This is a very comprehensive and well written paper on the JMAero. I don't have too much to add to the other reviews. The task at hand is straightforward: Provide an overview of the model components and provide verification. At least in regard to aerosol optical thickness they have done so admirably. I could follow what they were doing quite well, the figures were well done and they provide comprehensive error stats. One minor thing that needs mentioning, is that the verification really is toward AOT. When NRL composed its reanalysis paper (Lynch et al., 2016), we called it an AOT reanalysis even though it is a full 4 dimensional aerosol simulation because quite frankly AOT was all we could verify against. How these things pan out for surface concentrations is another kettle of fish, and perhaps the authors should note that. Indeed, not only is modeling surface concentrations (or concentrations at any level) difficult, but finding appropriate verification data is even more difficult. I don't hold the authors accountable to the honest facts, but they should mention it. Similarly, the JMAero does multi-bin particle size distributions (8 bins per specie), but they do not discuss at all how these bins interact, or if at the end of the day it buys them anything as al verification projects onto AOT metrics. This is ok of the paper is about AOT, but that should simply be stated up front in the abstract, introduction and conclusions. Other than this I have a series of equally minor comments that the authors might find helpful. Be well, Jeffrey S. Reid, US Naval Research Laboratory.

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.

As the reviewer pointed out, we validated the quality of the reanalysis with only AOD measurements. We added the text in the Abstract, Introduction, Conclusions, and Future directions as follows:

Abstract: "This paper describes the aerosol transport model, the data assimilation system, the observation data, and the set-up of the reanalysis and examines its quality with AOD observations."

Introduction: ", and validated its quality with AOD observations."

Conclusions: "A global aerosol data assimilation system was developed based on a global aerosol transport model, MASINGAR mk-2, and a 2-dimensional variational data assimilation method, and validated by AOD measurements."

Future directions: “In the present paper, the reanalysis was validated with only the AOD observations. Further validation is also needed. Comparison with vertical AOD distributions…”

1) Abstract, line 14; Please be clear on an r of 0.96 against the assimilated data. Also, this is just a personal preference, I tend to prefer r2 to r because r2 represents the fraction of variance explained. The authors are of course free to present how they wish, but too often r is used to put a happy face on things. If you want the best of both worlds, you can present bias and rms deviation separately

I stressed the results from comparison with observation that used for the assimilation as follows:

“Comparisons with MODIS AODs that used for the assimilation showed that the reanalysis showed much better agreement than the free run (without assimilation) of the aerosol model and improved under- and overestimation in the free run, thus confirming the sanity of the data assimilation system.”

Thank you for your suggestion. I understand your concern. However, in the present paper, we used r throughout the manuscript (main body, appendix, figures, and tables), and need a complete revision to unify into r2. In the next study, we carefully select the statistics to meet the aim of validation.

2) Please be clear when you refer to “size” if you mean radius or diameter. For example, (Page 6, Line 22) states size ranges from 0.2 to 20 um, but then says the dry radii are from 0.136 to 8.5 (line 3), which implies the original numbers were diameter. Traditionally aerosol science is in cgi units diameter, but it certainly is up to you. Please keep it consistent throughout the paper.

Thank you for your useful comment. We use “diameter” in the corrected manuscript.

“For mineral dust particles, the model uses a size bin method that logarithmically divides the particle size range from 0.2 to 20 μm into 10 size classes. The volumetric mean diameters of dry particles in each size bin are 0.271, 0.430, 0.681, 1.08, 1.71, 2.71, 4.30, 6.81, 10.8, and 17.1 μm.”

3) A little more discussion on where secondary OC comes from would help me understand the model better. As a by the way "mk-2 includes production from terpene" is stated on page 8, line 13. In the context of the paragraph in its isolation it is a bit of a non-sequitur. In just a couple of sentences can you please lay out how all primary and secondary POM production is handed with references? Also, no reference is given for the source function of primary POM or BC.

MASINGAR mk-2 does not treat secondary organic aerosol production explicitly but represents it implicitly by giving OC amounts produced from terpene. OC produced from terpene is assumed to be secondary organic aerosol and treated as hydrophilic. The emission of terpene is included in the emission inventory.

We added the following text in the revised manuscript.

"Although MASINGAR mk-2 does not calculate secondary organic aerosol production explicitly, this process is represented implicitly by giving OC production from terpene using emission data. Emission amount of terpene is provided by the emission inventory. OC produced from terpene is assumed to be secondary organic aerosol and treated as hydrophilic."

4) Can you please elaborate a little more on the paragraph starting Page 8, line 28 on how the coupling between AGCM and MASINGAR mk-2? For example what are the timescales of exchange? Are they run at exactly the same resolution? Is data assimilation between meteorology and aerosol particle handed at the same time or are aerosol particles handled after the fact?

Thank you for pointing out about the coupling. As the settings for the coupling depend on experiments, we added the detailed explanation about setting for coupling in the Section 2.4 (Experiment setup) rather than Section 2.1 (Overview of MASINGAR mk-2).

In present experiment, spatial resolution of the AGCM is the same to that of MASINGAR mk-2. Therefore, there is no spatial interpolation of the exchange variables. The timescale of exchange is the same to the time step of model (i.e., 900 sec); The AGCM and MASINGAR mk-2 exchange the variables at every time steps. At the aerosol assimilation time, the exchange is performed after the assimilation. This means that the AGCM receives the analyzed aerosol field from MASINGAR mk-2. We revised the Section 2.4 as follows:

"The AGCM has the same spatial resolution and time step to MASINGAR mk-2. The exchange of meteorological and aerosol variables between the AGCM and MASINGAR mk-2 through the coupler is performed every model time steps (i.e., every 900 seconds). At the assimilation step, the exchange follows the assimilation. This means that the AGCM receives the analyzed aerosol fields from MASINGAR mk-2."

5) On Section 2.2 (Data assimilation). Just a couple of things I am unclear about. First, how does JMAero handle the situation where AOT Obs say there should be a major event, and it is not at all in the model. This happens frequently due to a multitude of mesoscale forcing phenomenon or biomass burning. At NRL we use a climatology, and at GMAO they use the local displacement ensemble.

In the current version of JRAero, the background error was in proportion to the forecast AOD (i.e., first guess AOD; see Eq. (26)). This means that the background errors where the model did not predict aerosols became so small. Therefore, in that situation, the assimilation could not reproduce the major aerosol event because of the small background error.

Missing of aerosol sources (particularly for dust storms and biomass burnings) frequently causes the situation. Data assimilation both for initial condition and aerosol sources (i.e., emission inverse modeling; Yumimoto et al., 2008) will result in the better solution to the situation. Other effort should be application of an ensemble-based method with perturbed aerosol emissions.

We added discussion about this situation in Section 5 (Future direction) as follows:

"In the current version, the background error was in proportion to the forecast AOD (Eq. (26)), and became small where the model did not predict aerosols. Therefore, the analysis could not reproduce aerosol events that satellites observed but the model failed to predict (e.g., dust storms and biomass burning) due to the small background error. The ensemble-based estimate of the background error considering uncertainty in emissions will bring better analysis for this situation."

Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, a., Liu, Z. and Winker, D. M.: Adjoint inversion modeling of Asian dust emission using lidar observations, Atmos. Chem. Phys., 8(11), 2869–2884, doi:10.5194/acp-8-2869-2008, 2008.

6) Section 3.3.2, (Page 13 line 13). The papers describe the AERONET AOPs, but really it is just the AOT that are being used. AOPs implies the inversion products I think.

We corrected the manuscript.

7) Page 13, line 29. Listing of vertical levels is ambiguous, Are those the tops of the layers, or the layer thicknesses?

We corrected to the layer thickness (or box height) in the revised manuscript.

8) Page 15, line 5: To be fair, the AOT values are not that good either. It is hard to determine who is right when it comes to sea salt. . .

I agree with you. We need to check carefully where the errors come from. As you concerned, for instance, dust and anthropogenic aerosols could affect total AOD over Atlantic Ocean and northern Pacific Ocean. We corrected the text as follows:

"The distribution of the 5-year averaged increment (RA AOD minus FR AOD) (Fig. 6c) shows that, in general, assimilation increased AODs over the central Pacific Ocean, implying that in FR, MASINGAR mk-2 underestimated sea-salt aerosols."

9) Page 16, line 24 "70.0% of the deviations exceeded 0" I assume you then mean positive deviations? But then you said that overall the model is negatively biased. You might want to double check the language here.

In the figures, we defined the deviation as observed AOD minus modeled AOD. So, the positive deviation means that the model underestimates the observation.

10) Page 17, lien 21: Again, don't beat yourself up on Beijing. That and Kanpur have the worst performance in all global models (Sessions et al., 2105). This is a place where 2- and 3 d var is bound to fail. Need EnKF to make it work (http://onlinelibrary.wiley.com/doi/10.1002/2016JD026067/abstract). This I think is different from the overall low bias problem.

Thank you for your encouragement and suggestion.

As shown in Fig. 9, the model has the constant negative biases in the megacities mainly due to the insufficient anthropogenic emissions, the coarse model resolution and the missing of heterogeneous productions. We added more discussion about the negative biases caused by the model errors (also see reply to comment #1 of the reviewer #2).

"Multi-model inter-comparison studies (Kinne et al., 2006; Sessions et al., 2015) have pointed out that aerosol models having negative biases for high-AOD events is a common problem. Insufficient anthropogenic emissions data and model resolution for megacities are plausible reasons for 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."

The limited improvement by the assimilation is other problem. We have to take some steps about the low biases in the megacities both on the model and the assimilation. In the aspect of assimilation, as the reviewer suggested, the improvement of background error covariance is useful. We added the following texts that mention that the ensemble-based estimate of background error covariance has the possibility to overcome the problem referring the most recent results by Rubin et al. (2017).

Section 3.3.2: "The probability of a successful retrieval can be reduced during high-AOD events (Lynch et al., 2016); thus, fewer available satellite observations over megacities during high-AOD

events may also account for the negative biases in RA AODs. Rubin et al. (2017) applied an ensemble-based assimilation method to NAAPS and found that flow-dependent error covariances estimated by ensemble simulations utilized the AERONET AOD efficiently and brought better analyses at Beijing and Kanpur. Sophistication of the background error covariance and assimilation of additional observations have the potential to improve the analyses at the megacities."

Future directions: "At megacity and mountain sites, assimilation provided limited improvement, and positive and negative biases remained in the reanalysis. A plausible reason is the coarse model resolution, which is insufficient to resolve high-AOD events around megacities and local terrain effects in mountainous areas. Therefore, we plan to rerun the reanalysis with a finer resolution and check the performance of the model. Re-examination of the background error by an ensemble-based method (Yumimoto et al., 2016) and assimilation of additional observations (e.g., the AERONET AOD) have the potential to improve the analyses at the megacity sites (Rubin et al., 2017)."

Rubin, J. I., Reid, J. S., Hansen, J. A., Anderson, J. L., Holben, B. N., Lynch, P., Westphal, D. L. and Zhang, J.: Assimilation of AERONET and MODIS AOT observations using Variational and Ensemble Data Assimilation Methods and Its Impact on Aerosol Forecasting Skill, J. Geophys. Res. Atmos., doi:10.1002/2016JD026067, 2017.

11) Page 18, Line 33: This is because by nature the highest AOT events also have a lot of spatial variability and consequently it gets filtered out in the QA process.

I agree with your point. I added the point in the revised manuscript.
"The NRL-UND MODIS peak AOD values were also lower than the AERONET peak AODs, because larger spatial variability included in the dense aerosol events filtered the high AOD values out through the QA process."