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 Editor's comments

OVERALL

The manuscript deals with a hot topic, i.e. the constraining the aerosol Single Scattering Albedo (SSA) using satellite observations. This quantity is difficult to capture by observations and at the same time very important regarding the radiative forcing of aerosol. The proposed scheme to infer SSA information is new and of interest to the AMT community. The manuscript is suitable for publication in AMT after the issues below are addressed.

GENERAL COMMENTS

The manuscript does not read smoothly. Formulations need to be improved in many instances. The specific issues listed below cover a number of these instances but the list is not exhaustive.

The study is dealing with a single case (a plume of one specific emission event). It needs to be discussed how robust are findings.

We have more case studies for the SVR algorithm to prove its capability of SSA retrieval (Section 3.6 in the revised manuscript).

The study approach needs to be explained upfront more clearly. The choices regarding the source of data (UVAI, ALH, AOD, SSA) used for training the SVM-based scheme, for evaluating the SVM-based algorithms, and for evaluating the RTM-based algorithms, needs to be clarified (at a high level upfront, in detail in the specific sections).

In the last version manuscript, the support vector regression (SVR) method was not well-demonstrated in the manuscript, as we planned to more focus on the implementation and results.

But in the revised manuscript, we have restructured the manuscript. We have separated the RTM part (Section 2) from SVR part (Section 3). The SVR section now includes: theory of SVR (Section 3.1), feature selection based on OMAERUV-AERONET joint data set (Section 3.2), the training and testing data set (Section 3.3), the hyper-parameters tuning of SVR model (Section 3.4), the error analysis of SVR model (Section 3.5) and case applications (Section 3.6).

SPECIFIC COMMENTS

Line 12: It is not clear how SSA is retrieved, which algorithm is employed. Reference to "conventional radiative transfer simulations" is not sufficient.

It is not a specific algorithm. We fixed all other inputs in radiative transfer model except for the imaginary part of refractive index, then find the SSA by minimizing the difference between satellite retrieved UVAI and model simulated UVAI.

This sentence has been changed into: *In the first experiment, we retrieve SSA by minimizing the UVAI difference between observed ones and that simulated by a radiative transfer model. (line 11-13)*

Line 13: The approach to constraining the SSA retrieval is not clear. Is the ALH fixed in the forward model used in the SSA retrieval?

Yes, the ALH is taken from TROPOMI measurement, which is fixed. We used to try this method to retrieve SSA but the ALH is unknow for most cases, causing large uncertainties in SSA (Sun et al., 2018).

In the revised manuscript, we have rephrased the abstract: *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)*

Line 17: The sentence "In the second part of this paper, we propose. . ." is not clear. Clarify that the method relies on an empirical relation that has been established based on long-term datasets of UVAI, ALH and AOD based on the SVR concept. The term "data-driven" is misleading.

This sentence has been changed into: *In the second part of this paper, we propose an alternative method to retrieve SSA based on long-term record of collocated satellite and ground-based measurements using the support vector regression (SVR).* (*line 16-18)*

Line 20 (also caption Figure 8): AERONET does not "measure" SSA directly but retrieves it. Reformulate.

We have reformulated the term through the manuscript.

Eq. 1 is unclear and the variables are not introduced. Without more information the reader cannot guess how to interpret the superscribed labels "obs" and "Ray". It is recommended to explain the UVAI concept and highlight that the obtained index is sensitive to elevated absorbing aerosol.

We have added descriptions for each symbol. (line 31-35)

Line 41: What is meant by "with various spectral choices"?

It means AOD product is available at many wavelengths. As we consider it has nothing to do with comparison of ALH data availability (i.e. ALH is not wavelength-dependent). This sentence has been changed into: *There are plentiful AOD products with wide spatial-temporal coverage. (line 46)*

Line 68: "quantitatively determine" –> quantify.

This sentence has been changed accordingly: *Now with the operational TROPOMI ALH constraining forward simulations, it is expected to partly reduce the SSA retrieval uncertainty meanwhile quantifying the influence of assumed aerosol properties on the retrieved SSA. (line 67-68)*

Line 72: data-driven –> empirical. Proposed to reformulate "We propose an empirical [. . .]. ML algorithms learn [. . .]".

This sentence has been changed accordingly: *In the second experiment, we therefore propose an empirical method to predict aerosol absorption, based on the long-term records of collocated UVAI, ALH, AOD and absorbing aerosol optical depth (AAOD) using machine learning (ML) techniques. (line 70-72)*

Line 77: one piece of information is missing: does the training of an SVM requires less training data than ANN?

Yes, the SVM requires less training data than ANN method. We have added reference on this information: *Compared with other algorithms (e.g. the Artificial Neural Network), SVR is less sensitive to training data size and can successfully work with limited quantity of data (Mountrakis et al., 2011; Shin et al., 2005). (line 81-83)*

Line 77: ML and SVM seem to be used interchangeably, which not entirely correct: also ANN can be seen as ML tools.

This sentence is no more applicable. In the revised manuscript, we only use the term SVR in the parts related to SSA retrieval.

Line 79, 81: inconsistent use of singular and plural

We have uniformed the term into ‘SVM’ (support vector machines). SVR is a variant of SVM to solve regression problems. In the revised manuscript, we only use SVR.

Line 84: The term kernel functions is used as if it had been introduced already. Are these related to the support vectors?

Yes, the kernel functions are related to the SVR. The kernel function is an option for SVR method to solve either linear or nonlinear problems depending on whether the kernel function types. The parameters of a kernel function are determined during the training process. We have added introduction of SVR and its kernel in Section 3.1, the kernel hyper-parameter is determined in Section 3.4.

Line 107: reformulate “TROPOMI ALH retrieval is based on the pattern . . .”

This sentence has been changed into*: TROPOMI ALH is retrieved at oxygen A-band (759-770 nm), where the strong absorption of oxygen causes the highly structed spectrum. This feature is particularly suitable for elevated optically dense aerosol layers (Sanders et al., 2015; Sanders and de Haan, 2016) (line 133-135)*

Line 102 “For the forward radiative transfer calculations, the input aerosol profile is parameterized as . . .” is this choice consistent with the assumptions made in the ALH algorithm?

Yes, the same setting as ALH algorithm.

We have added reference on this information: *For the forward radiative transfer calculations, the input aerosol profile is parameterized according to the settings in ALH retrieval algorithm: a one-layered box shape profile, with central layer height derived from TROPOMI and an assumed constant pressure thickness of 50 hPa (Sanders and de Haan, 2016). (line 136-138)*

Line 129: The relevance of the reference to Herman & Celarier is not clear. Does the statement “A spectrally flat As is assumed . . .” apply to the OMI LER product? Or do you need to make this assumption?

In the previous version manuscript, I made this assumption as the wavelength dependence of the surface reflectivity between 340 and 380 nm is little (0.2%), but it is proved to be not generally true.

Thus in the revised version, I use the spectrally dependent surface albedo. Although in the radiative transfer calculation, due to the round-off, the results may not be significantly changed.

Section 3.1 falls short on an explicit and upfront specification of the source of the input data (such as AOD, ALH, UVAI) used for the RTM-based method. A discussion of temporal mis-registration between MODIS and TROPOMI data acquisitions is missing.

The content is now moved in Section 2.1.2. The UVAI and ALH come from the TROPOMI level 2 product and the AOD comes from the MODIS/AQUA level 2 collection 6. AQUA has a similar overpass time to that of S-5P (around 13:30 local time), which already has been included in the manuscript. The time difference in this case is only several minutes.

Line 156/157: it is not clear what are the implication of using surface reflectance data from the OMI LER for reproducing the UVAI using the RTM method. Please discuss. It is assumed that the surface reflectance generated within a UVAI product can be reproduced in a straight forward fashion if needed.

OMI LER surface reflectance is one of the inputs for radiative transfer calculation of UVAI. As currently surface albedo is not included in TROPOMI L2 UVAI product, we use OMI climatology instead (introduced in Section 2.1.2).

Line 161: The justification of the reporting wavelength of the retrieved SSA is not understood; in the end it is determined by the OMAERUV reporting wavelength?

The SSA retrieved by radiative transfer simulations can be reported at any wavelength you want only if you specify it in the configuration file used to run the radiative transfer model. We report SSA at 500 nm because the SSA retrieved by SVR is at this wavelength (both the OMAERUV SSA and AERONET SSA are available at 500 nm).

Line 173: Aerosol models cannot be a combination of a project and an algorithm. Rephrase.

This sentence has been changed into: *The aerosol models used for the Mie calculations are a combination of the aerosol models in ESA Aerosol_cci project (Holzer-Popp et al., 2013) and that in the OMAERUV algorithm (Torres et al., 2007; Torres et al., 2013). (line 106-107)*

Line 175: the phrase "The particle size distribution ..." needs grammatical/syntactic fixing

This sentence has been changed into: *We use the particle size distribution of the fine mode strongly absorbing aerosol of ESA Aerosol_cci project. The geometric radius (r_g) is 0.07 µm (effective radius r_{eff} of 0.14 µm) and the geometric standard deviation (σ_g) is 1.7 (logarithm variance $\ln\sigma_g$ of 0.53). (line 108-111)*

Line 179: subtype "BIO-1" is referred to without explanation/reference Line 177: real PART OF THE refractive index

This sentence has been changed into: *The real part of the refractive index (n) uses the same value as in the OMAERUV algorithm, which is set to be 1.5 for all subtypes and spectrally flat. We adopt the imaginary part of the refractive index at 388 nm (κ_{388}) of the OMAERUV smoke subtypes (except for BIO-1 whose κ_{388} is 0) in our study and add a subtype with κ_{388} equaling to 0.06. (line 111-114)*

Line 178: imaginary PART OF THE refractive index

See previous response.

Line 181: Sentence incomplete

This 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)*

Line 182 (also caption Figure 8): The specification of $\Delta\kappa$ in not clear. Clarify.

We actually explain the $\Delta\kappa$ in the latter part of this sentence: $\Delta\kappa$ is defined as the relative difference between κ_{354} and κ_{388}. We have added a formula in the revised manuscript: $\Delta\kappa = (\kappa_{354} - \kappa_{388})/\kappa_{388}$ *(line 118, Figure.4 caption and Table 1 title)*

Table 1: The specification of the imaginary part of the refractive indices $\Delta\kappa$ is unclear. For which reference wavelength are the numbers in the rightmost column valid? In the column "Refractive index imaginary part at 354 nm (k354) one expects an explicit list of values rather than a formula. Clarify.

The reference wavelength is 388 nm. $\Delta\kappa$ is the relative difference between κ_{354} and κ_{388} $((\kappa_{354} - \kappa_{388})/\ \kappa_{388})$. Also see previous response.

Here we select 9 different $\Delta\kappa$ values from 0% to 40% and 7 different κ_{388} values from 0.005 to 0.060, thus overall 63 values of κ_{354}. It is trivial to list all values of the refractive index imaginary part at 354 nm.

Line 186: Absorbing Ångström Exponent –> Ångström exponent

It has been changed accordingly through the manuscript. (line 121)

Line 207: "13-year measurement OMAERUV and AERONET measurements" rectify formulation.

This sentence has been moved to Section 3.2: *To start with, we collect the measurements of OMAERUV version 3 product (http://dx.doi.org/10.5067/Aura/OMI/DATA2004 last access: 17 October 2018) and AERONET version 3 level 1.5 inversion product (https://aeronet.gsfc.nasa.gov, last access: 4 June 2019) from 2005-01-01 to 2017-12-31. (line 250-252)*

Line 210: an OMI pixel is collocated –> OMI observations are considered as collocated

It has been changed accordingly: *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)*

Section 3.2.2 (Preparing training and testing data sets) is quite confusing. Please rewrite. Some terminology is used inconsistently (e.g. the terms "extra SVR", "adjusted ALH", "predicted ALH", or "ALH from OMAERUV" and "ALH from OMI"). Maybe introduce Table 2 and Flow charts (Figure 5) already at the beginning of the section.

As declared at the beginning of this document, the structure of the manuscript is changed. The original preparing training and testing data sets is moved to Section 3.2 and Section 3.3. The flow chart is still Figure.5.

Line 240, 259, 326: It is referred to "ALH from the OMAERUV" suggesting that ALH is generated by the OMAERUV algorithm. It is stated in the manuscript that these ALH values are actually taken from CALIOP. Please refer to "ALH from CALIOP" for clarity.

The ALH in OMAERUV product is not entirely from CALIOP climatology. Actually, we have explained that ALH in the OMAERUV product is a combination of CALIOP climatology and assumed ALH in the AOD retrieval (if the CALIOP climatology is not available): *The best-guessed ALH in the OMAERUV is either from CALIOP climatology or assumed ALH in the retrieval (if the CALIOP climatology is not available) (Torres et al., 2013). (line 246-247)*

The ALH from the OMAERUV just indicate the ALH provided in the OMAERUV. If use ALH from CALIOP, it seems that we take ALH directly from the CALIOP product, which is not the case.

Line 240: It is stated that the ALH from the OMAERUV product (actually from CALIOP) may not have sufficient quality. Clarify what is the concern. Co-location?

The co-location is one of the concerns, considering the OMI measurements are significantly affected by the row anomaly and the limited swath of CALIOP. Moreover, as described in the precious response, the ALH provided in OMAERUV product is not a real retrieval from measurements, but a best-guess based on CALIOP climatology and a priori assumptions.

We have added comment on the OMAERUV ALH: *As a result, one should keep in mind that the ALH from OMAERUV may suffer from the uncertainties of CALIOP climatology and a priori assumptions, and collocation error between OMI pixels and CALIOP footprint. (line 247-249)*

Line 242: What is meant with the OMI ALH? Is it the same as the ALH from the OMAERUV (CALIOP)?

It refers to the ALH from OMAERUV product. In the revised manuscript, we use the term OMAERUV ALH to make it clear.

Line 245: Is the TROPOMI ALH the one retrieved from the O2-A band?

Yes, TROPOMI ALH product is retrieved from the Oxygen A-band.

Line 249: It is stated that the "extra" SVR is trained on the Thomas fire case. A training sets should cover more than one case. Please discuss the validity of the approach.

As declared at the beginning of this document, the 'extra' SVR in the previous version manuscript is removed. We only use the OMAERUV-AERONET joint data source to train the SVR model and retrieve the SSA.

Line 253: it is noted that this "extra" SVR is a temporary intermediate step to obtain a better ALH". Please explain upfront the approach.

This part is no longer applicable as we have removed it from the revised manuscript. See the previous response and the description at the beginning of this document.

Line 255: It is stated that "there is no necessity to do this anymore once a reliable ALH product is accessible to build up training data sets, e.g. the TROPOMI ALH product that will be released in the near future". Clarify in which sense the training set using TROPOMI ALH is expected to outperform the training set using CALIOP ALH.

This part is no longer applicable as we have removed it from the revised manuscript. See the previous response and the description at the beginning of this document.

Line 259: What is meant with "The rule of thumb ratio is 70% versus 30%"? Eq. 3: Introduce the variable n.

This introduction on training/testing data set separation is now in Section 3.3. The empirical ratio to divide a data set into a training set (to training the SVR model) and a test set (to evaluate the generalization performance of the trained SVR model) is 70% to 30%.

n in the Eq.(3) is the number of samples. We have added this explanation in line 219.

Eq. 3: What is the dimensionality of omega? What is meant with ||omega||? Some kind of norm?

This content has been moved to Section 3.1, where we have added a new section briefly explaining the theory of SVR. ||…|| denotes the norm. (line 229)

Eq. 4: Introduce the variable x.

See Section 3.1. We have added this explanation in line 229.

Eq. 4: Why introduce the kernel function K? What is done with it?
Section 3.2.4 Data for case application: Please report the number of validation samples

The kernel function is aimed to solve either linear or nonlinear problems, depending on the kernel function types. This is introduced in Section 3.1 (line 232-235).

The number of samples in case applications are listed in Table 2 (California fire on 2017-12-12) and Table 4 (other case applications).

Figure 5: The figure shows at the same time SVR based ALH prediction and SVR based AAOD prediction, this is confusing. It would help to depict the two schemes for AAOD prediction in one flow chart, and the ALH prediction in a separate one.

As described at the beginning of this document, there is only one SVR model left, and Figure.5 has been changed accordingly.

Figure 6 (also Line 206): Why is the sign of the correlation coefficient not reported? Report the sign or justify and clarify in caption that |rho| is reported.

The priority is given to the magnitude of correlation rather than the sign, it is not a problem to show sign though. Besides, we use the Spearman's rank correlation coefficient instead of Pearson's correlation coefficient in order to deal with the non-linearity among different parameters.

Figure 6: For which ALH parameter are the correlations reported? For the predicted one or for the one from CALIOP?

The ALH in the training data set, i.e. the ALH reported in the OMAERUV product. The description on parameters in Figure.6 is in Section 3.2 line 261-263: *The parameters in OMAERUV-AERONET joint data set for feature selection consists of UVAI calculated by 354 and 388 nm wavelength pair, satellite geometries, surface conditions and ALH from OMAERUV, and SSA, AOD and AAOD from AERONET.*

Figure 7: The 3D plot is hard to interpret. Recommended to replace it with 2D scatter- plots (AOD versus ALH) where the UVAI is only color-coded.

As described at the beginning of this document, there is only one SVR model left, and this figure is no longer necessary to show. We have removed it from the revised manuscript.