.

Response to the comments from referee 1

Thanks for your comments on our manuscript. Here our response to each of your comments. The response is given in blue color.

Best regards,
The authors

General comment

Very complete assessment of the WRF-LES model in idealized conditions covering mean profiles and spectra at a 250-m boundary-layer mast. The main challenge in the study is in the inherent inconsistencies in the comparison of canonical ABL simulations with ensemble-averaged profiles. The study is based on the assumption that the observed profiles represent the horizontally homogeneous conditions that the LES model is based upon. This is not convincingly addressed or quantified so it deserves some additional discussion. The validation is quantified using normalized quantities based on u*. A suggestion is made to use other set of normalized profiles to avoid the intrinsic bias in u* in the computation of the RMSE.

Thanks for your general comment. We address how representative our measurements are when compared to canonical flows when responding to your first specific comment and your suggestion on normalizing the profiles in a different manner later when responding to the related specific comment.

Specific comments

P3. Section 2.1: I think you should justify why doing long-term statistics leads to mean profiles that can be considered good references for idealized ABL simulations. While bin averaging with long-term data will definitively help in the convergence of the mean profiles you don't guarantee that you are removing persistent heterogeneities from the mesoscale wind climate. To come up with reference profiles for canonical ABL you would want to filter out situations that deviate from quasi-steady and horizontally homogeneous conditions. Have the authors considering any filtering for such conditions? By filtering more you can decrease the value of the std-bars in the observed profiles to values that are closer to the ones from LES (e.g. Fig. 7 shows this discrepancy in the error bars). Even if you provide a reference about the measurements, please describe the case selection in more detail with regards to the binning process and number of samples in each case and discuss how you deal with these issues.

As the reviewer points out, we cannot guarantee that there are no mesoscale trends in the observations and as the reviewer suggests, one way of filtering out those periods is by deriving mesoscale tendencies from real-time WRF simulations. However, this is beyond the scope of the paper as we only performed idealized simulations. A discussion of this possibility is now added in the Discussion section. Following the recommendations of the reviewer, we now also extend Section 2.1 (with the description of the selection of the flow cases from the measurements) so that information with regards to the binning process, extension of the analyzed periods, and the amount of cases per stability condition is provided. It is important to note that in the error bars we show the standard deviation and not the standard error, which due to the large number of samples is very small. This is because here we are interested in the variability of both observations and simulations. Note also that the measurements extend over a 4 year period and so the statistics on each stability regime are robust. Filtering can indeed increase the variability of the observations; in Fig. 1, we show that when we only use 10-min periods that are between three consecutive 10-min periods showing the same flow characteristics (a simple stationary filter), the variability highly increases for neutral conditions. For these conditions, about a third of the data are filtered out with such a filter, and for stable and unstable conditions is nearly a fourth and one-eighth of the data, respectively. Also, if there was a strong mesoscale trend in the observations, it will be reflected in the low frequency range of the velocity spectra. However, it is particularly in that range where both simulations and observations compared the best. This is also now part of the discussion in the revised manuscript.

P4,83: Can you discuss if these stability functions have been found at the Østerild site and if they are also used in the surface layer model of WRF? Together with the roughness length, this may be relevant in the interpretation of surface fluxes.

These functions have been tested at Østerild (actually as it can also be seen in the normalized wind profiles in
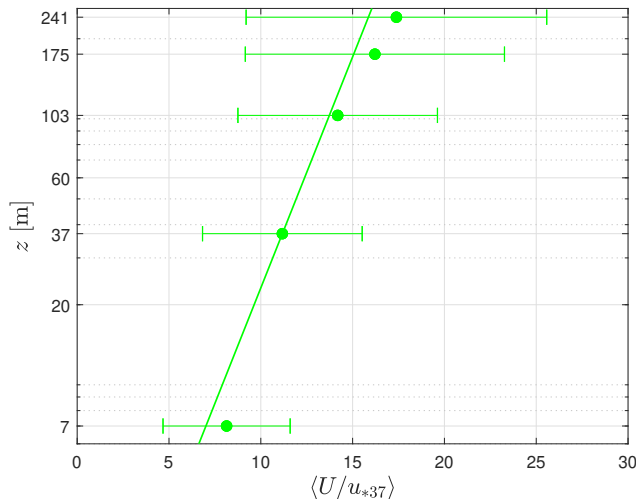
Figure 1: Normalized wind profile at Østerild with a simple non-stationary filter

Fig. 1 in the manuscript). And these functions were also used in the work of Peña (2019) at Østerild. This is now stated in the manuscript. As for WRF, the surface-layer scheme uses the same function under unstable conditions but a slightly stronger stability correction under stable conditions. However, at the first model level ($\approx$5 m), the difference between the correction from WRF and from this function is lower than 2%.

P4.Setup: There is no mentioning of any sensitivity analysis on the grid settings that would support the selected grid configuration. While the setup looks reasonable it would be better to include some discussion on the adequacy of these settings for canonical ABL simulations. In particular, it would be interesting to compare with the parametric study of LES model settings done by Mirocha et al (2018) that also includes idealized WRF simulations.

We did not perform grid sensitivity studies in a systematic matter and used this grid resolution and configuration based on the grid sensitivity study by Mirocha et al. (2018). They also found that the optimal grid aspect ratio $\Delta x/\Delta y = 3$ provided good comparison with two other models and with observations. As suggested, we now state this in the manuscript.

P5.Figure 2: The x-axis is the vertical level height. For those not familiar with WRF grid, please indicate if this height represents the cell center or the cell height.

As suggested, we now add this information in the caption of the figure.

P5.113: Can you discuss how you come up with the values of geostrophic wind (14 m/s in neutral and stable conditions and 8 m/s in unstable conditions)? I would suggest using a table to collect all the input quantities that you use to define the three flow cases.

This information is provided in Sect. 2.3 Setup: "The initial u_x-velocity for the three simulations was chosen so that it was close but slightly higher than the observed ensemble average of U at each of the stability regimes at the highest vertical level.". As suggested, we now add a table collecting the inputs for each of three flow cases.

P6.144: "The choice of the time to extract LES statistics depends on the type of boundary layer." Can you elaborate on how you select this time? Based on the discussion later on, the selection seems a bit arbitrary although I understand that the profiles never reach a steady state. Maybe, at this point, you should discuss this challenge. My impression is that you end up choosing profiles that have sufficiently developed to a quasi-steady state close to the values you have from measurements and "have the right look" for canonical ABLs.

As suggested, at the end of Sect. 3.1.3 we now discuss the challenge on the selection of the time due to the inherent type of simulation: "It is important to note that it is challenging to select the time for extracting the statistics due to the inherent unsteadiness of LES. For neutral conditions, we can objectively choose the period, but for unstable and stable conditions, this is a combination between reaching a quasi-steady state and the fluxes values observed at the surface mainly."

2

P7.Fig 3: Can you indicate (in the caption or in the legend) which time in the simulation corresponds to the blue line? A vertical line in the time series would also help.

As suggested, we now indicate in vertical lines the time of the selected period on the time series plots in Figs. 3–5

Figs 3/4/5: The TKE behaves similarly to the u* so you might as well skip it. In neutral conditions you may consider replacing it with the wind speed of the jet nose since this is used to select the neutral profile. In unstable and stable conditions you may consider replacing u* by the heat flux (as in Fig 5), or the Obukhov length (I'd use this one), to see how the profile evolves from the energy point of view.

As suggested, we now replace the tke by the jet nose speed time series in the neutral case. We also replace the tke by the heat flux time series in the unstable case. For the stable case, we now skip the tke and first model wind speed time series as well. The text has been modified accordingly.

P8.172: You should mention that the first model level is at around 5 m while the reference flux measurements are taken at 37 m. This difference may be significant in the value of the fluxes, specially in stable conditions. Why not using the closest level to 37 m (in all the plots) as the reference height for the time series? Why not showing the observed values of u* and heat flux in Figs. 3 and 4? In principle, instead of using the u* value produced by WRF at the surface, you could compute u* from the high-frequency time series as if they were sonic measurements to try to represent the same quantity (as it is done in section 3.4 for the fluxes). To improve consistency, you could introduce filters in the measurements to remove mesoscale trends and mimic the behavior of the sgs model in LES (as the low-pass filter displayed in the spectra). I understand that this is beyond the scope of the paper at this point but I think it is worth discussing how "canonical" are the measurements when we compare with idealized LES simulations since these quantities integrate different turbulent scales in simulations and measurements.

As suggested, we now remind the reader (at the end of the first paragraph in Sect. 3.1) that the observed fluxes are at 37-m height, whereas the surface simulated fluxes are derived from the first model level (about 5 m). We choose to show these simulated fluxes (and not those close to 37 m) because of what we have now further elaborated in Sect. 2.1 Selection of flow cases: by using the the 37-m fluxes at Østerild, the behavior of the dimensionless wind shear with dimensionless stability follows surface-layer scaling. Also, as suggested, we now show the observed fluxes in Figs. 3–5. It is important to notice that the roughness values are estimated based on the observed quantities at 37-m height as surface-layer scaling holds there. In the LES, surface-layer scaling is applied at the first model level and so although the heights between observations and simulations are not the same, the method to derive roughness (from observed fluxes) and that to simulate fluxes (from roughness) is consistent. This is also added in the text.

P11.Fig. 7: Because of the differences in the reference heights of u_* (and the underlying filter in LES), you are probably introducing a bias in the normalized profiles. Although less formal, maybe a simpler but more robust way of normalizing the profiles would be using the wind speed at 37 m instead of the friction velocity.

Normalizing by 37 m does not reduce the error bars much at the levels most far apart (237 m) but the idea of scaling with u_* is also to test that there is consistency between the observed and simulated fluxes (see the response to the previous comment).

P2.203: Please put the metrics of this section in a Table or put them together with those in Table 1.

As suggested, we now add the metrics in Table 1.

Figures 9,10,11; Table 1: As discussed before, these plots and error metrics depend on quantities that are normalized with the u* value, which may come with a bias due to using a different vertical level in measurements and observations. In addition, as pointed out by the authors, u* in the simulations depend on the value of the roughness length that is used as input which is difficult to quantify in reality. For these reasons, I think it is better to choose a reference height and normalize each profile by the corresponding value of each quantity at that reference height, e.g. U/U_{37}, uw/uw_{37}, etc (and exclude the 37 m level in the computation of RMSE).

Please, see the response to the the last three comments.

P18, 305: "The simulated means are always within the observed variability"... well, the variability is pretty large in the measurements, isn't it? Still, the agreement is pretty good considering the uncertainties in the

definition of "canonical" flow cases.

As suggested by the reviewer we now replace the sentence by: "The simulated means are always within the observed variability, but it should be noted that the latter is large; the observed variability at Østerild is lower than that from the 'canonical' flow observations performed at the 200-m tower at the SWiFT test facility in the US Southern Great Plains (Mirocha et al., 2018)"

P22.344: When running real-time simulations you can definitively extract mesoscale tendencies from WRF that you can use to filter out periods of strong heterogeneity. This can help you narrow down the case selection to obtain mean profiles that more closely match the idealized conditions of this study.

This is a good idea, which we will consider for future studies. This is also now discussed in the Discussion section

References

Mirocha, J. D., Churchfield, M. J., Muñoz Esparza, D., Rai, R. K., Feng, Y., Kosović, B., Haupt, S. E., Brown, B., Ennis, B. L., Draxl, C., Sanz Rodrigo, J., Shaw, W. J., Berg, L. K., Moriarty, P. J., Linn, R. R., Kotamarthi, V. R., Balakrishnan, R., Cline, J. W., Robinson, M. C., and Ananthan, S.: Large-eddy simulation sensitivities to variations of configuration and forcing parameters in canonical boundary-layer flows for wind energy applications, Wind Energ. Sci, 3, 589–613, 2018.

Peña, A.: Østerild: a natural laboratory for atmospheric turbulence, J. Renew. Sustain. Energ., 11, 063 302, 2019.