

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2022-134-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-134

Anonymous Referee #2

Referee comment on "Development of a regional feature selection-based machine learning system (RFSML v1.0) for air pollution forecasting over China" by Li Fang et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-134-RC2>, 2022

General Comments:

This paper presents the development of an improved machine learning based air quality nowcasting system. Instead of using all possible related features in the model training and predicting, they selected those general important and effective features. The feature selection is done using a computationally efficient ensemble method. Their nowcasting system is tested on the PM_{2.5} forecast on a national scale and validated to be superior than a CTM model and conventional MLs. Generally speaking, the paper is clearly written and well structure, their results are scientifically solid. I recommend accepting it after a minor revision. I also have questions and comments for the author that could help to improve their manuscript.

Major comments:

They have tested their regional feature selection-based ML nowcast system at a national scale and using several common ML models (RF, GB and MLP), which makes their results very sound. However, it is only tested at a 2019 winter season. I understand winter is the most severe polluted season there. Their system should be able to operate in a rolling forecast way. If extra training and testing are conducted at a less-polluted period/season, this study would be an excellent paper.

They should also explain the current machine learning model cannot fully replace model-based air quality forecasting systems, as ML models could not be trained and operated without inputs from the historical measurements. While for many rural regions, they are unavailable. The authors should explain this point clear.

Specific comments:

Page 6, Table 1: esolution to resolution

Page 12, line 5: computational complexity?

Page 17, line 1: and at a predicting horizon?