

Interactive comment on “Sensitivity analysis of PBL schemes by comparing WRF model and experimental data” by A. Balzarini et al.

Anonymous Referee #1

Received and published: 7 October 2014

Overall, I felt the subject has a large scientific relevance particularly to anybody using the WRF model for boundary layer studies, including for both air quality and wind energy research. While I have a few questions relating to the science performed in this study, most of the suggestions I have with regard to corrections are English language corrections, which at times made the paper awkward and confusing. However, the topic is very relevant and should be accepted after significant revisions to sentence structure and wording are considered. I think the author does a great job at summarizing the PBL schemes and their main differences, definitely something that is needed with so many new schemes coming online to the WRF community.

Scientific questions/considerations:

1. I was a little bit confused with the description of your WRF runs. You mention on line
C1932

18 of page 6139 that you ran WRF for February 2008 with a spin-up time of 5 days. Was there only one run that lasted the entire month of February for each PBL scheme? Or did you instead have a new run for each day during February 2008 for each PBL scheme? My concern is that I would think the PBL height and other boundary layer variables are highly sensitive to the time since model initialization and that conditions 5 days past model initialization might be significantly different than expected in the model simply as a result of the chaotic nature of the atmosphere. Can you comment a little more on your model setup/procedure? In addition, you mention your model has 27 vertical levels, how many are located in the boundary layer?

2. I notice you use the error statistics mean bias, RMSE, and correlation in your error analysis but when evaluating the different PBL schemes you really only talk about mean bias. It might be interesting to plot instead the mean absolute error (the mean of the absolute value of the differences). Often times the mean bias hides large errors that tend to cancel out and produce really small biases. However, model accuracy may be more appropriately measured in terms of mean absolute error where cancellation of errors can not occur. Perhaps your conclusions about which schemes are best will be different? The mean bias is most definitely a useful parameter to include, and it should not be eliminated from the paper, but MAE might give a better indication of PBL scheme superiority (for example, if one scheme has a lot of large errors, but they tend to cancel, the scheme can have no bias and be thought of as a good scheme, but in fact there are significant forecast deficiencies that are hidden that are brought forward when using MAE).

3. Can you comment on if you see similar error statistics for the coarser domains? There is debate in the community about if grids containing small grid spacing are necessarily better than coarse domains (some have even found the opposite with coarse domains providing better error statistics than the fine domains). Adding this would to the paper would be beneficial.

General comments on the text/language:

C1933

1. I would recommend using the term Numerical Weather Prediction (NWP) model over mesoscale meteorological model. While WRF definitely shines as a mesoscale model, it is being adapted for uses at all scales of NWP from LES to Global modeling. Skamarock (2008) uses the phrase NWP model.

2. Under section 2.1 you mention WRF is a non-hydrostatic meteorological model, however it can also be run hydrostatically.

3. Planetary Boundary Layer, Atmospheric Boundary Layer, PBL, ABL, and Boundary Layer are all used in the paper. I recommend the author stick with one of these for the entire paper (I would suggest PBL). The same can be said for the phrases underestimate, under predict, and under-predict. I think underestimate is appropriate.

4. Use grid spacing or grid length when talking about the horizontal grid spacing in WRF. Resolution implies that, for example, you can resolve a feature at 45-km in the horizontal with a grid box every 45-km when that is not the case. It is commonly cited that around 4-8 grid boxes are required to resolve a feature in NWP meaning your horizontal resolution with a grid spacing of 5-km is roughly 40-km.

There are numerous other grammatical issues that need to be corrected. The font sizes on the axis labels for the figures should be increased.

Interactive comment on Geosci. Model Dev. Discuss., 7, 6133, 2014.