Response to referee comments

We are grateful for the thoughtful and constructive feedback from the referee. We have revised the text in the manuscript to answer the referee's points and we believe this revision has improved the clarity and quality of our manuscript. This response provides a complete description of the changes that have been made in response to each comment. Referee comments are shown in plain text, author responses are shown in bold blue text. All line numbers in the responses refer to locations in the revised manuscript.

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.

Response: Thank you for taking time out of your busy schedule to review this paper, I really appreciate all your comments and suggestions. Please find my responses in below and my revisions/corrections in the re-submitted files. Thanks again.

Major/general comments:

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.

Response: Thank you for pointing out this, according to this comment, we added experiments considering the effects of YSU options, and the results are presented in section 4.3, lines 341-349:

"4.3 Impact of options in the YSU scheme

The impact of different options in YSU on wind-speed simulation is illustrated in Figure 15, the simulation with the best Taylor skill score in previous investigation is selected and referred to as YSU, three extra simulations with top-down mixing option turning off (No_mix), topographic correction option turning off (No_topo), and both options turning off (No_topo_mix) are conducted for comparison. The simulated wind speed increases when we turn off the individual or both options, which enlarges the overestimation of wind speed under stable conditions in our study (Figure 15a). Further investigation indicates that turning off the two options in YSU mainly degrades model performance with worse evaluation metrics, for example, the BIAS score increases from 0.36 m/s to 0.67 m/s in No_topo, to 0.43 m/s in No_mix, and to 0.69 m/s in No_topo_mix, the RMSE scores show similar degradation to BIAS by turning off the options in YSU.".

In the past few years, over North China, haze events are always accompanied by fog over the coastal areas, thus the top-down mixing option was turned on in the simulation.

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).

Response: Thanks for the comment, we totally agree with the reviewer that it is important to determine the impact of surface layer scheme, however, in the WRF model, the surface layer scheme is somehow tied to the PBL scheme, it is impossible to conduct systematic experiments using different surface layer schemes with the same PBL schemes, for example, YSU cannot run with surface layer scheme other than revised MM5 scheme, MYJ PBL scheme can only be used with MYJ surface layer scheme. In this study we designed the PBL experiments including both surface layer schemes and PBL schemes, thus the results of PBL schemes are actually the results of the combination of PBL and surface layer schemes.

For the revised MM5 scheme, it is compatible with a lot of PBL schemes, but not all, for example, it is not compatible with the TEMP PBL schemes.

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.

Response: Thank you for pointing out this, we removed figure 2 and added the distribution of cloud, geopotential height and winds as a new figure. The descriptions are in Lines 107-115: "Figure 2 depicts the distribution of geopotential height, winds, and cloud fraction during the study period, the geopotential height and winds data are from the ERA5 dataset (Hersbach et al., 2020), and the cloud fractions are from satellite observations of CLARA (CM SAF cLoud, Albedo and surface Radiation) product family (Karlsson et al., 2021). Figure 2 indicates that a weak high-pressure system persisted from 11 to 13 January, along with weak southwest wind in the study area, which would transport warm and wet air to the study area (Gao et al., 2016a), creating a favorable moisture condition for stable conditions and inhibiting pollutants dispersal (Zhang et al., 2014; Hua and Wu, 2022). Then the high-pressure system was replaced by strong northwest wind from 14 to 15 January 2019. The CLARA observations indicate cloud fraction exceeding 60% on 12 January at the study area, while for the rest of the time, cloud fraction is low. This stable event is used to investigate the impact of physical parameterizations of the WRF model."

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.

Response: Thanks for the comments, according to them, we added the comparisons in sections 4.1 (lines 318-330) and 4.3 (lines 341-349):

"4.1 Spatial distribution of wind field

Figure 13 shows the spatial distribution of observed and simulated wind fields during the study period, we choose 14:00 in local time as an example. The simulation using YSU, Dudhia-RRTM and WDM6 schemes is referred to as YSU, and the simulation using QNSE, Dudhia-RRTM and WDM6 is referred to as QNSE. YSU generally reproduces the wind field in the study area, especially in terms of wind speed. For example, the observed wind speed is lower on 13 January 2019, with values lower than 2 m/s in many stations, while on 15 January 2019, the observed wind speed is higher than 4 m/s in most of the stations. In the simulation with YSU, wind speed is about 2 m/s on 13 January 2019 and higher than 4 m/s on 15 January 2019 over the study area, which is close to the observation. On the contrary, simulation with QNSE fails to reproduce the distribution of wind speed, and shows strong overestimation, especially over the mountain areas of the study area (Figure 1a), for example, the peak wind speed in simulation with QNSE exceeds 20 m/s on 15 January 2019, which is more than five times greater than the observation, this overestimation is consistent with the large positive bias in previous investigation of Figure 3. For the wind-direction simulation, YSU shows degraded performance compared to wind speed, and generally fails to reproduce the wind-direction distribution for most of the stations, QNSE also fails to do so."

"4.3 Impact of options in the YSU scheme

The impact of different options in YSU on wind-speed simulation is illustrated in Figure 15, the simulation with the best Taylor skill score in previous investigation is selected and referred to as YSU, three extra simulations with top-down mixing option turning off (No_mix), topographic correction option turning off (No_topo), and both options turning off (No_topo_mix) are conducted for comparison. The simulated wind speed increases when we turn off the individual or both options, which enlarges the overestimation of wind speed under stable conditions in our study (Figure 15a). Further investigation indicates that turning off the two options in YSU mainly degrades model performance with worse evaluation metrics, for example, the BIAS score increases from 0.36 m/s to 0.67 m/s in No_topo, to 0.43 m/s in No_mix, and to 0.69 m/s in No_topo_mix, the RMSE scores show similar degradation to BIAS by turning off the options in YSU."

Minor/specific comments:

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

Response: Thank you for pointing out this, the errors were corrected in lines 14-15:"how different physical parameterizations impact simulated near-surface wind at 10-meter height over the coastal regions of North China"

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

Response: Thanks for pointing out this, it was revised to "model sensitivity is also impacted by ocean

***proximity and elevation*"**

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.

Response: Thanks for the comment, it was revised in lines 23-25: "for coastal stations, MYNN shows the best temporal correlation with observations among all PBL schemes, while Goddard shows the smallest bias out of SW-LW schemes"

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.

Response: Thank you for pointing out this, it was revised in line 50: "instead, parameterizations are needed to represent the effect of key physical processes, such as radiative transfer, turbulent mixing, and moist convection that occur at the sub-grid scale."

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

Response: Thanks for the comment, it was revised to "The impact of planetary boundary layer (PBL) schemes on wind simulations has been studied for many years".

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

Response: Thank you for pointing out this, the errors were corrected in line 61, we also revised the Abstract accordingly.

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).

Response: Thanks for the comment, it was revised to "The investigation is conducted using the WRF model at a grid spacing of 0.5 km" in lines 83-84.

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

Response: Thank you for pointing out this, it was revised in lines 84-86: "which belongs to the PBL "gray zone" resolution that is too fine to utilize mesoscale turbulence parameterizations and too coarse for a large-eddy-simulation (LES) scheme to resolve turbulent eddies (Shin and Hong, 2015; Honnert et al., 2016)"

9. L82: Should be, "findings" rather than "foundings".

Response: Thank you for pointing out this, it was corrected in line 91.

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

Response: Thanks for the comment, it was revised in lines 96-102: "The study area is located in the central section of the "Bohai Economic Rim", which is bordered to the southeast by the Bohai Sea and to the northwest by the Yan Mountains (Figure 1a). This region traditionally hosts heavy industry and manufacturing businesses and is a significant region of economic growth and development in North China (Song et al., 2020; Zhao et al., 2020). Air quality in this area has declined over the past decades, and the frequency of winter haze events has increased due to increased pollutant emissions and favorable stable weather conditions with lower wind speed (Gao et al., 2016a; Cai et al., 2017)."

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

Response: Thank you for pointing out this, the figure was removed and we added the synoptic description of this event in lines 107-115: "Figure 2 depicts the distribution of geopotential height, winds, and cloud fraction during the study period, the geopotential height and winds data are from the ERA5 dataset (Hersbach et al., 2020), and the cloud fractions are from satellite observations of CLARA (CM SAF cLoud, Albedo and surface Radiation) product family (Karlsson et al., 2021). Figure 2 indicates that a weak high-pressure system persisted from 11 to 13 January, along with weak southwest wind in the study area, which would transport warm and wet air to the study area (Gao et al., 2016a), creating a favorable moisture condition for stable conditions and inhibiting pollutants dispersal (Zhang et al., 2014; Hua and Wu, 2022). Then the high-pressure system was replaced by strong northwest wind from 14 to 15 January 2019. The CLARA observations indicate cloud fraction exceeding 60% on 12 January at the study area, while for the rest of the time, cloud fraction is low. This stable event is used to investigate the impact of physical parameterizations of the WRF model."

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

Response: Thank you for the comments, WRF V3.9.1.1 was released on August 28, 2017, and we used this version for a long time in different areas across China, the overall performance is satisfactory to us, thus we continue to use it in this study to support comparative studies with previous simulations. we assume that the WRF version does not affect our results.

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

Response: Thanks for pointing out this, it was revised in lines 123-124: "and the eta values of the first 10 levels are 0.996, 0.988, 0.978, 0.966, 0.956, 0.946, 0.933, 0.923, 0.912, and 0.901."

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.

Response: Thank you for the comments, we added the description of "gray zone' in the section 1 lines 84-86: "the PBL "gray zone" resolution that is too fine to utilize mesoscale turbulence parameterizations and too coarse for a large-eddy-simulation (LES) scheme to resolve turbulent eddies (Shin and Hong, 2015; Honnert et al., 2016)"

We also revised the sentences here as "<u>The horizontal grid spacing of 0.5 km is within the PBL gray zone resolution,</u> both PBL and LES assumptions are imperfect".

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

Response: Thanks for the comments, the references for the "gray zone" are added in lines 85-86: "the PBL "gray zone" resolution that is too fine to utilize mesoscale turbulence parameterizations and too coarse for a large-eddy-simulation (LES) scheme to resolve turbulent eddies (Shin and Hong, 2015; Honnert et al., 2016)", and the reference for the Deardorff SGS TKE scheme is added in line 142: "the 1.5-order turbulence kinetic energy closure model is used to parameterize motion at the sub-grid scale <u>(Deardorff, 1985)</u>".

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.

Response: Thanks for pointing out this, we removed it and rewrite the whole part in lines 151-160: "<u>Sixteen MP schemes were applied in this study (Table 1), Lin, WSM3, WSM5, ETA, WSM6, Goddard, SBU and NSSL1 schemes are the single-moment bulk microphysical scheme, which predicts only the mixing ratios of hydrometeors (i.e., cloud ice, snow, graupel, rain, and cloud water) by assuming particle size distributions. The other eight schemes (Thompson, MYDM7, Morrison, CMA, WDM6, NSSL2, ThompsonAA and P3) use a double-moment approach, predicting not only mixing ratios of hydrometeors but also number concentrations. Among them, two types of hydrometeors are included in WSM3 (cloud water and rain), three types of hydrometeors are included in ETA (cloud water, rain and snow) and P3 (cloud water, rain and ice), four types of hydrometeors are included in WSM5 and SBU (cloud water, rain, ice and snow), five types of hydrometeors are included in Lin, WSM6, Goddard, Thompson, Morrison, CAM, WDM6 and ThompsonAA (cloud water, rain, ice, snow and graupel), six types of hydrometeors are included in MYDM7, NSSL1 and NSSL2 (cloud water, rain, ice, snow, graupel and hail).</u>"

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

Response: Thank you for the comments, we removed this and rewrite the whole part in lines 151-160.

18. L144: Please define "ThompsonAA".

Response: Thank you for the comments, ThompsonAA is short for "Aerosol-Aware Thompson", and we explained it in Table 1.

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

Response: Thanks for the comments, it was explained in lines 169-171: "All data are screened before analysis in order to remove stations with data showing spurious jumps <u>(e.g., wind speed jumps to 0 m/s due to frozen sensor)</u>. After this filtering, 105 out of 132 weather stations (Figure 1a) remained, including 89 inland stations and 16 coastal stations."

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.

Response: Thank you for the comments, it was revised in lines 196-199: "reproduced by all schemes except <u>for QNSE, with which the wind speed change is considerably larger than with all other schemes during the simulation period. Almost all the PBL schemes overestimate wind speed by 1 m/s, however, for the QNSE scheme, the largest overestimation exceeds 10 m/s during the daytime on 11 and 15 January 2019.</u>"

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.

Response: Thank you for pointing out this, the claim about QNSE is only a speculation, to avoid misleading, the sentence was revised in lines 208-210 as "For the QNSE scheme, the maximum BIAS and RMSE scores for individual stations exceed 10 m/s and 16 m/s, <u>indicating that it has problems in reproducing wind speed under stable conditions over the study area</u>.", For MYNN, in version 3.9.1.1, EDMF is turned off by default.

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.

Response: Thank you for pointing out this, it should be "the simulated wind direction was calculated using the u and v- components, the bias in modeled wind direction can be attributed to bias in the u and v- component simulations", this is redundant, so we deleted it.

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

Response: Thanks for pointing out this, the errors are corrected in figures 5, 11 and 12.

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

Response: Thank you for pointing out this, we added the description of wind direction in lines 252-253: "For the wind direction simulation, LES combined with Dudhia-RRTM shows the best CORR score, while TEMF is the best scheme according to BIAS and RMSE scores".

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.

Response: Thanks for pointing out this, we added discussion of this in line 373: "Thus, model ensemble does not always provide the best performance".

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

Response: Thank you for pointing out this, it was revised to "According to the statistic scores, ENS4 reduces model bias by approximately half compared to ENSall" in lines 276-277.

27. L271-272: Please provide references.

Response: Thanks for the comment, it was revised in lines 279-280: "which impacts local low-level circulation patterns and the wind distribution (Yu et al., 2013; Barlage et al., 2016)."

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.

Response: Thank you for pointing out this, we made it clear in lines 294-295: "This degradation may be caused by the uncertainties from the prescribed SST in our simulation, which may require a better description of atmosphere-ocean coupling process in future model development".

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

Response: Thanks for the comment, previous investigation refers to the investigation in this study, so we revised it to "…for inland stations are generally consistent with those of previous investigations in this study," in lines 296-297 .

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.

Response: Thanks for pointing out this, the statement is misleading, so we removed it.

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.

Response: Thank you for the comments, we conducted further investigation of QNSE scheme in sections 4.1 and 4.2 (318-340):

"4.1 Spatial distribution of wind field

Figure 13 shows the spatial distribution of observed and simulated wind fields during the study period, we

choose 14:00 in local time as an example. The simulation using YSU, Dudhia-RRTM and WDM6 schemes is referred to as YSU, and the simulation using QNSE, Dudhia-RRTM and WDM6 is referred to as QNSE. YSU generally reproduces the wind field in the study area, especially in terms of wind speed. For example, the observed wind speed is lower on 13 January 2019, with values lower than 2 m/s in many stations, while on 15 January 2019, the observed wind speed is higher than 4 m/s in most of the stations. In the simulation with YSU, wind speed is about 2 m/s on 13 January 2019 and higher than 4 m/s on 15 January 2019 over the study area, which is close to the observation. On the contrary, simulation with QNSE fails to reproduce the distribution of wind speed, and shows strong overestimation, especially over the mountain areas of the study area (Figure 1a), for example, the peak wind speed in simulation with QNSE exceeds 20 m/s on 15 January 2019, which is more than five times greater than the observation, this overestimation is consistent with the large positive bias in previous investigation of Figure 3. For the wind-direction simulation, YSU shows degraded performance compared to wind speed, and generally fails to reproduce the wind-direction distribution for most of the stations, QNSE also fails to do so.

4.2 Vertical profile of wind speed

Figure 14 shows the observed and simulated vertical profile of wind speed at 08:00 and 20:00 during the study period, the location of the sounding station is illustrated in Figure 1. YSU reproduces the vertical structure of wind speed reasonably, with slightly larger model bias above the height of 15 km. Within the low levels below 2.5 km, simulated wind speed from the YSU scheme is close to the observation, with the bias lower than 2.5 m/s in most cases. Meanwhile, QNSE shows worse performance in reproducing the vertical structure of wind speed, with significant larger model bias compared to YSU. for example, QNSE overestimates the low-level (< 2.5 km) wind speed by about 10 m/s at 20:00 on 11 January 2019, and overestimate wind speed by 20 m/s at 20:00 on 11 January 2019. It is interesting to note that the simulation with QNSE is pretty similar to that with YSU at 08:00 during the study period, indicating that large difference between YSU and QNSE only occurs at specific time during the study period, which is also revealed in Figure 3a."

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.

***Response:** Thank you for pointing out this, we revised it to "to the PBL gray zone, which has rarely been used in previous simulation studies in China" in line 406.*

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

***Response:** Thank you for pointing out this, we revised it in lines 375-377: "further tuning of the parameters within the PBL schemes, such as turbulent kinetic energy (TKE) dissipation rate, TKE diffusion factor, and turbulent length-scale coefficients is needed."*

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

make them the same).

Response: Thanks for the comment, we make the range of y-axis identical in the revision.