This study developed a global aerosol data assimilation system based on a global aerosol transport model and a 2-dimensional variational data assimilation method. Validations for the reanalysis data were conducted and suggested that the accuracy of the reanalysis data were much higher than the free run of model. I think this paper is valuable for providing an efficient way to obtain high quality reanalysis AOD data through degrading the assimilation system from 3-dimensional to 2-dimensional. I also like the detailed description for the aerosol transport model, the data assimilation system, the observation data, and the set-up of the reanalysis for the global aerosol data assimilation system. I recommend it publish as a technical paper after addressing the following comments.

We thank a reviewer for careful reading our manuscript and for giving useful comments. We have deliberately and considered your comments. We believe that we have made adequate corrections and answers to your comments. In revised manuscript, the changes are highlighted by yellow markers.

1. P8, L19-21. The sulfate chemistry in MASINGAR mk-2 includes seven gas-phase reactions and two aqueous-phase reactions. Recently, some studies suggested the CTM model generally underestimated the sulfate, which may be caused by missing some key heterogeneous chemistry reactions in the model. Could you provide some discussions for this issue related to your sulfate modeling in MASINGAR mk-2.

I agree with you. The missing of the heterogeneous chemistry reactions is one of the major source of the error (the low bias) in the reanalysis product. It is particularly true in the megacities. For instance, Zhang et al. (2015) evaluated the impacts of heterogeneous chemistry with regional CTM in eastern and central China (urban and industrialized area) and indicated the significant role of heterogeneous chemistry in regional haze (PM2.5) formation. This partly explains the negative biases at megacities in industrializing countries found in the comparison with the AERONET AOD (Section 3.3.2). We added the following texts in the Section 3.3.2 to discuss sources of the negative biases.

"Zhang et al. (2015) evaluated the impact of heterogeneous chemistry with regional chemical transport model in eastern and central China (urban and industrialized area of China), and suggested the significant role of heterogeneous chemistry in regional haze formation. While the current version of MASINGAR mk-2 includes the nine gas-phase and two aqueous-phase reactions of the sulfate

chemistry, the implementation of the heterogeneous chemistry reactions is under development. The missing of the heterogeneous chemical productions may partly explain the negative bias."

We misstated the number of the sulfate chemical reactions included in the current MASINGAR mk-2. We corrected.

"it includes nine gas-phase reactions and two aqueous-phase reactions."

Zheng, B., Zhang, Q., Zhang, Y., He, K. B., Wang, K., Zheng, G. J., Duan, F. K., Ma, Y. L. and Kimoto, T.: Heterogeneous chemistry: a mechanism missing in current models to explain secondary inorganic aerosol formation during the January 2013 haze episode in North China, Atmos. Chem. Phys., 15(4), 2031–2049, doi:10.5194/acp-15-2031-2015, 2015.

2. P12, L3. The horizontal error correlation length L is set to 200 km. This may be small for the coarse model resolution of TL159. Could you give more explanation for that? Is that setting related to the localization scale showed in P11, L20-24?

Yes. Our setting about both the horizontal error correlation length and the localization scale are based on results from Zhang et al. (2008, JGR). They plotted spatial correlation between MODIS observation and their model results as a function of distance (Fig. 7 of Zhang et al. (2008, JGR)), and found that the SOAR function (Eq. (25) in our manuscript) with L = 200 km fit the correlation and the correlation become less than 0.05 when the distance is more than 1000 km. We set the horizontal error correlation length and the localization scale to 200 and 1000 km, respectively, following their results. They used their model with 1° × 1° horizontal resolution to calculate the correlation. Our model resolution of TL159 (about 1.1° × 1.1°) is almost the same and the setting of the horizontal error correlation length is reasonable. We unified the explanations about the horizontal error correlation length and add additional texts as follows:

"Zhang et al. (2008) calculated the spatial correlation between satellite observations and model forecasts (1° × 1° horizontal resolution) as a function of distance and found that the SOAR with *L* set to 200 km can fit the correlation and when the distance was more than 1000 km, the spatial correlation decreased to less than 0.05. On the basis of their results, we set the localization scale and the horizontal error correlation length to 1000 and 200 km, respectively."

3. P12, L27-28. Please provide more information for the observation error covariance matrix because it is a key issue for the quality of the reanalysis data. Are the standard deviations of the observation errors uniform over the whole model domain?

No. We used quantitative uncertainty estimation for each data point provided by the NRL-UMD MODIS AOD product. The error estimation includes the representativeness error based on variability of the L2 dataset and the observation error estimated empirically. We added the more details of the observation error as follows:

"We assumed that the observation error covariance matrix (**R**) was diagonal and assigned uncertainty of AOD provided by the NRL-UMD MODIS AOD product to the diagonal component of the observation error covariance matrix. The uncertainty includes empirical estimation of observation error and representativeness error based on variability of the L2 dataset (Zhang and Reid, 2006)."

4. P13, L30-31. Meteorological nudging was performed for the AGCM. What variables in the AGCM were nudged? Is the nudging conducted at the same time step as the AOD data assimilation?

The variables used for the meteorological nudging are horizontal wind components (longitudinal and latitudinal) and air temperature. The nudging forcing is applied at each timestep of the integration by temporary interpolating the variables by linear interpolation. We added these descriptions of the meteorological nudging to the revised manuscript.

"The nudging scheme is applied to the horizontal wind components and air temperature. The nudging term is applied at each time step of the integration by temporary interpolating the variables by linear interpolation."

5. P14, L23-25. Sensitivity experiment was conducted through reducing the background error covariance and the chi-square value in the sensitivity experiment was shown. I am interesting in the

impact of the change of the background error covariance on the accuracy of the reanalysis AOD data. Could you provide some results for that?

Both RA (analyzed) and FG (first guess) from the sensitivity experiment in which the background error covariances were uniformly decreased by 60% shows worse agreement with the MODIS AOD than the main experiment. RMSD (correlation coefficient) for RA and FG during 2011–2012 was degraded from 0.05 (0.93) and 0.098 (0.73) to 0.07 (0.87) and 0.10 (0.71), respectively. The independent validation with the AERONET AOD also shows that the sensitivity experiment obtained worse scores comparing with the main experiment. For chi-square value, the sensitivity experiment shows larger variation compared with the main experiment (0.027 versus 0.0059). These results imply that although the background and observation errors were persistently overestimated, the balance between background and observation errors was well-balanced and stable.

"Both RA and FG from the additional experiment obtained worse scores in the validations with MODIS and AERONET AODs than the standard experiment. For the χ^2 value, the additional experiment shows much larger variation (standard deviation). These results imply that although there were the persistent overestimates of background and observation errors, they were well-balanced and stable in the standard experiment."

6. P19, L23-24. There is a statement for improving the 6-24h forecast. But I have not seen the experiment for the 24h forecasting.

We employed the MODIS observation as the assimilation data. MODIS is onboard Low Earth Orbit (LEO) satellites and make observation on the same region once a day. What we want to mean by "short-term (6–24 hour) forecasting" is a short-term model simulation after the analysis to the next analysis (when the next MODIS observation is available) in a certain region. We corrected the text to "the short-term forecasting until the next analysis (until the next MODIS AOD is available)".