A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing

Andrew M. Sayer¹,², Yves Govaerts³, Pekka Kolmonen⁴, Antti Lipponen⁴, Marta Luffarelli³, Tero Mielonen⁴, Falguni Patadia¹,², Thomas Popp⁵, Adam C. Povey⁶, Kerstin Stebel⁷, and Marcin L. Witek⁸

¹GESTAR, Universities Space Research Association, Columbia, MD, USA
²NASA Goddard Space Flight Center, Greenbelt, MD, USA
³Rayference, 1030 Brussels, Belgium
⁴Finnish Meteorological Institute, Atmospheric Research Centre of Eastern Finland, Kuopio, Finland
⁵Deutsches Zentrum für Luft-und Raumfahrt e. V. (DLR), Deutsches Fernerkundungsdatenzentrum (DFD), 82234 Oberpfaffenhofen, Germany
⁶National Centre for Earth Observation, University of Oxford, Oxford, OX1 3PU, UK
⁷Atmosphere and Climate Department, NILU – Norwegian Institute for Air Research, 2007 Kjeller, Norway
⁸Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Correspondence: Andrew M. Sayer (andrew.sayer@nasa.gov)

Received: 16 August 2019 – Discussion started: 24 September 2019
Revised: 19 December 2019 – Accepted: 7 January 2020 – Published: 3 February 2020

Abstract. Recent years have seen the increasing inclusion of per-retrieval prognostic (predictive) uncertainty estimates within satellite aerosol optical depth (AOD) data sets, providing users with quantitative tools to assist in the optimal use of these data. Prognostic estimates contrast with diagnostic (i.e. relative to some external truth) ones, which are typically obtained using sensitivity and/or validation analyses. Up to now, however, the quality of these uncertainty estimates has not been routinely assessed. This study presents a review of existing prognostic and diagnostic approaches for quantifying uncertainty in satellite AOD retrievals, and it presents a general framework to evaluate them based on the expected statistical properties of ensembles of estimated uncertainties and actual retrieval errors. It is hoped that this framework will be adopted as a complement to existing AOD validation exercises; it is not restricted to AOD and can in principle be applied to other quantities for which a reference validation data set is available. This framework is then applied to assess the uncertainties provided by several satellite data sets (seven over land, five over water), which draw on methods from the empirical to sensitivity analyses to formal error propagation, at 12 Aerosol Robotic Network (AERONET) sites. The AERONET sites are divided into those for which it is expected that the techniques will perform well and those for which some complexity about the site may provide a more severe test. Overall, all techniques show some skill in that larger estimated uncertainties are generally associated with larger observed errors, although they are sometimes poorly calibrated (i.e. too small or too large in magnitude). No technique uniformly performs best. For powerful formal uncertainty propagation approaches such as optimal estimation, the results illustrate some of the difficulties in appropriate population of the covariance matrices required by the technique. When the data sets are confronted by a situation strongly counter to the retrieval forward model (e.g. potentially mixed land–water surfaces or aerosol optical properties outside the family of assumptions), some algorithms fail to provide a retrieval, while others do but with a quantitatively unreliable uncertainty estimate. The discussion suggests paths forward for the refinement of these techniques.

1 Introduction

The capability to quantify atmospheric aerosols from spaceborne measurements arguably goes back to 1972 with the launch of the Multispectral Scanner System (MSS) aboard the first Landsat satellite (e.g. Griggs, 1975; Kaufman and...
Sensors, which are able to provide extinction profiles; in many cases, two or more of each type of design, either identical or different, have been used to process AOD data products, and Table 2 lists the full names. Table 1 lists those sensors which have to date been used to retrieve total-column aerosol optical depth (AOD) across some part(s) of the ultraviolet (UV), visible, near-infrared and shortwave infrared, and tIR spectral regions, where aerosol particles are optically active; the most commonly reported is the mid-visible AOD at a wavelength in the range 500–565 nm. Some sensors are able to retrieve profiles of aerosol extinction, which may be integrated vertically to give partial- or total-column AOD (dependent on whether or not profiling is possible down to the surface). This proliferation, combined with geophysical and mathematic terminology, makes aerosol remote sensing an incredibly acronym-heavy field; indeed, instruments and algorithms are often referred to by their acronyms rather than full names. Table 1 lists those sensors which have to date been used to process AOD data products, and Table 2 lists those which are able to provide extinction profiles; in many cases, two or more of each type of design, either identical or with small modifications, have been flown. Where multiples of a given sensor have flown the date ranges indicate period(s) of continuous coverage as opposed to launch or decommission dates for individual instruments.

Retrieval algorithms are used to process the calibrated observations (referred to as level 1 or L1 data) to provide level 2 (L2) data products, consisting of geophysical quantities of interest. These L2 products are typically on the L1 satellite observation grid (or a multiple of it) and often further aggregated to level 3 (L3) products on regular space–time grids. For further background and a discussion of satellite data processing levels, see Mittaz et al. (2019). Table 3 provides acronyms and full names for some of the L2 processing algorithms which have been applied to L1 measurements from these instruments. Again, many of these algorithms have been applied (identically or with small modifications) to multiple sensors. This table is provided as a convenience to the reader to decode acronyms and decrease clutter in later tables and discussions; specific relevant details and references are provided later. Acronyms often summarize either the principle of the technique or the institution(s) which developed the algorithm. Some algorithms are not listed in this table as they do not have acronyms and are typically referred to by data producers or users by the sensor or mission name. Further, this is not an exhaustive list as numerous other approaches have been proposed in the literature; the criteria for inclusion and broader discussion in this study are that data have been (1) processed and (2) also made generally available for scientific use. Likewise, algorithms which provide aerosol properties as a by-product but not a focus (e.g. land–ocean surface atmospheric correction approaches) are not discussed as often the aerosol components are less detailed and/or used as a sink for other error sources in the algorithm (e.g. Kahn et al., 2016).

L2 retrieval algorithm development is typically guided by information content studies, sensitivity analyses, and retrieval simulations to gauge which quantities a given sensor and algorithmic approach can retrieve and with what uncertainty (e.g. Tanré et al., 1996, 1997; Hasekamp and Landgraf, 2007; Veihelmann et al., 2007; Young and Vaughan, 2009). As aerosol remote sensing is an underdetermined problem and there is considerable heterogeneity in the underlying (surface and atmospheric) conditions giving rise to the L1 signals, sensitivities and uncertainties are typically highly context-dependent. For example, the retrieval of AOD from optical sensors over a dark ocean surface is typically much easier than over a bright snow-covered surface. After an algorithm has been developed, these analyses are typically complemented by validation against reference data sets, most commonly AOD from Sun photometers such as part of the Aerosol Robotic Network (AERONET; Holben et al., 1998) over land and from handheld instruments deployed on ocean cruises in the Maritime Aerosol Network (MAN; Smirnov et al., 2009, 2011). The resulting uncertainty estimates provided by these studies and validation analyses are diagnostic; i.e. for a known true state they diagnose the retrieval error (difference between retrieved and true states). This is useful to identify the general tendencies for bias or loss of sensitivity under different conditions and assess potential ways to improve on them.

Increases in the quality of instrumentation, retrieval algorithms, models, and computational power have prompted an increasing desire for the provision of pixel-level uncertainty estimates in L2 aerosol data products. This has been driven in part by data assimilation (DA) applications, which need a robust error model on data for ingestion into numerical models (Benedetti et al., 2018), often in near-real time. Diagnostic uncertainty estimates are less useful here since the
true state is not known (only the retrieved state), so a prognostic (predictive) uncertainty model is needed instead. Early aerosol DA applications either treated diagnostic uncertainty estimates as prognostic ones (e.g. Collins et al., 2001; Matsui et al., 2004) or constructed their own prognostic error models as part of validation and bias-correction efforts (e.g. Zhang and Reid, 2006; Benedetti et al., 2009; Hyer et al., 2011; Shi et al., 2013). These uncertainty estimates are also valuable outside DA to identify when a retrieval is likely to be useful for a given purpose. As an example, air quality modelling also typically uses L2 retrievals and can benefit from these uncertainties. Climate applications often use L3 aerosol data for which uncertainty estimates have yet to be robustly developed; this is an important emerging area of research regarding both methods of aggregation and/or reporting (e.g. Levy et al., 2009; Kinne et al., 2017; Povey and Grainger, 2019; Sayer and Knobelspiesse, 2019) and the influence of sampling (e.g. Sayer et al., 2010b; Colarco et al., 2014; Geogdzhayev et al., 2014; Schutgens et al., 2016, 2017), and L2 uncertainty estimates will be an important input to this.

Driven by these needs, many AOD data sets now provide prognostic uncertainty estimates; in some cases these addi-

| Acronym | Instrument full name | Orbit(s) | Operation period(s) |
|---------|----------------------|----------|---------------------|
| ABI     | Advanced Baseline Imager | Geostationary | 2016+ |
| AHI     | Advanced Himawari Imager | Geostationary | 2014+ |
| AVHRR   | Advanced Very High Resolution Radiometer | Sun-synchronous | 1978+ |
| CAI     | Cloud–Aerosol Imager | Sun-synchronous | 2009+ |
| EPIC    | Earth Polychromatic Imaging Camera | Lagrange point | 2015+ |
| (E)TM   | (Enhanced) Thematic Mapper | Sun-synchronous | 1982+ |
| GOES Imager | Geostationary Operational Environmental Satellite Imager | Geostationary | 1978–2018 |
| GOCI    | Geostationary Ocean Color Imager | Geostationary | 2010+ |
| GLI     | Global Imager | Sun-synchronous | 2002–2003 |
| MERIS   | Medium Resolution Imaging Spectrometer | Sun-synchronous | 2002–2012 |
| MODIS   | MODerate resolution Imaging Spectrometer | Sun-synchronous | 2000+ |
| MSS     | Multispectral Scanner System | Sun-synchronous | 1972–2013 |
| OLCI    | Ocean and Land Color Instrument | Sun-synchronous | 2016+ |
| OLI     | Operational Land Imager | Sun-synchronous | 2013+ |
| SeaWiFS| Sea-viewing Wide Field-of-view Sensor | Sun-synchronous | 1997–2010 |
| SEVIRI  | Spinning Enhanced Visible and InfraRed Imager | Geostationary | 2004+ |
| VIIRS   | Visible Infrared Imaging Radiometer Suite | Sun-synchronous | 2012+ |
| VIRS    | Visible and Infrared Scanner | Precessing | 1997–2015 |
| (A)ATSR| (Advanced) Along-Track Scanning Radiometer | Sun-synchronous | 1991–2012 |
| CHRIS   | Compact High Resolution Imaging Spectrometer | Sun-synchronous | 2001+ |
| MISR    | Multispectral Imaging Spectrometer | Sun-synchronous | 2000+ |
| POLDER  | POLarization and Directionality of the Earth’s Reflectances | Sun-synchronous | 1996–1997; 2002; 2004–2013 |
| SGLI    | Second-generation Global Imager | Sun-synchronous | 2017+ |
| SLSTR   | Sea and Land Surface Temperature Radiometer | Sun-synchronous | 2016+ |
| AIRS    | Atmospheric Infra-Red Sounder | Sun-synchronous | 2002+ |
| GOME    | Global Ozone Monitoring Instrument | Sun-synchronous | 1995–2011 |
| IASI    | Infrared Atmospheric Sounding Interferometer | Sun-synchronous | 2006+ |
| OMI     | Ozone Monitoring Instrument | Sun-synchronous | 2004+ |
| OMPS NM | Ozone Mapping Profiler Suite Nadir Mapper | Sun-synchronous | 2012+ |
| SCIAMACHY| SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY | Sun-synchronous | 2002–2012 |
| TOMS    | Total Ozone Mapping Spectrometer | Sun-synchronous | 1978–1994; 1996–2005 |
| TROPOMI | TROPospheric Monitoring Instrument | Sun-synchronous | 2017+ |
Table 2. As Table 1, except for satellite instruments which have been used for aerosol extinction profiling.

| Acronym | Instrument full name                                      | Orbit(s)         | Operation period(s) |
|---------|----------------------------------------------------------|------------------|---------------------|
| Lidar   |                                                          |                  |                     |
| ALADIN  | Atmospheric LAser Doppler INstrument                     | Sun-synchronous   | 2018+               |
| CALIOP  | Cloud–Aerosol LIidar with Orthogonal Polarization        | Sun-synchronous   | 2006+               |
| CATS    | Cloud–Aerosol Transport System                           | Precessing       | 2015–2017           |
| GLAS    | Geoscience Laser Altimeter System                        | Polar (varied)   | 2003–2010           |
| LITE    | Lidar In-space Technology Experiment                     | Space shuttle    | 1994                |
|         | Limb or occultation profiler                             |                  |                     |
| GOMOS   | Global Ozone Monitoring by Occultation of Stars          | Sun-synchronous   | 2002–2012           |
| MIPAS   | Michelson Interferometer for Passive Atmospheric Sounding| Sun-synchronous   | 2002–2012           |
| OMPS LP | Ozone Mapping Profiler Suite Limb Profiler              | Sun-synchronous   | 2012+               |
| OSIRIS  | Optical Spectrograph and InfraRed Imaging System         | Sun-synchronous   | 2001+               |
| SAGE    | Stratospheric Aerosol and Gas Experiment                 | Precessing       | 1979–1982; 1984+    |
| SAM     | Stratospheric Aerosol Measurement                        | Precessing       | 1975; 1979–1993     |

2 Uncertainty estimates in current satellite aerosol data sets

2.1 Terminology

The International Standards Organization document often known as the GUM (Guide to Uncertainty in Measurement) provides standardized terminology for discussing uncertainties (Working Group 1, 2008). In the interests of standardization and in line with other treatments of uncertainty and error in remote sensing (e.g. Rodgers, 2000; Povey and Grainger, 2015; Loew et al., 2017; Merchant et al., 2017; Mittaz et al., 2019; von Clarmann et al., 2019), the GUM terminology is also adopted here. Terms are often used inconsistently in writing or informal conversation (in particular “error” and “uncertainty”), so to assist the reader, definitions of relevant terms are as follows (and see previously cited references).

- A **measurand** is a quantity to be determined (measured), in the case of this study the AOD.

- A **measurement** is the application of a technique to quantify the measurand, in this case the application of L2 retrieval algorithms to L1 satellite observations.

- The **measured value** is the output of the measurement technique, i.e. here the result of the L2 retrieval algorithm, often referred to as the “retrieved AOD”.

- The **uncertainty** is in the general sense an expression of the dispersion of the measurand. For most of the data sets discussed in this study it is presented as a 1 standard deviation (1σ) confidence interval around the retrieved value (which is defined as the standard uncertainty by the GUM). The true value of the measurand (AOD) is expected to lie within this confidence interval ∼ 68.4 % of the time (corresponding to 1 standard deviation, colloquially 1σ), following Gaussian statistics.
Table 3. Acronyms for some aerosol retrieval algorithms, data records, and/or institution names applied to one or more satellite instruments from Tables 1 and 2.

| Acronym | Algorithm full name |
|---------|---------------------|
| AAC     | Aerosols Above Clouds |
| ADV     | (A)ATSR Dual View |
| AerGOM  | Aerosol profile retrieval prototype for GOMOS |
| ASV     | (A)ATSR Single View |
| BAR     | Bayesian Aerosol Retrieval |
| CISAR   | Combined Inversion of Surface and AeRosol |
| DB      | Deep Blue |
| DT      | Dark Target |
| EDR     | Environmental Data Record |
| ESA     | European Space Agency |
| GACP    | Global Aerosol Climatology Project |
| GRASP   | Generalized Retrieval of Aerosol and Surface Properties |
| IMARS   | Infrared Mineral Aerosol Retrieval Scheme |
| JAXA    | Japan Aerospace eXploration Agency |
| LDA     | Land Daily Aerosol |
| LMD     | Laboratoire de Météorologie Dynamique |
| MAIAC   | Multi-Angle Implementation of Atmospheric Correction |
| MAPIR   | Mineral Aerosol Profiling from Infrared Radiances |
| MODACA  | MODIS Above-Cloud Aerosol |
| NOAA    | National Oceanic and Atmospheric Administration |
| OMACA   | OMI Above-Cloud Aerosols |
| OMAERO  | OMI Multi-wavelength AERosol product |
| OMAERUV | OMI AERosol UV product |
| ORAC    | Optimal Retrieval of Aerosols and Clouds |
| PMAp    | Polar Multi-sensor Aerosol product |
| SOAR    | Satellite Ocean Aerosol Retrieval |
| SU      | Swansea University |
| SYNAER  | SYNergetic AErosol Retrieval |
| ULB     | Université Libre de Bruxelles |
| xBAER   | eXtensible Bremen AErosol Retrieval |

The error is the difference between the measured and true values of the measurand, i.e. here the difference between true and retrieved AOD. Following the GUM convention, a positive error means that the measured value minus the true value is positive (and vice versa).

The error can only be known when the true value of the measurand is also known, which is rare. This is the province of validation exercises: Loew et al. (2017) note that in the remote sensing community (and adopted here), validation refers to a quality assessment of a data set, which is a different definition from that of the metrology community. While Loew et al. (2017) omit mention of aerosols, the points discussed there are applicable to aerosol remote sensing as well. They also note that some authors (e.g. Rodgers, 2000) have adopted a stricter definition of validation to also explicitly include the question of whether the theoretical characterization and obtained properties of the data are consistent; the aforementioned “validating the validation” framework developed in the present study is one component of this.

For validation exercises AERONET AOD data are often taken as a reference truth because the uncertainty on AERONET AOD data (around 0.01 in the mid-visible; Eck et al., 1999) is generally much smaller than that of satellite retrievals. This enables the diagnosis of retrieval errors at the times and locations of matchups with AERONET (or similar reference data), which are often generalized to infer the likely error characteristics of retrievals under various aerosol, surface, and geometric conditions. The implicit assumption is that such a generalization is possible, but it is important to bear in mind that validation data are spatiotemporally sparse and may underrepresent or omit certain factors relative to the real world (Virtanen et al., 2018).

In contrast to error, the uncertainty can be estimated for each individual measured value (retrieval). The term “expected error” (EE) is often used in the aerosol remote sensing literature (e.g. Remer et al., 2005; Kahn et al., 2010; Sayer et al., 2013) to define these prognostic and diagnostic estimates of the magnitude of the uncertainty, highlighting (viz. “expected”) the fact that it is a statistical quantity; in hindsight the term “estimated uncertainty” might have been less confusing. The uncertainty is a statement about the level of confidence (expected magnitude of the error), while the actual error is a realization drawn from the uncertainty distribution. By analogy, rolling a single unbiased die has a mean value (expectation) of 3.5, although this result is impossible to achieve on a single roll (which can take only integer values from 1 to 6). The various techniques which have been applied to provide prognostic estimates for AOD are discussed in Sect. 2.2, while Sect. 2.3 discusses those data sets for which only diagnostic uncertainty estimates are available. A difficulty, which this study aims to tackle, is how to tell whether these uncertainty estimates are quantitatively useful and reliable. Six “conditions of adequacy” have been proposed by von Clarmann et al. (2019) for temperature and trace gas profile uncertainty estimates, namely that they are the following: (1) intercomparable between instruments and/or error estimation schemes; (2) independent of vertical retrieval grid (often less relevant for aerosols); (3) usable to the reader not familiar with instrument or retrieval technical details; (4) documented and traceable; (5) validatable (part of the focus of this study); and (6) can be summarized without excessive additional data volume overhead. These are desirable from the point of view of aerosols as well.

2.2 Techniques for prognostic uncertainty estimates

Examples of existing prognostic uncertainty estimates for AOD or aerosol extinction data sets are given in Table 4. These fall into two broad camps: formal error propagation techniques accounting for individual terms thought to be relevant to the overall error budget and more empirical meth-
Table 4. AOD and extinction data sets providing prognostic uncertainty estimates as well as associated key references for uncertainty estimate calculation. Where applicable, algorithm names are given first with instrument names in parentheses. See Tables 1, 2, and 3 for acronyms.

| Data set | Key references for uncertainty calculation | Note |
|----------|------------------------------------------|------|
| ADV/ASV (ATSR2, AATSR) | Kolmonen et al. (2016); Kolmonen and Sogacheva (2018) | Jacobians at retrieval solution |
| AerGOM (GOMOS) | Vanhellemont et al. (2016) | Maximum likelihood with smoothness constraints |
| BAR (MODIS) | Lipponen et al. (2018) | Maximum likelihood, retrieves whole granule at once |
| CALIPSO | Young et al. (2013, 2018) | Propagation of contributions through lidar equation |
| CATS | Young et al. (2013) | Propagation of contributions through lidar equation |
| CISAR (CHRIS, SEVIRI) | Govaerts and Luffarelli (2018); Luffarelli and Govaerts (2019) | Optimal estimation with smoothness constraints |
| DB AAC (MODIS, SeaWiFS, VIIRS) | Sayer et al. (2016); Sayer et al. (2019b) | Maximum likelihood |
| DB land (MODIS) | Sayer et al. (2013) | Empirical expression from AERONET validation results |
| GOCI | Choi et al. (2018) | Empirical expression from AERONET validation results |
| GRASP (MERIS, POLDER) | Dubovik et al. (2011) | Maximum likelihood with smoothness constraints |
| IMARS (IASI) | Klüser et al. (2011) | Propagated measurement and forward model errors |
| JAXA (AHI) | Yoshida et al. (2018) | Optimal estimation |
| LDA (SEVIRI) | Govaerts et al. (2010) | Optimal estimation |
| LMD (AIRS, IASI) | Pierangelo et al. (2004); Capelle et al. (2014) | Parametric from sensitivity studies and validation |
| MAIAC (MODIS) | Lyapustin et al. (2018) | Propagated from uncertainty on surface reflectance |
| MAPIR (IASI) | Vandenbussche et al. (2013) | Optimal estimation |
| MIPAS | Günther et al. (2018) | Maximum likelihood with smoothness constraints |
| MISR dark water | Witek et al. (2018b) | Width of cost function distribution vs. AOD |
| MISR heterogeneous land | Martonchik et al. (1998, 2009) | Standard deviation of well-fitting aerosol models |
| MODACA (MODIS) | Meyer et al. (2015) | Maximum likelihood |
| NOAA EDR (VIIRS) | Huang et al. (2016) | Empirical expression from AERONET validation results |
| OMPS | Loughman et al. (2018) | Confidence envelope based on aerosol signal strength |
| ORAC (ATSR2, AATSR, SEVIRI) | Thomas et al. (2009, 2010, 2017) | Optimal estimation |
| PMAp | EUMETSAT (2015) | Standard deviation of aerosol models |
| OSIRIS | Bourassa et al. (2007); Rieger et al. (2019) | Optimal estimation |
| SAGE | McCormick et al. (2002) | Propagated measurement plus interfering species error |
| SU (ATSR2, AATSR, MERIS+AATSR) | North et al. (2010, 2017); Bevan et al. (2012) | Second derivative of error function |
| ULB (IASI) | Clarisse et al. (2013) | Propagated measurement and forward model errors |
| VAC MODIS, SeaWIFS, VIIRS | Csisár (CHRIS, SEVIRI) | Confidence envelopes through lidar equation |
| VCI | Csisár (CHRIS, SEVIRI) | Confidence envelopes through lidar equation |
| VIM | EUMETSAT (2015) | Standard deviation of aerosol models |
ods. The term “error budget” (not defined in the GUM, but in common colloquial use) here refers to, dependent on context, the overall collection of contributions to input or output uncertainty. Strictly, one might refer instead to “uncertainty budget” and “uncertainty propagation”, but for reader ease, the commonly used terms are adopted here.

2.2.1 Formal error propagation

The formal methods which have been applied to date are in general Bayesian approaches, which can be expressed in the formalism of Rodgers (2000), and are often referred to as optimal estimation (OE). OE approaches provide the maximum a posteriori (MAP) solution to the retrieval problem: maximization of the conditional probability \( P(x|y, x_a) \) of the retrieved state vector \( x \), where \( y \) and \( x_a \) represent the satellite measurements and any a priori information on \( x \), respectively. The MAP solution is achieved by minimization of a cost function \( J \), and the formalism allows for the calculation of various contributions to the total uncertainty \( \hat{S} \) on the retrieved state. OE accounts for uncertainty on the satellite measurements, retrieval forward model (e.g. atmospheric and surface structure assumptions, ancillary data), a priori information, and smoothness constraints (on e.g. spatial, temporal, or spectral variation of parameters). While notation differs between authors (see also Rodgers, 2000; Dubovik et al., 2011; Govaerts and Luffarelli, 2018), a general form of the cost function \( J \) can be written

\[
J(x) = (F(x) - y)^T S_y^{-1} (F(x) - y) + (x_a - x)^T S_a^{-1} (x_a - x) + x^T H^T S_y^{-1} H x + \ldots, \tag{1}
\]

where \( S_y \) and \( S_a \) are covariance matrices; \( S_y \) describes the measurement and forward model uncertainty, \( S_a \) describes the a priori uncertainty, and \( F(x) \) is the forward-modelled measurements. The third term represents a generic smoothness constraint on the state vector (which might be spatial, temporal, spectral, or otherwise), where \( H \) is a block diagonal matrix and \( S_y \) its associated uncertainty; the ellipses in Eq. (1) indicate the potential for the expansion of \( J \) to include additional smoothness terms. These smoothness constraints were first introduced in the context of aerosol remote sensing by Dubovik and King (2000) for AERONET sky-scan inversions. In recent years they have become more widespread in satellite aerosol remote sensing as more capable sensors (e.g. POLDER) and/or algorithms with increased (spatiotemporal, spectral, or directional) dimensionality of measured or retrieved quantities (Dubovik et al., 2011; Govaerts and Luffarelli, 2018; Shi et al., 2019) have been developed. Candidate algorithms for aerosol retrieval from information-rich future sensors also tend to use smoothness constraints (e.g. Xu et al., 2019). All these covariance matrices are assumed to be Gaussian, which may not always be true in practice.

Note that here \( S_y \) represents the total of measurement uncertainty, forward model uncertainty (due to approximations made in the radiative transfer), and the contribution of uncertainties in forward model parameters to the simulated signal at the top of the atmosphere (TOA). These model parameters are factors which affect the TOA signal but typically insignificantly enough to be retrieved. For example, many AOD retrieval algorithms ingest meteorological reanalysis to correct for the impact of absorbing trace gases (such as H2O) on the satellite signal at TOA (Pataidia et al., 2018) and to provide wind speed to calculate glint and whitecap contributions to sea surface reflectances (Sayer et al., 2010a). Sometimes these are represented in \( J \) instead by a “model parameter error” matrix denoted \( S_b \) and similar squared deviations, although mathematically since the terms in Eq. (1) are additive the two formalisms are equivalent if the model parameter uncertainty is transformed into measurement space and included in \( S_y \) (as is typically the case).

As \( S_y \) and \( S_a \) (etc.) are square matrices, correlations between wavelengths or parameters can (and, where practical, should) be accounted for. These terms often affect several satellite bands such that an error in e.g. reanalysis data ingested as part of an AOD retrieval would manifest in a correlated way between these bands. However, due to the difficulty in estimating these off-diagonal elements, in practice they are frequently neglected and the covariance matrices are often assumed to be diagonal (which does not, however, mean that \( \hat{S} \) is diagonal). Dependent on the magnitude and sign of these correlations, their neglect can lead to overestimates or underestimates in the level of confidence in the solution. When the cost function has been minimized, the uncertainty \( \hat{S} \) on the retrieved state is given by

\[
\hat{S} = \left(K^T S_y^{-1} K + S_a^{-1} + H^T S_y^{-1} H \right)^{-1}, \tag{2}
\]

where \( K \), known as the weighting function or Jacobian matrix, is the sensitivity of the forward model to the state vector \( \frac{\partial F(x)}{\partial x} \), typically calculated numerically. The \( 1\sigma \) uncertainty on the retrieved AOD is then the square root of the relevant element on the diagonal of \( \hat{S} \) (dependent on the contents of the state vector). Many current approaches in Table 4 omit a priori and/or smoothness constraints, in which case the corresponding terms in Eqs. (1) and (2) vanish. Only BAR and CISAR include both a priori and smoothness constraints. AerGOM, GRASP, and the MIPAS stratospheric aerosol data set use smoothness constraints without a priori on the aerosol state. Others (LDA, JAXA AHI, MAPIR, ORAC) use a priori but no smoothness constraints. Smoothness constraints are attractive for algorithms such as the GRASP application to POLDER, which includes the retrieval of binned aerosol size distribution and spectral refractive index (which are expected to be smooth for physical reasons), as well as those (e.g. BAR, CISAR, GRASP) moving beyond the independent pixel approximation to take advantage of the fact that certain atmospheric and/or surface parameters can be ex-
pected to be spatially and/or temporally smooth on relevant scales.

These smoothness and a priori constraints provide a regularization mechanism to suppress “noise-like” variations in the retrieved parameters when they are not well-constrained by the measurements alone, although there is a danger in that overly strong constraints can suppress real variability. As a result, a priori constraints on AOD itself are often intentionally weak compared to those on other retrieved parameters. Strictly, the MAP is a maximum likelihood estimate (MLE) only if the retrieval does not use a priori information, although it is often referred to as an MLE regardless (see Sect. 4.1 of Rodgers, 2000, for more discussion on this distinction). This distinction is made in the descriptions in Table 4.

The rest of the error propagation methods in Table 4, whether formulated as OE or not, are essentially propagating only measurement (and sometimes forward model) uncertainty through to the retrieval solution through Jacobians. MAIAc is a special case here because, rather than using the measurement uncertainty directly, it propagates the uncertainty of surface reflectance in the 470 nm band, which is thought to be the leading contribution to the total error budget (Lyapustin et al., 2018). It is important to note that the cost function and uncertainty estimate calculations in Eq. (2) are conditional on several factors.

1. The forward model must be appropriate to the problem at hand and capable of providing unbiased estimates of the observations. Typically if the forward model is fundamentally incorrect, and/or any a priori constraints strongly inappropriate, the retrieval will frequently not converge to a solution or have unexpectedly large $J$. For this reason, high cost values are often used in post-processing to remove problematic pixels (e.g. undetected cloud or snow) or candidate aerosol optical models from the provided data sets (Martonchik et al., 1998; Thomas et al., 2010).

2. The covariance matrices $S_y$, $S_a$, and $S_i$ (on measurements, a priori, and smoothness) must be appropriate; if systematically too large or small, uncertainty estimates will likewise be too large or small. These can be tested, to an extent, by examining the distributions of residuals (on measurements and a priori) and the cost function and comparing to theoretical expectations (e.g. Sayer et al., 2010a, 2012c).

3. The forward model must be approximately linear with Gaussian errors near the solution. This assumption sometimes breaks down if the measurements are uninformative on a parameter and a priori constraints are weak or absent, and the resulting state uncertainty estimates will be invalid. This can be tested (Thomas et al., 2009; Sayer et al., 2016) by performing retrievals using simulated data, perturbing their inputs according to their assumed uncertainties, and assessing whether the dispersion in the results is consistent with the retrieval uncertainty estimates.

4. The retrieval must have converged to the neighbourhood of the correct solution (i.e. near the global, not a local, minimum of the cost function), which can be a problem if there are degenerate solutions. In practice algorithms try to use reasonable a priori constraints, first guesses, and make a careful selection of which quantities to retrieve vs. which to assume (e.g. Thomas et al., 2009; Dubovik et al., 2011). Note that the iterative method of convergence to the solution is not important in itself.

A detailed further discussion on these conditions from the perspective of temperature and trace gas retrievals, which share some similar conceptual challenges to aerosol remote sensing, is provided by von Clarmann et al. (2019).

2.2.2 Other approaches

A particular challenge for the formal error propagation techniques is the second point above: how to quantify the individual contributions to the error budget necessary to calculate the above covariance matrices? This difficulty has motivated some of the empirical approaches in Table 4.

Sayer et al. (2013) used the results of validation analyses against AERONET to construct an empirical relationship (discussed in more detail later) expressing the uncertainty in MODIS DB AOD retrievals as a function of various factors. This basic approach was later adopted for other data sets, including GOCI and NOAA VIIRS EDR aerosol retrievals (Huang et al., 2016; Choi et al., 2018). This has some similarity to diagnostic EE envelopes, although importantly these prognostic estimates are framed in terms of retrieved rather than reference AOD. An advantage of this method is that, if AERONET can be taken as a truth and collocation-related uncertainty is small (Virtanen et al., 2018), it empirically accounts for the important contributions to the overall error budget without having to know their individual magnitudes. However, there are some disadvantages: if validation data are sparse or do not cover a representative range of conditions, there is a danger of overfitting the expression, and for an ongoing data set there is no guarantee that past performance is indicative of future results as sensors age and the world changes. For a quantity without available representative validation data, the method cannot be performed. Further, programmatically, it requires processing data twice: once to perform the retrievals and do the validation analysis to derive the expression and a second time to add the resulting uncertainty estimates into the data files. The LMD IASI retrieval has a similar parametric approach (Capelle et al., 2014), although as validation data are sparse, the parametrization draws on the results from retrieval simulations as well.

The MISR algorithms use different approaches. Both the land and water AOD retrieval algorithms perform retrieval
using each of 74 distinct aerosol optical models (known as "mixtures") and calculate a cost function for each. In earlier algorithm versions (Martonchik et al., 1998) uncertainty was taken as the standard deviation of AOD retrieval from mixtures which fit with a cost below some threshold. This is equivalent to assuming that aerosol optical models are the dominant source of uncertainty in the retrieval and that the 74 mixtures provide a representative sampling of microphysical and optical properties.

This approach was refined (for retrievals over water pixels) by Witek et al. (2018b) by considering the variation of retrieval cost with AOD for each model and transforming this to give a probability distribution of AOD, with the uncertainty taken as the width of this distribution. A similar approach has been proposed for the OMAERO retrieval by Kauppi et al. (2017), although it has not yet been implemented on a large scale. It has conceptual similarities with the propagation of measurement error in Eq. (2), except calculating across the whole range of AOD state space rather than an envelope around the solution and summing the results from multiple distinct retrievals (corresponding to the aerosol mixtures). These methods are, however, reliant on the set of available optical models being sufficient.

### 2.3 Examples of diagnostic uncertainty estimates

Available AOD data sets which do not currently provide prognostic uncertainty estimates are listed in Table 5. In these cases, algorithm papers typically summarize the results of sensitivity analyses to provide a rationale for choices made in algorithm development and to provide a summary of expected performance. Sensitivity analyses often include similar aspects to those employed in error propagation approaches: namely, characterization of the expected effects of uncertainties in sensor calibration and forward model limitations (e.g. assumed aerosol optical models, surface reflectance) on the retrieval solution, singly or jointly. In most cases these are provided for a subset of geometries and atmosphere–surface conditions. Compared to formal error propagation, this has the advantage of being easier to communicate to a reader concerned about a particular assumption (provided the results of the sensitivity analysis are representative), but on the other hand the summary results are specific to only the simulations performed, and real-world uncertainties may be more complicated, particularly when multiple retrieval assumptions are confounded.

Sensitivity analyses are often complemented by dedicated validation papers which summarize the results of dedicated validation papers which summarize the results of comparisons against AERONET, MAN, or other networks (e.g. Remer et al., 2005; Kahn et al., 2010); aerosol remote sensing is fortunate compared to other disciplines in that high-quality AOD validation data are fairly readily available. It is common for the results to be summarized in terms of EE envelopes or similar metrics; these envelopes are sometimes adjusted if pre-launch expectations prove too optimistic or
pessimistic (e.g. Levy et al., 2013). Diagnostic and prognostic uncertainty estimates should not be regarded as exclusionary; diagnostic analysis is useful to guide algorithm refinement and assess assumptions, and many data products which provide prognostic uncertainties also show the results of diagnostic validation activities. However, extending the data sets in Table 5 to also provide prognostic estimates would improve their specificity and utility for applications like DA.

2.4 Systematic and random contributions to uncertainty

Both the diagnostic and prognostic techniques typically (implicitly or explicitly) make the assumption that the sensor and retrieval algorithm are unbiased and that the resulting uncertainty estimates are unbiased and symmetric. However, it is well-known that many of the key factors governing retrieval errors are globally (e.g. sensor calibration, Lyapustin et al., 2014) or seasonally–regionally (e.g. aerosol optical model, surface reflection, cloud contamination, Eck et al., 2013; Zhao et al., 2013; Gupta et al., 2016) systematic and that true random error (i.e. propagated noise) is often small. While these systematic factors may partially cancel each other out over large ensembles of data (drawn from e.g. different regions, seasons, or geometries), this is not a given.

Uncertainty propagation approaches such as OE can in principle account for systematic uncertainty sources, as they (and any spectral or parameter correlations) can be included in the required covariance matrices. This can produce estimates of total uncertainty which are reasonable for an individual retrieval, but the true (large-scale) error distributions would then not be symmetric, lessening their value. Likewise, systematically biased priors can lead to systematically biased retrievals. As a result, it would be desirable to remove systematic contributions to the retrieval system uncertainty as far as possible. In practice this is often done through validation exercises, whereby diagnostic comparisons can provide clues as to the source of biases, which are then (hopefully) lessened in the next version of the algorithm. Distributions of the residuals of predicted measurements at the retrieval solution can also be indicative of calibration and forward model biases at the wavelength in question.

A possible solution to this is to perform a vicarious calibration, calculating a correction factor to the sensor gain as a function of time and band by matching observed and modelled reflectances at sites where atmospheric and surface conditions are thought to be well-known (e.g. thick anvil clouds, Sun glint, and AERONET sites). The derived correction factor then accounts for the systematic uncertainty on calibration and the radiative transfer forward model, although if this latter term is non-negligible then the vicariously calibrated gains will still be systematically biased (albeit less so for the application at hand). This has the advantage of transforming the calibration uncertainty from a systematic to a more random error source at the expense of creating dependence

on the calibration source and radiative transfer model. There is therefore a danger in creating a circular dependence between the vicarious calibration and validation sources as it can hinder understanding of the physics behind observed biases. Further, this has the side effect of potentially increasing the level of systematic error in other quantities or in conditions significantly different from those found at the vicarious calibration location if the forward model contribution to systematic uncertainty is significant (Kahn et al., 2016). Vicarious calibration is common within the ocean colour community (Franz et al., 2007), in which retrieval algorithms are in some cases more empirical and amenable to tuning than physically-driven aerosol retrieval algorithms. It has also been used for on-orbit calibration of instruments lacking on-board capabilities to track absolute calibration and degradation (e.g. Heidinger et al., 2010).

3 Statistical framework to evaluate pixel-level AOD uncertainty estimates

3.1 Background and methodology

The notation adopted herein is as follows. The AOD is denoted $\tau$; unless specified otherwise, references to AOD indicate that at 550 nm. The reference (here AERONET) AOD is $\tau_A$ and satellite-retrieved AOD is $\tau_S$. The $1\sigma$ estimated uncertainties on these are denoted $\epsilon_A$ and $\epsilon_S$, respectively. If the reference AOD is assumed to be the truth, then the error $\Delta_S$ on the satellite AOD is given by $\Delta_S = \tau_S - \tau_A$.

Figure 1 provides a simulation experiment to illustrate the relationship between AOD, uncertainty, and error distributions. Panel (a) is a histogram of AOD generated (1 000 000 points) assuming a lognormal distribution with geometric mean 0.2 and geometric standard deviation 0.35, which is a typical shape for many locations in North America and Europe (O’Neill et al., 2000). Panel (b) shows two distributions: in black is the distribution of the expected AOD uncertainty magnitude (often, as discussed before, called expected error or EE), assuming error characteristics of the MODIS DT land retrieval, $\epsilon_S = \pm (0.05 + 0.15 \tau)$ (Levy et al., 2013). This is obtained simply by multiplying the histogram in Fig. 1a by the magnitude of uncertainty $|\epsilon_S|$. The red line, in contrast, is the distribution of actual absolute retrieval errors (i.e. $|\tau_S - \tau_A|$), which would be expected to be seen in a validation exercise against AERONET if the expression for $\epsilon_S$ holds true. This red line is obtained by taking draws from the AOD distribution and then, for each, generating a normally distributed random number with mean 0 and standard deviation $\epsilon_S$ to provide the retrieval error (note that the absolute value of this retrieval error is shown in Fig. 1b).

An important nuance which bears repeating is that the distributions of estimated uncertainty and actual error in Fig. 1 are quite different in shape. This is because the estimated uncertainty distribution is one of the expectations of $\epsilon_S$ (given
the AOD distribution), while the distribution of errors is one of the realizations of (draws from) $\epsilon_S$. Recall again the distinction between the expectation of rolling an unbiased die (i.e. a result of 3.5) and the actual realization (result) of rolling a die (1, 2, 3, 4, 5, or 6). The latter distribution is broader. This illustrates why comparing errors and uncertainties on a 1 : 1 basis, or comparing distribution magnitudes, is not expected to yield agreement, and an evaluation of consistency requires a statistical approach. Figure 2 shows this more directly: there is little correspondence between the two on an individual basis.

When comparing satellite and reference data, the total expected discrepancy (ED) between the two for a single matchup, denoted $\epsilon_T$, should account for uncertainties on both the satellite and reference (here AERONET) data,

$$\epsilon_T = \sqrt{\epsilon_S^2 + \epsilon_A^2}, \tag{3}$$

adding in quadrature under the assumption that the uncertainties on satellite and AERONET AOD are independent of one another. One can then define a normalized error $\Delta_N$ as the ratio of the actual error to the ED, i.e.

$$\Delta_N = \frac{\Delta_S}{\epsilon_T} = \frac{\tau_S - \tau_A}{\sqrt{\epsilon_S^2 + \epsilon_A^2}} \tag{4}$$

In the ideal case $\epsilon_A \ll \epsilon_S$, in which case the shape of $\Delta_N$ is dominated by the uncertainty and errors on the satellite-retrieved AOD. If the uncertainties on satellite and reference AOD have been calculated appropriately and the sample is sufficiently large, then the normalized error $\Delta_N$ should approximate a Gaussian distribution with mean 0 and variance 1. Thus, the distribution of $\Delta_N$ can be checked in several ways against expected shapes for Gaussian distributions, for example the probability distribution function (PDF) and cumulative distribution function (CDF) as shown in Fig. 3.

The above distribution analyses are informative on the overall magnitude of retrieval errors compared to expectations (as well as, in the case of the PDF analysis, whether there is an overall bias on the retrieved AOD). However, alone they say little about the skill in assessing variations in uncertainty across the population. Taking things a step further, the data can be stratified in terms of ED and a quantile analysis performed to assess consistency with expectations. This is equivalent to taking a single location along the x axis in Fig. 2 and assessing the distribution of retrieval errors found for the points from that histogram. These, too, should follow Gaussian statistics.

An example of this is shown in Fig. 4. The data are divided by expected discrepancy $\epsilon_T$ into 10 equally populated bins, and within each bin the 38th, 68th, and 95th percentiles (i.e. approximate 0.5$\sigma$, 1$\sigma$, 2$\sigma$ points, following Gaussian statistics) of absolute retrieval error are plotted. If the uncertainties are appropriate, these should lie along the 0.5 : 1, 1 : 1, and 2 : 1 lines. This analysis provides a way of checking the validity of the uncertainty estimates across the spectrum from low to high estimated uncertainties as opposed to population-average behaviour (i.e. do the distributions of retrieval error change in the expected way as the estimated uncertainty varies?). The 68th percentile is of the most direct
Figure 3. (a) PDF and (b) CDF of normalized error distributions drawn from the numerical simulations in Fig. 1; theoretical (grey shading) and simulation (red) results lie on top of one another. Note that the CDF is of absolute normalized error. Dashed lines indicate various well-known percentage points of Gaussian distributions.

Figure 4. Expected AOD discrepancy against percentiles of absolute AOD retrieval error. Symbols indicate binned results from the numerical simulation; within each bin, paler to darker tones indicate the 38th, 68th, and 95th percentiles (approximate 0.5σ, 1σ, 2σ points) of absolute retrieval error. Dashed lines (0.5:1, 1:1, 2:1, respectively) show theoretical values for the percentiles of the same colour.

The binned analysis is similar to the assessment of forecast calibration in meteorology (Dawid, 1982). Note in a forecast sense that the term calibration refers to a comparison of forecast vs. observed frequencies or magnitudes, distinct from the common meaning of calibration to refer to radiometric accuracy in remote sensing. By further analogy to the forecast community (compare to the expressions in Murphy, 1988), a calibration skill score $s_{\text{cal}}$ can be defined,

$$s_{\text{cal}} = 1 - \frac{\sum_{b=1}^{B} \left( \epsilon_{T,b} - |\Delta_{S,b}| \right)^2}{\sum_{b=1}^{B} \left( |\Delta_{S}| - |\Delta_{S,b}| \right)^2},$$

where $|\Delta_{S,b}|$ is the 1σ absolute retrieval error in bin $b$ (Fig. 4) over $B$ bins total. This compares the observed squared discrepancy from the 1:1 line in Fig. 4 with that which would be obtained if a data user assumed that the retrieval uncertainty was equal to the mean absolute retrieval error ($|\Delta_{S}|$) from a validation exercise at that location, which is what might be done in the absence of pixel-level uncertainty estimates. This skill score is computed using binned values rather than individual matchups due to the previously discussed nature of the relationship between uncertainty and error (Figs. 1, 2). The highest possible score is 1, and a score of 0 indicates that the uncertainty estimates do not have greater skill than simply assuming the average retrieval error. If the magnitudes of $\epsilon_T$ are in error then it is possible for $s_{\text{cal}}$ to take unbounded negative values, in which case the uncertainties are said to be poorly calibrated (Dawid, 1982). This is quite a difficult test for a data set as a positive skill score requires that both the magnitudes of the uncertainty and the variations in both uncertainty and error must be accurate. This may be particularly difficult if the error does not vary much at a given location. As a result $s_{\text{cal}}$ should not be used as a single metric in isolation but rather examined in a broader context.

Figures 3 and 4 provide the basis for the framework proposed in this study. An earlier version of this method was designed during the development and assessment of prognostic uncertainty estimates for MODIS DB retrievals by Sayer et al. (2013). It has been further advanced through discussions at annual AeroSat meetings. These ideas have been further practically applied to NOAA VIIRS AOD data by Huang et al. (2016), to GOCI data by Choi et al. (2018), to retrievals of absorbing aerosols above clouds against airborne measurements by Sayer et al. (2019b), and to the latest MISR product over ocean by Witek et al. (2019). The idea of looking at normalized retrieval error distributions was also explored for AOD by Popp et al. (2016) and Kinne et al. (2017) when evaluating ESA Climate Change Initiative (CCI) aerosol products and in a more general sense (with cloud-top height as an example) by Merchant et al. (2017). Indeed, the method is not restricted to AOD, although AOD has the advantage of

---

**Atmos. Meas. Tech., 13, 373–404, 2020**

www.atmos-meas-tech.net/13/373/2020/
comparatively readily available, high-quality reference data in AERONET and other networks.

3.2 Practical application to satellite data products

3.2.1 AERONET data used and matchup criteria

Here, the reference AOD $\tau_\lambda$ is provided using level 2.0 (cloud-screened and quality assured) direct-Sun data from the latest AERONET version 3 (Giles et al., 2019). As AERONET Sun photometers do not measure at 550 nm, the AOD is interpolated using a second-order polynomial fit to determine the coefficients $a_0$, $a_1$, and $a_2$ for each measurement,

$$\log(\tau_\lambda) = a_0 + a_1 \log(\lambda) + a_2 \log(\lambda)^2,$$

where $\lambda$ is the wavelength. All available (typically four) AOD measurements in the 440–870 nm wavelength range are used in the fit, which is more robust to calibration problems in individual channels than a bispectral approach and accounts for spectral curvature in $\log(\tau_\lambda)$ (Eck et al., 1999; Schuster et al., 2006). The uncertainty on mid-visible AOD is dominated by sensor calibration and is $\sim 0.01$ (Eck et al., 1999). The sampling cadence is typically once per 10 min in cloud-free, daytime conditions but is more frequent at some sites.

Data from a total of 12 AERONET sites, listed in Table 6, are used here to assess the AOD uncertainty estimates in various satellite data sets. This is evenly split to provide six sites to evaluate AOD retrievals from algorithms over land and six over water. Each category is further split; three sites are described as “straightforward”, for which the AOD retrieval problem is comparatively uncomplicated (i.e. likely no significant deviations from retrieval assumptions) and so the uncertainty estimates might be expected to be reasonable, and three sites are “complex”. These complex sites were chosen as they have complicating factors which are not well-captured by existing retrieval forward models and might be expected to lead to breakdowns in the techniques used by the retrieval algorithms to provide uncertainty estimates.

The reasons for identifying a particular site as complex are as follows. Over land, Ilorin (Nigeria) and Kanpur (India) can exhibit complicated mixtures of aerosols with distinct optical properties and vertical structure (Eck et al., 2010; Giles et al., 2012; Fawole et al., 2016). Many AOD retrieval algorithms, in contrast, assume a single aerosol layer of homogeneous optical properties. Pickle Lake (Canada) is in an area studded by lakes of sizes similar to or smaller than satellite pixel size. This might be expected to interfere with data set land masking (which often determines algorithm choice) and surface reflectance modelling in a non-linear way (Carroll et al., 2017). Over water, Cape Verde (on Sal Island, officially the Republic of Cabo Verde) is characterized by frequent episodes of Saharan dust outflow; these particles have complex shapes, which are often approximated in AOD retrieval algorithms by spheres or spheroids. This assumption leads to additional uncertainties in modelling the aerosol phase matrix and absorption cross section, which are larger than for many other aerosol types and may not be accounted for fully in the retrieval error budget (Mishchenko et al., 1997; Kalashnikova et al., 2005). ICIPE Mbita (hereafter Mbita, on the shore of Lake Victoria in Kenya) is similar to Pickle Lake but for water retrievals; i.e. it allows for the sampling of nominal water pixels which may be influenced by partial misflagging of coastlines, 3-D effects from the comparatively bright land, and outflow into the water affecting surface brightness. Finally, Venice (Italy) is in the northern Adriatic Sea, slightly beyond the outflow of the Venetian lagoon, and its water colour is strongly divergent from the Case 1 (brightness tied to chlorophyll $a$ concentration; Morel, 1988) assumption employed by most AOD retrieval algorithms.

This breakdown is inherently subjective as all retrievals involve approximations; the dozen sites chosen are illustrative of different aerosol and surface regimes but not necessarily indicative of global performance. The purpose of this study is to define and demonstrate the framework for evaluating pixel-level uncertainties and provide some recommendations for their provision and improvement. It is hoped that, with growing acceptance of the need to evaluate pixel-level uncertainties, this approach can be applied on a larger scale. The sites were chosen as they are fairly well-understood and have multi-year data sets (data from all available years were considered from the analysis). Note that some of the satellite data sets considered here do not provide data at some sites for various reasons (discussed later).

The matchup protocol is as follows. AERONET data are averaged within $\pm 15$ min of each satellite overpass (providing $\tau_\lambda$) and compared with the closest successful satel-
Table 6. AERONET sites used and their categorization.

| Site                                      | Latitude (° N) | Longitude (° E) | Complexity   |
|-------------------------------------------|----------------|-----------------|--------------|
| For land algorithm evaluation             |                |                 |              |
| Avignon                                   | 43.93          | 4.88            | Straightforward |
| Goddard Space Flight Center (GSFC)        | 38.99          | −76.84          | Straightforward |
| Palencia                                  | 41.99          | −4.52           | Straightforward |
| Ilorin                                     | 8.48           | 4.67            | Complex       |
| Kanpur                                     | 26.51          | 80.23           | Complex       |
| Pickle Lake                                | 51.45          | −90.22          | Complex       |
| For water algorithm evaluation            |                |                 |              |
| Ascension Island                          | −7.98          | −14.41          | Straightforward |
| Midway Island                              | 28.21          | −177.38         | Straightforward |
| University of California Santa Barbara (UCSB) | 34.42        | −119.85         | Straightforward |
| Cape Verde                                 | 16.73          | −22.94          | Complex       |
| International Centre of Insect Physiology and Ecology (ICIPE) Mbita | −0.43 | 34.21 | Complex |
| Venice                                     | 45.31          | 12.51           | Complex       |

The satellite retrieval which has a pixel centre within 10 km of the AERONET site. This provides \( \tau_5 \) and \( \epsilon_s \). Each satellite data set’s recommended quality assurance (QA) filtering criteria are applied as provided in the data products. The AERONET uncertainty, \( \epsilon_A \), is taken as the quadrature sum of the AERONET measurement uncertainty (\( \pm 0.01 \); Eck et al., 1999) and standard deviation of the AERONET measurement uncertainty (\( \pm \epsilon_{AERONET} \)). Dependent on the site sensor, this additional filtering step removes around 10%–60% of potential matchups; Fig. 5 shows an example for MISR over-water retrievals at the Ascension Island site. As a reminder, the focus here is not on validating the AOD but rather validating the AOD uncertainty estimates (vertical lines in the figure).

These matchup criteria are stricter than what is commonly applied for AOD validation (e.g. Ichoku et al., 2002), which typically averages AERONET data within \( \pm 30 \)–60 min and satellite retrievals within \( \sim \pm 25 \) km; the smaller spatiotemporal window and additional filtering criteria decrease the potential (unknown) contribution of collocation uncertainty to \( \epsilon_A \), which increases as the collocation criteria are loosened (Virtanen et al., 2018). The reasoning behind taking the nearest, rather than average, satellite retrieval is similar: averaging would have the potential to decrease the apparent retrieval error, which would make the comparison less useful for evaluating \( \epsilon_s \). Weakening these criteria could increase the data volume for analysis at the expense of increased collocation-related uncertainty, and there is no objective way to determine universal optimal thresholds. However, in the future, site-specific criteria could be guided by analysis of high-resolution (spatiotemporal) model simulations and surface observations.

This work considers satellite AOD products from seven algorithm teams; five of these contain both land and water retrievals (albeit sometimes with different algorithms), while two only cover land retrievals. Only pixels retrieved as land are used for comparison with AERONET data from land sites in Table 6, and vice versa for water sites. These data sets are briefly described below, and the reader is referred to the references cited here and in Tables 4 and 5 for additional information. Note in the discussion that the term “pixel” refers to individual L2 retrievals, sometimes referred to “superpixels” in the literature as they are often coarser than the source L1 data.

### 3.2.2 MODIS data sets

Four of the data sets (three land, one water) are derived from MODIS measurements; there are two MODIS sensors providing data since 2000 and 2002 on the Terra and Aqua satellites, respectively. The sensors have a 2330 km swath width, which is advantageous in providing a large data volume for analysis. Since launch, the MODIS aerosol data products have included AOD from the DT algorithm family, which has separate algorithms for water and vegetated land pixels (Levy et al., 2013). These data sets provide only diagnostic uncertainty estimates of the form \( \epsilon_s = \pm (a + b \tau_A) \); in practice (and here) these are often treated as if they were framed instead in terms of \( \tau_5 \) with the same coefficients \( a \) and \( b \) when a prognostic estimate is needed. For retrievals over land, \( \epsilon_s = \pm (0.05 + 0.15 \tau_A) \), which is consistent with the expected performance of the algorithms at launch (Remer et al., 2005). Over water, the estimate has been revised since launch to \( \epsilon_s = \pm (0.03 + 0.1 \tau_A) \). Limited validation based on Collection 6 data by Levy et al. (2013) suggested that there...
might be an asymmetry to the envelope with the 1σ range over water being from −0.02 − 0.1τA to +0.04 + 0.1τA. This has not yet been corroborated by a global validation of C6 or the latest Collection 6.1 (C6.1), and it is also plausible that calibration updates in C6.1 may have ameliorated some of this bias. As a result the symmetric envelope is used here.

The DB algorithm retrieves AOD only over land and was introduced to fill gaps in DT coverage due to bright surfaces such as deserts (although it has since been expanded to include vegetated land surfaces as well). The latest version is described by Hsu et al. (2019). Prognostic AOD retrieval uncertainties are estimated as described in Sayer et al. (2013),

\[ \epsilon_S = \pm \left( \frac{a + b\tau_S}{\mu_0 + \mu} \right), \]

where \( \mu_0 \) and \( \mu \) are the cosines of solar and view zenith angles, respectively, and \( a \) and \( b \) are coefficients depending on the QA flag value, sensor, and (since C6.1) surface type. The latest values of \( a \) and \( b \) are given by Hsu et al. (2019).

BAR also performs retrievals only over land; it uses the same radiative transfer forward model as DT but reformulates the problem to retrieve the MAP solution of aerosol properties and surface reflectance simultaneously for all vegetated pixels in a single granule (Lipponen et al., 2018). This includes both a priori information and spatial smoothing constraints. Uncertainty estimates are provided organically by the MAP technique (Eq. 2). Note that BAR data are only available at present for 2006–2017.

For all MODIS products, data from the latest C6.1 are used. All products are provided at nominal (at-nadir) 10 km horizontal pixel size. Identical algorithms (and approaches for estimating uncertainty) are applied to both Terra and Aqua measurements, and the results of the evaluation were not distinguishable for Terra and Aqua data. For conciseness and to increase data volume Terra and Aqua data are not separated in the discussion going forward.

3.2.3 MISR data sets

The MISR sensor also flies on the Terra platform and consists of nine cameras viewing the Earth at different angles, with a fully overlapped swath width around 380 km (Diner et al., 1998). The latest version 23, used here, provides AOD retrievals at 4.4 km horizontal pixel size. Both land and water retrievals (Garay et al., 2017; Witek et al., 2018b) attempt retrieval using each of 74 candidate aerosol mixtures, although they differ in their surface reflectance models and uncertainty estimates. The overland “heterogeneous surface” retrieval estimates uncertainty as the standard deviation of AOD retrieved using those aerosol mixtures which provide a sufficiently close match to TOA measurements (Martonchik et al., 1998, 2009). The “dark water” approach (Witek et al., 2018b) looks at the variation of a cost function across the range of potential AOD and aerosol mixtures,

\[ f(\tau) = \frac{1}{N} \sum_{m=1}^{N} \chi_m^2(\tau), \]

where the sum is over \( N = 74 \) aerosol mixtures and \( \chi_m^2 \) is a cost function similar to the first term of Eq. (1). The uncertainty \( \epsilon_S \) is then taken as the full-width at half maximum of \( f(\tau) \), which is often found to be monomodal and close to Gaussian (Witek et al., 2018b). Note that MISR does not provide retrievals over Mbiba or Venice as the dark water algorithm logic excludes pixels within the matchup radius used here as too bright and unsuitable; thus, the approach cannot be evaluated at those sites.

3.2.4 ATSR data sets

The ATSRs were dual-view instruments measuring near-simultaneously at nadir and near 55° forward. ATSR2 (1995–2003) and AATSR (2002–2012) had four solar and three infrared bands, with approximately 1 km pixel sizes and a 550 km swath (although ATSR2 operated in a narrow-swath mode over oceans). Their predecessor ATSR1 lacked three of the solar bands and so has not been used widely for AOD retrieval. In 2016 the first of a new generation of successor instruments (the SLSTRs) was launched; SLSTR has several additional bands, a rear view instead of forward, the native spatial resolution of solar bands is finer, and the swath broader (Coppo et al., 2010). This study uses two data sets derived from this family of sensors.

ORAC is a generalized OE retrieval scheme which has been applied to multiple satellite instruments. Here, the version 4.01 ATSR2 and AATSR from the ESA CCI are used (Thomas et al., 2017), along with an initial version 1.00 of data from SLSTR. ORAC provides AOD retrievals over both land and ocean surfaces; the retrieval approaches are the same except for the surface reflectance models, which also inform the a priori and covariance matrices. Over water, surface reflectance is modelled according to Sayer et al. (2010a) with fairly strong a priori constraints. Over land, two approaches have been implemented in ORAC; the one used here is a model developed initially for the SU (A)ATSR retrieval algorithm (North et al., 1999), which assumes that the ratio between forward and nadir surface reflectance is spatially invariant and has very weak a priori constraints. Note that AOD and aerosol effective radius have weak and strong a priori constraints, respectively. Retrievals are performed at native resolution, and cost functions and uncertainty estimates are as in Eqs. (1) and (2) without smoothness constraints. ORAC simultaneously retrieves aerosol and surface properties, performing an AOD retrieval for each of a number (here, 10) of candidate aerosol optical models (mixing four components defined by the aerosol CCI; Holzer-Popp et al., 2013) and choosing the one with the lowest cost as the most likely solution. Retrievals passing quality checks (Thomas
et al., 2017) are then averaged to a 10 km Earth-referenced sinusoidal grid.

ADV uses the ATSR dual view over land to retrieve the contribution to total AOD from each of three aerosol CCI components (with the fraction of the fourth dust component prescribed from a climatology) by assuming that the ratio of surface reflectance between the sensor’s two views is spectrally flat. This has some similarity with the North et al. (1999) approach, except for ADV the ratio is estimated from observations in the 1600 nm band at which the atmosphere is typically most transparent, rather than being a freely retrieved parameter (Kolmonen et al., 2016). Over the water, the algorithm only uses the instruments’ forward view as this has a longer atmospheric path length and is less strongly affected by Sun glint. Because of this, the water implementation is often called ASV rather than ADV (Table 3), although for convenience here the term ADV is used throughout. Water surface reflectance is modelled as a combination of Fresnel reflectance and the chlorophyll-driven model of Morel (1988). The land and water algorithms treat other factors (e.g. aerosol optical models) in the same way. Unlike ORAC, ADV aggregates to a 10 km grid before performing the retrievals. ADV uncertainty estimates are calculated using Jacobians at the retrieval solution, i.e. the first component of Eq. (2), with $S$ assumed diagonal. The uncertainty on the TOA measurements is taken as 5 %, which is somewhat larger than that assumed by ORAC, so ADV is implicitly adding some forward model uncertainty into this calculation. Version 3.11 of the data sets (Kolmonen and Sogacheva, 2018), also from the ESA aerosol CCI, is used here.

Aside from pixel and/or swath differences, for both ADV and ORAC the implementation of the algorithms is the same for the three sensors. Matchups from the two (for ADV) or all three (for ORAC) sensors are combined here in the analysis to increase data volume due to the similarity in sensor characteristics and algorithm implementation. Note, however, that the difference in viewing directions between (A)ATSR and SLSTR (i.e. forward vs. rear) means that different scattering angle ranges are probed over the two hemispheres, which influences the geographic distributions of retrieval uncertainties. For both of these data sets, a large majority of matchups (75 % or more) obtained are with AATSR, as the ATSR2 mission ended before the AERONET network became as extensive as it is at present, and the SLSTR record to date is short. The results do not significantly change if only AATSR data are considered.

3.2.5 CISAR SEVIRI

Unlike the other data sets considered here, the SEVIRI sensors fly on geostationary rather than polar-orbiting platforms. This analysis uses data from the first version of the CISAR algorithm (Govaerts and Luffarelli, 2018) applied to SEVIRI aboard Meteosat 9; due to computational constraints, only SEVIRI data for 2008–2009 have been processed and included here. This sensor has a sampling cadence of 15 min and observes a disc centred over North Africa, covering primarily Africa, Europe, and surrounding oceans. The horizontal sampling distance is 3 km at nadir, increasing to around 10 km near the limits of useful coverage. This sampling means that several of the AERONET sites (GSFC, Kanpur, Midway Island, Pickle Lake, UCSB) are not seen by the sensor and cannot be analysed.

CISAR is also an OE retrieval scheme, which in its SEVIRI application accumulates cloud-free measurements from three solar bands over a period of 5 d and simultaneously retrieves aerosol and surface properties, reporting at each SEVIRI time step. Surface reflectance is modelled following Rahman et al. (1993) over land and Cox and Munk (1954a, b) over water, although the retrieval approach is otherwise the same between the two surface types. It employs a priori data and several smoothness constraints, so uncertainty estimates (Luffarelli and Govaerts, 2019) broadly follow Eq. (2).

3.3 Results

With the above criteria, the number of matchups $n$ obtained for each AERONET site with each data set is shown in Table 7. This additionally includes the long-term climatological mean (March 2000–February 2019) daytime cloud fraction $f_C$ from MODIS Terra, taken from the C6.1 level 3 monthly product (MOD08_M3) for the 1° grid cell in which the AERONET site lies. The cloud-masking approach is described by Frey et al. (2008), with more recent updates listed in Sect. 3 of Baum et al. (2012). Data from Terra are used as
Table 7. Number of matchups obtained for each AERONET site and data set, together with climatological cloud fraction.

| Data set | AERONET site | Straightforward sites | Complex sites |
|----------|---------------|-----------------------|---------------|
|          | Avignon | GSFC | Palencia | Ilorin | Kanpur | Pickle Lake |
| ADV      | 266     | 199  | 98      | 89     | 100    | 57         |
| BAR      | 1793    | 2088 | 2017    | 893    | 1087   | 1119       |
| CISAR    | 1749    | 0    | 868     | 493    | 0      | 0          |
| DB       | 3045    | 3010 | 1924    | 1144   | 1493   | 1068       |
| DT       | 2519    | 2409 | 1774    | 895    | 1250   | 529        |
| MISR     | 241     | 271  | 203     | 82     | 153    | 108        |
| ORAC     | 344     | 326  | 200     | 105    | 104    | 106        |
| Cloud fraction | 0.50 | 0.57 | 0.55    | 0.68   | 0.55   | 0.67       |

| Data set | AERONET site | Straightforward sites | Complex sites |
|----------|---------------|-----------------------|---------------|
|          | Ascension Island | Midway Island | UCSB | Cape Verde | Mbita | Venice |
| ADV      | 30     | 43    | 81    | 59     | 57    | 137     |
| CISAR    | 210   | 0     | 0     | 716    | 336   | 1442    |
| DT       | 748   | 443   | 1812  | 768    | 341   | 2698    |
| MISR     | 74    | 59    | 196   | 115    | 0     | 0       |
| ORAC     | 66    | 79    | 135   | 143    | 68    | 257     |
| Cloud fraction | 0.59 | 0.63 | 0.34  | 0.72   | 0.42  | 0.58     |

The majority of the aerosol data sets, like Terra, have a late-morning overpass time.

To make the counts more comparable between sites a sampling-corrected count \( \hat{n} \) can be calculated,

\[
\hat{n} = n \frac{\cos(\phi)}{(1 - f_C)} \frac{m_S}{m_A},
\]

where \( \phi \) is the site’s latitude (important as for polar-orbiting satellites a given latitude is overflown proportional to \( 1/\cos(\phi) \)), \( m_S \) the number of months of the satellite record, and \( m_A \) the number of months during the satellite record for which the AERONET site was in operation. For example, CISAR data used here cover the period 2008–2009 \( (m_S = 24) \); for these years, AERONET data at Ascension Island are available for 5 months in 2008 and 11 in 2009 \( (m_A = 16) \). Equation (9) thus provides a first-order estimate of the number of matchups which would have been obtained in the absence of clouds (as the data sets consider cloud-free pixels only), an equal rate of being overflown, and with the AERONET site in constant operation through the satellite lifetime. Normalizing each satellite data set to the maximum of \( \hat{n} \) across sites (to account for swath width and mission length differences, which determine total counts) provides a relative measure of how often each data set provides a valid retrieval at each location; the resulting relative sampling frequencies are shown in Fig. 6. This measure will be used in the ongoing discussion. Note that as CISAR is applied to geostationary SEVIRI data, the factor of \( \cos(\phi) \) is omitted (since each point on the disc is sampled once per scan, and each point outside the disc is never seen).

Graphical evaluations of the pixel-level uncertainties are shown in Figs. 7 and 8 for land and water retrievals, respectively. In both of these the left-hand column shows CDFs of absolute normalized error \( |\Delta_N| \) against theoretical expectations (see Fig. 3b), and the middle and right columns show the ED \( \epsilon_T \) and twice ED binned against the 1\( \sigma \) and 2\( \sigma \) points of absolute retrieval error \( |\Delta_S| \), respectively (see Fig. 4). Due to the very different sampling between data sets and sites (Table 7), the number of bins is taken as the lesser of \( n/20 \) or \( n^{1/3} \) (rounded to the nearest integer). This choice is a balance between well-populated bins to obtain robust statistics and the desire to examine behaviour across a broad range of \( \epsilon_T \). These figures also include an estimate of the digitization uncertainty on the binned values: for example, in a bin containing 100 matchups, the uncertainty on the 68th percentile (1\( \sigma \) point) binned value shown is taken as the range from the 67th to 69th matchup in the bin. For the MODIS-based records (which have the highest sampling) this digitization uncertainty is often negligible, but for others (ADV, MISR, ORAC) it is sometimes not.

A further way to look at the data is provided by Fig. 9, which shows the mean and standard deviation of \( \Delta_N \) for each data set and AERONET site; for unbiased retrievals with perfectly characterized errors (see Fig. 3a) the results should fall at coordinates \( (0, 1) \). This is a complement to the previously
Figure 7. Evaluation of pixel-level uncertainty estimates for overland retrievals. Each row corresponds to a different AERONET site, and colours are used to distinguish data sets. The left-hand column shows a CDF of the absolute normalized retrieval error $|\Delta N|$ (see Fig. 3b), and the middle and right columns show $1\sigma$ and $2\sigma$ expected discrepancy ED vs. absolute retrieval errors $|\Delta S|$ (see Fig. 4), respectively. In the left column, theoretical expectations are shaded grey; in the others, the $1:1$ line is indicated dashed in grey, and vertical bars indicate the uncertainty on the bin value, as described in the text.
Figure 8. As Fig. 7, except for AERONET sites used for over-water retrieval evaluation.
shown CDFs as it also provides measures of systematic bias in the AOD retrieval and systematic problems in estimating error magnitude: horizontal displacement from the origin indicates the relative magnitude and direction of systematic error, and vertical displacement indicates a general underestimation or overestimation of the typical level of error. Further, it shows how closely (or not) results from the different sites cluster together. For a larger-scale analysis of hundreds of AERONET sites, this type of plot could be expanded to a heat map. The CDFs in Figs. 7 and 8 assess the overall magnitude of normalized errors and the shape of the distribution, while the binned ED assesses the overall skill in the specificity of the estimates. In these figures, the top and bottom three rows show sites expected to be straightforward or complicated test cases for the uncertainty estimate techniques (Table 6). Table 8 provides the overall calibration skill scores for 1σ error at each site (Eq. 5), plus the coefficient of determination $R^2$ (where at least three bins were available) between binned uncertainty and 1σ error from the middle columns of Figs. 7 and 8. Together, these facilitate a visual and quantitative evaluation of the pixel-level uncertainty estimates.

### 3.3.1 Land sites

Turning to the land sites (Fig. 7), all the techniques show some skill in that the ED generally increases with retrieval error. There is, however, considerable variation between sites (which points to the utility of considering results site by site for this demonstration analysis) and data sets. For the straightforward sites, there is an overall tendency for the uncertainty estimates to be too large. This may indicate that the retrieval error budgets are a little too pessimistic; since overall errors and uncertainties also tend to be small at these sites, it is also possible that the uncertainty on the AERONET data (which can be a non-negligible contribution to ED here) is overestimated. A notable exception here is MISR, for which uncertainty estimates are very close to theoretical expectations. This implies that the overall assumptions made by this technique (that the principle contribution to error is in aerosol optical model assumptions, and the 74 mixtures provide a representative set such that the standard deviation of retrieved AOD between well-fitting mixtures is a good proxy for uncertainty) are valid. A second exception is CISAR, which more significantly overestimates the uncertainty, indicating that the retrieval is more robust than expected. For these sites the binned plots of 1σ and 2σ retrieval error vs. ED look similar, suggesting that, within each bin, the retrieval errors are roughly Gaussian (even if the magnitudes of uncertainty are not perfectly estimated). MODIS DT tends to overestimate uncertainty on the low end and underestimate on the high end, suggesting (at least for these sites) that the first and second coefficients in the expression $\epsilon_S = \pm (0.05 + 0.15\tau)$ may need to be decreased and increased, respectively.

For the complex land sites, the picture is different. At Ilorin, MODIS DB and ADV tend to overestimate uncertainty, while the others underestimate it. This site was chosen as a test case because of the complexity of its aerosol optical properties, which are more absorbing than assumed by many retrieval algorithms and can show large spatiotemporal heterogeneity due to a complex mix of sources (Eck et al., 2010; Giles et al., 2012; Fawole et al., 2016). Using aircraft measurements, Johnson et al. (2008) found mid-visible single-scattering albedo (SSA) from smoke-dominated cases between 0.73 and 0.93, with a central estimate for the smoke component of 0.81. DB has a regional SSA map with more granularity (Hsu et al., 2019), while the other algorithms do not contain sufficiently absorbing particles, leading to a breakdown in their uncertainty estimates when strong absorption is present.

The most absorbing component in the MISR aerosol mixtures has an SSA of 0.80 at 558 nm; mixtures including this...
component have SSA from 0.81 to 0.96, and all other MISR mixtures have SSA > 0.90 (Tables 2, 3 of Kahn et al., 2010). In smoke cases retrievals are biased low and the uncertainty estimates are too narrow because the set of candidate aerosol mixtures is not representative of optical properties at this location. MODIS DT and BAR (which uses the same optical models as DT) assume a fine-mode-dominated model with mid-visible SSA of 0.85 for December–May and 0.90 for June–November (Fig. 3 of Levy et al., 2007); this is mixed with a less absorbing coarse-dominated model, so they suffer from similar issues. CISAR retrieves AOD by a combination of aerosol vertices in SSA-asymmetry parameter space; the most absorbing (for SEVIRI’s 640 nm band, which is the shortest wavelength) has SSA around 0.79 (Fig. 4 of Luf-farelli and Govaerts, 2019). Due to the spectral curvature of smoke SSA, this would imply a weaker effective absorption in the mid-visible. ADV and ORAC share aerosol components prescribed by the aerosol CCI (Holzer-Popp et al., 2013); the most absorbing fine-mode component has mid-visible SSA around 0.80, although this is also always mixed with more weakly absorbing fine-mode (which have SSA of 0.98) and coarse-mode particles in varying proportions, so in practice the assumed SSA is always higher (Tables 1 and 2 of Thomas et al., 2017). It may be that ADV is providing reasonable estimates at this site despite this to the somewhat larger assumed forward model uncertainty than ORAC. For Kanpur, except for MISR (which has similar issues as Ilorin) and CISAR (as SEVIRI does not observe the site), these issues are lessened. This may be because, while Kanpur has similar complex mixed aerosol conditions, the components are overall less strongly absorbing and so these issues are less acute, with a typical SSA (similar to that of Ilorin in mixed, as opposed to smoke-dominated, conditions) around 0.89 (Giles et al., 2012). The issues with MISR may imply that the wrong mixture(s) are being selected here.

The case at Pickle Lake is more diverse: similar to the straightforward sites, MODIS DT, DB, and BAR all over-estimate uncertainty. ADV and MISR are fairly close to theoretical values; despite this, their skill scores are fairly low (Table 8) as the magnitudes of their uncertainties are not perfect and the range of 1σ retrieval errors is fairly small. All these algorithms provide retrievals significantly less often than would be expected by the site’s cloud cover, latitude, and AERONET availability (Fig. 6). This implies that the algorithms may be coping with a potential violation of assumptions (i.e. land mask issues from numerous small lakes) by simply not providing a retrieval at all. ORAC underestimates uncertainties at this site but provides retrievals relatively more frequently than the other data sets. As the land–sea mask is determined at full (1 km) resolution and used to set the surface model, it is likely that some of the pixels within the 10 km grid will be affected by misflagging and/or mixed surface issues, contributing to additional errors which are not being caught by these quality checks. Which

Table 8. Calibration skill scores $s_{\text{cal}}$ and coefficient of determination $R^2$ from binned 1σ uncertainties in Figs. 7 and 8.

| Data set   | AERONET site          | 
|-----------|-----------------------|
|           | s_{cal} | R^2 | s_{cal} | R^2 | s_{cal} | R^2 | s_{cal} | R^2 | s_{cal} | R^2 | s_{cal} | R^2 |
| Land calibration skill scores; $R^2$ |
| Straightforward sites | Complex sites |
| Avignon | GSFC | Palencia | Ilorin | Kanpur | Pickle Lake |
| ADV | <0 | 0.58 | <0 | 0.87 | <0 | 0.72 | 0.29 | 0.99 | <0 | 0.99 | 0.08 | 0.99 |
| BAR | <0 | 0.55 | <0 | 0.94 | <0 | 0.57 | <0 | 0.37 | <0 | 0.84 | 0.11 | 0.88 |
| CISAR | <0 | 0.67 | – | – | <0 | 0.075 | <0 | 0.94 | – | – | – | – |
| DB | <0 | 0.98 | 0.57 | 0.98 | 0.65 | 0.96 | <0 | 0.99 | <0 | 0.97 | 0.61 | 0.86 |
| DT | 0.57 | 0.89 | 0.47 | 0.89 | 0.53 | 0.92 | <0 | 0.04 | 0.69 | 0.91 | 0.50 | 0.99 |
| MISR | 0.84 | 0.85 | 0.97 | 0.99 | 0.93 | 0.98 | 0.62 | 0.75 | <0 | 0.38 | 0.87 | 0.96 |
| ORAC | <0 | 0.82 | <0 | 0.70 | <0 | 0.36 | 0.82 | 0.95 | <0 | 0.88 | <0 | 0.84 |
| Water calibration skill scores; $R^2$ |
| Straightforward sites | Complex sites |
| Ascension Island | Midway Island | UCSB | Cape Verde | Mbita | Venice |
| ADV | <0 | – | <0 | – | 0.40 | 0.79 | <0 | 0.91 | <0 | 0.35 | 0.11 | 0.70 |
| CISAR | <0 | 0.11 | – | – | – | – | <0 | 0.42 | <0 | 0.28 | <0 | 0.33 |
| DT | 0.72 | 0.87 | 0.38 | 0.95 | 0.45 | 0.99 | 0.73 | 0.93 | 0.62 | 0.92 | 0.63 | 0.98 |
| MISR | 0.52 | 0.94 | <0 | 0.45 | 0.80 | 0.97 | 0.78 | 0.94 | – | – | – | – |
| ORAC | <0 | 0.84 | <0 | 0.48 | <0 | 0.06 | <0 | 0.92 | <0 | 0.047 | <0 | 0.063 |
behaviour is more desirable (no data vs. more uncertain data than expected) is a philosophical and application-dependent matter. As it lies outside the SEVIRI disc, CISAR provides no retrievals at this site.

Aside from DB, DT, and MISR, skill scores (Table 8) are in most cases negative; for the former two the uncertainty estimates are somewhat empirical and not independent of the AERONET data, so the fact they are fairly well-calibrated is not surprising. Despite this $R^2$ is typically not negligible (although the small number of bins means the estimates of $R^2$ are somewhat uncertain). This implies that, while the absolute magnitudes of estimated uncertainty are often too small or large, the techniques do show some skill at predicting which retrievals are comparatively less or more uncertain at a variety of locations. Neither $s_{cal}$ nor $R^2$ should be over-interpreted in terms of site-to-site variations, as these depend strongly on the number of bins, the range in estimated uncertainties, and the range in actual retrieval errors at a given site. The main points of note are whether $s_{cal} > 0$ and whether there is a positive association between binned uncertainty and error.

### 3.3.2 Water sites

For the water sites (Fig. 8), only five satellite data sets are available – also recall that the MODIS DT uncertainty envelope is narrower than over land, and the MISR uncertainty is a PDF based on a cost function composited over AOD and aerosol mixtures rather than (as over land) a simple standard deviation. At the straightforward sites there is some commonality with the land sites. Specifically, the MISR approach works fairly well, CISAR overestimates uncertainty (although of the three, only Ascension Island is within the SEVIRI disc), and MODIS DT slightly overestimates uncertainty overall, with a tendency to overestimate on the low end and underestimate on the high end. In general a similar picture is also seen in terms of $s_{cal}$ and $R^2$: most data sets are not well-calibrated, although there is skill at assessing variations in uncertainty at individual sites.

ADV and ORAC are more systematic in their underestimation of uncertainty over water compared to over land, although as the over-water errors are often fairly small in absolute terms, they appear fairly large in relative terms. This difference in the ATSR-based records between land and ocean sites is intriguing. ADV assumes 5% uncertainty in the TOA signal, while ORAC includes separate measurement and forward model terms for a slightly lower total uncertainty overall (typically 3%–4% dependent on band and view), which in part explains ORAC’s larger normalized errors. The common behaviour either implies (1) that the calibration of the sensors may be biased or more uncertain than expected for these fairly dark ocean scenes or (2) that the over-water surface reflectance models or (for ORAC) their uncertainties (either in their contribution to forward model error in $S_o$, or the strength of the a priori constraint in $S_a$) might be less reliable than assumed. Figure 9 implies that there is a significant systematic error source in ORAC contributing to a positive bias over water. A thorough comparison between the two data sets using the matchups collected here is difficult due to the fairly low data volumes involved, especially for ADV. ADV provides significantly fewer retrievals overall than ORAC (for both land and water), implying stricter pixel selection and/or retention criteria; this is consistent with ESA CCI validation analysis of earlier versions of these data sets by Popp et al. (2016) and Kinne et al. (2017).

Despite the expected complexities at Cape Verde from mixtures of low-level sea spray and higher-altitude nonspherical mineral dust (Mishchenko et al., 1997; Kalashnikova et al., 2005), the error characterization at this complex site does not appear different from that obtained at the more straightforward sites. Interestingly, these algorithms seem more selective about when to provide retrievals at the three straightforward sites than they are at Cape Verde (Fig. 6). The reasons for this are unclear unless the estimate provided by $\hat{h}$ (Eq. 9) is not a good approximation for these sites; each is close to the coast and all should be roughly equally affected by Sun-glint sampling-related losses.

Mbita is in some sense the inverse of the land site Pickle Lake, and similar comments apply. MODIS DT uncertainties are reasonable, although the data volume is fairly low relative to expectations from Fig. 6. ADV and ORAC retrieve more frequently and perform well but with more high-error outliers than expected, likely due to mixed or misflagged land–water pixels. CISAR retrieves with a similar frequency at Mbita as Ascension Island (that is, less than expected but no less so than at the straightforward site). Looking at the binned ED vs. error, the errors for the $1\sigma$ points (Fig. 8n) are slightly overestimated and those for the $2\sigma$ points (Fig. 8o) underestimated, implying more extreme outliers than expected and indicating possible surface contamination issues. Note that MISR does not provide retrievals at this site as the algorithm does not consider Lake Victoria to be dark water.

Venice is sampled close to the expected rates by ADV, CISAR, MODIS DT, and ORAC (Fig. 6), and again it is excluded by MISR due to the bright, turbid water. Here, the CISAR $1\sigma$ retrieval error is $\sim 0.05$ and the $2\sigma$ error is about double that, regardless of the ED; the uncertainty estimates do not show skill overall. As SEVIRI’s wavelengths (640, 810, 1640 nm) are less strongly affected by water turbidity than the other sensors, the issues causing complexity here may not apply, and the overall tendency for CISAR to report too large an uncertainty may be dominating. ADV and DT results are reasonably in line with expectations, implying either that the turbid water is not a hindrance for the algorithm or that the additional uncertainty from this factor is compensated for by lower uncertainties in some other aspect of the algorithm. ORAC tends to more strongly underestimate the retrieval uncertainty. The water surface reflectance model (Sayer et al., 2010a) is based on low-turbidity Case I water (Morel, 1988), so it is likely providing a low-biased a
priori for the retrieval with too strong a constraint, leading to a high bias in AOD retrievals with overly high confidence in the solution, which becomes large when expressed in normalized terms.

4 Conclusions and path forward

Pixel-level uncertainty estimates in AOD products are an important complement to the retrievals themselves to allow users to make informed decisions about data use for data assimilation and other applications. Ideal estimates are prognostic (predictive), and these are increasingly being provided within data sets; when they are absent, diagnostic estimates can be used as a stopgap. This study has reviewed existing diagnostic and prognostic approaches, provided a framework for their evaluation against AERONET data, and demonstrated this framework using a variety of satellite data products and AERONET sites. It is hoped that this methodology can be adopted by the broader community as an additional component of data product validation efforts. Several conclusions about the performance of these existing estimates follow.

1. All tested techniques show skill in some situations (in that the association between estimated uncertainty and observed error is positive, and on average magnitudes are reasonable), although none are perfect, and there is no clear single best technique. Small data volumes for some sensors and locations limit the extent to which performance in the high-uncertainty regime can be probed.

2. The points in Fig. 9 tend to cluster by data set more strongly than by site. This implies that some of the quantitative limitations in the uncertainty estimates provided within the current data sets are large-scale issues (e.g. persistent underestimate or overestimate of some aspect of the retrieval error budget). Further, as the performance at expected straightforward vs. complex AERONET sites was not always distinct, these limitations (or other unknown factors) may at present be more significant error sources than the issues associated with the ground sites.

3. While skilful, the uncertainties are not always well-calibrated; i.e. they are often systematically too large or too small. If characterization of the error budgets of the retrievals cannot be significantly improved, it is plausible that a simple scaling (using e.g. averages of the standard deviations on the y axis in Fig. 9) could be developed to bring the magnitudes more into line with the expected values.

4. The formal error propagation techniques (employed here by BAR, CISAR, and ORAC) are very powerful. Their differing behaviour and performance illustrate the difficulties in appropriately quantifying terms for the forward model, a priori covariance matrices, and appropriate smoothness constraints. For these sites, CISAR tends to overestimate the uncertainty most strongly, BAR to overestimate slightly, and ORAC to underestimate (more strongly over water than land). The simpler approach taken by ADV (Jacobians from a flat 5% error on TOA reflectance) tends to be about right over land but also underestimates the true uncertainty over water.

5. The empirical validation-based MODIS DB approach works well but on average overestimates the total uncertainty and at these sites has little bias overall. That may indicate that the sites used here are coincidentally better-performing than the global results used to fit the expression. This points to the fact that the expression (which draws on AOD, geometry, quality flag, and surface types) captures many, but not all, of the factors relevant for quantifying total uncertainty.

6. The diagnostic MODIS DT approaches perform reasonably well if used instead as prognostic uncertainty estimates; they have a tendency to be insufficiently confident (overestimate uncertainty) on the low end and overconfident (underestimate uncertainty) on the high end. Despite the possibility for unphysical negative AOD retrievals in the DT land product, both land and ocean results indicate a systematic positive bias in the retrievals.

7. MISR’s two approaches (applied for land and water surfaces) are both based on diversity between different candidate aerosol optical models. They both perform well at most sites, although they have a tendency to underestimate the total uncertainty slightly. The implication from this is that the diversity in AOD retrievals from different candidate optical models does capture the leading cause of uncertainty in the MISR retrievals. The fact that they are underestimates does imply at least one remaining important factor which is not captured by this diversity, which could perhaps be a systematic error source such as a calibration or retrieval forward model bias.

More broadly, these results suggest paths for the development and refinement of pixel-level AOD uncertainty estimates for existing and new data sets. For algorithms attempting AOD retrievals from multiple candidate aerosol optical models, the diversity in retrieved AOD between these different models could be a good proxy for part of the retrieval uncertainty. The MODIS DT ocean and ORAC algorithms both perform retrievals for multiple optical models. As ORAC is already an OE retrieval, this aerosol-model-related uncertainty is one of the few components not directly included in the existing error budget, so it could perhaps be added in quadrature to the existing uncertainty estimate. MODIS DT provides only a diagnostic AOD uncertainty estimate; diversity between possible solutions (which draw from 20 possible combinations of four fine modes and five coarse modes)
A general principle behind the error propagation techniques is the assumption of Gaussian departures from some underlying forward model. When this is not true, the techniques tend to fail. The Ilorin case is one such example of this. Another is the higher-level issue of coastal or lake areas, as most algorithms make binary retrieval decisions with nonlinear implications (e.g. treat pixel as land or water for surface reflectance modelling), which cause problems if pixels are either misflagged or “contaminated” and contain mixed water or land. The algorithms tested here tend to deal with this in one of two ways. The first is simply to fail to provide a valid retrieval at all; in this case, the uncertainty estimates for available retrievals tend to be reasonable, although the data volume is significantly less than expected. The second option is to provide a retrieval but consequently provide a poor estimate (and typically an underestimate) of the associated uncertainty. Neither is entirely satisfactory. Performing retrievals at a higher spatial resolution with strict filtering might ameliorate these issues, as a smaller fraction might then be contaminated or misflagged; however, the resolutions of the sensor measurements and land mask (and its quality) place hard constraints on what could be achieved. Another option might be to attempt retrievals using both land and water algorithms for these pixels and either report both or an average (including the difference between them as an additional contribution to the uncertainty estimate). This would provide some measure of the potential effect of surface misclassification and at the least provide a larger uncertainty estimate to alert the data user about problematic retrieval conditions. A deeper understanding of the representativeness of AERONET sites on satellite retrieval scales would be useful to better understand the distributions of retrieval success rates and errors. This is a topic of current research (e.g. Schutgens et al., 2016, 2017; Kinne et al., 2013; Li et al., 2016; Schutgens, 2019) although often on a temporal basis or on coarser spatial scales than relevant for L2 validation.

A further difficulty in the assumption of Gaussian random errors is that sensor calibration uncertainty tends to be dominated by systematic effects rather than random noise. While in practice it is often (as in the algorithms assessed here) treated as a random error source, when it is a dominant contribution to the retrieval error budget it will tend to skew the retrievals toward one end of the notional uncertainty envelopes. This may explain some of the systematic behaviour along the $x$ axis in Fig. 9 within individual data sets (although the position along this axis is determined not only by the actual error, but also the estimated uncertainty). As discussed in Sect. 2.4, a pragmatic method for amelioration of this (if the forward model contribution to the systematic uncertainty cannot be significantly reduced by improvements to retrieval physics) would be to perform a vicarious calibration. Shipborne AOD observations were also used as one part of the MISR calibration strategy for low-light scenes (Witek et al., 2018a); if this removes the bulk of the systematic calibration error, it may help explain why the uncertainty estimation technique (dispersion in possible solutions with different aerosol optical model assumptions) generally works so well.

The framework for evaluating uncertainties here is general and not restricted to AOD. In practice, however, it is difficult to extend it to other aerosol-related quantities at the present time. For profiling data sets (such as lidar), uncertainties in extinction profiles are often strongly vertically correlated as the effects of assumptions propagate down the profile (Young et al., 2013). An assessment would also have to account for the vertical resolution of the sensors and compute appropriate averaging kernels (Rodgers, 2000); this is by no means intractable and has been done using ground-based lidar systems for aerosol properties (e.g. Povey et al., 2014) as well as other geophysical quantities (e.g. atmospheric temperature by Sica and Haefele, 2015). Possibly a stronger limitation is that there are relatively few validation-quality data sets (i.e. with significantly smaller uncertainty than the spaceborne sensor) to compare them to, so the ground-based contribution to the total expected discrepancy would not be negligible.

For the total column, other key quantities of interest include the Ångström exponent (AE), fine-mode fraction (FMF) of AOD, and aerosol SSA. The AE can easily be assessed using this framework, although AERONET AE itself can be quite uncertain in the low-AOD conditions which predominate in many locations around the globe (Wagner and Silva, 2008). In that case the expected discrepancy would include significant contributions from AERONET uncertainty, so the comparison would be less informative about the quality of the satellite uncertainty estimate. These issues are somewhat lessened in high-AOD conditions, however. Similar comments apply to AERONET FMF, which has an uncertainty of the order of ±0.1 in moderate- to high-AOD conditions and larger when AOD is low (O’Neill et al., 2003, 2006). The framework presented here would not become invalid in these cases (although it becomes statistically problematic for locations where FMF is close to the bounds 0 or 1) but would become a measure of the joint consistency of both satellite and AERONET uncertainties, rather than a test primarily of the satellite uncertainty estimates. Some of these issues are lessened if, instead of FMF, fine-mode AOD (i.e. the product of FMF and AOD) and coarse-mode AOD are used. While AOD is also positive definite, numerical issues associated with AOD near 0 can be removed if retrievals are performed in log space, reflecting the closer-to-lognormal distributions of AOD found in nature (O’Neill et al., 2000; Sayer and Knobelspiesse, 2019); ORAC, for example, retrieves AOD in log space.

Issues with SSA are somewhat more difficult; AERONET almucantar inversions have an uncertainty in SSA around
±0.03 under favourable conditions (moderate to high AOD and large solar zenith angle) but uncertainties can be significantly larger otherwise (Dubovik et al., 2000). Given that SSA (like FMF) is inherently bounded in the range 0–1, and most aerosol types have SSA in the visible spectral region around 0.8–1 (e.g. Dubovik et al., 2002), in practical matters this uncertainty is a significant fraction of the variability in the parameter to be observed. Further, the hard boundary of SSA = 1 means that the Gaussian statistics on which many uncertainty estimates and part of this framework rely will be less useful models of the real error characteristics. As such (similarly to FMF) it may be better to assess related optical properties, such as absorption AOD (AAOD), rather than SSA itself. This would address some of the statistical issues (plus AAOD is more directly connected to the radiative effect than SSA alone) but would not remove the underlying difficulty of accurate quantification of aerosol absorption, which remains both difficult to measure and difficult to retrieve from ground, airborne, or satellite remote sensing. Despite these difficulties with other aerosol properties (and the current limitations of techniques for quantifying AOD uncertainty), the routine provision, evaluation, and scientific use of prognostic AOD uncertainty estimates from satellite remote sensing will constitute an important step toward more optimal and robust applications of these data sets.

Data availability. AERONET data are available from https://aeronet.gsfc.nasa.gov (NASA AERONET team, 2020). MODIS DB and DT as well as MISR data are available from https://earthdata.nasa.gov (NASA Earthdata team, 2020). ADV and ORAC data are available from http://www.esa-aerosol-cci.org (ESA Aerosol CCI team, 2020). CISAR data are available from http://www.icare.univ-lille1.fr/archive (ICARE archive team, 2020). BAR data are available on request to the co-authors AL and TM.

Author contributions. AMS conceptualized the study, provided MODIS DB data, performed the analysis, and led the writing of the paper. ACP provided ORAC data. PK, AL, and TM provided ADV and BAR data. FP provided MODIS DT data. MW provided MISR data. YG and ML provided CISAR data. TP and KS provided general guidance and insight through ESA aerosol CCI and AeroSat validation and uncertainty characterization activities; TP also contributed significantly to table outlining and referencing approaches to uncertainty characterization. All authors contributed to editing the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The work of lead author Andrew M. Sayer was performed as development for the forthcoming NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission (https://pace.gsfc.nasa.gov, last access: 29 January 2020). ORAC data were generated by Roy G. Grainger (Oxford), Caroline R. Poulsen (RAL), co-author Adam C. Povey (Oxford), Simon R. Proud (Oxford), and Gareth E. Thomas (RAL) with the support of the European Space Agency’s Climate Change Initiative Aerosol Project, the Copernicus Climate Change Service, and the National Centre for Earth Observation. The AERONET team (led by Brent N. Holben, NASA GSFC) and site investigators and managers are thanked for the creation and maintenance of AERONET, which is an invaluable resource in the aerosol remote sensing enterprise. This research topic was initiated as part of the international AeroSat group of aerosol researchers, led by Ralph A. Kahn (NASA GSFC) and co-author Thomas Popp (DLR), who meet once a year to discuss and move toward solving common issues in the field of aerosol remote sensing. The authors are very grateful to attendees at AeroSat meetings over the past several years for numerous fruitful presentations and discussions on this and related topics. Additionally, Ghassan Taha (USRA), Stuart A. Young (CSIRO), and John Yorks (NASA GSFC) are acknowledged for discussions about profiling instruments. Finally, the authors would like to thank Oleg Dubovik (University of Lille), Robert C. Levy (NASA GSFC), and two anonymous reviewers for thoughtful and appreciated comments on this paper.

Financial support. This research has been supported by the NASA.

Review statement. This paper was edited by Alexander Kokhanovsky and reviewed by three anonymous referees.

References

Ahn, C., Torres, O., and Jethva, H.: Assessment of OMI near-UV aerosol optical depth over land, J. Geophys. Res.-Atmos., 119, 2457–2473, https://doi.org/10.1002/2013JD020188, 2013.
Baum, B. A., Menzel, W. P., Frey, R. A., Tobin, D. C., Holz, R. E., Ackerman, S. A., Heidinger, A. K., and Yang, P.: MODIS Cloud-Top Property Refinements for Collection 6, J. Appl. Meteorol. Clim., 51, 1145–1163, https://doi.org/10.1175/JAMC-D-11-0203.1, 2012.
Benedetti, A., Morcrette, J.-J., Boucher, O., Dethof, A., Engelen, R. J., Fisher, M., Flentje, H., Huneeus, N., Jones, L., Kaiser, J. W., Kinne, S., Mangold, A., Razinger, M., Simmons, A. J., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13205, https://doi.org/10.1029/2008JD011115, 2009.
Benedetti, A., Reid, J. S., Knippertz, P., Marsham, J. H., Di Giuseppe, F., Rémy, S., Basart, S., Boucher, O., Brooks, I. M., Menut, L., Mona, L., Laj, P., Pappalardo, G., Wiedensohler, A., Baklanov, A., Brooks, M., Colarco, P. R., Cuevas, E., da Silva, A., Escribano, J., Flemming, J., Huneeus, N., Jorba, O., Kazadzis, S., Kinne, S., Popp, T., Quinn, P. K., Sekiyama, T. T., Tanaka, T., and Terradellas, E.: Status and future of numerical atmospheric aerosol prediction with a focus on data requirements, Atmos. Chem. Phys., 18, 10615–10643, https://doi.org/10.5194/acp-18-10615-2018, 2018.
Dawid, A. P.: The Well-Calibrated Bayesian, J. Am. Stat. Assoc.,
Curier, R. L., Veefkind, J. P., Braak, R., Veihelmann, B., Torres,
Cox, C. and Munk, W.: Measurement of the roughness of the sea
Coppo, P., Ricciarelli, B., Brandani, F., Delderfield, J., Ferlet, M.,
Clarisse, L., Coheur, P.-F., Prata, F., Hadji-Lazaro, J., Hurtmans, D.,
Carroll, M. L., DiMiceli, C. M., Townshend, J. R. G., Sohlberg,
Capelle, V., Chédin, A., Siméon, M., Tsamalis, C., Pierangelo,
Bevan, S. L., North, P. R. J., Los, S. O., and Grey, W. M. F.: A
398 A. M. Sayer et al.: AOD uncertainty evaluation
Atmos. Meas. Tech., 13, 373–404, 2020 www.atmos-meas-tech.net/13/373/2020/
1982.
77, 605–610, https://doi.org/10.1080/01621459.1982.10477856,
2010.
https://doi.org/10.1029/2007JD008738, 2008.
plication to western Europe, J. Geophys. Res., 113, D17S90,
from OMI radiances using a multiwavelength algorithm: Ap-
Sun glitter, J. Mar. Res., 13, 198–227, 1954b.
44, 838–850, https://doi.org/10.1364/JOSA.44.000838, 1954a.
results, and perspectives, J. Quant. Spectrosc. Ra., 224, 474–511,
https://doi.org/10.1016/j.jqsrt.2018.11.024, 2019.
Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A.,
O’Neill, N. T., Slutsker, I., and Kinne, S.: Wavelength depen-
dence of the optical depth of biomass burning, urban, and desert
dust aerosols, J. Geophys. Res., 104, 31333–31349, 1999.
Eck, T. F., B. N. Holben, A. Sinyuk, R. T. Pinker, P. Goloub,
H. Chen, B. Chatenet, Z. Li, R. P. Singh, S. N. Tripathi, J. S. Reid,
D. M. Giles, O. Dubovik, N. T. O’Neill, A. Smirnov, P. Wang,
and X. Xia: Climatological aspects of the optical properties of
fine/coarse mode aerosol mixtures, J. Geophys. Res., 115, D19205,
https://doi.org/10.1029/2010JD014002, 2010.
Eck, T. F., Holben, B. N., Reid, J. S., Mukelabai, M. M., Piketh,
S. J., Torres, O., Jethva, H. T., Hyer, E. J., Ward, D. E., Dubovik,
O., Sinyuk, A., Schafer, J. S., Giles, D. M., Sorokin, M., Smirnov,
A., and Slutsker, I: A seasonal trend of single scattering albedo
in southern African biomass-burning particles: Implications for
satellite products and estimates of emissions for the world’s
largest biomass-burning source, J. Geophys. Res.-Atmos., 118,
6414–6432, https://doi.org/10.1002/jgrd.50500, 2013.
ESA Aerosol CCI team: Climate Change Initiative aerosol portal, available at: https://www.esa-aerosol-cci.org, last access: 29 January 2020.

EUMETSAT: Polar Multi-Sensor Aerosol Product: ATBD, Tech. rep., EUMETSAT, Darmstadt, Germany, available at: https://www.eumetsat.int/ (last access: 29 January 2020), report EUM/TSS/SPE/14/739904, v3B, 2015.

Fawole, O. G., Cai, X., Levine, J. G., Pinker, R. T., and MacKenzie, A. R.: Detection of a gas flaring signature in the AERONET optical properties of aerosols at a tropical station in West Africa, J. Geophys. Res.-Atmos., 121, 14513–14524, https://doi.org/10.1002/2016JD025584, 2016.

Flamant, P., Cuesta, J., Denneulin, M., Dabas, A., and Huber, D.: ADM-Aeolus retrieval algorithms for aerosol and cloud products, Tellus A, 60, 273–288, https://doi.org/10.1111/j.1600-0870.2007.00287.x, 2008.

Flamant, T., Stiegitz, H., Dabas, A., and Huber, D.: Aeolus L2A aerosol products: principle and first glimpse on performances, in: ESA Living Planet Symposium, Milan, Italy, available at: https://lps19.esa.int (last access: 29 January 2020), 2019.

Franz, B. A., Bailey, S. W., Werdell, P. J., and McClain, C. R.: Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry, Appl. Opt., 46, 5068–5082, https://doi.org/10.1364/AO.46.005068, 2007.

Frey, R. A., Ackerman, S. A., Liu, Y., Strabala, K. I., Zhang, H., Key, J. R., and Wang, X.: Cloud Detection with MODIS. Part I: Improvements in the MODIS Cloud Mask for Collection 5, J. Atmos. Ocean. Tech., 25, 1057–1072, https://doi.org/10.1175/2008JTECHA1052.1, 2008.

Fukuda, S., Nakajima, T., Takanaka, H., Higurashi, A., Kikuchi, K., and van Noije, T.: Model-based estimation of sampling-caused uncertainty in aerosol remote sensing for climate research applications, Q. J. Roy. Meteor. Soc., 140, 2353–2363, https://doi.org/10.1002/qj.3293-3292-2016, 2016.

Günther, A., Höpfnner, M., Sinnhuber, B.-M., Griessbach, S., Deshler, T., von Clarmann, T., and Stillier, G.: MIPAS observations of volcanic sulfate aerosol and sulfur dioxide in the stratosphere, Atmos. Chem. Phys., 18, 1217–1239, https://doi.org/10.5194/acp-18-1217-2018, 2018.

Hanel, R., Conrath, B., Hovis, W., Kunde, V., Lowman, P., Maguire, W., Pearl, J., Pirraglia, J., Prabhakara, C., and Schlachmann, B.: Investigation of the Martian environment by infrared spectroscopy on Mariner 9, Icarus, 17, 423–442, https://doi.org/10.1016/S0019-1035(98)00031-5, 1998.

Hasekamp, O. P. and Landgraf, J.: Retrieval of aerosol properties over land surfaces: capabilities of multi-viewing-angle intensity and polarization measurements, Appl. Optics, 46, 3332–3344, https://doi.org/10.1364/AO.46.003332, 2007.

Heidinger, A. K., Straka III, W. C., Molling, C. C., Sullivan, J. T., and Wu, X.: Deriving an inter-sensor consistent calibration for the AVHRR solar reflectance data record, Int. J. Remote Sens., 31, 6493–6517, https://doi.org/10.1080/01431161.2010.496472, 2010.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET: A federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, https://doi.org/10.1016/S0034-4257(98)00031-5, 1998.

Holzer-Popp, T., Schroeder, M., and Geisler, G.: Retrieving aerosol optical depth and type in the boundary layer over land and ocean from simultaneous GOME spectrometer and ATSR-2 radiometer measurements, I. Method description, J. Geophys. Res., 107, 4578, https://doi.org/10.1029/2001JD001203, 2002.

Holzer-Popp, T., Schroeder-Homscheidt, M., Breitkreuz, H., Martynenko, D., and Kläser, L.: Improvements of synergistic aerosol retrieval for ENVISAT, Atmos. Chem. Phys., 8, 7651–7672, https://doi.org/10.5194/acp-8-7651-2008, 2008.
Martonchik, J. V., Kahn, R. A., and Diner, D. J.: Aerosol Remote Sensing, Springer (Berlin), https://doi.org/10.1007/978-3-642-17725-5, 2013.

Levy, R. C., Remer, L. A., and Dubovik, O.: Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land, J. Geophys. Res., 112, D13210, https://doi.org/10.1029/2006JD007815, 2007.

Levy, R. C., Leptoukh, G. G., Kahn, R., Zubko, V., Gopalan, A., and Remer, L. A.: A Critical Look at Deriving Monthly Aerosol Optical Depth From Satellite Data, IEEE T. Geosci. Remote, 47, 2942–2956, https://doi.org/10.1109/TGRS.2009.2013842, 2009.

Levy, R. C., Matteo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmos. Meas. Tech., 6, 2989–3034, https://doi.org/10.5194/amt-6-2989-2013, 2013.

Li, J., Li, X., Carlson, B. E., Kahn, R. A., Lacis, A. A., Dubovik, O., and Nakajima, T.: Reducing multisensor satellite monthly mean aerosol optical depth uncertainty: 1. Objective assessment of current AERONET locations, J. Geophys. Res.-Atmos., 121, 13609–13627, https://doi.org/10.1002/2016JD025469, 2016.

Lipponen, A., Mielonen, T., Pitkänen, M. R. A., Levy, R. C., Sawyer, V. R., Romakkanen, S., Kolehmainen, V., and Arola, A.: Bayesian aerosol retrieval algorithm for MODIS AOD retrieval over land, Atmos. Meas. Tech., 11, 1529–1547, https://doi.org/10.5194/amt-11-1529-2018, 2018.

Loew, A., W. Bell, L. Brocca, C. E. Bulgin, J. Burdanowitz, X. Calbet, R. V. Donner, D. Gruber, T. Kaminski, J. Kinzel, C. Klepp, J.-C. Lambert, G. Schaepman-Strub, M. Schröder, and T. Verhoestel: Validation practices for satellite-based Earth observation data across communities, Rev. Geophys., 55, 779–817, https://doi.org/10.1002/2017RG000562, 2017.

Loughman, R., Bhartia, P. K., Chen, Z., Xu, P., Nyaku, E., and Taha, G.: The Ozone Mapping and Profiler Suite (OMPS) Limb Profiler (LP) Version 1 aerosol extinction retrieval algorithm: theoretical basis, Atmos. Meas. Tech., 11, 2633–2651, https://doi.org/10.5194/amt-11-2633-2018, 2018.

Luffarelli, M. and Govaerts, Y.: Joint retrieval of surface reflectance and aerosol properties with continuous variation of the state variables in the solution space – Part 2: application to geostationary and polar-orbiting satellite observations, Atmos. Meas. Tech., 12, 791–809, https://doi.org/10.5194/amt-12-791-2019, 2019.

Lyapustin, A., Wang, Y., Xiong, X., Meister, G., Platnick, S., Levy, R., Franz, B., Korkin, S., Hilker, T., Tucker, J., Hall, F., Sellers, P., Wu, A., and Angal, A.: Scientific impact of MODIS C5 calibration degradation and C6+ improvements, Atmos. Meas. Tech., 7, 4353–4365, https://doi.org/10.5194/amt-7-4353-2014, 2014.

Lyapustin, A., Wang, Y., Korkin, S., and Huang, D.: MODIS Collection 6 MAIAC algorithm, Atmos. Meas. Tech., 11, 5741–5765, https://doi.org/10.5194/amt-11-5741-2018, 2018.

Martonchik, J. V., Diner, D. J., Kahn, R. A., Ackerman, T. P., Verstraete, M. M., Pinty, B., and Gordon, H. R.: Techniques for the retrieval of aerosol properties over land and ocean using multiangle imaging, IEEE T. Geosci. Remote, 36, 1212–1227, https://doi.org/10.1109/36.701027, 1998.

Martonchik, J. V., Kahn, R. A., and Diner, D. J.: Retrieval of aerosol properties over land using MISR observations, in: Aerosol Remote Sensing Over Land, edited by: Kokhanovsky, A. A. and de Leeuw, G., Springer, Berlin, 2009.
data/files/ATBD-POL-aerosol-2018.pdf (last access: 29 January 2020), 2018.
Murphy, A. H.: Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient, Mon. Weather Rev., 116, 2417–2424, https://doi.org/10.1175/1520-0493(1988)116<2417:SSBTE2.0.CO;2, 1988.
Nakajima, T. Y., Figurashi, A., Hakajima, T., Fukuda, S., and Kata-giri, S.: Development of Cloud and Aerosol Retrieval Algorithms for ADEOS-II/GLI Mission, Journal of The Remote Sensing Society of Japan, 29, 60–69, https://doi.org/10.11440/rssj.29.60, 2009.
NASA AERONET team: AERONET web portal, available at: https://aeronet.gsfc.nasa.gov, last access: 29 January 2020.
NASA Earthdata team: AERONET web portal, available at: https://earthdata.nasa.gov, last access: 29 January 2020.
North, P., Heckel, A., Davies, W., S., B., and Grey, W.: Algorithm Theoretical Basis Document (ATBD), ATSR-2 and AATSR, al-gorithm ATSR-SU, Tech. rep., ESA, available at: http://www.esa-aerosol-cci.org (last access: 29 January 2020), ESA Climate Change Initiative, version 4.3, 2017.
North, P. R., Briggs, S. A., Plummer, S. E., and Settle, J. J.: Retrieval of Land Surface Bidirectional Reflectance and Aerosol Opacity from ATSR-2 Multispectral Imagery, IEEE T. Geosci. Remote, 37, 526–537, https://doi.org/10.1109/36.739106, 1999.
North, P. R. J., Grey, W., Heckel, A., Fischer, J., Preusker, R., and Brockmann, C.: MERIS/AATSR Synergy Algorithm for Cloud Screening, Aerosol Retrieval, and Atmospheric Correction, Tech. rep., Swansea University, Swansea, UK, available at: http://ggluck.swansea.ac.uk/ftp/ATSR_SYN/synergy-aerosol-atbd-20100316-v42.doc (last access: 29 January 2020), ESRIN Contract No. 21090/07/I-LG, 2010.
O’Neill, N., Eck, T., Smirnov, A., Holben, B., and Thulasiraman, S.: Spectral Deconvolution algorithm technical memo, Tech. rep., NASA Goddard Space Flight Center, revision April 26, 2006, version 4, available at: http://aeronet.gsfc.nasa.gov/new_web/PDF/tauf_tauc_technical_memo1.pdf (last access: 29 January 2020), 2006.
O’Neill, N. T., Ignatov, A., Holben, B. N., and Eck, T. F: The lognormal distribution as a reference for reporting aerosol optical depth statistics; Empirical tests using multi-year, multi-site AERONET Sunphotometer data, Geophys. Res. Lett., 27, 3333–3336, https://doi.org/10.1029/2000GL011581, 2000.
O’Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N., and Thulasiraman, S.: Spectral discrimination of coarse and fine mode optical depth, J. Geophys. Res., 108, 4559–4573, https://doi.org/10.1029/2002JD002975, 2003.
Patadia, F., Levy, R. C., and Mattoe, S.: Correcting for trace gas absorption when retrieving aerosol optical depth from satellite observations of reflected shortwave radiation, Atmos. Meas. Tech., 11, 3205–3219, https://doi.org/10.5194/amt-11-3205-2018, 2018.
Pierangelo, C., Chédin, A., Heilliette, S., Jacquinet-Husson, N., and Armante, R.: Dust altitude and infrared optical depth from AIRS, Atmos. Chem. Phys., 4, 1813–1822, https://doi.org/10.5194/acp-4-1813-2004, 2004.
Popp, T., G de Leeuw, C, Bingen, C, Brühl, V, Capelle, A, Chedin, L, Clarisse, O, Dubovik, R, Grainger, J, Griesfeller, A, Heckel, S, Kinne, L, Klüser, M, Kossmaie, P, Kolmonen, L, Lelli, P, Litvinov, L, Mei, P, North, S, Pinnock, A, Povey, C, Robert, M, Schulz, L, Sokageva, K, Stebel, D, Stein Zweers, G, Thomas, L, G, Tilstra, S, Vandenbussche, P, Veekind, M, Vountas, and Y Xue: Development, Production and Evaluation of Aerosol Climate Data Records from European Satellite Observations (Aerosol_cci), Remote Sens., 8, 421, https://doi.org/10.3390/rs8050421, 2016.
Povey, A. C. and Grainger, R. G.: Known and unknown unknowns: uncertainty estimation in satellite remote sensing, Atmos. Meas. Tech., 8, 4699–4718, https://doi.org/10.5194/amt-8-4699-2015, 2015.
Povey, A. C. and Grainger, R. G.: Toward More Representative Gridded Satellite Products, IEEE Geosci. Remote S., 672–676, https://doi.org/10.1109/LGRS.2018.2881762, 2019.
Povey, A. C., Grainger, R. G., Peters, D. M., and Agnew, J. L.: Retrieval of aerosol backscatter, extinction, and lidar ratio from Raman lidar with optimal estimation, Atmos. Meas. Tech., 7, 757–776, https://doi.org/10.5194/amt-7-757-2014, 2014.
Rahman, H., Pinty, B., and Verstraete, M. M.: Coupled surface-atmosphere reflectance (CSAR) model, 2. Semiempirical surface model usable with NOAA Advanced Very High Resolution Radiometer Data, J. Geophys. Res., 98, 20791–20801, https://doi.org/10.1029/93JD02072, 1993.
Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, J. Atmos. Sci., 62, 947–973, 2005.
Rieger, L. A., Zawada, D. J., Bourassa, A. E., and Degenstein, D. A.: A multi-wavelength retrieval approach for improved OSIRIS aerosol extinction retrievals, J. Geophys. Res.-Atmos., 124, 7287–7407, https://doi.org/10.1029/2018JD029897, 2019.
Rodgers, C. D.: Inverse methods for atmospheric sounding: Theory and Practice, Series on Atmospheric, Oceanic and Planetary Physics-Vol. 2, World Scientific, 2000.
Sayer, A. M. and Knobelspiesse, K. D.: How should we aggregate data? Methods accounting for the numerical distributions, with an assessment of aerosol optical depth, Atmos. Chem. Phys., 19, 15023–15048, https://doi.org/10.5194/acp-19-15023-2019, 2019.
Sayer, A. M., Thomas, G. E., and Grainger, R. G.: A sea surface reflectance model for (A)ATSR, and application to aerosol retrievals, Atmos. Meas. Tech., 3, 813–838, https://doi.org/10.5194/amt-3-813-2010, 2010a.
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, Atmos. Chem. Phys., 10, 10705–10716, https://doi.org/10.5194/acp-10-10705-2010, https://doi.org/10.5194/acp-10-10705-2010, 2010b.
Sayer, A. M., Hsu, N. C., Bettenhausen, C., Ahmad, Z., Holben, B. N., Smirnov, A., Thomas, G. E., and Zhang, J.: SeaWiFS Ocean Aerosol Retrieval (SOAR): Algorithm, validation, and comparison with other data sets, J. Geophys. Res., 117, D03206, https://doi.org/10.1029/2011JD016599, 2012a.
Sayer, A. M., Hsu, N. C., Bettenhausen, C., Jeong, M.-J., Holben, B. N., and Zhang, J.: Global and regional evaluation of over-land spectral aerosol optical depth retrievals from SeaWiFS, Atmos. Meas. Tech., 5, 1761–1778, https://doi.org/10.5194/amt-5-1761-2012, 2012b.
Sayer, A. M., Thomas, G. E., Grainger, R. G., Carboni, E., Poulsen, C., and Siddans, R.: Use of MODIS-derived sur-
face reflectance data in the ORAC-AATSR aerosol retrieval algorithm: Impact of differences between sensor spectral re-
response functions, Remote Sens. Environ., 116, 177–188, https://doi.org/10.1016/j.rse.2011.02.029, 2012c.
Sayer, A. M., Hsu, N. C., Bettenhausen, C., and Jeong, M.-J.: Validation and uncertainty estimates for MODIS Collection 6 "Deep Blue" aerosol data, J. Geophys. Res., 118, 7864–7872, https://doi.org/10.1002/jgrd.50600, 2013.
Sayer, A. M., Hsu, N. C., Bettenhausen, C., Lee, J., Redemann, J., Schmid, B., and Shinozuka, Y.: Extending “Deep Blue” aerosol retrieval coverage to cases of absor-
bearing aerosols above clouds: Sensitivity analysis and first case studies, J. Geophys. Res.-Atmos., 121, 4830–4854, https://doi.org/10.1002/2015JD024729, 2016.
Sayer, A. M., Hsu, N. C., Lee, J., Carletta, N., Chen, S.-H., and Smirnov, A.: Evaluation of NASA Deep Blue/SOAR aerosol retrieval algorithms applied to AVHRR mea-
surements, J. Geophys. Res.-Atmos., 122, 9945–9967, https://doi.org/10.1002/2017JD026934, 2017.
Sayer, A. M., N. C. Hsu, J. Lee, W. V. Kim, O. Dubovik, S. T. Dutcher, D. Huang, P. Litvinov, A. Lyapustin, J. L. Tack-
et, and D. M. Winker: Validation of SOAR VIIRS over-water aerosol retrievals and context within the global satellite aerosol data record, J. Geophys. Res.-Atmos., 123, 13496–
13526, https://doi.org/10.1002/2018JD029465, 2018.
Sayer, A. M., Hsu, N. C., Lee, J., Kim, W., and Dutcher, S.: Validation, stability, and consistency of MODIS Collection 6.1 and VIIRS Version 1 Deep Blue aerosol data over land, J. Geophys. Res.-Atmos., 124, 4658–4688, https://doi.org/10.1002/2018JD029598, 2019a.
Sayer, A. M., Hsu, N. C., Lee, J., Kim, W. V., Burton, S., Fenn, M. A., Ferrare, R. A., Kacenelenbogen, M., LeBlanc, S., Pis-
tone, K., Redemann, J., Segal-Rozenhaimer, M., Shinozuka, Y., and Tsay, S.-C.: Two decades observing smoke above clouds in the south-eastern Atlantic Ocean: Deep Blue algorithm updates and validation with ORACLES field campaign data, At-
mos. Meas. Tech., 12, 3595–3627, https://doi.org/10.5194/amt-12-3595-2019, 2019b.
Schuster, G. L., Dubovik, O., and Holben, B. N.: Angstrom exponent and bimodal aerosol size distributions, J. Geophys. Res., 111, D07207, https://doi.org/10.1029/2005JD006328, 2006.
Schutgens, N., Tsyro, S., Grysynedt, E., Goto, D., Weigum, N., Schulz, M., and Stier, P.: On the spatio-temporal representa-
tive observations, Atmos. Chem. Phys., 17, 9761–9780, https://doi.org/10.5194/acp-17-9761-2017, 2017.
Schutgens, N. A. J.: Site representativity of AERONET and GAW remotely sensed AOT and AAOT observations, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-767, in review, 2019.
Schutgens, N. A. J., Grysynedt, 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, Atmos. Chem. Phys., 16, 6335–6353, https://doi.org/10.5194/acp-16-6335-2016, 2016.
Shi, C., Hashimoto, M., and Nakajima, T.: Remote sensing of aerosol properties from multi-wavelength and multi-pixel information over the ocean, Atmos. Chem. Phys., 19, 2461–2475, https://doi.org/10.5194/acp-19-2461-2019, 2019.
Shi, Y., Zhang, J., Reid, J. S., Hyer, E. J., and Hsu, N. C.: Critical evaluation of the MODIS Deep Blue aerosol optical depth product for data assimilation over North Africa, Atmos. Meas. Tech., 6, 949–969, https://doi.org/10.5194/amt-6-949-2013, 2013.
Sica, R. J. and Haefele, A.: Retrieval of temperature from a multiple-channel Rayleigh-scatter lidar using an optimal estimation method, Appl. Optics, 54, 1872–1889, https://doi.org/10.1364/OA.54.001872, 2015.
Smirnov, A., B. N. Holben, I. Slutsker, D. M. Giles, C. R. McClain, T. F. Eck, S. M. Sakerin, A. Macke, P. Coot, G. Zibordi, P. K. Quinn, J. Sciare, S. Kinne, M. Harvey, T. J. Smyth, S. Piketh, T. Zielinski, A. Proshutinski, J. I. Goes, N. B. Nelson, P. Larouche, V. F. Radionov, P. Goloub, K. Krishna Moorhly, R. Matarrrese, E. J. Robertson, and F. Jourdin: Maritime Aerosol Network as a component of Aerosol Robotic Network, J. Geophys. Res., 112, D06204, https://doi.org/10.1029/2008JD011257, 2009.
Smirnov, A., Holben, B. N., Giles, D. M., Slutsker, I., O’Neill, N. T., Eck, T. F., Macke, A., Coot, P., Courcoux, Y., Sakerin, S. M., Smyth, T. J., Zielinski, T., Zibordi, G., Goes, J. I., Harvey, M. J., Quinn, P. K., Nelson, N. B., Radionov, V. F., Duarte, C. M., Losno, R., Sciare, J., Voss, K. J., Kinne, S., Nalli, N. R., Joseph, E., Krishna Moorhly, K., Covert, D. S., Gulev, S. K., Milin,sky, G., Larouche, P., Belanger, S., Horne, E., Chin, M., Remer, L. A., Kahn, R. A., Reid, J. S., Schulz, M., Heald, C. L., Zhang, J., Lapina, K., Kleinman, R. G., Griesfeller, J., Gaiti, B. J., Tan, Q., and Diehl, T. L.: Maritime aerosol network as a component of AERONET – first results and comparison with global aerosol models and satellite retrievals, Atmos. Meas. Tech., 4, 583–597, https://doi.org/10.5194/amt-4-583-2011, 2011.
Tanré, D., Herman, M., and Kaufman, Y. J.: Information on aerosol size distribution contained in solar reflected spectral radiiances, J. Geophys. Res., 101, 19043–19060, https://doi.org/10.1029/96JD00333, 1996.
Tanré, D., Kaufman, Y. J., Herman, M., and Mattoo, S.: Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiance measurements, J. Geophys. Res., 102, 16971–
16988, https://doi.org/10.1029/96JD03437, 1997.
Thomas, G. E., Poulsen, C. A., Sayer, A. M., Marsh, S. H., Dean, S. M., Carboni, E., Siddans, R., Grainger, R. G., and Lawrence, B. N.: The GRAPE aerosol retrieval algorithm, Atmos. Meas. Tech., 2, 679–701, https://doi.org/10.5194/amt-2-679-2009, 2009.
Thomas, G. E., Poulsen, C. A., Siddans, R., Sayer, A. M., Carboni, E., Marsh, S. H., Dean, S. M., Grainger, R. G., and Lawrence, B. N.: Validation of the GRAPE single view aerosol retrieval for ATSR-2 and insights into the long term global AOD trend over the ocean, Atmos. Chem. Phys., 10, 4849–4866, https://doi.org/10.5194/acp-10-4849-2010, 2010.
Thomas, G. E., Poulsen, C. A., and Povey, A. C.: Algorithm Theoretical Basis Document (ATBD), AATSR, Oxford-RAL Aerosol and Cloud (ORAC), Tech. rep., ESA, available at: http://www.
esa-eos-aerosol-cci.org (last access: 29 January 2020), ESA Climate Change Initiative, version 3.0, 2017.
Torres, O., Bhartia, P. K., Herman, J. R., Ahmad, Z., and Gleason, J.: Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation: Theoretical basis, J. Geophys. Res., 103, 17099–17110, https://doi.org/10.1029/98JD00900, 1998.
Torres, O., Bhartia, P. K., Herman, J. R., Sinyuk, A., Ginoux, P., and Holben, B.: A Long-Term Record
of Aerosol Optical Depth from TOMS Observations and Comparison to AERONET Measurements, J. Atmos. Sci., 59, 398–413, https://doi.org/10.1175/1520-0469(2002)059<0398:ALTROA>2.0.CO;2, 2002.

Torres, O., Jethva, J., and Bhartia, P. K.: Retrieval of Aerosol Optical Depth above Clouds from OMI Observations: Sensitivity Analysis and Case Studies, J. Atmos. Sci., 69, 1037–1053, https://doi.org/10.1175/JAS-D-11-0130.1, 2012.

von Clarmann, T., Degenstein, D. A., Livesey, N. J., Bender, S., Bräuer, M., Butz, A., Compermelle, S., Damadeo, R., Dueck, S., Eriksson, P., Funke, B., Johnson, M. C., Kasai, Y., Keppens, A., Kleinert, A., Kramarova, N. A., Laeng, A., Payne, V. H., Rozanov, A., Sato, T. O., Schneider, M., Sheese, P., Sofieva, V., Stiller, G. P., von Savigny, C., and Zawada, D.: Estimating and Reporting Uncertainties in Remotely Sensed Atmospheric Composition and Temperature, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2019-350, in review, 2019.

Vandenbussche, S., Kochenova, S., Vandeaele, A. C., Kumps, N., and De Mazière, M.: Retrieval of desert dust aerosol vertical profiles from IASI measurements in the TIR atmospheric window, Atmos. Meas. Tech., 6, 2577–2591, https://doi.org/10.5194/amt-6-2577-2013, 2013.

Vanhellemont, F., Mateshvili, N., Blanot, L., Robert, C. É., Bengen, C., Sofieva, V., Dalaudier, F., Tétard, C., Fussen, D., Dekemper, E., Kyrölä, E., Laine, M., Tamminen, J., and Zehner, C.: AerGOM, an improved algorithm for stratospheric aerosol extinction retrieval from GOMOS observations – Part 1: Algorithm description, Atmos. Meas. Tech., 9, 4687–4700, https://doi.org/10.5194/amt-9-4687-2016, 2016.

Veihelmann, B., Levelt, P. F., Stammes, P., and Veefkind, J. P.: Simulation study of the aerosol information content in OMI spectral reflectance measurements, Atmos. Chem. Phys., 7, 3115–3127, https://doi.org/10.5194/acp-7-3115-2007, 2007.

Virtanen, T. H., Kolmonen, P., Sogacheva, L., Rodríguez, E., Saponaro, G., and de Leeuw, G.: Collocation mismatch uncertainties in satellite aerosol retrieval validation, Atmos. Meas. Tech., 11, 925–938, https://doi.org/10.5194/amt-11-925-2018, 2018.

Wagner, F. and Silva, A. M.: Some considerations about Ångström exponent distributions, Atmos. Chem. Phys., 8, 481–489, https://doi.org/10.5194/acp-8-481-2008, 2008.

Weaver, C. J., Joiner, J., and Ginoux, P.: Mineral aerosol contamination of TIROS Operational Vertical Sounder (TOVS) temperature and moisture retrievals, J. Geophys. Res., 108, 4246, https://doi.org/10.1029/2002JD002571, 2003.

Witek, M. L., Diner, D. J., Garay, M. J., Xu, F., Bull, M. A., and Seidel, F. C.: Improving MISR AOD Retrievals With Low-Light-Level Corrections for Veiling Light, IEEE T. Geosci. Remote, 56, 1251–1268, https://doi.org/10.1109/TGRS.2017.2727342, 2018a.

Witek, M. L., Garay, M. J., Diner, D. J., Bull, M. A., and Seidel, F. C.: New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water, Atmos. Meas. Tech., 11, 429–439, https://doi.org/10.5194/amt-11-429-2018, 2018b.

Witek, M. L., Garay, M. J., Diner, D. J., and Smirnov, A.: Oceanic aerosol loading derived from MISR’s 4.4 km (V23) Aerosol Product, J. Geophys. Res.-Atmos., 124, 10154–10174, https://doi.org/10.1029/2019JD031065, 2019.

Working Group 1: Evaluation of measurement data – Guide to the expression of uncertainty in measurement, Tech. rep., Joint Committee for Guides in Metrology, available at: http://www.iso.org/sites/JCGM/GUM-introduction.htm (last access: 29 January 2020), Tech. Rep. JCGM 100:2008, 2008.

Xu, F., Diner, D. J., Dubovik, O., and Scheckner, Y.: A Correlated Multi-Pixel Inversion Approach for Aerosol Remote Sensing, Remote Sens., 11, 746, https://doi.org/10.3390/rs11070746, 2019.

Yoshida, M., Kikuchi, M., Nagao, T. M., Muramaki, H., Nomaki, T., and Higurashi, A.: Common retrieval of aerosol properties for imaging satellite sensors, J. Meteorol. Soc. Jpn., 96B, 193–209, https://doi.org/10.2151/jmsj.2018-039, 2018.

Young, S. A. and Vaughan, M. A.: The retrieval of profiles of particulate extinction from Cloud Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) data: Algorithm description, J. Atmos. Ocean. Tech., 26, 1105–1119, https://doi.org/10.1175/2008JTECHA1221.1, 2009.

Young, S. A., Vaughan, M. A., Kuehn, R. E., and Winker, D. M.: The Retrieval of Profiles of Particulate Extinction from Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) Data: Uncertainty and Error Sensitivity Analyses, J. Atmos. Ocean. Tech., 30, 395–428, https://doi.org/10.1175/JTECH-D-12-00046.1, 2013.

Young, S. A., Vaughan, M. A., Garnier, A., Tackett, J. L., Lambeth, J. D., and Powell, K. A.: Extinction and optical depth retrievals for CALIPSO's Version 4 data release, Atmos. Meas. Tech., 11, 5701–5727, https://doi.org/10.5194/amt-11-5701-2018, 2018.

Zhang, J. and Reid, J. S.: MODIS Aerosol Product Analysis for Data Assimilation: Assessment of Over-Ocean level 2 Aerosol Optical Thickness Retrievals, J. Geophys. Res., 111, D22207, https://doi.org/10.1029/2005JD006898, 2006.

Zhao, T. X.-P., Chan, P. K., and Heidinger, A. K.: A global survey of Angström exponent distributions, Atmos. Chem. Phys., 13, 373–404, 2020 www.atmos-meas-tech.net/13/373/2020/