

Referee Comments

Title: Impact of physical parameterizations on wind simulation with WRF V3.9.1.1 under stable conditions at PBL gray-zone resolution: a case study over the coastal regions of North China

Author(s): Entao Yu et al.

MS No.: gmd-2022-53

MS type: Model evaluation paper

Iteration: Revised submission

General Comments

The authors have addressed most of my comments and I am happy to see that many of my suggestions were adapted in the revised manuscript. The quality of the analysis has improved; however, the discussion and figures can still be improved upon, some corrections are required (in some parts the text does not seem to agree with what the figures show), and a few concerns remain to be addressed and clarified.

Specific Comments

- “The stable weather event”:
 - Based on Fig. 2 to me it seems that the study area is under the influence of the high-pressure system only on Jan 11th. Afterwards the study area experiences *northwest* winds (lines 110/111 state “southwest”) as the high-pressure system weakens and moves west. Except on Jan 11th, I am not convinced that the conditions are stable based on Fig. 2. The subplot for Jan 13th even indicates a weak shortwave and a shift towards positive vorticity over the study area. Please also frame the area covering D03 to guide the reader where to find the “study area” in this figure. Consider assessing other variables that indicate static stability (e.g., vertical profiles of potential temperature).
 - Line 111: “which would transport warm and wet air to the study area” – weak winds do not transport air masses very far and there is no major body of water to the west of the study area. Where does the “wet air” come from? And why are moist conditions favorable for stable conditions (line 112)?
 - Under stable conditions, wind direction is (a) more variable, hence, more challenging to predict accurately, and (b) less important, as wind speeds are weak. The manuscript should frame the results accordingly.
- This manuscript includes numerous figures with subplots. I think the authors should reconsider which figures are meaningful to include in the manuscript. I.e., is it essential for the reader to see Fig 5b vs. a short description that CORR scores are indistinguishable among MP schemes at a precision of .XX?
For all figures, ensure that Figure labels are readable on a printed paper size. (Larger font size is needed for most figures, but in particular, Fig 2, 11 & 12.)

- Figure 2:
 - Is cloud fraction data available from ERA5? Although the observational data product is interesting, for your study it might be also (maybe even more) important to see whether the initial condition model produced clouds. Several cloud products do not necessarily have to be shown in the manuscript, but it would be nice if differences among observed, ERA5 and WRF cloudiness would be explained and discussed in the manuscript.
 - Could you please add a box framing the “study area” (i.e., D03)? Is it necessary to show these maps at this scale? Perhaps the area could be a more zoomed in on the study area. (Also note that the labels are not readable.)
 - With respect to the existing topography in the area, 925hPa is a very low level. Consider showing 750 or 500hPa geopotential heights instead, which can be more descriptive for synoptic situations.

- Figures with error bar plots (Fig. 3-8): Please explain why the error bars (blue) sometimes have values outside of the range of station values (orange dots). Consider whether the metrics should be calculated over all stations (and lead times?), then averaged over simulations, or whether the metrics should be calculated over all lead times, then averaged over stations and simulations. Accordingly, show error bar ranges either across stations or simulations or both.

- It is challenging for the reader to remember which acronym belongs to each parameterization type. Therefore, it would be helpful to mention parameterization types more often throughout the manuscript please. E.g., in Fig. 11: Please use same scales on y-axis and add a label on the left for the corresponding parameterization types.

- Figure 13: The many overlapping vectors are hard to decipher. Consider using wind barbs instead of vectors or less dense vector distributions with smaller reference vector. At the scale of this figure, it is also difficult to identify direction and size of the simulated wind vectors. In figure 13 and 14 it is unclear why QNSE is shown when it was previously determined to perform poorly, and the authors decided to disregard it from further investigation.

- Figure 14: Why do you show and discuss the wind speeds up to a height of 20 km (stratospheric altitudes)? How relevant is that for your study and analysis?
I further suggest re-ordering the subfigures and showing all profiles at 8:00 at the top and 20:00 at the bottom.

- Lines 214-216: “Further comparison indicates that all PBL schemes strongly overestimate the speed of north wind compared to the observations, which is the main cause of positive bias in wind speed (Figure 3).” Please elaborate. The wind speed bias could well be attributed to several wind directions. When and where does north wind occur in the simulations? Could this value be from an isolated (e.g., high elevation) station?

- Line 170: “wind speed jumps to 0 m/s due to frozen sensor” - Is there a clear “jump” when a sensor freezes? How do you distinguish between real 0m/s in wind speed vs 0m/s resulting from a frozen sensor? Do you consider measured temperatures?
- Results and conclusions: Instead of discussing each metric separately and listing the best schemes, I think it would be better to present the big picture to the reader, and discussing the schemes that have better overall scores across metrics.
- Line 291-293: “WRF simulates wind speed less accurately for coastal stations compared to inland stations.” and further “the BIAS and RMSE scores are generally worse for coastal stations compared to inland stations” – actually Fig. 11 shows that bias is consistently worse for inland stations and RMSE is a tie. Revise!
Is the larger ensemble spread for coastal stations (the ensemble being less confidence) maybe a better representation for the forecast than the narrower spread for inland stations that have a larger bias (the ensemble being overconfident)?
- Lines 302-311: Could the distribution of station elevations also be included in Fig. 1? Since different marker shapes are already used for inland vs coastal station, different marker colors could be used for the elevation categories.
I would still like to see a synoptic discussion explaining why the high elevation station have lower wind speeds.
It is interesting that WRF wind speeds are similar or increase while observations decrease with increasing elevation.
- The current conclusions are more of a summary. And the current discussion section contains mainly content that belongs into the results section. Since the actual discussion is limited (I only see two discussion points: the need for bias correction and other potential sources of errors), I suggest combing the discussion with the current conclusions and renaming this section to “summary and discussion”.

Technical Corrections

Line 16: “We performed 640 ensemble simulations using multiple combinations of”

Lines 17/18: “Model performance is evaluated using measurements from 105 weather station observations”

Line 19: Where does the sensitivity to land surface models fit in?

Lines 37/38: To avoid repetition: “The haze events are most frequent in boreal winter and are closely related to local weather conditions, with haze forming in regions with low wind speeds”

Lines 38-42: “Projections of future climate change suggest that global temperatures will increase, and the frequency of conducive and weather conditions conducive to severe haze is projected to will increase substantially in response to the climate change, which in turn may increase the frequency of

~~haze event over~~ affecting North China (Cai et al., 2017). However, numerical models ~~always~~ often show large bias in wind predictions over China [...]"

Lines 44-46: "In recent years, numerical models have been used extensively to study *and forecast* the weather and climate over China, as they have high spatial and temporal resolutions, and employ sophisticated physical parameterization schemes that can reproduce ~~detailed~~ atmospheric and land surface processes"

Lines 46/47: "However, *studies* mostly focus on temperature or precipitation, and only a few ~~studies~~ have attempted to simulate winds over China" The models simulate all variables, but *studies* focus on certain variables.

Line 52: "[...] as it *can* strongly influences ~~the~~ model results" Your study shows that it does not always influence model results "strongly"

Line 57: "~~A lot of~~ Many studies"

Line 59: [New sentence] "For example [...]"

Line 63: "is used instead. [New sentence] The YSU scheme also shows [...]"

Line 65: "~~There are also some~~ Other studies suggesting that MYNN and ACM2 are more appropriate [...]"

Lines 67-69: The word "affect" is used 3 times in two sentences – see if you can alter language more.

Lines 70/71: The citation of Cheng et al. (2013) feels out of place – it considers a different season and location than this study. Maybe use it as general citation for the sentence in line 67 or with the other citations in line 69 if appropriate.

Lines 76/77: "The *combination of* physical parameterizations are also vital to wind simulation, as ~~they may alter~~ the processes of atmosphere-land interactions, radiation transport, and moist convection *interact*, and *may* amplify the uncertainties in wind prediction."

Line 81: "parameterization schemes. [New sentence] ~~To the best of~~ our knowledge, the sensitivity"

Line 105: New paragraph

Figure 1: Why is there such a large difference in station number between TangShan City and QuinHuangDao City?

Figure 2 caption: "The daily averaged geopotential height (contour *lines*, units: gpm) and winds (vectors, units: m/s) at 925 hPa and cloud fraction (shading, units: %) ~~at 925 hPa~~ during 11 *through* 15 January 2019"

Line 130: "All ~~the~~ simulations ..."

Lines 148/149: "In this study, ETA, QNSE, MYNN, Pleim-Xiu, and TEMF SL schemes are chosen separately for PBL schemes of MYJ, QNSE, MYNN, ACM2, and TEMF, *respectively(?)*." All the SL schemes need references too.

Table 1: Either clarify in the caption that the schemes that share rows are not specifically assigned to each other (except for SW-LW), or change back to the previous layout, with columns “Parameterization Type”, “Scheme” and “Reference”, then in the rows list all PBL, MP, SW-LW schemes.

Line 176/177: “including the Pearson’s correlation coefficient (CORR)” or “including the Pearson’s correlation coefficient (CORR)”

Lines 183/184: Perhaps also mention what the perfect scores for the other metrics are.

Line 188: “... vector notation approach. [New sentences] Circular correlation coefficient ...”

Line 194: “Figure 3a shows the time series of observed and simulated wind speeds at local time. *Model wind speeds are shown for different PBL schemes averaged over all other parameterization types.*” Also please clarify whether the day starts with the background shading or at the tick marks.

Line 200: “*smaller difference*” or better “more similar to measurements”

Line 203: “... Figure 3b-d. [New sentence] MYJ shows the best CORR score of 0.96; [semicolon] MYNN, ACM2 and UW are *next best*...”

Line 223/224: They are not “the same”, but “very similar” (perhaps use “to a precision of 0.XX [fill in the correct precision number]”)

Line 235: “Strong overestimation” seems extreme considering that total errors are often within a 1m/s margin and not much different from the other two schemes. Maybe: “RRTMG and CAM show larger overestimation than Dudhia-RRTM and Goddard at daytime peaks.”

Line 242: “BIAS of -15.7” – be consistent with rounding precisions (please check everywhere in the manuscript)

Line 248/249: “the results are expected to be consistent with evaluations using other MP schemes” – why? If you see differences between Fig 9a and 9b, is it not plausible that there would also be differences with other MP schemes?

Line 251: “... Dudhia-RRTM or Goddard. [Period]”

“when [YSU?] is applied with RRTMG, the BIAS and RMSE scores show an obvious increasement compared with that with Dudhia-RRTM” – I can’t follow.

Line 252: “[New paragraph] For the wind direction simulation ...”

Line 257/258: “this indicates that a systematic variation of parameterizations is not necessary” & “this indicates that a systematic variation of parameterizations is important when focusing on these variables.” (line 261) – These are my words from the previous review – it was meant as a comment on the authors’ motivation for this study, however, to the reader this wording is confusing here. Please use your own words to rephrase this depending on what you wish to say.

Figure 10: Please note in the figure caption that only MP schemes are labeled because all configurations use the same PBL and SW/LW schemes. You may also highlight in your discussion that the best individual schemes can reduce the bias of the ensemble by ~50%, which is significant.

Line 275: Be consistent with abbreviations ENSall in test body vs ENS(576) in Fig. 10. (I prefer ENS(576).)

Line 281/282: ~~“The effects of these parameters on the model results are presented below.”~~ Redundant. Skip.

Line 284/285: “Figure 11 compares the results of wind speed for coastal stations (closer than 5 km from the shoreline, 89 stations in total) and inland stations (over 5 km from the shoreline, 16 stations in total), the locations of these stations are shown in Figure 1a.” – In Fig 1a looks there are more inland stations than coastal stations – should it be 89 inland & 16 coastal station?

Line 289/290: ~~“, especially among the MP schemes during 11-13 January 2019 with lower wind speed”~~ It looks like this is valid everywhere.

Sections 4.1-4.4 should be part of the results (under section 3)

Line 325: “which is *similar* to the observations”

Line 337/338: Do you mean Jan 12th for one of the two? To me it looks almost like 30m/s on the 12th, and like 20m/s on the 11th.

Line 336: This sentence states generally “significant” differences between YSU and QNSE with an example; later the text says they are “pretty similar”.

“For example” (capitalize at the beginning of a new sentence)

Lines 362-365: This fits in better in the methods sections where the metrics are first introduced.

Lines 367-370: This fits in better with section 3.1.5 where the ensemble results are shown.

Line 381: “during a relatively long period of stable conditions” – see my comment above

Line 391: “*followed* by MYNN”?

The authors should try to vary their language more. Many sentences start with “further investigation” and “further comparison”. This phrase can be omitted.