Review of "Simulation performance of different planetary boundary layer schemes in WRF V4.3.1 on wind field over Sichuan Basin within "Gray zone" resolution" by Want et. al.

This study performed sensitivity experiments using four PBL schemes over the complex terrain in Sichuan Basin at the "Gray zone" resolution. The results show that while wind direction can be well reproduced and is not very sensitive to the PBL schemes, wind speed shows more sensitivity. The QNSE scheme had the best performance in reproducing the temporal variation, whereas the MYJ scheme had the smallest model bias. Using K-means classification, the authors concluded that the performance of the schemes is influenced by circulations. Predicting near-surface winds has practical importance and remains an ongoing challenge, especially over complex terrains. The choice of PBL has a significant impact on model performance. Therefore, this study is significant in this regard. However, the present form of analysis can be improved. I would overall recommend a major revision before it can be considered for possible publication.

Major comments

1. Since the authors emphasize this is a case study, one would expect case-by-case analysis. However, most analyses focus on bulk statistics or aggregate the data in some ways. The cases were selected solely based on wind speed exceeding 6 m/s. Is there any reason why this threshold is used? The length of each case should also be clarified.
2. The distribution probability analysis is a good way to evaluate the bulk features. How are the two parameters used in the Weibull distribution function connected to the distribution properties? Is the 10-min or event average used in the Weibull analysis? Please clarify.
3. The performance of PBL schemes can be influenced by many factors such as model assumptions, weather conditions, and local stability. Events with similar statistical errors do not directly reflect that they resulted from similar driving factors. Instead of classifying the events based on their statistical errors, I would suggest the opposite approach – classify the weather conditions and link the model errors to them.

Specific comments

Line 110-113: Please elaborate on why it is important to run the model at the "gray zone" resolutions?
Line 45: change to "winds"
Line 69: Please add WRF version.
Line 83 and other places: Add a space between the number and units.
Line 105-107: Please add reference to this statement.
Line 126-127: Why the case study is novel?
Line 133: Please replace "*" with "×".
Line 167: Change to "model configuration".
Table 1: What surface scheme was used?
Line 189: Why is 6 m/s selected as a threshold to select the cases? How long does a case last, a day?
Line 208: Suggest using Bias which is more commonly used.

Figure 2: What do the shading mean and dashed line mean? I assuming the dashed line is the threshold, which is 6 m/s in the text, but 5 m/s is showing in the figure. Please clarify. Again, from this figure, many of the cases were associated with diurnally varying winds while some cases were not. It would be interesting to see what synoptic scale/local conditions drive those wind patterns, and evaluate the PBL schemes' performance associated with those conditions.
Line 286: Please list some examples for the studies.
Line 290: Assuming the mean and median were calculated over the events. Please clarify.
Line 303-304: Please clarify that the "median" of the MJY ME is 0.96 m/s.
Line 307-319: This doesn't explain why MYJ is better in mean metrics while QNSE is better in variation. Since there is a suspicion that the performance of the PBL schemes differs under different stabilities, I'd suggest calculating the statistical metrics over different stabilities.
Line 330: Change to "10 m".
Figure 8: Looks like some points belonging to Cluster 1 is more close to the centroid of Cluster 2?