The paper presents the OSSE experiments results to demonstrate the benefit of GEMS ozone observations in future applications. Both the methods and the data assimilation results with additional OMI and surface synthetic observations are well presented. However, improvement can be made if the authors can address the following concerns.

We thank the reviewer for the thoughtful comments. Our point-by-point responses to the comments are in blue. We have also carefully considered and addressed each comment in the revised manuscript.

General:

When the influence of assimilation frequency is investigated, it is not clear how data assimilation experiments with a longer assimilation time window of 3-hr are carried out. Are the hourly surface station observations averaged inside the 3-hr time window? Are the satellite data inside the 3-hr time window assumed to be valid at one particular instance? If so, when are they supposed to be valid?

Response: In Exp 5–8, we reduce the number of assimilated observations and only assimilate the daytime synthetic observations at discrete 3 h time steps (*i.e.*, 01:00, 04:00, 07:00, and 10:00 UTC) of the GEOS-Chem model rather than taking the average inside a 3 h time window. The satellite observations are assumed to be valid and assimilated at the corresponding hour when available.

The authors found that sometimes the data assimilation has negative effects. For instance, "In Japan and Mongolia, the assimilation of GEMS data generally contributes to a deterioration of simulated ozone and even counteracts the positive impact of surface observations when performing the joint assimilation." It is not impossible to encounter such cases. When this happens, it is probably worth to investigate the reason for such a behavior. With the current OSSE setting, it is probably not too hard to investigate the underlying causes.

Response: Accepted.

We have revised this part in the revised Section 3.1 (lines 299–303) as follows:

"However, the influence of assimilation efforts is complicated in East Asia, such as in Japan and Mongolia where the *a priori* ozone and its bias are relatively low. In this case, adding the synthetic GEMS observations results in a slight deterioration of simulated ozone and even counteracts the positive impact of surface observations when performing the joint assimilation. We attribute this partly to the improper specification of model errors and the spatial spread of observational information *via* transboundary transport."

Specific:

Line 20: It is probably better to replace “data assimilation better represents” to “data assimilation improves”

Response: Corrected.

We have replaced “better represents” with “improves”. Please see line 20 in the revised manuscript.

Line 22: RMSE is a accuracy metric rather than a precision measure.

Response: Corrected.

We have replaced “precision improvements” with “significant improvements”. Please see line 22 in the revised manuscript.

Line 113: Is “optimal estimation” the same as “optimal interpolation”?

Response: Optimal estimation is a regularized matrix inverse method based on Bayes’ theorem. It is commonly used in geosciences to solve different kinds of inverse problems, particularly for atmospheric sounding (Rodgers, 2000). Optimal interpolation, however, is a relatively simple but useful method of data assimilation, which is a particular kind of inverse problem (Brasseur and Jacob, 2017).

Equation 5: It would be better to use "y" for variables in observation space.

Response: Corrected.

We have replaced “$\hat{\boldsymbol{x}}^{obs}$” with “$\boldsymbol{y}$” in the revised Section 2.3. Please see lines 196–197 in the revised manuscript.

Line 196: Xap should be in a vector in observation space, but it appears as a state vector. It is better to clearly differentiate state and observation vectors.

Response: Accepted.

$\boldsymbol{x_{ap}}$ is the *a priori* profile (vector) used in the satellite retrieval procedure (Eq. 1 in Section 2.2). In the data assimilation procedure, the *a priori* profile $\boldsymbol{x_{ap}}$, and averaging kernels $\mathbf{A}$ from satellite retrievals (Section 2.2) are used for the calculation of observation operator $\mathbf{H}$ for satellite measurements as follows. This procedure aims to remove the dependence of the analysis on the model-retrieval comparison (Miyazaki *et al.*, 2012, 2020) following our previous work (Shu *et al.*, 2022).

$$\mathbf{H}\boldsymbol{x}^b = \boldsymbol{x}_{ap} + \mathbf{A}\left(\mathbf{S}\boldsymbol{x}^b - \boldsymbol{x}_{ap}\right)$$

We have added the definition of the state vector ($\boldsymbol{x}$) and use $\boldsymbol{y}$ for the observation vector in the revised Section 2.3 (lines 195–196) to differentiate state and observation vectors as follows:

"At each assimilation time step, we calculate the optimal estimate $\widehat{\boldsymbol{x}}^a$ of the true ozone concentrations ($\boldsymbol{x}$, state vector) as a weighted average of the model forecast $\boldsymbol{x}^b$ and the observation $\boldsymbol{y}$."

Line 211: It is reasonable to assume no correlation between surface station observations. But it is probably questionable to assume no correlation for satellite observations.

Response: Accepted.

Since the horizontal resolution of all the synthetic observations (satellite and surface observations) is much finer than that of the GEOS-Chem model (0.5°×0.625°), we adopt the super-observation approach to produce more representative data and reduce the horizontal observation error correlations (Miyazaki *et al.*, 2012; Barré *et al.*, 2015; Ma *et al.*, 2019). The approach is to average the observations (including errors and averaging kernels) across each 0.5° latitude × 0.625° longitude bin. In addition, it is also computationally cheaper to use super-observations in satellite data assimilation. We have previously discussed this in Section 2.3.

We have rephrased this part in the revised Section 2.3 (lines 213–218) to avoid ambiguity as follows:

"$\mathbf{R}$ is the observation error covariance matrix, including the contributions from the measurement error and the representativeness error. Since the horizontal resolution of all synthetic observations (GEMS, LEO satellite, and surface observations) is much finer than that of the model, we apply a super-observation approach to produce more representative data and reduce the horizontal observation error correlations (Miyazaki *et al.*, 2012; Ma *et al.*, 2019). A super-observation is generated by averaging all the observations (including errors and averaging kernels) within the same 0.5° latitude × 0.625° longitude GEOS-Chem model grid. Thus, $\mathbf{R}$ is assumed to be diagonal, that is, the observation errors are not correlated."

Figure 6d: What do the two different shades of color represent in the lower two panels?

Response: Corrected.

We have revised the caption of Fig. 6 (lines 729–736) to make it clear as follows:

"Figure 6. Comparison of the averaged diurnal cycle of (a) surface ozone, (b) mean bias, and (c) RMSE, as well as (d) the histogram (in percentage) of the temporal correlation coefficient of hourly ozone at validation grids (Fig. S4), as simulated by the *control run* (*a priori*, in blue) and three *assimilation runs* (Exp 1–3 in Table 1, in green, yellow, and red, respectively) relative to the *nature run* ("True", black line in panel a) for June 2020. In panel (d), the green, yellow, and red bars respectively represent the frequency (%) of the temporal correlation coefficient of simulated ozone between the three *assimilation runs* (Exp 1–3) and the *nature run*, which is in comparison to that of the *control run* (in blue). The blue bars are the same in the three sub-panels to better illustrate the improvements in the temporal correlation of simulated ozone relative to the *control run*. The mean values of the temporal correlation coefficient (colored the same as lines in panel a) at all validation grids are inset."

Reference:

Barré, J., Gaubert, B., Arellano, A. F. J., Worden, H. M., Edwards, D. P., Deeter, M. N., Anderson, J. L., Raeder, K., Collins, N., Tilmes, S., Francis, G., Clerbaux, C., Emmons, L. K., Pfister, G. G., Coheur, P.-F., and Hurtmans, D.: Assessing the impacts of assimilating IASI and MOPITT CO retrievals using CESM-CAM-chem and DART, J. Geophys. Res. Atmos., 120, 10,501–10,529, https://doi.org/10.1002/2015JD023467, 2015.

Brasseur, G. P., and Jacob, D. J.: Modeling of Atmospheric Chemistry, Cambridge University Press, Cambridge, 2017.

Ma, C., Wang, T., Mizzi, A. P., Anderson, J. L., Zhuang, B., Xie, M., and Wu, R.: Multiconstituent Data Assimilation With WRF-Chem/DART: Potential for Adjusting Anthropogenic Emissions and Improving Air Quality Forecasts Over Eastern China, J. Geophys. Res. Atmos., 124, 7393–7412, https://doi.org/10.1029/2019JD030421, 2019.

Miyazaki, K., Bowman, K. W., Yumimoto, K., Walker, T., and Sudo, K.: Evaluation of a multi-model, multi-constituent assimilation framework for tropospheric chemical reanalysis, Atmos. Chem. Phys., 20, 931–967, https://doi.org/10.5194/acp-20-931-2020, 2020.

Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., and Boersma, K. F.: Simultaneous assimilation of satellite NO2, O3, CO, and HNO3 data for the analysis of tropospheric chemical composition and emissions, Atmos. Chem. Phys., 12, 9545–9579, https://doi.org/10.5194/acp-12-9545-2012, 2012.

Rodgers, C.: Inverse Methods for Atmospheric Sounding: Theory and Practice, 2000.

Shu, L., Zhu, L., Bak, J., Zoogman, P., Han, H., Long, X., Bai, B., Liu, S., Wang, D., Sun, W., Pu, D., Chen, Y., Li, X., Sun, S., Li, J., Zuo, X., Yang, X., and Fu, T.-M.: Improved ozone simulation in East Asia via assimilating observations from the first geostationary air-quality monitoring satellite: Insights from an Observing System Simulation Experiment, Atmos. Environ., 274, 119003, https://doi.org/10.1016/j.atmosenv.2022.119003, 2022.