

Geosci. Model Dev. Discuss., referee comment RC2  
<https://doi.org/10.5194/gmd-2021-253-RC2>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on gmd-2021-253

Anonymous Referee #2

---

Referee comment on "Deep-learning spatial principles from deterministic chemical transport models for chemical reanalysis: an application in China for PM<sub>2.5</sub>" by Baolei Lyu et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-253-RC2>, 2021

---

### General Comments

In the manuscript, a new data fusion paradigm is developed to estimate PM<sub>2.5</sub> reanalysis fields from station observations by a deep learning framework to learn multi-variable spatial correlations from Chemical Transport Model (CTM) simulations. The model includes an explainable PointConv operation to pre-process isolated observations and a regression grid-to-grid network to reflect correlations among multiple variables. Compared with previous data fusion methods of PM<sub>2.5</sub> reanalysis, the proposed fusion framework can fuse multi-variable observations from different monitoring networks (even when they are not spatially aligned at collocations) and the model training does not rely on observations. The deep learning data fusion model framework is novel and can reasonably generate spatio-temporally complete fused fields of PM<sub>2.5</sub> using observations at sparse locations. I would recommend publication in Geoscientific Model Development after consideration of the following comments.

### Specific comments

1. For the proposed fusion framework, why are only the predictions of PM<sub>2.5</sub> concentrations, relative humidity (RH) and wind speed (WS) together with the surface height of Digital Elevation Model (DEM) and land use and land cover (LULC) used to train the deep learning network?
2. Line 247: "This model was fitted with model simulation data by learning daily spatial patterns from long-term CTM simulations." When applying the fusion model, how long period of CTM simulated data is required at least for the network training to obtain the simulated spatial correlations?

3. Although, as it is said in lines 258-260, CTM simulation theoretically do not need to be very accurate in the model inputs, an accurate or reasonable spatial correlations (or spatial patterns) simulated by the CTM models is necessary for the model deep-learning. There are very limited information on the CTM simulation data used in the study. Have the simulated PM2.5 spatial patterns been evaluated? How about the performance? Please give some necessary introduction or relevant reference.

#### Technical comments

1. Lines 83-84: "Each of these data items at each were assigned...", the word of "site" or "station" is missed after "at each".

2. Line 185: "(Figure S2 in the SI)", Figure S2 is not found in the SI. Please check it.