Reply to Referee #1:

(Original comments in italics and reply in indented paragraphs. Note that all the figure numbers in our reply are referred to as in the ACPD version before revision.)

This paper brings together an interesting set of long-term observations and global aerosol model simulations, to show a general consistency of the model results with these observations. The analysis presented is exhaustive and robust. The paper reads much like a model validation paper – focusing on trends. This work without doubt merits to be published in ACP, but the authors could try to condense the technical discussions somewhat, and pay more attention to improving the abstract, discussion and conclusions to bring out the key findings better and make the paper more accessible.

Specifically, I would recommend to emphasize more:

1) What are the new outcomes and progress? In my perception this work seems of higher quality and the model performing better than other studies I have seen- but the paper needs to demonstrate this better by referring back to older papers.

> In the Introduction, we mentioned more than a dozen previous studies on regional or global aerosol trends. We have added the following sentence to emphasize the goal and uniqueness of this paper:
>
> "Expanding from previous, abovementioned aerosol trend studies, this study provides an comprehensive synthesis and assessment of aerosol variations over the last three decades (1980-2009) in different regions of the globe through a global model analysis of multiple-platform datasets, with a goal of determining the anthropogenic and natural contributions to the multi-decadal changes."

2) Currently the paper is very lengthy- which makes it sometimes difficult to understand what point the authors want to make. It would be very good if the authors could summarize at the end of each section what are the overall findings and robustness. I also suggest to construct a summary table where the main findings regarding regional trends are presented, including a quantitative and qualitative discussion.

> We have condensed the paper to (1) limit the regional trend discussions to include 4 pollution regions, 4 dust regions, 2 biomass burning regions, and 6 oceanic regions for their higher relevancy to our purpose, (2) reduce the AERONET section and only show the satellite and model evaluation with AERONET data, and (3) consolidate the model-data comparisons and trend analysis of satellite and surface measurements into the same sections (i.e., over land or over ocean), and (4) shorten the model description to include only necessary information to assist the later discussions. More complete figures and model descriptions are now moved into the Supplemental Material.
>
> We have also added a "synopsis" section at the end of section 3 to summarize the overall findings. We did not construct a summary table, as these points have been made in the synopsis as well as in the conclusion section.

3) Where are model results in agreement with the multiple sets of observations (with uncertainty margins) and what are the aspects that we do not understand well.

We now include a “satellite data diversity” panel and a “model minus satellite mean” panel in Figure 4 to show the locations where satellite data agree or disagree with each other and where the discrepancies between model and satellite occur. We also add the corresponding discussions on these features in the text.

4) If possible highlight where this specific model stands in a the crowd of other global model results- or with other words- what aspects can be expected to be model specific- and what aspects can be expected to more generally reflect our knowledge on changes in aerosol.

The intercomparisons of multiple models, including GOCART, have been presented in many publications (e.g., Textor et al., 2006, Kinne et al., 2006, Schulz et al., 2006, Koch et al., 2009, Huneeus et al., 2011, Koffi et al., 2012, etc.), and the lead author of the present paper (Mian Chin) is also working with the AeroCom modelers to compare model results from the AeroCom hindcast experiments on the 30-year trends. Therefore, in this manuscript we prefer to focus on the GOCART model results, meanwhile we will mention in the conclusion that some findings in this study could be model-dependent.

Detailed comments:

p. 19752: I find the title not very adequately covering the contents of the paper: there is relatively little on sources; it doesn’t tell that this is a study covering the globe; it is about surface and column aerosol. Suggest: A global model analysis of regional surface and column aerosol trend variability from 1980-2009.

Thanks for the suggestion to make the title more appropriate for the contents. We have changed the title to: “A global model analysis of multi-decadal regional aerosol variability from 1980 to 2009”, to emphasize the multi-decadal and regional analysis nature. We do not feel necessary to add more specifics in the title, such as “column” or “surface”.

p. 19753 l.4 global satellite observations and ground based networks mainly covering Europe and Network, with some information elsewhere.

Currently the sentence is “observations from multiple satellite sensors and ground-based networks”. Not clear if the referee recommends changes of the sentence?

p. 19753 l.7 Consistent quantitative or qualitative? Would be good to give numbers.

We have modified the sentence as “…is generally in the same direction as the anthropogenic emission changes but with smaller degree, because the change of natural aerosols, especially dust, over these regions can dampen the magnitude of anthropogenic changes.”

p. 19753 l. 10-15 Does the model accurately describe the dust fluctuations?

Yes, in section 5.2 (discussion).

p. 19753 l. 15 tropical North Atlantic: do you mean to say here that the dust is more effectively removed? Can you say something on the relatively importance of the two processes?

The negative correlation of dust with precip in North Atlantic suggests that lower dust is likely, as least partly, because of more precipitation that removes dust, and the positive correlation between dust over North Atlantic and Sahel dust emission implies that the dust transported to N Atlantic is reduced partly because of the lower Sahel emission. Separate these two effects will require model sensitivity and/or tagged region runs, which are not available yet. Therefore this is not a definitive results but are a part of discussion to suggest the importance based on the tightness of the correlation.

p. 19753 l. 19: Is it meant global-land average and global-marine average? Or just global average?

We meant global land, global ocean, and entire globe averages. We have changed the phrase to "global-scale average values" to be more consistent with the context in previous sentences that mentioned global land and ocean.

p. 19753 l. 20 In general regional assessments are useful- especially for aerosol with lifetimes of days-to one week. Global aerosol trends are mostly relevant for climate issues ('global dimming','global brightening'). I think the authors want to say something like that opposing regional trends make any estimate of global trends over the last 30 years difficult.

Now the sentence reads "…as the global-scale average values conceal the regional changes that are in opposite directions, and thus are not sufficient for assessing aerosol changes and their impacts in the past 30 years."

p. 19754-55 The introduction already contains quite some details on measured trends- I am wondering if it is not more appropriate to include details in the measurement section, and try to sketch the general picture first.

The introduction section is intended to provide some background information in addition to introducing the present study. The measurement section is mostly technical. We have shortened the introduction section but the general picture of what have been observed remains in the introduction section, if the referee agrees with this approach.

p. 19758 l. 3 Size? aerodynamic diameter?

Changed to "radius".

p. 19758 Do the satellite retrievals assume a similar non-spherical treatment of dust. With other words are the model and satellite (and Aeronet) results more comparable?

MISR and AERONET use non-spherical properties for dust in their retrieval – AERONET assumes spheroids whereas MISR has several non-spherical-shaped models used in their retrieval. MODIS assumes non-spherical dust shape in the land retrieval but spherical in the ocean retrieval. There is no one single universally used dust shape in remote sensing retrievals. Our non-spherical dust optical calculation code is provided by Oleg Dubovik, who developed the AERONET non-spherical dust retrieval method.

p. 19759 The authors are preparing us for some issues regarding the biomass burning emissions. Likewise can they say something about this extrapolation of 2000 emissions with scenario calculations? What are the inaccuracies?

The A2-ACCMIP anthropogenic emissions were chosen because it is a commonly used datasets in the global modeling communities and also because of its extended time coverage. We recognize the drawbacks of linear interpolation in the 5- or 10-year intervals. We have added the following sentences in the emission section to provide justification:

"We choose to use this emission dataset because of its broad acceptance in modeling communities and extended time coverage." "Although such linear interpolation does not necessarily resolve the interannual variations between individual years, it should capture the multi-decadal anthropogenic emission trends."

p. 19759 On my printer equation 1 didn't show well. Please check if this is general problem.

The equation does not seem to be a general problem as all the coauthors could see it. In any case, it is now moved to the supplemental material (SM) as one referee suggested.

p. 19760 The description of dust scheme is interesting but detailed. Move to appendix? I think the main issue is that the dynamic source model has improved by including surface wetness. Do other models include this as well?

The details are now moved to the SM. Most models do not include dynamic dust source, to our knowledge.

p. 19760 emission variability how does it compare to other published studies? For instance the Pozzoli et al. (ACP. 2011); seem to have a much larger variability of global mineral dust; with an approximately 10 % relative standard deviation. Further emissions were much lower than the ones in this study. Can the authors discuss why they think the emissions here are better. As a technical remark I think it is difficult to see the regional variability of the emissions- it would be great to see similar regional plots – or perhaps just make the data available in a spreadsheet for future modeling activities by other groups? Please pay attention that 2b in the ACP version will be readable (I had to enlarge it a lot in the on-line ACPD version).

Regarding the dust emission, it has huge temporal variability on time scales of sub-daily and daily, and even monthly, but much weaker interannually. Our 1980-2009 mean dust emission is 3234 Tg/yr, which is 4.3x higher than the 1980-2005 mean shown in Pozzoli et al. (ACP 2011), and our annual emission standard deviation is163.4 Tg/yr, which is 2.3x higher than Pozzoli et al. (ACP 2011). However, the ECHEM5 model used by Pozzoli et al. only includes dust sizes up to 0.5 micron in radius while GOCART considers a much wider range of size, up to 10 micron in radius. The difference in particle size range can easily explain the differences between GOCART and ECHEM5 calculated dust emission mass amount and standard deviation. A recent study on AeroCom multi-model dust intercomparison has shown that the dust over North Africa and North Atlantic from ECHAM5 is much too low compared with satellite data of dust AOD, while GOCART shows some overestimation over land (Kim et al., submitted to JGR, 2013). Note that the mean dust emissions from ECHAM5 and GOCART are still within the range of AeroCom A (phase I) multi-model range, which is 672-4040 Tg/yr (Textor et al., 2006).

We have also added the mean and standard deviation of 30-year emissions on each panel in Figure 2a to provide a context of the interannual variability of these emissions. We have also added text regarding our emissions and the AeroCom A emission range.

Yes, we will make the regional emissions shown in the first 4 panels in Fig. 2a available in SM.

p.19761: Something needs to be said about the fact that the use of of-line oxidants doesn't assume the separation of natural / current (standard) conditions, thus avoiding one source of non-linearity. But what about non-linearities in aerosol dynamics? How accurate is the attribution of the difference of the full simulation and the natural simulation to 'fossil/biomass burning'. Although I understand that the authors are not keen on performing a 3rd set of simulations (fossil/biomass) alone, I would recommend that verification of this assumption with a couple of years of sensitivity studies is needed.

Non-linearity in aerosol dynamics is not an issue in the model experiments since our off-line aerosol simulations do not interact with meteorology and do not simulate aerosol microphysical processes. We had performed simulations with FF+BB alone in the past (from an earlier version of the model) and compared to the results from the standard minus natural runs, and we found difference was rather small (within a few percent). It is widely accepted that it is more accurate to estimate the component contribution by the difference between the simulation that included that component and the one that exclude it (e.g., HTAP SR simulations).

p. 19762 Please explain better the difference between the two AVHRR datasets. In this section a discussion on the accuracy of the datasets is needed. The authors choose not to use ATSR-2 (Thomas et al) any reason?

We will modify Table 2 to include the accuracy of each satellite sensors. Currently we do not have the ATSR-2 data on hand (as we mostly concentrated on the US products that are available to the whole world), but we will try to get the ATSR-2 data to include in the revision. Our preliminary attempt does not seem to lead us to a well documented location to download the data that were processed with a consistent algorithm. If the referee has more information we would appreciate it in order to obtain the long term data.

p. 19764 'few' sites. How many and which ones? Figure 3 seems to suggest that there are many sites- but perhaps this overview of all AERONET sites, not of which ones were used?

We have revised the text saying "less than 10 sites". We also highlight the 8 long-term data sites on Figure 3. We have revise the manuscript showing only the comparisons with these 8 sites for time series, although we use all data to evaluate the model and satellite (text revised accordingly).

p. 19765 Section 4.1 gives trust in the general performance of the model (explain why 2001 was chosen?). This collection is truly impressive (make sure that it remains readable in final version). I think it is useful to have the global/ocean/land averages in the panels. It is somewhat hard to read from Figure 5a/5b the corresponding accuracy of regional averages- the plots are quite busy. While I understand there are some issues with the TOMS (over land); but what are skills of the model to re-produce variability over the period 2001-2010 when MOIDS/MISR/Seawifs data were available. And how to interpret the data? Is it fair to say that were various satellite datasets correspond (e.g. region 9; 13); the model is not performing very well, but what to say about SAS and EAS; where the models seem to correspond better with MISR/SEAWIFS than with MODIS.

> We have added "2001 was chosen because of all the above satellite data are available in that year (TOMS ended after 2001)." Yes we realize the figure will be more readable in ACP full-page layout. We have also limited our comparisons to 10 land and 6 ocean regions and made the plots in Fig. 5 larger for easier reading, and added statistics showing the agreement between model and satellite data.

p. 19768/figure 5b the correspondence at northern hemispheric ocean regions is remarkable, as is the lack of correspondence in the SH. Somewhere the authors declare that the contribution of seasalt is not large, but the problems in figure 5b region 9- 12 seem to suggest something else. Is the seasalt source function adequate, is the variability of the assimilated winds in the SH (driving Seasalt) enough? Or should the authors look more into DMS, which is a function of wind but also plankton variability. Please comment.

> We have added two panels in Fig. 4, one is satellite "diversity" and the other is the model "outbound", i.e., model is outside of the satellite range. They are added also to address another referee's comments. We have also discussed the bias from both satellite and GOCART compared to the MAN sunphotometer data (Smirnov et al., AMT 2011) over the oceans. For SH tropical regions, we don't think adding more sea salt could be the solution, since our comparisons with the Miami data show that the model is higher, not lower, in the tropical remote regions. It is also difficult to think larger DMS emission because the tropical open ocean area is a biologically-barren region to have low DMS in the sea water, winds are calm to not promote DMS emission from water to air, and our sulfate (mostly from DMS in remote ocean) in the remote ocean is not lower than the Miami data based on our comparisons. Lacking of organics in the tropics could be a possible cause, but there is no reliable data globally to confirm. On the other hand, the low AOD over the tropics is approaching the satellite detection limit, so satellite data are usually biased high, as the MAN data suggested. For the SH mid-latitude, GOCART is not too low – it is actually a little higher than the MAN sunphotometer data from Smirnov's paper, although it is lower than the current MODIS and MISR (but could be in a much better agreement with the new version of MODIS, per Levy et al. AMT 2013 paper). We have included some of these comments in the text.

General comment: the interannual variability is strongly determined by Pinatubo eruption- and it seems accurately reproduced by the model. I wonder what picture would emerge when removing the stratospheric sulfate contribution from the model (and equally form the observations)? What fraction of the interannual variability would be reproduced?

It is difficult to remove stratospheric sulfate from satellite AOD data (also see our reply on the end of page 8 to the top of page 9). As for the model, the interannual variability or trends of FF+BB and dust can be seen with the red and orange bars, respectively, in Figure 5. On the other hand, Pinatubo has very minor influences on surface sulfate concentrations.

p. 19769 l. 12-16 it is difficult to see these trends by eye- it would be good if more qualitative statements can be made- including the comparison with the emissions trends and earlier studies.

Comparisons with the anthropogenic emission trends are discussed in section 5.1 and shown in Fig. 9. Quantitative numbers are given in Table 4 for the four time slices (1980, 1990, 2000, 2009). Now we have converted Table 4 to a figure to convey the information, as another referee suggested. We also add more quantitative discussion.

p. 19770 l. 29 Here an issue with Modis-Terra is mentioned- elsewhere issues with other instruments were mentioned. This is of course not surprising. I would however find it useful if the authors could work on a more general way to evaluate their model results. When are the retrievals and trends in retrievals robust enough to make statements on discrepancies and consistencies of the model evaluations. In this particular case a statement is missing on why an off-set would exist in SEA-SHL-SAM-ANZ and not elsewhere?

The text regarding the MODIS Terra and Aqua offsets has been removed as an action to shorten the paper. In any case, this is rather technical and has a lot to do with the instrument calibration and the offset is seen different with location and time. Levy et al. has papers describing these issues and we decide to leave this part out in our paper.

P. 19771: General comment: the choice of large ocean regions precludes the analysis of outflow regions trends (e.g. North America); but also the Arabian peninsula- where large satellite trends were observed- but somehow not so visible in this analysis. Can the authors look into this?

Yes, the size of region will determine the magnitude of trend, but there is no best choice of a region domain. We decided on these domains with an intension of balancing the characteristics, the extent of coverage, and the number of regions. Averaged over a large ocean domain will dampen the magnitude of changes near the coastal area, but we are trying to convey the geographical information of the changes in the maps shown in Figure 6a and 6b. We feel in this way both the regional trends and location of the changes can been seen from a combination of the figures showing regional averaged time series and the maps.

The AOD increase over Arabian Peninsula (labeled as “MDE” in Fig. 5) is clearly seen from the SeaWiFS data (Fig. 5, also the diff maps in Fig. 6). MISR also shows the increase although with a lesser amount. GOCART agrees with the direction of change (increase) but the magnitude is lower, as shown in both Fig. 5a and 6b.

p. 19771: The two explanations offered on missing sources would probably deteriorate the model performance.

Agreed. Jaegle et al. (2011) showed an improved agreement with satellite data in the tropics after incorporating the temperature-dependent adjustment of sea salt emission that increases the emission in the tropics but decreases it at high latitudes. As we have shown and now also mentioned in the text, model comparisons with the U Miami data of sea salt measurements in the tropical region does not seem to suggest that our sea salt is too low. By increasing sea salt emission, our agreement with the U Miami data would indeed deteriorate. Adding primary organic emissions might help, to some extent, improve the agreement with satellite AOD data in the tropics. But, as always, there is hardly a global, universal fix that would improve the agreement between the model and every kind of data.

p. 19772 l. 13: Do I understand correctly that the author find a large contribution of volcanoes (Pinatubo; El Chichon) in maritime aerosol. Or do the authors want to say that AOD over marine regions in those periods is determined by stratospheric aerosol (not maritime aerosol). I would urge the authors to present an appropriate evaluation of surface seasalt aerosol – a number of years ago S. Gong has presented a rather rigourous evaluation- data could be retrieved from him?

We have clarified what we meant by "maritime" aerosol (which is different from "marine" aerosol). Now the text reads: "Interestingly, the composition from GOCART indicates that sea salt is not the major components for column AOD over the ocean…".

The University of Miami group (Prof. Prospero) has provided us with the sea salt concentration measurements, along with other aerosol species, at more than 10 stations. Earlier time periods from these data were used by Gong et al. We have compared model output with all data and showed the comparisons at 4 sites in Fig. 8. Now we include all plots in SM. Overall, the modeled sea salt is slightly higher than the U Miami data (overall stat in Table 3). Considering most sea salt is concentrated in the marine boundary layer, it seems our sea salt is not too far off, unless the optical properties that convert sea salt mass to sea salt AOD is wrong (i.e., mass extinction efficiency too low), or our hygroscopic growth is too weak. These are unfortunately hard to verify. However our AOD in the Southern Ocean, which consists large amount of sea salt, does not seem to be too low.

p. 19773l. 29 Again I suggest the authors bringing this extensive and interesting evaluation to some point on what we learned on the model performance? What is the more likely case?

We have added the model results that are outside of the satellite AOD range (new Fig. 4b) to show the locations that model is higher or lower than all satellite data. As for ocean specific, we have added a sentence quoting the results from Smirnov et al., 2011 that shows a general high AOD bias from both MODIS and MISR compared to the MAN ship-borne sunphotometer measurements. The overall model performance is summarized in the text.

p. 19774 I agree with statements made on the caveats of the choice of 2 years at the beginning and end of the respective periods. As earlier indicated, why didn't the authors try to correct for stratospheric aerosol after eruption. Using model results to derive tropospheric aerosol columns for both model/satellite would be one way. Using SAGE to get measured stratospheric aerosol could also be a possibility. More realistic trends analysis would be possible.

Using SAGE to remove AVHRR stratospheric fraction is difficult, because of insufficient overlapping spatial coverage from two very different types of measurements. There will be quite large uncertainties in the resulting "tropospheric residue" by subtracting SAGE from AVHRR. Besides, Pinatubo also have large influences in the upper troposphere to skew the tropospheric trends. The choice of two years is much constrained by (1) avoiding major volcanic influences, (2) satellite data overlapping period (TOMS ended in 2001, MISR and MODIS started in 2000, so there is only two-year overlap between TOMS and MISR/MODIS and we would like to include these two years which also serve as the beginning time for the 2nd time segment), and (3) a near neutral ENSO index after averaging of the selected years (see the figure in the Reply to Referee #2 that explains the selection of the years to average). These criteria determine the beginning and ending years for each time segment. Even though 2-year avg is relatively short, the trend features are robust over pollution and dust regions. What cannot be inferred to as "trend" is the areas with volcanic activity (e.g., near Japan) and biomass burning regions, where interannual variability is large and the selection of 2 years would just show these variabilities instead of trends.

p. 19777 l. 28 discussion on uncertainties should be elsewhere.

Removed.

p. 19780-82 The authors show snapshots of a couple of stations with surface observations. While this is useful- the authors should explain why these particular stations were chosen. How valid is the site-specific discussion for a larger set of data? I understand that a comparison of a larger set of observations is analysed in Table 3. This table should mention the amount of stations and requirements regarding record length, data completeness, etc were made.

The number of stations and observation period of each network have been given in Table 2, however the actual record length varies from station to station. We now make it clearer that the stations chosen in Fig. 8 for their geographic locations, length of observations, and completeness of species.

p. 19785 The relationships between emissions and AOD are interesting (Figure 9) especially when similar analysis can be performed for a range of models. Important factors determining these relationships are indeed local emissions and formation, long- range transport- and removal. I wonder why the authors have not utilized the simulations performed in HTAP to get a handle on the fraction of aerosol column/surface aerosol that is transported from elsewhere. Also why is the 'zero' anthropogenic emission case not include in the analysis. A sensitivity analysis with 1 meteo year- and an high/low emission case could give insights in the sensitivity to chemistry; whereas two different meteo years and 1 emission case could do likewise for the role of meteorology. With other words there are some opportunities to quantify the drivers behind the regional differences in these relationships.

Thanks for the suggestion regarding the HTAP work – we have added text regarding the HTAP general results about transport from extra regional sources, although the regional domains in HTAP 2010 report are different from those in this study.

The reason that "zero" anthropogenic emission is not used in Figure 9 is because the anthropogenic was turned off globally so effects of extra regional transport will not show up in the regional relationship. For example, since all BC is from FF and BB sources, the global "zero" emission will have only one point on the BC panel in Figure 9, which is [0, 0] (zero emission, zero concentration, zero AOD).

Sensitivity analysis: In the last HTAP analysis I (Mian Chin) did analyze model results from the standard, 20% reduction, and zero emission model runs (SR1, SR6, and SR6z) from multiple HTAP models with 1 meteorological year, and found slight non-linearity of sulfate concentration to the change of SO2 emission, but BC and OC responded to the emission change quite linearly. Unfortunately that part (figure + text) was not included in the HTAP 2010 report due to the page limitation. We have also conducted a 30-year run that use the same anthropogenic and biomass burning emission but let the met to vary. We plan to devote that sensitivity analysis in another manuscript focusing on analyzing the role of meteorology on aerosol variation and intercontinental transport. As for the present paper, our intention is to discuss the relationship, rather than to perform in-depth analysis on these relationships.

p. 19789 Again an interesting analysis that would be almost worth an deepened analysis in a separate paper.

Yes, we are working on that!

Reply to Referee #2:

(Original comments in italics and reply in indented paragraphs. Note that all the figure numbers in our reply under each point are referred to as in the ACPD version before revision.)

This manuscript documents the results comparing the aerosol variations and trends simulated by GOCART model during 1980–2009 against the observations from multiple satellite sensors and ground-based networks. Results show that the AOD and surface aerosol concentration reduced in Europe, Russia and North America, but increased in East Asia and South Asia, while the global mean AOD shows little changes over land and ocean in the past three decades because of the concealing of opposite trends in different regions. The study highlights the need of regional-scale assessment for aerosol, including its concentration, optical properties and radiative forcing.

Overall the results are interesting and evaluation for AOD and aerosol concentrations are valuable to the GOCART and even a boarder community. The analysis on dust emission and loading is insightful. While I believe this study merits to be published in ACP, the manuscript can be much better organized and presented in a more condensed way. The current version of manuscript is very lengthy but poorly organized, which make it very difficult for readers to catch up the major points of this paper. Here below are some comments that may help authors to better organize the paper and present the data in a context that would yield more scientific insights.

(1) Usually we first have initial science issues we want to address and major points we want to make in mind before we formally start writing a paper, then we only select the relevant results that serve to address those issues or points to be included in the paper. The authors often tend to present whatever they have evaluated and analyzed, but the reality is that readers often only quickly go through the conclusions and abstracts, and read the relevant figures/tables. For this paper, it is not clearly to me what are new progress and insights? Is this a first 30-year simulation of GOCART? And a newer version of GOCART? If yes, what aspects are new? How is the model performance comparing with previous version and other AeroCom models? I would suggest significantly refine and shorten the paper (see my other comments 2 and 3) by removing the results that are not relevant to the key points you want to make. Instead, more clearly list the major findings and progress in the Conclusions and other related parts.

> We have modified the introduction to make it more clear our contribution after mentioning previous studies: "Expanding from previous, abovementioned aerosol trend studies, this study provides an comprehensive synthesis and assessment of aerosol variations over the last three decades (1980-2009) in different regions of the globe through a global model analysis of multiple-platform datasets, with a goal of determining the anthropogenic and natural contributions to the multi-decadal changes. We also use the model to examine the relationships between emission, surface concentration, and column AOD, and factors controlling the long-term variations of dust aerosols."
>
> We have condensed the paper to (1) limit the regional trend discussions to include 4 pollution regions, 4 dust regions, 2 biomass burning regions, and 6 oceanic regions for their higher relevancy to our purpose, (2) reduce the AERONET section to only show the regional statistics on the comparisons with AERONET data, and (3) consolidate the model-data

comparisons and trend analysis of satellite and surface measurements into the same sections (i.e., over land or over ocean), and (4) shorten the model description to include only necessary information to assist the later discussions. More complete figures and model descriptions are now moved into the Supplemental Material.

(2) Too many subregions. While I understand the rationale using 27 sub-regions in the study, some regions show very similar variability and trend and some regions are not mentioned at all in the discussion. It might be too much work if I am asking to regroup subregions and redo analysis, but it may make sense to suggest only selecting those typical regions (you really mentioned in the discussion) over land and ocean to present in related figures/tables. 8-9 regions over land and 4-6 regions over ocean probably will make the presentation much clearer and quality of figures (e.g. Figure 5) much better.

See above – we now showing 10 regions over land and 6 regions over ocean. Figures for other regions are now moved to SM.

(3) AERONET data. Section 4.3, including Figures 7 and Table 3, present the evaluation of model AOD with AERONETS data. It's not clear to me what the major conclusions are obtained from this part. Authors provided a Table 3 and ask readers to summarize based on this table, but I believe this should be authors' job. The model significantly underestimates the AOD over the Mexico City and Kanpur, authors suggest this is because of missing aerosol types, emissions, PBL, or coarse resolution, etc., which are just very general suspects. In fact, it should be one of objectives of this study to narrow-down the suspicious list with further in-depth analysis. Given the great spatial variability of aerosol over those polluted regions with complex terrain, I don't think it is so meaningful or insightful to compare the in-situ point measurement with the coarse-resolution model results. I would suggest remove this section, unless authors can wrap up some insightful and solid conclusions after further in-depth analysis are done.

Detailed and/or in-depth analysis of model problems over Mexico City and Kanpur would require a separate, devoted effort. We have done several analysis on this, particularly over India, and it is clear at this time that the missing aerosol types (e.g., nitrate), too low RH in the model, and coarse resolutions are indeed all contributing to the low bias of model AOD and surface concentration. There is also information on possibility of emission being to low because of some sources are not accounted for. However, at this time, we cannot quantify the responsibility of each process. In fact one of the coauthors in the present paper has been funded to investigate the model problems over South Asia specifically, so we should have a better answer and solution in the near future.

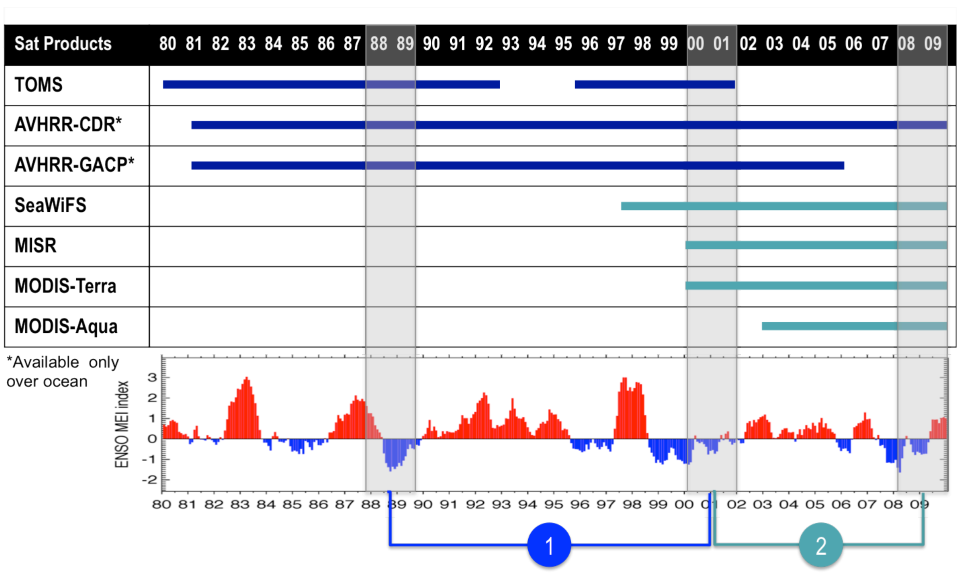
Taking the Referee's comments and meanwhile considering the role of AERONET AOD as a "standard" for satellite and model evaluation, we now limit our comparisons with AERONET on regional statistics rather than comparing the monthly time series over many sites in Fig. 7. This also helps reduce the length of the paper.

(4) Table 4. Given the reality that fewer readers can have patience to read such a detailed table and find the main conclusions from it, I would suggest to use a smartly- designed figure to replace the Table 4 to facilitate to capture the major information that authors want to delivery.

We are now displaying the information in Table 4 in a figure.

(5) Global pattern change. Section 4.2.3 (figure 6) uses the difference of two 2–year averages to represent the global change of AOD. Due to the large inter-annual variability of AOD mainly resulting from the inter-annual variability of meteorology, it is not possible to quantify the statistical significance of the differences for both model and satellite data if only two years of data are used. The comparison of model and observation over the regions where the difference between two periods is not statistically significant for either model or satellite data is not meaningful, that's why in climate community the simulations have to be run for decades or ensemble simulations are needed. So I suggest select at least 5 years to get an average and do the statistical test for difference, and discussion only focus over the region where the difference is statistically significant.

Our choice of the two-year average is constrained by, as indicated in the text, the need to (1) avoid large volcanic influences, (2) include maximum satellite datasets with their overlapping time period (TOMS ended in 2001 but MODIS and MISR started in 2000 so there is only two-year overlap between TOMS and MISR/MODIS and we would like to include these two years which is important as they serve as the end time for the 1st time segment and the beginning time for the 2nd time segment), and (3) have a near neutral ENSO index after averaging of the selected years to minimize the meteorological differences between the beginning and ending time period. The figure below helps explain our selections of the 2-year period (this figure is now included in the SM). Even though 2-year avg is relatively short, the features are robust over pollution and dust regions. What cannot be inferred as "trend" are areas with volcanic activity (e.g., near Japan) and biomass burning regions, where interannual variability is large and the selection of 2 years would just show these variabilities instead of trends. We have strengthened the arguments of how to read the features from the 2-year averaged maps.



(6) Multiple satellite datasets. To highlight the uncertainty associate with the observations, this paper (e.g. Figure 4) provides several AOD datasets from different satellite sensors, which is something that should be encouraged. However, the challenging for this approach is it will become more difficult to obtain an assertive conclusion (whether model underestimates or overestimates) if the model result is located within the bounds of observations (e.g. model AOD

is smaller than MODIS but larger than MISR). I don't have a perfect idea to figure out this issue, one suggestion that authors could do, is to identify the regions with higher confidence (smaller standard deviation among multiple datasets), which can be done by calculating the inter-observation datasets standard deviation (e.g. Figure 4). If the subregion discussed is located in the high-confidence area, then the conclusions can be more assertive and reliable. Otherwise we have to be more conservative for the discussion over the regions with lower-confidence (larger inter-satellite disparity).

We have now added two panels in Fig. 4 showing (1) satellite data diversity (standard deviation of satellite datasets divide by the multiple satellite mean x 100%) and (2) location and amount of model outbound of satellite max or min. These panels illustrates where the satellite data agrees the best or worst and where the model is most likely to be too high or too low compared with these satellite data. We have also added to the text on the robustness of the data, in the context of evaluation with the AERONET data in each region.

Reply to Referee #3:

(Original comments in italics and reply in indented paragraphs. Note that all the figure numbers in our reply under each point are referred to as in the ACPD version before revision.)

Comment on the paper "Multi-decadal variations of atmospheric aerosols from 1980 to 2009: sources and regional trends" by M. Chin et al. submitted to ACP.

This paper examines the variations of aerosols during the last 3 decades using results mainly from the GOCART model and observations either from several satellite instruments or from ground-based networks. The paper is rather lengthy and it is not always easy to follow the authors. I think a more condensed version of section 4 with focus onregions presenting significant variations during the last 3 decades will demonstrate the work done by the authors making the paper easier to follow. I suggest the publication of the paper once the following comments are addressed.

Major commnets:

(1) In Section 4, a table similar to Table 4 is needed, presenting the AOD values for the satellite instruments. This table will clearly show the regions where there is agreement about the trends/variations among the satellite instruments and the model, in addition to Figures 5 and 6, but providing more quantitative results. These comparisons will permit to authors to concentrate to regions with significant trends/variations, rather than presenting all of them in details. Given the uncertainties and the contradictory results (in some regions) between the different datasets, I hardly see the utility in doing so.

> We have now eliminated the Table 4 as the Referee #2 suggested. We convert the information in Table 4 to a figure, and will add the satellite information on the figure as well. The challenge of listing the satellite data is that there is no single datasets covers 30-year span with both land and ocean information. The more reliable (EOS) satellite data are only available from 2000, which cannot be used to draw a conclusion of trends over the past 3 decades, although they are useful for showing the last 10-year variations. In addition, the anthropogenic fraction can only be obtained from the model.

(2) Although, the validation of the model (and the satellite data) against ground-based observations is very useful, by showing just results for some stations is not enough to corroborate or reveal any contradictions presented in the previous sections about the trends/variations from the model and the satellites. Especially, by presenting only global results, as in the case of AERONET or island stations, the authors contradict themselves (see the last sentence of the abstract about regional changes), because the capability of the model to reproduce the observations is not the same in all regions nor in all times. I understand that this is a difficult task due to gaps in the time series, but the authors can use the stations with the more complete data sets, especially within regions where the majority of stations indicate similar trends/variations. Also, they could add some columns in Table 3 to indicate the temporal evolution of the aerosols concentrations/AOD in the specific regions (similar to what they have done for the model in Table 4).

Thanks for the suggestion – we will select the stations with at least 15 years of data and to assess the variations/trends in specific regions. We will also show the statistics of comparisons with all station data (AERONET and/or concentrations) as a model evaluation separated by regions.

Minor comments:

1) Page 19753, line 23: Add also terrestrial radiation to the solar radiation, already mentioned.

Added.

2) I propose to put the majority of section 2.2 in the appendix as to me provides technical details, except from the last paragraph.

We have now shorten the model description and emission sections (section 2.1 and 2.2) and put details in the Supplemental Material (SM).

3) I think it will be useful for the discussion to provide the accuracy for the model and the observation datasets in Sections 2 and 3, as already done for AERONET.

We will add in the text about available estimates of the accuracy or uncertainty of some datasets. It depends on many factors, such as instrument, measurement technique, sampling conditions, aerosol amount, derivation methods, etc. As for the model, the "accuracy" can only be evaluated against the data, and the accuracy varies with location, time, and composition. We will talk about those based on our previous and current comparisons.

4) The scale of the different regions/stations in Figures 4, 5, 6 (for AVHHR-CDR), 7, 8, 9 and 10 changes. Provide a note. I believe that you should use the same scale (Figure 4) or homogenize as possible (e.g. same scale in the rows of Figure 5 etc.).

The scale in Figure 4 is the same except the largest number shown at the end of the highest scale. We have now made a note in the Figure caption. For the regional trend time series plots, we now use the same scale for the panels in the same rows.

5) Page 19775, line 6: There is no obvious decrease from TOMS above Russia and Europe due to little coverage (see also your comment below, line 21-22).

We have changed the sentence to "TOMS shows an AOD increase over Asia and southern Africa, a strong decrease over northern Africa and Brazil, and a hint of a decrease over Russia and Europe from its rather limited spatial coverage at high latitudes."

6) Pages 19781-19782, Over the Arctic: Why present the two stations Barrow and Spitsbergen from the moment that their dataset is smaller? What is the additional information to the other two stations?

We have removed Barrow and Spitsbergen from the figure.

7) The Polar regions are not indicated in Table 1 and Figure 1, although appear in Table 4 and in Section 5.3. Also, you are using the surface measurements of Arctic sites (Section 4.4.1), but these are located over other regions (CAN and EUR). Please clarify/modify appropriately.

We have removed the polar regions from Table 4 (which is now replaced by a new Figure). We also make it clear that the Arctic sites shown in this paper are located in either CAN or EUR but within the Arctic Circle.

Technical comments:

1) Page 19762, line 24: Change 2012 to 2013 for the paper of Zhao et al.

Changed.

2) Page 19779, line 14: The reference of Pan et al. is missing.

We have changed the text on page 19779 to "(Pan, X., et al., Evaluation of aerosol simulations in multi-models over South Asia, manuscript in preparation, 2013)". We will update the status of such reference as the present manuscript progresses.

3) Page 19800, line 3: The paper of Kalashnikova et al., in now published in AMT.

Updated.