Referee #1

Air pollution, especially PM2.5, in India is a hot topic. However, studies are hampered by limited availability of data in India. As correctly pointed out in the Introduction that satellite or ground-based measurements have limitations in temporal continuous or spatial coverage. This paper analyses long-term reanalysis MERRA-2 datasets of PM2.5, tries to overcome the limitations and perform analysis of the space-time variability of surface PM2.5 over India during 2000-2017. However, as pointed out in the prereview by the two referees, the datasets (and quality control) used in this study are not detailed introduced. In this ACPD version, I see little improvements of the description/discussion of the datasets, quality control and validations of the used model results, and representativeness of the observations. I feel less confident about the drawed conclusion, if the authors do not understand that how are the datasets produced, how is the quality controlled and what are the limitations and representative-ness of the data used. I would recommend this manuscript for publication in ACP only if the following concerns can be nicely addressed.

We are grateful to the reviewer for providing the insightful comments on our manuscript. We have addressed all the comments and suggestions provided by the reviewer. We discussed the bias correction methods and the MERRA-2 datasets in great details with relevant references. Potential sources of uncertainty and probable limitations are also discussed clearly. Our point-by-point responses for the all the comments are mentioned below.

Comment Response = Red colour
Information in revised MS = blue colour

Comment:1
Satellite observed AOD550 with MISR, MODIS and AVHRR, plus ground-based AOD data from AERONET were assimilated in the MERRA2 reanalysis dataset. However, all these datasets are column parameters. I understand the assimilation of these datasets can improve the radiative forcing simulation directly. But, how does the assimilation of column parameters improve the simulation of surface PM2.5 concentration in MERRA2?

Response: We would like to thank the reviewer for raising this issue. We have introduced the following information in the revised MS section 2.1 [Line 141-162] as given below:

Both the aerosol and meteorology observations jointly assimilated in the GEOS-5 model. The AOD assimilation process includes the cloud screening and bias correction of space-based instruments (e.g. MODIS and AVHRR) with ground-based instrument AERONET. The GEOS-5 model also assimilates the non-biased corrected AOD from MISR (over the bright surface). GEOS-5 assimilates the real-time (for every 3 hours) quality controlled AOD observation by using the local displacement ensemble (LDE) methodology (Buchard et al., 2015, 2017). LDE methodology is introduced to correct the misplaced aerosol plumes by considering the various aerosol properties such as speciation and vertically distribution of aerosols. On the other hand, the regions where LDE is not applied, AOD increment factor has been implemented, which represents a vertical scaling factor, more detail can be found in Buchard et al. 2017.

In addition to this, Buchard et al. (2017) also performed some experiments with assimilation and without assimilation of AOD product over the global region. The results showed that aerosols with the assimilation system provides better estimates of surface $PM_{2.5}$ as they are well correlated with satellite-derived $PM_{2.5}$ from van Donkelaar et al. (2010). However, some discrepancies can be observed due to the lack of the nitrate in the GOCART module and low emission of OC (organic carbon) in MERRA-2 dataset. The carbonaceous MERRA-2 $PM_{2.5}$ are mainly from biomass burning, anthropogenic sources and plant matter (Buchard et al., 2017; Randles et al., 2017). Here, we assume that the contribution of OC concentration from biogenic sources is very less because the aerosol emissions are mainly from biomass burning and anthropogenic sources in Northern India (Gani et al., 2019).

We understand that there could be some uncertainty during the simulation process of $PM_{2.5}$ in MERRA-2 which is also true for any other exposure model, but since this provides hourly dataset, we believe that they can be utilized after applying bias correction. We have included detailed discussion of the dataset and the bias correction technique in the revised MS.

References:

Buchard, V., Da Silva, A.M., Colarco, P.R., Darmenov, A., Randles, C.A., Govindaraju, R., Torres, O., Campbell, J. and Spurr, R., 2015. Using the OMI aerosol index and absorption aerosol optical depth to evaluate the NASA MERRA Aerosol Reanalysis. Atmospheric Chemistry and Physics, 15(10), p.5743.

Buchard, V., da Silva, A.M., Randles, C.A., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J. and Winker, D., 2016. Evaluation of the surface PM2. 5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States. Atmospheric environment, 125, pp.100-111.

van Donkelaar, A., Martin, R.V., Brauer, M., Kahn, R., Levy, R., Verduzco, C. and Villeneuve, P.J., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. Environmental health perspectives, 118(6), pp.847-855.

Gani, S., Bhandari, S., Seraj, S., Wang, D.S., Patel, K., Soni, P., Arub, Z., Habib, G., Hildebrandt Ruiz, L. and Apte, J.S., 2019. Submicron aerosol composition in the world's most polluted megacity: the Delhi Aerosol Supersite study. Atmospheric Chemistry and Physics, 19(10), pp.6843-6859.

Comment:2

Following the above comment, I think the impact of AOD assimilation on surfacePM2.5 concentration would be strongly depended on PBL simulation, which is also a key topic analysed in this study. However, how is the PBL simulated and how is the top of PBL defined in the MERRA2; what kinds of meteorological datasets are assimilated in the model to improve the PBL simulation over India; and how good is the performance of the PBL in MERRA2 compared with observations or improved by assimilation,...etc? All kinds of these questions are not discussed in the paper.

Response: We would like to thank the reviewer for this suggestion. We have incorporated the information in the revised MS [Line 247-269] as given below:

According to Rienecker et al. (2011), two types of schemes are introduced in the GEOS-5 model for simulating the atmospheric boundary layer in the MERRA-2 reanalysis dataset. The first scheme (Louis et al., 1982) is based on the planetary stable condition in which no planetary boundary layer of clouds involved. The second scheme (Lock et al., 2000) is based on the unstable or cloud-topped planetary boundary layer condition involved. Additionally, the GEOS-5 model uses two more schemes based on orographic conditions such as orographic gravity wave drag (McFarlane 1987) and non-orographic waves (Garcia and Boville 1994).

MERRA PBL heights are also diagnosed by the turbulence parameterization in the atmospheric general circulation model, based on the eddy diffusivity coefficient for heat. The PBL height is diagnosed as the lowest level at which the diffusion coefficients drop below a value of 1 m^2/s. There are various meteorological variables (temperature, wind, and humidity) also introduced in the GEOS-5 model to assimilate the PBL heights. Regarding the performance of the GEOS-5 model for PBL height, Jordan et al. (2010) observed a good (r = 0.7) correlation of PBL height between CALIOP satellite and GEOS-5 model over the western hemisphere (e.g. American and African region). However, some disagreement was also observed over the equatorial Pacific Ocean, where GEOS-5 PBL height is lower than the CALIOP PBL height. In addition to PBL height, the other physical processes such as aerosol mixing and hygroscopic growth etc. could also contribute to the uncertainty in the aerosol model (Randles et al., 2017, Schutgen et al. 2010). With the calibration, we were able to reduce this low bias substantially.

References:

Jordan, N.S., Hoff, R.M. and Bacmeister, J.T., 2010. Validation of Goddard Earth Observing System-version 5 MERRA planetary boundary layer heights using CALIPSO. Journal of Geophysical Research: Atmospheres, 115(D24).

McFarlane, N.A., 1987. The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. Journal of the atmospheric sciences, 44(14), pp.1775-1800.

Garcia, R.R. and Boville, B.A., 1994. "Downward control" of the mean meridional circulation and temperature distribution of the polar winter stratosphere. Journal of the atmospheric sciences, 51(15), pp.2238-2245.

Randles, C. A., Da Silva, A. M., Buchard, V., Darmenov, A., Colarco, P. R., Aquila, V., Bian, H., Nowottnick, E. P., Pan, X., Smirnov, A., Yu, H., and Govindaraju, R.: The MERRA-2 Aerosol Assimilation. Technical Report Series on Global Modeling and Data Assimilation 44, NASA Global Modeling and Assimilation Office. [Available online at https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/docs/.], 2016.

Schutgens, N.A.J., Miyoshi, T., Takemura, T. and Nakajima, T., 2010. Sensitivity tests for an ensemble Kalman filter for aerosol assimilation. Atmospheric Chemistry & Physics Discussions, 10(3).

Comment:3

I agree with the second referee that most of the CPCB monitoring are in the urban area: the representativeness of CPCB dataset needs to be carefully discussed. Due to CPCB observations possibly represent the urban condition, how suitable they are for direct comparisons with the MERRA2 global model results with relatively coarse resolution? Based on this thinking, I am kind of agree with the second referee that "the bias correction/calibration methodology is overfitting the model data". This question was not discussed in the ACPD version, I feel we need to think about it more carefully in the next revised version

Response: We appreciate the reviewer for pointing out this issue. We have calculated the MERRA-2 $PM_{2.5}$ by using some scaling factor (Line 173) and further modified the previous bias correction method in the revised MS. This has addressed the overfitting issue pointed out by both reviewers. We personally went through the CPCB data for quality check. The detailed information of the bias correction can be found in the revised MS [Line 198-222] that is given as:

2.3 Calibration of MERRA-2 $PM_{2.5}$ with CPCB

We calibrated hourly MERRA-2 $PM_{2.5}$ with coincident $PM_{2.5}$ data from 75 CPCB sites across the country for the period 2009-2017, as CPCB data are available only for these periods. The uncalibrated MERRA-2 $PM_{2.5}$ shows a correlation of 0.57 (significant at 95% CI) with coincident in-situ $PM_{2.5}$ (left panel of Figure 1). For bias correction, we used the percentile based bias correction methodology. We divided the MERRA-2 data at 10 percent interval and then calculated the relationship (r = 0.9) between median bias at every 10 percentile ranges between the two datasets (central panel of Figure 1). Then we adjusted MERRA-2 data with the calibration factors of the respective percentile ranges. Bias-corrected MERRA-2 at every grid (Figure S1) and median $PM_{2.5}$ at every 50 $\mu g\ m^{-3}$ interval (right panel of Figure 1) show improved correlation with the in-situ data. We note that MERRA-2 $PM_{2.5}$ is still underestimated at very high concentration (i.e. >300 $\mu g\ m^{-3}$); but since most of the country does not have any ground-based monitoring, we proceeded with our analysis with the calibrated MERRA-2 $PM_{2.5}$ to examine the diurnal pattern over India.

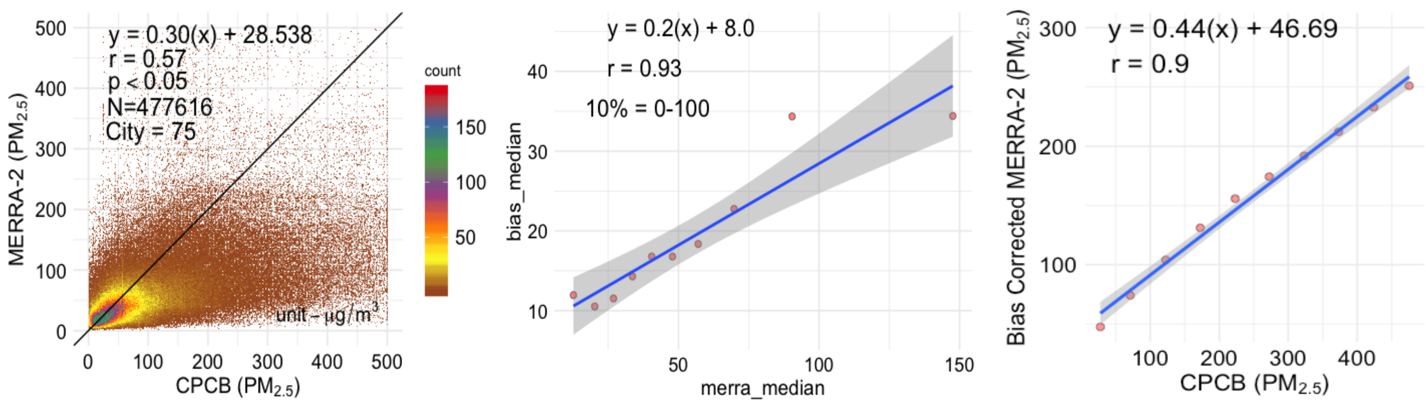

Figure 1. Scatter plot between (left) uncalibrated MERRA-2 and in-situ $PM_{2.5}$ data, (middle) median bias in MERRA-2 and in-situ $PM_{2.5}$ at every 10 percentile ranges, and (right) calibrated

median $PM_{2.5}$ from MERRA-2 and in-situ at every 50 μg m^{-3} interval. Spatial and temporal matching is done by averaging data from all ground-based monitoring sites falling within a single MERRA-2 grid for 1-hr duration.

Additionally, we would like to report that, modeling oriented studies either underestimate or overestimate the values with respect to the ground-based or space-based observational values. However, it is also true that models are also capable of capturing the more or less same spatial/temporal pattern with the observational data (in-situ and remote sensing). Therefore, in our study, we would like to point out that the MERRA-2 dataset has also shown the more or less same PM2.5 diurnal with in-situ CPCB over the Indian region.

Comment:4
Furthermore, based on my limited experience with CPCB dataset, it seems the quality of $PM_{2.5}$ concentration from CPCB is questionable. How is the CPCB dataset quality controlled, this question was raised by the second referee, however, still not addressed. There are some Indian cities have $PM_{2.5}$ observations from US diplomatic missions, which are generally believed to be of high quality and could help with quality validation of CPCB observations over these cities

Response:
It is difficult to objectify accuracy of the ground-based sensors based on perception. There is no study documenting whether the embassy monitors are more accurate than CPCB. Moreover, embassy monitoring stations (assumed to have high quality $PM_{2.5}$ data) operate the instrument at the embassy ground only which are located in urban centers (Pant et al., 2018). Therefore, they don't provide heterogenous environment. The other monitoring networks such as SAFAR and MAPAN are not available in the public domain.

We took caution in handling CPCB data. We personally went through the entire raw data, and eliminate data from the period where it showed spurious and wild fluctuations. We understand the importance of data quality by the regulatory agencies, but we cannot do more than this. WHO and exposure data used in GBD (Shaddick et al., 2018) also had to calibrate with CPCB network.

Reference:
Pant, P., Lal, R.M., Guttikunda, S.K., Russell, A.G., Nagpure, A.S., Ramaswami, A. and Peltier, R.E., 2019. Monitoring particulate matter in India: recent trends and future outlook. Air Quality, Atmosphere & Health, 12(1), pp.45-58.

Comment:5
MERRA2 dataset is simulated with EDGARv4.2 global emission inventory, as described by the paper. I suppose the EDGAR inventory for year 2012 (the latest one) was used. How well the 2012 inventory represent the condition of the period 2000-2017? As reported by lots of studies, between 2012 and 2017 the Indian emissions have changed a lot. And as described by the MERRA2 aerosol dataset developer (Buchard et al., 2017) that assimilation cannot correct for deficiency due to missing emissions. The uncertainty of emission inventory would propagate to the MERRA2 reanalysis data. And biomass/agriculture burning is believed to be a large contributor of surface PM2.5 over India. How is this burning source considered in the MERRA2 simulation.

Response:
MERRA-2 dataset includes different emission inventories (Table 1), in-situ and satellite datasets. The years in which emission inventories (EDGAR) are not available, different satellites (MODIS-Terra and Aqua, MISR) and in-situ data (AERONET) are utilized to simulate the aerosols in the MERRA-2 datasets along with fire emission inventories (QFED and GFED). We agree that there are obvious uncertainty in the simulations as emission is not periodically updated, but this is true for any modelling exercise. Globally, no emission inventory is updated annually. We believe that since MERRA-2 assimilates satellite and ground-based data, it should capture the temporal heterogeneity/variation. Further, our bias correction has brought the data closer to the observations.

Table 1. According to Randles et al. (2017), the aerosol species and their sources and inventories used are given as

Aerosol Type	Sources
Dust	Wind driven
Sea Salt	Wind Driven
Volcanic SO2	AeroCom Phase II
Biomass Burning	Scaled RETROv2
SO2, SO4, POM and BC	Scaled GFEDv3.1 QFED 2.4-r6
Anthropogenic SO2	EDGARv4.2 (Energy + non Energy)
Anthropogenic SO4, BC, and POM	AeroCom Phase II

We also thank the reviewer for asking about the burning source in the MERRA2 simulation. We have incorporated the information in the revised MS [Line 129-138] is given as:

Randles et al., (2017) and Buchard et al. (2017) also mentioned that Quick Fire Emissions Dataset (QFED) emission inventory (based on the MODIS fire radiative power) has been used from 2010 onward, but from the period of 1997-2009, the Global Fire Emission Dataset (GFEDv3.1) emission inventory has been introduced to simulate the aerosol emission from biomass burning in MERRA-2 reanalysis dataset (Darmenov and da Silva 2015; Randerson et al., 2006; van der Werf et al., 2006). Some scaling factor was also introduced in the QFED to minimize the uncertainty among the biomass burning emission inventories (Buchard et al., 2017). So, use of two different inventories for different time period could also be responsible for the uncertainty in fire-based MERRA-2 $PM_{2.5}$.

References:

Buchard, V., da Silva, A.M., Randles, C.A., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J. and Winker, D., 2016. Evaluation of the surface PM2. 5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States. Atmospheric environment, 125, pp.100-111.

Darmenov, A. and da Silva, A., 2013. The quick fire emissions dataset (QFED)–documentation of versions 2.1, 2.2 and 2.4. NASA Technical Report Series on Global Modeling and Data Assimilation, NASA TM-2013-104606, 32, p.183.

Randles, C.A., Da Silva, A.M., Buchard, V., Colarco, P.R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J. and Shinozuka, Y., 2017. The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. Journal of Climate, 30(17), pp.6823-6850.

Randerson JT, Liu H, Flanner MG, Chambers SD, Jin Y, Hess PG, Pfister G, Mack MC, Treseder KK, Welp LR, Chapin FS. The impact of boreal forest fire on climate warming. science. 2006 Nov 17;314(5802):1130-2.

Van Der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G.J., Kasibhatla, P.S. and Arellano Jr, A.F., 2006. Interannual variability of global biomass burning emissions from 1997 to 2004.

Commnet:6

As described in the paper that OC are secondary aerosols in MERRA2 dataset. I would like to know how is the secondary organic aerosols simulated or represented in the GEOS5/MERRA2 model. Since, OC contributed about half of fine particles mass in Delhi (possibly other IGB regions as well) based on recent observations (Gani et al., 2019). The correct simulation of OC secondary formation processes would be critical for the accuracy of MERRA2 aerosol dataset. Some comments on the validation of OC simulation within MERRA2 would be helpful.

Response:6

We would like to thank the reviewer for pointing out this comment. We have modified the revised MS [Line 158-162] and incorporated as:

In the MERRA-2 dataset, the carbonaceous $PM_{2.5}$ are from biomass burning, anthropogenic sources and plant matter (Buchard et al., 2017; Randles et al., 2017). Here, we are assuming that the contribution of OC concentration from biogenic sources is very less because the aerosol emissions are mainly from biomass burning and anthropogenic sources in Northern India (Gani et al., 2019).

Regarding the validation of OC emissions from MERRA-2, Bali et al., 2017 compared the OC emissions from MERRA-2 with ECMWF-GFAS (Global Fire Assimilation System) datasets. The study observed the same spatial distribution of OC emission in both the reanalysis datasets but with the high estimation of OC values in MERRA-2.

References:

Buchard, V., da Silva, A.M., Randles, C.A., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J. and Winker, D., 2016. Evaluation of the surface PM2. 5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States. Atmospheric environment, 125, pp.100-111.

Randles, C.A., Da Silva, A.M., Buchard, V., Colarco, P.R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J. and Shinozuka, Y., 2017. The MERRA-2 aerosol *reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. Journal of Climate, 30(17), pp.6823-6850.*

Gani, S., Bhandari, S., Seraj, S., Wang, D.S., Patel, K., Soni, P., Arub, Z., Habib, G., Hildebrandt Ruiz, L. and Apte, J.S., 2019. Submicron aerosol composition in the world's most polluted megacity: the Delhi Aerosol Supersite study. Atmospheric Chemistry and Physics, 19(10), pp.6843-6859.

Comment:7

Some technical correction: a) line 128. 'three different sensors' should be four sensors in total if count AERONET monitoring as well. b) line 108. I feel the word 'propose' might be inappropriate. This assimilation/reanalysis approach has been widely use dover other regions, the contribution of this study is used a reanalysis dataset to analyse the spatial-temporal variation of surface PM2.5 over India.

Response:7

a) has changed in the revised MS as per referee suggestion.
b) has changed in the revised MS as per referee suggestion.