We appreciate the reviewer for his/her thoughtful comments.

This paper describes an excellent concept for assessing the shortcomings of different aerosol retrievals from space by (a) comparing satellite retrievals to AERONET data and (b) cross-comparing MISR vs. the MODIS Dark Target and MODIS Deep Blue products. It is an inspiring read and a very important contribution to the field. It starts out with the more traditional approach of a satellite product validation using AERONET, and continues with a consistency analysis of MISR vs. MODIS products. Other than previous studies where, e.g., level-2 products or level-3 products from different satellites are compared regardless of the different data sampling, filtering, cloud mask etc., the comparison here makes an effort of a more quantitative comparison by (1) separating MODIS DT (the 'standard' MODIS algorithm) and MODIS DB and (2) by doing comparisons within 6-hour data windows per 0.5x0.5degree grid-box only if both MODIS and MISR provided a valid aerosol retrieval. This is a different, and probably more promising, strategy compared to a comparison of climatologies derived from individual satellite instruments because it allows a closer look at the details. Another compelling aspect about this study is the effort to engage the community. This is done by offering, e.g., kml files such that the user can visualize the study's results in different ways. This could be a model for future publications of this kind.

We really appreciate the kind encouragement from the reviewer.

(1) Using r^2 as an indicator for the goodness of a correlation in Figure 1, Table 1, and Figure 2a/b is the appropriate tool in statistics. However, and this is a common misconception, it does not necessarily reveal the level of confidence in the products without taking into account the number of data points and their distribution across the parameter space. To put it bluntly: Fitting a line to two data points gives r^2=1, but this is meaningless. Conversely, fitting a line to a lot of data points that are quite well correlated may produce a lower r^2 the more points are added. r^2 is only the first step in the regression analysis. In order to make a statement about the consistency of two datasets, this needs to be translated into a a confidence level of some sort within the framework of statistical analysis. Since the authors introduce such a nice new sampling concept, it is somewhat disappointing that they then resort to the rather rudimentary approach of using r^2, and only that, as indicator, since the aggregated data offers much more with little effort.

We fully agree with this, and generally there are plenty of data points for regression. Our examples in Figure 1 really are pretty representative. We have not seen any cases of isolates skewing the regression, and this issues, coupled with the non-linearity associated with multiple scattering. Indeed, if isolates were an issue, the correlations and regressions shown in figure 4 would not likely be regionally correlated. But the point for clarity is very well taken. We have included not only the number of data points but also the confidence intervals for the correlation coefficients in Table 1. We have also included figures (Figure 4g-h) to show the confidence intervals for the correlation coefficients in Figure 4a-b.

2) The paper argues in numerous places that somehow, a large intercept (sometimes wrongly referred to as 'interception') between retrievals is indicative of "problems with surface reflectance", whereas a slope different from unity is indicative of "microphysics" (i.e., aerosol model) problems. While this makes sense intuitively, and is probably (hopefully) described in one of the cited papers, it should at least be mentioned on what grounds this assertion can be justified.

This is also a good comment on where we need to clarify. A regression is ultimately composed of contributions from static and dynamic factors. From a TOA radiation point of view, surface albedo over land is the strongest static factor-ignoring of coarse issues of BRDF. Aerosol radiance is much more dynamic. Hence, the large intercept values are possibly due to inaccurate representation of surface characteristics [Kohn et al., 2010; Hyer et al., 2011]. We have modified the discussion in the text to highlight these references. "Regions with large intercept values indicate locations where surface reflectance values may be underestimated by the MODIS DT retrievals, typically in arid and semi-arid regions with high surface reflectance (Kahn et al., 2010; Hyer et al., 2011)." Indeed, Hyer et al., 2011 was able to correct the retrievals based on a climatological lower boundary condition, as has the Harvard group with their US products. This said, a positive Y intercept has also been an indicator for cloud contamination. Over land, Hyer et al., (2011) found the cloud screening for MODIS to be quite robust. Over ocean however, this is more of an issue. Similarly, the calibration differences between MODIS and MISR, results in a small but perceptible Y intercept difference over water (Zhang and Reid 2011). Looking at Figure 4 however, we can clearly see how Y intercept covaries with core geographic features of albedo and mountain ranges. This is even more clear when one uses the Google Earth KML provided in the supplemental materials. But ultimately, full attribution and derivation of a better lower boundary condition is worthy of several more papers.

We have also added the following discussion (for the large intercept values) based on Dr. Garary's suggestion: "Yet it may also be partially due to inaccurate representations of particle properties, especially non-spherical aerosol particles (Liu et al., 2007)."

We have also added references to indicate that slope changes are likely associated with aerosol microphysical properties [Hyer et al., 2011; Shi et al., 2011].

(3) The language is highly 'heterogeneous' across the manuscript. Some of the sections, especially, the introduction, are brilliant; others are, in the worst places, sometime incoherent and are crowded with awkward sentences or plain mistakes (examples given below).

We have revised the paper accordingly based on the reviewer's comments.

Specific comments:

p4297,l25:"The more complicated land surface also reduces the degrees of freedom in available microphysics models." - This needs to be explained. Why should that be?

This is because observations are obtained from a number of channels with limited degrees of freedom. Increasing the degree of complexity of the surface will reduce the number of degrees of freedom for solving aerosol properties. The easiest example of this would be the contrast between the over ocean and land aerosol retrieving algorithms. More aerosol properties are reported in the over ocean aerosol products than in the over land aerosol products because of the relatively simpler surface characteristics of the ocean.

*p4299,l7: '0.03+-0.15*AOD' - unclear what this means - is 0.03 an absolute error, or 15% a relative error? Should there be a bracket around 0.03+-0.15? In essence, why are there two numbers for the error? On the other hand, "0.05 or 0.2*AOD" on line 3, page 4300 makes much more sense.*

p4299,l7: "0.03±0.15×AOD" and "0.05±0.20×AOD" are the uncertainties for the MODIS over ocean and over land AOD retrievals respectively. These numbers were reported by the MODIS aerosol team [e.g. Remer et al., 2005]. It means that above the base uncertainties of 0.03 and 0.05 there are relative uncertainties of 15% and 20% of AOD values.

p4301: Fig 1: It is probably explained somewhere, but what's the date range of the AERONET-satellite comparisons? The section before only talks about satellite data, not AERONET.

p4301: Fig 1: The data for MISR, MODIS DT and DB are from 2000 to 2007. We added " for 2000 to 2007" after "comparisons" at page 4301.

l23: "Scores" need to be introduced - slopes? intercepts? correlation coefficients?

We have changed "scores" to "slopes, intercepts, r^2, number of data points, and the 95% confidence interval of correlation coefficients"

ll27: What is a 'reasonable' correlation?

We have modified the sentence to "xxx reasonable correlations with the collocated AERONET data. For example, other than the Maricopa and Sollar Village sites, r^2 values of above 0.6 are found between MODIS (DT and DB), MISR, and AERONET AODs."

p4302,l2: 'areas dominated by different aerosol species' - What are those areas?

We have changed "areas dominated by different aerosol species" to "areas dominated by different aerosol species (e.g., dominant dust aerosol particles over Kanpur and biomass burning aerosol particles over Mongu),"

p4302,l6: 'The influence of lower boundary conditions is less evident in MISR- AERONET than the MODIS-AERONET comparisons.' – How so? Which parameter distinguishes the impact of surface vs., say, aerosol type on the bias between satellite vs. AERONET? At some point later in the manuscript, it is mentioned in a side note that this is done by looking at slope and intercept as indicators (this should be better explained). On that note, what is the explanation for MISR performing sometimes "better" and sometimes "worse" than MODIS when compared to

AERONET in Table 1? Since MISR does better with surfaces (even if they are bright), it should, at least following the theory brought forward later, have smaller intercepts, provided that the surface around AERONET sites is perfectly well known. Another question on that note: Was any segregation of data by type done, based on the AERONET data? And if so, where is it shown? Such segregation would (arguably) be a better way of deciding whether 'the surface' or 'the aerosol type' has the largest impact on MODIS/MISR/AERONET discrepancies.

This is good for us to clarify. The "segregation of data by type based on AERONET data" has been performed by a few previous studies as well as a few papers that are currently in prepare. Examples of such studies are the long-term comparisons of MODIS over ocean (Shi et al., 2011), MODIS over land (Hyer et al., 2011), MISR (Kahn et al., 2010; Shi et al., manuscript in prepare) aerosol products with AERONET data. These studies suggest that a high bias at low AOD ranges over brighter surfaces is observed with the MODIS DT over land product but is not as noticeable with the MISR product, which indicates that MISR AOD retrievals are less affected by surface albedo related bias. This is possibly due to the fact that MISR has observations at high viewing angles with much longer slant paths that enhance the aerosol detection capability to, but not limited to, higher reflectance surfaces (eg. Hsu et al., 2006). As mentioned in Shi et al., [2011], the influence of lower boundary conditions is mostly constrained to small AOD ranges (e.g. AOD < 0.2), and at the high AOD range, the effects of aerosol microphysics dominates.

To highlight that this issue has been explored in other studies, we added references at the end of the sentence: "… is less evident in the MISR-AERONET than in the MODIS-AERONET comparisons [Hahn et a., 2010; Hyer et al., 2011; Shi et al., 2009; Shi et al., 2011 in prepare].

We have also included: "Regions with large intercept values indicate locations where surface reflectance values may be underestimated by the MODIS DT retrievals, typically in arid and semi-arid regions with high surface reflectance (Kahn et al., 2010)."

We haven't done segregation of data by type for the AERONET data, as in situ measurements are needed for segregating the data by type. However, segregating by data type is outside of the scope of this study and is therefore not discussed.

l11: What is an 'insufficient number of data points'?

We have modified the sentence from "Note that the black regression line for MODIS is not provided in the Maricopa plot due to an insufficient number of data points." to "Note that the black regression line for MODIS is not provided in the Maricopa plot due to an insufficient number of data points as well as a scattered and non-linear pattern of data distribution that makes a linear regression less representative." We have also removed the corresponding slope value from Table 1.

p4306, l2-5; l6-8: Here is the core description of 'high intercept' = 'surface problems' vs. 'slope' = 'microphysics'. (major point number (2) above). This needs to be explained - not even a reference is given here. If this is indeed true, demonstrate how this show differently with

MISR/AERONET vs. MODIS/AERONET because under several circumstances, MISR should perform better than MODIS because MISR does better for bright surfaces than MODIS, and it actually retrieves aerosol type along with other aerosol parameters, as opposed to MODIS where it is fixed and relies on a climatology - or is this a misunderstanding. It might be worthwhile pointing out these differences in the operational MODIS and MISR algorithms were explained briefly in the beginning. Even if most reader are familiar with it, it doesn't hurt to repeat.

Agreed. We have modified the paper with the following:

We have changed "Regions with high intercept values are most likely attributed to surface characteristics, because all of these regions are semi-arid and have relatively high surface reflectance" to "Regions with large intercept values indicate locations where surface reflectance values may be underestimated by the MODIS DT retrievals, typically in arid and semi-arid regions with high surface reflectance (Kahn et al., 2010)."

Since the two concepts "surface problem" affecting AOD retrievals at a low AOD range (e.g. AOD < 0.2) and the aerosol microphysics bias affecting AOD retrievals at a high AOD range (e.g. AOD >0.2), have been explored by various papers such as Shi et al., 2011; Hyet al., 2011; Kahn et al., 2010, we have added these references to the end of line 8.

p4307,l12-18: How were the regions segregated into the different 'problem groups' defined above? Certain parameters, a combination of parameters, or by empirical arguments?

It is based on existing studies (Hyer et al., 2011; Shi et al., 2011; Kahn et al., 2011) as well as surface albedo. Poor performance in AOD retrievals over a dark surface could be associated with microphysical influences (type2) and we categorize regions as type 3 if the surface is bright.

p4310,l24-25: How so - how were the regions identified? (Relates to a number of previous comments as well.)

It is based the existing studies (Hyer et al., 2011; Shi et al., 2011; Kahn et al., 2011) as well as determined empirically from this analysis.

p4311, l10-12: If problems are related to surface reflectance, how would AERONET sites help then?

Beside being used as a validation tool, AERONET data can be used to improve retrieval schemes. For example, the C5 MODIS retrieving algorithm used AERONET observations for fine tuning their surface albedo estimations [Levy et al., 2007].

l17: 'Our regressions show that...are robust.' - What is robust? What does 'robustness' mean. Avoid statistical slang and say what parameter this relates to - i.e.: high r^2 etc. (although, as pointed out above, r^2 alone is not necessarily sufficient!)

We have removed this sentence to avoid confusion.

Technical/Language Comments:

p4296: What is "greater south america"?

We have removed “greater”.

p4297:l7: ’surface-observtion-data-poor regions’ - excessive use of hyphenation looks awkward

We have rephrased the wording to “regions with poor surface observation data”

l10: Check ’Complicating...bias’ - something missing? word order?

We would argue that this is a preference in writing style.

l20: conditionS –> conditions

Done.

p4298: l8: bias –> biases?

Done.

l14: ’causes of the discrepancies should be collected’ –> awkward, collecting not the appropriate word. (we argue this before, and you insist to use collected. So I leave this to you)

We have changed: ’causes of the discrepancies should be collected’ to: ’causes of the discrepancies should be studied’

l29: ’heterogeneity ’ should be ’discrepancy’ or ’inconsistency’

We have change *’heterogeneity’ to ’inconsistency’*

p4299,l4: hyphenate near-UV, near-IR, fine-more, near-surface (farther down)

Done.

l2: can be very effectively used: word order?

We have changed “can be very effectively used” to “can be effectively used”

l14: retrieving process –> retrieval process

Done.

p4301, l1: "our results begin by presenting..." does not appear to be correct language; the entire paragraph should probably be reworded, it sounds awakward.

We have rewritten the whole paragraph.
“In this paper, our results begin by presenting example regressions of satellite AOD to AERONET from eight important geographical regions. Then, to understand the size of the bias features, we find the ratio of MODIS retrievals to MISR aerosol optical depth retrievals, and we study the spatial patterns of different products through spatially and temporally collocated comparisons. From these results, we return to our original eight comparisons and discuss limitations in the spatial coverage of current ground-based observations in the problematic regions identified from our results.”

l19,20: replace ’wavelengths’ with ’channels’?

Done.

p4302,l25: ’Satellite products correlate well’ - with each other? with AERONET? What’s a ’good’ correlation?

We have modified ’Satellite products correlate well’ to ’Satellite products correlate well with each other’. We consider a ‘good’ correlation as $r^2 > 0.6$.

l27: ’The question now becomes’ –> ’The question then/therefore becomes’?

We have changed ’The question now becomes’ to ’The question therefore becomes’

*p4303:l3-l8: Bad explanation of what’s shown in the Figure. It took many times to understand. Also, the caption of the figure is not helpful: What’s ’MISR AOD that corresponding to operational MODIS DT’ - this is grammatically incorrect and doesn’t make sense. During the first read-through I thought that Figure 2a is the *average* of the MISR and MODIS retrieval - which obviously is not true. But that’s the way it is described somewhere in the text.*

We have rewritten the sentence as:
“Figure 2 shows the three-year averaged spatial plots of AOD from the collocated MISR and MODIS Collection 5.1 retrievals. For example, to construct Figure 2a, only MISR AOD retrievals that have been collocated with MODIS AOD retrievals were used. The detailed collocation steps are described in section 2. By using only MISR AOD retrievals that have been collocated with MODIS AOD values, Figure 2a could be different from the three-year averaged MISR AOD plot that used all available MISR data.”

l7-8: ’Shown in Fig. 2a, the commonly ... are visible.’ awkward

We have rewritten the sentence from “Shown in Fig. 2a, the commonly acknowledged continental scale aerosol features are visible. Heavy smoke” to “Shown in Fig. 2a, regions of high AOD that are likely associated with heavy smoke aerosol plumes are seen over South America, South Africa, and Indonesia, with dust plumes are visible over North Africa and the Middle East.”

p4304: l29: ’ratio values’ –> ’ratios’

We have changed ’ratio values’ to ’ratios’

p4305,l18: ’The uncertainties...due to ratios from small values...’ Sounds awkward, also needs to be expanded.

We have removed this sentence to avoid confusion.

l22: ’data have MISR AOD values’ - sounds like the data own the values - there should always be a better word than ’have’

We have changed ’data have MISR AOD values’ to “data with MISR AOD values”

l23: Please change all occurrences of ’interception’ to ’intercept’.

Done.

p4307,l4: ’In summation’ - one of the avoidable English mistakes – it should be ’In sum’

Done. We have made this change.

l29: ’at the visible spectrum’ –> over? throughout?

We have changed “*’at the visible spectrum” to “in the visible spectrum”*

p4308,l3-7: run-on sentence

We have rewritten the sentence:
“For example, numerous field campaigns have been conducted over regions such as 5–10°S and 60–70°W of South America (e.g., Reid et al, 1998, 2005, SCAR-B and SMOCC campaigns), where many AERONET data are available. Even with extensive in situ and ground based observations, such regions may also reveal the difficulty of fully understanding aerosol properties and their spatial/temporal variations from limited ground and in situ observations.”

l17-19: ’The use of...retrievals.’ does not make sense. What does it mean?

Both low bias and high bias are observed in various regions. A global based analysis could average out such regional biases.

p4310,l2: 'indexes' –> 'indices'

Done.

p4311, l1: 'The AERONET has data' - fix English

We have changed "The AERONET has data from the UAE that helps address the Eastern Arabian Peninsula" to "The AERONET data from the UAE could be used to study retrieval related issues over the Eastern Arabian Peninsula".

l2: 'helps address the Arabian Peninsula' - what is addressed? need to state a problem, not a region.

Done. See the previous response.

l3: 'invite further experiments' - awkward; people invite, data are not alive.

We consider this a personal preference.

l2: Break out 'UAE' (United Arabic Emirates?)

Done.