AEROCOM/AEROSAT AAOD & SSA study, part I: evaluation and intercomparison of satellite measurements

Nick Schutgens[1], Oleg Dubovik[2], Otto Hasekamp[3], Omar Torres[4], Hiren Jethva[5], Peter J.T. Leonard[6], Pavel Litvinov[2], Jens Redemann[7], Yohei Shinozuka[8,9], Gerrit de Leeuw[10,11], Stefan Kinne[12], Thomas Popp[13], Michael Schulz[14], and Philip Stier[15]

[1]Department of Earth Science, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, the Netherlands
[2]Laboratoire d'Optique Atmosphérique, CNRS/Université Lille, Villeneuve d'Ascq, France
[3]SRON Netherlands Institute for Space Research, Utrecht, The Netherlands
[4]Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[5]Universities Space Research Association-GESTAR, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[6]ADNET Systems, Inc., Suite A100, 7515 Mission Drive, Lanham, MD 20706, USA
[7]School of Meteorology , University of Oklahoma, Norman, USA
[8]Universities Space Research Association, Columbia, Maryland, USA
[9]NASA Ames Research Center, Moffett Field, California, USA
[10]Finnish Meteorological Institute (FMI), Climate Research Programme, Helsinki, Finland
[11]currently at: Royal Netherlands Meteorological Institute (KNMI), R&D Satellite Observations, De Bilt, the Netherlands
[12]Max-Planck-Institut für Meteorologie, D-20146 Hamburg, Germany
[13]German Aerospace Center (DLR), German Remote Sensing Data Center Atmosphere, Oberpfaffenhofen, Germany.
[14]Norwegian Meteorological Institute, P.O.Box 43, Blindern, 0313 Oslo, Norway.
[15]Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, UK.

Correspondence: Nick Schutgens (n.a.j.schutgens@vu.nl)

Abstract. Global measurements of absorbing aerosol optical depth (AAOD) are scarce and mostly provided by the ground network AERONET (AErosol RObotic NETwork). In recent years, several satellite products of AAOD have been developed. This study's primary aim is to establish the usefulness of these datasets for AEROCOM (AEROsol Comparisons between Observations and Models) model evaluation with a focus on the years 2006, 2008 and 2010. The satellite products are super-observations consisting of $1^o \times 1^o \times 30^{\mathrm{min}}$ aggregated retrievals.

This study consists of two papers, the current one that deals with the assessment of satellite observations and a second paper that deals with the evaluation of models using those satellite data. In particular, the current paper details an evaluation with AERONET observations from the sparse AERONET network as well as a global intercomparison of satellite datasets, with a focus on how minimum AOD (Aerosol Optical Depth) thresholds and temporal averaging may improve agreement between satellite observations.

All satellite datasets are shown to have reasonable skill for AAOD (3 out of 4 datasets show correlations with AERONET in excess of 0.6) but less skill for SSA (Single Scattering Albedo; only 1 out of 4 datasets shows correlations with AERONET in excess of 0.6). In comparison, satellite AOD shows correlations from 0.72 to 0.88 against the same AERONET dataset. However, we show that performance vs. AERONET and inter-satellite agreements for SSA improve significantly at higher AOD. Temporal averaging also improves agreements between satellite datasets. Nevertheless multi-annual averages still show systematic differences, even at high AOD. In particular, we show that two POLDER products appear to have a systematic SSA difference over land of ~ 0.04, independent of AOD. Identifying the cause of this bias offers the possibility of substantially improving current datasets.

We also provide evidence that suggests that evaluation with AERONET observations leads to an underestimate of true biases in satellite SSA.

In the second part of this study we show that, notwithstanding these biases in satellite AAOD and SSA, the datasets allow meaningful evaluation of AEROCOM models.

1 Introduction

Aerosol is an important component of the Earth's atmosphere that affects the planet's climate, the biosphere, and human health. Aerosol particles scatter and absorb sunlight as well as modify clouds. Anthropogenic aerosol changes the radiative balance and influences global warming (Angstrom, 1962; Twomey, 1974; Albrecht, 1989; Hansen et al., 1997; Lohmann and Feichter, 1997, 2005). It may negatively affect solar power generation (Li et al., 2017; Labordena et al., 2018). Aerosol can transport soluble iron, phosphate and nitrate over long distances and provide nutrients for the biosphere (Swap et al., 1992; Vink and Measures, 2001; McTainsh and Strong, 2007; Maher et al., 2010; Lequy et al., 2012) . Aerosol can penetrate deep into lungs and may carry toxins or serve as disease vectors (Dockery et al., 1993; Brunekreef and Holgate, 2002; Ezzati et al., 2002; Smith et al., 2009; Beelen et al., 2013; Ballester et al., 2013).

Aerosol reflects visible radiation from the Sun, and some aerosol also absorbs it (Dubovik et al., 2002; Omar et al., 2005). The species that absorb the most visible sunlight are, in order of importance: black carbon, dust and brown carbon. Of these, black carbon is expected to exert a significant positive radiative forcing on the climate (Bond et al., 2013; Myhre et al., 2013). Absorbing aerosol's impact is mostly through heating of the atmospheric profile (direct effect) and subsequent stabilisation or instabilisation (Johnson et al., 2003) of the boundary layer (semi-direct effect). This affects cloud formation (Koren et al., 2008; Brioude et al., 2009) and precipitation (Hodnebrog et al., 2016; Samset et al., 2016; Hodzic and Duvel, 2018). In particular over bright surfaces (ice, deserts, clouds) the forcing due to absorbing aerosol can be significant (Haywood and Shine, 1995; Graaf et al., 2012; Tegen and Heinold, 2018).

On regional scales, biomass burning smoke has been implicated in increased tornado severity (Saide et al., 2015) while dust was observed to reduce cyclones (Chen et al., 2016), black carbon may affect the Hadley cell circulation (Allen et al., 2012; Tosca et al., 2013), and black carbon deposition can reduce glacier albedo (Thomas et al., 2017; Zhang et al., 2017; Dang et al., 2017) which may speed up glacier melt.

Currently, absorbing aerosol can be measured in a number of ways. AERONET (Holben et al., 1998) is a global but spatially sparse network of sun photometers that includes two scanning protocols (almucantar and hybrid) that allow inversion of measured radiances into particle size distributions and refractive indices (Dubovik and King, 2000). From this inversion, columnar AAOD can be derived. There are also networks (Laj et al., 2020) of (filter-based) absorption photometers, as used in EMEP (European Monitoring and Evaluation Programme), ACTRIS (Aerosol, Clouds and Trace Gases Research Infrastructure) and IMPROVE (Interagency Monitoring of Protected Visual Environments). These networks are concentrated in Europe and North America, and there is no global coverage. Moreover, these are surface measurements that do not measure the full atmospheric column. Finally, absorption photometers like the SP2 were used on flight campaigns like HIPPO (Schwarz et al., 2010, 2013; Wang et al., 2014). Again, this yields spatially sparse in-situ observations of absorbing aerosol. While these measurement networks have proven to be very important to our understanding of absorbing aerosol, a satellite derived AAOD would contribute greatly by adding spatial context in regions with ground-based instruments, and measurements in regions without such instruments. As it now stands, we have almost no observations of absorbing aerosol over the oceans, in particular in continental outflow regions.

However, in recent years a number of satellite AAOD products have been developed, often based on POLDER (Polarization and Directionality of the Earth's Reflectances) measurements. For example, Lacagnina et al. (2015) used POLDER data to evaluate SSA from AEROCOM models over oceans; Peers et al. (2016) evaluated over ocean above-cloud SSA in AEROCOM models for the African fire season; Lacagnina et al. (2017) estimated the global direct radiative effect of aerosol and Hasekamp et al. (2019b) estimated aerosol-cloud interactions. Chen et al. (2018, 2019) assimilated POLDER AOD and AAOD observations to estimate aerosol emissions while Tsikerdekis et al. (2021) showed the benefit of jointly assimilating POLDER AOD, AAOD and SSA observations. Kacenelenbogen et al. (2019) used combinations of A-TRAIN sensors to infer AAOD over clouds and estimate short-wave direct aerosol effects.

The challenge in retrieving AAOD from satellite is made clear by the challenge in retrieving AAOD from AERONET measurements. AERONET AAOD observations are known to be more uncertain than AOD observations. Dubovik et al. (2000) estimated that AERONET SSA uncertainties for AOD ≤ 0.2 at 440 nm would be at least 0.05, using numerical sensitivity tests. A recent in-depth estimate of the uncertainty in Inversion V3 data (Sinyuk et al., 2020) for four different sites suggested SSA uncertainties at AOD (at 440 nm) $= 0.2$ from 0.037 to 0.048 at 440 nm and from 0.035 to 0.045 at 675 nm. It is not clear whether these uncertainties should be interpreted as site-specific biases or random errors. This distinction matters as random errors can be reduced through appropriate averaging of data. Large differences between AERONET SSA at low AOD and in-situ measurements were indeed confirmed by Andrews et al. (2017). Even at higher AOD (≥ 0.5), Dubovik et al. (2000) suggested SSA errors of at least 0.03. Sinyuk et al. (2020) suggest smaller SSA uncertainties of 0.017 to 0.023 at 440nm and 0.015 to 0.026 at 675 nm for AOD (at 440 nm) = 0.6 . Given the challenges in satellite remote sensing compared to ground-based remote sensing, satellite AAOD and SSA products can be expected to have large errors as well.

GCOS requirements (WMO, 2011) for SSA specify an accuracy within 0.03 and a stability per decade within 0.01, for a horizontal resolution of 5–10 km and a temporal resolution of 4^{hr}. These requirements appear based on typical regional and yearly variations in SSA. However, SSA requirements are different for different applications (monitoring, trends, model evaluation, process studies) while the GCOS requirements are meant to provide a general broad estimate (Popp et al., 2016). In part 2 of our study we will show that current satellite AAOD and SSA capabilities allow useful evaluation of models.

For measurements to be useful in model evaluation, their errors after averaging (spatially, temporally) need to be smaller than the model errors the observations should be able to identify. A traditional evaluation of satellite datasets with AERONET data is unlikely to establish this, partly because the model aspect is ignored, partly because AERONET covers some very interesting aerosol source regions (e.g. oceans, most deserts and boreal fire scapes) only sparsely. In the first part of this study (the current paper) we complement the traditional evaluation with a satellite intercomparison (in itself not unusual) to broaden our understanding of satellite performance over diverse regions. In the second part (a follow-up paper), we present a novel analysis that combines satellite evaluation & intercomparison with model evaluation, and allows assessment of model biases in the context of satellite biases.

We will use satellite data aggregated over $1^o \times 1^o \times 30^{\mathrm{min}}$ as it allows spatio-temporal collocation amongst datasets (satellite, AERONET, AEROCOM) which should strongly reduce representation errors in our analyses (Schutgens et al., 2016b, a). All analyses, even of multi-year averages, will start from spatio-temporally collocated datasets.

This paper is the result of discussions in the AeroCom (AEROsol Comparisons between Observations and Models, https://aerocom.met.no) and AeroSat (International Satellite Aerosol Science Network, https://aero-sat.org) communities. Both are grass-roots communities, the first organised around aerosol modellers, and the second around retrieval groups. They meet every year to discuss common issues in the field of aerosol studies.

The observational datasets used in this study are described in Sect. 2. The collocation and analysis methodology are described in Sect. 3. A first look at the satellite datasets is presented in Sect. 4. Evaluation of satellite AOD, AAOD and SSA with AERONET is performed in Sect 5 and a more detailed intercomparison of satellite data is shown in Sect. 6. A summary and conclusions can be found in Sect. 7.

2 Datasets

2.1 Remote sensing data

Original satellite L2 data (estimates of geophysical variables on the spatio-temporal sampling pattern of the radiances, see also Mittaz and Merchant (2019)) were aggregated unto a regular spatio-temporal grid with spatio-temporal grid-boxes of $1^o \times 1^o \times 30^{\mathrm{min}}$. The resulting super-observations ($1^o \times 1^o \times 30^{\mathrm{min}}$ aggregates) are more representative of global model grid-boxes ($\sim 1^o - 3^o$ in size) while allowing accurate temporal collocation with other datasets. At the same time, the use of super-observations significantly reduces data amount without much loss of information (at the scale of global model grid-boxes). A list of products used in this paper is given in Table 1. A colour legend to the different products can be found in Fig. 1. More explanation of the aggregation procedure can be found in Appendix A.

Super-observations of AOD and AAOD at the same location and time were derived from the same set of L2 data and therefore measure the exact same scene (note an exception for GRASP dataset described below).

The main data are AOD and AAOD at 550 nm, the wavelength at which models typically provide (A)AOD. If (A)AOD was not retrieved at this wavelength, it was logarithmically interpolated or extrapolated from surrounding wavelengths.

2.1.1 FL-MOC

FL-MOC (Fu Liou - MODIS, OMI, CALIOP) is a technique for combining CALIOP aerosol backscatter, MODIS spectral AOD, and OMI AAOD retrievals for estimating full spectral sets of aerosol radiative properties (SSA, asymmetry parameter and AOD). It is not a retrieval per se but a consistent reinterpretation of the combined data within their stated uncertainties. Details are given in Kacenelenbogen et al. (2019, Appendix A). In brief, FL-MOC uses the L2 retrieved aerosol properties as input to a simple look-up table retrieval of aerosol types and concentrations, under the assumption that aerosol properties are consistent with the L2 aerosol observations within the stated uncertainties of each sensor's retrieval. This technique also assumes that the surface reflectance and clouds are properly treated in the underlying retrievals.

Over land, FL-MOC uses OMAERUV AAOD, over ocean OMAERO AAOD. OMAERO is an advanced multi wavelength UV-VIS algorithm that uses 17 wavelengths in the 331-500 nm range in order to calculate the aerosol optical depth and to discriminate between various types of aerosols. It is an extension of the near UV TOMS method (see the OMAERUV product) to a wider wavelength range. The OMAERO algorithm is applied over all surface types, however its primary objective is to derive aerosol properties over the oceans due to the limited availablity of spectral surface reflectivity databases over land.

2.1.2 OMAERUV

The Ozone Monitoring Instrument (OMI) on the EOS-Aura satellite was deployed in July 2004. It is a high resolution spectrograph that measures the upwelling radiance at the top of the atmosphere in the ultraviolet and visible (270–500 nm) regions of the solar spectrum (Levelt et al., 2006). It had a 2600 km wide swath and provides daily global coverage at a spatial resolution varying from 13×24 km at nadir to 28×150 km at the extremes of the swath. OMI hyperspectral measurements are used as input to inversion algorithms to retrieve ozone vertical distribution and column amounts of O_3, NO_2, SO_2, HCHO, BrO, and OClO. OMI observations are also used to retrieve information on aerosols and clouds.

Aerosol properties in the near UV are derived from OMI observations at 354 and 388 nm (Torres et al., 2007). The OMI UV aerosol algorithm (OMAERUV) takes advantage of the large sensitivity to aerosol absorption in the near UV discovered in the mid-90's using heritage TOMS instruments (Herman et al., 1997), and the low reflectance of all ice/snow free terrestrial surfaces, which facilitates the aerosol characterization over all arid and semi-arid regions of the world. The OMAERUV two-channel algorithm simultaneously retrieves AOD and SSA at 388 nm. The main sources of uncertainty are assumed aerosol layer height, and cloud contamination, the latter associated with the sensor's coarse spatial resolution. The OMAERUV fifteen-year record of AOD has been validated with AERONET observations (Torres et al., 2013; Ahn et al., 2014). The SSA record has also been evaluated by comparisons to AERONET and SKYNET (`https://www.skynet-isdc.org/index.php`) ground-based retrievals (Jethva et al., 2014, 2019).

2.1.3 POLDER-SRON

The POLDER-3 instrument was a multi-angle, multi-wavelength polarimeter flying aboard the Polarization & Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar (PARASOL) satellite. It was launched in 2004 and was a part of the satellite constellation A-Train until 2009. Initially designed to be operated for 2 years, POLDER-3 performed its measurements until late 2013, when it was decommissioned. PARASOL provides measurements of a ground scene under (up to) 16 different viewing geometries in 9 spectral bands (443, 490, 565, 670, 763, 765, 865, 910, 1020 nm). Linear polarization measurements (Stokes parameters Q and U) are performed in 3 spectral bands (490, 670, 865 nm). Its spatial resolution at the nadir was about 6 km, and its swath width was 2400 km.

An advanced retrieval algorithm making full use of the information content of the multi-angle photopolarimetric observations from POLDER-3/PARASOL has been developed at SRON-Netherlands Institute for Space Research. The algorithm has large flexibility in defining the aerosol properties included in the retrieval state vector (Fu and Hasekamp, 2018). The aerosol size distribution is described by the sum of an arbitrary number log-normal functions, called modes, where for each mode the effective radius (reff), effective variance (veff), aerosol column number, real and imaginary parts of the refractive index (in the form of coefficients of spectrally dependent functions), fraction of spherical particles assuming the mixture of spheres and spheroids proposed by Dubovik et al. (2006), and the Aerosol Layer Height can (in principle) be retrieved. In the setup used in the present study, the POLDER-SRON algorithm yields the different microphysical characteristics of a bi-modal aerosol size distribution (fine and coarse mode), with the fraction of spheres only be retrieved for the coarse mode (fine mode assumed to consist only of spheres) and the Aerosol Layer Height is fixed to 1km. For retrievals over ocean, the state vector also includes the wind speed, chlorophyll-a concentration, and white-cap fraction, while for retrievals over land, the state vector includes the parameters describing the surface BRDF (Bidirectional Reflectance Distribution Function) (Litvinov et al., 2011). The retrieval is based on an iterative fitting of a linearized radiative transfer model (Hasekamp and Landgraf, 2005) to the PARASOL data, using a cost function containing a misfit term between the forward model and measurement and a regularization term using a priori estimates of values of some of the retrieved parameters. The algorithm, including an application to PARASOL measurements over ocean, is described in Hasekamp et al. (2011). More recent refinements are described by Stap et al. (2015); Wu et al. (2015); Lacagnina et al. (2015); Fu and Hasekamp (2018); Fu et al. (2020). Retrieval results from the SRON algorithm have been used for aerosol type determination by Russell et al. (2014), in studies related to aerosol absorption and direct radiative effect by Lacagnina et al. (2015, 2017), and aerosol-cloud interactions by Hasekamp et al. (2019b), and data assimilation by Tsikerdekis et al. (2021). Currently, the algorithm has been applied to one year (2006) of global aerosol data.

2.1.4 POLDER-GRASP

For a description of the POLDER instrument, see the previous subsection.

GRASP (Generalized Retrieval of Aerosol and Surface Properties) is a unified retrieval algorithm for atmosphere properties from diverse remote sensing observations (Dubovik et al., 2011, 2014), based on earlier work by Dubovik and King (2000); Dubovik et al. (2002, 2006) for AERONET Inversions.

In the current paper, retrievals from the so-called “models” dataset are used. Aerosol is assumed to be an external mixture of five different aerosol components which are retrieved together with spectral parameters of surface BRDF and BPDF (Bidirectional Polarisation Distribution Function). The aerosol is assumed to be a mixture of spherical and non-spherical particles. Each fraction is characterized by particle size distributions similarly to AERONET retrievals. The non-spherical component is modeled as a mixture of randomly oriented spheroids with fixed shape distribution (Dubovik et al., 2006). The details of the “models” approach are discussed by Lopatin et al. (2020) and Chen et al. (2020).The actual inversion uses multi-pixel retrieval (Dubovik et al., 2011) where horizontal pixel-to-pixel variations of aerosol and day-to-day variations of surface reflectance are enforced to be smooth.

The full archive of POLDER/PARASOL observations was retrieved using GRASP and can be found at https://www.grasp-open.com. In addition to the “models” dataset, two other datasets are available (“optimized” and “high-precision”) that use slightly different assumptions in the retrieval. The detailed discussion and validation of all three 0.1 degree PARASOL/GRASP retrievals are provided by Chen et al. (2020). The "models" dataset used in this paper is considered the most applicable for a wide range of circumstances.

The dataset used in the current paper is aggregated to 1 degree spatial resolution (details are listed at https://www.grasp-open.com). The "models" dataset provides AOD and AAOD aggregated from slightly different L2 samplings: an additional minimum AOD threshold is used when aggregating AAOD. To select data of higher quality, AAOD retrievals were used only for cases with sufficient aerosol loading. The same AOD threshold is used for SSA as well. Specifically, a minimum AOD (at 440 nm) threshold of 0.3 over land and 0.02 over ocean were applied (the threshold over ocean is probably too low to assure high quality AAOD but higher thresholds result in significant data loss).

In the current study we prefer to use aggregated AOD and AAOD data that describe the exact same scene, and this is the case for the FL-MOC, OMEARUV and POLDER-SRON datasets mentioned earlier. For the GRASP product, we decided to assume that the aggregated SSA represents the same scene as the AOD aggregate and recalculated an AAOD from that AOD and SSA. Consequently, the AAOD product (indicated as GRASP-M) presented in this paper is different from the AAOD found in the official L3 "models" product. In-situ measurements (Delene and Ogren, 2002; Andrews et al., 2011, 2017; Schmeisser et al., 2018) have suggested a change in SSA at lower AOD so our SSA assumption may introduce additional biases. However, GRASP-M AAOD evaluated better against AERONET than "models" AAOD which showed a high bias vs. AERONET due to the aforementioned minimum AOD threshold.

For this study the L3 GRASP data were additionally filtered based on the FittingResidual field which was required to be smaller than 0.05 (over Land) or 0.1 (over Ocean). This subset evaluates substantially better for AOD retrievals and somewhat better for AAOD retrievals than the full dataset.

2.1.5 AERONET

AERONET (Holben et al., 1998) DirectSun V3 L2.0 (Giles et al., 2019; Smirnov et al., 2000) and Inversion V3 L1.5 & 2.0 data were downloaded from `https://aeronet.gsfc.nasa.gov`, logarithmically interpolated to values at 550 nm and aggregated by averaging over 30 minutes. The DirectSun dataset contains only AOD (at multiple wavelengths). These observations are based on direct transmission measurements of solar light and have a low uncertainty of ± 0.01 (Eck et al., 1999; Schmid et al., 1999), at 400nm and larger.

The Inversion dataset contains AAOD and SSA (at multiple wavelengths) based on measurements of scattered solar light from multiple directions. This inversion uses radiative transfer calculations (Dubovik and King, 2000) and yields larger errors than the DirectSun measurements. In particular, Dubovik et al. (2000) showed that SSA errors decrease with increasing AOD and estimated 440nm SSA errors of ± 0.03 for water-soluble aerosol at 440nm AOD ≥ 0.2 although for dust and biomass burning aerosol higher AOD ≥ 0.5 were needed. These error estimates were based on numerical calculations. A recent in-depth estimate of the uncertainty in Inversion V3 data (Sinyuk et al., 2020) suggested those thresholds to be 440nm AOD > 0.3 and ≥ 0.45, respectively. For an examination of the impact of geometrical configuration on SSA observations, see Torres et al. (2014). Schafer et al. (2014) showed that AERONET SSA retrievals were lower by 0.011 than flight campaign data (on average). Andrews et al. (2017) also compared flight campaign measurements to AERONET SSA and found that the data were usually within the expected errors, although at low AOD ≤ 0.2 significantly lower SSA values were observed by AERONET. A confounding issue for the evaluation of SSA (or, for that matter, AAOD) datasets is that there is no established gold standard.

The Inversion dataset also contains AOD (from Direct Sun retrievals) which is actually used in the inversion. Here we only use those AOD values in the Inversion dataset that have corresponding AAOD and SSA values, so that aggregate values always describe the same scene.

Inversion L2.0 is a subset of L1.5 (which contains almost 30× more observations), based on further cloud screening and the requirement that AOD at 440nm ≥ 0.4. This last criterion results in a minimum AOD at 550nm of 0.25 in the Inversion L2.0 product.

Since an individual AERONET site is not necessarily representative of a $1^o \times 1^o$ grid-box, satellite evaluation may be negatively affected. To select only sites with high representativity we use a list published in Kinne et al. (2013) as described in Schutgens et al. (2020), where we also tested this representativity (using 14 satellite AOD products). The Kinne list was developed with the AERONET DirectSun product (i.e. AOD) in mind but a high-resolution modelling study by Schutgens (2020) suggests that spatial representativity for AOD and AAOD observations can differ substantially for individual sites. We chose to use the Kinne list because it also includes information on maintenance quality, likely more important for Inversion than DirectSun retrievals.

2.1.6 How independent are these satellite products?

An interesting question is how independent these satellite products are.

The GRASP and SRON algorithms are independent retrieval codes with many specific differences in the implementation. First, in the present study POLDER-SRON retrieves parameters of bi-modal lognormal size distribution and complex refractive index for each size mode, while POLDER-GRASP-M retrieves the concentrations of five aerosol components with assumed properties of each component (Chen et al., 2020; Lopatin et al., 2020). Second, GRASP and SRON use the same mathematical function for the BRDF over land (Litvinov et al., 2011) but estimate the parameters to this function independently. In both algorithms, aerosol and surface properties are estimated simultaneously. Third, there are significant differences in use of a priori constraints. POLDER-SRON follows Phillips-Tikhonov regularization (Phillips, 1962; Tikhonov, 1963) including a priori estimates for most of the retrieved state vector parameters (a globally constant value is used) and a flexible strength of the regularization term. The GRASP algorithm is based on the least-square multi-term approach (see Dubovik et al. (2011)) and uses several a priori constraints simultaneously. Specifically, GRASP "models" uses smoothness constraints on the spectral dependence of surface BRDF parameters. Fourth, the SRON algorithm retrieves from measurements of individual pixels while the GRASP algorithm retrieves from measurements of multiple pixels simultaneously, applying spatio-temporal constraints in the process. For example, over land constraints were used to limit temporal variability of retrieved BRDF parameters as well as spatial variability of aerosol retrieved parameters (see Dubovik et al. (2011); Chen et al. (2020)).

The FL-MOC product uses OMAERUV AAOD as input over land but FL-MOC only uses OMAERUV AAOD as an a-priori and assigns this a sizeable uncertainty. CALIOP backscatter is expected to provide a constraint on SSA, and consequently AAOD. As a matter of fact, our analysis shows that FL-MOC and OMAERUV exhibit rather low correlations for AAOD (and SSA). This suggests that the OMAERUV a-priori does not lead to a strong dependency of FL-MOC on OMAERUV. On the other hand, it also suggests that at least one of these products contains sizeable errors.

3 Collocation & analysis methodology

To evaluate and intercompare the remote sensing datasets, they will need to be collocated in time and space to reduce representation errors (Colarco et al., 2014; Schutgens et al., 2016b, 2017). In practice this collocation is another aggregation (performed for each dataset individually) to a spatio-temporal grid with slightly coarser temporal resolution (1 or 3 hours, the spatial grid-box size remains $1^o \times 1^o$). This is followed by a masking operation that retains only aggregated data if it exists in the same grid-boxes for all involved datasets. More details can be found in Appendix A.

We need to allow some flexibility in the time separation between data (here 3 hours) to ensure sufficient numbers of collocated data pairs for further analysis. Schutgens et al. (2020) showed that shorter time separations greatly limited the number of pairs but did not substantially alter the correlation of satellite AOD with AERONET. On the other hand, longer time separations appear to negatively affect the correlation of satellite AAOD with AERONET, see Fig. 2. The analysis shows that satellite AOD correlation with AERONET Inversion data slowly decreases as the collocation criterion is relaxed from 3 to 24 hours. However, satellite AAOD shows a sharp drop in correlation with AERONET at 6 hours (OMAERUV is the exception, the correlation is already low and barely changes). We surmise this is due to plumes of absorbing aerosol drifting over the sites, requiring tight temporal constraints on collocation. Consequences of this finding will be further discussed in Sect. 7.

As the FL-MOC dataset, based on CALIOP measurements, is smaller than the other satellite datasets, we were compelled to collocate FL-MOC with AERONET within 2^o instead of 1^o. Even so, the data count for the FL-MOC evaluation is low.

After spatio-temporally collocating two or more datasets, the data may be further averaged in space and/or time for analysis purposes. Spatio-temporally averaged SSA is *always* derived from averaged AOD and AAOD:

$$\overline{\mathrm{SSA}} = 1 - \overline{\mathrm{AAOD}}/\overline{\mathrm{AOD}}. \tag{1}$$

During the evaluation of products with AERONET, a distinction will be made between either land or ocean grid-boxes in the common grid. A high resolution land mask was used to determine which $1^o \times 1^o$ grid-box contained at most 30% land (designated an ocean box) or water (designated a land box). Most ocean boxes with AERONET observations will be in coastal regions, with some over isolated islands.

3.1 Taylor diagrams

A suitable graphic for displaying multiple datasets' correspondence with a reference dataset ('truth'), is provided by the Taylor diagram (Taylor, 2001). In this polar plot, each data point (r, ϕ) shows basic statistical metrics for an entire dataset. The distance from the origin (r) represents the internal variability (standard deviation) in the dataset. The angle ϕ through which the data point is rotated away from the horizontal axis represents the correlation with the reference dataset, which is conceptually located on the horizontal axis at radius 1 (i.e. every distance is normalised to the internal variability of the reference dataset). It can be shown (Taylor, 2001) that the distance between the point (r, ϕ) and this reference data point at $(1, 0)$ is a measure of the Root Mean Square Error (RMSE, unbiased). A line extending from the point (r, ϕ) is used to show the bias versus the reference dataset (positive for pointing clock-wise).The distance from the end of this line to the reference data point is a measure of the Root Mean Square Difference (RMSD, no correction for bias).

3.2 Uncertainty analysis using bootstrapping

Our estimates of error metrics are inherently uncertain due to finite sampling. If the sampled error distribution is sufficiently similar to the underlying true error distribution, bootstrapping (Efron, 1979) can be used to assess uncertainties in e.g. biases or correlations due to finite sample size. Bootstrapping uses the sampled distribution to generate a large number of synthetic samples by random draws *with replacement*. For each of these synthetic samples, a bias etc. can be calculated and the distribution of these biases provides measures of the uncertainty, e.g. a standard deviation, in the bias due to statistical noise. Bootstrapping has been shown to be reliable even for relatively small sample sizes (that is the size of the original sample, not the number of bootstraps), see Chernick (2008). In this study, the uncertainty bars in some figures were generated by bootstrap analysis.

If the sampled error distribution is different from the true error distribution, bootstrapping will likely underestimate uncertainties. Sampled error distributions may be different from the true error distribution because the act of collocating satellite and AERONET data favours certain conditions. E.g. the effective combination of two cloud screening algorithms (one for the satellite product, the other for AERONET) may favour clear sky conditions and reduce our sampling of errors due to cloud contamination. This uncertainty due to sampling is unfortunately hard to assess, see e.g. Schutgens et al. (2020).

As an example of uncertainty due to sampling, we present Fig. 3 in which an evaluation of the current satellite AOD data with Inversion L2.0 data (only those AOD that have corresponding AAOD inversions, which constrains AOD at 440nm > 0.4) shows substantial shifts compared to DirectSun L2.0. As the uncertainty ranges indicate, the changes in biases are *not* due to statistical noise. Neither is this due to differences in collocated DirectSun and Inversion L2.0 AOD values, that agree very well. Rather, the issue is that AERONET Inversion data are an unrepresentative subsample of the DirectSun data (Inversion data are skewed to high AOD). It is unclear what this means for the AAOD and SSA evaluation but readers should be aware of this unaccounted-for sampling issue that may introduce biases.

3.3 Error metrics for evaluation

We will use the usual global error statistics (bias, standard deviation, Pearson correlation, regression slopes), treating all data as independent. Regression slopes were calculated with a robust Ordinary Least Squares regressor (OLS bisector from the IDL `sixlin` function, Isobe et al. (1990)). This regressor is recommended when there is no proper understanding of the errors in the independent variable, see also Pitkänen et al. (2016).

4 A first look at the satellite products

Multi-year averages of satellite AAOD and their differences are shown in Fig. 4. The AAOD maps can only be compared with caution, as they are derived from products with different temporal sampling. The differences, on the other hand, are based on collocated data and confirm major features. The products all agree on a major AAOD hotspot from (likely) African Savannah biomass burning. Three products agree on AAOD hotspots in China and India, that are known polluted regions (OMAERUV, which is relatively featureless, is the exception. We surmise this is due to the large pixelsize of the OMI instrument, see Table 1, which will not resolve small scale structure in AAOD. The existence of such small scale structure was inferred from Fig. 2). POLDER-GRASP-M and OMAERUV show a clear AAOD hotspot due to Amazonian biomass burning. POLDER-GRASP-M estimates relatively high values over land, and the ocean at high northern latitudes. OMAERUV shows relatively low AAOD over land but high over the entire ocean. FL-MOC clearly estimates higher AAOD over the Sahara than either POLDER-GRASP-M or OMAERUV. POLDER-SRON estimates relatively high AAOD over the Rocky Mountains, the Andes and Australia. Unfortunately, even in multi-year averages significant differences in regional AAOD between the products are observed, in excess of 50%. Figure S1 shows the corresponding SSA maps. As expected, POLDER-GRASP-M has relatively low SSA and OMAERUV relatively high SSA over land. FL-MOC has the highest SSA over ocean of all products. As the satellite AOD are fairly similar, lower values of AAOD translate into higher values of SSA.

One caveat is that AAOD and SSA retrievals are likely to be better (more accurate and precise) at high AOD. In the above analysis, no account was taken of AOD levels and the products were discussed as they are. The impact of AOD will be discussed later, when discussing the evaluation with AERONET in Sect. 5.2 and the satellite intercomparison in Sect. 6.

5 Evaluation of satellite products with AERONET

Taylor plots of the performance of the satellite products are shown in Fig. 5. Satellite AOD is evaluated against AERONET DirectSun L2.0. Satellite AAOD & SSA, are evaluated against AERONET Inversion L2.0 (which constrains AOD at 440nm > 0.4 and provides much less data than DirectSun). All products show high correlation with AERONET AOD ($r \geq 0.76$), although the correlations found are lower than those found in Schutgens et al. (2020) for several MODIS Aqua products (0.87-0.88). Correlations for AAOD and SSA are lower than for AOD suggesting that it is more challenging to retrieve absorbing qualities.

Interestingly, POLDER-SRON's SSA correlates significantly better with AERONET than POLDER-GRASP-M's but this is a sampling effect: once both products are collocated together, POLDER-GRASP-M's SSA correlation with AERONET increases from 0.41 to 0.69. The explanation for this is not entirely clear, although it turns out that POLDER-GRASP-M evaluates poorer with AERONET for 2010 than for 2006 and 2008 (POLDER-SRON is currently limited to 2006, see Table 2.1). Although the poorer evaluation for 2010 can be seen in AOD, AAOD and SSA, it is only statistically significant for SSA.

The impact of statistical noise on the AAOD evaluation is explored in Fig. 6. Using a bootstrapping technique, the spread in correlation and standard deviation were explored. For most datasets, the results seem fairly robust, except for FL-MOC which yielded only 24 data points. A proper intercomparison of products requires collocation (of *all* the satellite data), which reduces available cases even further. Figure S2 shows that results are not very different from Fig. 5, but the statistical noise increases substantially. The sampling noise on such a small subset should be even larger, see also Fig. 3 and Schutgens et al. (2020). For a sense of perspective, 48 data points represents less than 0.0008% of the total POLDER-GRASP-M data amount used in this paper.

5.1 Evaluation and intercomparison of AOD

In Fig. 7, we provide more detail on the satellite AOD products and their evaluation against AERONET DirectSun L2.0 AOD. In the central column, we show the products themselves, averaged over 1, 2 or 3 year(s), depending on availability (see Table 1). Note that the products exist for different years and even for the same years products will have different temporal samplings so comparisons should be made with caution (Colarco et al., 2014; Schutgens et al., 2016a). In the left and right column, we show satellite data collocated with AERONET. On the left-hand side is a scatterplot of the data (with associated statistics provided) and on the right-hand side is a map of multi-year difference with AERONET (provided at least 32 data points were available per site).

The scatter plots show good correlation with AERONET. The POLDER products show higher correlations and slopes closer to one (1) than FL-MOC and OMAERUV. Nevertheless, differences in evaluation seem rather small, which unfortunately cannot be said for the global distributions of AOD. POLDER-GRASP-M has rather high AOD over land and OMAERUV has rather high AOD over ocean (note that the satellite data themselves are not collocated). The multi-year differences with AERONET suggest that OMAERUV overestimates everywhere except in some regions with strongly absorbing aerosol. An intercomparison of satellite AOD with Aqua-DT is presented in Fig. S3 and suggests typically higher estimates over (Southern Hemisphere) Land for the POLDER products and over Ocean for OMAERUV. Note that Aqua-DT is not without significant regional biases, see Schutgens et al. (2020).

Figure 8 shows results when bias (sign-less) and correlation per site (that yielded at least 32 collocations) are averaged over all sites, for each satellite product. The same 52 sites are used for all datasets although each product is individually collocated with AERONET. For FL-MOC, no site provided at least 32 observations and it is not included in the analysis. For POLDER-SRON, only 18 sites provided at least 32 collocated observations and it was similarly excluded. As was also shown in Schutgens et al. (2020), OMAERUV shows rather large biases compared to the other AOD products. POLDER-GRASP-M, on the other hand, shows the smallest bias. The filtering of GRASP retrievals described in Sect. 2.1 plays a significant role in this result (without filtering, POLDER-GRASP-M shows a bias twice as large).

5.2 Evaluation of AAOD and SSA

Figure 9 provides more detail on the evaluation of satellite (A)AOD & SSA products against AERONET Inversion L2.0 (which constrain AOD at 440nm > 0.4). In the first three columns, we show scatter plots for respectively AOD, AAOD and SSA. In the last column we show SSA differences with AERONET as a function of AERONET AOD (Inversion L1.5). All products underestimate AERONET AOD and AAOD, although only by a small amount in the case of POLDER-GRASP-M. More importantly, AAOD correlations can be low as 0.34 (OMAERUV) and regression slope can deviate substantially from 1 (0.6 for OMAERUV). In contrast, some products underestimate SSA while others over-estimate it. Due to data sparsity (e.g. for POLDER-GRASP-M, the count dropped from 10454 to 423), it is not possible to do an analysis per AERONET site (as was done for AOD) and see how the global bias relates to regional biases. Bootstrap analysis suggest that results are fairly robust against statistical noise (except FL-MOC, see also Fig. 6).

The right-most column in Fig. 9 shows SSA difference as a function of (AERONET) AOD. To ensure the largest possible range in AOD values Inversion L1.5 instead of L2.0 is used. Especially at lower AOD, this dataset will have larger errors in AAOD and SSA than L2.0. Interestingly, as AOD increases, all satellite products seem to agree better with AERONET (for FL-MOC, the bin with largest AOD values is affected by a very low data count). This is of course as one would expect. For smaller AOD, there is increasingly more spread although the difference distribution remains fairly unbiased. The exception is POLDER-GRASP-M which shows increasingly lower SSA than AERONET at low AOD. We suggest that it is rather unlikely that three different satellite products have a similar SSA bias at low AOD as AERONET (and hence show no bias in the difference with AERONET) and that this low bias in POLDER-GRASP-M analysis is real. However, a better understanding of the nature of errors (bias vs. random) in AERONET SSA at low AOD is desirable.

Summarizing, there is skill in satellite AAOD and SSA but compared to AOD the correlations with AERONET are substantially lower. POLDER-SRON is the exception, with similar and fairly high correlations (~ 0.75) for all three parameters. However, it seems to underestimate AAOD by $\sim 25\%$ at high AAOD (slope of 0.76 in the AAOD scatter plot). OMAERUV appears to show the largest deviations from AERONET (low correlations and slopes) but its overall error statistics (mean and standard deviation) is not too different from the other products. Results for FL-MOC may be a statistical fluke due to the low data count. POLDER-GRASP-M shows quite high correlations for AOD (0.86) and AAOD (0.6) with reasonable slopes but has a very low correlation with AERONET for SSA (0.41), but this seems to depend strongly on sampling as discussed at the start of this section. In addition, it appears to systematically underestimate SSA at low AOD. Yet another aspect to this dataset (not visible in any of the analysis shown) is that it appears to have a hard SSA cut-off as SSA values larger than 0.99 do not occur.

A profound problem is the paucity of data. Even for POLDER-GRASP-M, we can only evaluate its performance (against AERONET) for less than 0.006% the total number of available observations. Is this sufficient to make meaningful statements about the performance of a product *at large*? In Schutgens et al. (2020), we showed that the process of collocation can skew error statistics (by changing the sampling) to the point that it becomes hard to meaningfully distinguish performance of several products. That study was done for AOD which allows much higher numbers of collocated data with AERONET than AAOD.

To elucidate this, we compare the difference in SSA between the two POLDER products (collocated within 3 hours, considering AOD ≥ 0.25 only) for three different samplings. First, we look at global POLDER SSA statistics. Secondly, we look at POLDER SSA statistics over AERONET sites only. Thirdly, we look at POLDER SSA statistics that are collocated with AERONET observations. Figure 10 shows the associated difference distributions. Using various non-parametric statistical tests (Mann-Whitney U, Student's t, Kolmogorov-Smirnov) we can show that the distribution means for the first and third sampling are significantly different. Not only that, but the mean difference in SSA for the first sampling is 2.6 as large (-0.043 vs. -0.017) as for the third sampling. As POLDER-SRON is biased high and POLDER-GRASP-M is biased low vs AERONET, the corollary to this is of course that at least one of the products has a larger bias vs the truth globally than can be seen in the AERONET observations. Conversely this suggests that the AERONET Inversion dataset does not allow a truly global evaluation of satellite datasets: it provides a sub-sample with skewed statistics of SSA errors. Incidentally, it is the temporal sub-sampling enforced by collocation with AERONET observations that causes the largest shift in the difference distribution (POLDER measurements over AERONET sites show a similar SSA distribution as the global dataset). It is possible that the SSA difference is partly driven by cloud contamination which we know is present in these satellite datasets (Schutgens et al., 2020) and may be ameliorated when a third cloud masking (from AERONET) is applied (through the collocation of data).

6 Intercomparison of satellite AAOD and SSA

To get a better appreciation of the satellite products, we now present a global intercomparison. To start with, Fig. 11 shows SSA differences between two products as a function of their mean AOD. As in Fig. 9, these differences become smaller (i.e. show a smaller spread) at higher AOD, as expected (intercomparisons with FL-MOC are the exception). However, satellite SSA values still exhibit random differences of 0.03 or larger for AOD $\gtrapprox 1$, as also confirmed by the AERONET evaluation. In addition, substantial biases remain.

The previous analysis was global but substantial differences can be seen between land and ocean scenes. For instance, the SSA bias between the POLDER products over land, does not decrease at lower AOD but remains fairly constant. A more detailed analysis can be found in Fig. 12 which shows biases, correlations and regression slopes for different products. Unsurprisingly, correlations and slopes tend to improve with minimum AOD, while biases may remain fairly constant (POLDER products), decrease (OMAERUV vs POLDER-GRASP-M) or even increase (FL-MOC). As a consequence it should be challenging to determine an AOD threshold above which products can be expected to perform within certain parameters. A similar analysis for AAOD can be found in Fig. S4.

A final analysis concerns multi-year averages of these products. Model evaluation will be done on such averages and it may be useful to better understand the agreement (or lack thereof) between products in that case, even though the aforementioned biases are unlikely to be much reduced. Figure 13 shows an intercomparison of three products (FL-MOC is excluded due to its low data count). The analysis shows statistics of the intercomparison of multi-year averages of SSA, as a function of two thresholds: a minimum AOD and a minimum number of super-observations during three years (per $1^o \times 1^o$ grid-box). The underlying super-observations were always collocated (to within 3 hours) before temporal averaging took place. We see that, in general, correlations increase and standard deviation in the difference decrease when either threshold increases. The improvement with increasing AOD has already been discussed and is due to better signal-to-noise conditions for the retrieval schemes. The improvement with increasing number of observations (used in the temporal averaging) can be interpreted as a significant random error in either product being lessened through averaging. In general, the AOD threshold has a more profound impact but the number of observations threshold allows more flexibility (by choosing a longer time-series to work with, smaller SSA differences (up to a point!) may be achieved).

However, biases between products can be quite robust as is particularly clear for the POLDER products. The decreasing bias for OMAERUV vs. POLDER-SRON (and, incidentally, the sudden jump in correlation for AOD > 0.4) is not really a sign of a better agreement between products at high AOD. Under these conditions, most observations come from the African dust and biomass burning regions. POLDER-SRON retrieves very reflective dust and very absorbing biomass burning aerosol while OMAERUV retrieves fairly reflective dust and fairly absorbing biomass burning aerosol. Consequently, global SSA bias decreases due to a balancing of very different biases over these regions while similar spatial patterns yield high correlations. Maps of the SSA difference between the POLDER products as a function of minimum AOD can be seen in Fig. S5. A higher minimum AOD mostly constrains data to a smaller portion of the globe but does not affect local biases greatly.

7 Conclusions

In this study, we evaluate several remote sensing datasets of AAOD and SSA, from a variety of sensors (CALIOP on CALYPSO, OMI on Aura, POLDER on PARASOL), in preparation of an AEROCOM model evaluation. This is the first global study to intercompare satellite remotely sensed products of AAOD (and SSA).

The evaluation of the products (daily aggregates over $1^o \times 1^o$) is done through comparison with AERONET DirectSun (AOD) and Inversion (AAOD and SSA) observations. To minimize sampling issues, satellite products and AERONET data are collocated in time and space, within 3 hours and 1 degree. One interesting finding is that AAOD evaluation requires a tighter temporal collocation criterion than AOD, with steep declines in correlation found for temporal collocation after 3 hours or more. We interpret this to be due to absorbing aerosol primarily being found in plumes. While we do not explore this further, this high temporal variability in observed AAOD may affect model evaluation as well. It could suggest that models need emissions with diurnal profiles, and output at higher frequencies than daily to obtain the best possible agreement with observations.

All satellite AOD products show significant correlation with AERONET ($0.76 \leq r \leq 0.86$). Global biases are not very different from those found in an earlier study of traditional products (Schutgens et al., 2020). However, when considering typical multi-year biases per AERONET site, there is a suggestion that POLDER-GRASP-M has smaller biases than these traditional products (there is a hint this may also be true for POLDER-SRON but paucity of data makes this analysis less certain). In contrast, OMAERUV shows the largest (and mostly positive) biases in AOD. Compared to Aqua-DT (Dark Target), the four products studied in this paper tend to estimate higher AOD over most of the land.

Results for AAOD are more diverse, with generally lower correlations ($0.34 \leq r \leq 0.78$) than for AOD. For most products, SSA correlates significantly worse with AERONET than AAOD. All products show an improvement in SSA with regards to AERONET at higher AOD. POLDER-GRASP-M is noted for a low bias in SSA at low AOD.

The two POLDER products perform better against AERONET than the other two products, with typically (but not always) higher correlations, smaller biases and regression slopes closer to one (1) for all three parameters AOD, AAOD and SSA. However, dearth of measurements makes it very difficult to 1) meaningfully compare evaluation metrics amongst the products and 2) draw global conclusions. Theoretical evidence (Hasekamp and Landgraf, 2007; Hasekamp, 2010; Hasekamp et al., 2019a) suggests that retrieval schemes for absorptive properties will benefit from using polarisation measurements at multiple view angles which would support the idea that the POLDER products perform better. In addition, the OMAERUV product is based on measurements from a sensor with substantially larger pixels than POLDER and will struggle to resolve the fine-scale structure of aerosol plumes.

An intercomparison of multi-year satellite AAOD and SSA suggests significant biases across the globe. Differences of 50% in multi-year averages of AAOD are not unusual. OMAERUV shows lower AAOD over land than the other products, but slightly higher AAOD over ocean. FL-MOC shows significantly higher AAOD over the Sahara and POLDER-GRASP-M is noted for a high AAOD at high Northern latitudes, both over land and ocean. POLDER-SRON has much higher AAOD than the other products over high-altitude regions. Many of these regions are unfortunately poorly instrumented with AERONET sites. Satellite SSA does agree better at high AOD, as was also observed for AERONET, although dearth of data means this can not be firmly concluded for FL-MOC. However, correlations for super-observations are often lower than 0.6, even at high AOD (0.75). Over ocean, SSA products tend to correlate better than over land. The two POLDER products correlate better than any other satellite pair ($r =\sim 0.8$ over ocean for AOD > 0.75). In addition to high AOD, we show that temporal averaging also improves agreement between satellite products, although it is not possible to give recommendations that work well with all products and for all regions. Even so, biases between products exist at high AOD after substantial temporal averaging.

Most surprisingly, POLDER-GRASP-M and POLDER-SRON show a fairly systematic difference in SSA (-0.04), independent of AOD (there are regional variations). For low AOD (< 0.1) cases over ocean, this systematic difference becomes small in the global average because of two opposite biases organised roughly (!) by hemisphere (see also Fig. S1). Identifying the cause of this bias may lead to substantial improvements of both products (or at least one of them). Based on a comparison with AERONET data, we suggest that cloud contamination is a possible candidate.

Throughout the paper, we have given examples of how limited sampling of observations (especially AERONET) constrains our ability to understand the true error statistics of satellite AAOD and SSA. The most prominent example is a much reduced systematic difference (-0.017) between POLDER-GRASP-M and POLDER-SRON SSA as seen in an evaluation with AERONET Inversion L2.0 observations, as compared to the global satellite dataset (-0.04). This suggest that biases inferred from an AERONET evaluation will be smaller than those actually present in the satellite products. To increase available SSA observations, one could use Inversion L1.5 data (which includes SSA at low AOD) and sample it to L2.0 AOD measurements (which, unlike SSA, exist at low AOD), thereby benefitting from the better L2.0 cloud screening. Especially if follow-up studies can show that inversion errors at individual sites behave as random errors (amenable to temporal averaging) and not systematic biases such an intermediate product might be very useful.

This paper is one part of a two paper study into the use of satellite AAOD and SSA for aerosol model evaluation. In its companion paper, we use the datasets introduced in the current paper to evaluate AEROCOM (AEROsol Comparisons between Observations and Models) models. It turns out that robust and consistent evaluation of the models is possible, notwithstanding the biases in the satellite data we have detailed in the current paper. The main reason seems to be that model biases (and the diversity in those biases) are even larger than satellite biases. Hence these satellite AAOD and SSA products are very useful: in regions with AERONET sites, they provide spatial detail lacking in a surface network; in regions without AERONET sites, they are the only datasets of observed AAOD and SSA available.

Code and data availability. All remote sensing data is freely available. Analysis code was written in IDL and is available from the author upon request.

Author contributions. NS designed the study, with the help of GL, TP, SK, MS and PS, and carried it out. OD, OH, OT, HJ, PL, JR and YS provided the remote sensing data. NS prepared the manuscript, with the help of all co-authors.

Competing interests. No competing interests are present

Acknowledgements. We thank the PI(s) and Co-I(s) and their staff for establishing and maintaining the many AERONET sites used in this investigation. The figures in this paper were prepared using David W. Fanning's Coyote Library for IDL. The work by N. Schutgens is part of the Vici research programme with project number 016.160.324, which is (partly) financed by the Dutch Research Council (NWO). NS thanks Tom Eck, Greg Schuster and Kostas Tsigaridis for insightful discussions on the use of AERONET observations. We would also like to thank four anonymous reviewers for attentive reading of our manuscript and many useful comments.

References

Ahn, C., Torres, O., and Jethva, H.: Assessment of OMI near-UV aerosol optical depth over land, Journal of Geophysical Research : Atmospheres, 119, 2457–2473, https://doi.org/10.1002/2013JD020188.Received, 2014.

Albrecht, B. A.: Aerosols, cloud microphysics, and fractional cloudiness, Science, 245, 1227–1230, 1989.

Allen, R. J., Sherwood, S. C., Norris, J. R., and Zender, C. S.: Recent Northern Hemisphere tropical expansion primarily driven by black carbon and tropospheric ozone, Nature, 485, 350–354, https://doi.org/10.1038/nature11097, http://www.nature.com/doifinder/10.1038/nature11097, 2012.

Andrews, E., Ogren, J. A., Bonasoni, P., Marinoni, A., Cuevas, E., Rodríguez, S., Sun, J. Y., Jaffe, D. A., Fischer, E. V., Baltensperger, U., Weingartner, E., Coen, M. C., Sharma, S., Macdonald, A. M., Leaitch, W. R., Lin, N., Laj, P., Arsov, T., Kalapov, I., Jefferson, A., and Sheridan, P.: Climatology of aerosol radiative properties in the free troposphere, Atmospheric Research, 102, 365–393, https://doi.org/10.1016/j.atmosres.2011.08.017, http://dx.doi.org/10.1016/j.atmosres.2011.08.017, 2011.

Andrews, E., Ogren, J. A., Kinne, S., and Samset, B.: Comparison of AOD, AAOD and column single scattering albedo from AERONET retrievals and in situ profiling measurements, Atmospheric Chemistry and Physics, 17, 6041–6072, https://doi.org/10.5194/acp-17-6041-2017, http://www.atmos-chem-phys.net/17/6041/2017/, 2017.

Angstrom, B. A.: Atmospheric turbidity , global illumination and planetary albedo of the earth, Tellus, XIV, 435–450, 1962.

Ballester, J., Burns, J. C., Cayan, D., Nakamura, Y., Uehara, R., and Rodó, X.: Kawasaki disease and ENSO-driven wind circulation, Geophysical Research Letters, 40, 2284–2289, https://doi.org/10.1002/grl.50388, http://doi.wiley.com/10.1002/grl.50388, 2013.

Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z. J., Weinmayr, G., Hoffmann, B., Wolf, K., Samoli, E., Fischer, P., Nieuwenhuijsen, M., Vineis, P., Xun, W. W., Katsouyanni, K., Dimakopoulou, K., Oudin, A., Forsberg, B., Modig, L., Havulinna, A. S., Lanki, T., Turunen, A., Oftedal, B., Nystad, W., Nafstad, P., De Faire, U., Pedersen, N. L., Östenson, C.-G., Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K. T., Overvad, K., Ellermann, T., Eeftens, M., Peeters, P. H., Meliefste, K., Wang, M., Bueno-de Mesquita, B., Sugiri, D., Krämer, U., Heinrich, J., de Hoogh, K., Key, T., Peters, A., Hampel, R., Concin, H., Nagel, G., Ineichen, A., Schaffner, E., Probst-Hensch, N., Künzli, N., Schindler, C., Schikowski, T., Adam, M., Phuleria, H., Vilier, A., Clavel-Chapelon, F., Declercq, C., Grioni, S., Krogh, V., Tsai, M.-Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Katsoulis, M., Trichopoulou, A., Brunekreef, B., and Hoek, G.: Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project, The Lancet, 6736, 1–11, https://doi.org/10.1016/S0140-6736(13)62158-3, http://linkinghub.elsevier.com/retrieve/pii/S0140673613621583, 2013.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., Deangelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, Journal of Geophysical Research Atmospheres, 118, 5380–5552, https://doi.org/10.1002/jgrd.50171, 2013.

Brioude, J., Cooper, O. R., Feingold, G., Trainer, M., Freitas, S. R., Kowal, D., Ayers, J., Prins, E., Minnis, P., McKeen, S. A., Frost, G. J., and Hsie, E.-Y.: Effect of biomass burning on marine stratocumulus clouds off the California coast, Atmospheric Chemistry and Physics, 9, 14 529–14 570, https://doi.org/10.5194/acpd-9-14529-2009, http://www.atmos-chem-phys-discuss.net/9/14529/2009/, 2009.

Brunekreef, B. and Holgate, S. T.: Air pollution and health., Lancet, 360, 1233–42, https://doi.org/10.1016/S0140-6736(02)11274-8, http://www.ncbi.nlm.nih.gov/pubmed/12401268, 2002.

Chen, C., Dubovik, O., Henze, D. K., Lapyonak, T., Chin, M., Ducos, F., Litvinov, P., Huang, X., and Li, L.: Retrieval of Desert Dust and Carbonaceous Aerosol Emissions over Africa from POLDER / PARASOL Products Generated by GRASP Algorithm, Atmospheric Chemistry and Physics, 18, 12 551–12 580, 2018.

Chen, C., Dubovik, O., Henze, D. K., Chin, M., Lapyonok, T., Schuster, G. L., Ducos, F., Fuertes, D., Litvinov, P., Li, L., Lopatin, A., Hu, Q., and Torres, B.: Constraining global aerosol emissions using POLDER / PARASOL satellite remote sensing observations, Atmospheric Chemistry and Physics, 19, 14 585–14 606, 2019.

Chen, C., Dubovik, O., Fuertes, D., Litvinov, P., Lapyonok, T., Lopatin, A., Ducos, F., Derimian, Y., Herman, M., Remer, L. A., Sayer, A. M., L evy, R. C., Hsu, N. C., Descloitres, J., Li, L., Torres, B., Karol, Y., H errera, M., Herreras, M., Aspetsberger, M., Bindreiter, L., Marth, D., Hangle r, A., and Federspiel, C.: product from, Earth System Science Data, 2020.

Chen, D., Liu, Z., Davis, C., and Gu, Y.: Dust Radiative Effects on Atmospheric Thermodynamics and Tropical Cyclogenesis over the Atlantic Ocean Using WRF/Chem Coupled with an AOD Data Assimilation System, Atmospheric Chemistry and Physics, pp. 17 917–7939, https://doi.org/10.5194/acp-2016-744, http://www.atmos-chem-phys-discuss.net/acp-2016-744/, 2016.

Chernick, M.: Bootstrap Methods : A Guide for Practitioners and Researchers, John Wiley & Sons, Inc., Hoboken, New Jersey, 2nd edn., 2008.

Colarco, P. R., Kahn, R. A., Remer, L. A., and Levy, R. C.: Impact of satellite viewing-swath width on global and regional aerosol optical thickness statistics and trends, Atmospheric Measurement Techniques, 7, 2313–2335, https://doi.org/10.5194/amt-7-2313-2014, 2014.

Dang, C., Warren, S. G., Fu, Q., Doherty, S. J., Sturm, M., and Su, J.: Measurements of light-absorbing particles in snow across the Arctic, North America, and China: Effects on surface albedo, Journal of Geophysical Research: Atmospheres, pp. 149–168, https://doi.org/10.1002/2017JD027070, http://doi.wiley.com/10.1002/2017JD027070, 2017.

Delene, D. J. and Ogren, J. A.: Variability of Aerosol Optical Properties at Four North American Surface Monitoring Sites, Journal of Atmospheric Sciences, 59, 1135–1150, 2002.

Dockery, D., Pope, A., Xu, X., Spengler, J., Ware, J., Fay, M., Ferris, B., and Speizer, F.: An association between air pollution and mortality in six U.S. cities, The New England Journal of Medicine, 329, 1753–1759, 1993.

Dubovik, O. and King, M. D.: A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, Journal of Geophysical Research: Atmospheres, 105, 20 673–20 696, https://doi.org/10.1029/2000JD900282, http://doi.wiley.com/10.1029/2000JD900282, 2000.

Dubovik, O., Smirnov, A., Holben, B. N., King, M. D., Kaufman, Y. J., Eck, T. F., and Slutsker, I.: Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) Sun and sky radiance measurements, Journal of Geophysical Research, 105, 9791–9806, https://doi.org/10.1029/2000JD900040, http://doi.wiley.com/10.1029/2000JD900040, 2000.

Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanre, D., and Slutsker, I.: Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations, Journal of Atmospheric Sciences, 59, 590–608, 2002.

Dubovik, O., Sinyuk, A., Lapyonok, T., Holben, B. N., Mishchenko, M., Yang, P., Eck, T. F., Volten, H., Mun, O., Veihelmann, B., Zande, W. J. V. D., Leon, J.-f., Sorokin, M., and Slutsker, I.: Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006619, 2006.

Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanré, D., Deuzé, J. L., Ducos, F., Sinyuk, A., and Lopatin, A.: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations, Atmospheric Measurement Techniques, 4, 975–1018, https://doi.org/10.5194/amt-4-975-2011, 2011.

Dubovik, O., Lapyonok, T., Litvinov, P., Herman, M., Fuertes, D., Ducos, F., Lopatin, A., Chaikovsky, A., Torres, B., Derimian, Y., Huang, X., Aspetsberger, M., and Federspiel, C.: GRASP: a versatile algorithm for characterizing the atmosphere, https://doi.org/10.1117/2.1201408.005558, http://spie.org/x109993.xml, 2014.

Eck, T. F., Holben, B. N., Reid, J. S., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne, S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, J. Geophysical Research, 104, 31 333–31 349, 1999.

Efron, B.: Bootstrap methods: another look at the jackknife, The annals of Statistics, 7, 1—-26, 1979.

Ezzati, M., Lopez, A. D., Rodgers, A., Vander Hoorn, S., and Murray, C. J. L.: Selected major risk factors and global and regional burden of disease., Lancet, 360, 1347–60, https://doi.org/10.1016/S0140-6736(02)11403-6, http://www.ncbi.nlm.nih.gov/pubmed/12423980, 2002.

Fu, G. and Hasekamp, O.: Retrieval of aerosol microphysical and optical properties over land using a multimode approach, Atmospheric Measurement Techniques, 11, 6627–6650, 2018.

Fu, G., Hasekamp, O., Rietjens, J., Smit, M., Noia, A. D., Cairns, B., Wasilewski, A., Diner, D., Seidel, F., Xu, F., Knobelspiesse, K., Gao, M., and Silva, A.: Aerosol retrievals from different polarimeters during the ACEPOL campaign using a common retrieval algorithm, Atmospheric Measurement Techniques, 13, 553–573, 2020.

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, Atmospheric Measurement Techniques, 12, 169–209, 2019.

Graaf, M. D., Tilstra, L. G., Wang, P., and Stammes, P.: Retrieval of the aerosol direct radiative effect over clouds from spaceborne spectrometry, Journal of Geophysical Research: Atmospheres, 117, 1–18, https://doi.org/10.1029/2011JD017160, 2012.

Hansen, J., Sato, M., and Ruedy, R.: Radiative forcing and climate response, Journal of Geophysical Research, 102, 6831–6864, 1997.

Hasekamp, O. P.: Capability of multi-viewing-angle photo-polarimetric measurements for the simultaneous retrieval of aerosol and cloud properties, Atmospheric Measurement Techniques, 3, 839–851, https://doi.org/10.5194/amt-3-839-2010, 2010.

Hasekamp, O. P. and Landgraf, J.: Linearization of vector radiative transfer with respect to aerosol properties and its use in satellite remote sensing, Journal of Geophysical Research (Atmospheres), 110, 4203–+, 2005.

Hasekamp, O. P. and Landgraf, J.: Retrieval of aerosol properties over land surfaces : capabilities of multiple-viewing-angle intensity and polarization measurements, Applied optics, 46, 3332–3344, 2007.

Hasekamp, O. P., Litvinov, P., and Butz, A.: Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/10.1029/2010JD015469, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD015469, 2011.

Hasekamp, O. P., Fu, G., Rusli, S. P., Wu, L., Noia, A. D., aan de Brugh, J., Landgraf, J., Smit, J. M., Rietjens, J., and Van Amerongen, A.: Journal of Quantitative Spectroscopy & Radiative Transfer Aerosol measurements by SPEXone on the NASA PACE mission : expected retrieval capabilities, Journal of Quantitative Spectroscopy and Radiative Transfer, 227, 170–184, https://doi.org/10.1016/j.jqsrt.2019.02.006, https://doi.org/10.1016/j.jqsrt.2019.02.006, 2019a.

Hasekamp, O. P., Gryspeerdt, E., and Quaas, J.: Analysis of polarimetric satellite measurements suggests stronger cooling due to aerosol-cloud interactions, Nature Communications, https://doi.org/10.1038/s41467-019-13372-2, http://dx.doi.org/10.1038/s41467-019-13372-2, 2019b.

Haywood, J. M. and Shine, K. P.: The effect of anthropogenic sulfate and soot aerosol on the clear sky planetary radiation budget, Geophysical Research Letters, 22, 603–606, 1995.

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, Journal of Geophysical Research: Atmospheres, 102, 16,911–16,922, 1997.

Hodnebrog, Ø., Myhre, G., Forster, P. M., Sillmann, J., and Samset, B. H.: Local biomass burning is a dominant cause of the observed precipitation reduction in southern Africa, Nature Communications, 7, 11 236, https://doi.org/10.1038/ncomms11236, http://www.nature.com/doifinder/10.1038/ncomms11236, 2016.

Hodzic, A. and Duvel, J. P.: Impact of biomass-burning aerosols on the diurnal cycle of convective clouds and precipitation over a tropical island, Journal of Geophysical Research: Atmospheres, 123, 1017–1036, https://doi.org/10.1002/2017JD027521, http://doi.wiley.com/10.1002/2017JD027521, 2018.

Holben, B., Eck, T., Slutsker, I., Tanré, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sensing of Environment, 66, 1–16, https://doi.org/10.1016/S0034-4257(98)00031-5, http://linkinghub.elsevier.com/retrieve/pii/S0034425798000315, 1998.

Isobe, T., Feigelson, E. D., Akritas, M. G., and Babu, G. J.: Linear regression in Astronomy I, The Astrophysical Journal, 364, 104–113, 1990.

Jethva, H., Torres, O., and Ahn, C.: Global assessment of OMI aerosol single-scattering albedo using ground-based AERONET inversion, Journal of Geophysical Research : Atmospheres, 119, 9020–9040, https://doi.org/10.1002/2013JD020188.Global, 2014.

Jethva, H., Torres, O., and Yoshida, Y.: Accuracy assessment of MODIS land aerosol optical thickness algorithms using AERONET measurements over North America, Atmospheric Measurement Techniques, 12, 4291–4307, 2019.

Johnson, B. B. T., Shine, K. P., and Forster, P. M.: The Semi-direct Aerosol Effect : Impact of Absorbing Aerosols on Marine Stratocumulus, Quarterly Journal of the Royal Meteorological Society, 130, 2003.

Kacenelenbogen, M. S., Vaughan, M. A., Redemann, J., Young, S. A., Liu, Z., Hu, Y., Omar, A. H., Leblanc, S., Shinozuka, Y., Livingston, J., Zhang, Q., and Powell, K. A.: Estimations of global shortwave direct aerosol radiative effects above opaque water clouds using a combination of A-Train satellite sensors, Atmosphere - Ocean, 19, 4933–4962, 2019.

Kinne, S., O'Donnel, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., Rast, S., Giorgetta, M., Eck, T. F., and Stevens, B.: MAC-v1: A new global aerosol climatology for climate studies, Journal of Advances in Modeling Earth Systems, 5, 704–740, https://doi.org/10.1002/jame.20035, http://doi.wiley.com/10.1002/jame.20035, 2013.

Koren, I., Martins, J. V., Remer, L. A., and Afargan, H.: Smoke invigoration versus inhibition of clouds over the Amazon, Science, 321, 946–949, 2008.

Labordena, M., Neubauer, D., Folini, D., Patt, A., and Lilliestam, J.: Blue skies over China : The effect of pollution- control on solar power generation and revenues, PLOSonec, 13, 1, 2018.

Lacagnina, C., Hasekamp, O. P., Bian, H., Curci, G., Myhre, G., Noije, T. V., Schulz, M., Skeie, R. B., Takemura, T., and Zhang, K.: Aerosol single-scattering albedo over the global oceans : Comparing PARASOL retrievals with AERONET , OMI , and AeroCom models estimates, Journal of Geophysical Research: Atmospheres, 120, 9814–9836, https://doi.org/10.1002/2015JD023501.Abstract, 2015.

Lacagnina, C., Hasekamp, O. P., and Torres, O.: Direct radiative effect of aerosols based on PARASOL and OMI satellite observations, Journal of Geophysical Research : Atmospheres, 122, 2366–2388, https://doi.org/10.1002/2016JD025706, 2017.

Laj, P., Bigi, A., Rose, C., Andrews, E., Myhre, C. L., Coen, M. C., Lin, Y., Wiedensohler, A., Schulz, M., Ogren, J. A., and Fiebig, M.: A global analysis of climate-relevant aerosol properties retrieved from the network of Global Atmosphere Watch (GAW) near-surface observatories, Atmospheric Measurement Techniques, 13, 4353–4392, 2020.

Lequy, É., Conil, S., and Turpault, M.-P.: Impacts of Aeolian dust deposition on European forest sustainability: A review, Forest Ecology and Management, 267, 240–252, https://doi.org/10.1016/j.foreco.2011.12.005, http://linkinghub.elsevier.com/retrieve/pii/S0378112711007365, 2012.

Levelt, P. F., Hilsenrath, E., Leppelmeier, G. W., Oord, G. H. J. V. D., Bhartia, P. K., Tamminen, J., Haan, J. F. D., and Veefkind, J. P.: Science Objectives of the Ozone Monitoring Instrument, IEEE Trans. on Geoscience and Remote Sensing, 44, 1199–1208, 2006.

Li, X., Wagner, F., Peng, W., Yang, J., and Mauzerall, D. L.: Reduction of solar photovoltaic resources due to air pollution in China, Proceedings of the National Academy of Sciences, 114, 11 867–11 872, https://doi.org/10.1073/pnas.1711462114, 2017.

Litvinov, P., Hasekamp, O., and Cairns, B.: Remote Sensing of Environment Models for surface reflection of radiance and polarized radiance : Comparison with airborne multi-angle photopolarimetric measurements and implications for modeling top-of-atmosphere measurements, Remote Sensing of Environment, 115, 781–792, https://doi.org/10.1016/j.rse.2010.11.005, http://dx.doi.org/10.1016/j.rse.2010.11.005, 2011.

Lohmann, U. and Feichter, J.: Impact of sulfate aerosols on albedo and lifetime of clouds : A sensitivity study with the ECHAM4 GCM, Journal of Geophysical Research, 102, 13,685–13,700, 1997.

Lohmann, U. and Feichter, J.: Global indirect aerosol effects : a review, Atmospheric Chemistry and Physics, 5, 715–737, 2005.

Lopatin, A., Dubovik, O., Fuertes, D., Stnchikov, G., Lapyonok, T., Veselovskii, I., Wienhold, F. G., Shevchenko, I., Hu, Q., and Parajuli, S.: Synergy processing of diverse ground-based remote sensing and in situ data using GRASP algorithm : applications to radiometer , lidar and radiosonde observations, Atmos. Meas. Tech. Discussions, 2020.

Maher, B., Prospero, J., Mackie, D., Gaiero, D., Hesse, P., and Balkanski, Y.: Global connections between aeolian dust, climate and ocean biogeochemistry at the present day and at the last glacial maximum, Earth-Science Reviews, 99, 61–97, https://doi.org/10.1016/j.earscirev.2009.12.001, http://linkinghub.elsevier.com/retrieve/pii/S0012825210000024, 2010.

McTainsh, G. and Strong, C.: The role of aeolian dust in ecosystems, Geomorphology, 89, 39–54, https://doi.org/10.1016/j.geomorph.2006.07.028, http://linkinghub.elsevier.com/retrieve/pii/S0169555X06003564, 2007.

Mittaz, J. and Merchant, C. J.: Applying principles of metrology to historical Earth observations from satellites, Metrologia, 56, 2019.

Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevåg, A., Lamarque, J.-F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J.-H., Zhang, K., Zhang, H., and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, Atmospheric Chemistry and Physics, 13, 1853–1877, https://doi.org/10.5194/acp-13-1853-2013, http://www.atmos-chem-phys.net/13/1853/2013/, 2013.

Omar, A. H., Won, J.-g., Winker, D. M., Yoon, S.-c., Dubovik, O., and Mccormick, M. P.: Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements, Journal of Geophysical Research-Atmospheres, 110, https://doi.org/10.1029/2004JD004874, 2005.

Peers, F., Bellouin, N., Waquet, F., Ducos, F., Goloub, P., Mollard, J., Myhre, G., Skeie, R. B., Takemura, T., Tanré, D., Thieuleux, F., and Zhang, K.: Comparison of aerosol optical properties above clouds between POLDER and AeroCom models over the South East Atlantic Ocean during the fire season, Geophysical Research Letters, 43, 3991–4000, https://doi.org/10.1002/2016GL068222, 2016.

Phillips, P.: A technique for the numerical solution of certain integral equations of the first kind, J. Assoc. Comput. Mach., 9, 84–97, 1962.

Pitkänen, M. R. A., Mikkonen, S., Lehtinen, K. E. J., Lipponen, A., and Arola, A.: Artificial bias typically neglected in comparisons of uncertain atmospheric data, Geophysical Research Letters, 43, 10,003–10,011, https://doi.org/10.1002/2016GL070852, http://doi.wiley.com/10.1002/2016GL070852, 2016.

Popp, T., De Leeuw, G., Bingen, C., Brühl, C., Capelle, V., Chedin, A., Clarisse, L., Dubovik, O., Grainger, R., Griesfeller, J., Heckel, A., Kinne, S., Klüser, L., Kosmale, M., Kolmonen, P., Lelli, L., Litvinov, P., Mei, L., North, P., Pinnock, S., Povey, A., Robert, C., Schulz, M., Sogacheva, L., Stebel, K., Zweers, D. S., Thomas, G., Tilstra, L. G., Vandenbussche, S., Veefkind, P., Vountas, M., and Xue, Y.: Development, production and evaluation of aerosol climate data records from European satellite observations (Aerosol_cci), Remote Sensing, 8, https://doi.org/10.3390/rs8050421, 2016.

Russell, P. B., Kacenelenbogen, M., Livingston, J. M., Hasekamp, O. P., Burton, S. P., Schuster, G. L., Johnson, M. S., Knobelspiesse, K. D., Redemann, J., Ramachandran, S., and Holben, B.: A multiparameter aerosol classification method and its application to retrievals from spaceborne polarimetry, Journal of Geophysical Research: Atmospheres, 11, 9838–9863, https://doi.org/10.1002/2013JD021411, 2014.

Saide, P. E., Spak, S. N., Pierce, R. B., Otkin, J. A., Schaack, T. K., Heidinger, A. K., Da Silva, A. M., Kacenelenbogen, M., Redemann, J., and Carmichael, G. R.: Central American biomass burning smoke can increase tornado severity in the U.S., Geophysical Research Letters, 42, 956–965, https://doi.org/10.1002/2014GL062826, 2015.

Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Andrews, T., Faluvegi, G., Fläschner, D., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J. F., Olivié, D., Richardson, T., Shindell, D., Shine, K. P., Takemura, T., and Voulgarakis, A.: Fast and slow precipitation responses to individual climate forcers: A PDRMIP multimodel study, Geophysical Research Letters, 43, 2782–2791, https://doi.org/10.1002/2016GL068064, 2016.

Sayer, A. M., Thomas, G. E., Palmer, P. I., and Grainger, R. G.: Some implications of sampling choices on comparisons between satellite and model aerosol optical depth fields, Atmospheric Chemistry and Physics, 10, 10 705–10 716, https://doi.org/10.5194/acp-10-10705-2010, http://www.atmos-chem-phys.net/10/10705/2010/, 2010.

Schafer, J. S., Eck, T. F., Holben, B. N., Thornhill, K. L., Anderson, B. E., Sinyuk, A., Giles, D. M., Winstead, E. L., Ziemba, L. D., Beyersdorf, A. J., Kenny, P. R., Smirnov, A., and Slutsker, I.: Intercomparison of aerosol single-scattering albedo derived from AERONET surface radiometers and LARGE in situ aircraft profiles during the 2011 DRAGON-MD and DISCOVER-AQ experiments, Journal of Geophysical Research: Atmospheres, 119, 7439–7452, https://doi.org/10.1002/2013JD021166.Received, 2014.

Schmeisser, L., Backman, J., Ogren, J. A., Andrews, E., Asmi, E., Starkweather, S., Uttal, T., Fiebig, M., Sharma, S., Eleftheriadis, K., Vratolis, S., and Bergin, M.: Seasonality of aerosol optical properties in the Arctic, Atmospheric Chemistry and Physics, 18, 11 599–11 622, 2018.

Schmid, B., Michalsky, J., Halthore, R., Beauharnois, M., Harnson, L., Livingston, J., Russell, P., Holben, B., Eck, T., and Smirnov, A.: Comparison of Aerosol Optical Depth from Four Solar Radiometers During the Fall 1997 ARM Intensive Observation Period, Geophysical Research Letters, 26, 2725–2728, 1999.

Schutgens, N., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M., and Stier, P.: Will a perfect model agree with perfect observations? The impact of spatial sampling, Atmospheric Chemistry and Physics Discussions, 16, https://doi.org/10.5194/acp-2015-973, http://www.atmos-chem-phys-discuss.net/acp-2015-973/, 2016a.

Schutgens, N., Partridge, D. G., and Stier, P.: The importance of temporal collocation for the evaluation of aerosol models with observations, Atmospheric Chemistry and Physics, 16, 1065–1079, https://doi.org/10.5194/acp-16-1065-2016, 2016b.

Schutgens, N., Tsyro, S., Gryspeerdt, E., Goto, D., Weigum, N., Schulz, M., and Stier, P.: On the spatio-temporal representativeness of observations, Atmospheric Chemistry and Physics, 17, 9761–9780, https://doi.org/10.5194/acp-2017-149, https://www.atmos-chem-phys-discuss.net/acp-2017-149/, 2017.

Schutgens, N., Sayer, A. M., Heckel, A., Hsu, C., Jethva, H., Leeuw, G. D., Leonard, P. J. T., Levy, R. C., Lipponen, A., Lyapustin, A., North, P., Popp, T., Poulsen, C., Sawyer, V., Sogacheva, L., Thomas, G., Torres, O., Wang, Y., Kinne, S., Schulz, M., and Stier, P.: An AeroCom – AeroSat study : intercomparison of satellite AOD datasets for aerosol model evaluation, Atmospheric Chemistry and Physics, 20, 12 431–12 457, 2020.

Schutgens, N. A. J.: Site representativity of AERONET and GAW remotely sensed aerosol optical thickness and absorbing aerosol optical thickness observations, Atmospheric Chemistry and Physics, 20, 7473–7488, 2020.

Schwarz, J. P., Spackman, J. R., Gao, R. S., Watts, L. A., Stier, P., Schulz, M., Davis, S. M., Wofsy, S. C., and Fahey, D. W.: Global-scale black carbon profiles observed in the remote atmosphere and compared to models, Geophysical Research Letters, 37, https://doi.org/10.1029/2010GL044372, 2010.

Schwarz, J. P., Samset, B. H., Perring, A. E., Spackman, J. R., Gao, R. S., Stier, P., Schulz, M., Moore, F. L., Ray, E. A., and Fahey, D. W.: Global-scale seasonally resolved black carbon vertical profiles over the Pacific, Geophysical Research Letters, 40, 5542–5547, https://doi.org/10.1002/2013GL057775, 2013.

Sinyuk, A., Holben, B. N., Eck, T. F., Giles, D. M., Slutsker, I., Korkin, S., Schafer, J. S., Smirnov, A., Sorokin, M., and Lyapustin, A.: The AERONET Version 3 aerosol retrieval algorithm , associated uncertainties and comparisons to Version 2, Atmospheric Measurement Techniques, 13, 3375–3411, 2020.

Smirnov, A., Holben, B. N., Eck, T. F., Dubovik, O., and Slutsker, I.: Cloud-Screening and Quality Control Algorithms for the AERONET Database, Remote Sensing of Environment, 73, 337–349, 2000.

Smith, K. R., Jerrett, M., Anderson, H. R., Burnett, R. T., Stone, V., Derwent, R., Atkinson, R. W., Cohen, A., Shonkoff, S. B., Krewski, D., Pope, C. A., Thun, M. J., and Thurston, G.: Public health benefits of strategies to reduce greenhouse-gas emissions: health implications of short-lived greenhouse pollutants., Lancet, 374, 2091–103, https://doi.org/10.1016/S0140-6736(09)61716-5, http://www.ncbi.nlm.nih.gov/pubmed/19942276, 2009.

Stap, F. A., Hasekamp, O. P., and Röckmann, T.: Sensitivity of PARASOL multi-angle photopolarimetric aerosol retrievals to cloud contamination, Atmospheric Measurement Techniques, 8, 1287–1301, https://doi.org/10.5194/amt-8-1287-2015, https://www.atmos-meas-tech.net/8/1287/2015/, 2015.

Swap, R., Garstang, M., Greco, S., Talbot, R., and Kallberg, P.: Saharan dust in the Amazon Basin, Tellus, 44B, 133–149, https://doi.org/10.1034/j.1600-0889.1992.t01-1-00005.x, 1992.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophysical Research, 106, 7183–7192, 2001.

Tegen, I. and Heinold, B.: Large-Scale Modeling of Absorbing Aerosols and Their Semi-Direct Effects, Atmosphere, 9, https://doi.org/10.3390/atmos9100380, 2018.

Thomas, J. L., Polashenski, C. M., Soja, A. J., Marelle, L., Casey, K. A., Choi, H. D., Raut, J. C., Wiedinmyer, C., Emmons, L. K., Fast, J. D., Pelon, J., Law, K. S., Flanner, M. G., and Dibb, J. E.: Quantifying black carbon deposition over the Greenland ice sheet from forest fires in Canada, Geophysical Research Letters, 44, 7965–7974, https://doi.org/10.1002/2017GL073701, 2017.

Tikhonov, A.: On the solution of incorrectly stated problems and a method of regularization, Dokl. Akad. Nauk SSSR, 151, 501–504, 1963.

Torres, B., Dubovik, O., Toledano, C., Berjon, A., Cachorro, V. E., Lapyonok, T., Litvinov, P., and Goloub, P.: Sensitivity of aerosol retrieval to geometrical configuration of ground-based sun/sky radiometer observations, Atmospheric Chemistry and Physics, 14, 847–875, https://doi.org/10.5194/acp-14-847-2014, http://www.atmos-chem-phys.net/14/847/2014/, 2014.

Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P. K., Veefkind, P., and Levelt, P.: Aerosols and surface UV products from Ozone Monitoring Instrument observations : An overview, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/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, Atmospheric Measurement Techniques, 6, 3257–3270, https://doi.org/10.5194/amt-6-3257-2013, 2013.

Tosca, M. G., Randerson, J. T., and Zender, C. S.: Global impact of smoke aerosols from landscape fires on climate and the Hadley circulation, Atmospheric Chemistry and Physics, 13, 5227–5241, https://doi.org/10.5194/acp-13-5227-2013, 2013.

Tsikerdekis, A., Schutgens, N. A. J., Hasekamp, O. P., and Amsterdam, V. U.: Assimilating aerosol optical properties related to size and absorption from POLDER / PARASOL with an ensemble data assimilation system, Atmospheric Chemistry and Physics, 21, 2637–2674, 2021.

Twomey, S.: Pollution and the planetary albedo, Atmospheric Environment, 8, 1251–1256, 1974.

Vink, S. and Measures, C.: The role of dust deposition in determining surface water distributions of Al and Fe in the South West Atlantic, Deep Sea Research Part II, 48, 2787–2809, https://doi.org/10.1016/S0967-0645(01)00018-2, http://linkinghub.elsevier.com/retrieve/pii/S0967064501000182, 2001.

Virtanen, T. H., Kolmonen, P., Sogacheva, L., Rodríguez, E., Saponaro, G., and Leeuw, G. D.: Collocation mismatch uncertainties in satellite aerosol retrieval validation, Atmospheric Measurement Techniques, 11, 925–938, 2018.

Wang, Y., Sartelet, K. N., Bocquet, M., and Chazette, P.: Modelling and assimilation of lidar signals over Greater Paris during the MEGAPOLI summer campaign, Atmospheric Chemistry and Physics, 14, 3511–3532, https://doi.org/10.5194/acp-14-3511-2014, http://www.atmos-chem-phys.net/14/3511/2014/, 2014.

Watson-Parris, D., Schutgens, N., Cook, N., Kipling, Z., Kershaw, P., Gryspeerdt, E., Lawrence, B., and Stier, P.: Community Intercomparison Suite (CIS) v1.4.0: A tool for intercomparing models and observations, Geoscientific Model Development, 9, https://doi.org/10.5194/gmd-9-3093-2016, 2016.

WMO: SYSTEMATIC OBSERVATION REQUIREMENTS FOR SATELLITE-BASED DATA PRODUCTS FOR CLIMATE; 2011 Update Supplemental details to the satellite-based component of the “Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC", Tech. rep., 2011.

Wu, L., Hasekamp, O., van Diedenhoven, B., and Cairns, B.: Aerosol retrieval from multiangle, multispectral photopolarimetric measurements: Importance of spectral range and angular resolution, Atmos. Meas. Tech., 8, 2625–2638, https://doi.org/10.5194%2Famt-8-2625-2015, 2015.

Zhang, Y., Forrister, H., Liu, J., DIbb, J., Anderson, B., Schwarz, J. P., Perring, A. E., Jimenez, J. L., Campuzano-Jost, P., Wang, Y., Nenes, A., and Weber, R. J.: Top-of-atmosphere radiative forcing affected by brown carbon in the upper troposphere, Nature Geoscience, 10, 486–489, https://doi.org/10.1038/ngeo2960, 2017.

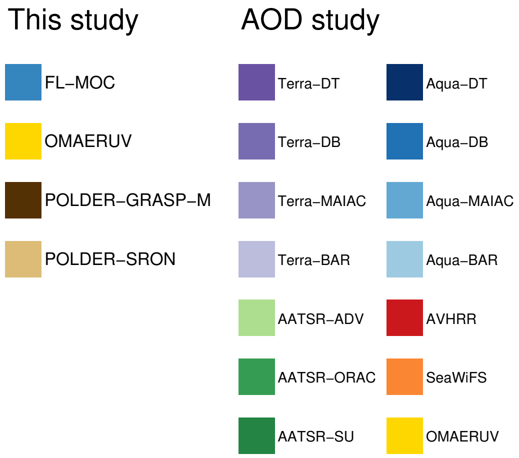

Figure 1. Colour legend used throughout this paper to designate the different satellite products, for both this study and the AOD study in Schutgens et al. (2020).

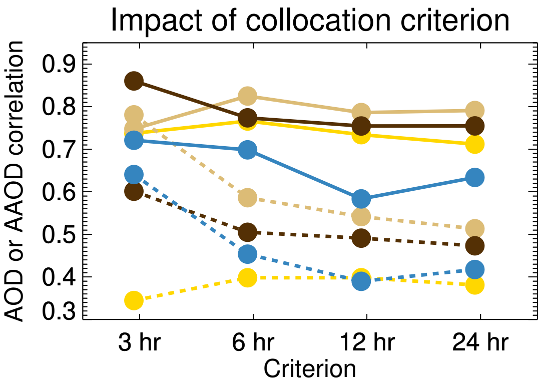

Figure 2. Correlation of satellite AOD (solid) and AAOD (dashed) with AERONET Inversion L2.0 data, as a function of temporal collocation criterion. Colours indicate satellite product, see also Fig 1. Satellite products were individually collocated with AERONET.

Appendix A: Generic aggregation and collocation

The aggregation of satellite L2 products into super-observations in this paper, and the subsequent collocation of different datasets for intercomparison and evaluation used the following scheme.

Assume a homogenous L2 dataset with times and geo-locations and observations of AOD and AAOD. Homogenous means that AOD and AAOD are available for the same times, geo-locations and wavelengths. Each observation has a known spatio-temporal foot-print, e.g. in the case of satellite L2 retrievals that would be the L2 retrieved pixel size and the short amount of time (less than a second) needed for the original measurement.

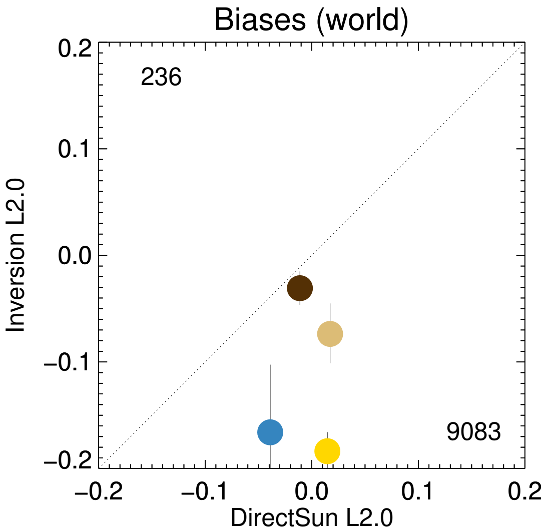

Figure 3. Global biases in four satellite AOD datasets depending on the chosen reference dataset (DirectSun or Inversion). Colours indicate satellite product, see also Fig 1. Numbers in upper left and lower right corner indicate amount of collocated data, averaged over all products. Error ranges indicate 5-95% uncertainty ranges based on a bootstrap analysis, see Sect. 3.2. Satellite products were individually collocated with AERONET, within 3 hours.

Satellite L2 data are aggregated into super-observations as follows. A regular spatio-temporal grid is defined as in Fig. A1. The spatio-temporal size of the grid-boxes (here $1^o \times 1^o \times 30^{\mathrm{min}}$) exceeds that of the footprint of the L2 data that will be aggregated. All observations are assigned to a spatio-temporal grid-box according to their times and geo-locations. Once all observations have been assigned, observations are averaged by grid-box. It is possible to require a minimum number of observations to calculate an average. Finally, all grid-boxes that contain observations are used to construct a list of super-observations as in Fig. A2. Only times and geo-locations with aggregated observations are retained. As the original L2 dataset was homogeneous, so is the resulting L3 dataset.

Station data is similarly aggregated over $1^o \times 1^o \times 30^{\mathrm{min}}$. Point observations will suffer from spatial representativeness issues (Sayer et al., 2010; Virtanen et al., 2018; Schutgens et al., 2016a), but the representativity of AERONET sites for $1^o \times 1^o$ grid-boxes is fairly well understood (Schutgens, 2020), see also Section 2.1.5. These aggregated L3 AERONET data will also be called super-observations.

Different datasets of super-observations can be collocated in a very similar way. Again a regular spatio-temporal grid is defined as in Fig. A1 but now with grid-boxes of larger temporal extent (typically $1^o \times 1^o \times 3^{\mathrm{hr}}$). Because this temporal extent is short compared to satellite revisit times, either a single satellite super-observation or none is assigned to each grid-box. A single AERONET site however may contribute up to 6 super-observations per grid-box (in which case they are averaged). After two or more datasets are thus aggregated *individually*, only grid-boxes that contain data for both datasets will be used

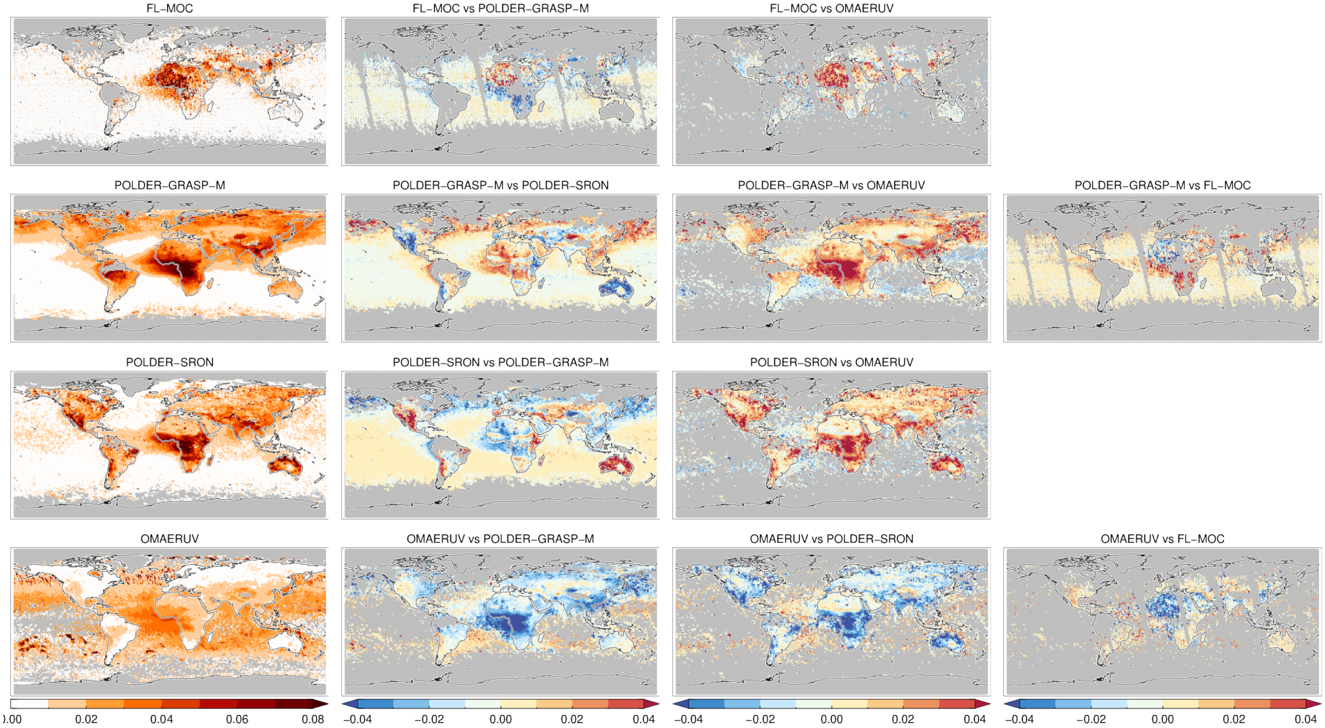

Figure 4. Global maps of AAOD for four products, and their differences. AAOD differences are based on collocated data (within 3 hours). Note that the products are available for different years, e.g. POLDER-SRON and FL-MOC do not overlap. No minimum AOD was required.

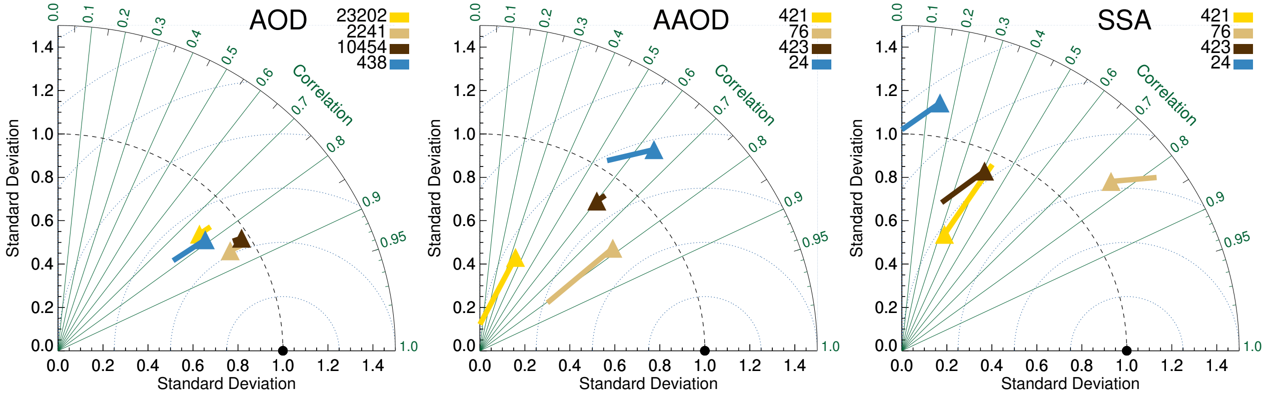

Figure 5. Taylor diagrams 3.1 for the satellite products. AOD is evaluated against AERONET DirectSun L2.0, AAOD and SSA are evaluated against AERONET Inversion L2.0. Colours indicate satellite product (see also Fig. 1), numbers next to coloured blocks indicate amount of collocated data. The lines extending from the data points indicate the bias. Products were individually collocated with AERONET, within 3 hours.

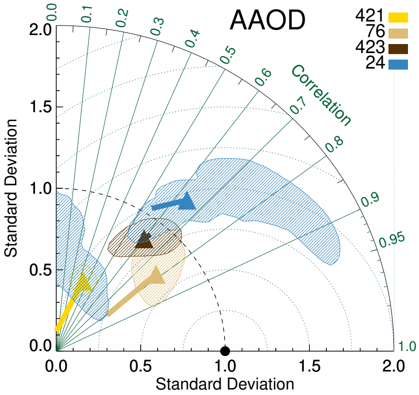

Figure 6. Impact of statistical noise on the correlation and internal variability of satellite AAOD products, using bootstrapping. Shaded regions indicate $5\% - 95\%$ uncertainty range of correlation and standard deviation (uncertainty in bias is not shown). Colours indicate satellite product, see also Fig 1, numbers next to coloured blocks indicate amount of collocated data. Satellite products were individually collocated with AERONET Inversion L2.0 within 3 hours.

to construct two lists of aggregated data as in Fig. A2. Those two lists will have identical size and ordering of times and geolocations and are called collocated datasets. By choosing a larger temporal extent of the grid-box, the collocation criterion can be relaxed.

As the super-observations are on a regular spatio-temporal grid and collocation requires further aggregation to another regular but coarser, grid, the whole procedure is very fast. It is possible to collocate 7 products from afternoon platforms over three years using an IDL (Interactive Data Language) code (that served as a prototype for CIS) and a single processing core in just 30 minutes (Schutgens et al., 2020). This greatly facilitates sensitivity studies.

Starting from super-observations, a 3-year average can easily be constructed by once more performing an aggregation operation but now with a grid-box of $1^o \times 1^o \times^{\mathrm{yr}}$. If two *collocated* datasets are aggregated in this fashion, their 3-year average can be compared with minimal representation errors. This allows us to construct global maps of e.g. multi-year AOD difference between two sets of super-observations.

A software tool (the Community Intercomparison Suite) is available for these operations at `www.cistools.net` (last accessed on December 20, 2019) and is described in great detail in Watson-Parris et al. (2016).

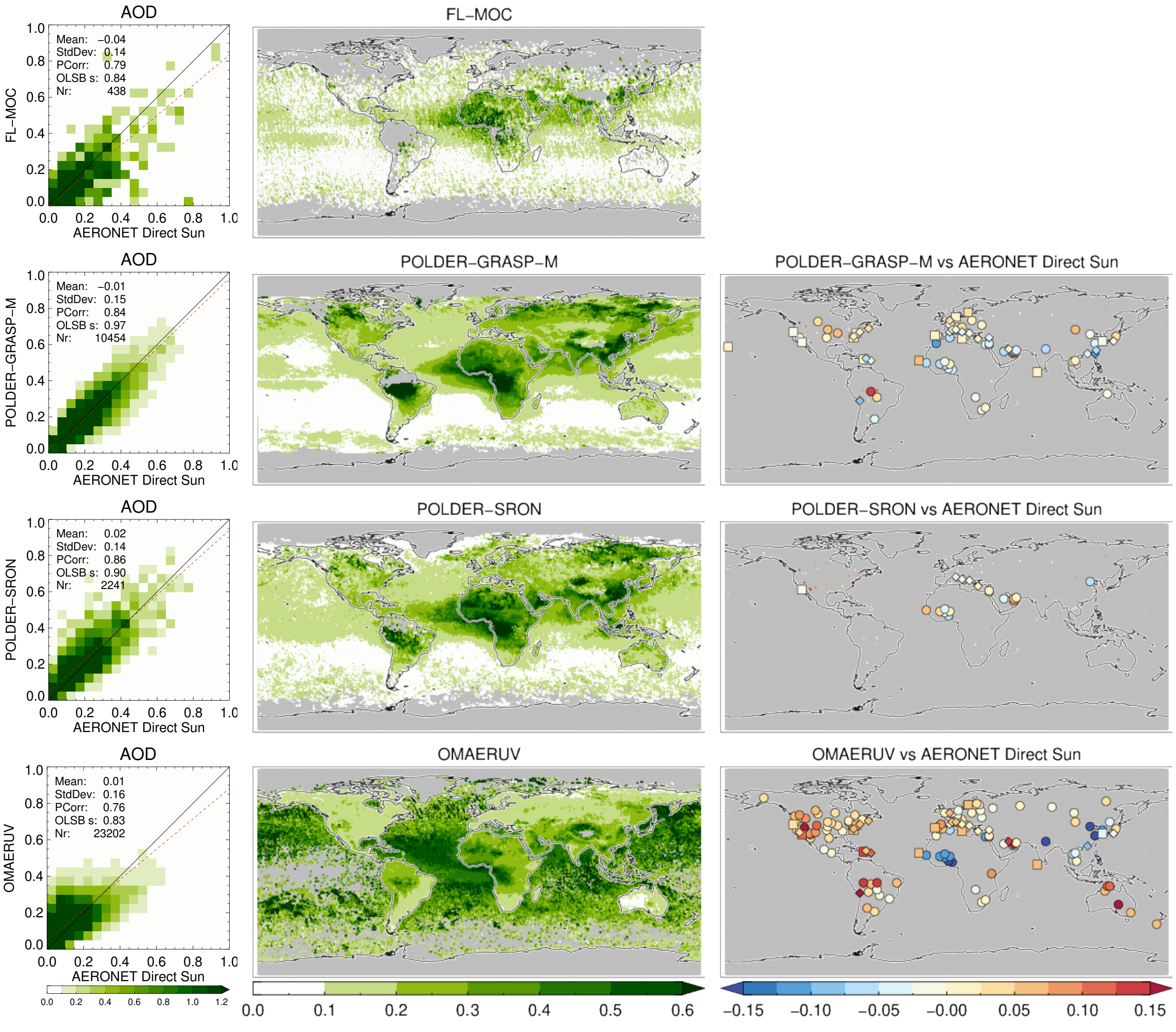

Figure 7. For the four satellite products are shown: a scatter plot of individual super-observations versus AERONET (the colour indicates amount of data in percentages, see Sect. 3.3 for an explanation of the metrics); a global map of the three-year AOD average; a global map of the three-year AOD difference average with AERONET (if site provided at least 32 observations; land sites are circles, ocean sites are squares, diamonds are the remainder). For FL-MOC, insufficient data prevent the plotting of a difference map. Products were individually collocated with AERONET DirectSun L2.0 within 3 hours.

Table 1. Remote sensing products used in this study

Platform	Overpass [hr]	Sensor	Swath [km]	Pixel [km]	Product	(A)AOD[1] 550nm	Years	References
Aqua/AURA/ CALIPSO	1:30PM	MODIS/OMI/ CALIOP	1	1	FL-MOC[1]	R	2007, ’08	Kacenelenbogen et al. (2019)
AURA	1:30PM	OMI	2600	18	OMAERUV v1.8.9.1	E	2006, ’08, ’10	Ahn et al. (2014); Jethva et al. (2014)
PARASOL	1:30PM[2]	POLDER	1600	6.18	POLDER-GRASP-M v1.2	I	2006, ’08, ’10	Dubovik et al. (2011); Chen et al. (2020)
PARASOL	1:30PM[2]	POLDER	1600	6.18	POLDER-SRON	I	2006	Hasekamp and Landgraf (2005) Hasekamp et al. (2011)

[1] This product uses a combination of Aqua-MODIS, OMI and CALIOP observations

[2] PARASOL started drifting away from Aqua at the end of 2009.

[3] Interpolated or Extrapolated to 550 nm, depending on surface type; or Retrieved at 550 nm

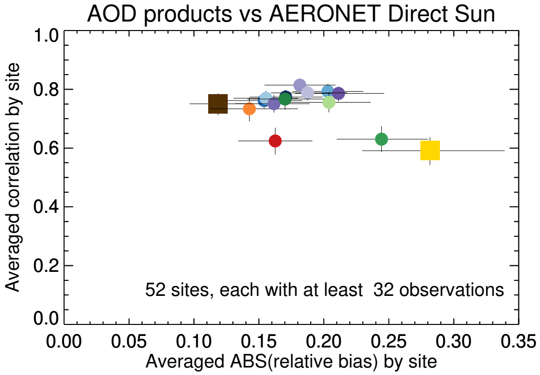

Figure 8. Evaluation of satellite products with AERONET per site, averaged over all sites. Squares indicate products used in the present study, circles indicate products used in Schutgens et al. (2020). Error bars indicate 5-95% uncertainty range based on a bootstrap analysis (see Sect. 3.2) of sample size 1000 (the bootstrap was performed on the contributing AERONET sites). Colours indicate satellite product, see also Fig. 1. Products were individually collocated with AERONET DirectSun L2.0 within 3 hour. All products use the same sites, each of which produced at least 32 collocations. POLDER-SRON and FL-MOC were excluded from this analysis due to lack of data.

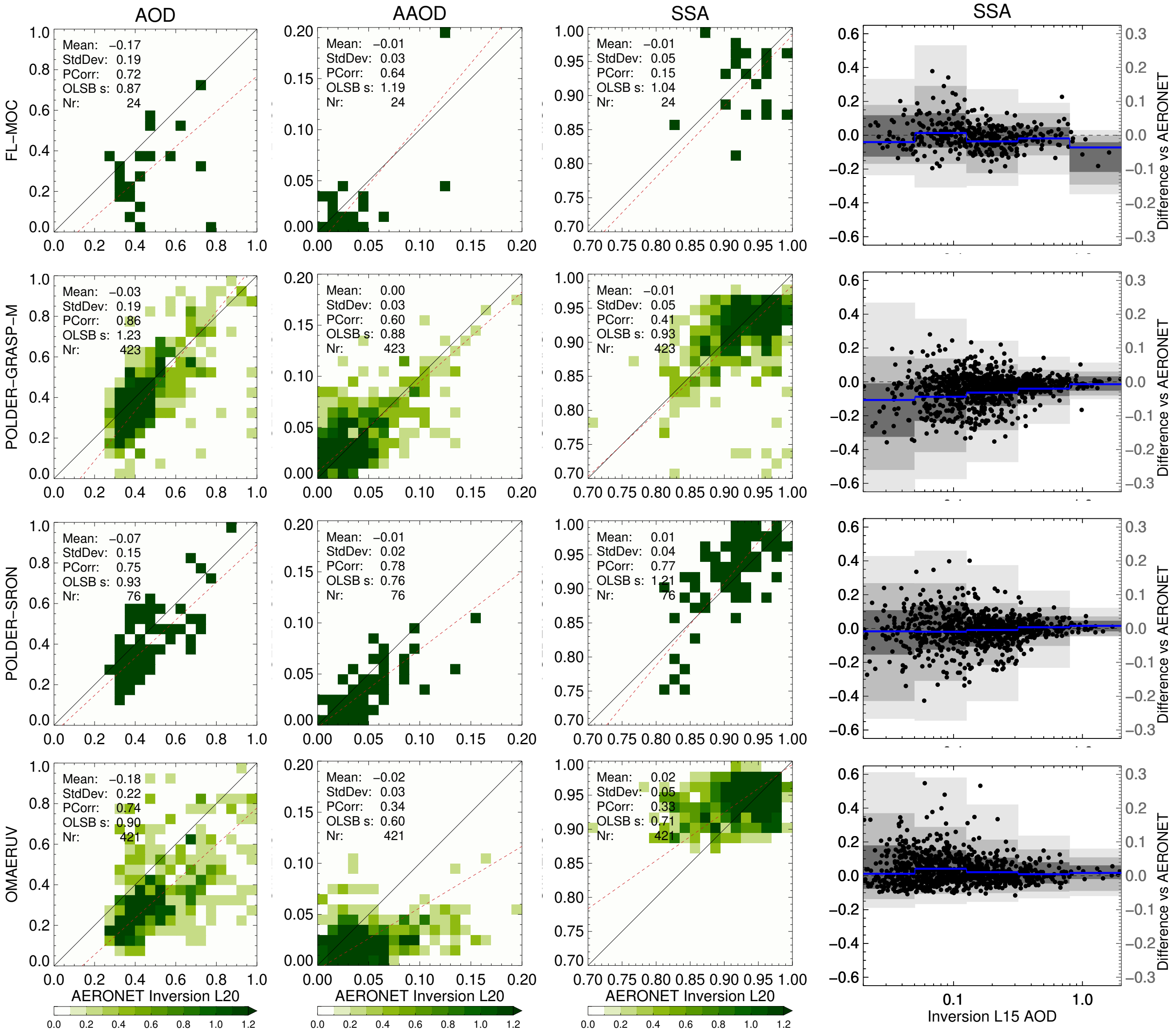

Figure 9. Evaluation of super-observations of AOD, AAOD and SSA for the satellite products. SSA is also evaluated as a function of AOD (binned). In the three left-most figures, the colour indicates amount of data in percentages; for an explanation of the metrics, see Sect. 3.3. The right-most column uses two vertical axes: the left-hand side is used for individual data points (sub-sampled), the right-hand axis is used for the grey-scale distribution $(9, 25, 50, 75, 91\%$ quantiles) and the median difference (blue line). Products were individually collocated with AERONET Inversion L2.0 within 3 hour, except the right-most column which used Inversion L1.5.

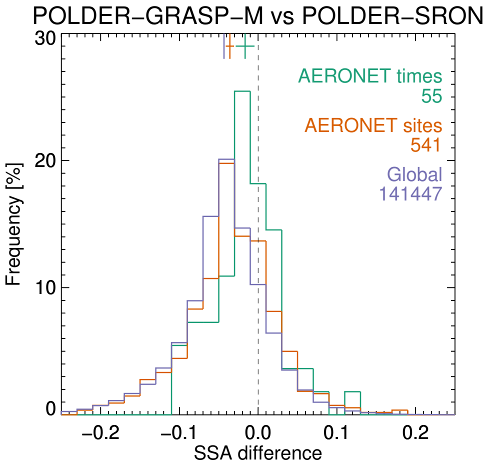

Figure 10. SSA differences POLDER-GRASP-M vs. POLDER-SRON for three different samplings: all available data, data available over AERONET sites that provide Inversion L2.0 data, data available at the times and locations of Inversion L2.0 data. The vertical coloured lines at the top show distribution means and the short horizontal lines extending from the middle show 2σ ranges. The dashed vertical line shows zero difference. Number of collocated data are indicated in the figure as well. This analysis suggests that an evaluation with AERONET would underestimate the actual difference between the two products. In all cases, data was collocated within 3 hours and a minimum AOD > 0.25 was required.

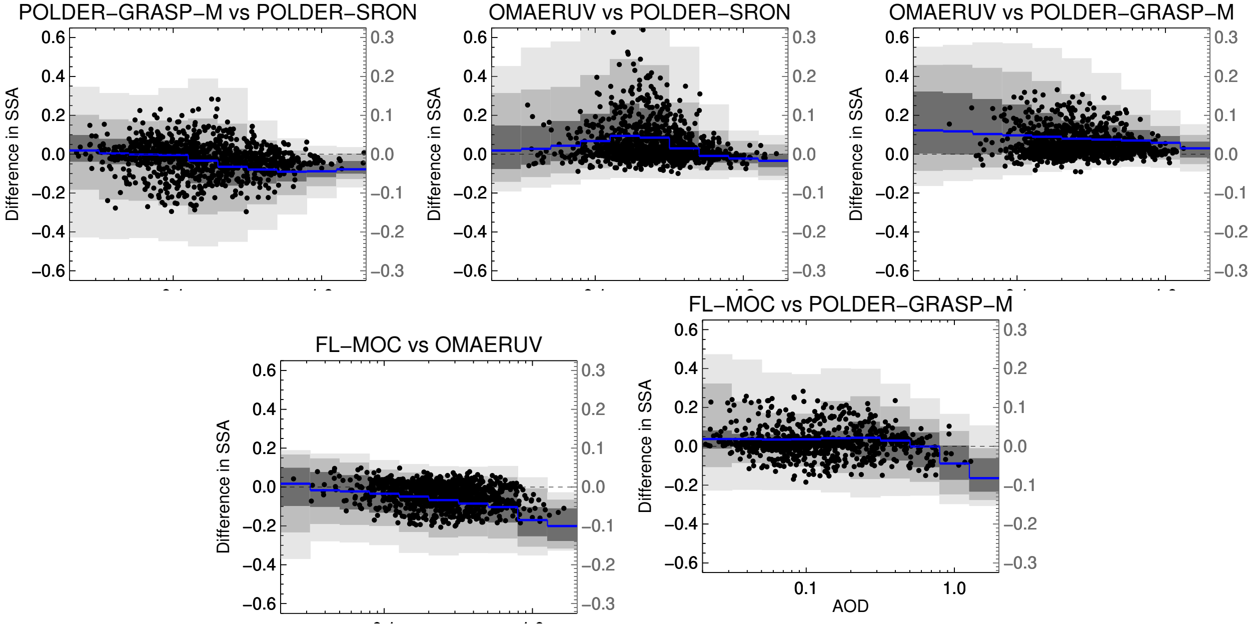

Figure 11. Difference in satellite product SSA as a function of AOD (averaged over both products). Two vertical axes are used: the left-hand side is used for individual data points (sub-sampled), the right-hand axis is used for the grey-scale distribution $(9, 25, 50, 75, 91\%$ quantiles) and the median difference (blue line). Data were collocated within 3 hours.

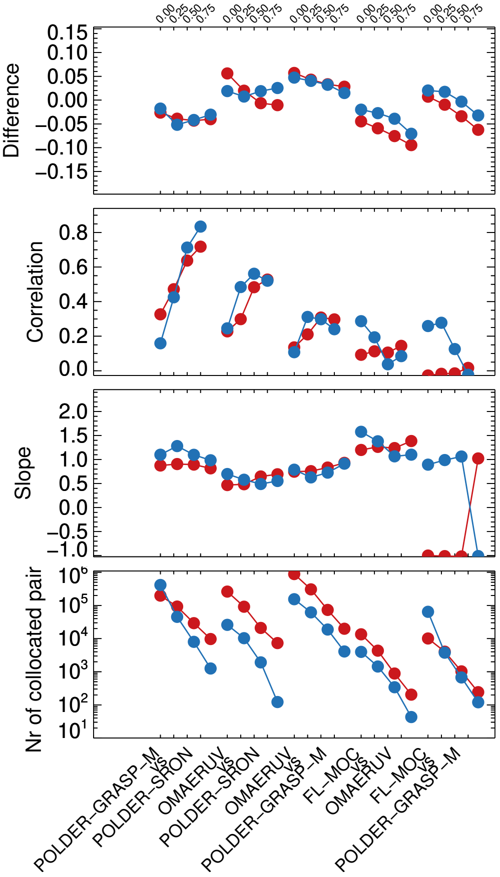

Figure 12. Comparison of different pairs of satellite SSA, over land (red) and ocean (blue), for different thresholds of minimum AOD (0.0, 0.25, 0.5, and 0.75). The data were collocated within 3 hours.

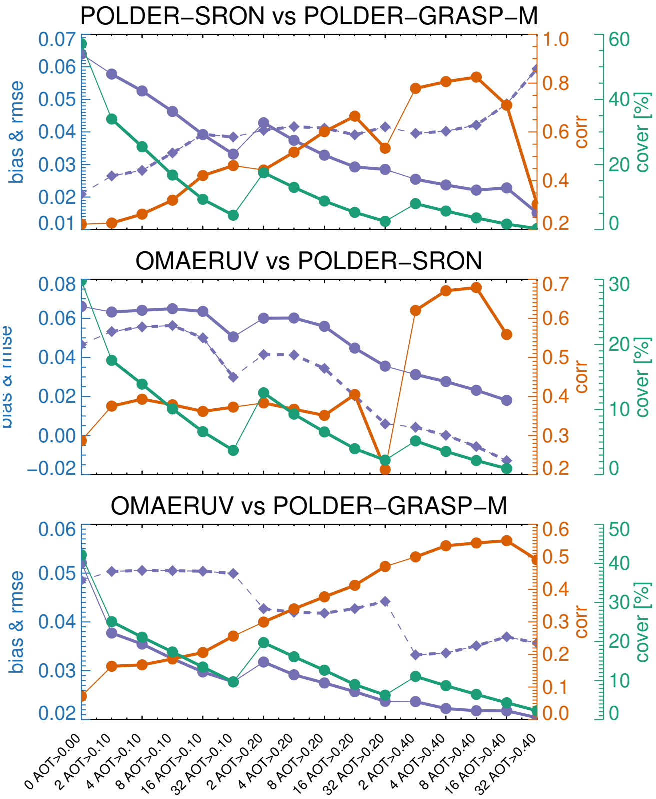

Figure 13. Intercomparison of SSA satellite products after multi-year averaging, as a function of minimum AOD and number of collocated observations (thicker lines group cases with the same minimum AOD but increasing number of observations). Bias uses a dashed line, and RMSE a solid line. Cover is defined as fraction of surface area covered by data. FL-MOC is not present due to scarcity of observations. The data were collocated within 3 hours.

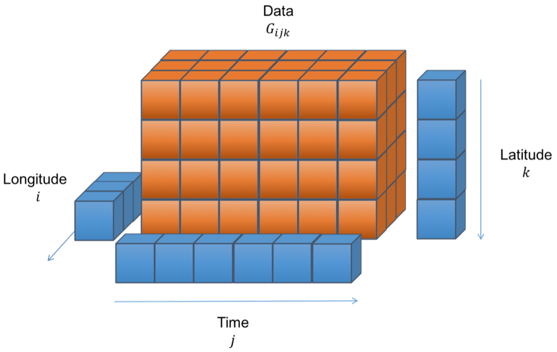

Figure A1. A regular spatio-temporal grid in time, longitude and latitude. Such a grid is used for the aggregation operation that is at the heart of the collocation procedure used in this paper. Grid-boxes may either contain data or be empty. Note that data may refer to any combination of observations, e.g. AOD at multiple wavelengths or AOD and AAOD at 550 nm. However, the dataset is homogenous. Reproduced from Watson-Parris et al. (2016).

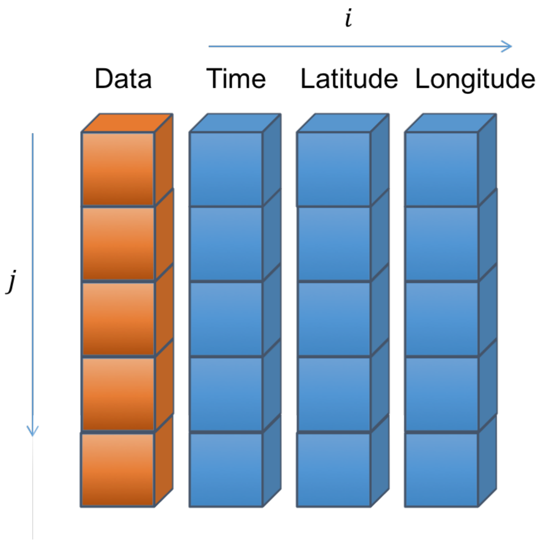

Figure A2. A list of data. Such a list is the primary data format used for the observations in this paper. Note that data may refer to any combination of observations, e.g. AOD at multiple wavelengths or AOD and AAOD at 550 nm. However, the dataset is homogenous. Reproduced from Watson-Parris et al. (2016).