Referee comment on "Evaluation of a forest parameterization to improve boundary layer flow simulations over complex terrain. A case study using WRF-LES V4.0.1" by Julian Quimbayo-Duarte et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-274-RC1, 2021

The authors present a model evaluation manuscript for evaluating WRF in LES mode above a forest in the vicinity of a double ridge. Modeled wind component velocities are compared to observations at 100 m tall meteorological towers and LIDAR observations. The authors find that using a forest parameterization in the LES improves the agreement between model and observations. The applied forest parameterization is a canopy aware LES in which canopy elements are modeled as a drag coefficient for model layers that are assumed to be within the forest canopy (2-3 bottom most layers). While I don't think that there are technical errors in the manuscript, I have several comments and questions about their work that the authors should address before final publication.

General comments:
1) Canopy aware LES (also sometimes called canopy resolving LES) has become reasonably mainstream in the last 10 years, with many community LES (i.e. WRF, DALES, PALM) having developed capabilities to represent canopy drag forces. It have been shown that such treatment, is capable of improved representation of turbulent statistics within and above the canopy (e.g. TKE, sigma_w, sweeps/ ejections). It has been shown that a sufficient number of within canopy model layers are needed to do so (order 10 layers) and that results are best when model resolutions are isotropic (i.e dx=dy=dz). Given the current setup of 2-3 model layers and dx/dy of 200 m in d03 (which would produce a grid aspect ratio of roughly 10:200), I am wondering to what extent the canopy drag formulation would be capable of improving representation of turbulent structures above the forest.
This brings me to my main question, which I could not find addressed in the manuscript: To what extent is the comparison between the forest parameterization and non-forest a true apples to apples comparison? The non-forest LES uses $z_0$ from the land-use dataset (I assume this is the built in WRF one, that may not be very accurate at such high resolutions), while the forest parameterization uses a direct sink for momentum. A better comparison may be to apply a similar momentum sink to the first model layer only, or to first calibrate the model's $z_0$ to observed values.

2) LLJ cases: The model evaluation against the LIDAR data is done as instantaneous cross section plots. While I can see that there the forest parameterization exhibits more dampening of lee-waves and has somewhat lower near surface wind speeds, it is generally hard to make out finer differences in the plots. I would suggest that the authors think about how to better visualize these model results. For example, one might want to focus on the d04 domain and would also want to zoom in closer to the surface. Alternatively, difference plots would also be appreciate. Lastly, I am wondering, given the 5 minute output interval, why comparisons are done for a single snapshot in time. Given the amount of output data, it would be interesting to see whether there is a better quantitative comparison possible. Furthermore, Wagner et al 2019a presents a model evaluation paper with respect to d03 domain in non-forest mode. Compared to the model evaluation for d03 in the current manuscript, this evaluation appears to be much more detailed and I am wondering why this is the case.

3) I am not familiar with the data from the field campaign that is available to the authors, but I am wondering whether there is a missed opportunity in not having any kind of turbulent quantities that the LES is evaluated against (e.g. from the towers or the LIDARs). Usually, the fact that LES are capable of partially resolving turbulence and to provide realistic profiles of turbulent statistics that can be compared against observations is seen as one of the crucial advantages of the LES. In this manuscript, there is no consideration of turbulent quantities, which could be used to evaluate the actual LES.

4) Relationship to retracted Wager et al 2019b: The authors mention in the submission that this work builds on the retracted Wager et al 2019b work. A cursory look at the discussion of this manuscript shows that two reviewers had issues with the WRF setup in the sense that they would have liked to see either a nudging or periodic restart for the long-term simulation. I would like to know whether these comments have been heeded in the present work. Similarly, reviewers of the W19b manuscript would have liked to see the full namelist of the WRF for added transparency/reproducibility. I feel that something like this would be particularly valuable given the objectives of GMD. Lastly, since W19b was retracted, it should not be referenced in this manuscript.

Specific comments:

P1L14: "The grids are now fine enough to fully resolve the atmosphere with techniques such as large eddy simulation (LES)" > I take issue with this statement, as it is incorrect. LES partially resolves turbulence, but this is nowhere near fully resolving (especially given the model resolutions of 200 and 40 m in this paper). This needs to be rephrased.
Introduction: The authors present a literature review of canopy aware LES, that is somewhat odd and starts with Dutton et al 2008 (they do cite Shawman and Schuh, 2003 later), but there are plenty of papers that could/should be referenced including:
Shaw & Patton, 2003: https://doi.org/10.1016/S0168-1923(02)00165-X
Dwyer et al. 1997: 10.1023/A:1000301303543