Review of “Evaluation of large-eddy simulations forced with mesoscale model output for a multi-week period during a measurement campaign” by Heinze et al. submitted to ACP.

This manuscript reports analysis of multiday LES simulations of realistic weather phenomena. The forcing comes for a NWP model and results are compared to variety of observations because the simulation period coincides with extensive field observations. I feel this is a very interesting and useful study. It will no doubt lead to many similar follow-up studies where specific aspects will be tested more extensively. The focus in the current manuscript is on the boundary (BL) layer structure (BL height in particular). I find this a little unfortunate as I would like to see more emphasis on simulated clouds as this I feel cloud simulation was at least partly the motivation to go to LES. Perhaps hints offered by Figs. 3 and 4 are sufficient for now.

Overall, I feel this paper should be published after some revisions and clarifications in response to my specific comments below. Some of the comments concern aspects of the models and methodologies beyond the scope of this manuscript, so I do not expect them to be addressed in responses. I do not need to see the revised manuscript.

Specific comments:

1. P1 L20. The key problem with stably stratified BL is the turbulence intermittency and wave-turbulence coupling.

2. P2 L17. I do not think the discussion following the statement "in various single-column and cloud-resolving modeling studies" has anything to do with CRM studies. I suggest revising it to read "in various single-column and LES studies".

3. Section 2.1 and in other places in the text. The terminology applied is not correct. The Boussinesq approximation refers to applying density perturbations only in the gravity term. Thus, most (if not all) nonhydrostatic atmospheric models are Boussinesq (both anelastic and compressible). A more appropriate distinction would be shallow convection approximation (in which case the incompressible assumption is valid and density is constant)) versus deep convection approximation (in which case anelastic or compressible equations are needed). So PALM applies shallow convection approximation (and obviously cannot work properly for deep convection), whereas UCLA-LES is anelastic.

4. A more general comment on model variables (no need to respond, just think about this for the future). Liquid water potential temperature is a perfect variable for shallow nonprecipitating convection. For deep convection, two problems arise. One, the exact formulation is rather cumbersome and I doubt it is used in the two models. Second, when precipitation is considered, then it is not conserved and its sources need to be considered. In contrast, equivalent potential temperature (or moist static energy, MSE) is conserved as long as ice processes are excluded. Should then MSE be used?

5. Another comment on the model. I do not think the water variable is the “total water specific humidity”. I think it is (or should be) the mixing ratio. Please do a simple math

exercise with the density of total water, dry air density, and total air density to show that mixing ratio is an appropriate variable when sources (e.g., of water vapor) are considered. The continuity equation for the specific humidity has an extra multiplier for the source term. The equation for the mixing ratio does not, and thus is preferable. Of course in practice the differences are miniscule and can be neglected.

6. P. 4. I think the fundamental differences between Eulerian thermodynamics in UCLA LES and Lagrangian approach in PALM should be better exposed in point 4. These are more significant than the authors realize I think (see below). Also, is PALM really applying saturation adjustment (the paragraph starting on P4L15) and Seifert/Beheng 2-moment microphysics? This is the Lagrangian Cloud Model, so it uses superdroplets, correct?

7. This is perhaps the most significant comment. Formulation of the forcing methodology in section 2.2 comes out of nowhere. I realize that similar methodology has been used by others, but I still think that the methodology can be better explained (perhaps with the exception of the nudging). Please have a look at section 2 in Grabowski et al. (JAS 1996, p. 3684-3709) that formally derived the forcing terms in case of evolving large-scale conditions and later applied it in cloud-resolving model simulations. Perhaps some of the forcing terms derived in that paper are missing in the way LES is forced. Just a thought…

8. The domain shown in Fig. 1 features quite a significant topography. Is then the periodic domain without topography justified? Can one use COSMO output to show if the topography has some effect (e.g., by comparing COSMO results north and south from the observation sites). Does the hill NE to the observation sites affect the observations? Are any observation made from the top of the hill?

9. Section 2.3 does not mention how surface fluxes are calculated. Is the same surface temperature and humidity assumed throughout the LES domain? How variable are surface conditions (e.g., soil type, soil moisture) over the LES domain. Do such considerations matter?

10. I like section 2.4. I think it is important to realize the magnitude of various terms and their evolutions. A small technical comment: the red and orange lines are not easy to distinguish. One of them can be green.

11. Fig. 3 is very nice. It also shows that BL tends to be systematically colder in LES than in observations. Is that a coincidence, or is this true for the entire simulation?

12. P. 12, paragraph below Table 2. I think the origin of the difference between clouds simulated by the two LES models goes beyond just the advection scheme. See comment 6 above. For instance, PALM predicts supersaturation, correct?

13. P13L35. I think this is 5 cm/s, not 50 (this would agree with captions to Fig. 5 and 6).

14. Fig. 6. Despite taking entire page, details of the figure are difficult to see. I suggest breaking the figure into 3 separate figures.

15. The integrated water vapor is commonly referred to as the precipitable water (PW).

16. P20L12-16. I am not sure about the discussion of mesoscale circulations. Since the model excludes topography and (I assume) applies homogeneous surface conditions, the simulated mesoscale circulations may be significantly weaker than in nature.

17. Figure 8 shows that impact of some of the changes is smaller than standard deviations which suggests that the differences may not be statistically significant. How is the standard deviation defined in those plots?