Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-253-RC1, 2021

Comment on gmd-2021-253

Anonymous Referee #1

Referee comment on "Deep-learning spatial principles from deterministic chemical transport models for chemical reanalysis: an application in China for $PM_{2.5}$" by Baolei Lyu et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-253-RC1, 2021

In this paper, the authors developed a deep-learning based method for fusing observational data into simulated air pollution concentration fields. The method requires considerable amount of efforts in preparation, but model execution is fast and the results are favorable, making it suitable for operational air quality forecasting platforms. Overall, this paper is well-structured, fluently written, and the topic fits the scope of this journal. I only have a few comments:

Is there any reason why only RH and WS were included as meteorological variables, but other important variables such as precipitation and boundary layer height were not included? Is it because of limitations in computational resources?

In model training, actual observation data is not used, rather a random sampling of 1500-2500 data points were used. In evaluation, actual observational data were used. It would benefit the readers if the authors could provide more justifications on: 1) Why 1500-2500 data points were chosen, and why the number of data points varies among years; 2) Why random samples of data points were used in training, instead of actual observational data; 3) Will the biases and errors in CTM simulation impact training results?

In lines 173-174, the authors mentioned that the kernels are generally isotropic but some anisotropic characteristics are evident. What are the expected impacts of such anisotropicity? Does including additional training variables such as wind direction help addressing this anisotropic issue?