As a result of the review process, the manuscript has been modified significantly. Mayor changes are:

1) Section 2 of the paper has been extended to include a brief but detailed description of the TropOMAER algorithm. It includes a description of the UVAI calculation as well as a summary of the AOD/SSA retrieval process.
2) Section 3 on the validation of retrieval results using AERONET observations also changed considerably. The original validation analysis consisting of a direct validation of TROPOMI AOD results to AERONET observations at 12 sites was replaced with an approach that allows the separate evaluation of retrieved product improvement as a result of instrument enhancement and algorithmic improvement. AERONET observations 12 sites are used as an aggregate. A three way validation exercise is then carried out: 1) AERONET vs OMI, 2) AERONET vs TROPOMI using heritage (OMI) cloud mask, and 3) AERONET vs TROPOMI using VIIRS-based cloud mask. Inter-comparison for validations 1 and 2 highlights the effect of improved instrumental capabilities, whereas differences in validations 2 and 3 indicate retrieved product improvement due to algorithmic upgrades.
3) The revised paper (to be available soon after the submission of replies to reviewers' comments) contains 13 figures (five more than in the original version).

In the reply below the reviewer's comment is in black and our answer in blue.

Reply to Comments by Reviewer 2

This paper briefly introduces a TropOMI aerosol data set based on heritage OMI UV algorithms by the Torres group (OMAERUV and OMACA). This provides UV aerosol index (UVAI), aerosol optical depth (AOD), and single scattering albedo (SSA). A comparison of AOD and SSA against data from selected AERONET sites is presented, along with a few case studies of extreme events. The concept of the paper is in scope for AMT. The quality of language is good. The topic is important because OMI is ageing and TropOMI is the next generation of this type of sensor (OMPS on SNPP and JPSS has some aerosol capabilities but is in other ways worse than OMI).

However, honestly, the current paper feels more like a conference proceedings or an article for a Letters journal than a full scientific paper. It is brief and does not go into much detail. For a focused journal like AMT I think something much more technical is needed. Though I realise I am proposing a fair amount of work, I prefer that the authors expand this analysis rather than resubmit elsewhere, because I think a thorough accounting for TropOMI's capabilities for UV aerosol remote sensing is needed and is Interactive more or less missing from the literature. The authors are the right people to do this comment because they are the most expert with their data products. I know it is annoying when reviewers ask to do more work, but there is not enough content here to justify publication and I don't think that the article as written satisfies the scope a reader would reasonably expect. Case studies are one thing but by nature are typically unusual events and so looking at them may not give a representative picture of the data set as a whole. I recommend major revisions and would like to review the revision.

The paper has been significantly extended to address the issues raised in the review process.

My main suggestion for expansion is to give a detailed comparison between OMI and TropOMI results.

The original evaluation analysis involving AERONET-TROPOMI comparison of aerosol derived products have been converted into a three-way AERONET-TROPOMI-OMI over the same period.

OMI-TROPOMI results are compared for individual events as well as in terms of monthly averages for three representative regions as well as seasonal (summer) global averages.

Users familiar with OMI need to know whether we can use TropOMI for the same types of research, and to what extent the same caveats/biases are found. Right now this is not answered in a thorough way. One big advantage of TropOMI over OMI is the spatial resolution. I would expect that this is important because those cases where the UV technique works well (absorbing aerosols) are also often strong and heterogeneous events. So the finer spatial resolution might mean both (1) less cloud contamination and (2) better AOD/SSA retrievals, because top of atmosphere radiance is not linear in AOD, so by resolving more spatial structure you become less sensitive to sub-pixel variations. If this is true in practice, great. If not, this needs to be shown and understood. It is briefly discussed in Section 3.1 but not supported by the plots shown, only by briefly mentioning other references. Here are some suggestions for relevant analyses to include:

The revised version of the paper specifically addresses the issues addressed by the reviewer as explained below.

(1) Show global maps so we can see how similar the big picture looks from both sensors. In my view the time series in Figure 5 isn't sufficient here because both data sets are heavily spatially and temporally averaged in it.

Because of the so-called row anomaly of the OMI sensor that reduces OMI's daily coverage to about 50%, OMI-TROPOMI global daily maps are not the best way visual comparison. We show OMI-TROPOMI comparison on daily, monthly regional, and global seasonal temporal scales.

In Figure 1 of the revised version of the manuscript we show a comparison of OMI, NASA-TROPOMI and KNMI-TROPOMI UVAI on August 18 over North America. To our knowledge, except for UVAI, no other TROPOMI aerosol products are available.

Side-by-side maps of OMI and TROPOMI retrieved SSA and AOD for the same event are shown on Figure 8.

A two-year time series of monthly-averaged OMI and TROPOMI AOD and AAOD (absorbing aerosol optical depth) over three regions are shown on Figure 4.

OMI and TROPOMI summer 2018 seasonal global maps are compared in Fig 6, and a scatter plots of OMI TROPOMI UVAI monthly mean values is shown on Figure 7.

(2) Include OMI in some of the case studies (e.g. visual inspection of maps).

OMI graphics similar to the TROPOMI images have been added to the discussion of the 2018 California and Pacific northwest fires.

(3) OMI validation results could be presented alongside the TropOMI data. I know the validation has been published elsewhere but it will be clearer to the reader if plots are shown next to one another with the same axis range, etc.

The focus of the comparison to AERONET has changed from the narrowly focused AOD validation exercise in the original version of the paper, to an analysis of the instrumental and algorithmic differences throughout the use of independent ground-based observations. The combined AERONET data aggregate from observations the 12 sites, is compared to satellite observations as follows. An evaluation of instrument-related improvements is done by comparing AERONET measurements to three satellite-based data sets:1) OMAERUV, 2) TropOMAER with heritage (i.e., OMAERUV) cloud screening, and 3) TropOMAER with VIIRS cloud mask.

A comparative analysis of evaluations 1 and 2 shows the impact of enhanced instrumental capabilities, whereas the analysis of evaluations 2 and 3 highlights the effect of using the VIIRS cloud mask which is the only TropOMAER algorithmic modification.

(4) Directly plot (as a scatter density diagram) the AOD and/or UVAI from OMI and TropOMI, for collocated pixels (i.e. same scene, same time, similar geometry) at level 2 resolution. The orbits should overlap frequently. Then we can see if there's much scatter, if it's a straight line or not, etc. I don't know how much collocated data is needed to get a meaningful comparison – perhaps the case studies give enough, perhaps it has to be done on a month's worth of data. MODIS or VIIRS data could be useful for extra context (and filtering); I know and the manuscript mentions that the TropOMI orbit choice makes it possible to take advantage of SNPP VIIRS for e.g. cloud masking.

Because of the row anomaly the orbital overlap the reviewer describes is very cumbersome and time consuming. Figure 7 shows a scatter plot of seasonally averaged UVAI for the data mapped in Figure 6.

We believe the OMI-TROPOMI comparative analysis at daily, monthly regional, and seasonal temporal presented offers a complete analysis of the equivalence and compatibility of these two data sets. Additional comparisons involving other sensors are beyond the scope of this manuscript intended as a paper on first results of the ported algorithm and not yet a consolidated product.

The above comments and suggestions all apply (potentially) to the DSCOVR-EPIC sensor, too, although OMI is the more well-known and mature record so probably makes better sense to baseline against. Though I would certainly be happy to see a three-way (OMI, EPIC, TropOMI) comparison.

We will certainly carry additional comparison to other satellite products in the near future.

Other comments on the study are as follows:

Introduction or section 2.1: somewhere here it would be good to contrast TropOMI capabilities (e.g. spatial/spatial) with OMI and maybe TOMS and EPIC, since those are the main comparative products. The introduction mentions GOME and SCIAMACHY but those are less relevant since the authors' algorithms are from TOMS/OMI heritage and EPIC data are shown later. Maybe mention OMPS too as while a step backwards from OMI in terms of spatial resolution, it is used for UVAI and is the US operational follow-on for that. I know that there are TropOMI products in development on the Dutch side too – it's not clear to me whether those are public yet, but if so, there may be value in comparing and contrasting with those too.

The TOMS, EPIC and OMPS records are included in the discussion.

Section 2.2: if I understand correctly this section states that (1) there is a 5-10% calibration difference between OMI/OMPS and TropOMI in the relevant bands in the standard calibration, and (2) because of this the authors do their own vicarious calibration. Is that right? Either way, this could be worded a little more clearly. What is the difference between the sensors after the vicarious calibration?

The vicarious calibration brings the TROPOMI and OMI closer in measured reflectance terms as evidenced by the AOD validation presented here that shows overall consistency between the two records. The revised version of the manuscript contains an improved description of the vicarious calibration procedure.

Section 3: clear statements and references about AERONET data products and versions used need to be made. For example, I assume this is version 3 level 2.0 direct Sun (Giles et al AMT 2019) and inversions (Sinyuk et al AMT 2020). However this does not appear to be actually stated in the paper. If this was not the versions used, the analyses should be repeated using the latest data versions.

Yes, AERONET data version 3, level 2.0 was used. It has been clearly stated in the revised version of the paper.

Section 3.1: if the authors really believe that a relative uncertainty of 30% on TropOMI AOD is true, then by definition they should not be using linear least squares regression fits, because a relative uncertainty means that the assumption of constant variance of errors is broken. See for example standard statistics textbooks or web pages such as https://statisticsbyjim.com/regression/heteroscedasticity-regression/ . This issue could be addressed with weighted least squares. Ideally also the uncertainty on AERONET AOD (I think 0.02 in this spectral region) should be accounted for in the fitting. Also, if you expect a relative uncertainty then RMSE is not the best metric to be reporting since that is scale-dependent…others like relative RMSE would be more appropriate to quote instead/as well (and this would help tell you if it is really 30%). The statistical analysis here is not very appropriate. The authors may have used this type of analysis before but that does not mean it is ok to do something again if it is wrong.

TROPOMI's retrieval uncertainty is probably lower than the quoted 30% value. This is actually a conservative TOMS/OMI based estimate that includes the combined effect of the uncertainty on assumed aerosol layer height (smoke and dust layers) and sub-pixel cloud contamination. At TROPOMI's much finer spatial resolution the cloud contamination component should be significantly lower. Actual uncertainty is still to be determined pending remaining calibration issues as discussed in this manuscript. We appreciate the reviewer's observation on the appropriateness of using linear square regression (LQR) fits . LQR analysis have been used as a standard method of validating satellite AOD retrievals The use of this common approach facilitates the relative comparison of the same physical parameter measured by large variety of sensors and retrieval algorithms.

The reported LQR parameters in this manuscript based on relatively small sample of observations are only intended to illustrate relative improvement in the accuracy of retrieved parameters associated with TROPOMI enhanced instrumental and algorithmic capabilities with respect to OMI. We do not expect the conclusion of our analysis to change if a more refined fitting approach was used. This is by no means an exhaustive validation exercise of the TROPOMI record for which a lot more AERONET observations are needed.

Section 3.2: the authors use a 6 hour time window (3 hours each side) for the SSA comparison because morning/evening almucantar inversions have lower uncertainty than midday ones. The untested assumption here is that SSA does not vary much throughout the day. Ok, but version 3 also introduced hybrid scans which were specifically developed to solve this problem by sampling a larger air mass and scattering angle range during the middle of the day. This could be checked by using the hybrid inversions as well and seeing if you get the same results.

Hybrid scan availability is limited to specific sensor types. In general, reliable AERONET SSA retrievals are done for AOD (440 nm) > 0.40. That limitation significantly reduces the number of SSA measurements available for comparisons to satellite retrievals. Using hybrid scans only further reduces data available.

The hybrid scans are certainly useful to examine the issue of diurnal variability. We will consider using them in future specific validation efforts.

Also, an explanation is needed for how the authors split the data into the three aerosol type categories for Figure 2 and the discussion.

The aerosol typing is described in a new section of the paper that describes the algorithm as suggested by reviewer 1

Section 4: this feels like advertising. I agree that TropOMI results look impressive but (aside from a brief mention of AERONET AOD) there is no way to know how 'real' they are. This section feels like something you might put on a webpage or brochure to attract attention to your new data set, rather than a detailed scientific analysis. I am not sure what is best to do here. For a journal like AMT I'd rather than space was devoted to more technical, large-scale comparisons. Perhaps this aspect could be split off for a Letters journal. Or, expanded with more context from meteorology and other (space or suborbital) data records and submitted separately to ACP. I know this is a joint special issue but the content still needs to match the journal. It does not really fit here, and there's not enough detail presented to consider this paper an authoritative reference for these case studies.

We disagree with the negative connotation of the term 'advertising' as used by the reviewer. As a matter of fact, this entire paper, not just section 4, as well as all science papers, are intended to introduce and advertise the availability of a new science products or ideas. That is the role of the scientific literature. The problem is when false advertisement takes place. Hopefully, the preceding three sections of the paper on algorithm description and evaluation of derived products give the reader some confidence to treat as 'real' the discussed practical applications of the derived products in section 4.

Figure 6 and associated text: I'm not sure that it makes sense to show the EPIC results on the left panel. That's a different sensor, different resolution, different observation geometry (backscatter for EPIC). UVAI is sensitive to all of these things. Also, what is the scaling referred to in the left panel? That is not mentioned in the paper.

Left panel Figure 6 has been excluded as it does not add much to the discussion without going into an additional explanation and description of the EPIC sensor. The EPIC application referred to in this paper is discussed in detail in the quoted literature.

I expect that the general point about the two events will still stand but it's not clear how much of the systematic difference (and scatter on the left panel) are a function of real differences in the smoke in the two events and how much is contributed by sensor differences. The paper is far too sparse in detail for a reader to judge, which makes the comparison less instructive.

Figure 6 left panel has been removed.

Figure 6 legend: is the black dot in the left panel legend (12 km) meant to be a black line like in the right panel? If so, formatting should be consistent. If not, the difference needs to be explained.

Figure 6 left panel has been removed.

Section 5: "The NASA TropOMAER aerosol algorithm is a modified version of the one applied to OMI observations." Wait, what? Section 2 describes the OMI approach but doesn't clearly state that there are modifications. What are these modifications, why were they made, what effect does this have on the results, and will they be back-ported to OMI? This all needs to be addressed in the paper.

Do not panic. The only modification is the use of the VIIRS cloud mask whose effect in retrieval results has been discussed.