it is compatible with all of them (I may be wrong here). Regardless, it would be good to run an additional simulation to determine the impact of the surface layer scheme (which I suspect is more important than the microphysics scheme under stable conditions).

2. Case study: The authors select a 4-day study period when stability conditions were stable to

evaluate the WRF model; however, there is only a few sentence discussion about the case in Section 2.1. Although the authors do conduct many simulations, this is still a case study, and unfortunately, the authors do not present any large-scale meteorological information. It would be good to know the synoptic pattern and what type of evolution occurred; I imagine there is a pattern change over the course of the event since the regional wind speeds went from ~1 m/s to ~5 m/s according to Fig. 2. Moreover, the authors consider the impact of microphysical schemes in WRF even though this is a stable case. Are the authors anticipating cloud effects? Despite the relatively small impact of the microphysics options (e.g., Fig. 5), there are noticeable differences on 2019-01-12. I think the authors should include some metric of observed clouds (e.g., satellite images) since clearly the model is producing clouds.

3. Physical explanations: By and large, this paper reports on the model performance with respect to near-surface wind speed and direction. However, I think the authors do not provide any physical explanations for any of their results. Ultimately, this ends up limiting the applicability of the study to other stable events in different seasonal periods and geographical locations. It would be good to address questions such as: why is QNSE so different from the other PBL schemes? How does the YSU topographic correction affect the forecast? Additionally, linking the low-level wind results to the PBL vertical structure (e.g., wind, temperature/stability, and moisture), as well as x-y spatial variability in the wind field, would be useful. For the planview, the authors could zoom in on a region where they see the largest differences in PBL schemes and overlay their observations. I think these comparisons are especially important for the different PBL schemes because the authors highlight the gray zone, which is where the choice of turbulent mixing tends to dominate the solution.

<u>Minor/specific comments:</u>

1. Abstract, L14: Should be, “near-surface” rather than “surface”.

2. Abstract, L24: Please be more specific by what you mean by “land type”.

3. Abstract, L25: “For example, for the weather stations located in coastal regions.” This is an awkward sentence; it would be good to combine with another sentence or rewrite.

4. L47: Should add, “that occur at the sub-grid scale” after you say “key physical processes”. Also, some of the processes listed (i.e., planetary boundary layers and cloud microphysics) aren’t really processes. Please be more specific or re-phrase the beginning of the sentence.

5. L51: The impact of PBL schemes on wind simulations has been studied for many years (probably >60 years at this point).

6. L55: Should be, “horizontal grid spacing” rather than “horizontal resolution”.

7. L77: I would say, “high resolution mesoscale simulation” rather than “very high resolution” considering dx=500 m is not high resolution with respect to typical LES scales (10s of meters).

8. L77: Please explain briefly what is mean by “gray zone” and provide references.

9. L82: Should be, “findings” rather than “foundings”.

10. L88-92: Some references would be nice here, especially for an audience who is not familiar with this region.

11. L99-100: Please add references for why this wind direction indicated a weakened East Asia winter monsoon.

12. L103: Why not use a newer version of WRF? Version 3.9.1.1 was released years ago (August 2017, I believe).

13. L108-109: Were eta levels specified? It would be important to report the vertical grid spacing near the surface.

14. L126: Discussing the “gray zone” here is fine, but please reference this section in the introduction so that the reader is not confused by what you mean by the terminology.

15. L126-130: Please provide references for the “gray zone” discussion as well as the Deardorff SGS TKE scheme.

16. L139: I disagree that the “Lin scheme is a sophisticated scheme”, considering it is single moment and there are several double (or higher order) moment schemes.

17. L142: These microphysics schemes are not new, as they are over 10 years old now.

18. L144: Please define “ThompsonAA”.

19. L154-155: How exactly are the data screened? Please explain.

20. L187: Only in the extreme case of QNSE did the absolute difference exceed 10 m/s; however, as written currently it sounds like many simulations have large differences.

21. L202-204: Do you have proof for the claim about QNSE? Are there other studies that support this claim? I wonder if MYNN used EDMF or just ED in these simulations. I know that in newer versions of WRF, EDMF is default for EDMF.

22. L209-210: I don't understand this sentence. Wind direction depends upon the u and v-components, so one can have the correct wind speed but not the correct wind direction because the components are incorrect.

23. Fig. 5: There are repeat colors in this figures, please fix this.

24. L252-253: It would be good here to briefly confirm/reiterate the best combination schemes for wind direction.

25. L265-266: The difference in statistical measures between the ensemble and WDM6 is very small, suggesting the ensemble is not improving things too much.

26. L266-267: "reduces model bias by approximately half", compared to what?

27. L271-272: Please provide references.

28. L287: This is a broad statement. What exactly do you mean by it with respect to future model developments? Perhaps some discussion should be added to Section 4.

29. L289: "previous investigations", please add references.

30. L305: "surface topography does not induce new uncertainties into the simulation", this is simply not true. What about errors associated with horizontal diffusion and topographic data sets? These actually worse as terrain becomes steeper/more complex.

31. L321-322: In general, the QNSE finding is surprising, given that it was originally developed for stable conditions. Do your results agree with other studies that have used QNSE under stable conditions? Discussion here is needed.

32. L336-337: This broad statement about these “gray zone” resolutions being used “rarely” in previous simulation studies is not true. There are dozens (if not hundreds) of papers looking at gray zone modeling, especially in the last ~5-10 years. Please be more specific with your statement.

33. L342: To which PBL scheme parameters are you referring? It would be good to give some examples.

34. Figs. 9 and 10: Please note in the caption that the y-axis range is not the same (alternatively, make them the same).