The editor's / reviewers' comment is in black, the author's response is in blue.

According to editor's and reviewers' comment, the structure of the manuscript has to be changed for better reading. As a guidance, we describe here the major changes:

- We used to describe SSA retrieval using radiative transfer simulations and SVR in parallel, which caused troubles in reading. Now in the revised manuscript, we separate the two methods thoroughly. Section 2 includes everything about the SSA retrieved by radiative transfer simulations, and Section 3 contains all information on SVR retrieval.
- There used to be 2 SVR models: one uses the OMAERUV-AERONET joint data set (UVAI, ALH from OMAERUV and AOD, AAOD from AERONET) to train the SVR model, we call it as the SVR trained by the original training data set; another uses the same training data but with adjusted ALH to replace the ALH in OMAERUV, we called it the SVR trained by the adjusted training data set. The adjusted ALH is using an intermediate SVR trained by TROPOMI ALH. Thus, there used to be 3 SVR models in the previous version manuscript. The SSA retrieved by the adjusted training data set is slightly better than that retrieved from the original training data set (OMAERUV-AERONET joint).
  The original purpose to adjust the ALH is because the OMAERUV ALH is not retrieval but is guessed either from CALIOP climatology or a priori assumptions from AOD retrieval. We used to adjust it with TROPOMI ALH to make it more like observations. But the SSA retrieved by the SVR with the original training data set is acceptable, meanwhile the adjusted ALH causes many confusions. Thus, in the revised manuscript, we have removed the process of adjusted ALH and the SVR trained by the adjusted training data set. There is only one SVR model in the revised manuscript, which is trained by the OMAERUV-AERONET joint data.
- We used to employ AEORNET version 2 inversion product to evaluate our SSA retrievals, and to construct the training data set for SVR method. According to Omar Torres's comment, we have replaced it with AERONET version 3 inversion product. The results and conclusions may change to some extent.
- We used to have only one case study in the manuscript as it was the only one available at that time. Now, we have searched through the recent half year since 2018 November and added cases as long as there are collocated TROPOMI UVAI and ALH, MODIS AOD and AERONET measurements available.
- We have included MERRA-2 aerosol reanalysis (Appendix C) as an independent reference to analyze the spatial variability of retrieved SSA in Section 3.6.3.

The structure of the revised manuscript is as follows:

Section 1 Introduction

Section 2 Experiment 1: SSA retrieval using radiative transfer simulations
Section 2.1 Radiative transfer simulation setup
Section 2.1.1 Aerosol models
Section 2.1.2 Inputs from satellite
Section 2.2 SSA retrieved by radiative transfer simulations

Section 3 Experiment 2: SSA retrieval using support vector regression
Section 3.1 Support vector regression
Section 3.2 Feature selection based on OMI and AERONET observations
Section 3.3 Preparing training and testing data sets
Section 3.4 SVR hyper-parameter tuning
Section 3.5 Error analysis
Section 3.6 Case applications
Section 3.6.1 California fire event on 12 December 2017
Section 3.6.2 Other case applications
Section 3.6.3 Spatial variability of retrieved SSA

Section 4 Conclusions
Appendix

Response to anonymous referee #1's comments

The paper tackles the important issue of the impact of assumptions about aerosol layer height and spectral dependency of the aerosol refractive index on the quantification of aerosol SSA in the ultraviolet. With this aim in mind, the Authors compare the results of the "standard" KNMI retrieval scheme to those of a novel retrieval based on support vector machines (SVM), trained with real observations, on a particular scene of an aerosol smoke plume observed by TROPOMI. The comparison, which uses AERONET SSA as a benchmark, reveals that some assumptions made in the KNMI standard retrieval look problematic, and that the SVM based method is able to circumvent the problem and return more realistic values for the SSA.

GENERAL COMMENTS

While the scientific result of this paper is certainly interesting, I think there are a number of issues that need to be addressed before the paper can be published. First of all, I agree with the Editor's opinion that the manuscript does not read smoothly. The explanation of the SVM algorithm is difficult to follow, fails to mention important information (what's a support vector, what's a kernel) and makes it difficult for a reader to understand what is going on. In the description of the pre-processing it is not always easy to understand which quantity comes from which product (e.g., surface reflectance). The actual description of what was done to train the SVR for the retrieval of the AAOD is also confusing. Till Section 3.2.3 I was convinced that only a SVR is trained for the retrieval of AAOD, but at the end of Section 3.2.3 I get to know that there are two, and I don't fully understand why. In general, I think that the description of the entire process flow and of the logic behind it needs to be made more intelligible.

Finally, I have some concerns on validation. Testing the proposed method on a single scene basically means that the validation of the method is done against only one measurement. While the agreement between the SVM-based retrieval and AERONET looks excellent for the case shown, it would be important to see if this result is con- firmed by looking at some more high aerosol loading events, which I guess should be possible to find, with ~1.5 years of TROPOMI observations now available. Below are some point-by-point comments.

As declared at the beginning of this document, we have restructured the manuscript, with a separated section on SVR (Section 3), and only keep one SVR model to avoid misunderstanding. For more information on the structure modifications and other changes, please see the overview at the beginning of this document.

There was only one case available when we were preparing this manuscript. In the revised manuscript, we have added other fire events happened recently, as long as there are collocated TROPOMI, MODIS and AERONET measurements available.

SPECIFIC COMMENTS

- Abstract, L16. Do you mean inappropriate assumptions on the spectral dependency of the SSA?

Actually, it is the inappropriate assumption on the spectral dependency of the imaginary part of the refractive index causes the disagreement between retrieved SSA and AERONET SSA.

The sentence has been changed into: *With the recently released ALH product of S-5P TROPOMI constraining forward simulations, a significant gap in the retrieved SSA (0.25) is found between radiative transfer simulations with spectral flat aerosols and strong spectral dependent aerosols, implying that inappropriate assumptions on aerosol absorption spectral dependence may cause severe misinterpretations of aerosol absorption. (line 13-16)*

- L29. After Eq. 1 it would be useful to recap what are typical values of the UVAI for absorbing and non-absorbing aerosols.

We have added the explanation: *Positive UVAI indicates the presence of absorbing aerosols, while the negative or near zero values imply non-absorbing aerosols or clouds (Herman et al., 1997). (line 35-37)*

– L37 and L46. Jeong and Su (2008) and Chimot et al. (2017) cannot be found in the references.

We have added the references accordingly.

- L72, “Another advantage”. “Another” with respect to what?

Sorry for the misunderstanding. The sentence has been changed into: *From our perspective, ML techniques can avoid making assumptions on poorly-understand aerosol micro-physics as that in the first experiment. (line 75-76)*

- L81. Format reference correctly.

The reference format has been changed accordingly.

- L83. Yao et al. (2008) cannot be found in the references.

We have added it in the reference.

- L83, “. . . as it only depends on a subset of training data”. WHAT exactly depends on a subset of training data? Also, here you mention the term “epsilon-insensitive loss” but don’t say what it is, thus after this sentence the reader is really none the wiser about what you mean.

SVR attends to find an optimal hyperplane that maximizes the margin of tolerance (i.e.ε) in order to minimize the error. The error within the margin does not contribute to the total loss function. Thus, we say SVR only depends on a subset of training data and its loss function is ε-insensitive.

More introduction on SVR is in the newly added Section 3.1.

- L84. Again the same problem. You mention “kernel functions”, but if you don’t say what they are and what they have to do with SVMs, then this sentence is of no use at this point.

The kernel function is a property of SVR to solve linear or non-linear problems, depending on the kernel functions.

More introduction on SVR kernel is in the newly added Section 3.1.

- L86. Mountrakis et al. (2011), Noia and Hasekamp (2018) cannot be found in the references.

We have added them in the reference.

- L86, “consist” -> “consisting”?

We have changed it accordingly.

- L90, “expresses” -> “discusses”

We have changed it accordingly.

- L99 and L110. What is the point of indicating the date of last access for a dataset that is only internally available?

It is just on the command of the journal.

- L109. Sanders and de Haan (2016) is not in the references.

We have added it in the reference.

- L125. Earlier you said that the TROPOMI product has a "scene albedo" A sc. What is the difference between A sc and A s? Then later, at L168, you say that you filter your data for A sc. Does this come from TROPOMI or from OMI then? I don't get it, I think all this is confusing.

The scene albedo (A_sc) is the total albedo of the scene (contributed by clouds, aerosols, surface, etc.) while the surface albedo (A_s) is only the albedo of surface. A_sc comes with TROPOMI L2 UVAI product, while the A_s is not provided in this product. Instead we use A_s from OMI climatology. For the radiative transfer simulation of UVAI, A_s is required rather than A_sc.

We used to A_sc to filter our data in order to reduce impacts of clouds. Now in the revised manuscript, we use the TROPOMI FRESCO cloud support product to filter the clouds (Section 2.1.2). The pre-processing criteria has been changed into*: θ_0 larger than 75°, $UVAI_{354,388}$ smaller than 1, AOD_{550} smaller than 0.5 or CF larger than 0.3. (line 151-152)*

- L142, Dubovik et al. (2000), Dubovik and King (2000) are not in the references.

We have added them in the reference.

- L165-166. While the reason for excluding large SZAs looks clear, why are the other two criteria introduced? Please discuss.

The other two criteria are to exclude effects due to non-absorbing compositions and lower measurement confidence (smaller aerosol signal).

The criteria in the revised manuscript also includes the FRESCO cloud fraction <= 0.3 to reduce effects from clouds: *Before implementing radiative transfer calculations, pre-processing excludes pixels meeting at least one of the following criteria: θ_0 larger than 75°, $UVAI_{354,388}$ smaller than 1, AOD_{550} smaller than 0.5 or CF larger than 0.3. (line 150-152)*

- L181, "a strong spectral dependence . . . aerosols" -> "absorption by biomass burning aerosols in the near-UV has a strong spectral dependence".

The sentence has been changed into: *Many studies have shown evidence that absorption by biomass burning aerosols in the near-UV band has a strong spectral dependence (Kirchstetter et al., 2004; Bergstrom et al., 2007; Russell et al., 2010). (line 115-116)*

- L199, "by the testing data" -> "on the testing data"

The sentence has been changed accordingly.

- Feature selection. It looks to me like you decided to train the SVR using only quantities that have a strong linear correlation to the SSA. In this way, though, you may be discarding some quantities that have some nonlinear relationship to the SSA which does not show up in the linear correlation coefficient. Please discuss.

We have replaced the Pearson correlation coefficient with the Spearman's rank correlation coefficient in the revised manuscript. The Pearson correlation assesses linear relationships, while the Spearman correlation assesses monotonic relationships (whether linear or not). The feature selection is re-written in Section 3.2.

- L209-L210. Please explain the reasons behind these filters for UVAI and ALH.

We used to exclude samples with UVAI < 0.8 and pixels with extreme high ALH but low UVAI, in order to exclude situations where strong absorbing aerosols layering at low altitude (because the California fire 2017-12-12 is elevated plume). But in other cases (added in the revised manuscript), where aerosol layering are more close to the surface. As a result, in the revised manuscript, no constraint on UVAI and ALH applied.

The criteria has been slightly changed in the revised manuscript, where only SZA and clouds are considered: *OMI pixels with θ_0 larger than 75° or cloud fraction larger than 0.3 are excluded.(line 252-253)*

- L246-248, sentence "This is realized . . . predicted". You want to replace the OMI ALH with a value that is closer to the one that would have been retrieved by TROPOMI. But then why is OMI the target and TROPOMI the input? I was expecting it to be the other way around.

This step is no longer applicable as we have deleted this part to avoid confusions. There is only one SVR model in the revised manuscript, which is the SVR trained by the OMAERUV-AERONET joint data set, with UVAI and ALH from OMAERUV, and AOD and AAOD from AERONET. More description can refer to the overview at the beginning of this document.

- L248-249, sentence "It should be noted . . . SVR". Please discuss why have you chosen to train this ALH-adjusting SVR on the Thomas fire and not on the dataset for the AAOD retrieval SVR.

Similar to the previous response, this step is no longer applicable as we have deleted this part to avoid confusions. More description can refer to the overview at the beginning of this document.

- L260. I don't get what you mean by "We fit the SVR for AAOD prediction to both data sets".

Similar to the previous response, this step is no longer applicable as we have deleted this part to avoid confusions. More description can refer to the overview at the beginning of this document.

- L262-264. I am lost here. Up to this point I was convinced that you trained two SVMs: one to adjust OMI ALH to the TROPOMI value and one to predict AAOD from UVAI, ALH and AOD, and that the goal of the ALH-adjusting SVM was to allow the use of OMI data to train the SVM for TROPOMI. Now I learn that there is a third SVM. It looks to me like this sentence contains new information, so it does not just "summarize the section". Please make sure that this is better explained in the paper, because it makes it really difficult to follow the discussion.

As described at the beginning of this document, there is only one SVR model in the revised manuscript. The step that adjust ALH in OMAERUV to the TROPOMI value is no longer applicable. Although in the previous version manuscript, the adjusted ALH leads to slightly better SSA retrieval, but the retrieval from SVR trained by the original OMAERUV ALH is acceptable enough.

The re-written SVR content is in Section 3, and the procedure of SVR is summarized in flow chart Figure.5.

- L273, "the nonlinear transformation" -> "a nonlinear transformation"

The sentence has been changed accordingly.

- L275. Either shed some light on the connection between the concept of kernel and the training of SVMs, or avoid mentioning kernels at all.

The kernel function is described in Section 3.1 in the revised manuscript. The kernel function is a property to solve either linear or non-linear problems, depending on the function types.

- L275. You should make it clear that the Mercer theorem sets the conditions for a function to be admissible as a kernel in positive semi-definite a SVM (basically, it says that the function should give rise to a positive-definite kernel matrix).

The sentence has been changed into: *where $K(x_i, x)$ is the kernel function that is positive semi-definite in order to satisfy Mercer's theorem. (line 234)*

- L280. At line 276 you start the paragraph with "It is clear that", but actually point 3 is not clear at all from what you say. Nowhere before this line have you introduced the concept of support vector, nor have you explained what you mean by its "influencing area".

The introduction on SVR and its relevant concepts, 'influencing area', 'support vector', etc. are in Section 3.1 in the revised manuscript.

- L282. It would be better to move Section B of the supplement to an appendix in the main paper. Supplement should be used for additional figures and data, not for theoretical explanations.

We have moved this content as part of manuscript in Section 3.4 SVR hyper-parameter tuning.

- L282-283. Before saying that you are using radial basis function kernels, it may be useful to say that these are among the functions that satisfy Mercer's theorem. You can do this at the end of the previous paragraph (L276). Also, I would advise to write down the expression of the RBF kernel, so that the reader can better appreciate what is the parameter sigma that you mentioned earlier.

The expression of RBF kernel is in Section 3.4 Equation (11).

- L328. I get a bit confused by the distinction between the validation pixels and the rest of the plume. Are the validation pixels those in the small horizontal strip near the AERONET site in Fig. 9? You may want to indicate that in the paper.

Yes, the mean values of these pixels are used to compare with AERONET. We have replaced the validation pixels with AERONET-collocated pixels. The collocation is the distance within 50 km and the time difference within 3 hours: *Then OMI observations are considered as collocated with an AERONET site if their spatial distance is within 50 km and their temporal difference is within 3 hours. (line 253-255)*

- L352, "trained by the adjusted ALH" -> "trained using the adjusted ALH".

The sentence is no longer applicable.

- L353, "to quantify" -> "of quantifying"

The sentence has been changed accordingly.

- L366, "representative" -> "well known"

The sentence has been changed accordingly.

- P10, References. The first reference looks incorrectly formatted.

The reference format has been changed.