

Geosci. Model Dev. Discuss., referee comment RC2 https://doi.org/10.5194/gmd-2021-356-RC2, 2021 © Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.

## Comment on gmd-2021-356

Anonymous Referee #2

Referee comment on "Development of a deep neural network for predicting  $6\Box$ h average PM<sub>2.5</sub> concentrations up to 2 subsequent days using various training data" by Jeong-Beom Lee et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-356-RC2, 2021

In this paper, the deep neural network (DNN) model, observation data and prediction data of numerical model are used to predict the average fine particulate matter (PM2.5) concentration in 6 hours for a three-day period. The performance of DNN model is compared with that of Korean multi-scale air quality (CMAQ) modeling system. The work is reasonable and complete. In general, the reviewer finds this article well-prepared with decent documentation of their motivation, methods, and results. It is recommended that the article be accepted after the following modifications.

Major Comments:

1, Line 75~80: "In addition, the membership function was used to reflect temporal information." More information is needed about "membership function". How does this function reflect temporal information?

2, Line 145-155: The authors want to predict PM2.5 within 3 days. Why do you need to add the time information ("adjacent month") of the next month that hasn't happened yet in Eq. (2)? If you know the information of next month, you can predict PM2.5 of next month. Why only forecast PM2.5 within 3 days. This is very difficult to understand.

3, In Section 2.2, what are the super parameters of DNN model? Why only use fivestacked-layer DNN model? Generally, a neural network model with more than 8 hidden layers is considered as a deep neural network (Hinton et al., 2012).

4, Line 210~215: The input data of the three experiments (DNN-OBS, DNN-OPM and DNN-ALL) are not very clear. Why should the predicted PM2 5 into the model (DNN-ALL)? Reason?

5, Line 230~240: There's something wrong with Eq. (18). The commonly used expression for F1-score is (2\*ACC\*Recall)/(ACC+Recall). F1-score is for one category. My understanding is that there are four categories (Good, Moderate, Bad and Very bad). Has anyone else used it like this? More explanation is needed.

6, In Table 2, why are 925hPa and 850hPa variables selected? Why not consider 700hPa and 500hp variables? Is there any reason?

7, Table 5 only provides the performance of the model in the test set (January–March 2021), and it is recommended to add the performance of the model in the training set (2016 to 2018) and validation set (2019).

8, In Table 5: The DNN-ALL model uses the forecast variable (F\_PM2.5 predicted by CMAQ). However, IOA of F\_PM2.5 in CMAQ is 0.9, 0.9 and 0.85 respectively, and IOA of PM2.5 in DNN-ALL is 0.95, 0.9 and 0.86 respectively. Could it be understood that compared with CMAQ, IOA in DNN-ALL model is improved by 0.05, 0.0 and 0.01 respectively? More explanation is needed.

9, In Table 6: From T04 to T11, why does the indicators (RMSE and IOA) not decrease monotonically? The IOA of T09 is larger than T04.

Meanwhile, the mean IOA of D+2 is 0.79 ((0.77+0.85+0.74+0.80)/4.0), and IOA of D+2 in table 5 is 0.86, What are the reasons for the unequal values?

Minor Comments:

1, Line 19: "IOA" should be "index of agreement (IOA)". The first abbreviation needs to give the complete name. Please check other parts of the paper.

2, Line 100 Figure 3: Nested-grid is often used in models. It is recommended to combine figure 2 and figure 3 into one figure.

3, Add the temporal and spatial resolution of the variables in Tables 1 and 2.

4, The font in Figure 4 is too small to see clearly.

5, It is suggested to add the content between DNN model and other machine learning models (https://doi.org/10.5194/amt-14-5333-2021, https://doi.org/10.1016/j.scitotenv.2021.150338).