RE-REVIEW OF 'JOINT RETRIEVAL OF SURFACE REFLECTANCE AND AEROSOL PROPERTIES WITH CONTINUOUS VARIATIONS OF THE STATE VARIABLES IN THE SOLUTION SPACE: PART 2: APPLICATION TO GEOSTATIONARY AND POLAR-ORBITING SATELLITE OBSERVATIONS'

My primary comments concern your response our opinions of Fig. 14.

- "*The bias between the CISAR retrieval and the AERONET data is shown in Fig. 2, which shows different performances for SEVIRI and PROBA-V. These differences show that the bias does not only depend on the CISAR algorithm itself, but also on the quality of the processed data.*"

  Your SEVIRI and PROBA-V implementations use a different number of channels, so it is entirely reasonable for them to exhibit different bias profiles. This does not excuse the fact that your algorithm exhibits biases of around 50% in your own validation (and an independent validation would be expected to find worse comparisons). I appreciate that the two satellites you use don't exhibit the radiometric quality of MODIS or AATSR. That is expressed (as I would expect) through the large uncertainty in your products relative to MODIS.

  The PV-LAC validation report (https://earth.esa.int/documents/700255/2632405/PV-LAC_ATMO_VR_v2.2.pdf/4c46403b-bfe5-4208-bfdf-2d42585d6589), prepared by Erwin Wolters from VITO (De Vlaamse Instelling voor Technologisch Onderzoek) through an independent verification, does not show much worse results, keeping in mind that the algorithm improved since then.

  However, Fig. 2 of your response only worsens my opinion of your results. A calibration offset should result in a retrieval bias that is (roughly) independent of optical depth. Lower quality detectors should give a wider scatter (which, admittedly, you appear to have). Cloud contamination should result in a positive bias. Yes, your results for SEVIRI are within the GCOS requirements for a particular range but your biases for AOD < 0.2 are almost 100%. As AOD is log-normally distributed, this region carries significant weight.

  We recognize the limitations of CISAR retrieval for low AODs when applied on SEVIRI and PROBA-V data. However, the bias for AOD = 0.1 is 54% for SEVIRI and 48% for PROBA-V. For AOD=0.15 the bias decreases to 18% for SEVIRI and 27% for PROBA-V.

  I think we see Fig. 2 very differently. I expect you see the SEVIRI bias as a straight line around zero, that drops off above 0.7 due to the small volume of data and cloud contamination. I look that that line and, neglecting the last point, see a linear downwards trend. PROBA-V does the same, but with a different gradient. I can live with data with a large RMS—average could reduce it. I can live with data that has a bias — one can subtract it. Your data has a slope. I'd need both coefficients of ax+b to bias correct your data and that's difficult. To my mind, an algorithm that sruggles to retrieve both small and large AOD provides little of use.
  So what do I think should be done? It would be inhuman to ask you to rubbish your own data in your own paper. Your revisions do better represent the quality of this data, but

you primarily blame the instruments and cloud. If those were the only problem, you should resubmit the paper after applying it to a better instrument.
I continue to believe that your algorithm is conceptually interesting. What I need is a discussion of what you intend to do next. What aspects of the algorithm are you working on? Where do you think the problem lies?

The sentence "The cloud mask omission errors impact on the AOT overestimation at low optical thickness deserve additional work." has been moved at line 516 and the following lines have been added afterwards:
"In order to reduce the impact of cloud contamination in the AOT retrieval, a new version of the CISAR algorithm is under development in the framework of the ESA-SEOM ConsIstent Retrieval of Cloud Aerosol Surface (CIRCAS) project (www.circas.eu). The new version of CISAR aims to retrieve both the AOT and the Cloud Optical Thickness (COT), overcoming the need of an external cloud mask. Within the CIRCAS project CISAR will be applied to observations acquired by the Sea and Land Surface Temperature Radiometer (SLSTR) on-board Sentinel-3."

- *It can be seen there only few points correspond to AOT È 0.8 (less than 5% of the total number of observations), affecting the reliability of the statistics for high values of AOT. The histograms have been added in Fig. 14.*

  Though I agree that large AOD events are rare, they can be very important. Large dust plumes seed the equatorial ocean, large fires impact air quality over entire continents, and volcanic eruptions affect international air travel. Most algorithms have trouble with high AOD, where the fundamental assumptions of such retrievals begin to break down, but they are an area of active development.

  As stated in the manuscript, we believe that the processing of more data would be necessary to increase the confidence in the results for high AOT values.

- *The overestimation rapidly decreases as the AOT approaches values of about 0.2.*
  This is unimportant as the line has to cross the axis somewhere.

  It is not clear to me what the reviewer means by this.

- *We are not aware of any algorithm capable of delivering a good AOT product from PROBA-V over land surfaces.*
  I'm not aware of anyone having tried as I'd never heard of the instrument before reading this paper.

  The operational product includes an atmospheric correction method. However, when compared with CISAR retrieval during the PV-LAC project, it did not show promising results.

- With regard to my own point 33, your Fig. 14 shows the comparison of CISAR to AERONET through boxplots. In the supplement, you show a comparison of CISAR to MODIS through 2D histograms. I vastly prefer the later form of plot and was requesting a second version of Fig. 14 in the style of Fig. S1.

  During the Aerosol-CCI project, it was strongly advised to show boxplots for the AOD retrieval, hence our choice.

A few other thoughts:

- I'm not overly happy with the assumptions you make in §2.4 but they're rational. I will note, though, that on L148 you state you set the uncertainty to a 'high arbitrary value'. Unity is not high. When I wish to avoid setting a prior for a variable, I use a prior uncertainty of 10^8.

  Reviewer#1 seemed to believe the opposite. From his annotate pdf "2.0 for the coarse mode is unreasonably high". An uncertainty of 1.0 associated with values < 0.2 (this is the cases for the AOD climatology values) can be considered as a high value.

- You reference an unusual number of reports and conference presentations. I know private companies struggle to justify publication costs but I'm slightly concerned that many background details for this algorithm haven't been peer reviewed.

  We will be happy, in the future, to cite this paper and its companion.

- I think I now understand §5. I'm not fond of this manner of qualitative quality filtering, but it is commonly used so I'll not comment further.

- Is Fig. 12 actually binned (i.e. showing the average correlation for all retrievals with QI in a certain range)? If so, it would be useful to represent those ranges on the plot (e.g. the step histograms of matplotlib).

  The final QI is rounded to one decimal place; therefore no binning is performed in Fig. 12.

All the reviewer's suggestions in the pdf were implemented, with the exception of:

- "CISAR has been applied to SEVIRI and PROBA-V observations acquired over 20 AERONET stations". The reviewer suggestions was to replace "over" with "from". However, the satellite observations are not acquired from AERONET, but over an area surrounding the AERONET station.
- "PROBA-V satellite mission is intended to ensure the continuation of the Satellite Pour l'Observation de la Terre 5 (SPOT5) VEGETATION products since May 2014". The reviewer suggestion was to replace "since" with "begun in". However, it was preferred to replace "since" with "starting from".
- "FASTRE uncertainty is in the range of 1% - 3% (Table 6), which is smaller or equal to the instrument radiometric noise." The reviewer suggestion was to replace "smaller

or equal” with “equivalent”. However, this would change the meaning of the sentence.

- The reviewer suggests eliminating the word “however” at L406. Nonetheless, it is needed to relate to the previous sentence and explaining why, despite the poor radiometric performances, the AOT retrieval from the two instruments is meaningful.