We thank the reviewer for the thorough review and constructive comments. Several interesting and significant changes/additions were made to the manuscript in accordance with the reviewers' suggestions. Consequently, this revision makes this paper differ from the past work in that:

(i) MODIS AOD retrieved from both Land and Ocean algorithm are assessed separately over the same set of AEROENT stations along the coast, and it is shown that AOD retrievals over the coastal region from Land or Ocean algorithm both have larger uncertainties than their counterparts in other (e.g., over open ocean or over non-coastal land) regions;

(ii) the assessment includes characterization of the PDF (mean and standard deviation) of biases and its statistical fitness with PDF from AERONET;

(iii) the impact of sediments (pigments and suspended particulate matter) on the AOD bias are analyzed, and is shown to be important;

(iv) further evaluation of correction for wind speed and cloud impact is conducted, and the implication of such correction for the trend analysis is studied with analysis over different AERONET satiations.

Since the manuscript has gone major revision, it is recommended that please first read the revised manuscript as a whole (attached through separate entry in the interactive discussion) and then read our replies.

Response to reviewer #1:

- **Comment:** "I recommend the manuscript undergoes major revisions and then a second round of peer-review. In brief, I recommend three main strands for revision, justified in the specific comments. Firstly, due to limitations of the current MAPSS software, it would be much better if the analysis were done with the "central point" rather than "box average" method. Secondly, the authors could request MODIS Collection 6 data from the MODIS science team (or wait for general release), and repeat the analysis with that, which would be worthwhile. This will cause a delay but the quality of science should be most important factor in these decisions. Additionally, using Collection 6 would doubtless increase the readership of the paper. Thirdly, a lot of the repeated work from previous studies (quality flags, wind speed) could be omitted or shortened, and then some new section showing a scientific application of the filtered/bias-corrected data could be added. The paper as it stands now is not too long, and that would add something more original to the study."

  Response: Firstly, the MAPSS use of "central point" vs. "box average" method was evaluated in "Petrenko, M., Ichoku, C., and Leptoukh, G.: Multi-sensor Aerosol Products Sampling System (MAPSS), Atmos. Meas. Tech., 5, 913–926, doi:10.5194/amt-5-913-2012, 2012". We justify our use of the "box average" because that study found differences in the MODIS AOD between the two methods to be small. Secondly, while we agree with the reviewer that use of Collection 6 can increase the readership of the paper, we also want to emphasize that the data we analyzed here have a time span of nearly 10 years and is global in scope. This analysis

of long-term data is needed to study how well the MODIS AOD product captures the frequency distribution of the AEROENT's counterparts along the coastal region. An analysis of 1-2 years of data (from collection 6) would not fulfill this purpose. We're using collection 6 (in collaboration with R. Levy) for air quality purposes; but a 10-year reprocessing to Collection 6 has not been completed at the time of this writing (November 2012), not to mention that they are not in the MAPSS database. Thirdly, in the updated manuscript, the analyses of quality flags and wind speed have been shortened, but two new sections have been introduced that respectively examines the effects of sediments on the retrievals, and the impact of empirical correction to the trend analysis of annual AOD. The impact of the pigment and suspended matter in the coastal water on the MODIS AOD bias is clearly demonstrated.

- **Comment:** "Abstract: This is quite long and some of the text is not needed for an abstract. The text may change based on the revision of this paper, but my suggestion based on the current version would be: "Coastal regions around the globe are a major source for anthropogenic aerosols in the atmosphere, but the underlying surface characteristics are not favorable for the Moderate Resolution Imaging Spectroradiometer (MODIS) algorithms designed for retrieval of aerosols over dark land or open-ocean surfaces. Using data collected from 62 coastal stations worldwide from the Aerosol Robotic Network (AERONET) from 2002–2010, uncertainty assessments are made for coastal aerosol optical depth (AOD) retrieved from MODIS aboard the Aqua satellite, from the Collection 5 dataset. It is found that coastal AODs (at 550 nm) characterized respectively by the 'Dark Target' land algorithm, ocean algorithm, and AERONET all exhibit a log-normal distribution. After filtering by quality flags, the MODIS AOD is highly correlated with AERONET (with R2=0.8), but only fall within the expected error envelope greater than 66% of the time for the land algorithm. Furthermore, the MODIS AODs show statistically significant discrepancies from their respective counterparts from AERONET in terms of mean, probability density function, and cumulative density function. Without filtering with quality flag, the MODIS land and ocean AOD dataset can be degraded by 30–50% in terms of mean bias. Overall, the MODIS ocean algorithm overestimates the AERONET coastal AOD by 0.021 for AOD < 0.25 and underestimates it by 0.029 for AOD > 0.25. This dichotomy is shown to be related to the ocean surface wind speed and cloud contamination effects on the satellite aerosol retrieval. The Modern Era Retrospective-Analysis for Research and Applications (MERRA) reveals that wind speeds over the global coastal region (with a mean and median value of 2.94ms^{-1} and 2.66ms^{-1}, respectively) are often slower than 6ms^{-1} assumed in the MODIS Ocean algorithm. An empirical scheme for correcting the bias of AOD retrieved from the MODIS Ocean algorithm is formulated and is shown to be effective over the majority of the coastal AERONET stations." "

  Response: The abstract has been nearly re-written to reflect the new content in the manuscript (such as effect of sediment and trend analysis). Manychanges recommended by the reviewer have been completed.

- **Comment:** "Page 5207, lines 17-26: You may consider the paper by Hsu et al (ACPD, 2012), which looks at AOD trends from SeaWiFS both over land (as you note, previous studies were ocean only) and ocean: Hsu, N. C., Gautam, R., Sayer, A. M., Bettenhausen, C., Li, C., Jeong, M. J., Tsay, S.-C., and Holben, B. N.: Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010, Atmos. Chem. Phys. Discuss., 12, 8465-8501, doi:10.5194/acpd-12-8465-2012, 2012."

  Response: The results from Hsu et al. have been examined and included in the text.

- **Comment:** "Page 5208, line 10: I suggest "uncertainty" rather than "accuracy" here, as this is the uncertainty confidence envelope, while "accuracy" (and "precision") have specific technical meanings which are different. This should be checked throughout the manuscript."

  Response: "Uncertainty" is now used throughout the manuscript to avoid any confusion.

- **Comment:** "Page 5208, lines 12-14: This discussion in the context of the MODIS uncertainty is not directly relevant and misleading. The uncertainty envelope is defined only for an instantaneous MODIS retrieval, and cannot be quoted in the context of trend analysis. Unless you know the systematic and random components of the uncertainty, and how they change over time, you cannot propagate the instantaneous uncertainty into the trend. And these things are not known quantities. For example, if there were a bias of 0.1 in MODIS AOD all the time but no other source of error you could still use it for trend analysis because the uncertainty has no time dependence, even though the total uncertainty might be larger than the trend. Additionally, for trend analysis, there are a whole other set of factors, such as temporal compositing strategy, spatial/temporal sampling, etc, which play a role and determine how easily any trend can be identified. This sentence must be reworded."

  Response: That paragraph is now written. Please see manuscript for details.

- **Comment:** "Page 5208, line 29: the official name is "Dark Target" rather than "Dark Land"; it would be preferable for the authors to use this, to minimize confusion, but not a critical issue so long as they are self-consistent."

  Response: Since both ocean and dark land are in the category of dark target, to better separate the algorithm difference, we use the dark land to refer land algorithm.

- **Comment:** "Page 5209, lines 22-25: This is another example of an incorrect statement which is misleading and must be corrected. There is no loss of quality information in the combined dataset, precisely because it is a union of the land and ocean datasets, the same quality flags apply as for the separate datasets!"

  Response: The text has been modified to more clearly state that no quality flags are reported in the Land_And_Ocean product and in order to use a quality flags for processing the individual Land and Ocean products must also be used. The more descriptive wording should alleviate any ambiguity to the meaning.

- **Comment:** "Page 5210: The discussion on wind speed is a bit long, particularly as, as you note, it has already been analysed and known about in multiple studies for several years now. If you wish to include this analysis in your paper, it would be sufficient just to cite those studies and say you will examine whether the same results hold over coastal areas as over the ocean as a whole. As mentioned previously, doing the analysis with Collection 6 data and seeing whether the wind-speed dependence has been removed successfully would be more useful."

  Response: The wind speed analysis has been extensively shortened and where appropriate more focus is given to the difference between the coastal and non-coastal impacts. We also add a note in the revision that effect of wind speed on surface reflectance is added in algorithm for Collection 6, although further evaluation is needed to study if the wind-speed dependence has been removed successfully.

- **Comment:** "Section 2.3: The MODIS ocean product provides two AOD datasets: from the best-fitting aerosol model, and from the average of several well-fitting aerosol models. Which is used here? This should be stated."

  Response: The average of several well-fitting aerosols models is used here. The information has been added to the manuscript in Section 2.1.

- **Comment:** "Page 5211, line 24 and page 5212, top: The acronym 'LUTs' should be defined as first use. I think this paragraph can probably be deleted and the algorithm papers just referenced. You are only describing the land algorithm here, not the ocean algorithm (although you don't state this in the text), and the information you give here is not used later in your discussion."

  Response: That paragraph is revised, and 'LUT' is not used in the manuscript anymore.

- **Comment:** "Page 5212, lines 21-24: This sentence is hard to follow and be reworded. Just say that using quality flag 3 over land and 1, 2, or 3 over ocean should give

agreement within EE 66% of the time on a global basis. The way it is written makes it sound as if each retrieval is validated and then assigned a quality flag, which is obviously not the case, and could be misleading."

Response: The sentence has been revised.

- **Comment:** "Page 5214, lines 1-7: Are you just using all AERONET sites? There are some (e.g. Mauna Loa, Izana) which you should exclude from your analysis, as they are e.g. high altitude sites which are not representative of their larger region on MODIS retrieval spatial scales, so not a useful validation site for MODIS data. Using everything blindly will make it more difficult to draw meaningful conclusions as discrepancies will arise for reasons other than retrieval error. Some years ago Stefan Kinne (MPI-Hamburg) was compiling a list of AERONET sites he thought were representative of their larger-scale environment. I don't know whether that list was published as part of any study but it could be worth asking him. Certainly you should consider only a subset of the available AERONET sites."

  Response: Yes, all AERONET sites were included for the study. In light of the comment, an analysis excluding high altitude sites such as Mauna Loa, and Izana was completed during revisions. By excluding those sites correlations between the different MODIS algorithms and AERONET changed by less than 2%, the wind speed correction was also negligibly changed. For those reasons we keep our analysis that was conducted prior to revisions.

- **Comment:** "Page 5214, lines 15-19: you say there is "little difference" between the two MAPSS analysis methods yet elsewhere you state as "significant" differences which are of order 0.01-0.02, i.e. within or near the AERONET uncertainty. You should quantify here exactly how different the two methods are. Also, as you note, the MAPSS "average" method does not unambiguously look at the effect of quality flags because it takes the mode (rather than subsetting for each quality flag). This appears to be a limitation of MAPSS. If you are really interested in looking at the effect of quality flags, as you do later, this implies you should really use the "central" rather than "average" method. I am sure data volume would be sufficient."

  Response: Firstly, the manuscript has been updated with some quantitative results from Petrenko et al. 2012 that reveal the similarity between the "mean" and "central" methods. However, Petrenko et al. 2012 is the main resource for that particular analysis. Secondly, it is necessary to evaluate a quality control for the "mean" method because of its wide use in the scientific community. The results of the QA filter described in the manuscript are targeted at the scientists in the research community that utilize the "mean" method. Thirdly, as in your previous comment, if MODIS AOD is meant to be representative to the large environment, a mean method perhaps is better than central method. Along the coast, there can also many cases where the characteristic of aerosol plume around AERONET site may change within 1 hour, and

so using spatial average to compare with temporal average have been the strategy for conducting the AOD evaluation in the past, and is also adopted here.

- **Comment:** "Page 5215, lines 8-11: I mention this issue here but it applies at several later points too. I know ordinary least squares (OLS) linear regression is popular in our community. But it is really the wrong thing to do here, and your fits will be skewed. This is a well-known issue presented in statistical textbooks. Being a popular technique does not mean it is a good technique to use. The assumptions for the regression you have done are that the relationship is linear (maybe ok, although Figure 3 shows nonlinearities) and that the noise about the linear relationship is Gaussian and the same size across the range of the independent variable (here, AERONET AOD). As you note several times already in the manuscript, the MODIS uncertainty has a dependence on AOD, so you have already stated in your manuscript one reason why OLS regression is not an appropriate technique (i.e. scatter at low AOD and at high AOD are different)! In addition, the uncertainties at low AOD are more likely to be biased high than low, because negative AOD retrieval is not permitted over ocean and you say later you throw negative points out over land. This will result in the linear fits being biased to overestimate at low AOD and underestimate at high AOD–exactly as observed in parts of this study and others. Sampling is also extremely non-uniform along the AOD axis, because AOD distributions vary (as you note) approximately lognormally. You should not use OLS regression but instead another technique (there are several options) which is statistically appropriate. Just because others have published with such erroneous techniques does not mean you should fall into the same trap. In fact doing it properly would hopefully serve as an inspiration to others in the future."

  Response: It is not the aim of our analysis to promote the regression results. Rather, we like to promote the other statistical methods to evaluate the MODIS retrievals. Throughout the updated manuscript, where appropriate, we have conducted analysis about the PDF, CDF, and bias analyses. The OLS regressions shown in this manuscript is presented for the purpose of comparing our results to those found previously.

  The uncertainty envelope of MODIS AOD is meant to describe the likely range of the uncertainty, but it doesn't necessarily describe the actual uncertainty characteristics associated with each individual retrieval. While we agree that the retrieval uncertainty depends on AOD value itself, it appears that the uncertainty is nonlinear with respect to the MODIS AOD, which contrast with what the uncertainty envelope equation would suggest. Indeed, in the revision we do find that the distribution of AOD bias is Gaussian, regardless of high AOD or low AOD.

- **Comment:** "Page 5215, lines 11-15: It would be interesting to see the distributions of the bias. The mean is a handy statistic but if you have outliers, skewedness and so on, other quantities such as median and standard deviation/interquartile range can be more informative. I suggest showing these distributions in the revised manuscript. The

distribution of error, and/or the distribution of error relative to the expected error (which one would hope is Gaussian but could well not be) would be good to see."

Response: The bias distributions (frequencies) have been added for each of the MODIS products (Land, Ocean, Land_And_Ocean, and out corrected product). The results of this extra analysis have been added into the revised manuscript, and yes, it is nearly Gaussian, as you hoped. Again, our intention is not to just focus on the mean, but to look at the PDF, and CDF, and that should address the issues related to skewedness and extreme cases. It is noted that the empirical correction not only reduces the bias, but also reduce the standard deviation or spread in the biases, which is interesting.

- **Comment:** "Page 5216, lines 6-8: This is not a good thing to do. If you are throwing away negative retrievals, which you know are low-biased, you will therefore skew your analysis towards reporting a more positive bias than is really the case. I understand that you can't do a lognormal fit with negative values, so, ok, you can throw out those 400 points for this, but they should be included for the rest of the analysis of the paper. Also, what proportion of your total sample is 400 points? Sampling is given in Table 2, but mentions the relevant number in the text here too."

  Response: The total number of retrievals is now also provided in the manuscript, 400 points is less than 1% of the total retrievals. It is also more clearly explained that the negative retrievals are only removed for the log-normal fit, not for the other analyses. Finally, the retrieval of the negative AOD is not physical. It is not a good idea to keep what is non-physical for the sake of a good statistic result.

- **Comment:** "Page 5218, lines 7-23: See comments about OLS previously; an alternative method should be used. These R2 and regression fits would be much clearer, and the paragraph more readable, if presented in a table."

  Response: Table 3 has been added to the revised manuscript that includes the regression and correlation information. See our previous response regarding OLS.

- **Comment:** "Page 5219, lines 11-12: This sentence is jarring. You are basically saying you use 0.25 as a threshold because another paper used 0.2. I don't think you need to justify a choice of 0.25 as something in that region fits with common sense. So, I'd either delete the mention of Levy et al (2010) or else adopt their 0.2 threshold (which I expect won't change things much)."

  Response: This statement has been removed.

- **Comment:** "Pages 5519-5521: here you commence looking at differences and calculating statistical significance. But remember that AERONET AOD has an

uncertainty of 0.01-0.02, which is of a similar size to the mean differences you are claiming as "significant". In this light, can you say it is scientifically significant? On page 5220-5221 you say “MODIS is not very accurate in modeling the actual nature as represented by AERONET” and that “MODIS does not model the actual nature represented by the AEROENT AOD observations”. Aside from the previous point of whether “accurate” is the correct word here, this conclusion is not necessarily warranted from a scientific point of view. The relevant question is whether the differences are important for a given application (as mentioned in the General Comments, I am not sure what the end goal of the authors is from this study). Are we interested in typical values or extremes? It might be that one is consistent between the two datasets and the other not. The CDFs in Figure 6 look very similar, with much of the difference contributed by points off the left-hand side of the plot (also see later comment on that figure)–and as you say in the text you are throwing some negative values out over land. Would just removing the average wind-bias offset from the MODIS ocean AOD make the two statistically consistent? It looks like it might help, certainly. I suspect for many applications these differences are not important. Be careful not to confuse statistical and scientific significance. You can detect two things are statistically different (e.g. the distribution of sizes of grapes in two bags) but it does not mean it is important (both may provide a tasty snack). The discussion should be extended and these aspects discussed. Remember also that you are not looking at the MODIS or AERONET data directly here. You have a spatial average of MODIS and a temporal average of AERONET, and averaging is going to change the shape of the PDFs/CDFs, dependent both on the noise in each dataset and also on the spatial/temporal variability of each dataset. You don’t really mention this aspect, although it is discussed in several other satellite AOD validation studies. So this analysis and conclusion is misleading. This is another reason it would be much better to use the “central” MAPSS method: you can look at quality flags directly, and it removes the effect of spatial/temporal averaging from the comparison of PDFs/CDFs. You claim that in your analysis you are going beyond previous studies. As I said in the general comments, there is a nice study buried in here, but to get to it you are going to need to redo the calculations and add to the discussion, particularly in this section. Otherwise it is not really adding anything useful over previous studies.”

Response: One of the main questions we aim to answer after this study is whether there is a trend in extreme aerosol events, and the impact the extreme events may have on climate studies. For that purpose, the statistical tests (and findings) must include the distribution of the AOD retrievals instead of just bias. The “mean” vs. “central” comparison was completed in Petrenko et al. 2012 and their finding was that the “mean” and “central” method provided similar results. The overall discussion in these sections has been updated to incorporate the goal of evaluating the AOD trend and climatic implications. In addition, while AERONET AOD has an uncertainty of 0.01-0.02, we don’t have any evidence that this uncertainty is systematically positive and negative, and this uncertainly is for instantaneous AODs. However, the frequency of MODIS AOD bias, while Gaussian, is not centered at zero, and the bias overall is systematic. This can be important for the climate studies. Likewise, CDF and PDF distributions are useful for looking at overall change of aerosol climatology, and it

should be evaluated against AERONET. At this point, we can only say from a statistical point of view the difference and fit between the two products. Its scientific significance, as you pointed out, depends on applications, and should be evaluated by the science community.

In the summary section of the manuscript, we stated that "It should be noted that while our analysis of retrieval error sources is based upon the physical reasoning and supported by the statistical results, the statistical significance is mainly evaluated from a mathematical point of view. Implication of these statistical results to the applications of AOD for climate studies or air quality monitoring should be interpreted with caution because each application has its own requirement of the data accuracy and own tolerance of uncertainty."

- **Comment:** "Section 4 (page 5221, lines 19-24): This is the same misleading statement the authors make earlier and must be corrected. Again, the LandAndOcean quality flag is the same as that for the Land dataset for pixels over land, and the same as that for the Ocean dataset for pixels over ocean. I question the value of this section and think it can be removed or summarized in a single sentence, unless you repeat the analysis with Collection 6 data, or do something else which is really new. The MODIS team recommends applying the quality flags, precisely because if you don't the data are of lower quality. This has been shown in previous studies, and doesn't need to be rehashed here. Also, the point about OLS regression applies again here."

  Response: An updated description has been provided for the Land_And_Ocean data sets, with reference to the MODIS level 2 atbd. Two different Land_And_Ocean data sets are available in collection 5.1, "Image_Optical_Depth_Land_And_Ocean" which has no QAC threshold, and "Optical_Depth_Land_And_Ocean" which requires quality flags > 0 over land, and ≥ 0 over ocean (MODIS level 2 atbd). However, unlike the Land or Ocean datasets, the combination Land_And_Ocean does not retain (or report) the quality flag information.

- **Comment:** "Page 5223: Yes, Zhang and Reid (2006) was about the Collection 4 product. However, Shi et al (2011), which you cite elsewhere, is a similar analysis for the Collection 5 product (reexamining clouds, wind etc). So, this has already been extensively investigated on a global basis (and with similar results to the coastal analysis) to create their group's data assimilation MODIS products, which include this additional cloud filtering. This section is retreading old ground. This is another aspect which would be much more useful with Collection 6 data, for which the analysis has not yet been done, and cloud contamination issues should hopefully be smaller."

  Response: This section of the manuscript has been shortened and revised. New discussions have been added to the manuscript that shed light on the differences between the trend from the standard MODIS product and the corrected product. Performing the analyses on Collection 6 data will be the goal of future research.

- **Comment:** "Page 5224-5226: This is retreading old ground from the Zhang/Reid group papers, and does not provide new insights. I suggest removing it, or waiting and repeating the analysis with Collection 6 data."

  Response: Similar to the previous response, this section has been shortened and revised, and new discussions provided. We also studied the impact of empirical correction on the bias PDF and AOD trend analysis, although the study on trend analysis here can be further detailed in the future analysis.

- **Comment:** "Page 5227, lines 1-8: If 46 out of 62 coastal sites have a significant relationship, the more interesting question is: what about the 16 which don't? Which are they, and how/why are they different? Perhaps the dominant aerosol source is not marine at those 16? I suggest discussing these sites in more detail, which may provide new insights concerning these regions which could be useful for other studies. They are shown in Figure 7 but there is no interpretation offered, and there should be."

  Response: The 16 AERONET sites are discussed in the revised manuscript. An brief discussion about the characteristics of these sites is given and a possible source of uncertainty is introduced. "For those 16 AERONET sites that do not show a statistical significant correlation between MODIS bias and wind speed are, they have two main characteristics in common: 1) the MODIS AOD correlation with AERONET AOD is less than the average correlation for the coastal group; 2) all of the AERONET sites are close to the coastline (i.e. within 5 km) except Bac_Lieu which is ~8.5 km from the coastline. (1) suggests that the retrieval errors at these sites are not systematic, and (2) indicate that the rough ocean surface model may not be appropriate to estimate the surface reflectance at the first place, which is supported by the analysis in the following section."

- **Comment:** "Page 5227, lines 11-13: You mention "complex surface characteristics". This is true for e.g. the 1 km pixels which straddle the land/ocean boundary, or even the 10 km retrievals which include water, beach, vegetation, urban areas, etc. But since you are averaging a 55 km box, much of these areas will include the same types of terrain that you see in non-coastal areas. So I don't think the statement you make in this sentence is a fair assessment of what is tested in this work. Again, using the "central" MAPSS method rather than "average" would be an improvement because you specifically would be looking at these complicated heterogeneous regions."

  Response: We now add the analysis of sediment in the revised manuscript. Note, this same type of layout is with respect to AERONET, not in the eye of MODIS. Depending on the view geometry, the actual surface location that are included in the retrieval can very different, especially in terms of the fraction of water and land, even

in 10X10 km box. The 'complex' is meant for MODIS retrieval. Over open ocean or inland, the 10X10 km box can be just 100% of vegetation or sea water.

- **Comment:** "Pages 5227-5228: The rest of the conclusion is somewhat brief. In the end you state that bias correction improves the agreement between MODIS and AERONET and should be done for trend analysis and data assimilation. Well, this is what the Zhang/Reid et al analyses have shown for some years, which you cite, and has been used to produce their data assimilation MODIS dataset. So, what is new from this study? Again, using Collection 6 data, and including some further analysis where you make use of your bias-corrected dataset, would make this much more interesting and useful."

  Response: We have done more analysis to illustrate the implication of this study for the trend analysis. In addition, we also show the impact of sediment in the residual bias, as well as the evaluation of MODIS land AOD along the coastal, and the fitness of CDFs and PDFs. So, the revised manuscript have several parts that are not done in Zhang/Reid.

- **Comment:** "Tables 1, 2: Mean bias, as mentioned earlier in the review, is not a very useful metric by itself. What about the median bias, standard deviation of bias, etc? Such other quantities would also be worth putting in the tables."

  Response: A new figure has been added to the revised manuscript that displays the bias frequency and includes the standard deviation of bias. It is found that, along with the mean, the standard deviation of the bias is reduced after the empirical correction. We looked at the median bias, but found that the results of mean bias are similar as median bias.

- **Comment:** "Figure 1: I think there is an inconsistency between figure labels and text here (e.g. B is labeled LandAndOcean but the fourth, not second, plot listed in the caption). I would suggest it would be more informative to include the numbers from this plot (plus median AOD, as well as mean) in a table, and instead plot these relative frequency distributions on top of each other with lines. That would enable a more direct comparison of exactly where in AOD space the distributions are different. Also, these bins are somewhat coarse: I suggest narrowing them to see whether there is any finer-scale structure."

  Response: You are correct, the figure panels are now properly arranged. The bin width was chosen based on the formula from Wilks 2011, $h = \frac{c \cdot IQR}{n^{1/3}}$, where h is bin width, n is the number of samples, IQR is the interquartile range, and c is a constant assumed to between 2.0 and 2.6. c is chosen to be 2.2 for our distributions.

- **Comment:** "Figure 2: How do these results change if you use finer bins? Since you do not have a large number of bins, it is easier to find a statistical agreement between observed and theoretical lognormal distributions. It would be more convincing if the bin size was smaller and you still found statistical significance."

  Response: See previous response for bin width selection.

- **Comment:** "Figure 3: These would be clearer if you plotted scatter density plots, e.g. Figure 1 of the Levy et al (2010) MODIS validation paper. That lets you see where the bulk of the data actually are, rather than a cloud of points. Again, the OLS linear regression should be replaced by a more appropriate technique. Look at any of the top panels, especially the top-right. It is obvious that the OLS linear fit is not a good model for what is happening at high AOD."

  Response: See our previous comment about OLS regression. It is debatable if the points with high AODs should be given more weight or less weight in the regression analysis. So, we look into MODIS AOD CDF and PDFs, and evaluate their fit with AERONET. The number density overlay can sometimes mask the actual data points, and the CDF and PDF can shows clearly how good the fit in different data range is.

- **Comment:** "Figures 6, 9D: I have some concerns with these. First, if I understand correctly, these show the lognormal fit distributions of the data, rather than the actual data themselves. Given the departures from the lognormal fits (very evident in e.g. Fig 2D), I think you are introducing a non-negligible uncertainty by using the approximate fit distributions, especially given your criterion for what is a "significant" difference is very small. So, this is somewhat misleading and could be affecting your conclusions. It would be much better to use the actual distributions. I understand these are not continuous distributions, but you could plot the CDF of the binned data here, which would be good enough (particularly if you take my suggestion of narrower bins). That would be a much fairer treatment. Or, you could do this type of analysis using QQ plots instead of the CDF. As it stands, I worry that comparing the CDF of distributions which your data only approximately match, as opposed to CDFs of your actual data, is just making things difficult. Secondly, the inference from these figures is that about 15% of the time the AOD is less than 0.01 in all datasets. This is surprising. I would check your code. Your Figure 2 does not support this (almost nothing below 0.018), and I also looked at AERONET data for a few sites (Lanai, COVE) and found almost no points with AOD at 550 nm < 0.01. I note your Figure 2 is natural logarithm, while Figures 6 and 9D appear to be base-10 logarithm. Perhaps somewhere in your calculations one was incorrectly used, introducing the error."

  Response: The figures no longer display the lognormal fit of the distributions. The actual distributions are used for each of the AOD products and AERONET. There was a problem in the computer code, base-10 logarithms were used instead of natural logarithms. This has been corrected with the new figures.

- **Comment:** "Figure 7: There are several island sites which appear to be missing from this plot, e.g. from cursory examination Arica, Lanai, Midway Island, Ascension Island, Reunion, Tahiti, plus the Australian coastal sites (Darwin appears in the top two panels, but not the bottom two; this should be checked). Perhaps more. Is this just because there were no valid land retrievals for them? These are fairly well-established AERONET sites with lots of data, and some are on islands significantly larger than the 10x10 km nominal MODIS retrieval size. I suggest checking up on this."

  Response: The selection criteria for our coastal sites are presented in Section 2.3. In the revised manuscript more detail is provided about this selection/categorization. We only use coastal sites that have at least 15 (high quality) retrievals from both the Land and Ocean MODIS algorithms during collocated AERONET AOD measurements. This reduces the number of AERONET sites but allows us to focus on the differences in trend analysis between the corrected and standard MODIS products. In addition, only sites that show a statistical significant p-value (<0.05) are displayed in the original Figure 7 (now Figure 9); this is stated in the figure caption.