**Review of "Improving boundary layer flow simulations over complex terrain by applying a forest parameterization in WRF" by Johannes Wagner, Norman Wildmann, and Thomas Gerz, Manuscript number: wes-2019-77**

The manuscript describes the use of a canopy parameterization for characterizing surface roughness in WRF model simulations. Using the parameterization improves various aspects of the low-level flow in the challenging site of Perdigão, Portugal.

The topic of the manuscript is engaging and useful. The influence of forest roughness length on wind simulations has not been studied before in this way and with the WRF model.

The goals of the paper are exciting and worth publishing, but the methodology is generally not appropriate and, in many places, not well explained. Furthermore, the analysis of the model simulations lacks some depth. Two more significant points stand out:

Major issue 1). Performing 48-day simulations without re-initialization or nudging to control the error growth in the model domain is inappropriate. It does not conform to the best practices in atmospheric modelling. As recommended by Warner (2011), one should "understand the limitations to the predictability of the phenomena being modelled". After about 72 hours, the WRF simulations probably lost most of their predictability. Luckily, the prominent topographic features at Perdigão are a dominant control of the evolution of the flow. Thus, it is not essential if WRF captures all the details of the synoptic flow correctly. It is probably only necessary to get the right large-scale flow direction for the terrain to force the small-scale flow in the right way. However, the use of such long un-nudged simulations gives the wrong impression to the readers. The setup of the simulations is not appropriate for what the manuscript wants to show.

I understand that the goal of the paper is not to show that forest canopies improve weather forecasts (of course, it probably does). But calculating correlation coefficients that include all masts, different heights and different sites together give that impression. (I will come back to this point in comment 2). So, I suggest downplaying (or entirely removing) the correlations and focus on a more interesting analysis of the results.

Major issue 2). I think it a pity to combine all masts and heights in Figure 9. I believe there is much more to be understood. The added friction of the forest parameterization reduces the winds in places, but Figure 9 cannot show that. I think a more in-depth analysis of the results, via for example wind speed distributions at relevant sites and heights will be much more exciting and valuable. I don't see the point of the analysis of potential temperature either.

My recommendation is that the manuscript might be acceptable after significant revisions: explaining and supporting the decisions made in the model setup and expanding the analysis of the long simulations.

In addition to the significant issues above, I also have some minor comments:

1. Page 2, L4 and elsewhere: should be "real case", not "real-case". The dash is not necessary.

2. Page 2, L8: "This representation of surface friction becomes critical...". Which representation? It is not clear; please explain.
3. Page 2, L10: I would write "simulation", and not "forecasts".
4. Page 2, L20: The term "huge" is not a formal adjective for scientific writing. Please replace by "large".
5. Page 3, Table 1: TSE is not yet explained in the text.
6. Page 4, L9: "The four wind scanners WS1 to WS4 were performing range-height..." is the wrong verb tense. It should be "The four wind scanners WS1 to WS4 performed range-height..."
7. Page 4, L20-21: "National Center for Atmospheric Research Earth Observing Laboratory (NCAR EOL) (UCAR/NCAR - Earth Observing Laboratory, 2019)" is defined here, but already used in L5.
8. Page 5, Figure 1. Sorry, but I cannot see the difference between the red and magenta dots. Maybe use a different symbol?
9. Page 6, L10. You write that you set the lowest model level to 80m and 40 m in D1 and D2. What was the rationale for that choice? PBL schemes are known to be sensitive to this height (see Shin et al. 2011), mainly because a height of 80 m could often be above the surface layer. Also, a model top of 200 hPa is too low for the possible convective atmosphere during the late sprint and not recommended by the WRF model developers (see WRF Best Practices, link below).
10. Page 6, L11-12: Do I understand that the ST and LT simulations did not use the same PBL scheme? Why? Is the choice of PBL scheme in the inner LES domains important? Please justify with appropriate references the WRF parameterizations used.
11. Page 6, L15-19: You do not give enough detail in the description of how the topography and land use of the WRF model domains were generated. Also, the provided namelists (which use USGS classes and gtopo30) do not match the description in the manuscript (CORINE and ASTER). What kind of filtering (interpolation type and smooth options) was used? How accurately the D3 and D4 terrain match the real topography of the Perdigão site?
12. Page 7, L22. "amount of the ... wind vector", maybe the "module" will be a better word?
13. Page 8, L6: I think the equation is better written as Lm = 1.69 (LAI/h). Constants should come first.
14. Page 8, L9-10: You write "... LAI is retrieved from the CORINE land use dataset". This statement is incorrect. CORINE is only a land-use dataset, and there are no LAI values associated with it. Do you mean that you used the WRF landuse table? Does this include the seasonal variations? Besides, there are several versions of the CORINE dataset. Please provide a reference. Also, WRF has only tables for USGS and MODIS land categories. How did you use the CORINE data?
15. Page 8. Why did you use a random perturbation to the height of the forest? Laser scans of the area of Perdigão exist and were part of the field experiment data. As far as I know, the data is available from the Univ. Porto.
16. Page 8. I am missing a section that describes how the WRF model output was processed to compare with the lidar and mast measurements. There is a line in Figure 5, but I would like to see more. For example, what height was used in the vertical interpolation?

17. Page 8, L15-16: "...the double ridge is completely covered by trees in the model ...". Do you mean trees cover the two ridges and the valley?
18. Page 11, L5-6: Why not write "Fig. 5-6 and Fig. 7-8 show snapshots of..."
19. Page 12, the caption to Figure 4: The colours are not contours, so you should not write "colour contour interval: 0.5 m/s". You provide a colour scale for that. Also, the time of the last cross-section does not match the one in the figure caption.
20. Figures 4-8: What is the source of the shaded hill at the bottom of each cross-section?
21. Page 21-22: The discussion regarding wind power densities in the conclusion section seems irrelevant at this point. Yes, winds are stronger without the forest parameterization, and they will give much larger power densities, there is no need to emphasize this one more time. A couple of sentences would be plenty.

References:


Warner, T.T., 2011: Quality Assurance in Atmospheric Modeling. *Bull. Amer. Meteor. Soc.,* **92**, 1601–1610, https://doi.org/10.1175/BAMS-D-11-00054.1

WRF Best Practices: https://www2.mmm.ucar.edu/wrf/users/tutorial/201907/chen_best_practices.pdf

Shin et al. 2011: Impacts of the Lowest Model Level Height on the Performance of Planetary Boundary Layer Parameterizations, https://doi.org/10.1175/MWR-D-11-00027.1