



The role of aerosol layer height in quantifying aerosol absorption from ultraviolet satellite observations

Jiyunting Sun^{1,2}, Pepijn Veefkind^{1,2}, Swadhin Nanda^{1,2}, Peter van Velthoven³, Pieternel Levelt^{1,2}

¹ Department of Satellite Observations, Royal Netherlands Meteorological Institute, De Bilt, 3731 GA, the Netherlands

5 ² Department of Geoscience and Remote Sensing (GRS), Civil Engineering and Geosciences, Delft University of Technology, Delft, 2628 CD, the Netherlands

³ Department of Weather & Climate Models, Royal Netherlands Meteorological Institute, De Bilt, 3731 GA, the Netherlands

Correspondence to: Jiyunting Sun (jiyunting.sun@knmi.nl)

10 **Abstract.** The purpose of this study is to demonstrate that S-5P/TROPOMI aerosol layer height (ALH) observations can be used to improve the single scattering albedo (SSA) retrieval from ultraviolet satellite observations. We take the Thomas Fire in southern California on 12 December 2017 as a case study. In the first part of this paper, we apply conventional radiative transfer simulations to retrieve the SSA. With forward simulations constrained by TROPOMI ALH, we can determine the uncertainty in SSA due to the assumed spectral dependence of refractive indices between two wavelengths of the near-
15 ultraviolet absorbing aerosol index (UVAI). A significant gap in the retrieved SSA (0.24) between radiative transfer simulations with ‘grey’ and ‘colored’ aerosols implies that inappropriate spectral dependences may cause severe misinterpretations of aerosol absorption. In the second part of this paper, we propose a data-driven method to quantify aerosol absorption from long-term measurements of UVAI, the aerosol optical depth (AOD) and ALH using support vector regression (SVR). We present the potential of TROPOMI ALH in this new method. The SVR predicted SSA (0.96 ± 0.01)
20 outperforms that predicted by radiative transfer simulations (0.90 ± 0.05), considering the AERONET SSA measurement is 0.96 and assuming that the aerosol absorption should be homogeneous within the plume (i.e. small SSA standard deviation). We thus believe that the upcoming TROPOMI ALH product can make it feasible to quantify aerosol absorption via data-driven methods, which would play an important role in constructing a long-term global SSA data set.

1 Introduction

25 The concept of the near-ultraviolet (near-UV) absorbing aerosol index (UVAI) initially came along with the ozone product of the Nimbus 7/Total Ozone Mapping Spectrometers (TOMS). It detects UV-absorbing aerosols by measuring the spectral contrast difference between a satellite observed and a model simulated Rayleigh atmosphere for a given wavelength pair (λ and λ_0) (Herman et al., 1997):

$$UVAI = -100 \left(\log_{10} \left(\frac{I_{\lambda}}{I_{\lambda_0}} \right)^{obs} - \log_{10} \left(\frac{I_{\lambda}}{I_{\lambda_0}} \right)^{Ray} \right), \quad (1)$$

The over four-decade heritage UVAI observations (1978 to present) has been widely used for aerosol research. It would be
30 beneficial to have a quantitative relationship between UVAI and other aerosol absorption properties that do not have such long-term global records, e.g. the single scattering albedo (SSA), which is the ratio of aerosol scattering to aerosol extinction. Aerosols are considered as the largest error source in radiative forcing assessments (IPCC, 2014), and SSA is one of the key parameters to reduce this uncertainty (Haywood and Shine, 1995).

The most straightforward approach to derive a relationship between UVAI and quantitative aerosol absorption properties like
35 SSA is through forward radiative transfer simulations. Lookup tables (LUTs) of simulated UVAI for various measuring geometries, aerosol properties, atmospheric and surface conditions are constructed by radiative transfer models (RTMs). Then SSA is derived by minimizing the difference between pre-calculated UVAI and satellite observed ones (Colarco et al., 2002; Hu et al., 2007; Jeong and Hsu, 2008; Sun et al., 2018). Hereafter, we refer to this method as the RTM-based retrieval.



Apart from pre-assumed aerosol micro-physics, the aerosol loading and the aerosol vertical distribution are two key
40 parameters in forward simulations of UVAI. The former is usually provided in terms of the aerosol optical depth (AOD).
There are plentiful AOD products providing wide spatial-temporal coverage with various spectral choices. By contrast, only
little information on the aerosol vertical distribution is available. The most well-known aerosol profile product is offered by
the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), but the number of measurements is limited because of its
narrow tracks (Winker et al., 2009). Passive sensors only measure columnar quantities but, in some cases, also provide the
45 aerosol layer height (ALH), a compact form of aerosol profile information indicating where most aerosols are located.
Chimot et al. (2017) present the feasibility of ALH retrieval using the oxygen (O_2-O_2) band at 447 nm of the Ozone
Monitoring Instrument (OMI), but so far it has not been run operationally yet.
Recently, a new ALH algorithm based on the near-Infrared (NIR) O_2 A-band has been developed for the Tropospheric
Monitoring Instrument (TROPOMI) on board the Copernicus Sentinel-5 Precursor (S-5P) (Sanders et al., 2015). TROPOMI
50 was launched on 13 October 2017. The instrument is equipped with both the UV-visible (270–500 nm) and the near-infrared
(NIR) (675–775 nm) channels (Veefkind et al., 2015), which makes it possible to interpret UVAI using corresponding ALH
measurements. Furthermore, TROPOMI has a wide swath of 2600 km, providing daily global coverage with a high spatial
resolution of 7×3.5 km² in nadir.

The purpose of this paper is to demonstrate the potential of the TROPOMI ALH product for quantifying aerosol absorption.
55 As the TROPOMI ALH product is not operationally available yet, we focus on the data of one of the largest wildfires that
happened in southern California in 2017, i.e. the Thomas Fire
(http://www.fire.ca.gov/current_incidents/incidentdetails/Index/1922). Ignited on 4 December 2017, the fire was expanded
quickly northwest by the strong and persistent Santa Ana winds and was fully under control on 12 January 2018. The precise
cause of the fire remains unknown, but a prolonged period of heat and absence of precipitation definitely contributed to this
60 devastating fire (<https://inciweb.nwcg.gov/incident/5670/>). We selected one day (12 December 2017) for our case study. As
shown in Fig. 1, a brown smoke plume produced by the Thomas Fire was blown away from the continent and transported
northwards. The major part of the plume was over the ocean with cloud free conditions, which is favorable for space-borne
aerosol observations.

We conduct two experiments to investigate the potential of TROPOMI ALH for quantifying aerosol absorption of the smoke
65 plume generated by the Thomas Fire. First, as concluded in our previous study (Sun et al., 2018), the absence of aerosol
vertical distribution information and an improper spectral dependence of aerosol absorption in the near-UV region can be
responsible for a large difference between estimated and measured SSA. Now with the TROPOMI ALH as a constraint, we
are able to quantitatively determine the influences of assumed wavelength dependent aerosol absorption on retrieved SSA.
Similar to our previous study (Sun et al., 2018), SSA retrieval in the first experiment is conducted by the RTM-based
70 method.

In the second part of this paper, we propose a statistics-based machine learning (ML) approach to predict aerosol absorption.
ML algorithms are data-driven and they learn the underlying behavior of a system from a given training data set. Another
advantage of this kind of method is that no a priori knowledge about the relationship between data is needed. Hence, ML
algorithms are particularly useful to address ill-defined inversion problems in the field of geosciences and remote sensing,
75 where theoretical understanding is incomplete but there is a significant amount of observations (Lary et al., 2015). Various
algorithms have been developed to deal with classification or regression problems, among which artificial neural networks
(ANNs) and support vector machines (SVMs) are most common. In this study we choose the latter due to relatively limited
availability of training data (nevertheless it is still large enough to apply ML techniques).

SVMs is a non-parametric statistical algorithm initially devised by Vapnik (1995) to solve classification problems, and later
80 extended to its regression variant, i.e. support vector regression (SVR) (Drucker et al., 1997). An extensive introduction on
SVR can be found in (Alex J. Smola and Scholkopf, 2004). In summary, SVMs algorithm is suitable to solve problems of



small training data sets with a high-dimensional feature space and can provide excellent generalization performance (Durbha et al., 2007; Yao et al., 2008). Besides, SVMs can efficiently avoid overfitting problems as it only depends on a subset of the training data (ϵ -insensitive loss). The wide choice of kernel functions also allows it to solve different non-linear problems.

85 SVMs have been applied extensively to solve remote sensing problems (Lary et al., 2009; Mountrakis et al., 2011; Noia and Hasekamp, 2018). Our second experiment is constructing a training data set consist of existing measurements of UVAI, AOD and ALH, from which the SVR learns the underlying relationship of the system and predicts the SSA. We will present the efficiency and accuracy of this empirical method by comparing it with the results of the RTM-based method.

This paper is organized as follows: section 2 introduces the data sets involved in this study; the first part of section 3
90 expresses the setup of the RTM-based method and the first experiment, i.e. a sensitivity study of SSA examining assumptions on the spectral dependence of near-UV aerosol absorption; the second part of section 3 focuses on the second experiment: how the training data set is prepared and how SVR parameters are determined; results and main conclusions are presented in sections 4 and 5, respectively.

2 Data sets

95 Data sets listed in this section are either used by the RTM-based method or the SVR-based method, or both. The pre-processing and detailed usage of those data are explained in section 3.

2.1 TROPOMI satellite data

In this study, we employ the TROPOMI Level 2 reprocessed UVAI product to quantify aerosol absorption for the target fire event (TROPOMI UVAI data on 12 December 2017 is only internally available, last access: 19 June 2018. UVAI offline
100 data can be accessed via <http://doi.org/10.5270/S5P-0wafvaf>). The TROPOMI UVAI is calculated for two different wavelength pairs. One uses the conventional 340 and 380 nm to continue the heritage of UVAI records from multiple sensors, and the other uses 354 and 388 nm in order to allow comparison with OMI measurements (D.C. Stein Zweers, 2016). In this study we retrieve the SSA based on the latter pair. Satellite measurement geometries (solar/viewing zenith angle θ_0/θ_v , solar/viewing azimuth angle φ_0/φ_v) and the surface pressure (P_s) included in the UVAI product are input for the radiative transfer calculations. The scene albedo (A_{sc}) from the same product is also used in the pre-processing as will be
105 described later.

TROPOMI ALH retrieval is based on the pattern of a highly structured spectrum with strong absorption of O₂ in the A-band (759–770 nm), which is particularly suitable for elevated optically dense aerosol layers (Sanders et al., 2015; A.F.J. Sanders and J.F. de Haan, 2016). The TROPOMI ALH data is still in a pre-operational phase of development and only internally
110 available (last access: 22 June 2018). The ALH is reported in both altitude and pressure. The ALH data for this fire event is available for pixels with UVAI values (calculated for the wavelength pair of 340 and 380 nm) larger than 1 to exclude pixels dominated by non-absorbing aerosols. For the forward radiative transfer calculations, the input aerosol profile is parameterized as a one-layered box shape profile, with central layer height derived from TROPOMI and an assumed constant pressure thickness of 50 hPa.

115 2.2 MODIS satellite data

The corresponding AOD of the plume generated by the Thomas Fire is taken from the Level 2 product MYD04 (Collection 6) from Aqua/MODIS (http://dx.doi.org/10.5067/MODIS/MYD04_L2.006, last access: 16 March 2018). Aqua has an overpass time similar to S-5P (13:30 local solar time). The AOD at 550 nm used in the RTM-based method is a combination of the Deep_Blue_Aerosol_Optical_Depth_550_Land and the Effective_Optical_Depth_Op550um_Ocean (Levy et al., 2013).
120 The retrieval uncertainty in AOD is only 3%–5% over the oceans whereas it is 5%–15% over land (Remer, 2005).



Note that in the second experiment, the MODIS AOD at 550 nm is converted to 500 nm using the Ångström Exponent (α) provided by the nearby AERONET site (i.e. UCSB, see section 3.4), as AOD in the training data set is reported at 500 nm. The training data set will be described in subsection 3.2.

2.3 OMI satellite data

125 Surface reflectance (A_s) is currently not provided in the TROPOMI UVAI product. Instead, we use the Aura/OMI Level 3
Lambertian equivalent reflectance (LER) monthly climatology calculated from measurements between 2005 and 2009
(Kleipool et al., 2008) (Kleipool, 2010) (<http://dx.doi.org/10.5067/Aura/OMI/DATA3006>, last access: 26 September 2018).
TROPOMI on S-5P and OMI on Aura have similar overpass times (both are approximately at 13:30 local time) and
measuring geometries (Levelt and Noordhoek, 2002) (Veefkind et al., 2015). A spectrally flat A_s is assumed between 354
130 and 388 nm according to Herman and Celarier (1997) (Herman and Celarier, 1997).
The OMAERUV is currently the only product containing a long-term UVAI with corresponding ALH (Torres et al., 2007)
(Torres et al., 2013). It is noted that the ALH reported in the OMAERUV product is not derived from the OMI observations,
but from a combination of climatology derived from the CALIPSO data and from a chemical transport model. In order to
construct the training data set from which the SVR algorithm can learn, we collect the Level 2 OMAERUV version 3
135 product (<http://dx.doi.org/10.5067/Aura/OMI/DATA2004>, last access: 17 October 2018) from 1 January 2005 to 31
December 2017. The training data set will be described in subsection 3.2.

2.4 AERONET data

The retrieved aerosol absorption is evaluated with the version 2 Level 1.5 inversion product (<https://aeronet.gsfc.nasa.gov>,
last access: 13 October 2018) (Holben et al., 1998) from the nearby AERONET site UCSB (119.845°W, 34.415°N) as
140 marked in Fig.1. SSA and the absorbing aerosol optical depth (AAOD) at 500 nm from AERONET are estimated by linear
interpolation of discrete measurements between 440 and 675 nm (Dubovik et al., 2000; Dubovik and King, 2000).
Fig.2 presents the time variation of AAOD, SSA and AOD during December 2017. The high AOD and AAOD indicate that
the AERONET site captured the plume on 12 December 2017. But only 2 samples are available on this day and both were
earlier than the TROPOMI overpass time (from 21:39:54 to 21:43:06 UTC). We choose the latter ground-based
145 measurement (at 19:55:07 UTC) to validate our retrieved SSA.
In addition, we collect the AERONET version 2 Level 1.5 direct sun and inversion product for all stations for the same
period as OMAERUV (1 January 2005 to 31 December 2017), to construct the training data set for the SVR-based method.
The training data set will be described in detail in subsection 3.2.

3 Methodology

150 This section introduces the procedure and technical concerns of the RTM-based method and the SVR-based method. The pre-
processing and detailed usage of data sets mentioned in section 2 are also explained.

3.1 The RTM-based method

Forward radiative transfer simulations are conducted by the KNMI developed radiative transfer model DISAMAR
(Determining Instrument Specifications and Analyzing Methods for Atmospheric Retrieval) (de Haan, 2011). Fig.3
155 illustrates the main inputs and the procedure. For each pixel, first, aerosol optical properties are computed by Mie theory for
pre-defined aerosol models. Then DISAMAR calculates UVAI using the corresponding AOD, ALH, satellite measuring
geometries (θ_0 , θ_v , φ_0 and φ_v), surface and environmental conditions (A_s and P_s) of the target pixel. For the detailed
implementation of these forward simulations, please refer to Sun et al. (2018). The output of the forward simulations is a
LUT of UVAI as a function of the input SSA at 500 nm (determined by the pre-defined aerosol models), which is fit by a



160 second order polynomial function. Finally, by specifying the corresponding satellite observed UVAI, the SSA of the target
pixel is estimated from the UVAI-SSA relationship. The retrieved SSA is reported at 500 nm in order to compare with the
results of the SVR method.

3.1.1 Pre-processing of the satellite inputs

Due to different spatial resolutions, TROPOMI ALH, OMI A_s climatology and MODIS AOD are resampled onto the
165 TROPOMI UVAI grid. Before implementing radiative transfer calculations, pre-processing excludes pixels meeting at least
one of the following criteria: θ_0 larger than 75° , $UVAI_{340,380}$ smaller than 1 (corresponding $UVAI_{354,388}$ smaller than 0.8) or
AOD₅₅₀ smaller than 0.5. As shown in Fig.1, the north part of the plume may be contaminated by underlying clouds, thus
pixels in this region (latitude over $36^\circ N$) with A_{sc} larger than 0.15 are also removed. In the end there 4808 plume pixels are
left. The pre-processed satellite data are presented in Fig.4 (the TROPOMI measurement geometries are given in terms of the
170 scattering angle, Θ). The south part of the plume is the most absorbing region, where the aerosol loading is high (AOD larger
than 2) and the aerosol layer has risen to about 4 km.

3.1.2 Aerosol models and sensitivity study construction

The aerosol models used in this study for the Mie calculations are a combination of the ESA Aerosol_cci project (T. Holzer-
Popp et al., 2013) and the OMAERUV algorithm (Torres et al., 2007; Torres et al., 2013). We assume a fine mode smoke
175 aerosol type and further divide it into 7 subtypes (Table 1). The particle size distribution employs the fine mode strongly
absorbing aerosol of ESA Aerosol_cci project: a geometric radius (r_g) of $0.07 \mu m$ (effective radius r_{eff} of $0.14 \mu m$) and a
geometric standard deviation (σ_g) of 1.7 (logarithm variance $\ln\sigma_g$ of 0.53). The real 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 refractive
index at 388 nm (κ_{388}) of the OMAERUV smoke subtypes (except for BIO-1) in our study and add a subtype with κ_{388}
180 equal to 0.06.

Many studies have shown evidence that a strong spectral dependence in the near-UV band of absorption by biomass burning
aerosols (Kirchstetter et al., 2004; Bergstrom et al., 2007; Russell et al., 2010). Accordingly, a constant 20% $\Delta\kappa$ has been
applied to all smoke subtypes in the recent OMAERUV algorithm (Jethva and Torres, 2011), where $\Delta\kappa$ is defined as the
relative difference between κ_{354} and κ_{388} . In this study, we investigate how the retrieved SSA responds to the assumed
185 spectral dependence by considering 9 different $\Delta\kappa$ values from 0% (i.e. 'grey' aerosols) to 40% (Table 1). This corresponds
to an Absorbing Ångström Exponent (α_{abs}) from 1 to 3.4 and from 1.3 to 4.7, depending on aerosol subtype. Note that the
 $\Delta\kappa$ is only applied between κ_{354} and κ_{388} . Aerosol absorption at wavelengths larger than 388 nm is set equal to that at 388
nm.

To summarize, the first experiment to explore the SSA sensitivity to the spectral dependence of aerosol absorption in the
190 near-UV band consists of 9 cases represented by different $\Delta\kappa$. Within each case, there are 7 pre-defined aerosol subtypes
with varying κ_{388} . Thus, we perform 63 forward simulations in total for each pixel.

3.2 The SVR-based retrieval

In this subsection, we propose an SVR-based method to derive SSA from existing measurements of UVAI, AOD and ALH
as a replacement of the RTM-based retrieval. The procedure is presented in Fig.5. As many other ML algorithms, the
195 major steps of SVR consist of feature selection, training and testing data preparation, hyper-parameters tuning and
application. We first collect parameters relevant for the derivation of SSA from UVAI, from which we then select a subset
based on our knowledge to construct the SVR model (feature selection). The SVR model is then fit to a training data set
containing the selected parameters (training process). The SVR model hyper-parameters are tuned until the generalization



performance of the SVR evaluated by the testing data set is satisfied (hyper-parameters tuning). Finally, the tuned SVR
200 model is used to predict the SSA for our target event (case application). Detailed descriptions on these procedures are
provided in the following subsections.

3.2.1 Feature selection based on OMI and AERONET observations

Although SVR is able to cope with high-dimensional input features, feature selection is still important for generalization
performance, computational efficiency and interpretational issues (Weston et al., 2001). Many sophisticated approaches have
205 been devised for feature selection (Guyon and Elisseeff, 2003). In this study we choose features based on the Pearson
correlation coefficients (ρ) between various parameters in collocated OMAERUV and AERONET measurements and our
empirical knowledge on aerosol absorption.

To start with, we collect the 13-year measurement OMAERUV and AERONET measurements as described in section 2.
OMI pixels with θ_0 larger than 75° , or cloud fraction larger than 0.1, or $UVAI_{354,388}$ smaller than 0.8, or pixels with extreme
210 high ALH but low UVAI are excluded. Then an OMI pixel is collocated with an AERONET site if their spatial distance is
within 50 km and their temporal difference is within 3 hours. To ensure consistency between the different measurement
techniques (ground-based and space-borne), samples are also excluded if the SSA difference between OMI and AERONET
is larger than 0.03, or the AOD difference between OMI and AERONET is larger than 5%. The AERONET SSA and AAOD
are linearly interpolated to 500 nm as OMAERUV reports AOD at this wavelength. In total 4003 samples are left. Fig.A1
215 and A2 in the Supplement, part A show the global distribution and the statistical distribution of the OMI-AERONET joint
measurements, respectively. Note that these are not restricted to biomass burning areas, but may also contain other aerosol
types.

The parameters in OMI-AERONET joint data set for feature selection consists of UVAI, geometries, surface conditions and
ALH from OMI, and SSA, AOD and AAOD from AERONET. Fig.6 presents the correlation coefficients matrix (absolute
220 values, $|\rho|$) of those parameters. Geometries and surface conditions show little correlation with aerosol absorption quantities.
Hence, we will not focus on them. Except for AAOD, SSA is barely dependent on any of the other parameters. This also
points out that the one-to-one numerical relationship between UVAI and SSA in the RTM-based method does not have a
firm physical basis. On the contrary, AAOD is correlated with UVAI and AOD, as it carries information on both aerosol
absorption and aerosol loading. Therefore, it is decided to predict AAOD from given UVAI and AOD and to derive SSA via
225 the relationship in Eq.(2) rather than to directly predict SSA from UVAI.

$$SSA = 1 - \frac{AAOD}{AOD} \quad (2)$$

One may also notice that the UVAI dependence on ALH in Fig.6 is insignificant (the correlation coefficient $|\rho|$ is only 0.3).
This is because the ALH from the OMAERUV product is either as pre-described as other aerosol properties in LUTs or
based on the monthly CALIOP ALH climatology. Another possible reason is the presence of non-smoke aerosols in the joint
data set, for which the relationship between UVAI and ALH may differ from that for biomass burning aerosols. By contrast,
230 the independently observed TROPOMI UVAI and ALH of our target case reveals a much higher correlation ($|\rho|$ is 0.63).
Although this value may be overestimated as it is calculated for a specific case, it seems more reasonable since a strong
UVAI dependence on ALH is well documented (de Graaf et al., 2005; Sun et al., 2018). Under the same conditions, a higher
aerosol layer can shield more Rayleigh scattering beneath the layer, resulting in larger UVAI. Thus, ALH is a necessary
parameter to prevent misinterpretations of aerosol absorption.

235 Based on the above analysis, we construct an SVR model with UVAI, AOD and ALH as input features and AAOD as output
to be predicted. Then the SSA is derived by Eq.(2).



3.2.2 Preparing training and testing data sets

As described in the previous section, the selected features consist of UVAI, ALH, AOD (inputs) and AAOD (output). A good training data set is crucial to ML algorithms, because they learn the underlying behavior of the system from it. One problem we are facing in this study is that the ALH from the OMAERUV product may not have sufficient quality to use. The influence of OMI ALH can be seen by comparing the relationship between 3 input features from TROPOMI-MODIS measurements (Fig.7a) and that from OMI-AERONET measurements (Fig.7b). One may recall that the OMI ALH is climatological whereas the TROPOMI ALH is directly derived from observations. As a result, we propose to create a ‘more observational’ ALH to replace the OMI ALH in the original OMI-AERONET joint data set. We numerically adjust the OMI ALH to the TROPOMI ALH to make the relationship between UVAI, AOD and ALH of OMI-AERONET joint data (Fig.7b) more similar to that of TROPOMI-MODIS data (Fig.7a). This is realized by an extra SVR model, where TROPOMI-MODIS data is the training data set and the ALH in the OMI-AERONET data is the target output to be predicted. We call this intermediate step the SVR for ALH prediction. It should be noted that this SVR is trained on the Thomas fire case, which has no overlap with the training data set for the other SVR.

As can be seen in Fig.5 (indicated in purple), apart from the TROPOMI UVAI and MODIS AOD, the input features for this extra SVR also include the TROPOMI θ_0 and reflectance measured at 354 nm (R_{354}^{obs}) since their correlations with ALH are also high ($|\rho|=0.67$ and $|\rho|=0.61$, respectively) compared to other parameters. Note that this extra SVR is a temporary intermediate step to obtain a better ALH which enhances the dependence of the OMI-AERONET UVAI on the ALH. It is a temporary solution only applied in this study to solve the specific problem of the lack of collocated UVAI and ALH observations. 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.

The ALH predicted by the extra SVR replaces the OMI ALH in the original OMI-AERONET joint data set. Both the original and adjusted OMI-AERONET joint data are partitioned into a training data set and a testing data set, respectively. The rule of thumb ratio is 70% versus 30%. The training data set containing the original OMAERUV ALH and the adjusted ALH are referred to as the original and adjusted training data set, respectively. We fit the SVR for AAOD prediction to both training data sets in order to investigate the importance of a reliable ALH input.

To summarize this section, 3 SVR models are applied in this paper: first, an SVR is used to predict ALH in an intermediate step to improve the quality of the original training data set. This is a temporary solution only applied in this paper due to the lack of observational information concerning aerosol vertical distribution. The remaining 2 SVR models are used to predict AAOD with the original OMI ALH and the adjusted ALH, respectively. Fig.5 shows the procedure of the SVRs for ALH (indicated in purple) and AAOD (indicated in green) prediction. Table 2 also summarizes the input features, output parameter and the corresponding data sources of the 3 SVR models discussed in this sub-section. All input features are scaled into the range between 0 and 1 (min-max normalization) before training.

3.2.3 SVR training process and hyper-parameters tuning

For the mathematical formulation of SVR algorithm one can refer to Smola and Scholkopf (2004). Briefly, SVR tries to find the coefficient ω and the bias b of a linear model by minimizing the function:

$$\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n L_{\varepsilon}(y_i, g(x_i)^T \omega + b) \quad (3)$$

, where y_i is true value and $g(\cdot)$ is the nonlinear transformation that maps x onto a m -dimensional feature space. The loss function in SVR is ε -insensitive loss (Eq.(4)), where ε is the width of insensitive zone within which the error is ignored (Vapnik, 1995).



$$L_{\varepsilon}(y_i, g(x_i)^T \omega + b) = \begin{cases} 0 & \text{if } |y_i - (g(x_i)^T \omega + b)| \leq \varepsilon \\ |y_i - (g(x_i)^T \omega + b)| - \varepsilon & \text{otherwise} \end{cases} \quad (4)$$

275 SVR can be solved by specifying a kernel function $K(x_i, x_j) = g(x_i)^T g(x_j)$ which is positive defined so that the Mercer's theorem is satisfied (Tuia et al., 2011).

It is clear that the generalization performance of the SVR depends on the following hyper-parameters: (1) the width of insensitive zone ε . The cost function does not consider errors in the training data as long as their deviation is smaller than ε , by which SVR can efficiently avoid overfitting issues; (2) the regularization constant C that determines the trade-off between
280 model complexity and the degree to which deviations larger than ε are penalized; (3) the kernel parameter σ that concerns the influencing area of support vectors. We adopt the methodology from (Cherkassky and Ma, 2004), where SVR parameter C and ε can be directly determined from the training data (Eq.(B1) and (B2) in the Supplement). We employ a radial basis function (RBF) kernel to take into account the non-linearity of the SVR models applied in this paper. The kernel parameter σ is determined by hyper-tuning on a testing data set (Durbha et al., 2007). More information on the SVR tuning procedure is
285 given in Part B of the Supplement.

The values chosen by the above methods are robust in our case (Fig.B1-B3 in the Supplement, part B), i.e. retaining a relatively low error while preventing overfitting. Table 3 summaries the settings of the SVR models determined by tuning procedure and the evaluation of the algorithm performance. All 3 SVR models present good generalization capabilities as the differences in root mean square error (RMSE) between the training data and the testing data are minor. The accuracy of the
290 SVR model for ALH prediction is 0.26 km. Fig.7c shows the relationship of UVAI, AOD and the SVR predicted ALH. The structure is more similar to that in Fig.7a and $|\rho|$ between UVAI and ALH increases from 0.30 to 0.61, which is sufficient to mitigate the impact of uncertainties of ALH in the OMAERUV product. Note that this value may be overestimated as the SVR for ALH prediction is only trained by a specific case due to the limited availability of TROPOMI ALH, but it is more reasonable compared with the original UVAI and ALH relationship in the OMI-AERONET data. The predicted ALH,
295 together with OMI UVAI, AERONET AOD and AAOD provides a new training data set for AAOD prediction, i.e. the adjusted training data set. The accuracy of SVR models for AAOD prediction trained by the original and the adjusted training data set are 0.01.

3.2.4 Data for case application

After determining the hyper-parameters, the SVR is applied to predict aerosol absorption in the case of the smoke plume
300 generated by the Thomas Fire event. The input UVAI, ALH and AOD are taken from TROPOMI UVAI from near-UV band, TROPOMI ALH from O₂ A-band and MODIS AOD from visible band, respectively. The predicted AAOD is further used to derive the SSA according to Eq.(2). The retrieved SSA within 50 km from AERONET site (UCSB) is validated. Information about each data product can be found in the corresponding sub-sections in Section 2.

4 Results and discussions

305 4.1 Evaluating the retrieved SSA with AERONET measurements

We first analyze the results of the RTM-based method. Fig.8a displays the mean SSA of all plume pixels retrieved by the RTM-based method as a function of $\Delta\kappa$. The retrieved aerosol absorption decreases with $\Delta\kappa$. This finding is in good agreement with Jethva and Torres (2011). 'Gray' aerosols require stronger absorption to reach the same level of UVAI than 'colored' aerosols. This also explains the high SSA standard deviation (filled area) in the cases with little or no spectral
310 dependence in aerosol absorption. The large variability in retrieved SSA (from 0.71 ± 0.09 to 0.94 ± 0.03) demonstrates that assumptions on the spectral dependence of near-UV aerosol absorption may significantly bias our interpretations of smoke aerosol absorption and should be carefully handled in forward radiative transfer calculations.



The retrieved aerosol absorption is evaluated with the nearby AERONET site (UCSB), whose SSA at 500 nm at 19:55:07 UTC is 0.96. There are 24 TROPOMI pixels located within 50 km from the UCSB site. Hereafter we call them validation
315 pixels. As illustrated in Fig.8b, the mean SSA of the validation pixels also increases with $\Delta\kappa$ and eventually levels off at around 0.96. The extremely low SSA (0.53) and high variation (± 0.37) retrieved for 'gray' aerosols prove that the spectral independence assumption is not recommended for smoke aerosols (at least in this case). The differences between the mean SSA of the validation pixels and the AERONET measurement are shown in Fig.8c. The retrieved SSA falls inside the uncertainty range of AERONET (± 0.03) (Holben et al., 2006) when $\Delta\kappa$ is larger than 15%. This indicates that the
320 assumption of a 20% $\Delta\kappa$ in the OMAERUV algorithm to describe the spectral dependence of aerosol absorption is adequate. Only a slightly better estimate is found when $\Delta\kappa$ equals to 25%, where the mean SSA of validation pixels is 0.002 lower than that of AERONET (Table 3).

On the other hand, the results from both SVR models are also in agreement with the ground-based measurements. SSA retrieved with the original training data set is approximately 0.01 lower than that of AERONET, whereas that predicted by
325 the adjusted training data set shows barely any difference to the AERONET measurements. It is apparent that the adjustment in OMAERUV ALH data leads to an improvement of the performance of SVR algorithm.

4.2 Comparing the results of the RTM-based method and the SVR-based method

From the small set of validation pixels, it is difficult to distinguish the RTM or SVR method is superior, because their results are all satisfactory if one considers the typical AERONET uncertainty (± 0.03). We further investigate the SSA retrieved by
330 the different methods for the entire plume.

As listed in Table 4, the significant difference between the mean SSA of the plume pixels and the validation pixels (0.896 vs 0.957) and the large plume standard deviation (± 0.045) reveal a large spatial variability in the RTM-retrieved results. By contrast, the SSA estimated by SVR presents a more homogeneous spatial feature. The plume mean SSA is similar to that of the validation pixels and both have small variabilities of about 0.01. The difference in spatial pattern is also reflected in
335 Fig.9. The RTM-estimated SSA (Fig.9a) shows strong heterogeneity in the horizontal direction. Although enhancing the spectral dependence of aerosol absorption can reduce the SSA standard deviation to some extent, the spatial variability is still considerable even when applying a $\Delta\kappa$ of 40% (± 0.03 , Fig.8b). Conversely, only small spatial variability is observed in the SSA predicted by the SVR-based method (Fig.9b and 9c).

The strong spatial variability in the RTM-retrieved SSA is mainly controlled by the heterogeneity of the UVAI (Fig.4a).
340 UVAI is a comprehensive variable dependent to many factors, as listed in Fig.3. The spatial pattern of UVAI is passed to the SSA through the one-to-one numerical relationship. This relationship may differ from one pixel to another as the algorithm focuses on one-pixel retrieval each time. However, in the SVR-based algorithms, the spatial variability of the intermediate output AAOD is cancelled out by the 3 input features, rather than dominated by UVAI alone. Furthermore, SVR predicts SSA for each pixel based on the common relationship between UVAI, AOD and ALH in the training data set.
345 SSA is an inherent aerosol attribute that is only determined by the size distribution and the refractive indices. In cloud-free cases, it is expected that micro-physical properties of smoke particles within the plume should be similar over short time periods as they were originated from the same source and generated under the same conditions. We therefore conclude that the empirical SVR-based models outperform the conventional RTM-based retrieval as the SVR-predicted SSA shows a more homogeneous spatial distribution.

350 To summarize, the validation results of both methods are satisfactory with respect to the AERONET uncertainty (± 0.03), nevertheless, the higher homogeneity in SVR-retrieved SSA shows that the SVR-based method is superior to the RTM-based method. The SVR trained by the adjusted ALH also shows a better retrieval compared with that trained by the original OMAERUV ALH. This reveals the feasibility to quantify aerosol absorption based on data-driven methods such as SVR, as



a replacement of radiative transfer calculations. The operational TROPOMI ALH product that is becoming available in the
355 near future will be the key to widely apply this empirical method and to build up a long-term global SSA data set.

5 Conclusions

The purpose of this paper was to investigate the potential role of the TROPOMI ALH product in quantitatively interpreting
the absorption of biomass burning aerosols from near-UV satellite observations. Firstly, we used the TROPOMI ALH to
quantify the influences of assumed spectral dependence of near-UV aerosol absorption (represented by the relative difference
360 between κ_{354} and κ_{388}) on the retrieved SSA. A significant gap in SSA (0.24) between ‘gray’ and ‘colored’ aerosols
($\Delta\kappa=0\%$ and 40% , respectively) demonstrates that inappropriate spectral dependences may significantly bias the retrieved
smoke aerosol absorption.

The TROPOMI ALH product also provides the opportunity to propose an alternative SSA retrieval method. This is a
statistical method based on the correlation between UVAI, AOD and ALH, and requires no a priori assumptions on aerosol
365 micro-physics, which is considered as one of the major error sources in RTM-based method. The new method was realized
by SVR, a representative ML algorithm. The SVR models were trained using 4003 co-located observations from OMI and
AERONET during the period from 2005 to 2017 throughout the world. An adjustment on OMI ALH was implemented to
enhance the quality of training data set. The results of the SVR-based retrieval are more convincing than that of the RTM-
based in terms of retrieved values as well as spatial features. According to the SVR algorithm, the mean SSA of the smoke
370 plume generated by the Thomas Fire on 12 December 2017 is 0.96 ± 0.01 .

The successful SSA retrieval in this study demonstrates the feasibility to quantify aerosol absorption directly from existing
measurements. So far, we have realized this retrieval method by applying SVR which is suitable to the relatively small
training data size. We choose input features based on Pearson correlation coefficients and our knowledge of the subject, and
analytically determined parameters of SVR models from training data set. In the future, more sophisticated feature selection
375 and SVR parameter tuning should be considered to investigate the robustness the algorithm. Moreover, individual training
data sets for different aerosol types may be necessary to enhance the retrieval accuracy. But the most important factor
determining the performance of this method is the availability of qualified training data. We believe that the upcoming
TROPOMI ALH product is key to construct a reliable training data set and will contribute greatly to construct a long-term
global SSA database. Its large amount observations will also boost the possibilities to explore other aerosol SSA retrieving
380 algorithms using ML techniques.

Acknowledgements. This work was performed in the framework of the KNMI Multi-Annual Strategic Research (MSO).
The authors thank NASA’s GES-DISC and LAADS DAAC for free online access of OMI and MODIS data. The authors
thank NASA Goddard Space Flight Center AERONET Project for providing the data from the UCSB AERONET station.
385 The authors also thank Swadhin Nanda for providing TROPOMI aerosol layer height product for this specific fire event.

References

- Alex J.Smola and Scholkopf, B.: A tutorial on support vector regression, *Stat. Comput.*, 199–222,
doi:10.1023/B:STCO.0000035301.49549.88, 2004.
- Cherkassky, V. and Ma, Y.: Practical selection of SVM parameters and noise estimation for SVM regression, *Neural Networks*,
390 17(1), 113–126, doi:10.1016/S0893-6080(03)00169-2, 2004.
- D.C. Stein Zweers: TROPOMI ATBD of the UV aerosol index., 2016.
- Drucker, H., J.C.Burges, C., Linda, K., Alex.Smola and Vladimir, V.: Support vector regression Machines, *Adv. Neural Inf.*
Process. Syst., 155–161, doi:10.1145/2768566.2768568, 1997.



- Durbha, S. S., King, R. L. and Younan, N. H.: Support vector machines regression for retrieval of leaf area index from
395 multiangle imaging spectroradiometer, *Remote Sens. Environ.*, doi:10.1016/j.rse.2006.09.031, 2007.
- Guyon, I. and Elisseeff, A.: An Introduction to Variable and Feature Selection, , 3(03), 1157–1182,
doi:10.1016/j.aca.2011.07.027, 2003.
- de Haan, J. F.: DISAMAR Determining Instrument Specifications and Analyzing Methods for Atmospheric Retrieval., 2011.
- Haywood, J. M. and Shine, K. P.: The effect of anthropogenic sulfate and soot aerosol on the clear sky planetary radiation
400 budget, *Geophys. Res. Lett.*, 22(5), 603–606, doi:10.1029/95GL00075, 1995.
- Herman, J. R. and Celarier, E. a.: Earth surface reflectivity climatology at 340–380 nm from TOMS data, *J. Geophys. Res.*,
102(D23), 28003, doi:10.1029/97JD02074, 1997.
- Herman, J. R., Bhartia, P. K., Torres, O., Hsu, C., Seftor, C. and Celarier, E.: Global distribution of UV-absorbing aerosols
from Nimbus 7/TOMS data, *J. Geophys. Res.*, 102(D14), 16911, doi:10.1029/96JD03680, 1997.
- 405 Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima,
T., Lavenu, F., Jankowiak, I. and Smirnov, A.: AERONET - A federated instrument network and data archive for aerosol
characterization, *Remote Sens. Environ.*, 66(1), 1–16, doi:10.1016/S0034-4257(98)00031-5, 1998.
- Holben, B. N., Eck, T. F., Slutsker, I., Smirnov, A., Sinyuk, A., Schafer, J., Giles, D., Dubovik, O. and Lille, U. S. T. De:
AERONET's Version 2.0 quality assurance criteria, *Int. Soc. Opt. Photonics*, 6408(6408Q), 2006.
- 410 IPCC: Climate Change 2014 Synthesis Report, Geneva. [online] Available from: http://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full.pdf, 2014.
- Jethva, H. and Torres, O.: Satellite-based evidence of wavelength-dependent aerosol absorption in biomass burning smoke
inferred from Ozone Monitoring Instrument, , 10541–10551, doi:10.5194/acp-11-10541-2011, 2011.
- Kleipool, Q.: OMI Surface Reflectance Climatology README, 2010.
- 415 Kleipool, Q. L., Dobber, M. R., de Haan, J. F. and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI
data, *J. Geophys. Res. Atmos.*, 113(18), 1–22, doi:10.1029/2008JD010290, 2008.
- Lary, D. J., Alavi, A. H., Gandomi, A. H. and Walker, A. L.: Machine learning in geosciences and remote sensing, *Geosci.
Front.*, (7), 3–10, 2015.
- Levelt, P. F. and Noordhoek, R.: OMI algorithm theoretical basis document, , I(August), 1–50, 2002.
- 420 Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F. and Hsu, N. C.: The Collection 6 MODIS
aerosol products over land and ocean, *Atmos. Meas. Tech.*, 6(11), 2989–3034, doi:10.5194/amt-6-2989-2013, 2013.
- Remer, L. a.: The MODIS Aerosol Algorithm, Products, and Validation, *J. Atmos. Sci.*, 62, 947–973, 2005.
- Sanders, A. F. J., Haan, J. F. De, Sneep, M., Apituley, A., Stammes, P., Vieitez, M. O. and Tilstra, L. G.: Evaluation of the
operational Aerosol Layer Height retrieval algorithm for Sentinel-5 Precursor : application to O 2 A band observations from
425 GOME-2A, , 4947–4977, doi:10.5194/amt-8-4947-2015, 2015.
- Sun, J., Veeffkind, J. P., Velthoven, P. Van and Levelt, P. F.: Quantifying the single scattering albedo for the January 2017
Chile wildfires from simulations of the OMI absorbing aerosol index, *Atmos. Meas. Tech. Discuss.*, 2(February), 1–24, 2018.
- T. Holzer-Popp, Leeuw, G. de, Griesfeller, J. and Martynenko, D.: Aerosol retrieval experiments in the ESA Aerosol cci
project, *Atmos. Meas. Tech.*, 6(8), 1919–1957, doi:10.5194/amt-6-1919-2013, 2013.
- 430 Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P. K., Veeffkind, P. and Levelt, P.: Aerosols and
surface UV products from Ozone Monitoring Instrument observations : An overview, *J. Geophys. Res.*, 112(D24S47), 1–14,
doi:10.1029/2007JD008809, 2007.
- Torres, O., Ahn, C., Chen, Z. and Space, G.: Improvements to the OMI near-UV aerosol algorithm using A-train CALIOP and
AIRS observations, , 3257–3270, doi:10.5194/amt-6-3257-2013, 2013.
- 435 Tuia, D., Verrelst, J., Alonso, L., Perez-Cruz, F. and Camps-Valls, G.: Multioutput support vector regression for remote
sensing biophysical parameter estimation, *IEEE Geosci. Remote Sens. Lett.*, 8(4), 804–808,



doi:10.1109/LGRS.2011.2109934, 2011.

440 Veefkind, J. P., Aben, I., McMullan, K., Förster, H., Vries, J. De, Otter, G., Claas, J., Eskes, H. J., Haan, J. F. De, Kleipool, Q., Weele, M. Van, Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H. and Levelt, P. F.: Remote Sensing of Environment TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, *Remote Sens. Environ.*, 120(2012), 70–83, doi:10.1016/j.rse.2011.09.027, 2015.

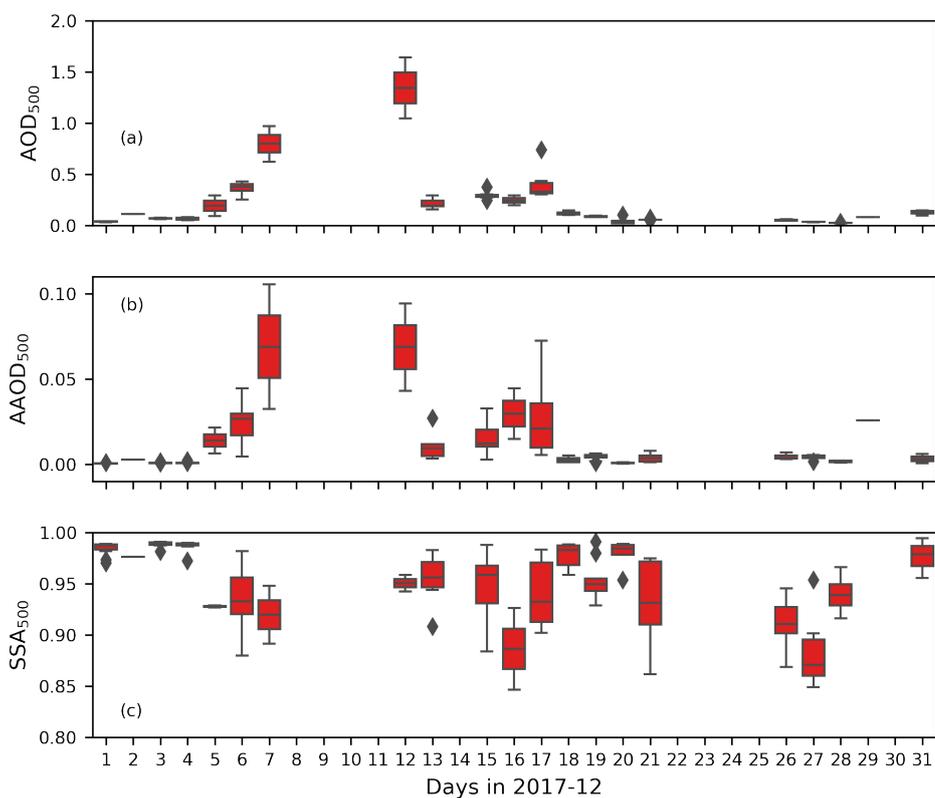
445 Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T. and Vapnik, V.: Feature selection for SVMs, *Adv. Neural Inf. Process. Syst.*, 668–674 [online] Available from: <http://papers.nips.cc/paper/1850-feature-selection-for-svms.pdf>, 2001.

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H. and Young, S. A.: Overview of the CALIPSO mission and CALIOP data processing algorithms, *J. Atmos. Ocean. Technol.*, 26(11), 2310–2323, doi:10.1175/2009JTECHA1281.1, 2009.

450

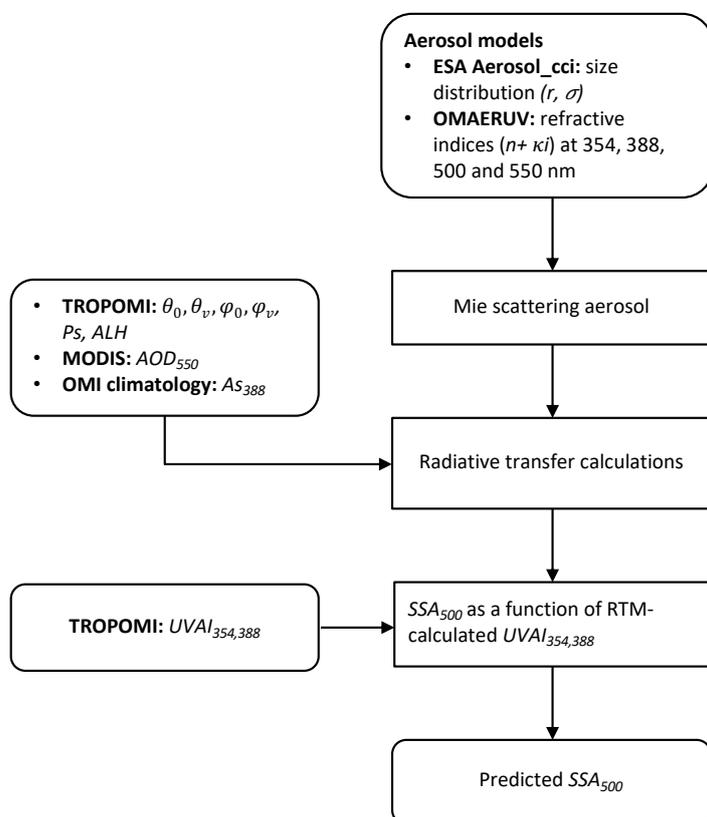


Figure 1: Smoke plume captured by Aqua/MODIS on 12 December 2017 (source: <https://gibs.earthdata.nasa.gov>). The red regions indicate fires and thermal anomalies, the majority of which are the Thomas Fire (119.091°W, 34.415°N). The white dot represents the AERONET site UCSB (119.845°W, 34.415°N).

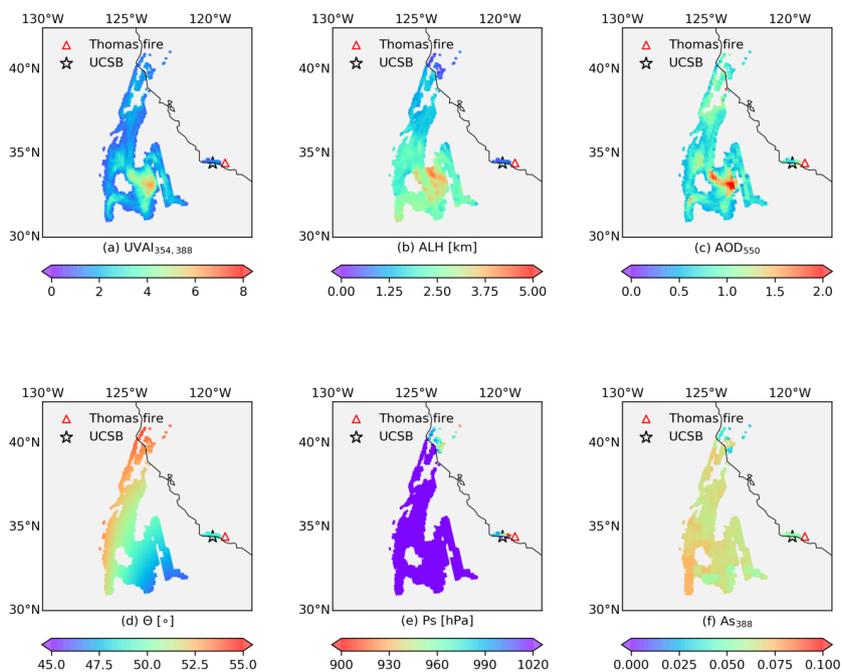


455

Figure 2: AERONET measurements in December 2017: (a) AOD at 500 nm; (b) AAOD at 500 nm; (c) SSA at 500 nm.



460 **Figure 3: Procedure of the RTM-based method: forward radiative transfer calculations and backward retrieval. Note that the aerosol profile is a one-layered box shape profile, with the central layer height set to be the TROPOMI ALH and an assumed constant pressure thickness of 50 hPa.**



465 **Figure 4: Data involved in RTM-based method: (a) TROPOMI UVAI calculated by reflectance at 354 and 388 nm; (b) TROPOMI ALH; (c) MODIS AOD at 550 nm; (d) TROPOMI scattering angle Θ (calculated from θ_0 , θ_v , φ_0 and φ_v); (e) TROPOMI surface pressure P_s ; (f) OMI A_3 climatology at 388 nm. All parameters shown here are projected onto TROPOMI UVAI grid.**

470

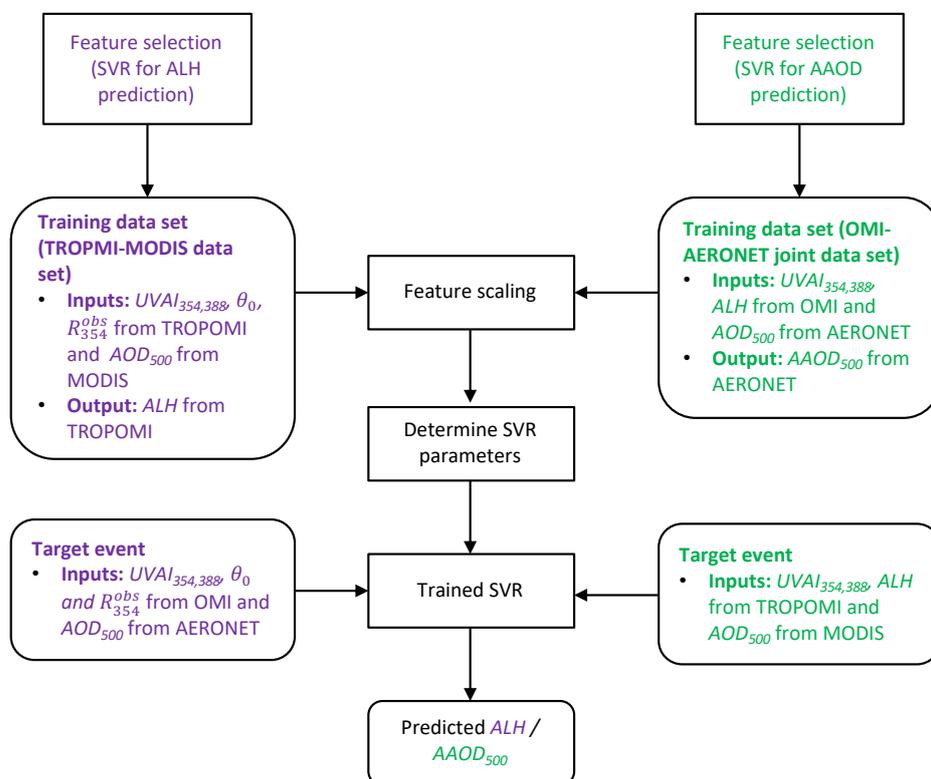
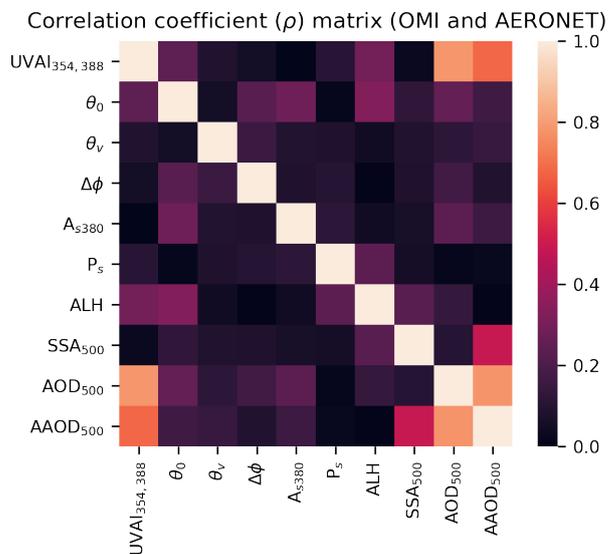


Figure 5: Procedure of the SVR-based method. The purple parts are the SVR for ALH prediction, and the green parts are the SVR for AAOD prediction, respectively. The black parts are general modules for an SVR model.



480 **Figure 6: Correlation coefficient matrix of parameters in the OMI-AERONET joint data set.**

485

490

495

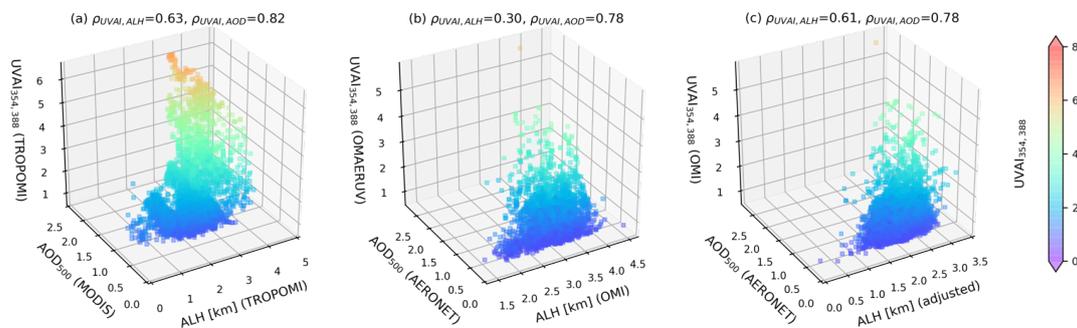


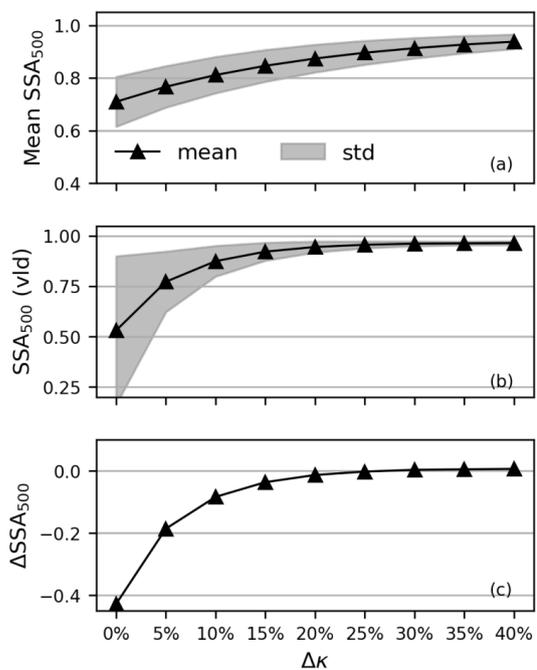
Figure 7: Relationship between UVAI, AOD and ALH: (a) TROPOMI UVAI, ALH and MODIS AOD; (b) OMI UVAI, ALH and AERONET AOD (original training data set); (c) OMI UVAI, adjusted ALH and AERONET AOD (adjusted training data set).

500

505

510

515



520 **Figure 8:** SSA retrieved by the RTM-based method as a function of $\Delta\kappa$ (the relative difference between κ_{354} and κ_{388}): (a) SSA
525 mean and standard deviation (filled region) of 4808 plume pixels; (b) SSA mean and standard deviation (filled region) of the 24
530 validation pixels; (c) absolute difference between the mean SSA of the 24 validation pixels and the AERONET measurement.

525

530

535

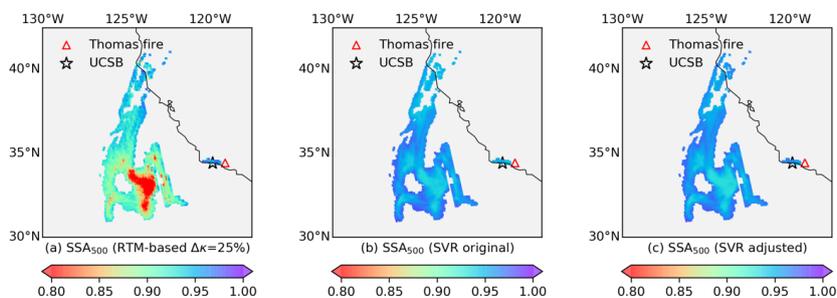


Figure 9: SSA retrieved by the: (a) RTM-based method; (b) SVR-based method with OMAERUV ALH; (c) SVR-based method with adjusted ALH.

540

545

550

555

560

565



570

Table 1 Aerosol models used in forward radiative transfer calculations.

Geometri c radius (r_g)	Effective radius (r_{eff})	Geometry standard deviation (σ_g)	Variance ($\ln\sigma_{eff}$)	Refractive index real part (n)	Spectral dependence ($\Delta\kappa$)	Refractive index imaginary part at 354 nm (κ_{354})	Refractive index imaginary part of other wavelengths (≥ 388 nm)
0.07 μm	0.14 μm	1.7	0.53	1.5	0%, 5%, 10%, 15%, 20%, 25%, 30%, 35% and 40%	(1 + $\Delta\kappa$) $\times\kappa_{388}$	0.005 0.010 0.020 0.030 0.040 0.048 0.060

575

580

585

590

595

600



Table 2 SVR algorithm schemes: input features, output and corresponding data sources.

	Input features	Output	Data source for training and testing	Data source for application	Comment
ALH prediction	UVAI, θ_0 , R_{354}^{obs} , AOD	ALH	TROPOMI (UVAI, θ_0 , R_{354}^{obs}) and MODIS (AOD)	OMI (UVAI, θ_0 , R_{354}^{obs}) and AERONET (AOD)	Temporary solution to improve OMAERUV ALH quality.
AAOD prediction (original training data set)	UVAI, ALH, AOD	AAOD	OMI (UVAI, ALH) and MODIS (AOD)	TROPOMI (UVAI, ALH) and MODIS (AOD)	OMI ALH is used in the training and testing process.
AAOD prediction (adjusted training data set)			OMI (UVAI), adjusted ALH and MODIS (AOD)		Predicted ALH is used in the training and testing process.

605

610

615

620

625

630



635 **Table 3 Settings and results of support vector regression. C , ϵ and σ^2 are the regularization constant, the width of the insensitive zone and the BRF kernel parameter. 3 statistical parameters are given between the truth and the predicted values: the correlation coefficient (ρ), the p-value of T-test, and the root mean square error (RMSE).**

SVR model	SVR model parameters			Training data set			Testing data set		
	C	ϵ	σ^2	ρ	T-test p-value	RMSE	ρ	T-test p-value	RMSE
ALH prediction	4.2571	0.0013	0.25	0.92	0.83	0.25	0.91	0.65	0.26
AAOD prediction (original training data set)	0.1010	0.0002	1	0.81	0.19	0.01	0.80	0.78	0.01
AAOD prediction (adjusted training data set)	0.1010	0.0002	1	0.79	0.19	0.01	0.78	0.48	0.01

640

645

650

655

660

665



Table 4 Retrieved SSA by the RTM-based methods and the SVR-based method. Δ SSA is the SSA difference between retrieved SSA and AERONET measured one.

Retrieval methods	SSA (plume pixels)	SSA (validation pixels)	Δ SSA
RTM-based with $\Delta\kappa=25\%$	0.896 ± 0.045	0.957 ± 0.017	-0.002
SVR with OMAERUV ALH	0.956 ± 0.010	0.949 ± 0.006	-0.010
SVR with adjusted ALH	0.956 ± 0.011	0.959 ± 0.007	-0.000

670

675

680

685

690

695

700

705