Response to Referee #1

Response) We thank the reviewer for their constructive comments and suggestions and have made several changes to the paper to address the issues raised. Reviewers' comments are shown in italics with our response shown after each.

The paper is an evaluation of the aerosol distributions simulated by the TOMAS aerosol scheme used in the GISS modelE2 global climate model. The evaluation is similar to most papers of its kind, but covers a larger number of observational datasets to address more aspects of the aerosol model (for example including size distributions). The paper is most interesting and useful to other aerosol model developers when it clearly explains the reasons behind the choices of parameters (e.g. size distribution of emissions, parameterisation of DMS emission rates, fraction of primary sulphate emissions, differences with bulk mass scheme, etc.) and behind model skill, or lack thereof, when compared to observations. This good level of explanation is not always present, however, and most of my comments below suggest improvements in that direction.
The paper could have taken a different – and arguably more original – direction, looking at how improvements in the host model, from II-prime to ModelE2, have affected the skill of the aerosol scheme at reproducing observed aerosol distributions with fidelity.
This is an important aspect of global aerosol modelling, as yet almost unexplored in the literature. The paper is long, with a large number of Tables and Figures. However, considering the breadth of the model evaluation, it is difficult to recommend shortening the discussion or removing figures (with the possible exception of Figure 25). The conclusion (section 6) is a good summary of the findings. I recommend publication after the following comments, aimed at improving the discussion, are addressed by the author.

1 Main comments
• *Section 2: Since the main motivation for developing ModelE2-TOMAS is to be able to use TOMAS in a better model than II-prime (page 5835, lines 8), one would have expected a more complete discussion of ModelE2 compared to IIprime, especially on those aspects that are relevant to the life cycle of aerosols and their radiative effects. So section 2 should be extended with a discussion of changes in cloud, precipitation, and transport schemes, summarising the improvements in those and how they are expected to impact on the quality of the aerosol simulation. A lot of a model's skill at simulating aerosols does not depend on the aerosol scheme itself, but on the host model – and this dependence has been little investigated in the literature so far. I would strongly encourage the authors to look into that aspect.*
Response) I agree with the reviewer that it would be very interesting to explore how aerosol model predictions depend on a host model configuration. It is disappointing to mention that because GISS-TOMAS has been unofficially retired and is not available to run at this time, it is not possible to compare two TOMAS models. Also, especially due to the lack of documents for the old II-prime model, it is difficult to describe how non-aerosol GCM components are different between the two GCM versions.

• *Section 5.2: The AeroCom 1 simulations are a decade old now, and comparing against those*

does not mean much – aerosol modelling has progressed since then, and comparing to older – presumably poorer – models does not really demonstrate skill. The comparison should therefore be restricted to AeroCom 2 models (comparing total carbonaceous if a more detailed split is not possible) and ACCMIP. Also, since emissions vary between the different studies, comparing absolute values for burdens and mass fluxes is not really useful: the focus should only be on lifetimes and relative contributions to deposition rates, with burdens given on Table 5 for information only.

Response) We have deleted the comparisons to AeroCom Phase 1 budgets except for the lifetime and removal rate coefficients comparison, as the reviewer suggested. We only use the ACCMIP multi-model budgets to compare the burden, source, and deposition rates to compare sulfate and elemental carbon. We can't use the AeroCom Phase 2 budgets from Myhre et al. (2013), mainly because their budgets are anthropogenic portion, which is the differences between the present-day run and the preindustrial run. As the changes in the revised manuscript are large, please see the revised manuscript for the changes.

2 Other comments

• *Page 5831: The title is misleading, because the paper is really about evaluation. Development details are delegated to previously-published papers. I would therefore drop the word "development" from the title.*

Response) We have modified the title to

"Evaluation of the global aerosol microphysical ModelE2-TOMAS model against satellite and ground-based observations"

• *Page 5833, line 16: Strictly speaking, the pre-industrial atmosphere matters only for the radiative forcing, not for the radiative effect.*

Response) Thanks for catching that. It's now corrected in the revised version.

• *Page 5834, line 11: "very accurately": I'm not sure what the point is here. Are the authors suggesting that single-moment representations are inherently less accurate in what they can simulate?*

Response) Yes – at least when compared for the same number of size sections. Single-moment sectional model assumes an average mass of particles in each bin to be constant and it is not changed over time, leading a substantial numerical error in a number prediction. In reality and in the two-moment scheme, the average particle mass in a bin is varied with time, and this flexibility leads to greater accuracy. A two-moment model does not assume the average mass in a bin to be constant and overcomes the problem stated above. Harrington and Kreidenweis et al. (1998) shows much more variables are need for a single-moment sectional model to predict accurate simulation. Also, Feingold et al. (1988) shows the superior of a two-moment method compared to a single-moment. We have included these references in the revised manuscript.

Debra Y. Harrington and Sonia M. Kreidenweis, "Simulations of Sulfate Aerosol Dynamics Part II: Model Intercomparison," Atmospheric Environment 32, no. 10 (May 1, 1998): 1701–9,

doi:10.1016/S1352-2310(97)00453-6.

G Feingold, S Tzivion, and Z Levin, "Evolution of Raindrop Spectra. Part I: Solution to the Stochastic Collection/Breakup Equation Using the Method of Moments," Journal of the Atmospheric Sciences 45, no. 22 (1988): 3387–99, doi:10.1175/1520-0469(1988)045<3387:eorspi>2.0.co;2.

• *Page 5837, section 2.1: What is the mixing assumption in the bulk model: external? Same question for TOMAS: how are aerosols mixed within each size bin (Page 5838, line 11)?*
Response) The bulk aerosol model (i.e., ModelE2-OMA in the revised manuscript) assumes external mixing, and the TOMAS microphysical processes assume internal-mixing state in a size bin. We included the following sentence in the revised manuscript.

"In general, TOMAS treats all aerosols as internally mixed during microphysics such as calculating condensation and coagulation rates. However, a portion of EC is treated as externally mixed for purposes of wet deposition."

• *Page 5838, line 17: Ammonium has just been discussed, so this statement seems redundant.*
Response) The bulk tracer NH_4 has been removed.

• *Page 5838, line 23: How is hygroscopic growth represented then?*
Response) For sulfate and sea-salt, we use a polynomial fit based on ISORROPIA, a thermodynamic equilibrium model for inorganic aerosols (Nenes et al., 1998). For organic carbon, it is based on the observations of Dick et al. (2000). We have added the following statements in Section 3.

"Water uptake by sulphate and sea salt is based on a polynomial fit based on ISORROPIA, a thermodynamic equilibrium model for inorganic aerosols (Nenes et al., 1998). For water uptake by hydrophilic OM, it is based on the observations of Dick et al. (2000)."

• *Page 5839, line 5: Why are the other kinds of nucleation unused? Ok, it is mentioned in section 4.2: it may be useful to point out here that the reason for not using a given parameterisation will be discussed later in the paper.*
Response) We have moved the following sentences from Section 4.2 to Section 3 (model description).

"The boundary-layer nucleation is off in all simulations because it tends to overpredict aerosol number concentrations in our model. Also we do not show any run with the ternary nucleation (Napari et al., 2002) because it overpredicts aerosol number concentration severely (not shown)."

• *Page 5839, lines 6–19: For the benefit of other model developers, it would be interesting to say how the size bin/cutoff configurations are selected. I assume it is a compromise between*

computation cost and fidelity of the model, but how is that latter quantified?
Response) The computation costs and numerical errors of the Fast TOMAS models are presented in Lee et al. (2013b). In Lee et al. (2013b), we compared the different versions of TOMAS models to an analytical solution of condensation and coagulation in box-scale and also compared the fast versions of TOMAS models to the original TOMAS model, which was our benchmark. We have added the following in Section 3.

"As discussed in Lee and Adams (2012), the Fast TOMAS reduces the computational burden by 2-3 times while generally predicts CCN concentrations within a few percent of the original TOMAS."

• *Page 5840, lines 1–3: Is the fraction of precipitating cloud water computed in each model layer? Does the wet deposition flux account for re-evaporation of precipitation?*
Response) Yes. The fraction of precipitating clouds water is computed for each model layer. Wet deposition accounts for the changes of tracers in precipitation-water by its re-evaporation. We have modified the following sentence.

"The fraction of activated aerosols removed by wet deposition is proportional to the fraction of cloud water that precipitates, **which is computed in each model layer. Wet deposition accounts of re-evaporation of precipitation.**"

• *Page 5882, caption of Figure 1: The caption could be improved to make clear that the Table lists aerosol and precursor emissions, and how nucleation is accounted for.*
Response) I think the comment is for Table 1. The Table 1 caption is now revised as below.

"Table 1. Aerosol and precursor gas emissions used in TOMAS and the bulk aerosol models and the nucleation scheme used in the TOMAS simulations"

• *Page 5841, lines 15–18: I understand the need for pragmatic choices like this one, but it would be useful to offer an explanation as to why the Nightingale et al. (2000) leads to an overprediction in the Southern Hemisphere in both aerosol schemes. Is it because of other aspects of the model?*
Response) Boucher et al (2003) simulates DMS using the Nightingale et al. (2000) scheme (used in ModelE2-OMA) using a global climate model, LMD-ZT. The overprediciton of DMS is not shown in Boucher et al. (2000). Their DMS mixing ratios at the same observation sites (Amterdam Island and Cape Grim) are about 50% of the ModelE2 DMS values and is more comparable to ModelE2-TOMAS DMS, which uses the Liss and Merlivat (1986) scheme. This seems to suggest that ModelE2 has a stronger surface wind speed than the LMD-ZT model used in Boucher et al. (2003), but we need to investigate this further in the future to confirm.

While doing this, we have updated the DMS measurements at Amsterdam Island and Cape Grim to cover a longer time period. The most noticeable change from this update is the observed DMS mixing ratios during the winter at Amsterdam Island, which are about two

times higher than our original DMS observation. This is because DMS has a large year-to-year variability (Sciare et al., 2000). With this revised data, the DMS in ModelE2-OMA is overpredicted by about 30% at Amsterdam Island (much less than before) and still more than a factor of two at Cape Grim. We have included the following in the revised text.

"The DMS concentrations seem to agree well against the observations when using the sea-air transfer function of Liss and Merlivat (1986), i.e, the case for ModelE2-TOMAS, but this run underpredicts during the winter season at the Amsterdam Island site and all season at Dumont site. Earlier, we mentioned that the global DMS budgets from Boucher et al. (2003) agree well with those from ModelE2 when using the same DMS emission parameterization. However, Boucher et al. (2003) shows better agreement to the same DMS measurements when using the sea-air transfer function of Nightingale (2000) at Amsterdam Island and Cape Grim (i.e., the case for ModeleE2-OMA), because their DMS mixing ratios from that simulations are actually closer to ModelE2-TOMAS. This may suggest that, at least over SH high latitude regions, the surface wind speed in ModelE2 is much stronger than that in LMD-ZT, resulting in higher DMS emissions and burden. We need to investigate further to find out a source for the difference though."

J. Sciare, N. Mihalopoulos, and F. J. Dentener, "Interannual Variability of Atmospheric Dimethylsulfide in the Southern Indian Ocean," Journal of Geophysical Research: Atmospheres 105, no. D21 (November 16, 2000): 26369–77, doi:10.1029/2000JD900236.

O Boucher et al., "DMS Atmospheric Concentrations and Sulphate Aerosol Indirect Radiative Forcing: A Sensitivity Study to the DMS Source Representation and Oxidation," Atmos. Chem. Phys. 3, no. 1 (2003): 49–65, doi:10.5194/acp-3-49-2003.

• *Page 5883, Table 2: Need to define GMD and GSD in the caption.*
Response) Added.

• *Pages 5884 and 5885: It would make sense to merge Tables 3 and 4, since they are analysed together.*
Response) We have merged the two tables.

• *Page 5846, lines 24–25: It would be useful to remind the reader that anthropogenic emissions are supposed to be representative of the year 2000, which justifies the choice of period for observations.*
Response) We would like to note that, as mentioned in Section 4.1, the year 2000 emissions for SO_2 include the anthropogenic emissions as well as the natural emissions such as volcanic eruptions and DMS oxidation. The year 2000 intends to represent the "present-day". The choice of observational period is not to evaluate only anthropogenic emissions but rather the "present-day" conditions.

• *Figure 2: The red and blue signs are difficult to distinguish. Perhaps use filled circles, or*

another, thicker, symbol? Also, the model overestimates concentrations by more than 10 times for 5 sites of the EMEP network. Do those 5 sites have some common characteristics that could be the signature of a specific weakness of the model?
Response) The modified Figure 2 uses a different symbol to make it more distinguishable. For the EMEP sites showing significant overpredictions, we could not find any common characteristics except that their observed concentrations are on the low side. This requires more detailed investigation, but, for now, we guess that those locations are influenced by overpredicted SO_2 emission nearby.

• *Page 5848, lines 4–5: This is interesting, but also worth an explanation. Why would DMS concentrations increase aloft? Lack of oxidants?*
Response) We suspect that the model has too strong vertical transport by clouds. In fact, we provided the explanation the below paragraph (Pg. 5848, L: 10-15). We have decided to delete the sentence and modified the following part to avoid any confusion.

(Old) "Considering the small DMS peak at 8 km and the elevated SO_2 in the upper/free troposphere"
(New) "Considering the small DMS peak at 8 km and the elevated SO_2 in the upper/free troposphere in the model (see Figs. 4 and 5)"
(Deleted) Interestingly, the model DMS seems to show an increase above ~8 km quite consistently throughout the sites.

• *Page 5848, lines 16–17: It looks like volcanic emissions help the model do the right thing in Mar-Apr, but not Aug-Oct. Is the agreement in Mar-Apr coincidental, or should one apply a seasonality to volcanic emissions? Also what about the peaks in observed SO2 visible above 4 km over Tahiti and Easter Island: that looks like transported aerosol layers. Are those mentioned in the papers on the PEM-Tropic campaigns? Where do they come from?*
Response) Please read below for our response. The following is now included in the revised manuscript.

"A large peak in the mid-troposphere at Hawaii in the models results from volcanic SO_2 emissions, while the observations show a similar peak only during March-April 1999, which is heavily influenced by volcanic emissions (Thornton et al., 1999). During August-October 1996, the observations at Tahiti and Easter Island show transport of volcanic SO_2 emissions in the middle and upper troposphere (Thornton et al., 1999), which the model does not capture. Since our model includes only continuous volcanic emission with a yearly resolution, our model fails to simulate variability in volcanic SO2 emissions at higher time resolution."

D. C. Thornton et al., "Sulfur Dioxide Distribution over the Pacific Ocean 1991–1996," Journal of Geophysical Research: Atmospheres 104, no. D5 (March 20, 1999): 5845–54, doi:10.1029/1998JD100048.

• *Page 5849, lines 5–6: TOMAS sulphate lifetime is one third longer than the AeroCom mean. I would not qualify that as "slightly longer".*
Response) Changed from "slightly longer" to "longer".

• *Page 5849, line 7: Are we to understand that weak dry deposition rates explain the longer lifetime of sulphate in TOMAS?*
Response) The longer SO_4 lifetime in TOMAS should be influenced by both weak wet deposition and dry deposition. Both deposition coefficients are small compared to the AEROCOM mean values (0.04 less for wet deposition and 0.027 less for dry deposition). When increasing the dry deposition coefficient to the AEROCOM mean alone, the overall SO_4 lifetime is decreased from 5.6 days to 4.8 days. Doing the same for wet deposition with no change in the dry deposition, the overall SO_4 lifetime decreases from 5.6 days to 4.5 days. We have added the following right after this.

"However, the longer overall sulphate lifetime is contributed by both dry and wet deposition, rather than dry deposition. When increasing the dry deposition coefficient to the AEROCOM Phase 1 mean alone, the overall lifetime is decreased from 5.6 days to 4.8 days. Doing the same for wet deposition with no change in the dry deposition, the overall lifetime decreases from 5.6 days to 4.5 days."

• *Page 5849, line 9: Is the remainder of wet deposition caused by large- scale precipitation? Also, I guess "moist convective clouds" are in fact simply convective clouds.*
Response) Yes, the remainder (73%) is by large-scale precipitation and the moist convective clouds are convective clouds. In case it could be confusing, we have dropped "moist" from moist convective clouds throughout the text. Also, we have included the following in Section 2 to make sure that the model clouds are distinguished into convective and large-scale stratiform clouds.

"In the model, clouds are distinguished into convective and large-scale stratiform clouds."

• *Page 5849, line 13: Sulphate lifetime is given at 5.6 days on line 6 and in Table 5.*
Response) It was typo. It is now corrected to 5.6 days.

• *Page 5850, lines 1–14: Comparing lifetimes of coarse mode aerosols is only meaningful if the models cover the same size ranges. Is that the case here?*
Response) We agree with the reviewer. We have deleted the comparison of our model sea-salt and dust particles lifetime to the AEROCOM Phase 1.

• *Page 5851, line 6: Readers are left to draw their own conclusion, here, so the paper should be more affirmative: is the dry deposition parameterisation in TOMAS better – or equivalently, are dry deposition rates in other models likely overestimated?*

Response) First of all, we'd like to make it clear that this is to explain why dry deposition rate and lifetime differ dramatically between the two models. Unfortunately it would be difficult to judge which parameterization is superior. One might think the TOMAS dry deposition is better, because it is more physically based parameterization (i.e. accounting for a size-dependent resistance and gravitational-settling velocity) than the one used in the bulk aerosol model. However, the dry deposition velocity assumed in the bulk aerosol model could be more comparable to some of the observations, as it is based on the observations from Wesely et al. (1985). More information is required to draw a reasonable conclusion.

About the second question, we can't say whether other models overpredict their dry deposition rates, not only because we don't have any measurement of dry deposition rates to judge and the topic is further complicated because a dry deposition rate is influenced by the dry deposition parameterization as well as many other factors (e.g., boundary layer scheme, and model surface definitions/properties).

• *Page 5896, caption of Figure 6: Are OC surface concentrations given in terms of [C] or [OM]? There is a factor 1.4 between the two in TOMAS.*
Response) The unit is in term of [OM]. The model simulates OM concentrations rather than OC concentrations, as it matters for particle size.

• *Page 5851, line 24: OM aerosol concentrations are high over North Hemisphere continents, including in regions I would not particularly associate with industry or biomass-burning, especially on an annual average (midwest US, central Siberia). Are those biogenic sources?*
Response) Yes. We have included the following sentence in the revised text.

"Due to the SOA formation, the OM concentrations over Midwest US and Central Siberia are also noticeably high."

• *Page 5897, caption of Figure 7: I presume that units are the same as in Figure 6?*
• *Page 5898, caption of Figure 8: Sulphate in ug[S] m-3 as before?*
Response) Thanks for pointing this inconsistency. The unit used in Figure 7 and 8 was $\mu g\ m^{-3}$. The Figure 6 is now also shown in $\mu g\ m^{-3}$ instead of $\mu g\ [S]\ m^{-3}$.

• *Page 5852, line 6: Figure 7 really shows zonal cross-sections – but the vertical aspect of the Figures is not discussed.*
Response) The following sentences are added to describe the vertical distributions.

"The two small spikes shown the EC and OM concentrations between 10° S and 10° N are a result of the injection height used in the biomass burning emissions. Since a significant amount of sulfate and OM are also formed through chemical reactions in the atmosphere, their vertical gradients are relatively small."

"A strong dust plume is shown at around 0°-30° N due to the large Northern African and

Middle East dust emissions, and a small dust plume at around 30° S due to Australian emissions. In the model, the PM10 concentrations in upper troposphere are dominated by sulfate and dust particles."

• *Page 5853, lines 18–20: Isn't that statement in contradiction with the statement on lines 11–14, where TOMAS overestimated sulphate?*
Response) We understand that they may sound as if we are contradicting each other. The referee #2 also raised a question for the line 11-14, and we have modified the manuscript. Please see our response to the referee #2's comments.

• *Page 5854, lines 1–13: The sea-salt evaluation lacks discussion. The overestimations inland are probably due to the model resolution, which would produce an abnormally large transport from the ocean. The low bias in the Tropics is more interesting: perhaps a consequence of poor simulation of near-surface wind speeds in the host model?*
Response) We have added the following discussions in the text.

"The overprediction of sea salt in continents may suggest that ModelE2 has a stronger transport from ocean to in-land, as there is no obvious overprediction over adjacent oceanic sites. The ModelE2-OMA model exhibits a particularly large overprediction over most SH sites. Both models tend to be biased significantly low at some of the sites near the tropics where the observed sea salt concentrations are high. Similar underprediction is also shown in mineral dust (see Fig. 13). This might be due to fast wet scavenging due to overpredicted precipitation in that area (see Fig. 9 in Schmidt et al., 2014)."

• *Page 5854, from line 14: Again, lack of discussion, this time for mineral dust aerosols. It is interesting that TOMAS seems to be doing better than the bulk aerosol model here. Could that better performance be explained by specific differences between the two models?*
Response) We have explained why ModelE2-TOMAS performs better over continents than ModelE2-OMA. The followings are included in the revised manuscript.

"This is due to the emission size assumptions (resulting in more clay emissions than ModelE2-OMA) and the differences in the deposition parameterizations (resulting in slightly longer lifetime for clay particles; ~ 9 days in ModelE2-TOMAS and 6.5 days in ModelE2-OMA)."

• *Page 5855, lines 24-26: Why not show the comparison against IMPROVE where the OC:OM ratio is consistent with the choice made in the model? The other comparison is of little interest and can be removed. Again, the paragraph could discuss the results of the comparison of EC and OC – the comparison is not too bad in fact. Emissions will probably be the main source of error for the networks used here, which are located relatively close to the source.*
Response) We now use OM/OC ration of 1.4 for IMPROVE sites, and Figure 15 has been updated. The following is either newly added or modified in the revised text.

"The simulated EC and OM in both models agree very well at the IMPROVE sites (for OM, LMNB=-0.17 to -0.08; for EC, LMNB=-0.08 to 0.05). Note that we applied an OM/OC ratio of 1.4 to the IMPROVE network to make it consistent with our model assumption, but the IMPROVE OM data provided to us was based on the ratio of 1.8. Over Europe, the model predictions are still reasonable (within a factor of 2-3), but the agreement is slightly worse than the IMPROVE sites. Since these sites are mostly adjacent to the source/emissions, the good agreement suggests that the emission inventory (used in this study) is well represented for these regions."

• *Page 5856, line 7: It does not mean that they don't contribute to PM2.5 – so seasalt and mineral dust aerosols may be partly to blame for the under-prediction.*
Response) Thanks for pointing this out. Doing more investigation, we found that the model PM2.5 in Oceania is little influenced by sulfate or dust and we also found a small error in PM2.5 calculation for the bulk aerosol model. Please see below for the modified text.

"The PM2.5 overprediction in Oceania is mainly due to too much fine mode sea-salt particles (the overall agreement in Oceana is little influenced by sulfate or dust particles). Note that the sea salt comparison to the Miami dataset (in Figures 11) shows severe underpredictions in several sites in Oceania because their concentrations are likely dominantly by coarse mode sea salt particles."

• *Page 5856, lines 16–19: Please give the full details of the satellite products used here: collection for MODIS, version for MISR, and whether level 3 (monthly, gridded distributions) were used.*
Response) We have included the following statement.

"Specifically, we use Terra MODIS Level 3 (MOD08_M3.051), Aqua MODIS Level 3 (MYD08_M3.051), and Terra MISR Level 3 (MIL3MAE4), which are monthly products with 1x1 degree resolution"

• *Page 5856, lines 25–26: This statement is unclear: does that mean "in the cloudfree fraction of gridboxes" or "in cloud-free gridboxes"?*
Response) It is cloud-free gridboxes. The following is now added.

"where clouds are not present." ➔ ""where clouds are not present (i.e. cloud-free grid-box only)."

• *Page 5857, first paragraph: It should be said that the satellite products do not seem to support the bands of large sea-salt AOD at high latitudes. In that respect, TOMAS does better than the bulk model. Also, the quality of the comparison will depend on the host model simulating clear skies in the right regions and seasons.*
Response) First of all, we want to clarify that both models and satellites show an enhanced AOD band in SH high latitude (compared to the neighboring areas). However, the magnitude

of the satellite AOD seems to be in between the TOMAS AOD and the OMA AOD. We agree with the reviewer that the AOD evaluation would be more meaningful when the host model simulates clear skies in right time. Our current model outputs have monthly resolution, but we think sub-daily model outputs using a satellite simulator seems to be more appropriate to evaluate model clear sky predictions with a satellite production.

• *Page 5859, line 15: Number concentrations from the bulk aerosol model could be derived from the simulated mass and prescribed size distributions, but the comparison would probably not be useful.*
Response) It is possible to compute number concentrations from the bulk-aerosol model. However, the number concentrations in bulk model are not useful, as they are quite inaccurate.

• *Page 5861, lines 2–4: Why is that surprising?*
Response) It was surprising because our previous study using GISS-TOMAS (Lee et al., 2009) did not show a large contribution of dust particles to CCN-sized particles. Instead, when including mineral dust, the model CCN(0.2%) is decreased by 10-20% over dust regions due to microphysical feedback by dust particles. The direct source of CCN by dust emissions was less significant. However, ModelE2-TOMAS shows a significant CCN concentration over dust regions. We added the followings in the revised text.

"This is opposite to the results from GISS-TOMAS (Lee et al., 2009), which shows 10-20% reduction in CCN(0.2%) when introducing mineral dust emissions. Despite the direct source of CCN-sized particles from dust emissions, CCN and ultrafine particles that grow to become CCN are scavenged via coagulation with coarse dust particles, and dust particles compete for condensable sulfuric acid, leading to a slower growth rate of ultrafine particles)."

Y H Lee, K Chen, and P J Adams, "Development of a Global Model of Mineral Dust Aerosol Microphysics," Atmospheric Chemistry and Physics 9, no. 7 (2009): 2441–58.

• *Page 5861, line 16: How was the 30% number obtained? By reducing the altitude where simulated aerosol numbers are taken until the agreement with observations is satisfactory? If so, isn't that being too easy on the model?*
Response) "30% reduction of the altitude" was chosen based on the improvement in the model CN seasonality. It indicates a problem in model orographic transport issue in free tropospheric sites, but this might be an inherent problem by the coarse grid resolution in GCMs.

• *Page 5861, line 25 and Figure 21: It is difficult to tell which of the three simulations does best. Is it BASE? If not, wouldn't that be a reason to redefine the scientific configuration of the BASE simulation, at least as far as nucleation is concerned? Figure 24 strongly suggests that LowNUC would make a good BASE.*

Response) We concluded that LowNUC is the better configuration than BASE in terms of predicting number concentration related to nucleation, but we prefer to use "the least perturbed case" as our BASE scenario.

• *Page 5862, line 6: It is difficult to reconcile Figures 21 and 23. The model seems to a have a high bias in aerosol number over Europe - something that Figure 21 doesn't really show. Should we conclude that this is specific to that region?*
Response) The measured size distribution data are obtained from Putaud et al. (2003), while the measured CN concentration used in Fig. 21 are obtained from Spracklen et al. (2010). The temporal coverage between two datasets are not necessarily same. We found out that except for Jungfraujoch, most sites used in Figure 23 have comparable CN concentrations to the ones used in Figure 21. For Jungfraujoch, the total CN from the size distribution data (from June 1997 to May 1998) is about a factor of two lower than the ones used in Figure 21 (from 1995 to 1999, 2003-2007). This seems indicate a strong year-to-year variability in CN at that site. We have included the following statement in the revised text to warn the difference.

"This data is obtained from Putaud et al. (2003) and, for the same sites shown in Figure 21, the temporal coverage used in Putaud et al. (2003) is not necessarily matched with them. Most sites are reasonably close to the dataset used in Figure 21. However, for Jungfraujoch, the total CN concentrations summed from the size distribution data, which covers from June 1997 to May 1998, is about a factor of two lower than the CN shown in Figure 21, which covers from 1995 to 1999 and 2003-2007."

• *Page 5863 and Figures 25 and 26: Both Figures show similar things, but Figure 26 does it better. With Figure 25, I guess the authors wanted to show the extend of observational variability, but that backfires since the observations are shown to provide little actual constraint. So perhaps Heintzenberg et al. (2000) is not suited for this kind of comparison.*
Response) We do not agree with the reviewer's point. Given that the observations compiled by Heintzenberg et al. (2000) are the collection of some 30 years of marine aerosol observations, the large observational variability (shown in the error bar) is somewhat expected. Their data could help to reveal some climatology of marine aerosols, and thus it is useful for GCM evaluation.

• *Page 5863, lines 22–24: It should be easy to check that convective clouds are indeed more frequent in ModelE2.*
Response) This should be easy to check. However, unfortunately, we don't have an access to GISS GCM II' output (GISS-TOMAS has been retired), so the exact comparison is not possible. Based on other relevant parameters (and some personal communications), we can confirm that the current version of model has more frequent moist convective clouds.

• *Page 5864, line 6: Does the "rest" cover non-marine environments, non- boundary layer locations, or both?*

Response) All CCN data compilation from Spracklen et al. (2011) is in the boundary layer. The rest refers to non-marine boundary layer. However, we have decided to use continental boundary layer (CBL) instead. We have modified the following in Section 5.6.

"In Fig. 27, the CCN data is divided into two groups: CCN in the MBL (marine boundary layer) and CCN in the CBL (continental boundary layer). Note that all CCN measurements used here are in the boundary layer.

• *Page 5864, lines 19–21: Again, the authors rely on the fact that the observational dataset does not provide a strong constraint. This is a bit unfortunate – could a better use of the dataset be made?*
Response) Unlike CTM models, the observations with short duration can't be a strong constraint for GCM. Since climatological CCN measurements are very rare, we performed the evaluations to the short-term CCN measurements, as this is still very useful for a quick check.

• *Page 5866, line 25–28: Saying that aerosol modelling produces large differences is not uninteresting, but if host model impacts had also been discussed in details in the paper, it would have been possible to tell which of the two is the dominant factor. Also, it is possible that the two aerosol schemes are affected by the host model in different ways. Host model effects would then be misattributed to aerosol modelling differences.*
Response) We agree with the reviewer that the impact of host model on aerosol model is an interesting subject, but this is not a scope of this paper. Our main goal of this paper is to evaluate ModelE2-TOMAS.

3 Technical comments
Page 5840, line 22: Typo: "InitiAtive".
Page 5842, line 12: Typo: Stier et al., 2005.
Page 5844, line 5: Delete "that".
Page 5886, caption of Table 5, typo: "standard" deviation.
Page 5853, line 1: The sentence does not read well. I suggest: "For details of the GBD PM2.5 dataset, the reader is referred to..."
Page 5859, line 4: Typo: "underprediction"
Response) Thank you for catching those mistakes. All Technical comments are now corrected as suggested.