The authors thank Referee #1 for useful and thoughtful comments and suggestion to improve the paper. Below we answer to the comments point-by-point. The referee comments are in **bold**. The pages, lines and figures reported correspond to the manuscript under discussion.

Response to Anonymous Referee #1

The paper addresses a very important topic in aerosol retrievals from satellite: the uncertainty of the selected aerosol model and the impact on the AOD. Indeed, this is one of the largest uncertainties for aerosol AOD and the proposed method is very appropriate. It builds on current methods and elegantly generalises the techniques using a sound approach. The paper is well written and well structured and I feel this paper could give an important contribution to a more accurate retrieval of AOD from space based instruments.

There are a few improvement that I deem necessary for this paper to be acceptable. In its present form it lacks a clear definite conclusion and recommendation. A very decent physical and mathematical framework is presented, however at the end the reader is left with a somewhat unsatisfactory feeling, not knowing whether the whole exercise was successful or not. For me the questions that are addressed here are: 1) does the AOD retrieval improve when a combination of aerosol models is allowed and combined using the Bayesian model evidence? 2) Does the model selection uncertainty give a better estimate of the AOD uncertainty than the current one?

The authors pose the questions and address them, but I see no clear answer for these questions. It's left hanging in the conclusion section. It says 'the posterior probability distribution can characterise the uncertainty more extensively than commonly given standard deviation'. Fair enough, but what does this mean? Is it better? Should we generally apply this method? Also from the provided sensitivity studies it is just not clear whether things work as expected (probably leading to the general inconclusive conclusion section).

What I lack is an answer to these questions (supported by evidence): Does the average AOD perform better than the standard one, when compared to AERONET? If not, is this reflected in a larger uncertainty? If yes, are the AERONET and OMI AOD retrievals consistent within this new the uncertainty?

If this could be adequately answered, i recommend this paper for publication.

We thank the Referee #1 for reviewing our manuscript and encouraging general comments.

These are very important question. We have now included more discussion of these points in the end of Section 5 (Discussion and Conclusions). For the sake of clarity and as suggested by Anonymous Referee #2, we made also some other modifications in Section 5.

In order to highlight the main targets of our study, we added in the beginning of Conclusions section:
1) to improve the retrieval error estimate (i.e. to produce more realistic uncertainty estimate),
2) to evaluate the model choice procedure and
3) to find more robust AOD estimate that is based on the average of the most appropriate aerosol microphysical models instead of on a single model chosen probably by chance.

The aim was not actually to develop retrieval algorithm or improve the existing one (OMAERO), but to examine the influence of aerosol model selection to the resulted AOD and uncertainty.

It is difficult to give clear answers to the Referee #1 questions since we have not done yet the comprehensive testing and validation. However, we can present conclusions and give some recommendations based on the set of test cases done so far and based on the application of the method to measurements of one instrument.

1) does the AOD retrieval improve when a combination of aerosol models is allowed and combined using the Bayesian model evidence? 2) Does the model selection uncertainty give a better estimate of the AOD uncertainty than the current one?
1) In general, combination of aerosol models by utilizing Bayesian model averaging approach improve the retrieved AOD. From the test cases we can see that, usually, the averaged posterior gives better AOD estimate than if based on one best model, when compared to AERONET.
2) The uncertainty that accounts for the model selection has more information about the difficulty in model selection and thus the uncertainty is more realistic.
We also considered the forward modeling uncertainty (i.e. model discrepancy) separately in order to take into account the imperfect forward modeling when fitting the LUT-based reflectance into observations.

'the posterior probability distribution can characterise the uncertainty more extensively than commonly given standard deviation'. Fair enough, but what does this mean? Is it better?
The posterior distributions of the best models and averaged posterior give more information about the uncertainty in model selection and in estimated AOD than one number and standard deviation can give. But presenting this pixel-wise uncertainty information, given by the posterior densities, in compact but still informative form is not clear. In addition, the other question could be what information of the uncertainty is needed or is sufficient to present.

We have now removed this sentence since it was unclear statement (Discussion and Conclusions, p12l15-15) and added the following expression:.
"Moreover, further study and discussion is needed to determine how to express the uncertainty information, provided by the posterior distribution, in more compact form.

Should we generally apply this method? Also from the provided sensitivity studies it is just not clear whether things work as expected
The method brings more information about the uncertainty and it can be used for evaluation of the model selection process e.g. study the influence of aerosol microphysical model selection on the estimated AOD.
So far, the sensitivity studies have brought interesting information about the uncertainty related to the model selection process, e.g. difficulty in model selection or lack of appropriate model LUTs. However, the case studies have also shown that the aerosol type selection works as expected, albeit with some exceptions e.g. resulted unexpected type of model.

Does the average AOD perform better than the standard one, when compared to AERONET? If not, is this reflected in a larger uncertainty? If yes, are the AERONET and OMI AOD retrievals consistent within this new the uncertainty?
The case studies reveal that the proposed method using averaged AOD was not better than the standard one (OMAERO) if compared only the retrieved AOD values to AERONET. But the proposed method got solution for more pixels than OMAERO. Also the retrieved, but LUT-dependent, Ångström exponents were in rather good agreement with the AERONET values.
The test cases also show that, in general, the larger uncertainty, i.e. posterior width, reflected the uncertainty in the retrieval. Also, when the deviation from the AERONET AOD was larger then the uncertainty was higher.

Minor comments:

p1l24 (and a few more): data is -> data are
Corrected

p3l25: referred -> referred to
Corrected.

p3l28: a cloudy ground pixel sounds strange. I would say a cloudy atmosphere pixel. Or just a cloudy pixel.
We changed "a cloudy ground pixel" to "a cloudy pixel".

p3l31: What is a wise quality?
This is a typo; "wise" removed

p4l4: Before the start of the new sentence, add 'For surface reflectivity,' (we used.. etc)
Added, thank you.

p4l11-13: you say: the band at 477 nm adds important info, yet you exclude it specifically. Why?
We excluded band at 477 nm since we did not need aerosol layer height information in our study. The other reason is based on experimental issue since we found that this band brought extra complexity when examined the modeled spectral reflectance fit to the observed reflectance.

We changed the order of the last two sentences in the revised manuscript and rephrased the sentence as:
"However, we omitted in our study the band 477 nm due to experimental purpose and since we did not need aerosol height information."

p5l3: Equation (1) is not just a 'formula'. Start this discussion with a physical description like: Assuming a Lambertian surface the contribution of the radiation at the TOA can be separated from that of the atmosphere (e,g. Chandrasekhar, 1960), viz. etc.
Thank you, we have rephrased the sentence as suggested. The sentence now reads:
"Assuming a Lambertian surface the contribution of the radiation at the TOA can be separated from that of the atmosphere (e.g. Chandrasekhar, 1960) leading to the equation for modeled reflectance as …"

p5l10: of the real -> of the aerosols in the real
Corrected.

p5l10-11: This forward model app error,.. Which one? You haven't described an error yet. Do you mean the difference between real and approx. reflectances? Then describe that.
By "forward model approximation error" we mean error that originates from forward model approximation. The beginning of the sentence has been revised to "Approximations in forward modeling …"

p5l13: This is strange: I would expect that a total (megs) error would be forward model error, noise (and perhaps more). Noise surely doesn't include forward model error? What is epsilon_obs? Noise or total? Rephrase l11.
Thank you for notifying this incoherent statement; "Measurement noise" is a wrong expression.
We have now removed this sentence since it is unnecessary here.

ε_{obs} is the measurement error (or noise) and $\varepsilon_{obs}(\lambda) \sim N(0, \sigma^2_{obs}(\lambda))$. We have clarified this in the revised manuscript.

To make clear, we have also expressed the measurement error standard deviation $\sigma_{obs}(\lambda) = R_{obs}(\lambda)/SNR$ (in p5l20) and changed the notation "$\sigma(\lambda)$" to "$\sigma_{obs}(\lambda)$" (in Eq. 3)

p6. Increase the size of eq. 4 and 5, like eq. 1. They are the basis of the paper.

Done.

p7l19: cover -> covers
Corrected.

p7l20: cover -> covers a
Corrected.

p7l22-25: Move this to section 2. And add a description of MODIS, which is introduced in the next paragraph.
The text part p7l22-25 introduces the AERONET data that are used for evaluating the case studies. That's why we would like to retain the AERONET description part in this Section 4 (Case studies and results).

As suggested, we have now added a description of MODIS in the end of Section 4 (p7). The added text reads:
"We tracked clouds and land scene for the case studies by utilizing true-color images from MODIS, on Aqua satellite, that has the equator crossing time only about 15 minutes earlier than OMI. The MODIS instrument is onboard both Terra and Aqua spacecraft. The data products derived from MODIS measurements include atmosphere (e.g. cloud mask and aerosol products), land, cryosphere and ocean products (see e.g. http://modis.gsfc.nasa.gov)."

p8l1 & Figure 1. The OMI pixels -> The OMI pixels that were analysed The location of the OMI pixels within the MODIS swath are not clear. In Figure 1 add the contours of the OMI pixels that are used in Fig 2-6.
P8l1 Corrected: "The OMI pixels" -> "The OMI pixels that were analysed".
Fig 1.: We added contours of the OMI pixels.
We also added in the figure caption: "The area of analysed OMI pixels is marked with red contours."

p8l2: The pixel has no data if -> No data are reported if the pixel is
Corrected

p8l18: pixel wise -> pixel-wise
Corrected.

p823: in the latter day case -> On the 27th,
Corrected.

p9l8-10: Figure 8 is superfluous. Remove it and on describe the results from it in the text. It will reduce the number of figures, which is needed anyway.
We agree. We removed Fig 8., rephrased the text accordingly and moved it to the beginning of the paragraph. The sentence reads now:
"The other types of models, e.g. weakly absorbing type, do not match as well as the selected best BB models."

p9l11-13;: Elaborate on this result. It is as important as the 16th.

As suggested, we have now included more discussion in the revised manuscript about the results of 27th case.

p9l14-22: Here's the first missing conclusion. So you compared the Angstrom exponents. Whats the conclusion from all this? Does it improve as expected or not. Describe this, instead of just showing numbers in a table. The table is just there to backup the story.

The conclusion from comparison with AERONET values is that the derived, even if LUT dependent, Ångström exponent values are in rather good agreement with the AERONET values (see Table 2). But in some cases we observed that agreement between the AOD values do not necessarily lead to agreement between the Ångström exponent values.

Unfortunately we cannot answer to the question "does it improve as expected or not" since we have only done the comparison between LUT-based derived α1(442-500 nm) (or α2) and AERONET α(440-675 nm).

We have now included the following description in the end of section 4.1 in the revised version:

"For Beijing case, in both days, the derived Ångström exponent value of the best model (α1) is in good agreement with the AERONET value. Even so, on the 16th of April α2 deviates more from the AERONET value although the estimated AOD, based on the second best model, is closer to the AERONET AOD values (see Fig. 7 left)."

For Africa case, the Figure 14 shows distribution of α1 (left) and α2 (right) values, respectively. We added the following description in the revised manuscript (p10):

"That is, the Ångström values are low where the desert dust type of models dominate. Correspondingly, in the coastal region where typically is smoke and urban polluted air the Ångström exponent is higher."

When we compare Ångström exponent values at locations of AERONET sites in Africa the agreement is generally good, except at DMN_Maine_Soroa site.

We added the following sentences related to Ångström exponent comparison (p10) for Agoufou:

"However, the derived Ångström exponent α1 has rather good agreement with the AERONET value (Table 2)."

and for DMN_Maine_Soroa:

"… but the derived Ångström exponents, α1 and α2, do not agree with the AERONET value (Table 2)."

and for IER_Cinzana and Saada:

"For the sites IER_Cinzana and Saada the best and the second best models have as good evidence (Fig. 15 right column) indicating that the selection of the best model happened by chance. Consequently, the derived α1 for Saada site is consistent with the AERONET value whereas the derived α2 for IER_Cinzana has better agreement than α1 with the AERONET value."

p9l24 & Figure 9: This figure is inadequate. Again the location of the OMI pixels is not clear. Merge MODIS quicklookd into one RGB image and overlay the OMI pixel contours.
We have now merger the two MODIS RGB images and added the contours of the OMI pixels. We also added in the figure caption: "The area of analysed OMI pixels is marked with red contours."

p10l3: the selection of the volcanic type is most probably.. : Most probably? Who is going to give a conclusive answer to that if not the authors themselves? First, indicate where the OMI pixels are in the MODIS RGB image as suggested above. Then, conclude whether or not this is due to the 'white area'. . . Do you mean cloud?
The contours of the OMI pixels added in the MODIS RGB image as suggested above.

We have now changed the notation "white area" as "cloud" and rephrased the text as:
"The selection of volcanic aerosol type as the only appropriate aerosol type happens for pixels located northeast from the Lake Chad where is seen cloud in the MODIS RGB image (Fig. 8)."

p10l10: perhaps indicating..: Again, why perhaps? Tell the reader whether there was dust or not. If not, why select this day? Surely a dust event can be easily found using OMI UVAI on a clear day. Indeed, 26 March 2008 shows low UVAI over the northern Sahara, so change this day and choose a day where you know what's going on and what aerosol model you should expect.
The criterion for selecting that date, 26 March, is almost cloud free scene over Northern and Central Africa thus providing large pixel area to study. The aim was to study the uncertainty in aerosol model selection and its effect on the results in "the normal aerosol situation" and we did not seek a special case with known dust or smoke event. The resulted aerosol types were what we expected i.e. dust in the north and urban pollution/smoke in the coast region.

We have now removed this imprecise statement (p10l10):
"The retrieved AOD estimates are rather small perhaps indicating that no dust event or active fires were going on."
in the manuscript since it is not relevant here.

p10l24-26. So what's the conclusion here? Is the posterior uncertainty better or the same in the case of one chosen model? Does the (new) high uncertainty include the difference between the two measurements, or is it too small?
The paragraph p10l24-26 considers results in one OMI pixel located around AERONET Agoufou site.

As a result there is only one selected model having a sufficient, even poor, fit to the measured reflectance. As expected, the large width of the posterior distribution, that is the averaged posterior distribution as well, indicates high

uncertainty in the model selection and thus in the retrieved AOD. Consequently, the answer to the first question is: even the method gives a solution that passed the goodness-of-fit test it does not ensure correctness of the result.
The answer to the second questions is: the posterior uncertainty is the same in case of one chosen model.

The retrieval uncertainty is high and still the posterior density does not cover the AERONET Agoufou AOD values (or daily average) but it covers the OMAERO AOD (1.557). However, the Ångström exponent values for AERONET daily average α(440-675 nm) and proposed method α1(442-500nm), i.e. 0.375 and 0.293 respectively, match quite well (Table 2).
It must be noted here that the AERONET measurements at the Agoufou site were made in the morning and the last one about 3.5 hours before OMI overpass time.

We have now added more discussion and rephrased the paragraph p10l24-28 in the revised version.

The conclusion section should be extended with a clear recommendation.
We have included more text for recommendation, and hopefully in clear way, in the Discussion and Conclusions section.