ESA Climate Change Initiative
aerosol_cci

Comprehensive error characterization report (CECR)
Version 3.2 (D1.3b)

Document reference:

Aerosol_cci2_CECR_v3.2.pdf

DOCUMENT STATUS SHEET

	FUNCTION	NAME	DATE	SIGNATURE
LEAD AUTHOR	Editor, author	Adam Povey	26/11/2015	
CONTRIBUTING AUTHORS	Co-authors	Christine Bingen Virginie Capelle Lieven Clarisse Don Grainger Andreas Heckel Thomas Popp Lars Klueser Pekka Kolmonen Pavel Litvinov Peter North Charles Robert Larisa Sogacheva Kerstin Stebel Gareth Thomas Sophie Vandenbussche	16/11/2015 17/11/2015 10/11/2015 07/11/2015 19/11/2015 10/11/2015 13/11/2015 19/11/2015 17/11/2015 19/11/2015 16/11/2015 19/11/2015 25/11/2015 26/11/2013 16/11/2015 15/07/2017	
REVIEWED BY	Co-science leader review	Gerrit de Leeuw Thomas Popp	30/11/2015 27/07/2017 17/08/2017	
APPROVED BY	Technical officer (ESA)	Simon Pinnock		
ISSUED BY	Project manager	Thomas Popp		

EXECUTIVE SUMMARY

This document summarizes the status of error characterization at the end of the aerosol_cci project. Error characterization means the identification of different sources of error, assessment of their behaviour, and calculation of the sensitivity of retrieval algorithms to each source.

An algorithm theoretical baseline document (ATBD) has been prepared for each algorithm involved in the aerosol_cci project, which details (amongst other things) the propagation and treatment of uncertainties. This report summarises the major principles and knowledge contained within those documents to compare the techniques used and contextualise them within the field of metrology. These uncertainty estimates are based on the developer's understanding of their retrieval and its sensitivity to the environment, with significant consideration given to the pre-launch calibration of each sensor. The differences between the various products and between their versions provide some insight into the nature of currently unquantified uncertainties.

Just as it is important to validate a data product to demonstrate that it is fit-for-purpose, it is necessary to validate the estimates of uncertainty to demonstrate that they represent the distribution of error. A procedure to do so by comparison against AERONET is outlined and results are presented for the three ATSR algorithms. These show that the uncertainty is partly underestimated (except for the SU ATSR algorithm, which is close to the optimal representation and except the ORAC algorithm over land, which has some overestimation). Overall, the values are better estimated over land than at coastal sites (which is understandable due to the complexity of coastal waters). The ATSR ensemble has some overestimation which is due to the dominating largest value of one of the contributing algorithms. For all algorithms, the uncertainties are reasonably stable over time and also in the overlap of the two sensors ATSR-2 and AATSR.

This document consists of 7 sections. After an introduction, it summarises the terminology of error characterisation (which is harmonized between all CCI projects). The sources of error are then generalised, followed by an outline of the uncertainty estimation techniques used in each of the algorithms in the project. Validation procedures for the products' uncertainties are presented for both Level 2 and 3 data. Some advice from data producers on appropriate use of aerosol data is provided, leading to final conclusions.

Issue	Date	Modified Items / Reason for Change
1	09.11.2015	Document outline, with sections 1-3 drafted
2.0	18.11.2015	Partner submissions integrated
2.1	26.11.2015	Final version prepared for submission reviewed by science leader (solely layout changes)
3.0	15.07.2017	Updates for algorithms where error propagation was revised (ORAC, AERGOM) and of the compelte section 5.2 (lv2 validation)
3.1	27.07.2017	science leader review (update of executive summary and conclusion, layout corrections)
3.2	17.08.2017	Science leader corrections in response to RIDs from ESA (support by K. Stebel)

TABLE OF CONTENTS

1 INTRODUCTION

This report aims to review the methods of uncertainty characterization used by the various algorithms within the aerosol_cci project. It will summarise the known sources of error in the aerosol products, overview the methods by which the resulting uncertainty is estimated, and present an initial validation of those methods by comparison against AERONET (ground-based sun photometer) observations.

The scope of this report is to provide an overview of the management of uncertainty, not to replicate the ATBDs for the algorithms. The report has been and will be updated regularly to represent the results of validation activities and sensitivity studies produced during the project.

1.1 References

1.1.1 Applicable Documents

[AD1] The Statement of Work, reference CCI-PRGM-EOPS-SW-12-0012, issue 1, revision 2, dated June 7th, 2013, and its specific annex C.

[AD2] The Contractor's Proposal reference 3010317 revision 1.2 , dated 04 April 2014

1.1.2 Reference Documents

[RD1] ATBD for FMI's ADV algorithm, v4.1, dated 21.04.2017.

[RD2] ATBD for Oxford and RAL's ORAC algorithm, v3.0, dated 07.04.2017.

[RD3] ATBD for SU's AATSR algorithm, v4.3, dated 15.05.2017.

[RD4] ATBD for LMD IASI algorithm, v2.1, dated 18.04.2017.

[RD5] ATBD for IMARS IASI algorithm, v.2.0, dated 16.12.2016.

[RD6] ATBD for ULB IASI algorithm, v1.7, dated 27.03.2017.

[RD7] ATBD for MAPIR IASI algorithm, v3.5, dated 01.03.2017.

[RD8] ATBD for GOMOS AERGOM algorithm, v5.0, dated 30.03.2017.

[RD9] Phase 2 Product Validation and Inter-comparison Report, version 3.31, draft 20.06.2017.

1.1.3 Academic References

Anderson, T.L. R.J. Charlson, D.M. Winker, J.A. Ogren and K. Holmén: Mesoscale variations of tropospheric aerosols, J. Atmos. Sci., 60, p.119-136, doi:10.1175/1520-0469(2003)060<0119;MVOTA>2.0.CO;2, 2002.

August, T., D. Klaes, P. Schlüssel, T. Hullberg, M. Crapeau, A. Arrisaga, A. O'Carroll, D. Coppens, R. Munro, and X. Calbet: IASI on Metop-A: Operational Level 2 retrievals after five years in orbit, J. Quant. Spectrosc. Radiat. Transfer, 113, p.1340-1371, doi:10.1016/jqsrt.2012.02.028, 2012.

Beers, Y.: Introduction to the theory of error, Addison-Wesley, Massachusetts, 1957.

Bevan, S., North, P., Los, S. & Grey, W. (2012). A global dataset of atmospheric aerosol optical depth and surface reflectance from AATSR. *Remote Sensing of Environment* 116, 199-210.

BIPM: Guide to the Expression of Uncertainty in Measurement (GUM), Bureau International des Poids et Mesures, http://www.bipm.org/en/publications/guides/gum.html, 2008.

Holzer-Popp T., G. de Leeuw, D. Martynenko, P. North, P. Kulmonen, L. Sogachewa, G. Thomas, D. Graigner, A. Heckel, W. von Hoyningen-Hüne, D. Ramon, F.-M. Breon, D. Tanre, S. Kinne, M. Schulz, J. Griesfeller, J. Decloitres, L. Klüser: Aerosol retrieval experiments in the ESA Aerosol_cci project, Atmos. Meas. Tech., 6, 1919–1957, doi:10.5194/amt-6-1919-2013, 2013.

Hughes, I.G. and T.P.A. Hase: Measurements and their uncertainties, Oxford University Press, Oxford, 2010.

North, P., Briggs, S., Plummer, S. & Settle, J. (1999). Retrieval of land surface bidirectional reflectance and aerosol opacity from ATSR-2 multiangle imagery. *IEEE Transactions on Geoscience and Remote Sensing* 37(1), 526-537.

Pierangelo, C., A. Chédin, S. Heilliette, N. Jacquinet-Husson and R. Armante: Dust altitude and infrared optical depth from AIRS, Atmos. Chem. Phys., 4, p.1813-1822, doi:10.5194/acp-4-1813-2004, 2004.

Pougatchev, N., T. August, X. Calber, T. Hultber, O. Oduleye, P. Schlüssel, B. Stiller, K. St Germain and G. Bingham: IASI temperature and water vapour retrievals – error assessment and validation, Atmos. Chem. Phys., 9, p.6453-5458, doi:10.5194/acp-9-6453-2009, 2009.

Povey, A.C. and R.G. Grainger: Known and unknown unknowns: uncertainty estimation in satellite remote sensing, Atmos. Meas. Tech., 8, 4699–4718, 2015.

Sayer, A.M., G.E. Thomas, and R.G. Grainger: A sea surface reflectance model for (A)ATSR, and application to aerosol retrievals, Atmos. Meas. Tech., 3(4):813–838, doi:10.5194/amt-6-813-2010, 2010.

Sofieva, V., Kan, V., and Dalaudier, F.: Influence of scintillation on quality of ozone monitoring by GOMOS, Atmos. Chem. Phys., 9, 9197–9207, doi:10.5194/acp-9-9197-2009, 2009.

Tamminen, J., Kyrölä, E., Sofieva, V. F., Laine, M., Bertaux, J. L., Hauchecorne, A., Dalaudier, F., D. Fussen, F. Vanhellemont, O. Fanton-d'Andon, G. Battor, A. Mangin, M.Guiriet, L. Blanot, T. Fehr, L. Saavedra de Miguel and R. Fraisse: GOMOS data characterisation and error estimation, Atmos. Chem. Phys., 10(19), 9505–9519, doi:10.5194/acp-10-9505-2010, 2010.

1.2 List of acronyms

AAI	absorbing aerosol index		LUT	look up table
(A)ATSR	(advanced) along track scanning radiometer		MAPIR	mineral aerosol profiling from thermal infrared radiances
ADV	ATSR dual view		ORAC	optimal retrieval of aerosol and cloud
AERONET	aerosol robotic network		pdf	probability density function
AOD	aerosol optical depth		rmse	root mean square error
ATBD	algorithm theoretical basis document		SCIAMACHY	scanning imaging absorption spectrometer for atmospheric cartography
BRDF	bidirectional reflectance function		SNR	signal to noise ratio
CCI	climate change initiative		SSA	single scattering albedo
ECMWF	European centre for mid-range weather forecasting		SU	Swansea University
ECV	environmental climate variable		TIR	thermal infrared
GOMOS	global ozone monitoring by occultation of stars		TOA	top of atmosphere
IASI	infrared atmospheric sounding interferometer		UV/VIS	ultraviolet/visible (parts of the spectrum)
IMARS	infrared mineral aerosol retrieval scheme			

2 DEFINITION OF TERMS

2.1 Describing error and uncertainty

A **measurement** is a set of operations intended to determine the value of a quantity. Following BIPM (2008), it is useful to define the term **measurand** as the particular quantity subject to measurement such that the phrases "true value of a quantity" and "value of the measurand" are synonymous.

Very few instruments directly measure the measurand. An instrument generally reports a quantity from which the magnitude of the measurand is estimated (e.g. an instrument sensitive to infrared light might be used to measure the temperature of an object). The process of measurement is intrinsically inexact. The difference between a measured value and the value of the measurand is called the error. Traditionally (e.g. Beers, 1975), the word "error" has also meant a numerical value that estimates the variability of the error if a measurement is repeated (i.e. a width of the distribution of possible errors). To avoid this ambiguity, the CCI program has adopted the BIPM (2008) definitions

- **error (of measurement)**: the result of a measurement minus a true value of the measurand;
- **uncertainty (of measurement)**: a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand.

The "true" value of the error is rarely known such that its magnitude is hypothetical. Error is frequently viewed as having a random and a systematic component, defined by BIPM (2008) as

- **random error**: result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions;
- **systematic error**: the mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions minus the true value of the measurand.

A more detailed discussion of these concepts applied to satellite remote sensing can be found in Povey and Grainger (2015).

Two qualitative terms not defined in BIPM (2008) but commonly used to describe a measurement (e.g. Beers, 1957, Hughes and Hase, 2010) are precision and accuracy, defined here as

- **precision**: qualitative measure of the (relative) magnitude of the random uncertainty;
- **accuracy**: qualitative measure of the (relative) magnitude of the systematic uncertainty.

Although it is not possible to compensate for random error, the resulting uncertainty in our estimate of the measurand can usually be reduced by averaging repeated, independent observations. The statistical distribution of random error can be described by a probability density function (pdf) for which the **expected value** (i.e. the average over the pdf) is zero. As the random error often arises from the addition of many effects, the central limit theorem suggests that a Gaussian distribution should be a good representation of the pdf. Hence, the uncertainty resulting from random error is commonly represented by the one-sigma standard deviation that would be obtained from repeated measurements of the same quantity under the same conditions. If N uncorrelated observations are available, the random component of uncertainty is their one-sigma standard deviation multiplied by a factor of $1/\sqrt{N}$. The smallest possible change in value that can be observed can be taken as half the uncertainty (which can also be used as the detection limit of the instrument).

In some circumstance, a correction can be applied to compensate for systematic errors. Afterwards, the expected value of the error is assumed to be zero (i.e. the correction leaves only random errors), but there will be random and systematic errors in the correction itself.

An **error budget** is a summary of the known sources of error in a measurement with estimates of their resulting uncertainty (and, preferably, information on how those uncertainties combine). Standard methods of error propagation (e.g. Hughes and Hase, 2010) can be used to transform uncertainties into measurement units. The total uncertainty is the combined total accounting for any correlation between component errors.

When multiple measurands are estimated simultaneously, their uncertainty may not be independent. Their mutual uncertainty is represented with a covariance matrix $\mathbf{S}_{ij} = \langle \sigma_i \sigma_j \rangle$, where each term is the expected value of the product of the uncertainties σ_i of the ith and jth measurands. If the measurands are independent then the off-diagonal terms are zero and the uncertainty on each measurand is given by the square-root of the corresponding diagonal element.

2.2 Validation of Measurements

Validation is the assessment of a measurement and its uncertainty. This is principally achieved by **external validation**, the comparison of a measurement to an independent measurement (from a different instrument). The independent estimate of the measurand is termed the **validation value**. The **discrepancy** is defined as the difference between the measurement and the validation value. A small average discrepancy (e.g. the root-mean-square) between the measurement and validation value is indicative of an accurate measurement but could also result from a fortuitous cancellation of error terms. The uncertainty is assessed by its ability to characterise the observed distribution of discrepancies.

It is only practical to report individual discrepancies for small data sets. For the large number of measurements typical in satellite remote sensing, validation involves statistically characterising the discrepancies. The behaviour of the instrument (or algorithm) is often expected to be a function of the conditions observed, so it is typical to characterize discrepancies separately over a number of "regimes" (e.g. land and sea). The choice of regimes could come from a cluster analysis of discrepancy (if the difference in regimes causes differences in systematic error) but more commonly comes from knowledge of the measurement process.

Consider a set of n measurements $\{x_1 \pm \delta x_1, x_2 \pm \delta x_2, x_3 \pm \delta x_3, \ldots x_n \pm \delta x_n\}$ together with a set of validation values $\{v_1 \pm \delta v_1, v_2 \pm \delta v_2, v_3 \pm \delta v_3, \ldots v_n \pm \delta v_n\}$. The statistical characterization of the discrepancies within a regime is made through three **quality parameters**.

- **Bias,** b, the mean value of the discrepancy,

  $$b = [\, \Sigma_{i=1}^{n} x_i - v_i \,] / n \, .$$

  The expectation value of the bias is the combination of systematic errors in the measurement and in the validation value. The bias can only be attributed to the measurement if the systematic error in the validation value is known. Ideally, the bias would be zero.

- **Chi-squared**, χ^2**,** the goodness of fit between the actual and estimated uncertainties of measurement and validation values,

  $$\chi^2 = [\, \Sigma_{i=1}^{n} (x_i - v_i)^2 / (\delta x_i^2 + \delta v_i^2) \,] / n \, .$$

  The expected value for χ^2 is unity. A value lower than this indicates the uncertainties attributed to the measurements or the validation values (or both) are too high. A value greater than unity indicates the uncertainties are too low.

- **Stability**, *s*, the change in bias with time,

$$s = [\, b(t + \Delta t) - b(t) \,] / \Delta t \,.$$

  Ideally, the stability would be zero over any timescale. In remote sensing the stability can display periodicity related to factors, such as instrument drift or solar illumination of the satellite (both over an orbit and seasonally). It is suggested that the stability is estimated at the same temporal scale that any trends in the data are calculated.

It may be that the quality parameters are independent of the observed conditions of measurement and apply at all locations and times. Then, the three quality values adequately characterize the quality of measurement. More commonly, the quality values vary so that a **validation table** is used to summarise the bias, χ^2 and stability over various regimes.

In some cases, **internal validation**, a comparison of measurements from a single instrument, can be used to assess the uncertainty. For example, repeated observations of an unchanged target should sample the distribution of error.

2.3 Comparing Measurements with a Model

Further understanding can be achieved through comparison of measurements with model output. A model field is sampled as if viewed by a satellite and the same quality parameters are calculated. However,

- the uncertainty in the model may not be reported and so has to be assumed, and
- the bias cannot be attributed to the model or measurements without reference to additional information.

If the model evaluates substantially different scales to the satellite an estimate of uncertainty due to interpolation must be included. If the model is at a coarser resolution than the measurements an approach could be to compare the model value with a (weighted) average of the measurements. In that case, correlations in the systematic uncertainty need to be considered.

The statistical comparison of model and measurement data must account for the influence of sampling. For example, the comparison of monthly time series from model output and averaged measurements may show discrepancies due to a lack of observations in certain regions, such as those with persistent cloud coverage.

3 SOURCES OF UNCERTAINTIES

3.1 Classification

Despite their extensive use, the classification of errors as random or systematic is limited. A random error can appear to introduce a systematic bias after propagation through a non-linear equation due to its asymmetric distribution, and the distribution of a systematic error has finite width. The use of these terms is better understood as synonyms for the non-technical meanings of noise and bias, respectively.

Regardless, users have an interest in the causes of uncertainty in a measurement. The source of an error affects how it is realised and its relative importance. Povey and Grainger (2015) proposed five classifications of error by their source.

- **Measurement errors**: These result from statistical variation in the measurand or random fluctuations in the detector and electronics. To assess these accurately requires the comparison of the instrument to a thoroughly characterised reference. The response may evolve over time, necessitating the periodic repeat of calibration procedures.
- **Parameter errors**: Retrievals use auxiliary information to constrain features of the environment not easily determined from the measurements. Parameters will be produced by an independent retrieval and have associated uncertainties that propagate into the results.
- **Resolution errors**: Aerosol is a continuous field in space and time, but satellite observations only sample it and aren't necessarily representative. Filtering procedures (for quality control) can further limit the sample. As aerosol retrievals are only performed in cloud-free conditions, the concept is also known as "fair-weather bias". Filtering can also remove exceptional events, as high aerosol optical depth (AOD) plumes often fail cloud clearing, producing a low bias in averages.
- **Approximation errors**: It is not always practical to evaluate the most precise formulation of a forward model. For example, the atmosphere may be approximated as plane parallel or look-up tables (LUTs) may be used rather than solving the equations of radiative transfer. Such approximations will introduce error, which can be assessed by comparing the performance of the rigorous and simplified forward models through simulated data.
- **System errors**: It is also not always possible to constrain every aspect of the environment with the available information. In this project, the type of aerosol observed is an important example. These properties are assumed and errors result from their inaccuracy. The errors are a non-linear function of the observed state and are known to be a significant source of uncertainty, but that uncertainty cannot be readily quantified.

Measurement and parameter errors are both intrinsic sources of uncertainty. Measurement errors affect the quantities measured and analysed by the retrieval. Parameter errors are propagated from auxiliary inputs, such as meteorological data or empirical constants. Resolution errors result from finite sampling of a constantly varying system. These can be important as satellites do not sample randomly but with a systematic bias due to the satellite's orbit and quality control or filtering. Approximation errors represent aspects of the analysis that could have been done more precisely but do not affect the fundamental measurand. System errors express choices in the analysis that alter the measurand. The system error results from the difference between the assumed system and reality.

3.2 Qualitative description of major sources of uncertainties

The following table provides an overview of the common sources of error across the various algorithms in the project, based on the description of the algorithms in their ATBDs [RD1-3].

Table 3-1: *Qualitative error budget for aerosol retrievals using nadir observations.*

Source of uncertainty	Description	Qualitative estimate of contribution
Cloud screening and safety zone	Capabilities depend on available spectral range (e.g. thermal bands are important); safety zone also masks elevated AOD around clouds	High for UV/VIS sensors, medium for stratospheric algorithms
Overpass time	Polar orbiting sensors provide typically one or two sun-synchronous overpass times per day	High when comparing to different sensors or against models
Land surface reflectance (BRDF)	Can be estimated from vegetation index and/or mid-infrared bands, drawn from a climatology or ECV, or retrieved alongside AOD from multi-view data	High for nadir-only sensors, with larger uncertainty at higher reflectances
Ocean surface reflectance	Estimated using white caps parameterisation and possibly a climatology of ocean colour	Medium
Calibration	Absolute radiance calibration is critical with spectral calibration being less critical due to the broad-band features considered	Medium
Aerosol optical properties	This includes spectral extinction, absorption, phase function and shape (degree of sphericity)	Medium to high for sensors with low information content, low for AOD < 0.15
Vertical aerosol profile	Different assumptions are made for different aerosol types but sensitivity at TOA is small for VIS/IR sensors, increasing in the TIR	Medium for UV observations and absorbing aerosol, low otherwise
Directional reflectance ratio	Ratio between nadir and forward views is transferred from mid-infrared to visible bands	Medium for multi-view sensors
Pixel size	Ranges from 1x1 km^2 for radiometers to 16x7 km^2 for polarization instruments to approximately 0.25x0.5° for spectrometers	Medium when pixels dimension approach 50 km (approximate scale of aerosol variation)
Temperature vertical profiles	Usually of very high accuracy and precision, but might be significantly affected by the presence of high absorbing aerosol load	Low to medium (only for TIR sensors)
Trace gas concentration profiles	Critical absorption bands are usually avoided	Low
Radiative transfer forward model	Typical accuracy < 1%	Low
Look-up table discretization	Uncertainty often a function of the number of discretization points	Low
Wind speed	Used to estimate ocean reflectance	Low
Sampling	Practically all sensors under-sample the aerosol fields in time; different samplings lead to bias between different products	Depends strongly on the repeat cycle of the sensor and its swath width
Aggregation to 10x10 km^2	Aims to improve the signal-to-noise ratio and exclude outliers	Reduces random error (but not systematic) and may decrease representivity of data

4 METHODS OF DETERMINING UNCERTAINTIES AT L2

In response to user requirements, all aerosol_cci products include a pixel-level assessment of the uncertainty. There are several approaches to evaluating such values.

- Analytical calculations, such as the traditional propagation of errors via Jacobians, assume that the errors are normally distributed, that the uncertainties on the retrieval inputs are accurately quantified, and that the formulation of the algorithm is physically valid (or, at least, that the distribution of error resulting from the formulation is known).
- Validation activities, such as comparison against the AERONET sun photometer network, can approximate the error in a measurement by the discrepancy between the retrieved and the validation values. This can be a useful investigation of uncertainty where little is known about the distribution of error but does not provide an independent estimate of the uncertainty.
- Theoretical information content analysis can be used when neither of the previous options is possible, such as before the launch of the satellite. Significant differences between pre-launch expectations and post-launch results could indicate inaccurate assumptions in the algorithm.
- Ensembles of different retrieval algorithms or different parameters may be useful in characterising errors that cannot be quantified directly. Aerosol type can be assessed in this manner, by attempting retrieval on one measurement with several types.
- Quality flags provide the user with qualitative uncertainty information, indicating where the formulation of the algorithm may not be valid (such that analytical calculations are not necessarily meaningful).

4.1 AOD from ATSR algorithms

Three algorithms are used within the aerosol_cci project to produce the AOD ECV. Their techniques for determining the pixel-level uncertainty differ and are briefly summarised here. (For a more exhaustive description, see their respective ATBDs. For descriptions of quality flags, see the user guides.) An attempt to homogenise (certain aspects of) the techniques was conducted, its results are assessed in section 5.2 .

4.1.1 Analytical uncertainty propagation in ORAC

The optimal estimation framework of the ORAC algorithm provides rigorous uncertainty propagation. A known approximation in the forward model can be accounted for within the measurement covariance matrix (as a forward model error). Uncertainty in parameters used by the forward model (e.g. the height distribution of aerosol) can also be included (as a parameter error).

The validity of the optimal estimation approach, and particularly the uncertainty on retrieved parameters derived using it, is dependent on the accurate characterisation of the measurement and *a priori* uncertainties. The measurement covariance matrix is based on the pre-launch characterisation of the AATSR channel accuracy, with additional terms to account for uncertainties in the ratios between the different surface reflectance terms as well as error from interpolation of the LUTs. These uncertainties are listed in Table 4-1.

In versions less than 4.0 (when retrieving superpixels), the reflectance uncertainties are divided by the square root of the number of individual observations included in each 10 km sinusoidal bin to give the standard error. If this value is found to be below the digitisation limit of AATSR, the limiting value is used instead. The interpolation and parameter errors are then added in quadrature, giving a diagonal

measurement covariance matrix. The uncertainty of the product is calculated from the diagonal of the retrieved covariance matrix after the final iteration.

In version 4.0 and greater (when retrieving at native resolution), the uncertainties are averaged in the same manner as the retrieved values, providing various statistical parameters for each sinusoidal bin.

The *a priori* and corresponding covariance matrices are set separately for each retrieved parameter. All retrieved parameters are assumed uncorrelated *a priori.*

- The *a priori* AOD is 0.1 but only weakly constrains the retrieval as its variance is 1.5. (As this is retrieved in $\log_{10}$ space, this is equivalent to one standard deviation bounds of 0.003 to 3.2.)
- The *a priori* effective radius is set to the literature value for the aerosol type (ranging from 0.91 to 1.2 µm) and there is a tighter constraint applied, with a variance of 0.15 on $\log_{10}(r_e)$.
- The *a priori* surface reflectance is set over ocean by the sea surface reflectance model derived by Sayer et al. (2010). To determine its variance, they considered an ensemble of evaluations of the model, with the input parameters perturbed by their estimated uncertainties. Based on that, an uncertainty of 5% is assumed for all channels.
- The surface reflectance over land is retrieved using the ORAC-DEV model. The *p* parameters have an *a priori* of 0.3 with a variance of 0.5 for the forward view and 0.01 for the nadir view (as this simply acts as a scaling factor). The *s* parameters have an *a priori* of 0.1 with variance 1.0.

Table 4-1: *Uncertainties applied to the shortwave ATSR-series channels in ORAC.*

Wavelength (µm)	**0.55**	**0.67**	**0.87**	**1.6**
Relative reflectance error	2.4%	3.2%	2.0%	3.3%
Minimum absolute error	0.0005	0.0003	0.0003	0.0003
Interpolation error	0.81%	0.67%	0.66%	0.68%
Parameter error (land, nadir)	1.61%	2.25%	2.97%	3.71%
Parameter error (land, for.)	1.19%	1.75%	2.79%	3.45%
Parameter error (sea, nadir)	2.00%	2.36%	2.63%	4.61%
Parameter error (sea, for.)	1.32%	1.50%	1.61%	2.94%

4.1.2 Error propagation approach in AATSR SU

While the original CCI baseline dataset (Bevan et al., 2012) used an empirical error approximation, within aerosol_cci this has been developed to give full analytic propagation of error for each retrieval. Over both land and ocean, the retrieval uses non-linear optimisation of an error function, of the form

$$X^2 = \sum_{\lambda=1}^{\lambda=4} \sum_{\Omega=0}^{\Omega=55} \frac{(M(\lambda,\Omega) - O(\lambda,\theta))^2}{\sigma^2_{M(\lambda,\Omega)} + \sigma^2_{O(\lambda,\Omega)}},$$

where $\sigma_{M(\lambda,\Omega)}$ and $\sigma_{O(\lambda,\Omega)}$ denote estimates of 1 s.d. uncertainty in model and observation of surface reflectance at waveband λ and view direction θ (nadir is here denoted by θ=0°, forward view by θ =55°). It is possible to also include the full covariance matrix into the X^2 formulation, but currently error in model and observations are approximated as uncorrelated between channels.

For correctly normalised value of chi sq, the estimate of 1 s.d. error in τ_{550} is derived from the second derivative (curvature) of the error surface near the optimal value:

$$\sigma_{\tau 550} = \left(\partial^2 X^2 \Big/ \partial^2 \tau_{550} \right)^{-0.5}.$$

The curvature term is estimated by a parabolic fit of the error function for surrounding values of τ_{550}.

4.1.2.1 Model uncertainty

Over land, model uncertainty was evaluated by inversion (against the test dataset) of surface BRDF values computed from a 3D Monte Carlo model. The 3D model and test dataset are described in North (1996) and North et al. (1999). Per channel errors in reflectance were estimated from this fit as,

$$\alpha = \{0.01, 0.01, 0.04, 0.02\}.$$

No significant angular dependence of model error was found.

Over ocean the surface model is based on the models of Cox and Munk (1954) for glint, Monahan & O'Muircheartaigh (1980) and Koepke (1984) for foam fraction and spectral reflectance, and Morel's case I water reflectance model. Inputs required *a priori* are surface wind speed W (ms^{-1}) and pigment concentration C ($mg\ m^{-3}$).

Uncertainty in ocean reflectance model $M_{ocean}(\lambda,\Omega,W,C)$ is given by

$$\sigma^2_{M\_ocean} = \sigma^2_{M\_ocean\_W} + \sigma^2_{M\_ocean\_C},$$

where

$$\sigma_{M\_ocean\_W} = \left| M_{ocean}(\lambda,\Omega,W+\sigma_W,C) - M_{ocean}(\lambda,\Omega,W,C) \right|$$

$$\sigma_{M\_ocean\_W} = \left| M_{ocean}(\lambda,\Omega,W,C+\sigma_C) - M_{ocean}(\lambda,\Omega,W,C) \right|.$$

Uncertainties in wind speed and pigment concentration are assigned values of 3 ms^{-1} and 0.1 $mg\ m^{-3}$

4.1.2.2 Observation errors

The per-channel observation error gives an estimate of the 1 s.d. uncertainty in derived land surface reflectance, and includes errors due to instrument calibration, radiative transfer model and LUT, and uncertainty in aerosol absorption parameterization.

$$\sigma^2_O = \sigma^2_{RT} + \sigma^2_{inst} + \sigma^2_{AerMod}.$$

Approximations for these are given by:

$$\sigma^2_{inst} = T_S(\lambda,\theta)(a_\lambda + b_\lambda R_{TOA}(\lambda,\theta)),$$

where

$$a = \{0.0005, 0.0003, 0.0003, 0.0003\}$$

$$b = \{0.024, 0.032, 0.02, 0.033\}.$$

Currently the *a* term is neglected. The term Ts gives scaling from TOA to surface at

$$T_S(\lambda,\Omega) = \frac{\delta R_{SURF}(\lambda,\Omega)}{\delta R_{TOA}(\lambda,\Omega)},$$

and is derived from the LUT coefficients at time of inversion

Based on Kotchenova and Vermote (2007), error due to RT at all channels is approximated as

$$\sigma_{RT}^2 = 0.006.$$

The error due to uncertainty in aerosol absorption is approximated by

$$\sigma_{AerMod} = 0.05 P_R(\lambda,\Omega).$$

where P_R denotes atmospheric path radiance, estimated from LUT values.

4.1.3 Calculation via Jacobians in ADV

The ADV aerosol retrieval is based on the determination of aerosol model by comparing the measured reflectance to that modelled. The uncertainties of the retrieved aerosol model decision parameters are determined. These errors are used to determine the uncertainty in the retrieved AOD. Other possible sources of error arise from the modelling, including uncertainty in the aerosol model selection (fine mode fraction, absorbing/non-absorbing fine particle mixture, dust fraction), LUT interpolation errors, and radiative transfer computation errors.

Formal treatment in the land retrieval (ADV algorithm) is based on the general equation in the inverse problem treatment. First, denote the parameters in the least squares problem of the ADV (see [RD1]) by

$$\mathbf{x} = \{b1, b2, L, \mathbf{r}\}, \quad (1)$$

and the problem equations by

$$f_i(x) = \frac{\rho^n(\lambda_i) - \rho^n(b_1,b_2,L,\lambda_i)}{T^n(b_1,b_2,L,\lambda_i)} - \frac{\rho^f(\lambda_i) - \rho^f(b_1,b_2,L,\lambda_i)}{kT^f(b_1,b_2,L,\lambda_i)} \quad (2)$$

where $\mathbf{r} = \{\rho^n(\lambda_i), \rho^f(\lambda_i)\}\ \forall\, i \in (1,3)$ is the measured nadir and forward reflectance, $b1$ is the fine mode fraction, $b2$ is the absorbing/non-absorbing fine particle mixture, and L is the LUT AOD level. Index i refers to the wavelengths; $i = \{1,2,3\}$. Note that the dust fraction is not retrieved; it is taken from an AEROCOM climatology. This formulation of the problem enables the determination of uncertainty for the decision parameters based on the measurement error. The formulation could take into account the effect of *a priori* information for $b1$, $b2$ and L but this is neglected as the error is assumed to come only from the measurements.

The least squares problem can be solved using a quasi-Newton method. The maximum likelihood point can be found using the iteration

$$\mathbf{x}_{n+1} = \mathbf{S}_X \mathbf{F}_n^T (\mathbf{S}_T + \mathbf{F}_n \mathbf{S}_X \mathbf{F}_n^T)^{-1} \boldsymbol{f}(\mathbf{x}_n),$$

where $\mathbf{F}$ is the gradient of $\boldsymbol{f}$, the equation that is minimised, at $\mathbf{x}_n$. For the final solution $\mathbf{x}_\infty$, the *a posteriori* covariance is

$$\mathbf{S}_{X'} = (\mathbf{F}_\infty^T \mathbf{S}_T^{-1} \mathbf{F}_\infty + \mathbf{S}_X^{-1})^{-1}, \quad (5)$$

where x_∞ is the solution of the minimized equation (3). Note that even though the ADV solution is not computed using the iteration scheme above, it is still useful for determining the *a posteriori* covariance.

The Jacobian matrix $\mathbf{F}$ is of the form

$$\mathbf{F}=\begin{pmatrix} \frac{\partial f_1}{\partial b_1} & \frac{\partial f_1}{\partial b_2} & \frac{\partial f_1}{\partial L} \\ \frac{\partial f_2}{\partial b_1} & \frac{\partial f_2}{\partial b_2} & \frac{\partial f_2}{\partial L} \\ \frac{\partial f_3}{\partial b_1} & \frac{\partial f_3}{\partial b_2} & \frac{\partial f_3}{\partial L} \end{pmatrix}. \quad (6)$$

All the partial derivatives are computed numerically as the evaluation of these values requires interpolation from the aerosol LUTs, for which analytical differentiation is inaccurate.

The covariance $\mathbf{S}_T$ considers only measurement errors. For AATSR this error is taken to be 5 % of the measured signal for each channel. The principal difficulty is that in equation (3) there are two measured values $\rho^n(\lambda_i)$ and $\rho^f(\lambda_i)$. The formulation in equation (5), however, takes account of the uncertainty of only one value in the covariance matrix. The solution is that as the nadir and forward relative measurement error is set equal to the larger absolute value is used. It would be useful to study individually the effect of the nadir and forward measurement error on the aerosol model parameters in future. In addition, when all errors are considered to be Gaussian in nature, modeling errors could be simply added to the measurement errors. Another simplification made is that measurement errors do not correlate. Thus, $\mathbf{S}_T$ is diagonal. This assumption does not hold true for the *a posteriori* covariance $\mathbf{S}_{X'}$. The uncertainty in the aerosol model parameters will be correlated.

The *a priori* covariance matrix for the aerosol model defining parameters $\mathbf{S}_X$ is neglected at the moment as the uncertainty contribution of the measurement error to these very parameters is the motivation of this treatment.

The AOD is determined for three wavelengths using the aerosol model defined by the three optimized decision parameters: $b1$, $b2$ and L. First, for all four aerosol types that are used the corresponding AOD is interpolated from their LUTs using the AOD level parameter L. Simultaneously, fine aerosol type is mixed from the non-absorbing and absorbing AOD using $b2$, and coarse aerosol type is mixed from the sea salt and dust AOD using the dust fraction. The final AOD is then mixed from the fine and coarse AOD using $b1$.

If we denote the AOD interpolation/mixing operator by p, then for wavelength i, the AOD is

$$\mathrm{AOD}_i = p_i(b1, b2, L). \quad (7)$$

The covariance of AOD is then

$$\mathbf{S}_{AOD} = \mathbf{P}\mathbf{S}_X\mathbf{P}^T.$$

The reported $\mathbf{S}_{\mathrm{AOD}} = \mathbf{P}\mathbf{S}_{X'}\mathbf{P}^T$, (8)

where $\mathbf{P}$ is now the Jacobian of the interpolation/mixing operator

$$\mathbf{P}=\begin{pmatrix} \frac{\partial p_1}{\partial b_1} & \frac{\partial p_1}{\partial b_2} & \frac{\partial p}{\partial L} \\ \frac{\partial p_2}{\partial b_1} & \frac{\partial p_2}{\partial b_2} & \frac{\partial p_2}{\partial L} \\ \frac{\partial p_3}{\partial b_1} & \frac{\partial p_3}{\partial b_2} & \frac{\partial p_3}{\partial L} \end{pmatrix} \quad (9)$$

The uncertainty estimate for AOD can be determined from the diagonal of $\mathbf{S}_{\mathrm{AOD}}$. The uncertainty of the dust fraction, which is not a retrieved parameter, could be added to $\mathbf{S}_{X'}$ if this uncertainty was assumed to be Gaussian.

This error treatment can be straightforwardly applied for the ASV algorithm used over ocean.

4.2 Dust AOD from IASI algorithms

Four algorithms are currently participating in a round-robin exercise within aerosol_cci to produce coarse mode (a.k.a. dust) AOD from IASI. As these algorithms are in an earlier stage of development, their estimation of uncertainty is not necessarily mature (and should not expected to be so). Regardless, each does produce an estimate and the techniques are briefly outlined in this section.

4.2.1 ULB

This section discusses the uncertainties of the ULB-IASI v6 dust AOD product at 10 µm. IASI is an infrared instrument and cannot directly measure AOD at 550 nm, but in the framework of this project an attempt was made to convert dust AOD at 10 µm to dust AOD at 550 nm. This conversion depends on composition and the precise particle size distribution, which are highly variable in both time and space. To avoid the introduction of extra assumptions and dependencies, the ULB IASI algorithm uses a constant conversion factor of two. Needless to say, it is practically impossible to accurately estimate the uncertainty on the resulting AOD product at 550 nm, but individual errors in the range [-50%, 100%] can be expected as a result of using a constant conversion factor (in addition to the uncertainty of the AOD at 10 µm).

Systematic uncertainties include biases introduced by the choice of size distribution, refractive index and forward model. Their estimation follows studies from the LMD group (Pierangelo et al., 2004). The size distribution and refractive index were each estimated to yield an error less than 10%. Other forward model errors are expected to be negligible. Biases in the neural network training are mostly below 10%. In view of this, the systematic uncertainty is conservatively estimated to be smaller than 30%.

The nature of the neural network simplifies the calculation of how errors on the input parameters translate to uncertainties of the final OD retrieval. The following uncertainties σ are attributed to the five input parameters,

1. Aerosol altitude. The standard deviation of the CALIOP heights, $\sigma_{ALT} = \sigma_{cal}$, is used.
2. IASI instrumental noise on R. The R-value has by definition an uncertainty of $\sigma_R = 1$.
3. IASI instrumental noise on the input baseline channels, $\sigma_{BL} = 0.28$ K.
4. Temperature profile. A value of $\sigma_{TEMP} = 1$ K has been applied for the whole profile (Pierangelo et al., 2004; August et al., 2012).
5. Humidity profile. A value of $\sigma_{HUM} = 10\%$ has been applied for the whole profile (Pierangelo et al., 2004; August et al., 2012).

Using these, and assuming that these uncertainties are indeed random, the total random uncertainty on the AOD at 10 µm due to dust is given by,

$$\sigma_{OD} = \sqrt{\left(\frac{\partial OD}{\partial A}\sigma_{ALT}\right)^2 + \left(\frac{\partial OD}{\partial R}\sigma_R\right)^2 + \left(\frac{\partial OD}{\partial B}\sigma_{BL}\right)^2 + \left(\frac{\partial OD}{\partial T}\sigma_T\right)^2 + \left(\frac{\partial OD}{\partial H}\sigma_{HUM}\right)^2}.$$

This value is calculated on a per-pixel basis and is part of the retrieval product.

4.2.2 LMD

The method used to retrieve aerosol properties is based on proximity recognition within pre-computed LUTs. Error estimation for such a non-parametric inversion is non-trivial. Currently, the standard deviation of the set of solutions that are consistent with the observation cannot be readily determined

because their distribution is in general not Gaussian (though their average value correctly represents the observation). Characterizing the precision/accuracy of this method requires (i) developing a (non-parametric) specific procedure and (ii) accounting for the error made in the conversion of the 10 µm AOD to visible AOD, a step necessary to compare against solar measurements in the absence of validation data in the infrared. The infrared (10 µm) to visible (500 nm) AOD ratio is defined as the ratio of the extinction coefficient at the two wavelengths. Evaluating this coefficient requires an accurate knowledge of the size distribution parameters, as well as the refractive index. In the absence of estimates of these parameters, the visible AOD is estimated during the inversion process by assuming constant values, which is not verified by observations and may therefore induce offset and random errors in the conversion.

A first approach to quantify uncertainties at 10µm (biases and standard deviation) was described in Pierangelo et al. (2004). Evaluation has been performed using synthetic retrievals within the LUT themselves. In this study, three complementary approaches will be used to estimate the uncertainties on the method.

- In complement to the method applied in Pierangelo et al. (2004), synthetic simulations (OSSE's) will be performed for a large number of situations derived from ECMWF reanalysis profiles, with surface temperature and surface emissivity chosen at random, across various values of optical depths and mean altitudes. Statistical analysis of the bias and standard deviation of the 10µm AOD retrieval would then provide an *a priori* estimation of the precision of the method.

- An empirical estimate of the uncertainties will be obtained by converting IASI AOD to the visible and comparing to AERONET coarse-mode AOD over more than eight years of collocated data around the world. This approach both quantifies the robustness of the method (by analysing the bias and standard deviation of the comparisons) and provides an estimate of the uncertainties associated with the conversion between infrared and visible AOD. However, correctly interpreting these results requires a careful analysis of the consequences of the difference in spatial resolution between a point-like observation from AERONET and the 12 km pixel of IASI, as well as of the impact of a difference in time of measurements (observations from AERONET are averaged over 1.5 hours around the IASI observation time).

- Finally, in order to obtain a more precise estimate of the uncertainty for each pixel, we are also developing "non-parametric" methods more relevant to our approach. Such statistics are not based on parameterized families of probability distributions and, unlike parametric statistics, make no assumptions about the probability distributions of the variables being assessed.

4.2.3 IMARS (DLR)

The IMARS (Infrared Mineral Aerosol Retrieval Scheme) algorithm for IASI is based on spectral pattern recognition of a predefined set of dust mixtures (with differing mineralogical compositions) to determine AOD (in the thermal infrared, TIR) and dust layer height (from the thermal contrast between the dust emission temperature and the skin temperature of the underlying surface). The uncertainty of the AOD retrieval at 10µm is estimated from the RMS deviation of the forward simulation from the measurement (at 35 hinge points in the atmospheric window). The simulation in this calculation is the average of all realizations considered, weighted by their probability of resembling the measurement. That probability assumes a normal distribution, with mean equal to the measured value and standard deviation derived from the spread of equivalent optical depths returned by the different dust mixtures.

The radiative transfer solution is approximated using a simple two-stream approach and a finite set of prescribed surface emissivity spectra are used in the different forward simulations. It is assumed that these uncertainties dominate relatively minor error sources, such as instrument noise or the atmospheric state (e.g. temperature, water vapour profiles, abundance of other trace gases), due to the maximum-brightness-temperature-in-bin approach used to prepare the IASI spectra within IMARS. Additional uncertainty is introduced by variation in the dust itself. The IMARS retrieval assumes that the mineralogical composition of the dust is included in the set of mixtures considered in the forward simulations. Although a comprehensive list of dust mixture analyses from field campaigns has been used to generate the representations, there are sources which produce dust with very specific mineralogies (for example, the Bodélé Depression in Northern Chad or the coastal dune fields in Kuwait). These may not be accurately represented by any of the mixtures used and the resulting error may not be reflected in the uncertainty. The same can be assumed for the particle size distribution of the dust.

Recent analyses have shown that the spectral evolution of dust extinction in the TIR is strongly impacted by particle shape. The IMARS scheme, being based on spectral pattern recognition, uses an approximation for non-spherical particles which has been tested against laboratory measurements, resulting in very high correlations (R^2 of about 0.95). Nevertheless, it is possible that extreme particle shapes may occur, which are not well represented by these assumptions. In such cases, the reported uncertainty will be an underestimate, especially as the transfer of AOD from TIR to 0.55µm strongly depends on the retrieved composition and particle size.

The conversion of dust AOD from TIR to 0.55µm uses the ratio of extinction efficiencies derived from the particular composition and size distribution used. Consequently, each dust realisation has its own TIR-VIS AOD ratio. For iron oxides, the TIR sensitivity is extremely low, but they are the dominant absorption in the visible range. It is not possible to address the iron oxide content of desert dust with TIR observations, so the (consistent) transfer AOD ratios have added uncertainty.

4.2.4 MAPIR

The optimal estimation used in the MAPIR algorithm allows for rigorous uncertainty propagation. However, the current algorithm does not consider uncertainties in parameters used by the forward model as the Jacobians for the parameters are not computed. (Aerosol parameters because the necessary physics is not yet coded and those linked to surface and atmospheric parameters to save time in this already very time-consuming algorithm.)

Uncertainties considered in the error budget are almost exclusively linked to the measurements. However, retrievals never converged using the real measurement uncertainty because it is extremely small compared to the expected uncertainty resulting from all parameters (especially atmospheric temperatures and aerosol parameters). We therefore empirically selected the value used for the measurements error such that retrievals would decently converge. This is detailed in the ATBD.

The *a priori* vertical profile of aerosols is currently characterized by its variance (a single value for the whole vertical profile, being part of the climatology) and a Gaussian vertical correlation of 1km length. The *a priori* surface temperature is arbitrarily characterised by a variance of 5%. This may seem very large, but the *a priori* value comes from IASI Level 2 retrievals, which are extremely biased in the presence of dust aerosols, especially above deserts. The forward model itself is a full RT calculation with minimal approximation, resulting in minimal uncertainties.

The uncertainties provided with the MAPIR product are in a very early stage and probably not yet very representative of the real distribution of error. Indeed, as the measurement uncertainty has been empirically adjusted to allow for retrievals to complete (in the absence of physical characterization of all parameters' uncertainties), the resulting AOD uncertainty is based on biased values, even though it is computed by a rigorous method.

A short sensitivity study has been undertaken on a small data subset. Table 4-2 summarises the mean and maximum AOD and layer altitude differences when perturbing various parameters within their "considered uncertainty". The surface temperature was the parameter leading to the highest uncertainty; it has now been included in the state vector in the retrievals.

Table 4-2: *Results of MAPIR sensitivity study.*

Parameter		Considered uncertainty	Mean (max) AOD ≠	Mean (max) Alt ≠
Aerosol	Refractive index	Different databases	10 (40) %	0.1 (1) km
	Size	From R_{eff} 2µm to 4µm	20 (50) %	0.6 (1.5) km
Surface emissivity	Land	0.04	0 %	0 km
	Ocean	0.05	5 (20) %	0.1 (1) km
Atmos	RH / T	Difference between IASI l2 version 4 and 6: max 10% / 4K.	10 (20) %	0.5 (1) km

4.3 Stratospheric aerosols from GOMOS

Random errors are significant in the GOMOS retrieval, unlike most satellite remote sensing retrievals which are dominated by systematic errors. The random errors in AerGom data are mainly due to the propagation of 1) measurement noise and 2) imperfect scintillation correction (see Tamminen et al, 2010 for more details).

1) GOMOS observes stars of different brightness and effective surface temperatures, which both affect significantly the SNR. One could consider GOMOS to be 180 different instruments with various characteristics dependent on the properties of the star viewed.

2) Stellar scintillation is the result of air density perturbations (generated by internal gravity waves and turbulence) causing refractive index perturbations, with characteristic scales from a few kilometres down to a dissipation scale. The GOMOS scintillation correction cannot remove isotropic turbulence (Sofieva et al, 2009) and the residual is significant. Imperfect scintillation correction causes artificial structures in the data, and these are the largest modelling errors in the GOMOS retrieval of the stratosphere (of the order of 30% around 10 km, 2-3% between 15-25 km, and 10-50% between 35-40 km).

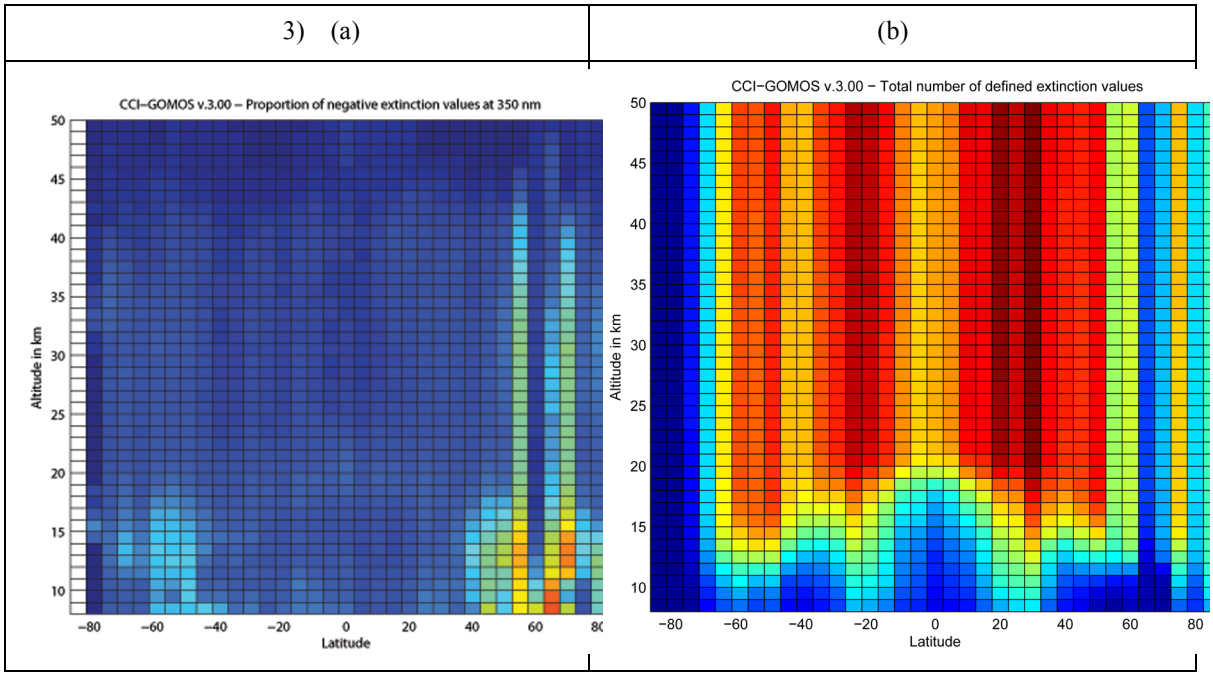


Figure 4-1: *(a) Proportion of negative extinction values at 350 nm, cumulated at all time and longitude values, as a function of latitude and altitude. A value is considered as negative if the sum of the extinction value and its associated uncertainty is negative. (b) Corresponding total number of definite extinction values.*

Other sources of uncertainty are the parameterization model (around 10% for aerosols below 35km, 10-15% above), the gaseous species cross-sections and their variation with temperature, uncertainty in the neutral density profile (<5% below 22km, 5-15% between 22-40 km), ray path computation, and neglected constituents. The stratospheric product includes the stratospheric AOD, which requires a precise determination of the tropopause height. The AOD uncertainty resulting from that of the tropopause height, calculated from ECMWF profiles, is of the order of 10-15%. Concerning the aerosol spectral parameterization, the spectral retrieval makes use of an aerosol extinction model based on the extinction values at a set of fixed wavelengths (in the last AerGOM version, v.3.00, a quadratic function in the inverse wavelength at fixed values of 350, 550 and 756 nm). In the case of cold stars, the choice of reference value at 350 nm is not optimal because it corresponds to a spectral range where the star emissivity is limited. This has an impact on the quality of the extinction retrieval at latitudes higher than 40°N, with a more frequent occurrence of negative extinction values at the lower wavelengths and at low altitude. As a worst case, Figure 4-1 shows that about 40% of negative values at 350 nm at 15 km and below. Moreover, a particularly high proportion of negative values is found at all altitudes of interest (z <~35 km) in the latitude bins centred on 52.5°N and 67.5°N, but these latter cases are characterized by a low measurement rate.

A natural framework for studying the uncertainties is obtained by describing the solution as a Bayesian *a posteriori* distribution. The posterior uncertainties are approximated with a Gaussian distribution (whose mean and covariance are computed). Because all species are retrieved at the same time and their cross-sections spectrally overlap, some correlation in retrieval errors is observed. These should be accounted for but are not currently. The results of propagating all sources of error mentioned in this section are shown in Figure 4-2 for a representative selection of stars of different magnitudes (M_s) and temperatures (T_s).

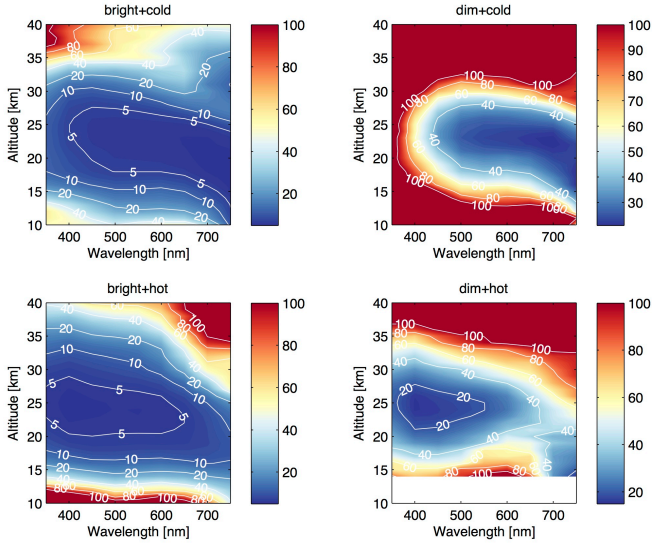


Figure 4-2: *Median relative uncertainty (in %) of AerGom aerosol extinction as a function of altitude and wavelength. These values were calculated for 300 observations made with various stars: bright (M_s < 1.5, left), dim (2.6 < M_s< 3, right), cold (Ts < 6000 K, up) and hot (Ts > 26000 K, down). These values were produced by AerGom v1.0.*

5 VALIDATION OF UNCERTAINTIES

5.1 Validation of products against AERONET

All aerosol_cci products have been validated, to varying extents, against the ground-based sun-photometer measurements provided by AERONET. These activities are summarised in detail in the Product Validation Report [RD4] and will only be briefly discussed here.

Direct sun AOD measurements at 550 nm are substantially more accurate than those produced by other techniques as there are fewer sources of error (e.g. there is no influence from the surface, the impact of multiple scattering is minimised using a long baffle). However, the observations include cloud filtering and are only representative of the subset of environments that contain an AERONET station (e.g. a consistently poor retrieval over remote mountainous regions would not be identified by this validation). Though AERONET provides an undeniably useful and high quality validation data set, it should not be considered to represent the "true" AOD. Despite these limitations, it is common (both in this project and generally) to perform validation activities as if AERONET measurements have no uncertainty for simplicity. An example of such results is shown in Figure 5-1.

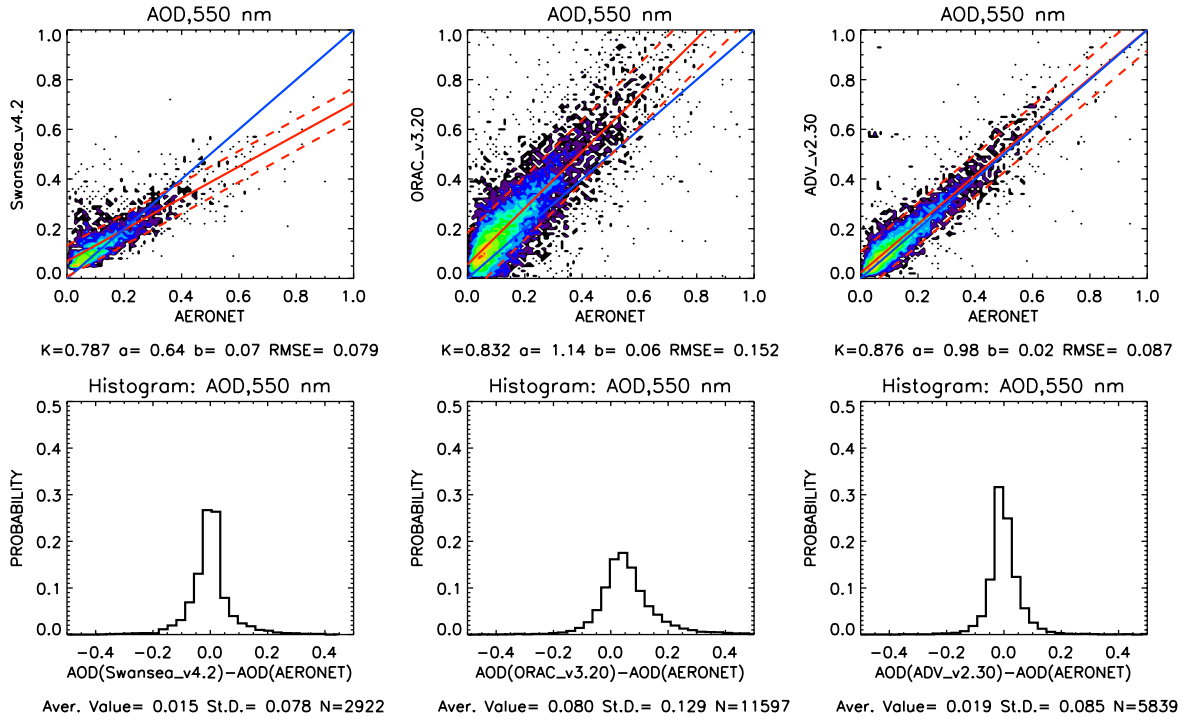


Figure 5-1: *Comparison of AOD at 550 nm as retrieved by the SU, ORAC, and ADV algorithms (left, middle, and right columns, respectively) over ocean from AATSR against AERONET direct-sun observations. Scatterplots (top) include the one-to-one line (blue) and a linear best-fit (red, solid) with associated uncertainty (red, dashed). The distribution of discrepancy between the two measurements is shown in the bottom row, while the correlation (K), gradient (a), offset (b), and RMS discrepancy (RMSE) are printed between the plots.*

5.2 Validation of quoted uncertainties

The uncertainty estimates produced by the ATSR algorithms can be validated by evaluating their ability to represent the distribution of discrepancy relative to AERONET, which are substantially more accurate than satellite retrievals. Neglecting the uncertainty in AERONET observations and possible issues with their ability to represent a satellite pixel area, the error in the retrieval can be approximated by the difference between the satellite and AERONET retrievals. To evaluate how well the pixel-level uncertainty represents the observed distribution of error, we consider the distribution of the ratio Δ between the difference between the satellite and AERONET retrievals and the ATSR uncertainty.

$$\Delta = \frac{AOD_{\mathrm{ATSR}} - AOD_{\mathrm{AERONET}}}{\sigma_{\mathrm{ATSR}}}.$$

If the uncertainty is a good representation of the discrepancy, **Δ** will be normally distributed and 68.3% of values should fall within the range [-1, +1], with zero mean and unit standard deviation. A non-zero mean indicates the presence of residual systematic errors (an issue for algorithm development). A standard deviation less than one indicates uncertainties are underestimated ($\Delta \leq 68.3\%$), which could result from neglecting an important source of error, while a standard deviation greater than one ($\Delta \geq 68.3\%$) indicates an overestimate, which is not necessarily problematic (for example, the subset of circumstances sampled by the AERONET observations used may not represent the global distribution that the uncertainty describes).

AOD pixel level uncertainties from three algorithms, SU (Swansea University) version 4.3, ADV/ASV (AATSR Dual/Single View) in brief denoted as ADV, version 2.30_plume, ORAC (Oxford RAL Aerosol and Cloud retrieval) version 4.01, and an Ensemble version 2.6, were evaluated. The co-location criteria were: 50 km spatial and 30 min. temporal difference; only AERONET sites below 2000 m were considered.

Figure 5.2 shows an example of the comparison of nominator and denominator of Δ for the Avignon AERONET station (43.9°N, 4.9°E, 32 m a.s.l) for AATSR data from 2002 onward. This site is located in an agricultural region 10 km east of the city of Avignon. It can be seen that the satellite uncertainties are an upper envelope of the differences between AERONET and satellite AOD. There is a minimum uncertainty of about 0.02 and 0.01 for SU v4.3 and ORAC v4.01 algorithm, respectively. Besides a number of reasonable good represented uncertainty estimates, ADV shows spurious low uncertainty values.

The comparison is shown in Figures 5-3 until 5.7. Data for ATSR-2 (left panels) and AATSR (right panels) are shown, within each group, the left pair consider inland AERONET stations while the right consider coastal stations. The upper panel show Δ against the AERONET AOD, the lower panel are histograms of the estimated uncertainty (blue) and difference compared to AERONET (red). The results of the comparison for the individual years are visualized in Figure 5.8, which shows the percentage of Δ within [−1, +1] per year for the various algorithms for ATSR-2 and AATSR over land and coastal sites. In Table 5.1 the results were summarized. Given are the percentages of Δ within [−1, +1] for the ATSR-2 and AATSR period, separated for the ATSR-2 and AATSR and the four algorithms evaluated. The percentage of Δ within [−1, +1] for the ATSR-2 AATSR overlapping period is shown in Table 5.3.

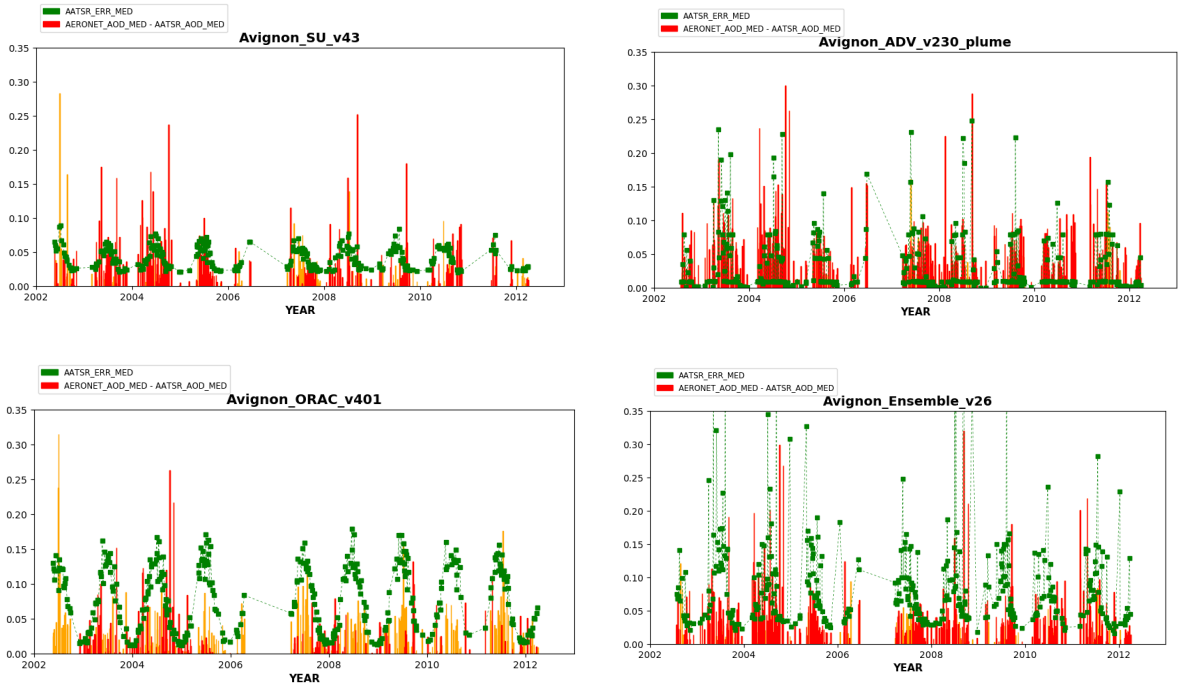


Figure 5-2: Example comparison of nominator and denominator of Δ for the Avignon AERONET station (43.9°N, 4.9°E, 32 m a.s.l). If the median of the AOD at the AERONET site is larger than the median of the AATSR AOD then red symbols are used, if AOD at site is smaller than the satellite AOD then orange symbols are used.

In Figure 5-3, the plots evaluate the v2.30_plume ADV product. As already expected from the example shown in Figure 5.2, the ADV values were substantially lower than the optimum percentage value of 68.3%, with Δ = 35 % and 30 % for ATSR-2 and AATSR over land, respectively. The analogue values over ocean are lower: 24 % for the ATSR-2 comparison and 19 % for AATSR. The v43 SU algorithm appears to be the most accurate representation of the observed distribution with Δ = 62 % and 65 % for ATSR-2 and AATSR over land, and Δ = 65 % and 67 % over ocean (see Figure 5.4). Figure 5.5 shows the values for the v401 ORAC algorithm, which are slightly higher, but as well seem a very reasonable representation of the uncertainty over land, while the values over ocean are too low. Δ = 67 % and 67 % for ATSR-2 and AATSR over land, and Δ = 20 % and 28 % over ocean (see Figure 5.4). As the uncertainties of the ensemble are calculated with error propagation, they are dominated by the higher values. Δ = 88 % and 86 % for ATSR-2 and AATSR over land, and Δ = 58 % and 67 % over ocean (see Figure 5.4).

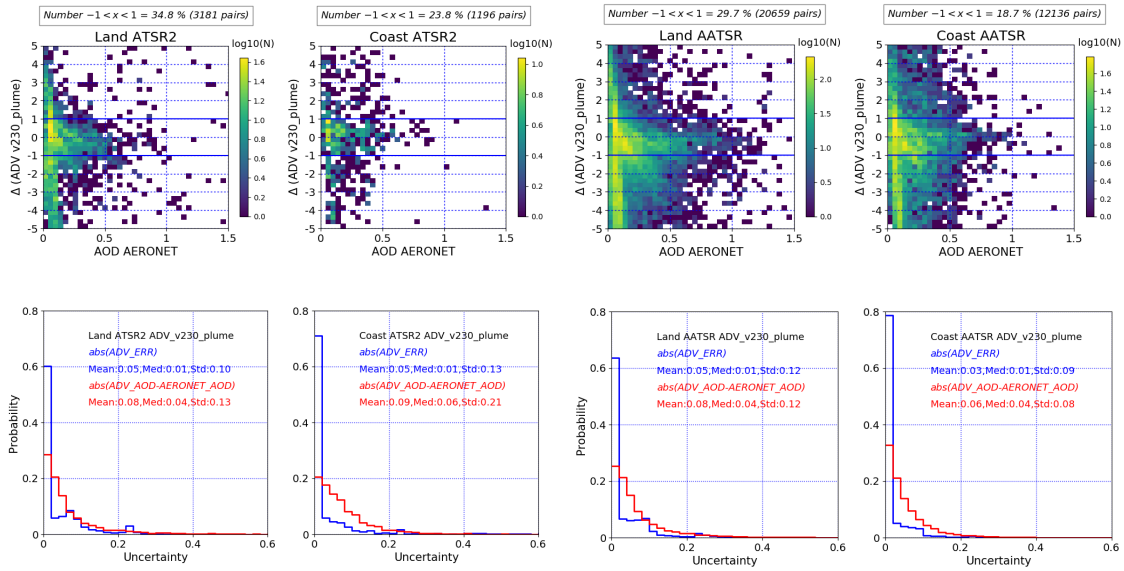


Figure 5-3: *Evaluation of estimated uncertainty in versions 2.30_plume of the ADV algorithm by comparison against AERONET. Data for ATSR-2 (left panels) and AATSR (right panels) are shown, within each group, the left pair consider inland AERONET stations while the right consider coastal stations.* Top: **Δ** *against the AERONET AOD. The percentage of points falling in the range [-1, 1] is written at the top, which would be 68.3% for a Gaussian distribution.* Bottom: *Histograms of the estimated uncertainty (blue) and difference compared to AERONET (red). The mean, median, and standard deviation of each distribution are written to the top.*

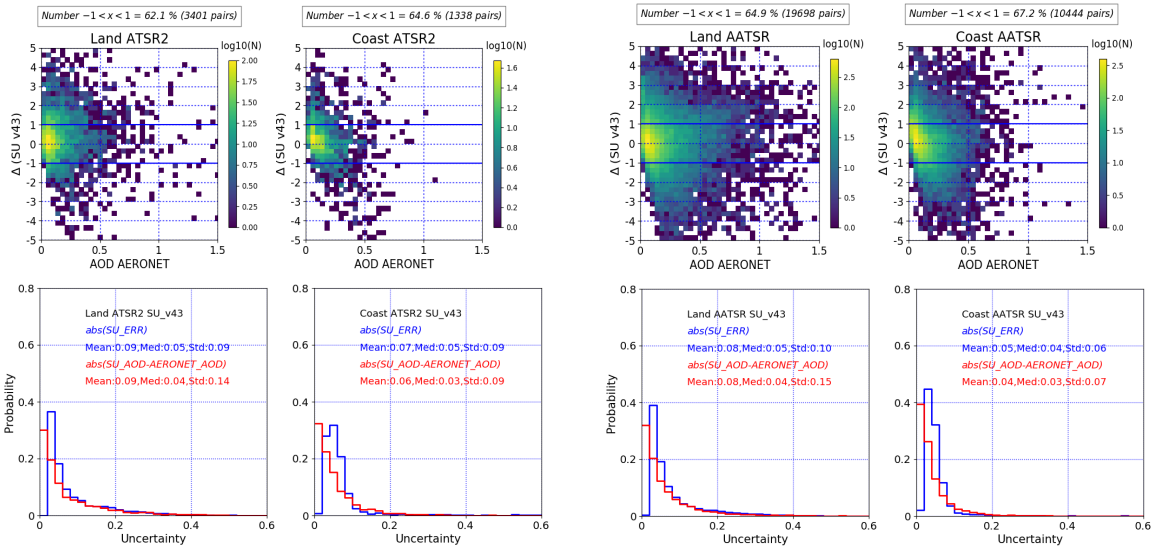


Figure 5-4: *As*

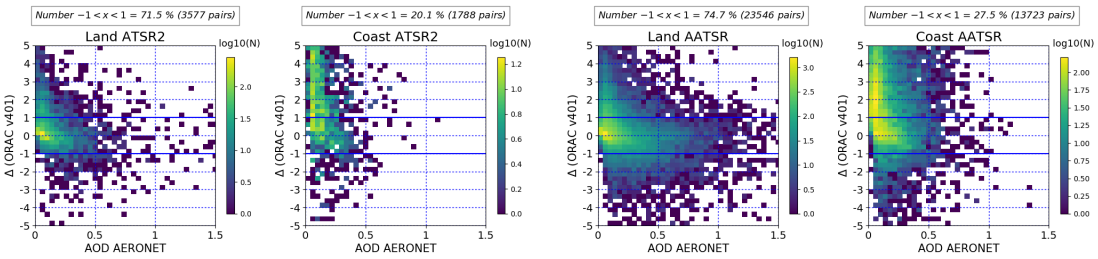


Figure 5-6: *As **Error! Not a valid bookmark self-reference.** but for the v2.6 Ensemble ATSR-2 and AATSR products.*

but for the v43 SU ATSR-2 and AATSR products.

Fi

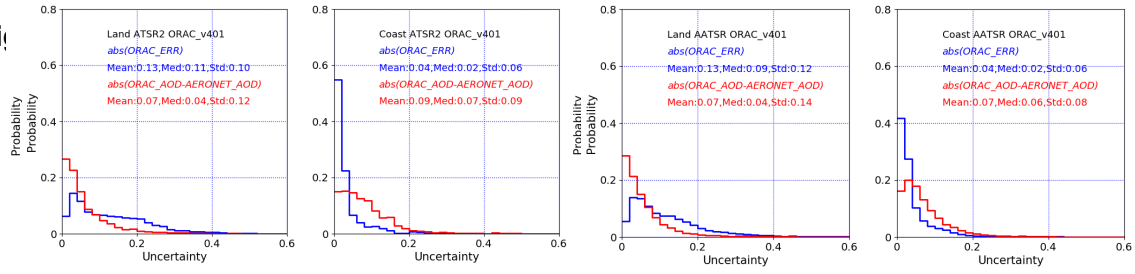


Figure 5-6: *As **Error! Not a valid bookmark self-reference.** but for the v2.6 Ensemble ATSR-2 and AATSR products.*

but for the v401 ORAC ATSR-2 and AATSR products.

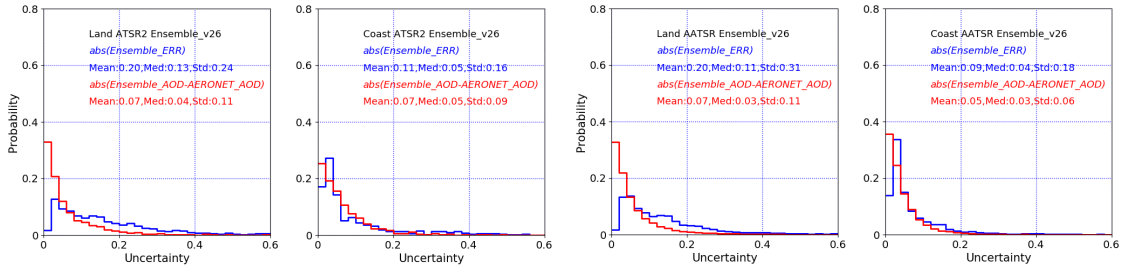
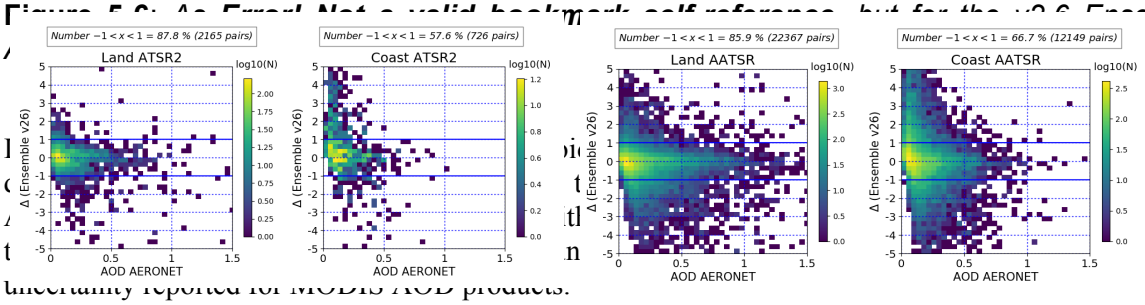


Figure 5.6: As Error! Not a valid bookmark self-reference. but for the v2.6 Ensemble



Considering the v401 ORAC and 2.30_plume ADV algorithms, uncertainties over land better reproduce the distribution of differences (with respect to AERONET) than those at coastal sites. This is not surprising as coastal waters are difficult to model as they tend to be shallow and contain sediments. The v43 SU algorithm over coastal regions matches the optimal value well.

Figure 5.7 shows the long term stability of the percentage of Δ within [−1, +1] over the 17 years of data recorded from ATSR-2 and AATSR. All algorithms do show a good long-term stability. Δ for the overlapping time period between the two satellites are given in Table 5.2. For the SU and the ORAC algorithm, we evaluated the periods between 06/02-05/03. For ADV and the Ensemble, the period covered was 07/02-12/02. There is only a few percent deviations founds, so considering the much lower number of co-located data during the earlier ATSR-2 period, we conclude that the matching between ATSR-2 and AATSR is reasonable for all algorithms.

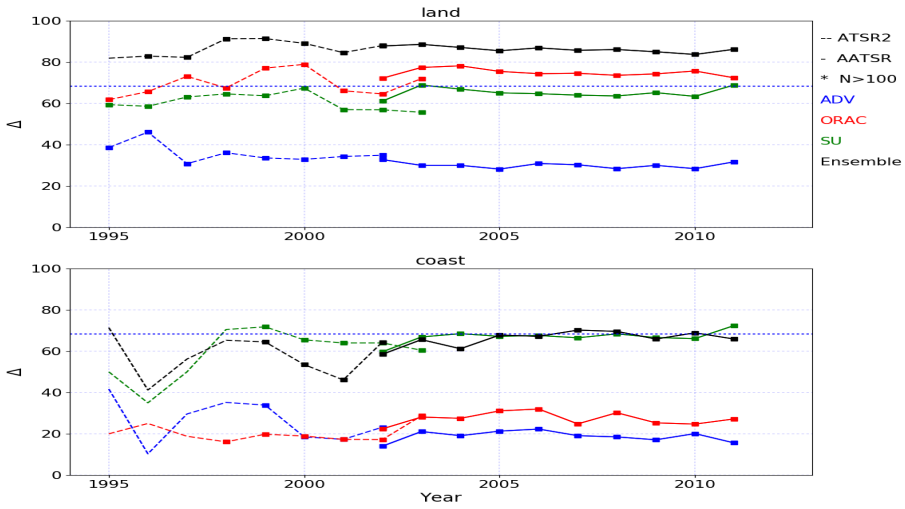


Figure 5-7: Percentage of Δ within [−1, +1] per year for the ADV v230_plume, ORAC v401, SU v43 and Ensemble v26 algorithm for ATSR-2 and AATSR over land and coastal sites. The black line shows the optimum percentage value. The dotted lines, without symbols, are showing years where less than 100 co-locations were found, so these comparisons are not significant.

Table 5.1: Percentage of Δ within [−1, +1] for the entire ATSR-2 and the entire AATSR period, separated for the two instruments and the four algorithms evaluated.

	Land		Ocean sites	
Algorithm	ATSR-2	AATSR	ATSR-2	AATSR
ORAC v401	67 % ± 06 %	75 % ± 02 %	20 % ± 04 %	28 % ± 03 %
SU v43	62 % ± 06 %	65 % ± 02 %	65 % ± 04 %	67 % ± 03 %
ADV/ASV 2.30_plume	35 % ± 05 %	30 % ± 03 %	24 % ± 11 %	19 % ± 04 %
Ensemble v26	88 % ± 04 %	86 % ± 02 %	58 % ± 10 %	67 % ± 05 %

Table 5.2: Percentage of Δ within [−1, +1] for the overlapping period of the two instruments ATSR-2 and AATSR. For the SU and the ORAC algorithm, we evaluated the periods between 06/2002-05/2003. For ADV and the Ensemble, the period covered was 07/2002-12/2002.

S, O: 06/02-05/03 A, E: 07/02-12/02	land		coast	
	ATSR-2	AATSR	ATSR-2	AATSR
ADV v230_plume	36 %	33 %	22 %	14 %
ORAC v401	66 %	75 %	24 %	24 %
SU v43	56 %	62 %	61 %	61 %
Ensemble v26	88 %	88 %	58 %	59 %

5.3 Uncertainties in Level 3 products

As discussed in Section 3, there are many sources of error and the relative importance of each is a function of the scene observed and the manner in which the measurement is made. For Level 2 AOD retrievals, the system errors caused by the choice of aerosol optical properties are often dominant, though measurement and parameter errors are important. When data are averaged to Level 3, the uncertainty due to the latter two sources of error is expected to decrease as those are random errors, but it is not clear how the system errors propagate. If only a single type of aerosol is observed within the data averaged, the error would be a systematic difference between the properties assumed and those observed, such that the error would not change with averaging. If multiple types of aerosol are observed, a variety of systematic errors will be combined. These will add together or cancel out in an unknown manner, resulting in a highly uncertain uncertainty.

Resolution errors are generally considered to be small at Level 2 because aerosol fields are expected to vary over days and tens of kilometres (Anderson et al., 1998) while measurements are over less than a second and 10 km. This is not true when the data are aggregated to Level 3, where grid cells cover hundreds of kilometres and either a day or month. The underlying variability of the aerosol field then becomes important.

5.4 Empirical investigation of the representation of uncertainty

At a workshop organised by this project, despite awareness of the sources of error in Level 3 aerosol products, there were few ideas for how to quantify the resulting uncertainties and little consensus over the validity of those proposed. As such, the aerosol_cci project has explored an empirical approach to the problem. A number of simple metrics which may relate to the uncertainty in Level 3 data were defined. Each AATSR algorithm produced realisations of the metrics for daily data over Mar, Jun, Sep, Dec of 2008. These were evaluated against collocated AERONET observations (as in Section 5.2) to assess if any metric presented a useful representation of their discrepancy.

The metrics used were, for N observations AOD_i with uncertainty σ_i:

1. The mean of the reported uncertainty in the pixels $\overline{AOD} = \frac{1}{N}\sum_i \sigma_i$, representing the confidence in the retrievals that went into this L3 pixel;
2. The standard deviation of the pixels $\sqrt{\sum_i \frac{(AOD_i - \overline{AOD})^2}{N-1}}$, approximating the natural variability in this L3 pixel;

3. The propagation of the uncertainties into the mean $\frac{1}{N}\sqrt{\sum_i \sigma_i^2}$, treating each retrieval as independent and contributing a random error;
4. The sum of 2 and 3, assuming that these terms represent the expected dominant sources of error;
5. A worst-case propagation $\frac{1}{N}[\sum_i(AOD_i + \sigma_i) - \sum_i(AOD_i - \sigma_i)]$, a very simplistic approach that should replicate twice the value of metric 1 if the distribution is symmetric.

The comparison is currently in progress, with preliminary results shown in Figure 5-3 to 7. Almost all the metrics are underestimates of the observed difference as fewer than 63% of the **Δ** fall in the range [-1,1]. Considering the histograms, metrics 1 and 2 clearly provide a poor representation of the distribution of error in all algorithms but the remainder have merit in at least one case. As metric 4

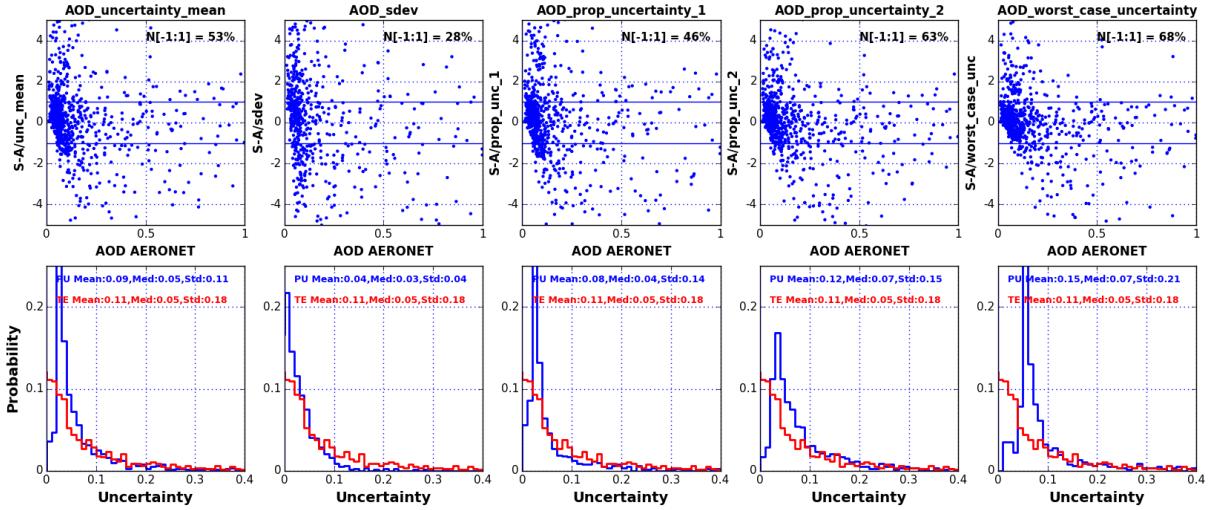


Figure 5-3: *Evaluation of five metrics of the uncertainty on Level 3 data (columns 1 to 5 being metrics 1 to 5, respectively) using the SU retrieval products. Plots are otherwise as in*

Figure 5-6: *As **Error! Not a valid bookmark self-reference.** but for the v2.6 Ensemble ATSR-2 and AATSR products.*

.

contains metric 3, this may imply that metric 4 is the most useful representation of Level 3 uncertainty (of those considered) but a more robust technique will be developed to evaluate the quality of each metric in detail.

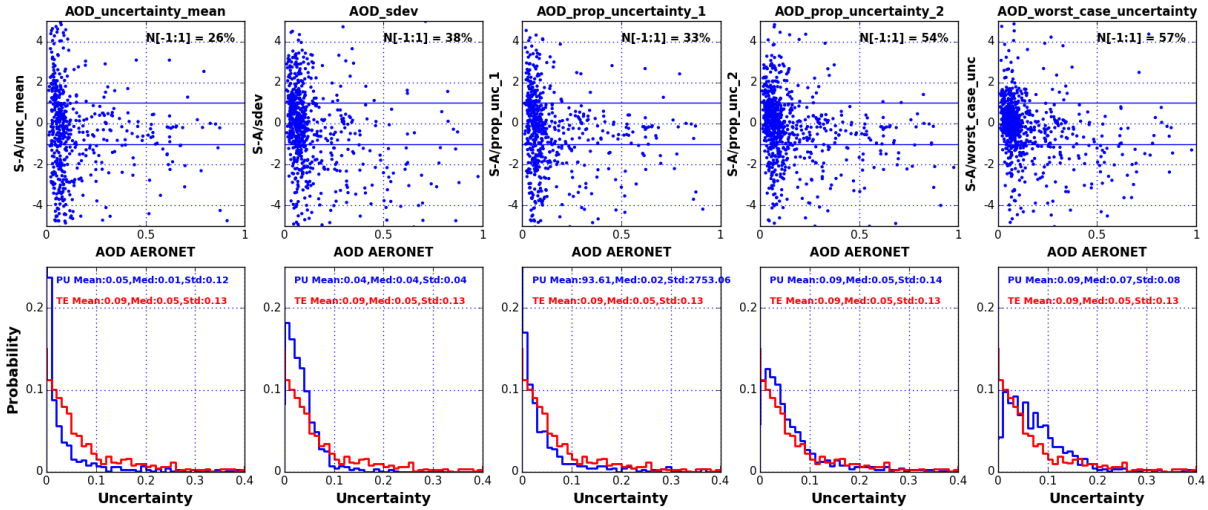


Figure 5-4: *As Figure 5-3 but for the ORAC retrieval products.*

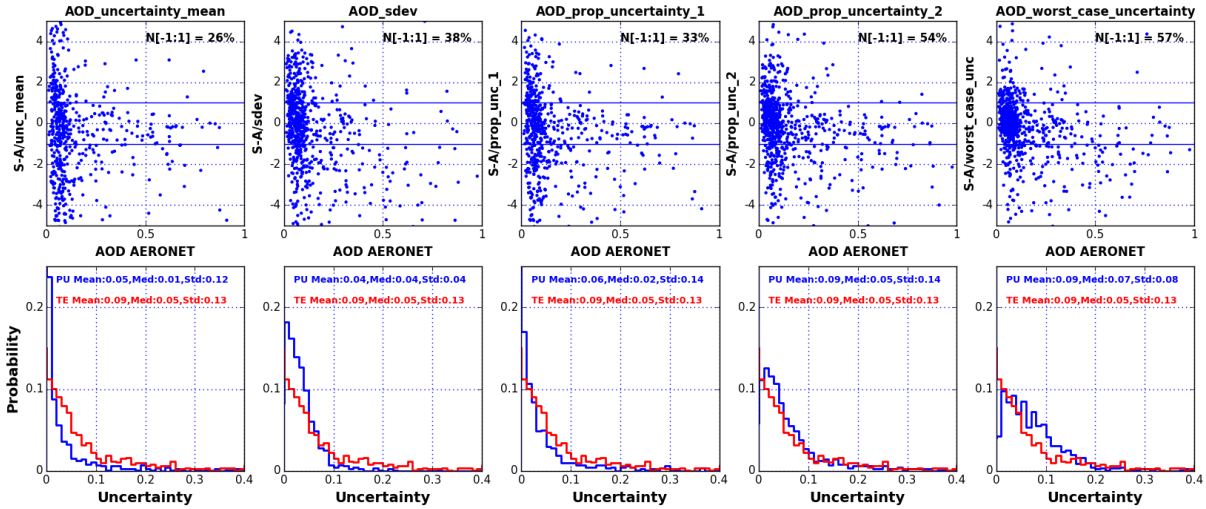


Figure 5-5: *As Figure 5-3 but for the ADV retrieval products.*

6 GUIDELINES FOR USING THE PRODUCTS

Irrespective of the data set used, users should:

- Read the short (~1 page) user guide. It summarises the key features and limitations of the data to provide a guide to avoid common confusions or mistakes.
- Be aware that flagging data as low quality indicates that there is little confidence that the algorithm is appropriate for that data. Depending on how conservative the data producer has chosen to be, that judgement may remove good data and/or exclude poor data and may introduce spatio-temporal artefacts. This is not to say that we recommend using low quality data, but users should be aware that quality flags are intended as qualitative statements not robust assessments of uncertainty.
- Use the Level 3 products prepared alongside the Level 2 data. The retrieval experts have carefully considered sampling issues that may not be easily managed in post-processing without additional information (that is not necessarily available).
- Feedback to the project and data producers as to the quality and utility of the products. Without input, it is impossible to produce more useful products.

7 CONCLUSIONS

An estimate of the uncertainty on a measurement is necessary to make appropriate use of the information conveyed by a measurement. Awareness of the importance of pixel-level uncertainty estimates is beginning to pervade the aerosol remote sensing community through efforts such as the yearly AeroSAT workshop. By comparing the performance of the algorithms evaluated within the aerosol_cci project to each other and validation data, the quality of their (preliminary) uncertainty estimates can be evaluated and potential areas for improvement can be illuminated.

This work still has to continue to better homogenise uncertainty calculations among the different algorithms for one sensor. The IASI algorithms are in an earlier stage of development, but are already considering the estimation of uncertainty – its validation is even more complex (need to use AERONET coarse mode AOD at 550 nm as proxy for IASI retrievals which include as challenging step the conversion from 10 micron to 550 nm.

End of the document