

Although numerous changes have been made to the manuscript that helped the readability and altogether clarity, which I thank the authors for, nevertheless there are still several significant inconsistencies that became even more obvious after the revision. Some of the answers were not satisfactory. For this reason, I suggest more major revisions before accepting this manuscript.

Major comments:

1. I believe, that even when using three different hours (three timesteps), the results are not robust. The simulation still covers only one flight, and I don't understand how adding two additional hours (Fig 3.) shows that the results are more robust. Quite the opposite, you clearly state that the wave field is mostly stationary, however parameterized TKE from TE parameterization changes more significantly. How are any following comparisons reliable then? I understand that you stated that more comprehensive statistical evaluation is planned for future work and that it is beyond the scope of this project. However, in that case it raises concerns regarding the significance of this manuscript.
2. You show comparison between the flight legs and simulations, where in the lower leg, there is some agreement between the two, but there are strong differences in the upper leg. In the observation, amplitudes of the waves stay the same or at some places possibly even grow – something we would expect in the absence of the wave breaking/dissipation. However, in the simulations, the amplitudes of the waves decrease significantly, suggesting that they dissipate more. Here would be useful comparison at different time with different flight as it would show if that is a consistent bias of the simulations or something specific to the current meteorological conditions for example. You state, that the observations are used not for thorough validation, but to estimate if the simulations provide realistic flow structures. But even that is hard to estimate just from Figure 4. You could provide other variables from both observations and simulations such as u and v , to estimate if the meteorological situation is well represented in the model. That would help to distinguish if the different wave characteristics in the upper leg are connected more to different winds or for example the turbulence parameterization. If the simulations do not agree with the observations, although that is to some degree expected, there should be more explanation as to why there are such differences.
3. Throughout the manuscript, you state that TE scheme is better, with more realistic outcomes. I don't understand how you show that. From the Fig. 7 and Fig. 9, there is obviously less TKE being produced by TE scheme, but why is that more realistic? Wouldn't we also expect some TKE being present at the UTLS? From Fig.9 it seems there is close to zero TKE being produced by the TE scheme in the higher altitudes. From the comparison with the flight measurements, it seems that there is more dissipation of the waves aloft in the simulations, which

should suggest more TKE being present there. Other variables such as momentum flux or velocity variance were not conclusive either as they are quite similar between the schemes due to very similar wave field. It would be useful here for example to calculate gravity wave drag from the momentum flux to diagnose if the levels of heightened TKE in the TKE scheme agree with the deposition of the drag.

Minor comments:

1. In the appendix, could you add definition of what the subscripts 1 and 2 mean in the equations A2-A4?
2. If the TKE values from the TKE scheme are indeed overestimated, you could use equation A1 to calculate the TKE values from the LES simulation. Although, as you stated, it wouldn't provide useful information in the higher altitudes, it could help validating the TKE from the other simulations near the surface.