We thank the reviewers (F. Seidel and one anonymous) for their helpful comments and suggestions on the manuscript. We have prepared a revised version of the manuscript, to address these comments, as indicated herein. Below, reviewer comments are in *italics*, and our responses in regular type below them.

In addition to the revisions to address reviewer comments, we made the following changes since the original submission of the manuscript:

- The reason for the delayed revision of the paper is that, towards the end of the Interactive Discussion period, we became aware that a new version 2 of the NASA VIIRS L1 data would shortly become available. We decided to wait and repeat the analysis with version 2 (rather than the previous version 1.1) L1 data to ensure that the paper remains up-to-date and consistent with the final public NASA data products, increasing its utility to other groups interested in the use of VIIRS data (from the NASA Land/Atmospheres SIPS). This also enabled extension of the time series up to the end of July 2016. (On July 27 2016, a change was made to MODIS calibration in some bands which had the potential to introduce a discontinuity in the results for some bands, so we chose this as the end point.) In the end, the difference between the version 1.1 and version 2 NASA L1 files and so the paper's results was small (<0.5%, which is within uncertainty estimates, in all cases except band M08). After discussion with the NASA VIIRS Characterisation Support Team (VCST), differences of this magnitude were expected, as was the larger change for M08. Trending in calibration is smaller in V2 than V1, which is also consistent with our analysis. VCST also confirmed that no further changes to the VIIRS calibration are anticipated at present, meaning that these results should hold for the foreseeable future.
- Several minor updates were made to the radiative transfer code, and the lookup tables were regenerated with these updates, plus a denser set of node points to further decrease interpolation errors. This makes the analysis more robust, although again the effect on the results presented is negligible.
- For the calibration exercise, the spread around the aerosol fine mode fraction (FMF) values tested to assess uncertainty is now ±0.2 rather than ±0.3. This change was made as we felt on reflection that range of the initial values, intended to represent a 1-standard-deviation variability around a true value, was too broad. This has various small effects on the analysis (mainly an increase in valid data volume for the final time series analysis) but all well within the uncertainty of the results.
- The results of the AERONET comparison section have changed slightly, due both to minor SOAR algorithm updates and an updated AERONET record since the original submission of the manuscript. Again, the main results and conclusions of the paper are not changed by these updates.

Reviewer 1 (Felix Seidel)

Summary and recommendation

This paper presents a correction to VIIRS data based on MODIS Aqua over cloud-free dark water surfaces. The applied method searches for matching observations within a 10 minute and 3 degrees viewing geometry envelope among other criteria, run a radiative transfer model (VLIDORT) to create a look up table (LUT) of simulated spectral TOA radiances normalized with the solar irradiance, find the best match between those results and the MODIS Aqua observations in terms of TOA radiances, assume those represent the truth, aggregate the data monthly, and finally compare the LUT to VIIRS observations to derive correction factors per VIIRS band as a function of time. The LUT has dimensions of viewing geometry, wind speed, Chlorophyl, AOD, and fine mode fraction. I understand that, for each observation, the first five parameters in Table 2 on page 13 are assumed to be known while the fine mode fraction is a retrieved parameter, which is forced to be the same between MODIS and VIIRS. The only free retrieved parameter is AOD with values allowed to range between 0 and 0.2 in order to retain a match. The paper shows a plot (Fig. 9 on page 18) of the correction factors for each VIIRS band as a time series to analyze trends and variances. The latter is used as an uncertainty estimate. The paper continues to show a comparison of the VIIRS retrieved AOD against AERONET, before and after the vicarious calibration factor are applied (Fig. 10 on page 23). Those results indicate something about the sensitivity of the radiometric differences between MODIS aqua and VIIRS on the retrieved AOD values. The results suggest that the (high) bias in VIIRS AOD retrievals with respect to AERONET decreased from 0.031 to 0.015 after the application of the correction factors. There are almost no corrections in the blue (Fig. 9 on page 18), which is interesting because the AOD in those bands decreased as shown in the Fig. 11 on page 23. Please explain this. It may be noted that this work bases on relative comparisons between MODIS aqua and VIIRS and against AERONET. The latter provides more of a direct observation than the satellite based retrievals, but some uncertainties remain. Since no in-situ measurements of TOA radiances or columnar AOD are possible, the conclusions in this paper rely on qualitative arguments about the improvement of AOD due to the application of the derived correction factors to the VIIRS data.

I think that this is a comprehensive and very well written paper. The methods used are not novel, but they are applied in a new and interesting way. This paper covers two main topics: cross-calibration between MODIS aqua and VIIRS and the effect of those corrections on AOD retrievals. This work is therefore relevant to both, the aerosol remote sensing and the instrument calibration community. I can see this method being used more often in the future to inter compare slightly different optical remote sensing instruments and to create homogenized datasets of satellite observations and retrieved geophysical quantities. In summary, this paper addresses two relevant questions within the scope of AMT and I therefore recommend to publish this article in AMT after all comments are addressed.

The reviewer provides a good summary of the study and we are glad that he sees the value of it. We hope that the revisions will satisfy the questions raised in the review since both reviewers made some similar points. Note that the comment about Figures 9 and 11 is addressed later, under the reviewer's specific comments.

The term 'vicarious calibration' is often used with respect to the absolute radiometric calibration of a satellite or airborne optical sensor under operational conditions where surface and atmospheric parameters are sufficiently known (ideally from in-situ measurements) such that the signal at the sensor level can be modeled using a RTM. This is partially true here as well. However, the work in this paper does not use independent surface and atmospheric measurements. It relies on another remote sensing sensor with comparable observational capabilities and accuracy. It effectively transfers the calibration of MODIS aqua to VIIRS, which is typically called 'cross-sensor calibration'. I therefore invite the authors to think about their use of the term 'vicarious calibration'.

This comment is fair enough – we agree that “vicarious” is most often used to refer to an absolute calibration, so have made this change to “cross-calibration” throughout.

Title

- *The title should clearly reflect the AOD related content of the paper.*
- *’M-bands’ should be spelled out: ’moderate-resolution’.*
- *’cross-calibration’ instead of ’calibration’, as suggested above.*

We have changed the M-band and cross-calibration parts of the title. However we did not include AOD in the title. This is both for length/readability of the title, and also the fact it may put off some potential readers who might assume the contents are only relevant for aerosol retrievals (when in fact they are relevant for any application aiming for MODIS-VIIRS radiometric consistency). We will however note that aerosols are already highlighted in the Abstract as well.

Page 1, line 6 I guess the percent numbers or correction factors apply to TOA sun-normalized radiances. Please clarify and state clearly throughout the paper to what those numbers refer to.

Yes, correct. All corrections are made to the level 1 data (sun-normalised radiance). This is mentioned in a few places but we have revised the text to state this more explicitly and more frequently.

Page 4, Eq. 1-4 This is a comment or a suggestion based on my own personal use of notation. It does not need to be addressed because the notation used in the paper is properly introduced and explained. However, I am confused with the use of I_i to express the spectrally integrated radiance per band, or radiance per band (flux per unit projected area per unit solid angle, L_i , [W/m2/sr]). I am used to I_i expressing intensity (flux per unit solid angle, [W/sr]). The same goes for Eq. 3 where $F_{0,i}$ expresses the spectrally integrated irradiance per band, or irradiance per band [W/m2]. I am used to F_0 or Φ_0 denoting the solar radiant flux (rate of flow of radiant energy or simply power, [W]). For example, on page 5, line 6 the term ’Sun-normalised radiance’ is introduced together with I/F_0, which is confusing to me as I would expect $L_i/E_{0,i}$. Personally, I would prefer a notation change throughout the paper as follows: • $I_i \Rightarrow L_i$ • $F_{0,i} \Rightarrow E_{0,i}$

Initially, we’d used the pairs L/I and E/F to highlight the difference between spectral and band-integrated properties, since a point to emphasise about the cross-calibration is that the spectral response functions are important and the sensor bands are not monochromatic. However we realise that this mixing of notation could make things less clear, and we have changed the notation at the reviewer’s suggestion to use consistently *L* and *E* (although we also note in the revised manuscript that *I* and *F* are commonly used for the same quantities within the community; favoured notation seems to differ between institutions). The difference between spectral and integrated quantities is still hopefully illustrated sufficiently by this notation, as the former include subscripted λ to indicate spectral dependence and the latter a subscripted i to denote sensor band i.

Subsection 3.1 Please discuss how cloud contamination issues are addressed. Clouds passing the cloud mask, such as smaller cumulus and thin cirrus, could potentially be a significant contribution to the uncertainties due to the large contrast with the dark water surface. Clouds are very variable in time and space, even within the envelope allowed for successful matchups. I suggest to mention clouds more specifically as well on page 24, line 14ff.

Clouds evading the cloud mask are addressed largely by using robust statistics (i.e. medians instead of means) and large-scale compositing of data rather than pairwise pixel-to-pixel comparisons. We feel the time matching, water vapour threshold, and 5 km exclusion zone around detected clouds are effective at minimising the extent of cloud contamination. We have extended the discussion in the paper as the reviewer requests, to go into this more at a few points.

Subsection 3.2 I suggest to shorten this subsection because it sidetracks from the main topic on this level of detail. Isn't the main point here is that gaseous corrections are being made implicitly using VLIDORT to compare the observations from MODIS and VIIRS?

We have retained this section. A lot of data users don't realise that trace gas absorption can be non-negligible, and that differences between similar sensors can also be non-negligible.

Figure 3 I am not sure how important this figure is. If it is, please explain why.

The point is that the gas corrections can be a similar order of magnitude to the calibration corrections, so it's important that you account for them. If you don't, the conclusions can be significantly in error. This was mentioned on page 10 lines 16-19, but we have strengthened the text in the revised version.

Figure 4 It may be worth reporting how the AOD at 550 nm is calculated if it was retrieved at another wavelength. Many AOD retrieval algorithms would combine the information from many available bands and not just a single band as shown here. It may also be worth mentioning that the purpose of using single band retrievals here is to reveal band to band calibration differences. Ideally, the AOD shown in Fig. 4b should all fall onto the same value. However, not only the calibration might affect those results, but also that the sensitivity to retrieve AOD. The latter depends largely on the combination of spectral aerosol extinction, the phase function and the surface reflectance spectra. Maybe it would be worth mentioning this even if only dark water surfaces are used in this work.

The aerosol optical model used is self-consistent, allowing the AOD to be reported at whichever wavelength is desired, even if that was not the precise wavelength used for AOD estimation. We have edited this part to improve clarity, and removed panel (b) of Figure 4, as we became concerned some readers may get confused between the "AOD estimation for correction" in the calibration stage done here and the "SOAR AOD retrieval algorithm/validation" analysis done later, or think that the calibration corrections are only useful for aerosol retrievals.

Page 16, line 22-32 I suggest to remove this paragraph as well as Figs 7 and 8. I believe that those detailed information are not directly relevant to the main findings. I can not identify a clear relation between the results shown in Fig. 9 and the number of matching observations over time. It could go to an Appendix, for example.

Again, we choose to retain this because we feel it makes an important point that although we refer to these as 1:30 pm Equatorial crossing times this is only an approximation. Over the course of manuscript preparation, we showed these plots to various people involved in MODIS/VIIRS retrievals and they were surprised by the time offsets and how they changed over the past five years. We have rescaled Figure 9 to be the same size as Figure 8, which we think makes it easier to see that this time difference is reflected in Figure 8 (albeit with a seasonally-varying component over the top)

Figure 11 Please explain in the text why the influence of the correction factors on AOD are strongest in the blue and getting weaker towards longer wavelengths, especially since the correction factors in those wavelengths (eg at 412, 440 and at 490 nm) are very close to 1. I would have expected larger correction factors at those shorter visible wavelengths to explain the relatively large impact on the spectral AOD.

This is because SOAR is a multispectral retrieval, using 7 wavelengths to simultaneously retrieve AOD, FMF, and aerosol model in a self-consistent way. It doesn't independently retrieve AOD at each wavelength. So one cannot map a change in one channel's calibration to an expected outcome in AOD. This was already noted on page 21 lines 16-19 of the original manuscript, although we have highlighted it again in the revision. Also note that SOAR does not use the 412/440 nm bands at present; if it did, it would be likely that the influence of the calibration corrections on the spectral shape of retrieved AOD would change (since, as the reviewer notes, the 412/440 nm bands appear not to need a calibration correction to bring them in line with MODIS Aqua).

Reviewer 2

This paper assesses the radiometric calibration of VIIRS based on comparisons with MODIS by using aerosol retrievals. It attempts to quantify both the biases, as well as the longterm trend in the calibration of VIIRS relative to MODIS. The study is well done, and the results will be very useful to both the calibration and the aerosol retrieval community. It serves as important feedback to the data producers for further improvements in the calibration. This study also complements the studies by the ocean color teams since both aerosol and ocean color have stringent requirements for calibration accuracy and stability. Therefore, the paper should be acceptable for publication after the following issues are addressed, since the current version lacks clarity in several areas.

We are glad that the reviewer finds this study useful, and thank them for their comments. We hope that this revised version will address the issues they found unclear concerning when we are talking about reflectance vs. AOD.

Page 1, Abstract. Line 4-9: "reflective solar bands is thought to be biases. . .": This needs some clarification. First, I assume that the biases you are referring to are the "biases in reflectance relative to MODIS". However, this can also be interpreted as "biases in AOT due to bias in reflectance". This may sound the same but they are not. The difference becomes significant when you start talking about +2% and -7% (line 7). Are these numbers in "reflectance", or in AOT? My understanding is that 1% bias in reflectance is not the same as 1% in AOT. This is a major concern throughout the paper, since it discusses both "reflectance" and "AOT" and often interchangeably. My recommendation is that in the revision, clearly state what it is when you discuss biases and trends. Otherwise, the results are hard to interpret and compare. Also note that 7% bias in "reflectance" would be significantly out of instrument performance specification, while 7% in bias in "AOT" does not have the same meaning.

All corrections are corrections to the calibration of the sensor (reflectance), not to retrieved quantities such as AOD. The AOD section at the end is only to shown the positive impact on downstream data products. This was mentioned in the original paper but we have added mentions of reflectance in a few more places in the revised version of the paper, to minimise the chance for confusion.

The second concern is the VIIRS L1 product used in the study. On line 9, it states "standard VIIRS NASA calibration". However, it is my understanding that there are several versions of the VIIRS NASA calibration. As far as I know, there was the LandPeate version which evolved into Land SIPS. There is the Atmospheric SIPS, and the ocean color version as well. The authors touched some of these different versions but still it is not clear which version is actually used in this study. Furthermore, even for the same SIPS, there are different versions due to lookup table upgrades. Therefore, it is important to state what specific data were used in this study, preferably very early in the paper. This also becomes important when discussing other versions such as the NOAA IDPS, as discussed later.

We believe the reviewer is slightly mistaken – VCST produce one set of VIIRS calibration coefficients/tables, which are applied to create the NASA Level 1 VIIRS data products by both the Land and Atmospheres SIPS (where SIPS is the new name for what was formerly known as PEATEs). (The Ocean Colour group go ahead and do their own vicarious calibration, but the source they work from is the same.) The data version used was stated on page 7 line 19, but in the revised version we have added text to indicate that this is the version used for processing of all NASA Atmosphere and Land VIIRS products.

Page 2, Line 33: “suggesting calibration biases in excess of 5% in some cases”. This statement oversimplified the situation and may be misleading for two reasons. First, the large biases between VIIRS and MODIS were found to be due to spectral response differences (such as M5, see discussion later). The biases are within 2% after spectral factors are taken into account. In fact, after reaching the “validated” maturity, all bands achieved 2% accuracy except M11, which had a performance exemption since there is no standard to compare against (same issue with MODIS which also has a very different spectral response). Second, the studies mentioned here are using the NOAA IDPS data, which is not the “Standard NASA data” that the current study is using so whatever bias values discussed may not be relevant to the current study. Mixing these different data sets only introduces confusion. Note that the NOAA IDPS data suffers from the issue of major calibration changes since launch especially before the validated maturity. The IDPS focuses on near real time (NRT) data production, and has no ability for recalibration on what was generated in the past.

I can see that the authors try to make an argument that VIIRS calibration may not be accurate, but I don’t think these arguments here are well founded for the reasons discussed above. Mixing the bias in AOT vs. “reflectance” doesn’t help either. Suggest focus on the specific data used in your study and the findings, which should be more convincing in reaching conclusions.

We noted in the Introduction and on page 20 that the different VIIRS data versions which existed meant that we could not apply corrections developed for those for our study. The text in these sections has been rewritten to make this point about IDPS more clearly – we certainly did not want to criticise the IDPS processing system and know that NOAA’s goals are different from NASA’s, which means that different approaches are used.

Page 3, Line 33: “describe a vicarious calibration of S-NPP VIIRS RSB against MODIS Aqua. . .”. I think what you are doing is a “vicarious comparison of the S-NPP VIIRS RSB calibration against MODIS Aqua”. The word “calibration” has specific meanings in the context of radiometric calibration, although it may be loosely interpreted as many different activities which can only introduce confusions.

Related to this and Reviewer 1’s comment, we have changed the text to refer to “cross-calibration” throughout.

Page 4, line 8, “close to one more MODIS band. . .”, should be “one of MODIS band”?

Good catch – this should have read “close to one or more MODIS bands”, and we have corrected this in the revised version.

Page 6, Figure 1: band M3 shows large differences in VIIRS and MODIS mainly because B3 is spectrally very different from that of the VIIRS band. Why didn’t the authors use a better matching band of MODIS B10? M3/B10 match much better which should reduce the uncertainties in the analysis significantly.

This was addressed on page 4 lines 8-14. In short, we do this because (a) the so-called MODIS 'land bands' 1-7 are what are used for MODIS land/atmospheres applications, and since we are in the Atmospheres community we are most interested in radiative consistency for atmospheric applications, and (b) saturation can be an issue for MODIS 'ocean' bands such as band 10 (although for our specific dark-water case, this is not important). It is possible that using MODIS B10 rather than B3 would decrease the uncertainties, but this channel appears to be quite stable and our estimated uncertainty is low, so we don't expect that there would be much to gain. Note the large change to this band is also consistent with independent analyses from the cloud teams.

Page 20, line 26: "is to darken the VIIRS channel by up to ~7%...". See comments in the beginning. This needs clarification on whether you are suggesting "reflectance 7%" or AOT 7%. The difference is fundamental. If you really found a 7% in reflectance in some VIIRS bands, this would lead to a major review in the VIIRS performance because it would suggest fundamental issues in the instrument which would be new to paper most people.

As detailed in the response to other comments, these corrections are to reflectance, not AOT. This has been added here in the revised manuscript for clarity. We agree that 7% seems quite large and discussed this issue a lot with MCST/VCST at the 2016 MODIS/VIIRS Science Team meeting and elsewhere before submitting the paper. Like us they were surprised but did not find fault with our analysis. Additionally, the NASA MODIS/VIIRS cloud team, doing an independent analysis of bright liquid water cloud scenes, found similar results to us (unpublished; K. Meyer, personal communication, 2017). It is also possible that some of the difference is attributable to a calibration issue with MODIS, since this is a cross-calibration rather than an absolute calibration. We have mentioned this possibility again as well, in the revised manuscript.

Page 23, lines 1-5: Several previous studies have shown the large biases of ~7% in M5 and ~3% in M7 (Uprety et al). However, it has been explained in those studies that the biases are mainly due to the spectral response differences compared to MODIS bands rather than due to VIIRS radiometric calibration. As discussed earlier, the authors should clarify whether the bias found in this study is AOT, or "reflectance". Also, whether spectral difference induced biases have been accounted for, before recommending a correction in the reflectance calibration.

The reviewer is correct that some figures quoted in past studies did not correct for spectral response function differences, which contribute to apparent differences between MODIS and VIIRS. Sections 2 and 3 and associated Figures describe the details of how we account for these spectral response function differences through the radiative transfer model and trace gas absorption calculations. In the revised version we mention again that we have accounted for these differences already. Some of the aforementioned cited studies did also account for these spectral differences; others did not, and in some cases it is unclear. Those that did still found unresolved calibration biases of up to several percent, in some cases.