Responses to the Comments of Referee #1

(1) The manuscript evaluates the PMCAMx model, comparing the aerosol optical depth (AOD) simulated by the model with observations from MODIS and AERONET. The manuscript fits perfectly the goals of the GMDD journal and the methodology and results are clear. However, as there are no space limitations for this journal, and being GMD(D) dedicated to technical and specific publication I would have expected to have a detail and complete description of the modeling system and the observational datasets used for the evaluation. However, to my point of view, this was not the case. Although I have no real comments on the methodology and results, the lack of detailed information raises serious doubts on the scientific relevance of such an evaluation. The authors should therefore add all the necessary information to the manuscript before this can be considered for publication.

We appreciate the constructive comments of the referee. We have followed the corresponding suggestions adding more information about the model and its inputs. There have been more than 10 papers that have been published describing PMCAMx and its evolution, so we necessarily rely on the corresponding references for a lot of the details. We believe that adding a complete and detailed model description (something that would require probably hundreds of pages) to every GMD paper would be clearly problematic.

Major Comments:

(2) Aerosol concentration evaluation: The aerosol optical depth can be considered as indirect method to evaluate the model performance, as this is normally estimated from the aerosol composition and the radiative properties of the aerosol components. Therefore, I find quite disturbing the absolute lack of any discussion in the capability of the model to reproduce the observed aerosol composition and concentration before to evaluate the AODs. Wrong aerosols compositions would still give reasonable AODs, but for the wrong reasons. Therefore, I urge the authors to evaluate also their aerosol composition results. For example, in the introduction the authors stated that "these errors are probably due to an underestimation of sulfates". I expect to be enough sulfate measurements in Europe to validate this statement, as example with the AirBase dataset (http://acm.eionet.europa.eu/databases/airbase/) which present up to hourly observations for single stations. Additionally, only PM1 evaluation of PMCAMx is mentioned, although this was "limited in space" (page 4, line 15). Therefore, I would recommend to first have a through evaluation of the aerosols fields against measurements (AirBase, EMEP...) before to dig in the detail of AOD. If this was probably published elsewhere, it is impossible to find such reference in this manuscript, also in the PMCAMx description. In the conclusions it has been mentioned that PMCAMx aerosols composition simulation was evaluated, but no publications have been listed.

The reviewer has unfortunately missed the following statement in lines 11-14 of page 4 of the original manuscript: "The PM_1 composition predictions of PMCAMx have been evaluated over Europe in May 2008 (Fountoukis et al., 2011). PMCAMx performance against airborne measurements was as good as its performance against the hourly ground measurements. More than 94% of the organic aerosol (OA) hourly values and more than 82% of the sulfate ones were reproduced within a factor of 2."
This paper provides a detailed evaluation of the ability of PMCAMx to reproduce the detailed EUCAARI campaign PM composition measurements over Europe. This

includes both ground measurements and airborne measurements during the EUCAARI flights. There were approximately 8500 measurements (data points) that were used in this evaluation. Please note that this is the same period as the one analyzed in the present work. We do not believe that the referee's statement about "the absolute lack of any discussion in the capability of the model to reproduce the observed aerosol composition and concentration" is justified.

In this work we exclude the periods with high dust (or in general coarse particles) concentrations, so PM_1 is the appropriate metric for composition evaluation. In the Fountoukis et al. (2011) paper we have used all the available PM_1 composition measurements in Europe for the corresponding period. This is better clarified in the revised paper.

A more detailed evaluation of the ability of PMCAMx to reproduce the organic aerosol composition during the same period has been published by Fountoukis et al. (Organic aerosol concentration and composition over Europe: insights from comparison of regional model predictions with aerosol mass spectrometer factor analysis, ACP, 9061-9076, 2014). A discussion of the findings of this evaluation exercise has been added to the revised paper.

We have added a new section in the paper focusing on just the published evaluations of the capability of PMCAMx to reproduce the PM composition and concentration over Europe to make sure that similar misunderstandings can be avoided.

(3) Period of simulation/analysis: The analysis is focusing on the period 1-29 May 2008. The first information on the period is on page 8, line 15, under the satellite description, which is possibly not the best location for such information. Nevertheless, it is somehow unclear to me why this period has been chosen. Why not the entire month of May? Why not another period of the 2008? Linked to this issue is also the poor description on how PMCAMx has been used to simulate the period of interest. As I am not familiar with the modeling system, it is difficult to me to understand the first sentence of page 6 "To limit the effect of the initial conditions on the results, the first six days of each simulation were excluded from the analysis". Are you referring to multiple simulations? Is the model re-initialized? Or was it a continuous simulation from which only May 2008 was extracted? These pieces of information are essential to put the model into context, but they are largely missing in the manuscript. Possibly few references would help the reader to gather the missing information, if the description of the model set-up would be too tedious. Nevertheless there are simply not there. Finally, it would have been interesting to make an analysis of an entire year, so to cover the different dynamical and chemical space, such as strong aerosol emissions in winter and strong photochemistry in summer. If that is a difficult task, at least few time-slice analysis for different seasons should be performed.

The May 2008 period was chosen for two reasons. First this was the period of the EUCAARI campaign focusing on a photochemically active period with summertime like concentrations. Detailed measurements of PM_1 composition both at the ground and aloft as well as a corresponding emission inventory (prepared by TNO) exist for that period. The second reason was that the ability of PMCAMx to reproduce these detailed PM_1 composition measurements has already been evaluated in previous work (Fountoukis et al., 2011; 2014) and therefore we could focus on the optical properties

of the fine particulate matter in this paper. The exact dates simulated were the same as in the previous publications for consistency. We have tried to clarify the reasons for this selection in the revised paper.

We have clarified the initialization procedure of the model. Given that the initial conditions are quite uncertain and the first few days are dominated by them, we are excluding the corresponding "start-up" period from the model evaluation.

Only one baseline model simulation was performed together with a number of sensitivity tests described in the paper. The corresponding confusing sentence has been rephrased.

We do intend to extend the current work to other seasons. However, each season is characterized by its own issues. For example, during winter there is significant wood burning activity over Europe and the corresponding emissions are probably seriously underestimated in a lot of countries. We have tried to address this issue recently in Denier van der Gon et al. (ACP, 15, 6503-6519) where we try to improve the wood burning emission inventories for the wintertime. The current work focuses on a photochemically active period with the AOD dominated by mostly secondary fine aerosol.

(4) MODIS data: The authors are using MODIS data collection 5.1. Although newer products are available since early 2015 (collection 6), it would be good to know exactly which products you are using. If I am not wrong, in the MODIS collection 5.1 in the AQUA platform both Deep-Blue and Dark-Target algorithm are available. Which one did you use? Additionally, you used "the union of Terra and Aqua MODIS AOD [...]". Could you explain what do you mean with union? How did you unify the two fields? Finally, you are using level 2 data and you calculated the monthly average for May 2008. Which spatial resolution did you use to create such field? How did you merge spatially the observations? As you were using monthly averages, why not using level 3 data? There is a severe lack of pieces of information here that are important to understand how these sensed AODs have been produced. I strongly suggest to fully rewrite the section with the additional information.

We have followed the reviewer's suggestion and we have rewritten the corresponding section of the paper, clarifying the MODIS products used and their processing for the evaluation in an effort to avoid any ambiguities.

Briefly, regarding the details, the Dark-Target algorithm products were used. By "union" we mean that data from both Terra and Aqua datasets were used in order to have better spatial and temporal coverage. We did not alter the values of the data records and we did not apply any sort of transformations. We have changed the corresponding text to make these clear to the reader. The MODIS AOD values, retrieved with spatial resolution $10x10km^2$, were collocated onto the grid of the PMCAMx modeling domain. Over the May 2008 simulation here were several values of MODIS AOD attributed to each cell of PMCAMx. Then, the monthly mean AOD for each grid cell was calculated by taking the average value of the MODIS AODs falling inside it. The L2 MODIS data are better suited to the temporal and spatial resolution of our model. Hourly PMCAMx concentration predictions were used are. Also, the filtering of the MODIS retrievals over Land and Ocean is performed prior to

monthly averaging. Therefore, it is necessary to use L2 MODIS data since they have the necessary temporal resolution. In any case L3 MODIS data are derived from the L2 data.

Minor Comments:

(5) Title: Why PMCAMx-2015? Is that a new version of PMCAMx? If so, it would be great to use the same naming convention through all the manuscript.

We have replaced "PMCAMx-2015" with "PMCAMx" in the title and throughout the manuscript to avoid unnecessary confusion.

(6) Page 5, line 16: The PMCAMx model is using results from WRF as meteorological forcing. Which frequency is needed? Could the author add some comments on the possible error introduced by the non exact dynamics? Is there any evaluation of the dynamics?

We now explain that we used hourly meteorological data from WRF as input to PMCAMx. We have also added details about the application of WRF. The performance of WRF for several air quality-relevant parameters (wind speed, wind direction, RH, temperature) in the same domain is discussed by Fountoukis et al. ACPD, 2016). The agreement with measurements was more than satisfactory.

(7) Page 5, line 20: Same for the emissions: is there any reference and comparison with other emissions dataset?

References for the development of the EUCAARI TNO emissions have been added.

(8) Page 6, line 1: Would be good to mention here the period covered by the simulation(s). Here reads as there are more than one. Could the author be more specific? (see major comments).

The first sentence in page 6 has been corrected to “To limit the effect of the initial conditions on the results, the first six days of the basecase simulation were excluded from the analysis”. Only one baseline model simulation for May 2008 was performed. We have included this information in the text to make it clear to the reader.

(9) Page 9, line 22: What do you mean with "The PMCAMx AODs have been calculated for exactly the same period as MODIS retrievals[...]?" Are you using model results at TERRA/ACQUA overpass (local time)? Are you using daily average for the periods where observations are available? Please specify.

To evaluate PMCAMx performance for the period of May 2008 we only used model results that correspond to the same Terra/Aqua overpass time. Monthly averages are calculated from the corresponding PMCAMx AODs and the MODIS retrievals. We do not use daily average values to make the comparison more exact. We have made the proper changes to the text in order to clarify this important point.

(10) Page 10, line 2: Does it make sense to compare this region when most of the data are masked due to the strong presence of dust aerosols there? Your data sample is strongly reduced, probably not allowing a great statistics here. The same is valid for Turkey and North Africa region.

This is a valid concern by the reviewer. We have added this point in the revised text. Since the data sample size is small in Turkey and North Africa the corresponding

comparisons provide little information. This is the reason that we have avoided discussing these regions in any detail in the paper.

(11) Fig. 4: You mentioned that the white areas mean that not enough dust-screened AODs are present. However, it seems to me that in the Po Valley the white area is much larger that what is present in Fig. 3 form MODIS and PMCAMx. Are you sure that here you are not masking additional values?

This was due to our choice of colors and scale. We have redrawn Figure 4 correcting this problem pointed out by the referee.

Remarks:

(12) To my knowledge the author "Meij" should read "de Meij". Please check the references.

We thank the reviewer for the correction. We have replaced "Meij" with "de Meij".

Response to the Comments of Referee #2

(1) Remote sensing measurements of aerosols represent a valuable complementary to surface in-situ data for CTM evaluation. Indeed, satellite observations provide finely resolved in space AOD data with global coverage, though being of somewhat varying quality due to assumptions involved in the retrieval algorithms. AERONET sunphotometers provide directly measured AOD at high time-resolution. Therefore, last decades those data have been increasingly widely used for model evaluations. In this work, the authors make use of MODIS and AERONET measured AOD to compare with results from PMCAMx-2015 model in order to get better insight in the model performance with respect to aerosol loads. Thus, the paper addresses relevant to the scope of GMD issues.

The article is very neatly and clearly written, and the methods applied are valid, but it does not offer any substantial novelty regarding ideas, data or methodology.
Some of the conclusions appear not to be satisfactorily well founded (i.e. regarding model performance with respect to the individual aerosol types based on AOD evaluation).

The title contain a proper reference to the model used, but does not indicate the short term (one month) and thus limited model evaluation. Besides, only levels of monthly mean AOD have been compared, rather than a complete evaluation. Therefore, I 'd suggest to use "comparison" instead of "evaluation". Also, I'd not advise to include rather hypothetical explanations (lines 22-25), but rather say that the probable reasons of disagreements are discussed in the paper.

In general, the paper is written in good language, the formulations are clear and the supplemented references are relevant and ample.

We do appreciate the constructive comments and suggestions of the referee. The major new methodological improvement in this work is the screening of the satellite retrievals for periods with high dust (or coarse particles in general) concentrations and the combination of the MODIS/AERONET datasets so that the conclusions can be more robust. This is now stressed in the revised manuscript.

We have followed the reviewer's suggestion and changed the word "evaluation" with "comparison" in the title of the paper.

It is clear that comparison of the predicted AOD with the MODIS/AERONET results can shed only limited light on the ability of a CTM to reproduce the composition of the aerosol. We have rephrased the corresponding sentences in the conclusions stressing that the performance of the model for AOD (combined with its performance for composition in the sites where there are ground and airborne PM composition measurements) can be used to derive some tentative conclusions about its composition performance. These are clearly limited to the components dominating the AOD in each area and either suggest problems or lack of major errors.

Other Comments:

(2) The considered period (May 2008) should be indicated in the Abstract and in Sections 2, 3.

We have added the considered period of May 2008 (EUCAARI campaign) in the Abstract and in Sections 2 and 3.

(3) I recommend to include a bit more complete summary of earlier evaluation of all aerosol components.

We have followed the reviewer's suggestion and added a new section in which we provide a more extensive summary of the results of the earlier published evaluations of PMCAMx for the same period focusing on PM composition (see also Comment 2 of Referee 1).

(4) Explain more clearly whether the model calculates size-resolved chemical composition or only size-resolved number density.

We now explain in the revised section 2 that PMCAMx simulates the composition of each size section and therefore predicts the size-resolved PM composition using in this application 10 size bins. PMCAMx calculates the aerosol number from the corresponding mass distribution while its sister model, PMCAMx-UF, simulates both the aerosol number and mass distributions explicitly.

(5) For comprehensive and robust model evaluation and better understanding model result more in-depth analysis should be performed, including spatial and temporal correlations, RMSE, STD etc.

We have calculated additional performance metrics for the model including the RMSE and STD. These provide limited additional insights compared to the four metrics that are currently used in the paper. This information has been added to the Supplementary Material. We agree that the spatial dependence of the performance of the model is useful. We found that the separation of the model domain in areas, given our emphasis on secondary aerosol, was the best way to approach this issue. For the temporal performance we have added in the revised paper some discussion focusing mainly on the average diurnal profiles of the AERONET AOD.

(6) I find the explanations of model vs observations AOD discrepancies by over/underestimation of a particular aerosol components a bit speculative. I would strongly recommend to also include (at least) aerosol evaluation with monitoring surface data in different regions (and airborne measurements if possible) to support the conclusions).

We do agree that these explanations are necessarily speculative. The recommendation of the reviewer is very useful. We have combined the discussion of the AOD performance of the model with its composition performance for the areas (central Europe, United Kingdom and Ireland, North Atlantic, Mediterranean) in which there were PM composition measurements. Combining these date sources does strengthen our conclusions regarding the model performance in these areas.

(7) P.2 lines 13-14: What is the temporal resolution of AERONET data?

The AERONET measurements have a variable temporal resolution varying from 15 min when the sun is high up in the sky to X h when the sun is closer to the horizon..

Measurements start at sunrise when the sun is at approximately 7.5 degrees above the horizon and end at sundown when the sun is once more at approximately 7.5 degrees. This information has been added to Section 3 describing the AERONET data.

(8) P.4 lines 13-16: provide biases for all aerosol species and even better for the regions included in your AOD discussion; only 4 sites with data for sulphates???

We have added the biases for all aerosol species and analyzed them by region thus synthesizing the AOD and PM composition information. We have included the data from both the ground and the airborne measurements and therefore our comparison includes four regions and thousands of data points.

(9) P.7 line 3: How is Mie theory applied for aerosol mass? line 10: Have you made tests on accounting for "brown carbon", i.e. absorbing OC (which is believed to make notable contribution)? Lines 19...Study period? time resolution of AERONET data? AOD at which wave length was used?

We have added a paragraph and the corresponding references clarifying the application of Mie theory of the aerosol size composition distribution simulated by PMCAMx. We have tested in a sensitivity study the effect of the potential absorption enhancement of the BC due to coatings by the other PM components and the effect on AOD for this area and period has been found to be quite small. Given that the biomass burning emissions in Europe during that period were low and that biomass burning is expected to be one of the major sources of brown carbon the effect is also expected to be small. This is explained now in the revised paper. We also clarify in this page the study period (May 2008) and the AERONET AOD wavelength (550 nm). The variable AERONET data time resolution is discussed in our response to Comment 7 above.

(10) P.8 line 7: location instead of part.

We have replaced "part" with "location".

(11) P.9 lines 4-6: I do not understand. Suggest to explain better, or just refer to the sources. Lines 22-23: times coinciding with the satellites' overpasses?

We have rewritten this rather confusing sentence to explain better the binning of the data points for the comparison of the MODIS AOD with the AERONET AOD shown in Figure S1. The comparisons with the MODIS AOD retrievals correspond exactly in space and time, so the times coincide with the satellites' overpasses. We have made the corresponding clarification in the paper.

(12) P.10 line 16: compared with

We have made the corresponding correction.

(13) pp. 11 lines 10-18: Given rather poor quality of emission data for those regions, I feel rather skeptic and "alarmed" about good agreement between model and measurements.

We were also expecting significant discrepancies between predicted and observed AOD over Russia given the uncertainty in the corresponding emissions. However, the agreement was quite good with both AERONET and MODIS. This rather surprising

result clearly requires additional investigation. This point is now stressed in the corresponding section.

(14) p.13 line 4: Rather sloppy formulation

We have rewritten the corresponding sentence.

(15) P.15 line 16-18: This is a rather unfair statement. MODIS data is particularly valuable due to its spatial coverage (besides the AOD errors are relatively small). Line 16: correct "complement" Line 21-22: please, elaborate, otherwise leave out. It's not needed unless model comparison with MODEI and AEROCOM lead to different conclusions.

We agree with the reviewer about the value of the MODIS data and the enormous value of the spatial coverage of the corresponding dataset. This sentence has been rephrased. We have corrected the typo in Line 16 and have deleted the potentially confusing sentence in Lines 21-22.

(16) P. 16 line 7: again "excellent" model performance using poor emission input is typically indicative of some kind of compensating errors. Lines 15-17: too speculative conclusion about model's excelling in calculating all of aerosol types.

In the revised manuscript we repeat at this point the uncertainty of the emissions in this region and the potential existence of some form of compensating errors. We have rephrased the sentence in Lines 15-17 to avoid misinterpretation of the corresponding findings.