

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-420-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-420

Anonymous Referee #2

Referee comment on "A comparative analysis for a deep learning model (hyDL-CO v1.0) and Kalman filter to predict CO concentrations in China" by Weichao Han et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-420-RC2>, 2022

In this manuscript, the authors evaluated the performance of a deep learning model and Kalman Filter in predicting CO concentration in China. The topic is interesting and fits the readership of GMD. However, the manuscript lacks details in the description of method, which hinders understanding of the results. A more thorough literature review of the performance of CO simulation and assimilation is also needed. Additional experiments should also be performed to demonstrate whether the DL model has good performance in predicting CO, see specific comments below.

Title: please clarify whether it is CO emission or concentration.

L20-21: please clarify whether the statistics in E. China come from the training data. It is not clear to me what is the difference between E. China and the independent observation just based on the description here.

L27: what is the cause of this? Does DL model have less capability in capturing the extreme values?

L38-43: since this manuscript is focused on CO, I don't see it necessary to discuss NO₂, PM_{2.5}, and O₃ here. Instead, readers are more interested in what the performance of CO simulations are compared to different observations.

L45-49: similar comment as the previous one. Data assimilation of other species is not relevant here. What are previous data assimilation efforts for CO? There are many of them and should be summarized here.

L71: I think you should introduce the focused species right at the beginning. Also, why choose CO for this comparison? Can the conclusions for CO be extended to other species considering their similarities and differences?

L100: what is the total number of grids?

L154: cost function is not defined.

L158-160: it is not clear what is being done here. Please clarify.

Section 3.1: It seems to me that DL can capture the temporal variations in 2019-2020 because the seasonality is unchanged compared to the training data, but when the magnitude changes in the test period (2019-2020), DL cannot capture. I am still doubtful how much information one can get from this DL model.

Fig2 & 3: please clarify are these the 90% training stations?

L213: I don't think the performance in 2015-2018 is even worth mentioning if you are using the 90% training stations, since this is the training data. A good consistency during this period does not mean anything regarding the model performance in prediction.

L230-231: If that is the case, you should evaluate the model using only 2019 data to demonstrate how well the model performs in a normal case, and then another evaluation for only early 2020 to show how much emission reductions affect the prediction.

L247: what is the sensitivity of KF results to model and observation errors? Is the poor performance of KF caused by the inappropriately assumed errors?

L259: what is the difference between KF in the training and testing period? Did you also feed in no observations for KF during the testing period? If that is the case, I don't even think this should be called KF, since the simulation purely depends on GEOS-Chem and initial condition.

