Review of "Can machine learning improve the model representation of TKE dissipation rate in the boundary layer for complex terrain?" by Bodini et al., for Geophysical Model Development.

Spuriously represented TKE dissipation rates in numerical weather prediction models are known to affect simulation results, especially for complex terrain. In the presented manuscript, this problem is addressed by investigating if machine learning techniques can help to improve the representation of the TKE dissipation rate in comparison to established parameterization schemes. For this purpose, the authors first demonstrate the deficiencies of the commonly used Mellor, Yamada, Nakanishi, and Niino (MYNN) parameterization for turbulence measurements, collected at 184 sonic anemometers during a 6-week field campaign in Perdigão, Portugal. Afterwards, three different machine learning methods are trained on the same dataset and the results are compared to MYNN. The study shows that the systematic bias of MYNN under stable conditions is significantly reduces with machine learning techniques.
The study is within the scope of GMD and addresses a relevant and interesting topic for the modelling community. The manuscript is well structured and comprehensibly written. Therefore, the paper merits publication after a few corrections.

General Comments

- it is surprising that land use and topography have almost no impact on the random forest algorithm. This feature of the machine learning algorithm is in contradiction to the actual importance of land use and topography on turbulence in nature, as already stated in the introduction. The authors should discuss in more detail this low sensitivity and give possible reasons. For instance, by looking at Figure 7 it can be seen that all measurement sites are located within or at the borders of a valley. Does this lead to a channeling of the wind field and consequently only to two occurring wind directions (more or less) in the dataset. This would result in a low upstream variability of h_{veg} and std(z_{terr}), possibly explaining their little impact on the random forest algorithm.
Is the impact of land use and topography also small for the other machine learning algorithms? A simple way to assess the sensitivity w.r.t. h_{veg} and std(z_{terr}) would be to just omit them as input features and look at the effect on RMSE and MAE. Did the authors do that and if yes, what was the outcome?

- If the low impact of land use and topography on turbulence in this study is caused by a channeling effect of the wind field, the question arises how representative the results really are. Against the background of an intended implementation of machine learning techniques into numerical weather prediction models (as stated by the authors in the conclusions), it is necessary that the method can be applied on a variety of different land use and topography conditions. The authors should therefore discuss in a bit more detail than they currently do in the conclusions how to achieve this. What are e.g. the data requirements that need to be fulfilled by other measurement datasets to account for the impact of different land use and topography conditions?
Furthermore, how would one incorporate the results of the machine learning algorithms in a numerical weather prediction model? In their reply to reviewer #1 the authors say that the model weights cannot be directly determined – but isn' t that just what one would need?

Specific Comments:

Lines 48 and 424: cite the accepted paper (Leufen & Schädler, 2019)

Lines 58, 105 and 435: change 'Nakanish' to 'Nakanishi'

Line 200 (Eq. 12): I guess there should be an n as upper limit in the summation over k.

Line 319: change ‘seems’ to ‘seem’.

Line 330: omit ‘ultrasimple’

Figures 3, 4, 8 and 9: I don’t think ‘density histogram’ is the appropriate name for this kind of scatter plot.