Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-200-RC2, 2021


Geoscientific Model Development Discussions

# *Interactive comment on* "Machine learning models to replicate large-eddy simulations of air pollutant concentrations along boulevard-type streets" *by* Moritz Lange et al.

**Anonymous Referee #2**

Received and published: 15 May 2021

This is a very well-conducted study and in my opinion, it is warranted for publication for the following reasons: 1) the main idea (regressive model identifying useful features for time-averaged pollutant prediction on 2d grid from LES data) is sound. 2) The methodology is well defined. The reasononing behind the choice of train and test datasets are explained. A well-established feature selection method is used. The pool of candidate features is also extensive. 3) The "concept detection" in section 3.4 is another interesting check on the validity of model.

Some comments for the authors to think about and address:



1) The issue with the negative output of the models is puzzling. The authors scale input features. Why not use a scaling of the target variable such that the output data is projected between [-min, max]. Every time the model is queried, the output of the model can go through the inverse transform of the scaling to scale it back to physical values. Could this solve the clipping of the output problem?

2) SVR results seem pretty reasonable too. In fact, they have equal num features with log-norm SVR and smaller bias and RMSE on PM2.5. Is there a reason the authors have chosen not to use/highlight that model

3) The dummy model while being a simple reference (no feature selection needed, no training, etc.). It is too much "dummied down" in my opinion to draw comparisons against, especially in the conclusions and abstract. My suggestion is to take the linear reg as the baseline. Not much would change as your three highlighted models still perform better in at least one sense (fewer features, better RMSE on one of the two tests). This would be a stronger comparison and reflects better on your results.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-200, 2020.