

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2022-164-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-164

Anonymous Referee #1

Referee comment on "Development of an LSTM broadcasting deep-learning framework for regional air pollution forecast improvement" by Haochen Sun et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-164-RC1>, 2022

In this paper, the authors introduce an air pollutant (ozone and PM_{2.5}) forecasting model system which based on the deep-learning method. With the implementation of ground observations and the outputs of 3D chemical transport model, this model can forecast more accurate concentrations of ozone and PM_{2.5}. Further, this model system can extend the prediction of air pollutants from individual station to a regional forecasting by considering the temporal characteristics of the time series and spatial relationships among different stations. By comparing the result with the observations, the results of this model show a more accurate and reasonable distributions of ozone and PM_{2.5}, which indicates that this model system can work as a feasible and efficient option to improve current forecast performance. I think the authors did an interesting work. And this study is within the scope of GMD journal. Some problems need to be solved before it can be published.

Comments:

- "The ground monitoring stations with at least 90% valid records" (Line 92~93) and "the ground monitoring stations with at least 95% valid recordswere selected as the source stations" (Line 95~96). How are these threshold values determined? Please describe detailed information about these threshold values determination methods.
- the data of source stations is also used for model training. What is the difference between the data of source stations and which of training stations?
- Line 114. I am confused that why the authors use the WRF-CMAQ data with the next two days to training the model?
- In this model system, large quantities of observations are used for training model. Please discuss the impact of the number of training stations and source stations on the model performance?