Remote Sensing of Environment 162 (2015) 295–315

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Evaluation of seven European aerosol optical depth retrieval algorithms for climate analysis

Gerrit de Leeuw [a,b,c,*], Thomas Holzer-Popp [d], Suzanne Bevan [e], William H. Davies [e], Jacques Descloitres [f], Roy G. Grainger [g], Jan Griesfeller [h], Andreas Heckel [e], Stefan Kinne [i], Lars Klüser [d,j], Pekka Kolmonen [a], Pavel Litvinov [m], Dmytro Martynenko [d], Peter North [e], Bertrand Ovigneur [f], Nicolas Pascal [f,l], Caroline Poulsen [k], Didier Ramon [l], Michael Schulz [h], Richard Siddans [k], Larisa Sogacheva [a], Didier Tanré [m], Gareth E. Thomas [g], Timo H. Virtanen [a], Wolfgang von Hoyningen Huene [n], Marco Vountas [n], Simon Pinnock [o]

[a] Finnish Meteorological Institute (FMI), Erik Palmenin Aukio 1, P.O. Box 501, FI-00101 Helsinki, Finland
[b] Department of Physics, University of Helsinki, Erik Palmenin Aukio 1, P.O. Box 64, FI-00014 Helsinki, Finland
[c] TNO Environment and Geosciences, Dept. of Air Quality and Climate, P.O. Box 80015, 3508 TA Utrecht, The Netherlands
[d] DLR German Aerospace Center, German Remote Sensing Data Center (DFD), D-82234 Oberpfaffenhofen, Germany
[e] Global Environmental Modelling and Earth Observation (GEMEO), Department of Geography, College of Science, Swansea University, Singleton Park, Swansea SA2 8PP, UK
[f] ICARE/CNRS, ICARE Data and Services Center, University Lille 1, Polytech Lille E206, 59655 Villeneuve d'Ascq Cedex, France
[g] Atmospheric, Oceanic & Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK
[h] Research and Development Department, Norwegian Meteorological Institute, Postboks 43 Blindern, 0313 Oslo, Norway
[i] Max-Planck-Institut für Meteorologie (MPI), Hamburg, Germany
[j] Augsburg University, Experimental Physics II, Universitätsstr. 1, D-86135 Augsburg, Germany
[k] STFC Rutherford Appleton Laboratory, Chilton OX11 0QX, UK
[l] HYGEOS, Euratechnologies, 165 Avenue de Bretagne, 59000 Lille, France
[m] Laboratoire d'Optique Atmosphérique, CNRS, Universite Lille-1, Bat P5 Cite Scientifique, 59655 Villeneuve d'Ascq, France
[n] Universität Bremen, Institute of Environmental Physics, Bremen, Germany
[o] ESA-ESRIN, Frascati, Italy

ARTICLE INFO

Article history:
Received 1 November 2012
Received in revised form 17 April 2013
Accepted 19 April 2013
Available online 10 October 2013

Keywords:
Aerosol retrieval algorithms
Aerosol optical depth
AATSR
MERIS
PARASOL

ABSTRACT

Satellite data are increasingly used to provide observation-based estimates of the effects of aerosols on climate. The Aerosol-cci project, part of the European Space Agency's Climate Change Initiative (CCI), was designed to provide essential climate variables for aerosols from satellite data. Eight algorithms, developed for the retrieval of aerosol properties using data from AATSR (4), MERIS (3) and POLDER, were evaluated to determine their suitability for climate studies. The primary result from each of these algorithms is the aerosol optical depth (AOD) at several wavelengths, together with the Ångström exponent (AE) which describes the spectral variation of the AOD for a given wavelength pair. Other aerosol parameters which are possibly retrieved from satellite observations are not considered in this paper. The AOD and AE (AE only for Level 2) were evaluated against independent collocated observations from the ground-based AERONET sun photometer network and against "reference" satellite data provided by MODIS and MISR. Tools used for the evaluation were developed for daily products as produced by the retrieval with a spatial resolution of 10×10 km^2 (Level 2) and daily or monthly aggregates (Level 3). These tools include statistics for L2 and L3 products compared with AERONET, as well as scoring based on spatial and temporal correlations. In this paper we describe their use in a round robin (RR) evaluation of four months of data, one month for each season in 2008. The amount of data was restricted to only four months because of the large effort made to improve the algorithms, and to evaluate the improvement and current status, before larger data sets will be processed. Evaluation criteria are discussed. Results presented show the current status of the European aerosol algorithms in comparison to both AERONET and MODIS and MISR data. The comparison leads to a preliminary conclusion that the scores are similar, including those for the references, but the coverage of AATSR needs to be enhanced and further improvements are possible for most algorithms. None of the algorithms, including the references, outperforms all others everywhere. AATSR data can be used for the retrieval of AOD and AE over land and ocean. PARASOL and one of the MERIS algorithms have been evaluated over ocean only and both algorithms provide good results.

1. Introduction

Satellite-based radiometers and spectrometers have been used for the observation of aerosol properties from space since more than

* Corresponding author at: Finnish Meteorological Institute, Climate Change Unit, P.O. Box 503, FI-00101 Helsinki, Finland. Tel.: +358 50 919 5458 (mobile); fax: +358 29 539 3146.
E-mail address: gerrit.leeuw@fmi.fi (G. de Leeuw).

0034-4257/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.rse.2013.04.023

three decades (e.g., de Leeuw & Kokhanovsky, 2009; Lee, Li, Kim, & Kokhanovsky, 2009). The data have increasingly been used for purposes such as air quality assessment (Hoff & Christopher, 2009; van Donkelaar et al., 2010), emission estimates (Huneeus, Chevallier, & Boucher, 2012), forest fires applications (Kaufman et al., 1998; Labonne, Bréon, & Chevallier, 2007; Sofiev et al., 2009), atmospheric correction of oceanic (Müller et al., in this issue) and terrestrial (Zelazowski, Sayer, Thomas, & Grainger, 2011) observations. In this paper we focus on the use of satellite instruments to provide aerosol observations for climate and climate change studies. In particular eight aerosol retrieval algorithms using data from different instruments, or a combination of instruments, are evaluated for their suitability to produce climate-relevant aerosol parameters. This study was undertaken in the context of the European Space Agency (ESA) Climate Change Initiative (CCI) (Hollmann et al., 2013) project Aerosol-cci (Holzer-Popp et al., 2013). Aerosol-cci focuses on European instruments and the results are evaluated against non-European instruments such as the Moderate Resolution Imaging Spectroradiometer (MODIS), the Multi-angle Imaging SpectroRadiometer (MISR), and model predictions.

Below, a brief overview is presented of the Aerosol-cci project. Background information on aerosols, their effects on climate and the use of satellite-based instruments to obtain information on aerosols is presented in Appendix A. This information has been used to develop algorithms to obtain information on aerosol properties from satellite-based instruments. The eight aerosol retrieval algorithms which were evaluated as part of the Aerosol-cci project are presented in Table 1 and their characteristics are briefly summarized Appendix B. Recent improvements to these algorithms are described in detail in Holzer-Popp et al. (2013) and briefly summarized in Appendix C. The main focus of this paper is the validation and evaluation of the aerosol retrieval algorithms in a round-robin (RR) exercise. Methods used for validation and evaluation are presented in Section 2 and the protocol used in the RR exercise is outlined in Section 3. The results from this exercise are presented in Sections 3.1–3.3 and discussed in Section 4. Conclusions are presented in Section 5.

Algorithms to retrieve aerosol properties from the radiance measured at the top of the atmosphere at different wavelengths, viewing angles and polarization, have been developed to optimally use the available information, based on different physical principles, cf. Kokhanovsky and de Leeuw (2009) and de Leeuw et al. (2011) for detailed descriptions of algorithms used for the retrieval of aerosol properties over land. However, the aerosol optical depth (AOD) obtained from different algorithms may vary widely and some algorithms may perform better than others, even when applied to the same data set from the same instrument. These differences are regionally dependent and there is no single algorithm that outperforms all others everywhere (cf. Kahn et al., 2009; van Donkelaar et al., 2010). The MODIS dark target algorithm (Levy, Remer, Mattoo, Vermote, & Kaufman, 2007) is most often used. It has been validated (Levy et al., 2010), provides two observations daily, each of them with near-global coverage, and the data are easy to access. Nevertheless, there are gaps, e.g., no data are available over bright surfaces such as deserts.

The basis for the assessment of aerosol retrieval algorithms is usually the comparison of the retrieval results, in particular AOD and the variation of AOD for a given pair of wavelengths expressed by the Ångström Exponent (AE), with independent data provided by AERONET, a federated network of ground-based sun photometers (Holben et al., 1998). Ground-based sun photometers provide accurate measurements of AOD (uncertainty ~0.01–0.02, Eck et al., 1999) because they directly observe the attenuation of solar radiation without interference from land surface reflections. The comparison of, e.g., MODIS and MISR AOD with AERONET data shows that the results from each instrument are within specification, and yet there are differences between them (Kahn et al., 2009). The performance of most of the European sensors prior to the start of the Aerosol-cci project was poorer than that of, e.g., MODIS or MISR, as indicated from a comparison of the AOD retrieved using the baseline algorithms with that obtained from either MODIS or MISR and with the AERONET AOD (Holzer-Popp et al., 2013). It is noted here that AERONET data is well screened for the presence of clouds so that a comparison of satellite-retrieved AOD with AERONET data does not provide a good test of the behavior of an algorithm in cases where cloudy pixels have not adequately been removed by cloud flagging.

The Aerosol-cci project was designed to provide essential climate variables (ECVs) for aerosols from satellite data (Holzer-Popp et al., 2013). To achieve this, the quality of current satellite aerosol products needed to be assessed and, when the quality was found to be insufficient, improved. Participating algorithms, focusing on European instruments, are listed in Table 1 and their most important characteristics are summarized in Appendix B. Aerosol parameters retrieved from other instruments such as MODIS or MISR were used for comparison, and this comparison provided a measure for the performance of the Aerosol-cci algorithms and their subsequent improvement. The initial focus of the Aerosol-cci project was on understanding differences between participating algorithms as a basis for their improvement. The baseline algorithms were those that existed at the start of the project and improvements were measured with respect to these, using several different methods described in Section 2. Tests were made for data from a single month (September 2008) as described in Holzer-Popp et al. (2013). The best version, as decided by each earth observation (EO) team for their own algorithm based on these tests, was used in a round robin (RR) test which encompassed four months in 2008

Table 1
Instruments and algorithms participating in the Aerosol-cci project. Providers, products and references for each algorithm are indicated. A brief description for each algorithm and references to full descriptions are provided in Appendix B.

Instrument	Algorithm	Provider	Products											
			Land	Ocean	AOD(n)	Type	FMF	Absorption	Dust	Uncertainty	Quality flag	Altitude	Surface reflectance	Cloud fraction
AATSR	ADV	FMI/UHEL	+	+	3/4	3	+	(+)	−	+	+	−	+	−
	ORAC	Univ Oxford/RAL	+	+	2	1	+	−	−	+	+	−	+	+
	SU	Univ Swansea	+	+	4	1	−	−	−	+	−	−	−	−
AATSR + SCHIAMACHY	SYNAER	DLR	+	+	4	3	+	+	+	+	+	−	+	+
MERIS	ESA standard (v8)	ESA	+	+	3	1	−	−	−	−	−	−	−	+
	BAER	Univ Bremen	+	+	1	0	+	−	−	−	+	−	−	+
	ALAMO	HYGEOS	−	+	2	1	+	−	−	−	−	+	−	−
POLDER	PARASOL	LOA	−	+	3	2	+	−	+	−	+	−	−	−

AATSR: Advanced Along-Track Scanning Radiometer.
SCHIAMACHY: SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY.
MERIS (MEdium Resolution Imaging Spectrometer).
POLDER: POLarization and Directionality of the Earth's Reflectances.
AOD(n), n = number of wavelengths.
Type: number of independent aerosol components which potentially can be retrieved.

(March, June, September and December) representing the different seasons. This paper describes the RR tests and results. Based on the RR exercise, the best algorithm, or combination of algorithms, will be selected to produce the global AOD for the whole year 2008 for further evaluation as regards the use of the products in climate studies. For more information on the aerosol-cci project, see: http://www.esa-aerosol-cci.org/.

The Aerosol-cci project is a consortium including 14 partners coordinated by DLR with FMI providing the science co-leader. The consortium consists of three teams. The EO team is responsible for algorithm development and improvement, the validation team is responsible for the validation and evaluation of the retrieval products, and the system engineering team is responsible for the actual processing of the data series and system design. The validation team is independent from the EO team (different partners) which ensures an independent and unbiased evaluation of the EO products. Furthermore, the validation team includes representatives of the global climate modeling community through AEROCOM and their feedback ensures that products will indeed be useful for climate studies. This aspect has proven to be of crucial value for the improvement of the retrieval algorithms. The system engineering team contributes the experience of data centers and experience with data format and data access issues. The Aerosol-cci project started in July 2010 and has duration of 3 years with a potential extension to 6 years.

2. Aerosol retrieval algorithm validation and evaluation

For validation of the retrieval algorithms used in Aerosol-cci, to evaluate their improvement, to select the most suitable algorithm for ECV production, and to assess the achievements as regards meeting user requirements, independent and objective methods are needed leading to quantitative scores. These scores are obtained by comparison with independent data sets. In this case these are provided by the ground-based sun photometer network AERONET (Holben et al., 1998) against which all satellite results, both those participating in the Aerosol-cci RR and the reference satellite data sets, are evaluated.

Three principal methods are used based on statistics for Level 2 (L2) and Level 3 (L3) products. In Aerosol-cci L2 products are the daily products as produced by the retrieval with a spatial resolution of 10×10 km^2 and L3 are daily or monthly aggregates (also referred to as mean or averaged data) provided on a spatial scale of $1^\circ \times 1^\circ$. The L2 and L3 products are available globally. L2 products contain for each pixel quality flags, or a level of confidence, set by the data provider as well as uncertainty estimates. L3 products contain for each pixel the statistics obtained during the aggregation process, such as standard deviation. In addition to these statistics-based methods, other metrics were used for evaluation such as bias, spatial coverage, number of data points globally and representation of features such as biomass burning aerosol plumes, the occurrence of desert dust, or anthropogenic pollution.

Other validation exercises include studies on uncertainty estimation and studies on the comparison with measurements of aerosol properties at ambient relative humidity (RH) (such as column integrated measurements with associated variations of ambient properties with height) with in situ measurements such as those made in the ground-based networks with controlled RH (cf. Zieger et al., 2011). These exercises were not part of the current RR and will not be reported here.

For the intercomparison of Aerosol-cci data sets, and for the comparison of Aerosol-cci data sets to other data sets in the ICARE archive (e.g., MODIS, model results, etc.), a multi-sensor visualization and analysis tool has been developed. All key parameters of each sensor product can be selected independently for visualization. For each product, a link to the product documentation is provided. For a given parameter, a unique color scale is used for direct visual inter-comparison. The geographic selection, date selection or product selection can be modified independently, while the other two selection criteria remain unchanged. All data sets are displayed in Plate-Carree projection to make inter-comparison and geographic selection straightforward. Aerosol-cci daily and monthly L3 products don't require any reprojection. Aerosol-cci L2 products are originally produced in sinusoidal grid, so they are re-gridded on-the-fly upon selection in the graphical interface. Additional interactive capabilities are available, such as display of data values, or X/Y plot comparison. The multi-sensor visualization and analysis tool is available from http://www.icare.univ-lille1.fr/browse/?project=cci.

An extract tool has been developed to interactively extract Aerosol-cci product values in the vicinity of validation sites. Several validation networks are supported, including AERONET. Single or multiple parameters from Aerosol-cci aerosol products can be selected for extraction. A time period and a search radius can be specified. For each selected validation site, and each overpass of the satellite, all data values found within the specified range are displayed, if any, along with the corresponding acquisition time and pixel location. Those extracted values can be directly compared to validation data off-line. The Aerosol-cci extract tool is available from http://www.icare.univ-lille1.fr/extract/cci.

2.1. L2 statistics

AOD and Ångström exponent of L2 data sets were compared with AERONET data using scatter plots and least-squares fits to the data. The comparisons were made for collocated satellite and AERONET observations, i.e. satellite pixels were selected within a spatial threshold of $+/-35$ km and a time frame of $+/-30$ min from AERONET measurements. Where available (ORAC, ADV and SYNAER), quality flags or confidence indicators in the products were taken into account to select high-quality pixels. Furthermore a distinction was made between retrievals over land and water. Round Robin MERIS (MEdium Resolution Imaging Spectrometer) datasets do not have a water/land flag, therefore the pixels over land and ocean for MERIS Standard and MERIS BAER were selected using the ORAC water/land flag.

2.2. L3 scoring

For L3 scoring, an evaluation routine has been developed to determine for a test data set a performance indicator, for cases when trusted reference data are available. Here the test data are daily L3 satellite data, the reference data are AERONET observations within half an hour of the particular satellite overpass. To simplify comparisons, all sun photometer data were gridded on the spatial $1^\circ \times 1^\circ$ resolution of the satellite data. Although in theory satellites should locally offer more than 100 samples for the four months, the available number of valid data points is smaller due to the presence of clouds. The number of samples is further reduced due to a limited swath (e.g., AATSR and MISR), stringent quality control measures (e.g. SU) or due to limited temporal coverage (e.g., SYNAER). In addition, also AERONET data were not available each day.

The selected performance indicator for L3 evaluation is based on a combination score, which separately investigates temporal variation, spatial variation and bias. The evaluation of the sub-scores is explained in Appendix D. Performance indicators E are defined to range from 0 for 'perfect' to 1 for 'poor'. Conversely, associated scores S ($S = 1 - E$) range from 0 for 'poor' to 1 for 'perfect'. This definition for the scores allows via sub-score multiplication the determination of an overall score S and of an overall performance indicator E.

$$\mathbf{E} = 1 - |\mathbf{S}| \text{ with } \mathbf{S} = \text{sign}\,(\mathbf{E_B}) * (1 - |\mathbf{E_B}|) * (1 - \mathbf{E_S}) * (1 - \mathbf{E_T}), \quad (1)$$

where $\mathbf{E_B}$ is the performance indicator for bias, $\mathbf{E_S}$ is the performance indicator for spatial variability and $\mathbf{E_T}$ is the performance indicator for temporal or seasonal variability. Note, that the sign of the bias defines the sign of the total score **S**. Each of the three sub-scores is based on statistics. Hereby, valid sub-scores require a minimum number of samples. Given sufficient data-pairs for test-data D and reference-data R, the bias

performance indicator $\mathbf{E_B}$ compares sums of associated (value-) ranks of an array that contains all elements from both D and R. If the rank sum associated with elements of D (D_{SUM}) is similar to the rank sum associated with elements of R (R_{SUM}), no bias is determined. However, if the two rank-sums differ, then a bias is identified, in strength and sign.

$$\mathbf{E_B} = w * [(D_{SUM} - R_{SUM})/(D_{SUM} + R_{SUM})], w = [_{IQ-R}D +_{IQ-R}R]/[_{IQ-A}D +_{IQ-A}R]. \quad (2)$$

Based on the average interquartile range (IQ-R) to interquartile average (IQ-A) ratio of both data-sets, a variability factor w is defined. The factor w is applied as weight to the bias performance indicator, to avoid an overemphasis in the case that all individual values are close to their average. The same factor w is also applied to both the spatial variability performance indicator $\mathbf{E_S}$ and the temporal variability performance indicator $\mathbf{E_T}$. The spatial variability performance indicator is based on data-pairs spread spatially at one instance, whereas the temporal variability score is based on time-series data-pairs at one specific location. When sufficient data-pairs are available, rank correlation tests are performed and the resulting rank correlation coefficient R_C defines the performance indicator.

$$\mathbf{E_S} = w * (1 - R_C)/2 \quad (3)$$

$$\mathbf{E_T} = w * (1 - R_C/))/2. \quad (4)$$

With this definition 100% correlation yields zero performance indicator (excellent), whereas 100% anti-correlation yields the maximum performance indicator of 1 (bad). Note, that randomness for temporal and spatial variability yields still scores of 0.5 (not zero).

Performance indicators of any test data, D, with respect to the reference data set, R, are determined in two parallel steps, at the smallest temporal resolution and at smallest temporal scale. In step 1, temporal and bias performance indicators are determined at each location, by applying available time-series data pairs at the smallest temporal resolution. In step 2, spatial and bias performance indicators are determined by exploring data pairs in their spatial context for each time step. The final bias performance indicator is averaged from both processing steps.

Since properties usually vary with longitude/latitude and surface conditions, the evaluation is regionally stratified. These regional scores can be combined via average weighting into a single global score. Thus, this method offers an assessment via a single global (or regional) score, while still maintaining regional diagnostics on bias the ability to match temporal and spatial variability.

2.3. Level3 validation using AEROCOM methods

The L3 validation of daily gridded products using AEROCOM tools is applied to the nearest satellite pixel value on a $1° \times 1°$ grid corresponding to daily mean AERONET values excluding mountain sites. The evaluation with the AEROCOM tools provides bias, histograms, scatter plots, time series graphs, zonal mean comparisons, and score tables. This analysis includes all pixels regardless of quality flags or confidence indicators. A specific focus was put into common data point filters between the AATSR algorithms. The ORAC land/sea mask was used for all retrievals to differentiate between land, coast and sea cases.

3. Round Robin exercise

The Round Robin exercise was set up for an independent and objective evaluation of the global retrieval results (AOD, AE) provided by each of the participating algorithms (Table 1, Appendix B). The versions of the algorithms used to provide these products were selected by each of the retrieval groups based on the exercises described in Holzer-Popp et al. (2013) and summarized in Appendix C. The results were evaluated using the tools described in Section 2. Based on these results, the independent validation team (see Introduction) provided an assessment of the statistical quality. Other considerations were data coverage and spatial patterns. In addition, the same tools (Section 2) were applied to data from MODIS Aqua, MODIS Terra and MISR, for intercomparison and as a measure of how well the Aerosol-cci algorithms perform in comparison to other satellite data sets which are often used in climate studies.

For an objective evaluation of the RR results, a protocol was developed using the following rules:

- evaluation was performed by independent Aerosol-cci partners, i.e. partners not directly involved in providing retrieval data: the validation team (Section 1);
- A set of criteria for selecting the best algorithm was developed beforehand:
  - using the statistics (L2), ranking based on scoring (L3), and L3 validation using AEROCOM tools, as described in Section 2
  - evaluation of performance on global and regional scales
  - evaluation of seasonal performance
  - evaluation of spatial coverage, reproduction of regional and global patterns and the occurrence of features such as desert dust and biomass burning plumes, anthropogenic pollution, etc.

Additional considerations for algorithm selection were:

- long-term application potential (follow-up or predecessor sensors)
- availability/quality of uncertainty information on pixel level
- ability to provide essential complementary data to available satellite data products
- technical criteria such as the operationality of algorithms (e.g., throughput, dependence to systematic external datasets, implementation efforts).

The rankings provided by the validation team, i.e. based on statistical results, are presented in Table 2 and discussed below.

3.1. Level 2 validation

For the L2 evaluation of AOD and AE the statistical measures evaluated were Pearson correlation coefficients, linear fit parameters, standard deviations (from linear fits and from AOD difference histograms), average differences, and number of AERONET sites and satellite pixels used. Examples of scatterplots between the satellite-retrieved AOD and AE vs. AERONET data are shown in Figs. 1 and 2, together with the fit parameters. These figures illustrate that there are differences amongst the several AATSR algorithms, both over ocean and over land, and between the AATSR and PARASOL results over ocean. In all cases, over-ocean satellite AOD is reasonably well correlated with AERONET, although outlyers are observed for ORAC, which may be due to insufficient cloud screening. Over land, SU AOD is well correlated with AERONET and the slope is close to 1, but for ADV and ORAC the correlation is less good than over ocean.

Correlations of AE are much smaller than for AOD, especially over land where in most cases the correlation is poor. Over ocean the correlations are better and the PARASOL AE seems to follow AERONET values reasonably well. It is not clear why the SU results are not at all correlated with AERONET over ocean and AE's are mostly very close to zero.

Criteria used for ranking of the L2 validation results are based on correlation coefficient, standard deviation and number of satellite pixels:

- The closer the linear Pearson correlation coefficient is to 1, the better the algorithm (both for AOD(550 nm) and AE).
- The smaller the standard deviation of the difference between retrieved and AERONET AOD, the better the algorithm (both for AOD(550 nm) and AE).
- Algorithm should provide a large enough number of retrieved pixels.

Table 2
Rankings of the Aerosol-cci algorithms (best = 3): summary of the results from the three independent validation and evaluation methods.

Validation criteria		Algorithm						
		ADV	ORAC	SU	SYNAER	ESA Standard	ALAMO	PARASOL
Algorithm version		v1.3/Set 3D	v1.1b	v3.0	v3.2	v8.0	v1.0	v0.23a
L2 validation results	Land	2	1	3	0	1	–	–
	Ocean	3	2	1	0	1	2	3
AEROCOM tools	Land	3	1	2	0	1	–	–
	Ocean	3	2	1	1	1	2	3
	Coastal	2	1	3	0	1	2	3
L3 scoring		3	1	2	–	–	–	–
Coverage of features (monthly AEROCOM maps)	Land	3	0	2	1	1	–	–
	Ocean	3	0	1	2	1	2	3

The application of the criteria leads to the following rankings for the algorithms using AATSR or MERIS data (see also Table 2):

AATSR over ocean: ADV, ORAC, SU, SYNAER
AATSR over land: SU, ADV, ORAC, SYNAER
MERIS over ocean: ALAMO, ESA standard

It is noted that BAER was not included because no products were available at the time the RR was conducted.

This ranking is based on the statistics provided in Table 3. These statistics show that from all participating algorithms, over ocean PARASOL has the best combination of high correlation, small standard deviation and large pixel number, but also AATSR ADV and MERIS-ALAMO have good correlations. Over land AATSR SU has good correlations, whereas MERIS has clearly weaker correlations and larger standard deviations, with only slightly larger number of pixels.

3.2. L3 scoring

The evaluation of L3 data as described in Section 2.2 was separately conducted for the 25 sub-regions shown in Fig. 3 (TransCom; Gurney et al., 2002). Within each of these 25 regions, at least 10 data-pairs were required for both the spatial and the temporal test in order to get a valid score. This required sufficient satellite data samples and also sufficient $1^\circ \times 1^\circ$ grid boxes in each region with AERONET coverage.

These data-pair requirements permitted only scores for the Northern Hemispheric land regions with sufficient AERONET coverage. Unfortunately, also for these regions collocated satellite and AERONET data were often so sparse that a valid score was not possible. Table 4 shows the resulting satellite AOD retrievals scores for the globe and for North America.

Table 4 indicates that the data volume of some of the Aerosol-cci AOD retrievals for the test period (four months in 2008) which matched to AERONET data is so sparse that no scores can be offered. Even those Aerosol-cci AOD products which allow scoring have much poorer coverage than MODIS and even MISR (which has a smaller swath of about 360 km compared to about 500 km for AATSR). This is also illustrated by the number of samples that contribute to the scores for North America, for which sub-scores for bias, temporal variability and spatial variability are listed.

Among the different Aerosol-cci AOD retrievals the AATSR products show the highest skill but total and sub-scores vary. However, the comparison of the scores is limited. Global scores are based on different

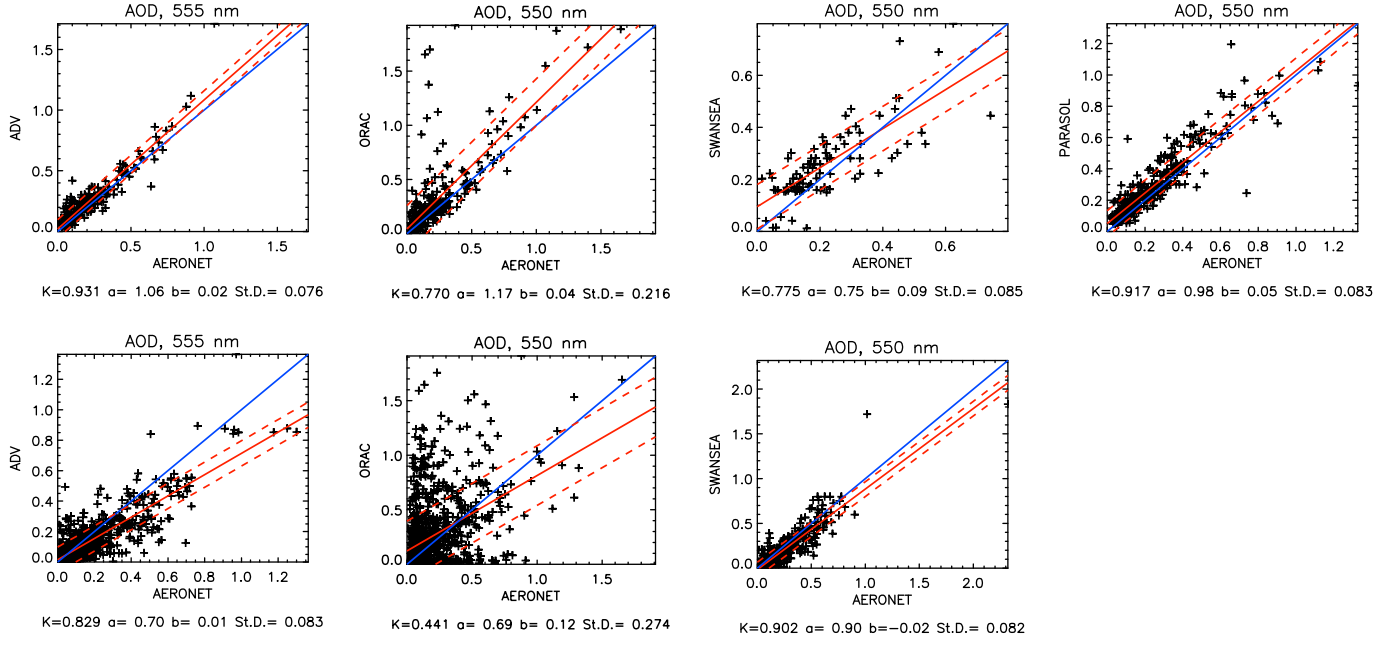

Fig. 1. Examples of scatterplots between the satellite-retrieved AOD vs. AERONET data. AOD over ocean (top row) and land (bottom row) were separated using the ORAC land/sea flag for all retrievals. The algorithm is indicated along the vertical axis. Statistics from a least squares fit of the type $y = ax + b$ are indicated in the legends at the bottom (see also Table 3): K is the correlation coefficient, a is the slope, b is the bias and St.D. is the standard deviation.

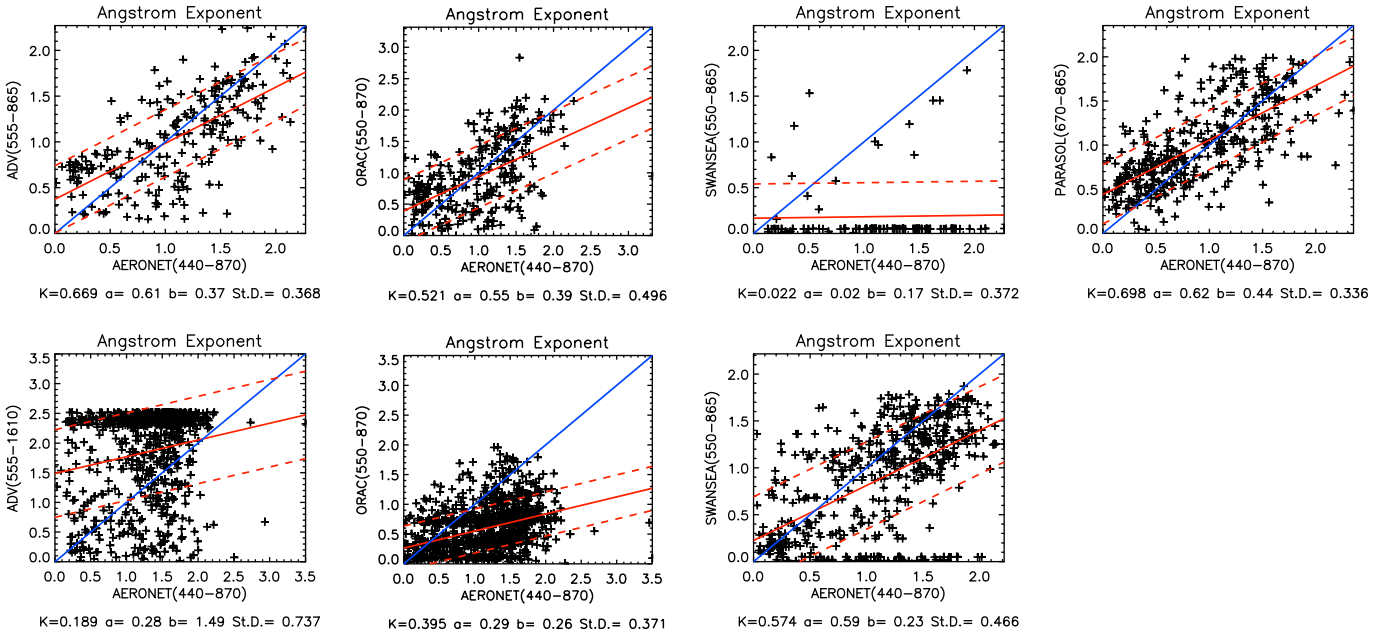

Fig. 2. Examples of scatterplots for Ångström exponents vs. AERONET data, for further explanation see Fig. 1.

numbers of regions. And also in more appropriate comparisons for North America, where almost all products supply a score, the underlying numbers of data-pairs differ.

For North America, ADV is ranked before SU and ORAC. The ADV score (.52) matches the MISR score and both the ADV, MISR and SU (.48) scores are better than the MODIS scores (.42/.39) which are particularly poor here. Looking at the sub-scores, the relatively low ORAC score (.38) has a bias score that is as good as in MISR or MODIS and clearly better than for SU. The sub-scores also indicate that ADV and SU display spatial distributions for North America which are superior among the examined data sets, even better than MISR or MODIS.

Calculated regional performance indicators, as well as contributing performance indicators due to spatial variability among MODIS, MISR and ADV are compared in Fig. 4. The same performance indicator comparisons among the three ATSR products are presented in Fig. 5.

Fig. 4 indicates that the ADV performance indicators for North America and Europe (where a sufficient amount of AATSR data are available) are as low as for MISR and better than for MODIS. However, as mentioned above, the data volume of MISR and ADV is much smaller than that offered by MODIS, mainly due to their narrower swaths. The ADV data volume is similar to that of the MISR data, despite the larger AATSR swath. We note that the value of satellite products is not only determined by its accuracy but also by temporal and spatial coverage.

The comparison of the three AATSR algorithms shows that the data coverage is poor for the SU product. Clearly efforts are needed to increase coverage in order to make these data sets more attractive to users. For regions with available scores it could be concluded that the ADV product scores best and that the ORAC product scores poorest, despite having a relatively low bias performance indicator. Still, this is just based on an analysis for two regions dominated by urban-industrial aerosol and there are many more facets to aerosol (e.g. dominance by dust or biomass burning).

Clearly these initial comparisons leave many open questions. The most worrisome aspect is that there are so many regions where no scores could be calculated for this limited data set. This can be addressed once data are provided for an entire year or more. Also reference data

Table 3
L2 validation statistics.

Instrument:			AATSR				MERIS		PARASOL
Algorithm:			ADV	ORAC	SU	SYNAER	ESA Standard	ALAMO	
Surface	Parameter	Metrix							
Land	AOD	cc	0.83	0.44	0.90	0.59	0.55		
		st dev	0.08	0.27	0.08	0.18	0.14		
		Bias	0.01	0.12	−0.02	0.17	0.17		
	AE	cc	0.19	0.40	0.57	−0.02	0.06		
		st dev	0.74	0.37	0.47	0.39	0.48		
		Bias	1.49	0.26	0.23	1.61	0.54		
		Number of pixels	738	1015	536	200	663		
Ocean	AOD	cc	0.93	0.77	0.78	0.36	0.67	0.82	0.92
		st dev	0.08	0.22	0.09	0.20	0.11	0.11	0.08
		Bias	0.02	0.04	0.09	0.16	0.11	0.15	0.05
	AE	cc	0.67[a]	0.52	0.02	0.25	0.32	0.39	0.70
		st dev	0.37	0.50	0.37	0.30	0.51	0.54	0.34
		Bias	0.37	0.39	0.17	1.54	0.58	1.09	0.44
		Number of pixels	221	285	99	61	262	103	384

AE(555–865).
[a] For ADV, AE(555–1610) yields similar statistics (cc = 0.66, st dev = 0.37, bias = 0.05), but the average AE is lower.

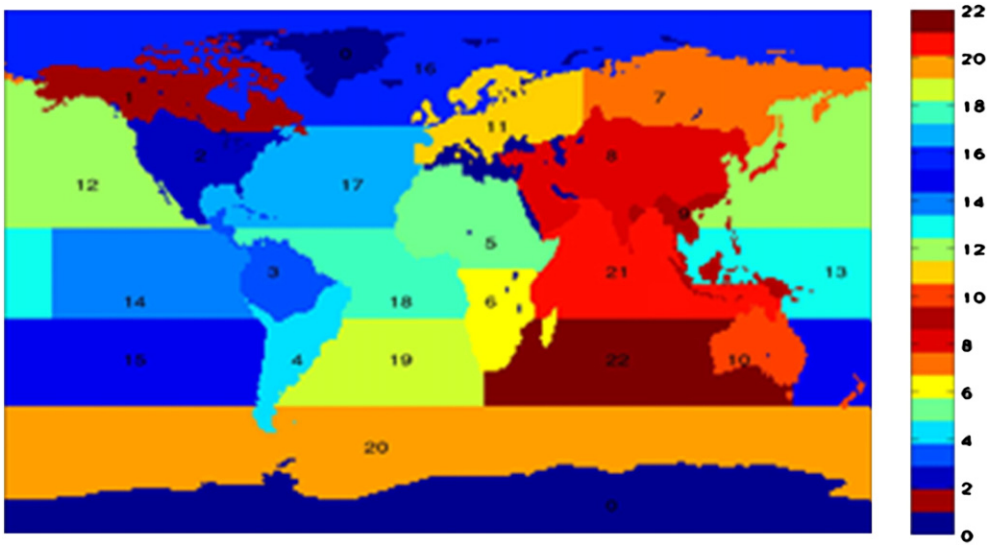

Fig. 3. Regional stratification of the globe following TransCom (Gurney et al., 2002).

over oceans are needed and will be added in future assessments (e.g. using data from the Maritime Aerosol Network (Smirnov et al., 2012) or using trusted and mature satellite AOD products).

3.3. Level3 validation using AEROCOM tools

The AeroCom tools allow for the selection of regions (World, Europe, China, India, E. Asia, N. Africa, N. America, S. America, World w/o mountains) and annual (only a four month average for 2008 in this RR), seasonal (represented by 4 different months), and monthly L3 (4 months in 2008) averages. A common 1° × 1° mask was established where valid data were available from all retrieval algorithms (AATSR and MERIS). In all regions this info is further refined using an ocean, coastal and land mask, based on whether a grid point was identified as purely ocean or land (using the ORAC L2 land/sea mask). Remaining grid points are defined as coastal. Altogether 8 × 4 regional selections are possible. For comparison, similar statistics are available for the annual averages of MODIS-Terra, MODIS-Aqua and MISR AOD data, with selection by region. For each selection a list is produced showing the statistics, cf. Table 5 as an example. Examples of the results are presented in Figs. 6–12.

3.3.1. Global AOD maps

Fig. 6 shows the global annual mean AOD maps for the algorithms participating in the Aerosol-cci RR, as well as reference AOD maps from MODIS v5.1 (Terra and Aqua), MISR and the AEROCOM median. As compared with the baseline algorithms (Holzer-Popp et al., 2013), the current results are much closer to each other and also closer to the references. Yet, also quite large differences are observed, both as regards the global coverage, the number of valid pixels (provided with the statistics given with Fig. 8), the spatial distributions and the features in each of the maps. Clearly, ADV provides the smallest global coverage, which is also reflected in the number of valid pixels which is smaller than for ORAC and MERIS Standard, but larger than for SU. The small number of ADV pixels is due to the facts that (a) ADV limits the retrieval to solar zenith angles smaller than 65° and (b) no retrieval is made over very bright surfaces (see discussion). The even smaller number of pixels provided by SU, in spite of the larger global coverage, is due to a stricter quality control. Further, there are clear differences in the global mean AOD (provided for each algorithm in the legend at the top at the right), which vary from 0.154 for ADV to 0.215 for SYNAER, as compared to MODIS mean AOD values of 0.189 (Terra) and 0.179 (Aqua) and MISR (0.176).

Over land there are clear differences in the AOD distributions, such as at high northern latitudes where the AOD provided by ORAC and MERIS Standard are clearly higher, SYNAER is a bit higher, and ADV provides distributions similar to those from the reference satellites. SU, on the other hand, provides AODs which are substantially lower. It is noted that the AEROCOM median shows somewhat lower AODs at

Table 4
Comparison of scores for different AOD satellite retrievals for year 2008 data for the months March, July, September and December. The larger the absolute value of the score, the better the performance, with the overall sign indicating the bias vs. AERONET. The left side presents scores for the globe, land and oceans. The right side presents total and sub-scores for North America. The sub-scores for North America were added because the global scores are difficult to compare as the number of contributing regions differs. For North America most data provide a score. Note that the total number of regions for which a score would be possible is 25 (Fig. 3). Also note that even for North America the number of data pairs varies strongly and is for some Aerosol_cci data so small that no score can be provided (regions contributing given in columns "areas" for global and "data pairs" for North America). For North America the scores are broken down to the sub scores for bias, temporal correlation and spatial correlation. No scores are given for SYNAER and ALAMO because the number of samples was too small.

	10+ samples	Global scores				North American scores					
		Ocean & land	Ocean	Land	No. of regions	Total	Bias	Temporal	No. of pairs	Spatial	No. of pairs
Reference	MISR v22	−.59	−.62	.57	4	.52	.88	.80	43	.74	33
	MODIS aqua	.58	.59	.58	13	.42	.85	.79	109	.62	105
	MODIS terra	.58	.59	.58	13	.39	.84	.77	111	.60	104
	SeaWiFS	−.55	.57	−.55	7	−.49	.84	−.81	57	.73	63
	OMI_UV	.43	.45	.40	13	.28	.72	.70	104	.57	103
AATSR	ADV v13	−.54	−.57	−.50	4	−.52	−.81	.77	42	.84	27
	SU v30	−.48	−.49	−.49	1	−.49	−.77	.76	27	.84	22
	ORAC v11	.33	.33	−.37	4	−.38	−.86	.62	56	.71	54
MERIS	Std	.37	.38	.38	1	.38	.69	.81.	22	.68	13
	Parasol v23	–	.65	–	3	–	–	–		–	–

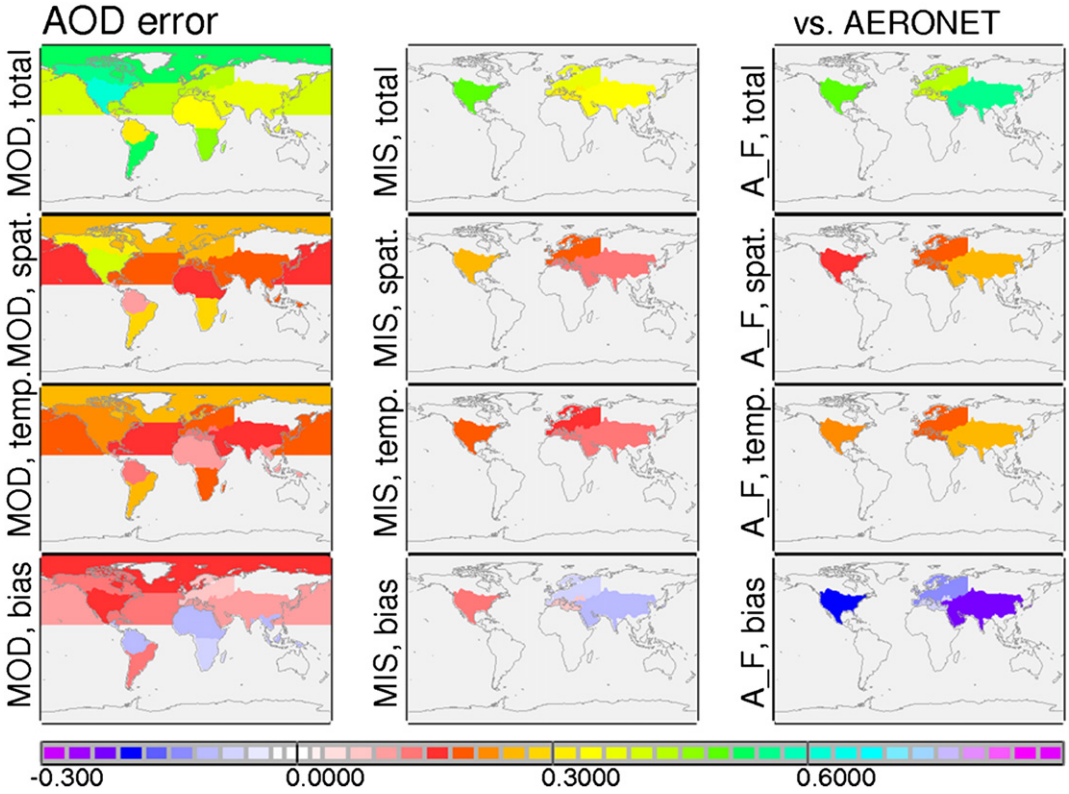

Fig. 4. Regional performance indicator comparisons between MODIS (left), MISR(center) and ADV (right). The first row shows the total performance indicator (E), the second row shows the spatial performance indicator E_S, the third one the temporal performance indicator E_T and the bottom row shows the bias performance indicator E_B along with the sign (positive or negative). Performance indicators are only displayed for regions with sufficient data-pairs to co-incident AERONET sun-photometer samples.

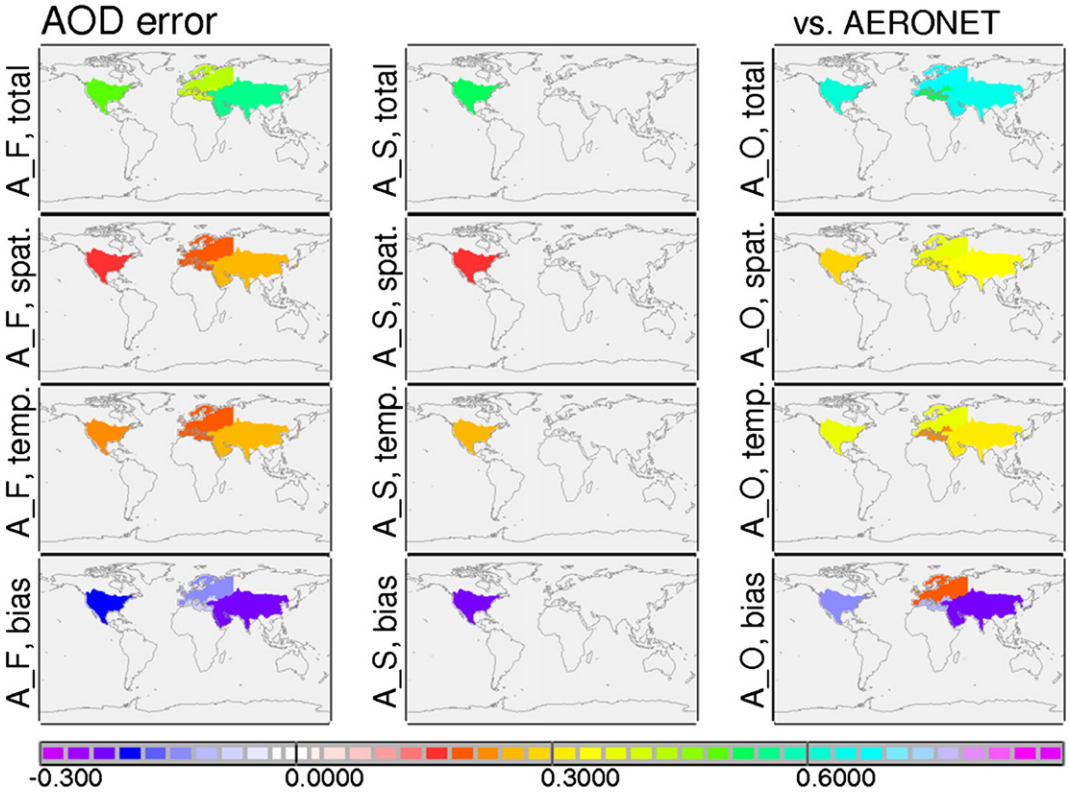

Fig. 5. Regional performance indicator (E) comparisons between AATSR algorithms ADV (left), SU (center) and ORAC (right). The first row shows the total performance indicator (E), the second row shows the spatial performance indicator E_S, the third one the temporal performance indicator E_T and the bottom row shows the bias performance indicator E_B along with the sign (positive or negative). Performance indicators are only displayed for regions with sufficient data-pairs to co-incident AERONET sun-photometer samples.

Table 5
Statistics from the L3 AEROCOM validation for the World Annual Average (4 months in 2008) AATSR ADV.v1.3-Set3D AOD data versus AERONET observations.

Number of valid observations	296
Mean AOD from AATSR	0.217
AERONET mean AOD	0.178
Spearman rank correlation	0.730
Pearson correlation coefficient	0.820
Spatial yearly mean correlation coefficient	0.784
RMS error	0.679
Slope fit forced through zero	0.102
Regression coefficient, slope	1.164
Regression constant, offset	1.060
STDDEV (Model) / STDDEV(Data)	0.028
Score (mean relative bias)	32%
Taylor score	0.897

northern latitudes, with a clear gradient over Siberia, than the reference satellites. Over western Europe, most AOD maps show enhanced values, higher than further north, except ORAC and MERIS Standard, while the reference satellites show no enhancement over western Europe with respect to northern latitudes, in contrast to the AEROCOM median. Over N. America the patterns are quite different between different satellites (both Aerosol-cci and references). The AOD is lower in the west for ADV, ORAC and SU, but there are clear differences between these algorithms as regards the patterns, and the values for ORAC are higher all over the continent. The lower AOD in the western USA is in agreement with the AEROCOM median. In contrast, the AOD in the west is higher than in the eastern USA for SYNAER and MERIS Standard, and this is also observed, although less clear, for MODIS while MISR shows no clear differences across the USA. It is noted that differences between MODIS and MISR AOD observations have been reported; e.g., van Donkelaar et al. (2010) noted that over the SW USA a large AOD enhancement was observed in the MODIS retrievals but not from MISR. Also over S. America there are large differences with a very high AOD over the northern part from ORAC and an overall high AOD from SYNAER and MERIS Standard. SU shows the largest spatial variations and ADV and the reference satellites are quite close in their AOD values with little or no gradients (on the scale on which AOD is displayed). Similar comments can be made over Africa, where all retrievals clearly show the biomass burning plumes, but with different intensity. Also there are clear indications of the Sahara desert dust plumes, however, the analysis of differences between algorithms is difficult because several algorithms do not provide data over bright surfaces such as the Sahara.

Over ocean there are also considerable differences. ORAC provides a clear pattern with very low AODs over most of the southern oceans and a transition across the tropics to the northern hemisphere. The low AOD values over ocean are in line with values reported by Smirnov et al. (2012) based on hand-held sun photometer observations on ships of opportunity as reported in the Maritime Aerosol Network (MAN). Unfortunately the MAN observations for 2008 were too sparse to be used in the Aerosol-cci RR validation. Low AOD values, but much less prominent, over the southern oceans are also observed in the SYNAER and MERIS ALAMO AOD maps, and in the southern Pacific in the AEROCOM median. Also MODIS Aqua, MERIS Standard and ADV indicate low AOD in the southern Pacific. AEROCOM further shows a clear band with enhanced AOD in the southern hemisphere between roughly 40° and 60°, which is reproduced to some extent by ORAC, somewhat less clear by ALAMO and weakly by the reference satellites.

The overall picture emerging from these maps is that the ADV AOD distribution is closest to that of the reference satellites, both over land and ocean, but ADV does not provide data at the higher latitudes resulting in a global coverage which is much less (ca. 30%) than for some other Aerosol-cci algorithms. The global mean AOD produced by the algorithms varies and very large differences occur locally; these local differences are to some extent canceled in the global mean. Therefore it is useful to also look at regional differences to learn the strengths and weaknesses of each algorithm and thus improve the algorithms. Features over land, such as forest fire, desert dust and anthropogenic

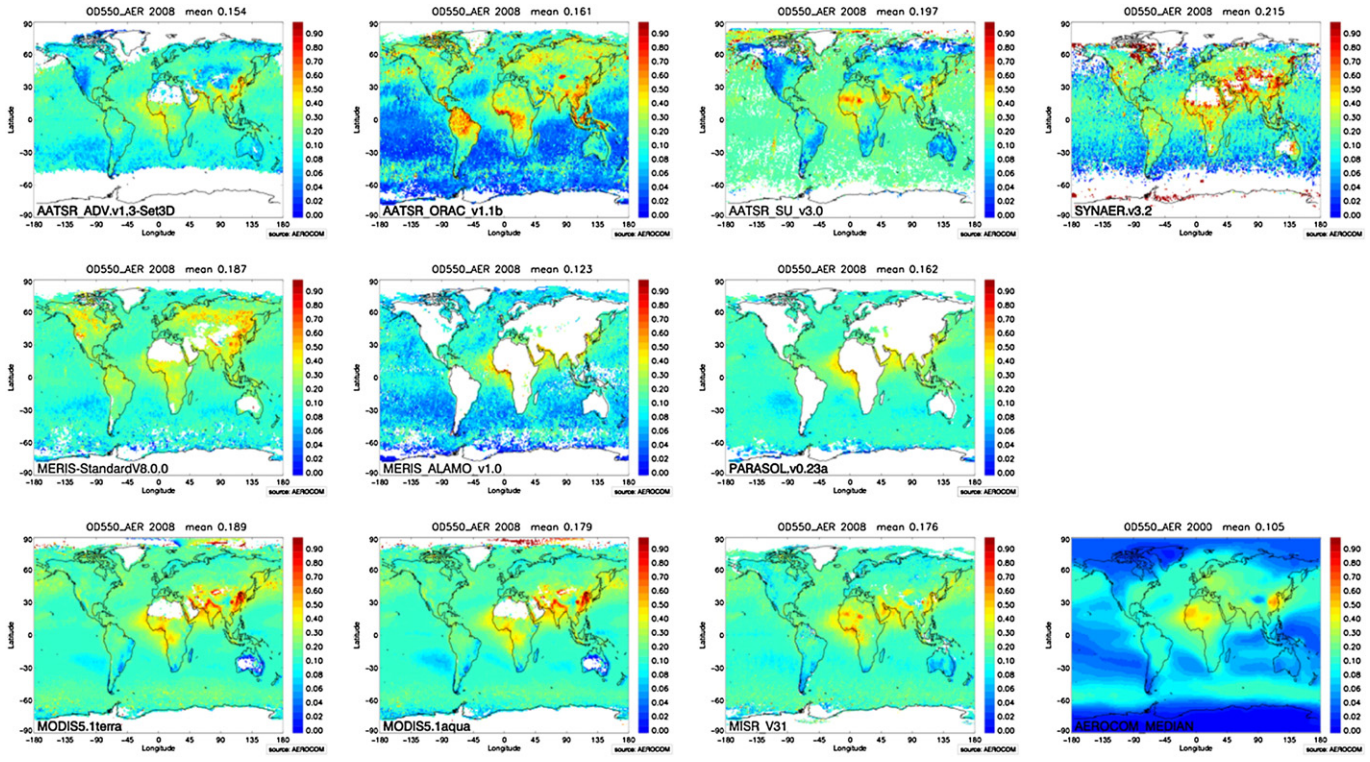

Fig. 6. Global annual mean AOD maps for all algorithms participating in Aerosol-cci as well as reference maps. Top: AATSR ADV, ORAC, SU, SYNAER; middle: MERIS Standard, ALAMO, PARASOL; bottom, MODIS Terra, MODIS Aqua; MISR, AEROCOM median. Note that ALAMO and PARASOL are only used over ocean.

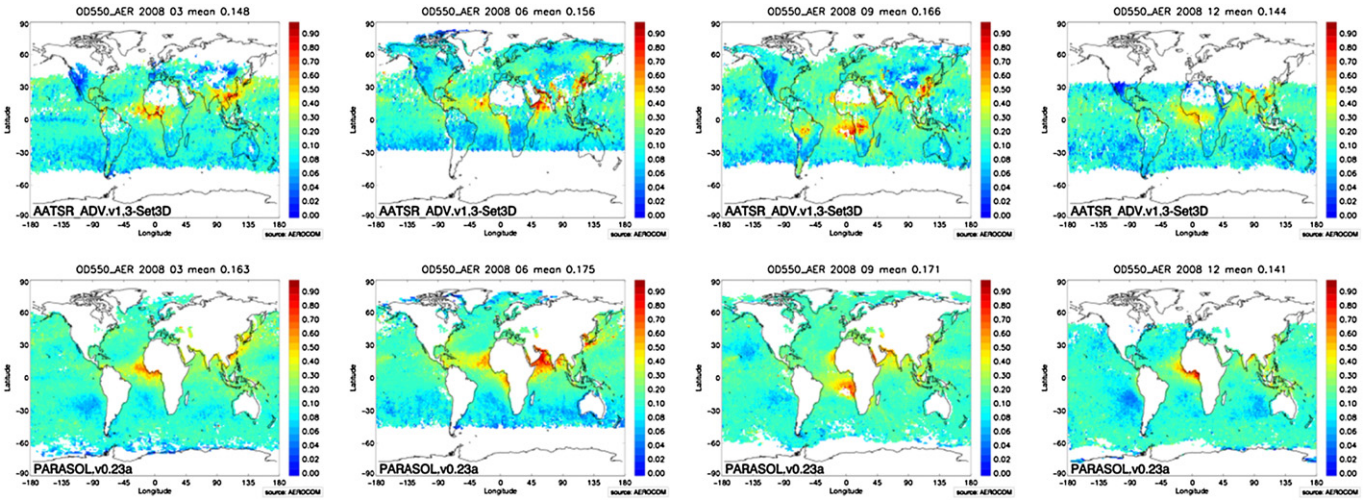

Fig. 7. Global AOD maps of the AOD retrieved using ADV for AATSR data (top row) and for PARASOL data (bottom row) for the months March, June, September and December (left to right), selected in this RR exercise to cover one month in each season.

pollution plumes usually are smoothly extended over ocean but in many cases land-sea transitions are visible. This is clearly a point for future research.

3.3.2. Monthly AOD maps

Monthly AOD maps, for one month in each season selected for this RR exercise, are presented for ADV and PARASOL in Fig. 7. The features are similar to those discussed in Fig. 6, but there are clear differences between seasons in relation to the production and removal of different aerosol types. This is most clearly illustrated with the biomass and desert dust plumes generated over Africa and transported over the Atlantic Ocean. There are also clear differences in the AOD distributions over the continents such as over Asia (China, India, deserts) and adjacent downwind oceans. Differences are also visible over N. America (features discussed in connection with Fig. 6) and over S. America which is likely connected with biomass burning in Amazonia. In addition, differences in coverage occur due to seasonal variation of the solar zenith angle.

3.3.3. Statistical evaluation of AOD retrieval results versus AERONET AOD

Examples of the AEROCOM statistical analysis of the Aerosol-cci results for the 4 months in 2008 are presented in Figs. 8, 9 and 10. Figs. 8 and 10 include MODIS Terra evaluation results as a reference. MODIS Terra was selected here because the overpass time is close to that of ENVISAT with AATSR and MERIS. Fig. 8 shows scatterplots of the retrieved AOD vs. AERONET values. The statistics are provided in the legend in the upper left corner of each plot, for clarity they are also reproduced in Table 6. The algorithm name is given along the vertical axis of each plot. The scatterplots illustrate the differences between the various algorithms and how much they deviate from the reference value. These differences are quantified, in a statistical sense, by the correlation coefficient, the bias and the rms. Fig. 8 shows scatterplots including data for all 4 months considered in Aerosol-cci for the whole globe, i.e. including land, ocean and coastal regions, whereas the bar charts in Fig. 9 differentiate between land and coastal for each month separately; there are not enough L3 collocations over ocean to provide meaningful statistics. The data shown in Fig. 9 have been used to provide a ranking between the four AATSR algorithms, see Table 7. The numbers in Table 7 are the number of months, out of a total of 4, when a certain algorithm performed best, 2nd best etc. based on two statistical parameters: correlation and RMS. The results show that over land ADV provides the best results, before SU, and in coastal areas SU ranks before ADV. ORAC is sometimes close.

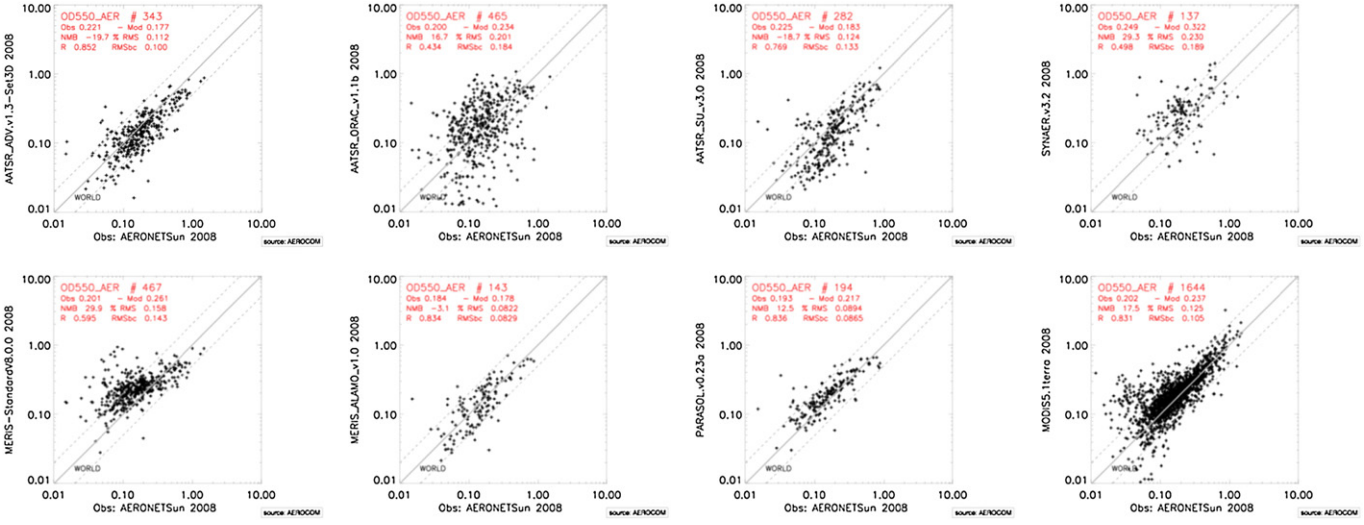

Fig. 8. Scatterplots of the AOD provided by each Aerosol-cci algorithm and by MODIS Terra, identified on the vertical axis of each figure plotted on a log-log scale versus AERONET AOD. The lines indicate the 1:1 and the factor 2 limits. Results from a least squares fit are provided in the legend at the upper left corner of each figure and are summarized in Table 6.

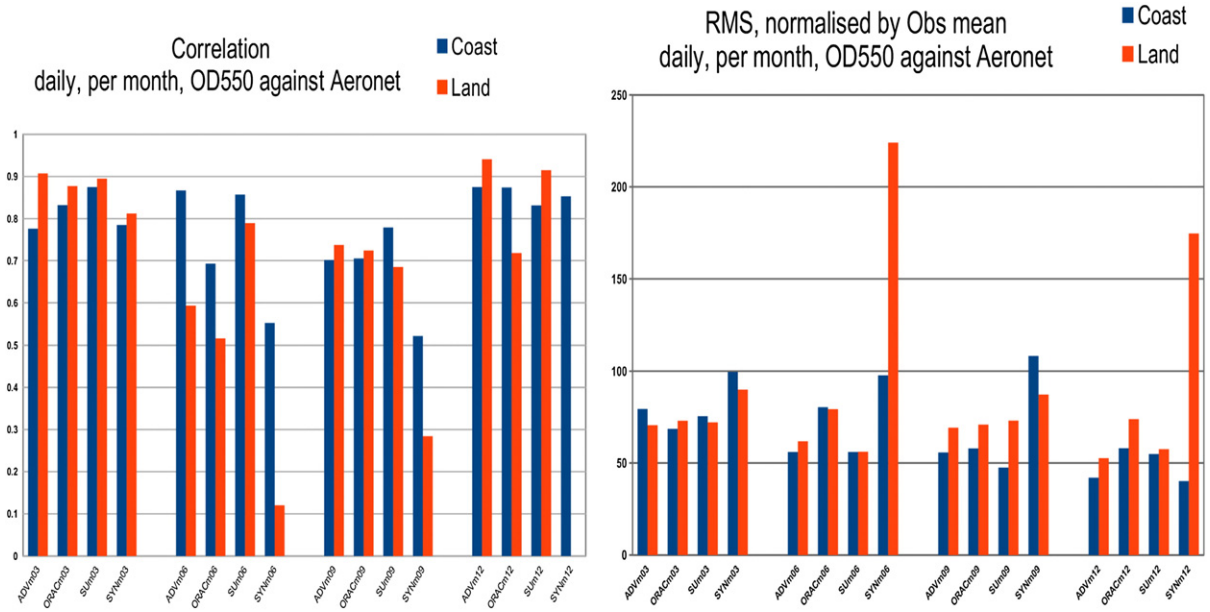

Fig. 9. Bar charts showing the seasonal variation (as given by 1 month in each season) of the correlation coefficient and RMS over land and in coastal regions for the 4 algorithms using AATSR data. Over ocean there are too few collocations to provide meaningful statistics for each month.

These numbers can therefore be used to provide a ranking, however, the statistics also provide a quantitative number, in a statistical sense, showing how large (or small) the differences between the algorithms are, which has been used to provide the ranking presented below.

Fig. 10 shows the statistics in a different way, as histograms of the frequency of occurrence of the AOD values retrieved from the satellite observations, compared with collocated AERONET observations. Ideally, the two curves should exactly coincide. However, even MODIS Terra, with 1644 collocations does not provide an exact coincidence and the lower AODs are on average somewhat overestimated whereas the higher AODs (around 0.2) are somewhat underestimated by MODIS Terra. For the Aerosol-cci algorithms, covering only 4 months and thus having many fewer collocations, the histograms show larger variations between bins. Yet, ALAMO and PARASOL, with over-ocean retrieval only, follow the AERONET pattern quite well, with a tendency for PARASOL to underestimate the lowest AODs. It is noted here that only AERONET data (i.e. land based) were used in this analysis, i.e. sun photometers situated at or near the coasts, which may result in some bias. From the other algorithms, the ADV-retrieved AOD histogram follows that of AERONET quite well In the other algorithms deviations are visible with either overestimation of the lower AOD values (SU) or underestimation (ORAC) whereas MERIS Standard appears to have the largest deviations for both small and large AOD.

3.3.4. Algorithm performance

Another way to use the statistics is to evaluate where algorithms perform well and where improvements are needed. An example is presented in Fig. 11 where the difference between the satellite and AERONET AOD observations, given by (Satellite − AERONET) / AERONET ∗ 100, is color-coded on the map for individual AERONET stations across the world. Blue indicates that the satellite is underestimating; red indicates that the satellite is overestimating. Light colors indicate that the differences are very small and as the color is darker the under or

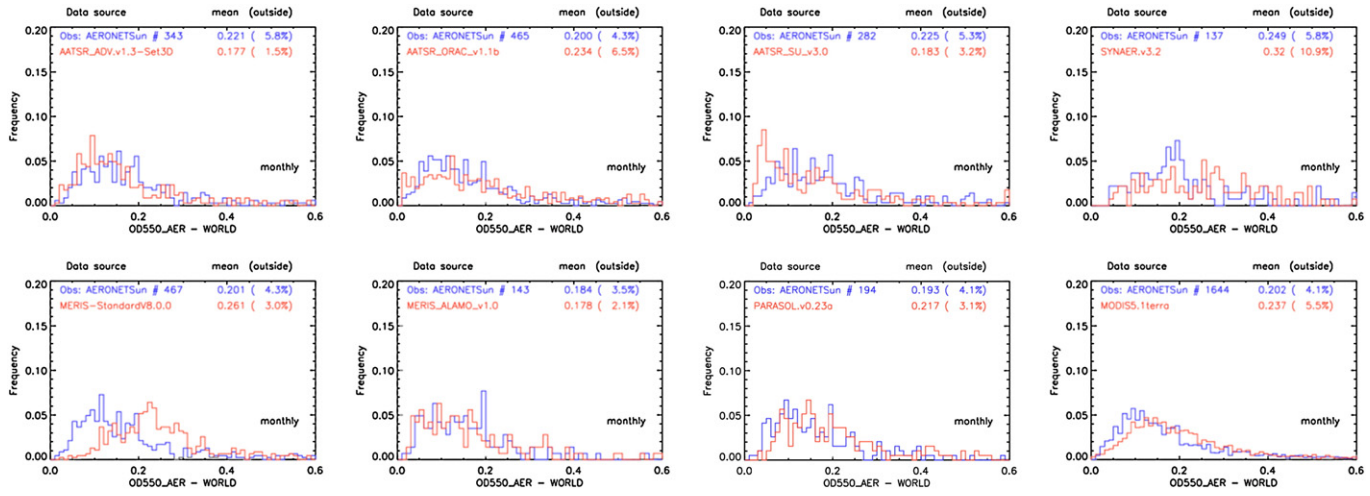

Fig. 10. Histograms showing the frequency of occurrence of the AOD values provided by each Aerosol-cci RR algorithm and MODIS Terra (in red), together with that of the AERONET reference values (in blue), together with the global mean value and the percentage of points with an AOD of greater than 0.6 which are not shown in the plot (indicated as "outside"). The number of collocations is given in the top left legend, top line, the algorithm is indicated in the second line.

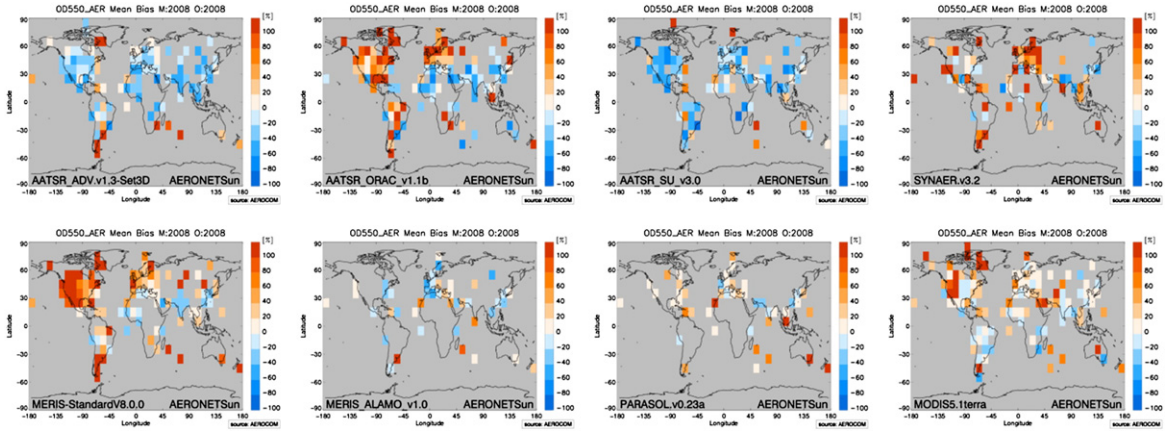

Fig. 11. Evaluation of the performance of the Aerosol-cci algorithms and, for comparison, MODIS Terra. The plots show the difference between the satellite-derived AOD and the AERONET AOD, as explained in the text, the color scales are given to the right of each map.

over-estimation is larger. For these plots, the globe has been gridded into boxes of 10° × 10°. For each grid box, the stations located there are taken into account, and their data is then averaged depending on the time period to be plotted. When smaller regions are considered instead of the whole globe, the grid is reduced to a 1° × 1° grid. The plot does not show how many stations are within the grid box. The locations of the grid boxes (especially at the plot for the whole globe) are not entirely correct, because while the outline of the continents and countries are correct according to the map projection, the grid boxes are not. This would require a reverse map projection which would destroy the analysis grid because of associated interpolations. The error becomes greater as the boxes are further away from the equator and the zero meridian.

The maps show that none of the algorithms, including the MODIS reference, is perfect everywhere. Improvements are needed, but areas for which the improvements are needed, and in which direction (under- or over-estimation), are different for each algorithm. Taking the AATSR algorithms as an example, ADV and SU appear to perform reasonably well over Europe, even though they tend to underestimate, whereas ORAC and SYNAER have large overestimation. The same pattern emerges over the USA, except that SYNAER appears to work quite well over the eastern USA. Almost all algorithms show a large overestimation at high latitudes, except ALAMO in the northern hemisphere. Because AERONET stations are located over land, or in coastal areas, this evaluation cannot be made over ocean.

Fig. 12 shows the zonal mean AOD for the Aerosol-cci algorithms and MODIS Terra, with AERONET for comparison. Together with Fig. 10, this figure illustrates the performance of each algorithm. As in Fig. 10, ideally the satellite-retrieved AOD would follow the AERONET observations, as for MODIS Terra in Fig. 12. Also the over-ocean only AODs provided by ALAMO and PARASOL show a quite good behavior. However, for the other algorithms, which include both land and ocean in the plots in Fig. 12, the trends are reproduced well with high AOD north of the equator and lower AODs toward the poles, but quantitatively there are differences. MERIS Standard is in the (sub-) tropics very close to AERONET, but at mid-latitudes the AODs are much higher than those from AERONET. Similar observations can be made for SYNAER, but the SYNAER AOD shows an increasing trend from south to north as opposed to all other observations. From the other AATSR algorithms, ADV deviates quantitatively most from AERONET, whereas ORAC follows AERONET quite well but peaks right at the equator and has much higher values both at −60° and in the far north. SU seems to give the best performance in this comparison except for the very high values in the far north.

The overall ranking resulting from the evaluation with the AEROCOM tools is given in Table 2.

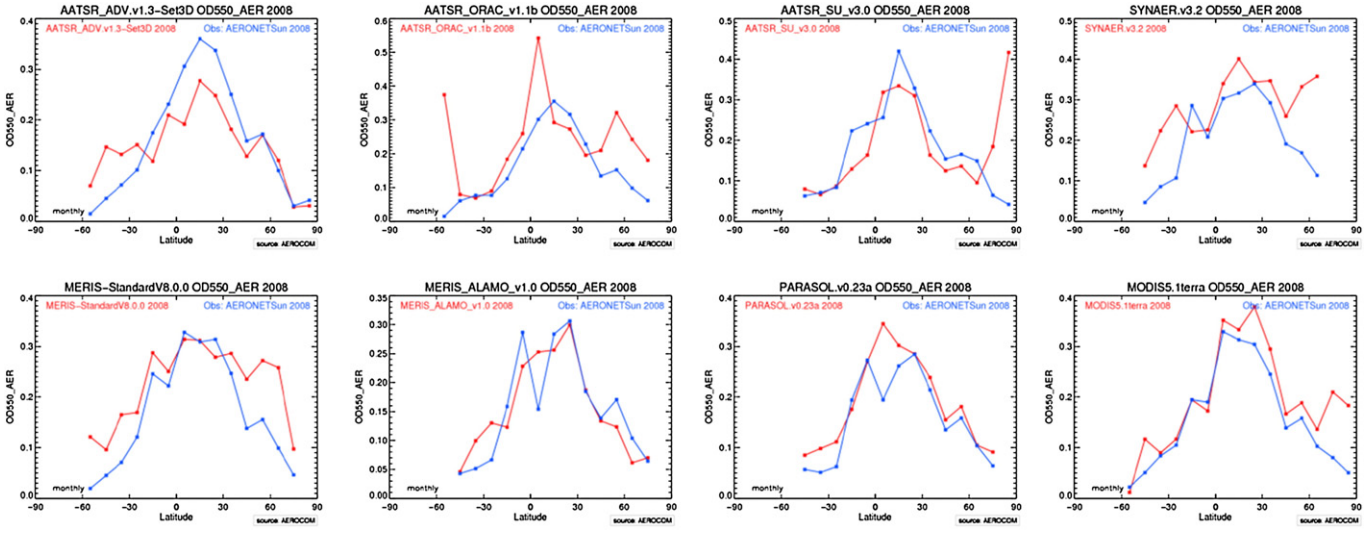

Fig. 12. Zonal mean AOD for the 4 months in 2008 for the Aerosol-cci algorithms and MODIS Terra (all in blue) and AERONET (in red).

Table 6
L3 statistical evaluation results (Fig. 8).

Algorithm	Number of collocations	Global mean AOD AERONET	Global mean AOD satellite	NMB (%)	RMS	R	RMS-bc
AATSR-ADV	343	.221	.177	−19.7	.112	.852	.100
AATSR-ORAC	464	.200	.234	16.7	.201	.434	.184
AATSR-SU	282	.225	.183	−18.7	.124	.769	.133
SYNAER	137	.249	.322	29.3	.230	.498	.189
MERIS-Standard	467	.201	.261	29.9	.158	.595	.143
MERIS-ALAMO	143	.184	.178	−3.1	.0822	.834	.0829
PARASOL	194	.193	.217	12.5	.0894	.836	.0865
MODIS Terra	1644	.202	.237	17.5	.125	.831	.105

4. Discussion

The combined effort of European aerosol retrieval teams, supported by MODIS and MISR retrieval specialists participating in workshops and discussion meetings, has resulted in an enormous improvement of the retrieval algorithms and the products resulting from them. These efforts have been described in Holzer-Popp et al. (2013) and are briefly summarized in Appendix C. This paper is focused on further improvement and algorithm inter-comparison with the goal to use the algorithms for climate studies. This application requires a very high accuracy as formulated by climate users, and the inclusion of uncertainties per pixel. To evaluate algorithm performance, as judged by the evaluation of their products, in this case mainly the AOD, methods have been developed as described in Section 2. AE was validated only for L2 and the results were mostly not good enough to justify the use of AE L3 data in the RR. The evaluation of the results is based on statistical methods applied on regional and global scales and subjective but informed methods based on existing knowledge of how aerosol properties vary. They show, quantitatively, the good performance of the PARASOL (v0.23a) and the MERIS ALAMO (v1.0) algorithms over ocean, and the improvement of the AATSR ADV, SU and ORAC algorithms for use over both land and ocean. The other algorithms, SYNAER and MERIS Standard, need further improvement before they can be used to provide parameters useful for climate studies. MERIS BAER needs further improvement with respect to cloud screening and the consideration of absorbing aerosols. This situation led to the conclusion that, in view of their good performance, the PARASOL and the MERIS ALAMO algorithms can be used for the retrieval of AOD over ocean and thus provide 10 years (MERIS) and 7 years and more (PARASOL) global data series.

For AATSR, all three algorithms using only AATSR data (ADV, ORAC, SU) show good performance, although there are regional and seasonal differences. However, there is not one algorithm which performs best everywhere, as shown by the rankings provided in Section 3 from each of the three different methods. Overall, ADV appears to provide the best scores, and compares most favorably with the reference satellite data sets, but although retrievals over very bright surfaces are made, no results are reported when the surface reflectance in the 1.6 μm wavelength band of the nadir camera exceeds 0.45. SU does provide retrievals over highly reflecting surfaces but the number of data points is very small, mainly due to the application of stricter quality control than ORAC and ADV. However, this stricter quality control does not lead to the highest scores everywhere. ORAC is potentially the most consistent algorithm in a mathematical sense, however both statistically and as regards the reproduction of features it performs less well than ADV and SU. Yet, the low AOD over ocean seems to be in line with results published by Smirnov et al. (2012). Nevertheless, ORAC does not rank highest over ocean, which may be due to the lack of independent validation data over open ocean which could confirm the low AOD observed by ORAC. Based on the current RR results, ADV ranks first, followed by SU with ORAC as third. Yet, the differences are so small, that the ranking may change when further improvements are implemented. Furthermore, the ranking may be influenced by uncertainties introduced by L3 sampling methods as discussed in Sayer, Thomas, Palmer, and Grainger (2010).

Much of the difference between algorithms and their scoring may be due to cloud masking. The different cloud masks used by each of the algorithms slightly complicates the like for like comparison especially as common filtering may not completely account for possible differences in cloud masks or thresholds used at the 10×10 km^2 retrieval level. As pointed out in the introduction, the comparison with AERONET data, which are well screened for cloud occurrence, does not provide a good test for how well clouds have been detected in satellite data and the results may be influenced by the occurrence of residual clouds.

Several recommendations resulted from the RR exercise. One of them was that, although the ADV algorithm overall ranking was best, the coverage was a problem and needed to be improved. Such improvement could be found from increasing the maximum solar zenith angle used in the ADV retrieval from 60° to 75° which would give a similar coverage as other algorithms and from the implementation of a module to model the reflectance of bright surfaces such as applied by SU. The latter has been implemented in ORAC V2, together with the ADV post-processing step (see Appendix B) and initial results are better than those described here. SU in turn has increased the number of pixels retrieved by using a less strict quality control together with the ADV post-processing. The implementation of these changes requires thorough testing and evaluation of the results to avoid loss of accuracy and production of erroneous results. Hence such improvements will be reported in subsequent papers. Thus a similar round robin exercise should be repeated with improved algorithms and the current conclusions should be regarded as a snapshot evaluation of a continuous algorithm development process.

Table 7
Ranking by level3 AEROCOM analysis for land and coastal areas. For ocean the number of collocations is too small to provide meaningful scores. See text for further explanation.

	Best	2nd best	Third	Fourth
Based on correlation				
Coast				
SU	2	1	0	1
ADV	2	0	1	1
ORAC	1	2	1	0
SYNAER	0	0	2	2
Land				
ADV	3	1	0	0
SU	1	2	1	0
ORAC	0	1	3	0
SYNAER	0	0	0	4
Based on rmse				
Coast				
SU	3	0	1	0
ADV	2	1	1	0
ORAC	0	1	2	1
SYNAER	0	1	0	3
Land				
ADV	3	1	0	0
SU	1	1	2	0
ORAC	1	0	0	3
SYNAER	0	1	2	1

5. Conclusions

A validation protocol and necessary tools to implement the protocol have been developed and were applied to eight algorithms for aerosol retrieval using AATSR (4, one synergistic with SCIAMACHY), MERIS (3) and PARASOL data. For reference, these tools were also applied to MODIS and MISR data. The application of these tools, to L2 and to L3 data using different statistical methods and scoring based on a combination of methods, revealed the strengths and weaknesses of each algorithm as well as a scoring of both the Aerosol-cci and the reference algorithms. A crucial issue is the dependence of validation scores on data filtering — this led to the development of a common point filter to assure the comparison of equivalent datasets.

The results show that PARASOL has the highest accuracy over ocean and covers features well. The AATSR algorithm ranking depends critically on filtering. Overall (features, validation) ADV and SU seem better than ORAC which provides some unrealistically high features. SU and ADV scores are similar over land with SU providing data over bright surfaces and ADV having a better coverage of features. Over ocean ADV seems best (except coasts). ADV data complemented with SU data over bright surfaces and coastal areas could provide best products. MERIS ALAMO performs well over ocean; MERIS standard has large overestimations of AOD over land. SYNAER overestimates the AOD, has lower coverage and accuracy but a rather good coverage of features (except in central Asia and high latitudes).

The scoring method shows that the AATSR algorithm results are close to or somewhat better than those from MODIS (and close or similar to MISR), but the number of points retrieved is much smaller than MODIS due to swath width. Obviously, this gap cannot be closed. However, the dual view provided by AATSR makes this instrument potentially better suited for aerosol retrieval over land. Also, it provides one of the longer time series with the combination of ATSR-2 and AATSR (1995–April 2012), continued with data from SLSTR (Sea and Land Surface Temperature Radiometer) planned to be launched in April 2014 as part of ESA's Sentinel-3 satellite. Sentinel-3 also has OLCI (Ocean Land Colour Instrument), which will extend the 10 years of MERIS observations.

Taking into account the results from the RR exercise, the improved algorithms will be used to provide a 1-year data set (2008) of Aerosol ECV products which, after validation using similar tools as described in this paper, will be offered to the climate modeling community for their validation and feedback as regards the use for climate studies. Taking these into account, the full 17 years of ATSR-2/AATSR is planned to be processed.

A round robin exercise for aerosol ECVs cannot be conducted using a fully automatic scoring since trade-offs between coverage and accuracy or between added value and accuracy need to be made. This requires scientific expertise and a team dialog to come up with conclusions which meet the standards of peer review by the scientific community. A strong user involvement in the whole validation and selection process is crucial to understand and take into account the user needs.

The cooperation of the EO community with the global modeling community has proven to be very important, in particular as regards the production of a data set in such a way that they are indeed useful for climate studies. The cooperation between the EO groups, which was the first time on a European level, has led to large improvement of almost all retrieval products. The initial gap with non-European products (in particular MODIS) has become much smaller.

Acknowledgments

The results presented in this paper were achieved within the Aerosol_cci project funded by the European Space Agency (ESA) as part of the ESA Climate Change Initiative (CCI). We acknowledge AERONET in situ aerosol data providers through which ground based measurements for the validation activities in the project were acquired, as well as ICARE and AEROCOM which played a crucial role in the data validation and data storage services. Aerosol retrieval work by several participants is supported by a range of other ESA projects than the Aerosol-Climate Change Initiative as well as EU FP6 and FP7 grants. ALAMO has been developed by HYGEOS and LOA with the collaboration of ICARE and with partial financial support of CNES "*Centre National d'Etudes Spatiales*".

Appendix A. Aerosol properties and their retrieval from satellite data

Atmospheric aerosol is formally defined as a suspension of particles and/or droplets in air. In the following we neglect the surrounding medium and refer mainly to the particles which are characterized by a certain radius (specified at a certain relative humidity, RH): dry (RH < 30–40% (WMO/GAW, 2003)), at RH = 80%, or their in situ value at ambient RH. Satellites observe aerosol properties in situ, usually integrating over the whole atmospheric column in which both RH and aerosol concentrations may vary strongly. Ground-based measurements are prescribed to be made at dry conditions (WMO/GAW, 2003). Aerosol particles may have sizes ranging from a few nm to several tens of μm, can be composed of a wide range of chemical species (organic matter, inorganic salts) which are either internally mixed (different species occur in one particle) or externally mixed (each particle is composed of one single species) and mixed forms of these. Each size range may have its own physical and chemical properties and based on these, different 'modes' are considered such as cluster (a few nm), nucleation (ca. 5 nm), Aitken (some tens of nm), accumulation (a few hundreds of nm) and coarse (larger than 500 nm) particle modes, where the numbers in parenthesis indicate dry mode radius (see Eq. (A1)). The particle size distribution describes the variation of the particle concentrations with size. The concentrations may be as large as 10^4 to 10^5 cm^{-3} for accumulation mode particles in polluted conditions or as small as 10^{-5} cm^{-3} for the largest particles (radius some tens of μm). Total concentrations, i.e. integrated over the whole size distribution, may vary from a few 10s cm^{-3} in very clean conditions to up to 10^5 cm^{-3} in polluted conditions. Particles can be directly produced by, e.g. mechanical (wind-blown dust, sea spray aerosol), biological (pollen) or combustion (traffic, industry, fires) processes, or they can be produced from precursor gasses by gas-to-particle conversion processes and nucleation. Atmospheric aerosol particles have a lifetime varying from hours to days, depending on their size, during which they undergo physical and chemical changes which in turn changes their chemical composition and their optical and physical properties. Of importance for climate and climate change are particles with dry radii between ca. 30 nm to several μm because these particles are most effective for scattering of radiation in the UV/VIS part of the solar spectrum, and because these particles can be activated to become cloud condensation nuclei (CCN) and thus affect cloud macro- and micro-physical properties which in turn affects cloud reflectance, lifetime and precipitation.

Aerosol size distributions are commonly approximated by multimodal log-normal size distributions (Seinfeld & Pandis, 1998), i.e.:

$$\frac{dN(r)}{d\ln r} = \sum_{i=1}^{2} \frac{N_i}{(2\pi)^{1/2}\ln\sigma_i} \exp\left(-\frac{\left(\ln r_i - \ln r_{gi}\right)^2}{2\ln^2\sigma_i}\right), \tag{A1}$$

where each log-normal mode is defined by three parameters: aerosol number concentration N_i, number mode radius $\bar{r}_{gi}$ and (geometric) standard deviation σ_i. Only aerosol particles with sizes larger than about 0.05 μm in radius (in situ) are optically active and therefore in satellite retrievals only these larger sizes need to be represented. As there is a cross-section minimum at radii of about 0.5 μm and the aerosol composition above and below that size is usually quite

different, in aerosol retrieval the size distribution is usually described as bi-modal rather than mono-modal. The smaller size mode (aerosol radii < 0.5 μm) of the assumed bi-modal distributions is referred to as fine mode and the large size mode (aerosol radii > 0.5 μm) is referred to as coarse mode.

Aerosols have a large impact on climate through their direct effects (scattering and absorption of solar radiation) and indirect effects (through their effect on cloud microphysical properties) on the radiation balance in the earth system. Studies on the effect of aerosols on climate were traditionally made by using chemical transport models (CTM) or global climate models (GCM), or their regional versions. In the last decade satellite observations have increasingly been used to provide observation-based estimates of the effects of aerosols on climate (e.g., Thomas, Chalmers, Harris, Grainger, & Highwood, 2012; Yu et al., 2006). Satellite observations offer the advantage of large spatial coverage with the same instrument and technique as implemented in an instrument-specific retrieval algorithm, at the cost of accuracy and temporal coverage offered by most ground-based observations. However, ground-based observations are representative for only a relatively small area around the observation site, mainly concentrated in certain areas, i.e. Europe, North America and some parts of other continents, while the oceans are sparsely covered. Satellite observations offer in principle global coverage, depending on swath width, in about one day to a week.

The effect of aerosol particles on solar radiation are determined by the particle size distribution and their size-segregated chemical composition, which together determine the angular scattering (expressed as the phase function), absorption and single scattering albedo (SSA, the ratio of scattering and the sum of total scattering and absorption), and the vertical variation of these parameters. Scattering and absorption together determine the extinction of solar light by aerosol particles and the extinction coefficient is the sum of the scattering and absorption coefficients. Changes in global, regional and local effects of aerosol particles can thus be determined by changes in these properties or a combination of them. The basic aerosol parameter retrieved from satellite-based observations is the aerosol optical depth (AOD, or τ), i.e. the column-integrated extinction coefficient specified for a certain wavelength, λ. AOD time series could thus be used to determine trends indicating changes on regional to global scales. However, this requires that AOD can be determined with sufficient accuracy to provide statistically significant trends. Such requirements have been formulated by GCOS (2011) and were further formulated as part of the aerosol-cci project described in the introduction. Another constraint for the use of time series based on satellite data is there representativeness due to limited temporal data availability (number of overpasses, only cloud-free conditions). In addition to AOD, other parameters are sometimes made available from satellite observations with a varying degree of reliability and accuracy. These parameters include the Ångström exponent (AE) describing the wavelength dependence of the AOD, the fine mode fraction (FMF) describing the contribution of particles with dry radii smaller than 0.5 μm to the total AOD, coarse mode fraction (CMF) describing the contribution of larger particles to the total AOD, aerosol type (i.e. parameters describing the aerosol size distribution and optical properties), absorbing aerosol index (AAI), SSA, absorbing aerosol optical depth (AAOD), aerosol layer height. The determination of these other parameters usually requires an AOD exceeding a certain value (e.g., Holzer-Popp, Schroedter, & Gesell, 2002a, 2002b; Kahn et al., 2010).

Instruments used for aerosol retrieval include spectrometers and radiometers with one or more wavebands in the Ultra-violet (UV), Visible (VIS) and Near Infra-Red (NIR) parts of the electromagnetic spectrum, i.e. those wavelengths most sensitive to the scattering of solar light by aerosol particles, with one or more viewing directions and in some cases with information on polarization of the scattered light. Wavelengths in the thermal infrared (TIR) are mainly used for cloud detection, i.e. together with shorter wavelengths they provide information on the occurrence of clouds which hinders the retrieval of aerosol properties; thus cloud-contaminated pixels are discarded from aerosol retrieval. Wavebands in the NIR and TIR also provide information on larger aerosol particles such as volcanic ash and desert dust. A challenge is to discriminate between desert dust and clouds, i.e. desert dust, although considered aerosol, is often inadvertently classified as cloud and thus discarded in the aerosol retrieval process. In addition, satellite-based lidars are used to provide information on aerosol properties. An overview of instruments and algorithms used for the retrieval of aerosol properties from space is provided in Kokhanovsky and de Leeuw (2009) and de Leeuw et al. (2011).

The first instruments which have been used for the retrieval of aerosol properties were launched over three decades ago and thus have the potential to be used for the provision of long time series of aerosol properties and for the analysis of aerosol trends. However, there are issues related to the use of different instruments, which may not be exactly the same, and their calibration. Furthermore, most instruments used for aerosol retrieval were not designed for that purpose and the information they provide is sub-optimal. Exceptions are MODIS, MISR and POLDER (POLarization and Directionality of the Earth's Reflectances). Nevertheless, instruments like the MEdium Resolution Imaging Spectrometer (MERIS), ATSR-2 (Along Track Scanning Radiometer) and AATSR (Advanced ATSR), SeaWiFS (Sea-viewing Wide Field-of-view Sensor), OMI (Ozone Monitoring Instrument) and AVHRR (Advanced Very High Resolution Radiometer), as well as instruments flying on geostationary satellites such as SEVIRI (Spinning Enhanced Visible and Infrared Imager), are currently used for aerosol retrieval. However, the results are often less accurate in comparison with dedicated aerosol retrieval instruments. This may be somewhat surprising in cases where the instrument characteristics are not limiting factors. For instance, the ATSR-2/AATSR instruments should potentially provide good results because of the dual view capability which allows for more effective decoupling of the surface and atmospheric contributions to the top of atmosphere (TOA) radiance than is possible with a single view, and because of the availability of wavebands from the visible to the thermal infrared facilitating effective cloud screening.

All instruments, also those dedicated for the retrieval of aerosol and cloud properties, do provide insufficient information to accurately determine all relevant aerosol properties, i.e. particle size distribution, size-dependent particle shape and chemical composition, mixing state, from which the optical properties could be determined. This is in part due to the lack of vertical resolution of spectrometers and radiometers. These instruments observe the effect of aerosol particles integrated over the whole atmospheric column while usually not only their concentrations may change with height but also their chemical composition. In addition, as indicated above, particle sizes change with varying relative humidity. Furthermore, the atmosphere may be stratified and in disconnected layers with different origin and different history the aerosol properties may be different. This situation is further complicated by the occurrence of absorbing particles, the effect of which on the AOD depends on the altitude at which they occur.

As a result, the retrieval problem is underdetermined, i.e. there are more unknowns than independent pieces of information to solve the radiative transfer equations and assumptions need to be made. These include assumptions on the aerosol properties, using simplified descriptions of size distributions and optical parameters and aerosol layer height. Furthermore the treatment of the surface is very important, in particular over reflecting surfaces where the surface contribution to the upwelling TOA radiance may be as strong as, or even much stronger than, the atmospheric contribution. Over ocean the retrieval is relatively simple because the ocean surface is dark at wavelengths in the NIR and an ocean reflectance model is often used to account for effects such as sun glint, waves, whitecaps or chlorophyll. Over land, forests are usually relatively dark at shorter wavelengths in the UV/VIS and at wavelengths in the UV all surfaces are dark. The latter is used in the MODIS deep blue algorithm (Hsu, Tsay, King, & Herman, 2004). For the retrieval of aerosol properties with instruments that

do not provide measurements at wavelengths in the UV, other assumptions need to be made.

Appendix B. Brief descriptions of aerosol retrieval algorithms used in the Aerosol-cci project

The aerosol retrieval algorithms used in the Aerosol-cci project, see Table 1, use data from AATSR and MERIS, both flying on ESA's Environmental satellite ENVISAT (2002–2012), and POLDER, flying on PARASOL which is part of NASA's A-train constellation. Aerosol-cci includes algorithms which use one single instrument and the SYNAER algorithm which synergistically uses data from AATSR and the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY). These algorithms provide information on column-integrated aerosol properties such as AOD and additional information which differs for each algorithm. An overview is presented in Table 1. In addition, the Ozone Monitoring Instrument (OMI) provides information on the aerosol absorbing index (AAI) and the Global Ozone Monitoring by Occultation of Stars (GOMOS) provides information on stratospheric aerosol profiles.

Each of these algorithms is extensively described in their respective ATBD (algorithm theoretical baseline document) provided on the Aerosol-cci website (http://www.esa-aerosol-cci.org/) and references provided in these. Brief summaries of the essential characteristics of each algorithm are provided below.

AATSR ADV and ASV

The ATSR-2/AATSR dual view aerosol retrieval algorithm, ADV, is based on Veefkind, de Leeuw, and Durkee (1998). The main feature of the ATSR instruments is the dual view which in ADV is used to effectively eliminate the contribution of the surface reflection to the TOA reflectance, using the k-ratio approach, and retain only the atmospheric path radiance. The k-ratio approach uses the ratio of the reflectances measured in the forward and nadir views, based on Flowerdew and Haigh (1995). The k-ratio is evaluated for the 1.61 μm channels and is assumed to be wavelength-independent. Over bright surfaces this approximation may not apply and the method is therefore limited to TOA reflectances at 1.6 μm wavelength of smaller than 0.45 at nadir. Furthermore, the contribution of aerosols to the AOD at 1.61 μm is in first approximation assumed to be negligible, but is given a value during the next iteration steps. This assumption does not hold in the presence of coarse mode aerosol such as desert dust. Aerosol retrieval over ocean is based on the single view algorithm, ASV, developed by Veefkind and de Leeuw (1998). The ocean surface is assumed dark at wavelengths in the NIR and an ocean reflectance model is used to correct for effects due to chlorophyll and whitecaps. Pixels for which the AATSR L1b GBT data indicates sun glint are excluded from retrieval. ADV and ASV use the cloud mask described by Robles Gonzalez (2003) (see also Curier et al., 2009), with a post-processing method based on comparison of neighboring pixels in a 3x3 pixels (L2) area. The post-processing effectively eliminates spatial inhomogeneity's such as those due to previously undetected clouds and cloud edges. The path radiance is used to retrieve the aerosol properties using a LUT approach with a combination of aerosol components described in Appendix C. The mixing ratio of these aerosol components, and thus the size distribution and optical properties, is varied to match the reflectances at each of the 3 (ADV) or 4 (ASV) wavelengths in the VIS and NIR. ADV and ASV products are AOD at 3 (ADV) or 4 (ASV) wavelengths, AE (needs AOD (550 nm) > 0.2 to obtain reasonable results) and mixing ratio, with SSA and surface albedo as research products. Not only the default resolution of 10×10 km^2 is used, but also 1×1 km^2 is used in certain studies. The latest version of ADV/ASV including many improvements made at FMI and the University of Helsinki (UHEL) and uncertainty characterization is described in Kolmonen et al. (2013).

AATSR ORAC

The Oxford-RAL Retrieval of Aerosol and Cloud (ORAC V1) algorithm is an optimal estimation (OE) retrieval scheme designed to provide estimates of aerosol optical depth and effective radius, cloud top pressure, height and temperature, cloud particle effective radius, cloud optical depth and cloud type (generally liquid water or ice) from multispectral imagery (Poulsen et al., 2012; Sayer et al., 2011; Thomas et al., 2009). The method fits all the shortwave forward and nadir radiances simultaneous using a forward model based on the DISORT radiative transfer code (Stamnes, Tsay, Wiscombe, & Jayweera, 1988). It is worth noting that the simultaneous retrieval of all state parameters provided by the OE method ensures that a physically consistent and numerically optimal estimate of the state is produced. The quality of fit to the radiances allows the quality of the retrieval to be judged a posteriori. In addition the error in the retrieved aerosol parameters is estimated by propagating both the measurement and forward model uncertainties into state space. Note that the dataset described here was produced by the ORAC V1 algorithm where an a priori surface BRDF is set using MODIS MCD43B BRDF products (Jin et al., 2003) over land and an ocean surface reflectance model over the ocean (Sayer et al., 2008). More recent processing with an updated surface model is currently under evaluation and initial results indicate a substantial improvement when compared to V1.

SU ATSR algorithm

The SU-ATSR algorithm has been developed at Swansea University for estimation of atmospheric aerosol and surface reflectance for the ATSR-2 and AATSR sensors. Over land, the algorithm employs a parameterized model of the surface angular anisotropy, and uses the dual-view capability of the instrument to allow estimation without a priori assumptions on surface spectral reflectance. Over ocean, the algorithm uses a simple model to exploit the low ocean leaving radiance at red and infra-red channels at both nadir and along-track view angles. The surface models are used to invert the 6SV model (Kotchenova & Vermote, 2007; Kotchenova, Vermote, Matarrese, & Klemm, 2006) to perform retrieval at 10 km resolution. The algorithm has been implemented on the ESA Grid Processing on Demand (GPOD) system to allow global processing and free download of AOD and surface reflectance. The method is documented in North, Briggs, Plummer, and Settle (1999), North (2002), Grey, North, Los, and Mitchell (2006), Grey, North, and Los (2006), Bevan, North, Grey, Los, and Plummer (2009) and Bevan, North, Los, and Grey (2012).

SYNAER

The synergistic aerosol retrieval method SYNAER delivers aerosol optical depth (AOD) and an estimation of the type of aerosols in the lower troposphere over both land and ocean by exploiting a combination of a radiometer and a spectrometer. The type of aerosol is estimated as percentage contribution of 4 representative aerosol components (sea salt, mineral dust, weakly absorbing accumulation mode and strongly absorbing accumulation mode aerosol). The high spatial resolution including thermal spectral bands of the radiometer permits accurate cloud detection. The SYNAER aerosol retrieval algorithm comprises of two major parts. In step 1 a dark field method exploits single wavelength radiometer reflectances (at 670 nm over land, at 870 nm over ocean) to determine 36 values of the aerosol optical depth and surface reflectance over automatically selected and characterized dark pixels for a set of 36 different pre-defined boundary layer aerosol mixtures. In step 2 the parameters retrieved in the first step are used to simulate spectra for the same set of 36 different aerosol mixtures with the same radiative transfer code after spatial integration to the larger pixels of the spectrometer. A least-squares fit of these calculated spectra at 10 wavelengths to the measured spectrum delivers the correct AOD

value (the one AOD retrieved in step 1 for the aerosol type selected in step 2) and – if a uniqueness test is passed – the most plausible spectrum and its underlying aerosol mixture. (Holzer-Popp, Schroedter-Homscheidt, Breitkreuz, Klüser, & Martynenko, 2008; Holzer-Popp et al., 2002a). Using a combination of 2 instruments with different scan patterns SYNAER can only provide global cloud-free coverage every 12 days and with large pixels of 60 × 30 km^2. However the combination of the 2 instruments has the potential to provide aerosol type information (qualitatively shown in Holzer-Popp et al., 2008). Although these method-inherent limitations mean a significant drawback in comparison to AATSR AOD products, SYNAER has been included into the Aerosol_cci project in order to qualify and improve its quantitative AOD results and thus eventually strengthen the aerosol type information.

MERIS ESA standard

The MERIS standard aerosol retrieval over land algorithm was designed to work over Dense Dark Vegetation (DDV) targets (Ramon & Santer, 2001; Santer, Carrere, Dubuisson, & Roger, 1999). A set of DDV Bidirectional Reflectance Function (BRF) models was assembled for 11 different biomes on Earth. DDV detection is based on a threshold on the Atmospherically Resistant Vegetation Index (ARVI) computed from Rayleigh corrected reflectances at 443, 665 and 865 nm. As DDV spatial cover is low, the aerosol inversion was extended to brighter surfaces called Land Aerosol Remote Sensing (LARS) targets (Santer, Ramon, Vidot, & Dilligeard, 2007). LARS spectral albedo can be predicted as it is linearly related to ARVI. Slopes and offsets of these linear regressions are stored in Look Up Tables for 1°x1° boxes and on a monthly basis. The aerosol retrieval consists in the inversion of the AOD at 443 and 665 nm that allow to reproduce the measured TOA reflectances at these wavelengths using pre-calculated aerosol scattering functions for aerosol models described by a Junge Power-Law (JPL) size distribution and a constant refractive index of 1.45–0.0i. The outputs of the algorithm are the AOD at 443 nm and the aerosol Ångström exponent derived between 443 and 665 nm.

Cloud contamination is the biggest issue of the product that is delivered at the same spatial resolution as the level 1B data (i.e. 1.2 km). The product, with a good spatial coverage now, has been validated only for the AOD at 443 nm. The Ångström exponent is not validated since the retrieved AOT at 665 nm is noisy. It is mandatory to move toward spatial resolution of 10 × 10 km^2 for the aerosol product in order to reduce cloud contamination and enhance the Signal to Noise Ratio (SNR) for the Ångström exponent retrieval. Finally there is a need for improving the LARS BRDF model.

MERIS ALAMO

The MERIS ALAMO (Aerosol Load and Altitude from MERIS over Ocean) algorithm has been primarily developed for aerosol altitude retrievals using MERIS data. Necessary inputs for altitude retrievals, such as aerosol optical properties, are derived in a first step with an initial assumption on the layer altitude. The cloud masking and AOD retrieval schemes are a close adaptation of the MODIS algorithm (Remer et al., 2005; Tanré, Kaufman, Herman, & Mattoo, 1997), using only the following MERIS bands: 510, 560, 665, 753.75 and 865 nm. Due to spectral characteristics of MERIS, ALAMO is limited to a maximum wavelength of 865 nm and only two pieces of information on aerosol properties can therefore be retrieved instead of three parameters with MODIS. MERIS aerosols products are retrieved with a spatial resolution of 10 × 10 pixels (12 × 12 km^2). This resolution allows (i) an adequate signal-to-noise ratio (SNR) for a better characterization of the aerosol type and (ii) rejection of pixels considered as non-valid through statistics criteria, in order to ensure the quality of the aerosol product. The aerosol products of ALAMO include the optical thickness and the mixing ratio of fine and coarse modes. Aerosol models used for ALAMO are the same as the ones used for the most current version of MODIS products.

In a second step the altitude of the aerosol layer is estimated using the MERIS O_2 A absorption channel and following the algorithm described in Dubuisson et al. (2009). A pixel reclassification is done after the altitude retrieval to remove high thin clouds based on a threshold on altitude and spatial variance of altitude.

MERIS BAER

The Bremen Aerosol Retrieval, BAER, has been developed to derive spectral AOD from multispectral satellite imagery such as from MERIS over ocean and land. It separates the spectral aerosol reflectance from the surface and Rayleigh path reflectances for the short-wave ($\leq$0.67 μm) TOA reflectance over land. Over ocean the whole spectral range of MERIS is utilized for the AOD retrieval.

The surface reflectance is estimated by a linear mixing of vegetation and non-vegetation spectra which are tuned by the Normalized Differential Vegetation Index (NDVI). Bidirectional Reflectance Distribution Function (BRDF) effects are taken into account using the Raman–Pinty–Verstraete model (Maignan, Bréon, & Lacaze, 2004). Finally BAER derives the target quantity, the AOD, using LUTs, created with rigorous radiative transfer model calculations, ensuring spectral smoothness for the retrieval over all channels (von Hoyningen-Huene, Freitag, & Burrows, 2003; von Hoyningen-Huene et al., 2011).

After specific adaptations it could be shown, that the approach is also successfully applicable to retrievals over bright surfaces such as deserts (Dinter et al., 2009).

PARASOL

The PARASOL algorithm is based on LUTs of the directional, spectral, and polarized radiances calculated for different aerosol models with different optical thicknesses, size distribution and refractive index. The choice of the models used to build the LUT is a key issue. The aerosol size distribution is assumed to be the sum of two contributions, one coming from small spherical (fine mode) aerosols and the other from large (coarse mode) aerosols (Herman, Deuzé, Marchand, Roger, & Lallart, 2005). Large particles can be either spherical, non-spherical or a mixture of both. The size distributions of spherical particles (small or large) are described by a log-normal function defined by two parameters, namely, a mean radius and a standard deviation σ. For large non-spherical aerosols, an experimental model is implemented in the LUT (Volten et al., 2001). The LUT are built with a radiative transfer code based on successive orders of scattering (Lenoble et al., 2007). The Stokes parameters are calculated at the top of the atmosphere and computations include multiple scattering in the atmosphere by molecules and aerosols and take into account the surface-atmosphere interaction.

Over ocean, the inversion scheme mainly uses the normalized radiances in the 865 nm channel, where the ocean color reflectance is zero, and in the 670 nm channel with a constant water reflectance of 0.001. The polarized Stokes parameters at 865 and 670 nm are also used for deriving the best aerosol model. Computations are performed with a rough ocean surface (Cox & Munk, 1954) and a wind speed of 5m/s. The foam contribution is calculated according to Koepke's model (Koepke, 1984) and a constant value of 0.22 for the foam reflectance.

Appendix C. Algorithm improvement

Aerosol retrieval is an underdetermined problem since the number of degrees-of-freedom, i.e. parameters describing the aerosol properties which determine the observed TOA radiances, is smaller than the number of observations. Hence assumptions need

to be made. The most important assumptions made in aerosol retrieval concern:

- cloud screening
- surface treatment
- aerosol optical properties and size distribution.

Aerosol retrieval can only be made for cloud-free sky because the high reflectance of clouds at wavelengths in the UV-NIR interferes with the aerosol reflectance and hence prohibits accurate retrieval of aerosol properties. Therefore, an accurate cloud mask has to be applied to screen all pixels for the occurrence of clouds and exclude them from retrieval. Currently all algorithms participating in Aerosol-cci use their own cloud detection procedures as described in Appendix B and the literature referenced there. The use of a common cloud flag for similar products is under study (Holzer-Popp et al., 2013). To further eliminate cloud-contaminated data, a post-processing step has been developed to effectively detect cloud edges as described in Appendix B for the AATSR ADV and ASV algorithms. This post-processing step results in a smoothly varying AOD across extended areas without sudden transitions. This post-processing step has been implemented in other algorithms (ORAC, SU) as well.

The treatment of the surface and accounting for contributions of surface reflectance to the radiance measured at TOA depends on the instrument properties, and how they are used. An overview of surface treatment and application to different algorithms such as the AATSR algorithms used in Aerosol-cci has been presented in Kokhanovsky and de Leeuw (2009), for MERIS BAER in de Leeuw et al. (2011), and for the other MERIS algorithms in the respective ATBDs. Therefore surface treatment will not be discussed here.

Apart from improved cloud screening, most progress has been made in harmonizing aerosol models and their use in the various retrieval algorithms. For the Aerosol-cci project, a simple set of four aerosol components has been developed consisting of two fine mode components, one of which has a complex refractive index representative of weakly absorbing aerosol particles and the other one represents strongly absorbing aerosol particles. The two coarse mode components are representative for the characteristics of desert dust and sea salt aerosol. Each component is thus described by a lognormal size distribution, defined by mode radius, effective radius, geometric standard deviation and variance, and by the complex refractive index (Table C1).

The two fine mode-types are extremes in terms of absorption and reality (in terms of absorption) is always a combination of these two types. The choice of the fine mode radius is based on an analysis of AERONET sun-photometer data which shows that the most frequent fine mode size (in terms of the effective radius) is near 0.14 μm. The coarse mode is dominated by two quite different aerosol types: spherical largely non-absorbing sea-salt and non-spherical absorbing dust. Based on an AERONET probability distribution for the coarse mode, the effective radius was set to 1.94 μm for these two coarse mode aerosol types. See Holzer-Popp et al. (2013) for more detail.

The optical properties of aerosol particles are usually calculated by application of a Mie code (Mie, 1908), which applies to spherical particles. However, for dust Mie codes cannot be applied because of the non-spherical shape of dust particles. In Aerosol-cci a T-matrix method was used assuming randomly oriented spheroids with aspect ratios between 1.44 and 3.0 (Dubovik et al., 2002; Sinyuk, Torres, & Dubovik, 2003). Although spheroids may be unable to represent the entire shape complexity for dust, this spheroid method is preferable over methods for spheres. The choice of the refractive index for dust is based on Volten et al. (2001). Observational data (Dubovik et al., 2002; Sinyuk et al., 2003) demonstrate that the dust absorbing strength is wavelength dependent, and decreases from the UV (imaginary refractive index, RFi, near 0.005) to the near-IR (RFi near 0.001). To avoid time-consuming computations during the retrieval, radiative transfer is computed in atmospheres with different aerosol components, for discrete AOD values and a range of discrete configurations (e.g., solar zenith angle, viewing angle), and the results are stored in a Look-Up Table (LUT). During the retrieval the optical properties for the relevant configuration are obtained by simple extrapolation of the LUT values.

For successful retrieval of the aerosol type by using a mixture of the four basic aerosol components presented in Table C1, additional information may be required on relationships between fine and coarse mode, between less and more absorbing fine mode and between dust and sea-salt components in the coarse mode. This information is supplied in terms of monthly $1^\circ \times 1^\circ$ climatological data derived from two sources, modeling and observations.

The modeling information was obtained by combining output of 14 different global models, with complex aerosol components which participated in AeroCom exercises, into 'AeroCom' median maps (Kinne et al., 2006). Based on these median maps, ratios between different aerosol components are defined. Dust and sea salt generally define the coarse mode; sulfate, organic matter and black carbon define the fine mode.

A global aerosol climatology was developed as an improvement of the AeroCom model median by adding AERONET (Holben et al., 1998) quality data in a merging process for AOD, Ångström exponent (describing the AOD spectral dependence) and single scattering albedo (describing the absorption potential). With observational ties data of this 'climatology' are recommended over data from 'modeling' alone.

This climatology is used as a priori for the occurrence of aerosol types/mixtures, per region and per month. In general the coarse mode component selected would be sea salt, except in the presence of desert dust which mainly occurs in certain areas. The choice of the fine mode component would also be based on the climatology and the two fine mode components, with equal microphysical properties, could be mixed to obtain the desired absorption properties (as provided by the SSA in the climatology). Using the occurrence of aerosol types, the retrieval algorithm computes the radiances at the top of the atmosphere which are compared with the satellite measurements. Based on this comparison the aerosol mixtures are adjusted and the procedure is iterated until convergence is reached and the most likely aerosol model providing the measured radiance is selected. With this model the AOD is computed. It is emphasized that the climatological AOD is not used in the retrieval process, and the aerosol mixtures are only used as a priori, except in sensitivity studies. The actual AOD and aerosol mixtures are retrieved based on the measured radiances.

Table C1
Log-normal parameters for two coarse and two fine mode aerosol components and their associated mid-visible refractive indices (mode number radius and standard deviation [or variance] define the effective radius, which is the 3rd moment to 2nd moment radius ratio).

Aerosol component	Refract index real p. (.55 μm)	Refract index imag p. (.55 μm)	Reff (μm)	Geom. st dev (σ_i)	Variance (ln σ_i)	Mode#. radius (μm)	Comments	Aerosol layer height
CM1: dust	1.56	0.0018	1.94	1.822	0.6	0.788	Non-spherical	2–4 km
CM2: sea salt	1.4	0	1.94	1.822	0.6	0.788	AOD threshold constraint[#]	0–1 km
FM1: weak-abs	1.4	0.003	0.140	1.7	0.53	0.07	(ss-albedo at 0.55 μm: 0.98)	0–2 km
FM2: strong-abs	1.5	0.040	0.140	1.7	0.53	0.07	(ss-albedo at 0.55 μm: 0.802)	0–2 km

Algorithm improvement was measured by application of the validation and evaluation exercises described in Section 2. These exercises were made for only one month, September 2008, a necessary restriction because of the time it takes to run the retrieval with different aerosol mixtures. Success was identified by comparison with the baseline algorithms and successive improvement after implementation of different aerosol models, the use of the AEROCOM median with different degrees of comprehensiveness (i.e. varying from completely free retrieval without any use of the climatology, to a full prescription of the aerosol mixing, and combinations thereof) and different cloud masks. In addition to these experiments, algorithms were also improved as regards coding and debugging and the retrieval products were improved by application of post-processing. Results from this study for 1 month are presented in Holzer-Popp et al. (2013).

Appendix D. Evaluation of L3 performance indicators

The applied single value scoring method is based on outlier resistant sub-scores for bias, spatial correlation and temporal correlation. All sub-scores are based on values ranks and not on values. First the scores are introduced and then the particular application is explained.

Bias sub-score S_B

Given data-pairs for test-data D and reference-data R, a bias performance indicator E_B for the range $[-1.0, +1.0]$ is determined the following way:

1. put the elements of test-data D and reference-data R into a single array A.
2. re-order elements in array A in increasing order and assign each element its rank
3. sum the ranks separately for elements of reference data R (R_{sum}) and test data D (D_{sum})
4. determine a weight w based on interquartile range and interquartile average of reference data R ($IQrange[R]$, $IQavg[R]$) and test-data D ($IQrange[D]$, $IQavg[D]$), in order to reduce errors, when the central variability relative to central values is relative small
5. determine bias performance indicator E_B and bias sub-score S_B, with the sign indicating the bias direction

$$E_B = w * \frac{D_{sum} - R_{sum}}{D_{sum} + R_{sum}}, \quad w = \max\left\{\frac{IQrange[D] + IQrange[R]}{IQavg[D] + IQavg[R]}, 1.0\right\} \tag{D1}$$

$$S_B = 1 - E_B, \quad if \quad E_B > 0 \qquad S_B = -1 + E_B, \quad if \quad E_B < 0. \tag{D2}$$

Sub-scores for spatial (S_S) and temporal (S_T) variability

Given n data-pairs for test-data D and reference-data R, spatial and temporal Spearman rank correlations are quantified via the correlation coefficients R_C. Then the spatial and temporal performance indicators E_S and E_T for a range $[+1.0, 0.0]$ are determined this way:

1. rank all elements of test data D in increasing order and assign each element its rank D_{rank}
2. rank all elements of the reference data R in increasing order and assign element rank R_{rank}
3. evaluate the Spearman rank correlation coefficient R_C
4. determine a weight w based on interquartile range and interquartile average of reference data R ($IQrange[R]$, $IQavg[R]$) and test-data D ($IQrange[D]$, $IQavg[D]$), in order to reduce errors, when the central variability relative to central values is relative small
5. determine performance indicators E_S and E_T and correlation sub-scores S_S and S_T

$$Rc = 1 - \frac{6\sum_{i=1}^{n} (\mathrm{Drank} - \mathrm{Rrank})_i^2}{n(n^2 - 1)} \tag{D3}$$

$$E = w * (1 - Rc/2), \quad w = \max\left\{\frac{IQrange[D] + IQrange[R]}{IQavg[D] + IQavg[R]}, 1.0\right\} \tag{D4}$$

$$S_S = 1 - E_S, \quad (spatial) \qquad S_T = 1 - E_T, \quad (temporal). \tag{D5}$$

Once all three sub-scores (S_B, S_S and S_T) are defined the overall single score (S) is defined via multiplication. Hereby the sign of the bias score is carried in the overall score S to indicate the bias direction. The total score also defines the overall performance indicator E.

$$S = S_B * S_T * S_S E = 1 - |S|. \tag{D6}$$

Application

For the evaluation of satellite retrieved AOD data, total scores and sub-scores are determined for individual regions (see Fig. 3) based on daily data. Hereby AERONET samples provide the reference data. Since AERONET data are not available in some regions, or not available in some regions in sufficient numbers, no scores could be provided in many regions. When in addition the satellite samples are relatively sparse, the number of regions with available scores is further reduced. For each region (1) temporal correlation and bias sub-scores are determined for individual sites (10 sites minimum) and (2) spatial correlation and bias sub-scores are determined for individual days with (10 minimum sites) distributed in that region. Only for both valid temporal and spatial correlation scores the total regional score is determined. Hereby the bias score is the average of both bias sub-scores. Finally, total regional scores are combined according the fractional global coverage for a final single score.

References

Bevan, S. L., North, P. R. J., Grey, W. M. F., Los, S. O., & Plummer, S. E. (2009). Impact of atmospheric aerosol from biomass burning on Amazon dry-season drought. *Journal of Geophysical Research*, *114*, D09204. http://dx.doi.org/10.1029/2008JD011112.

Bevan, S. L., North, P. R. J., Los, S. O., & Grey, W. M. F. (2012). A global dataset of atmospheric aerosol optical depth and surface reflectance from AATSR. *Remote Sensing of Environment*, *116*, 119–210.

Cox, C., & Munk, W. (1954). Statistics of the sea surface derived from sun glitter. *Journal of Marine Research*, *13*, 198–208.

Curier, L., de Leeuw, G., Kolmonen, P., Sundström, A. -M., Sogacheva, L., & Bennouna, Y. (2009). Aerosol retrieval over land using the (A)ATSR dual-view algorithm. In A. A. Kokhanovsky, & G. de Leeuw (Eds.), *Satellite aerosol remote sensing over land* (pp. 135–159). Berlin: Springer-Praxis978-3-540-69396-3.

de Leeuw, G., Kinne, S., Leon, J. F., Pelon, J., Rosenfeld, D., Schaap, M., et al. (2011). Retrieval of aerosol properties. In J. P. Burrows, U. Platt, & P. Borrell (Eds.), *The remote sensing of tropospheric composition from space* (pp. 259–313). Berlin, Heidelberg: Springer-Verlag978-3-642-14790-6. http://dx.doi.org/10.1007/978-3-642-14791-3.

de Leeuw, G., & Kokhanovsky, A. (2009). Introduction. In A. A. Kokhanovsky, & G. de Leeuw (Eds.), *Satellite aerosol remote sensing over land* (pp. 1–18). Berlin: Springer-Praxis978-3-540-69396-3.

Dinter, T., von Hoyningen-Huene, W., Burrows, J. P., Kokhanovsky, A. A., Bierwirth, E., Wendisch, M., et al. (2009). Retrieval of aerosol optical thickness for desert conditions using MERIS observations during the SAMUM campaign. *Tellus B*, *6*, 229–238 (ISSN 0280-6509).

Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M.D., et al. (2002). Variability of absorption and optical properties of key aerosol types observed in worldwide locations. *Journal of the Atmospheric Sciences*, *59*, 590–608.

Dubuisson, P., Frouin, R., Dessailly, D., Duforêt, L., Léon, J. -F., Voss, K., et al. (2009). Estimation of aerosol altitude from reflectance ratio measurements in the O2 A-band. *Remote Sensing of Environment*, *113*, 1899–1911.

Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., et al. (1999). Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *Journal of Geophysical Research*, *104*(D24), 31,333–31,349. http://dx.doi.org/10.1029/1999JD900923.

Flowerdew, R. J., & Haigh, J.D. (1995). An approximation to improve accuracy in the derivation of surface reflectances from multi-look satellite radiometers. *Geophysical Research Letters*, *22*, 1693–1696.

GCOS (2011). *Systematic observation requirements for satellite-based products for climate, 2011 update*. GCOS Report 154. WMO (http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf)

Grey, W. M. F., North, P. R. J., & Los, S. (2006). Computationally efficient method for retrieving aerosol optical depth from ATSR-2 and AATSR data. *Applied Optics*, *45*(12), 2786–2795.

Grey, W. M. F., North, P. R. J., Los, S. O., & Mitchell, R. M. (2006). Aerosol optical depth and land surface reflectance from multi-angle AATSR measurements: Global validation and inter-sensor comparisons. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(8), 2184–2197.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., et al. (2002). Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models. *Nature*, *415*, 626–630. http://dx.doi.org/10.1038/415626a.

Herman, M., Deuzé, J. -L., Marchand, A., Roger, B., & Lallart, P. (2005). Aerosol remote sensing from POLDER/ADEOS over the ocean: Improved retrieval using a nonspherical particle model. *Journal of Geophysical Research*, *110*, D10S2. http://dx.doi.org/10.1029/2004JD004798.

Hoff, R. M., & Christopher, S. A. (2009). Remote sensing of particulate pollution from space: have we reached the promised land? *Journal of the Air and Waste Management Association*, *59*, 645–675.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., et al. (1998). AERONET — A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*, 1–16.

Hollmann, R., Merchant, C., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., et al. (2013). The ESA climate change initiative: Satellite data records for essential climate variables. *Bulletin of the American Meteorological Society*. http://dx.doi.org/10.1175/BAMS-D-11-00254.1.

Holzer-Popp, T., de Leeuw, G., Griesfeller, J., Martynenko, D., Klüser, L., Bevan, S., et al. (2013). Aerosol retrieval experiments in the ESA Aerosol_cci project. *Atmospheric Measurement Techniques*, *6*, 1919–1957. http://dx.doi.org/10.5194/amt-6-1919-2013.

Holzer-Popp, T., Schroedter, M., & Gesell, G. (2002a). Retrieving aerosol optical depth and type in the boundary layer over land and ocean from simultaneous GOME spectrometer and ATSR-2 radiometer measurements, 1, Method description. *Journal of Geophysical Research*, *107*, 4578. http://dx.doi.org/10.1029/2001JD002013.

Holzer-Popp, T., Schroedter, M., & Gesell, G. (2002b). Retrieving aerosol optical depth and type in the boundary layer over land and ocean from simultaneous GOME spectrometer and ATSR-2 radiometer measurements, 2, Case study application and validation. *Journal of Geophysical Research*, *107*, 4770. http://dx.doi.org/10.1029/2002JD002777.

Holzer-Popp, T., Schroedter-Homscheidt, M., Breitkreuz, H., Klüser, L., & Martynenko, D. (2008). Improvements of synergetic aerosol retrieval for ENVISAT. *Atmospheric Chemistry and Physics*, *8*, 7651–7672.

Hsu, N. C., Tsay, S.C., King, M.D., & Herman, J. R. (2004). Aerosol properties over bright-reflecting source regions. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(3), 557–569.

Huneeus, N., Chevallier, F., & Boucher, O. (2012). Estimating aerosol emissions by assimilating observed aerosol optical depth in a global aerosol model. *Atmospheric Chemistry and Physics*, *12*, 4585–4606. http://dx.doi.org/10.5194/acp-12-4585-2012.

Jin, Y., Schaaf, C. B., Woodcock, C. E., Gao, F., Li, X., Strahler, A. H., et al. (2003). Consistency of MODIS surface BRDF/Albedo retrievals: 1. Algorithm performance. *Journal of Geophysical Research*, *108*, D54158. http://dx.doi.org/10.1029/2002JD002803.

Kahn, R. A., Gaitley, B. J., Garay, M. J., Diner, D. J., Eck, T., Smirnov, A., et al. (2010). Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network. *Journal of Geophysical Research*, *115*, D23209. http://dx.doi.org/10.1029/2010JD014601.

Kahn, R. A., Nelson, D. L., Garay, M. J., Levy, R. C., Bull, M.A., Diner, D. J., et al. (2009). MISR aerosol product attributes and statistical comparisons with MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, *47*, 4095–4114.

Kaufman, Y. J., Justice, C. O., Flynn, L. P., Kendall, J.D., Prins, E. M., Giglio, L., et al. (1998). Potential global fire monitoring from EOS-MODIS. *Journal of Geophysical Research*, *103*, 32215–32238.

Kinne, S., Schulz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S. E., et al. (2006). An AeroCom initial assessment optical properties in aerosol component modules of global models. *Atmospheric Chemistry and Physics*, *6*, 1815–1834.

Koepke, P. (1984). Effective reflectance of oceanic whitecaps. *Applied Optics*, *23*, 1816–1823.

Kokhanovsky, A. A., & de Leeuw, G. (2009). *Satellite aerosol remote sensing over land*. Berlin: Springer-Praxis978-3-540-69396-3 (388 pp.).

Kolmonen, P., Sundström, A. -M., Sogacheva, L., Rodriguez, E., Virtanen, T. H., & de Leeuw, G. (2013). Uncertainty characterization of AOD for the AATSR dual and single view retrieval algorithms. *Atmospheric Measurement Techniques Discussions*, *6*, 4039–4075. http://dx.doi.org/10.5194/amtd-6-4039-2013.

Kotchenova, S. Y., & Vermote, E. F. (2007). Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part II. Homogeneous Lambertian and anisotropic surfaces. *Applied Optics*, *46*(20), 4455–4464.

Kotchenova, S. Y., Vermote, E. F., Matarrese, R., & Klemm, F. J., Jr. (2006). Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance. *Applied Optics*, *45*(26), 6762–6774.

Labonne, M., Bréon, F. -M., & Chevallier, F. (2007). Injection height of biomass burning aerosols as seen from a spaceborne lidar. *Geophysical Research Letters*, *34*, 1–5. http://dx.doi.org/10.1029/2007GL029311.

Lee, K. H., Li, Z., Kim, Y. J., & Kokhanovsky, A. (2009). Atmospheric aerosol monitoring from satellite observations: A history of three decades. In Y. J. Kim, U. Platt, M. B. Gu, & H. Iwahashi (Eds.), *Atmospheric and biological environmental monitoring* (pp. 13–38). Springer Science+Business Media B.V.

Lenoble, J., Herman, M., Deuzé, J. L., Lafrance, B., Santer, R., & Tanré, D. (2007). A successive order of scattering code for solving the vector equation of transfer in the Earth's atmosphere with aerosols. *JQSRT*, *107*, 479–507.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., & Eck, T. F. (2010). Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *Atmospheric Chemistry and Physics*, *10*, 10399–10420. http://dx.doi.org/10.5194/acp-10-10399-2010.

Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E., & Kaufman, Y. J. (2007). Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *Journal of Geophysical Research*, *112*(D13), D13211. http://dx.doi.org/10.1029/2006JD007811.

Maignan, F., Bréon, F. -M., & Lacaze, R. (2004). Bidirectional reflectance of Earth targets: Evaluation of analytical models using a large set of spaceborne measurements with emphasis on the Hot Spot. *Remote Sensing of Environment*, *90*, 210–220.

Mie, G. (1908). Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Annals of Physics*, *25*, 377–445.

Müller, D., Krasemann, H., Brewin, R. J. W., Brockmann, C., Deschamps, P. -Y., Doerffer, R., et al. (2013). The ocean colour climate change initiative: Spatial and seasonal homogeneity of atmospheric correction algorithms. *RSE* (in this issue).

North, P. R. J. (2002). Estimation of aerosol opacity and land surface bidirectional reflectance from ATSR-2 dual-angle imagery: Operational method and validation. *Journal of Geophysical Research*, *107*. http://dx.doi.org/10.1029/2000JD000207.

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

Poulsen, C. A., Watts, P. D., Thomas, G. E., Sayer, A.M., Siddans, R., Grainger, R. G., et al. (2012). Cloud retrievals from satellite data using optimal estimation: Evaluation and application to ATSR. *Atmospheric Measurement Techniques*. http://dx.doi.org/10.5194/amt-5-1889-2012.

Ramon, D., & Santer, R. (2001). Operational remote sensing of aerosols over land to account for directional effects. *Applied Optics*, *40*, 3060–3075.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., et al. (2005). The MODIS aerosol algorithm, products, and validation. *Journal of the Atmospheric Sciences*, *62*, 947–973.

Robles Gonzalez, C. (2003). *Retrieval of aerosol properties using ATSR-2 observations and their interpretation*. (Ph.D. thesis). University of Utrecht.

Santer, R., Carrere, V., Dubuisson, P., & Roger, J. C. (1999). Atmospheric correction over land for MERIS. *International Journal of Remote Sensing*, *20*, 1819–1840.

Santer, R., Ramon, D., Vidot, J., & Dilligeard, E. (2007). A surface reflectance model for aerosol remote sensing over land. *International Journal of Remote Sensing*, *28*, 737–760.

Sayer, A.M., Poulsen, C. A., Arnold, C., Campmany, E., Dean, S., Ewen, G. B.L., et al. (2011). Global retrieval of ATSR cloud parameters and evaluation (GRAPE): dataset assessment. *Atmospheric Chemistry and Physics*, *11*, 3913–3936. http://dx.doi.org/10.5194/acp-11-3913-2011.

Sayer, A.M., Thomas, G. E., & Grainger, R. G. (2008). A sea surface reflectance model suitable for use in (A)ATSR aerosol retrieval algorithms. *Atmospheric Measurement Techniques*, *3*, 813–838. http://dx.doi.org/10.5194/amt-3-813-2010.

Sayer, A.M., Thomas, G. E., Palmer, P. I., & Grainger, R. G. (2010). Some implications of sampling choices on comparisons between satellite and model aerosol optical depth fields. *Atmospheric Chemistry and Physics*, *10*, 10705–10716. http://dx.doi.org/10.5194/acp-10-10705-2010.

Seinfeld, J. H., & Pandis, S. N. (1998). *Atmospheric chemistry and physics: From air pollution to climate change*. New York: Wiley & Sons0-471-17815-2.

Sinyuk, A., Torres, O., & Dubovik, O. (2003). Combined use of satellite and surface observations to infer the imaginary part of refractive index of Saharan dust. *Geophysical Research Letters*, *30*, 1081. http://dx.doi.org/10.1029/2002GL016189.

Smirnov, A., Sayer, A.M., Holben, B. N., Hsu, N. C., Sakerin, S. M., Macke, A., et al. (2012). Effect of wind speed on aerosol optical depth over remote oceans, based on data from the Maritime Aerosol Network. *Atmospheric Measurement Techniques*, *5*, 377–388. http://dx.doi.org/10.5194/amt-5-377-2012.

Sofiev, M., Vankevich, R., Lotjonen, M., Prank, M., Petukhov, V., Ermakova, T., et al. (2009). An operational system for the assimilation of the satellite information on wild-land fires for the needs of air quality modelling and forecasting. *Atmospheric Chemistry and Physics*, *9*, 6833–6847. http://dx.doi.org/10.5194/acp-9-6833-2009.

Stamnes, K., Tsay, S. -C., Wiscombe, W., & Jayweera, K. (1988). Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Applied Optics*, *24*, 2502–2509.

Tanré, D., Kaufman, Y. J., Herman, M., & Mattoo, S. (1997). Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. *Journal of Geophysical Research*, *102*, 16 971–16 988.

Thomas, G. E., Chalmers, N., Harris, B., Grainger, R. G., & Highwood, E. J. (2012). Regional and monthly and clear-sky aerosol direct radiative effect (and forcing) derived from the GlobAEROSOL-AATSR satellite aerosol product. *Atmospheric Chemistry and Physics Discussions*, *12*, 18459–18497. http://dx.doi.org/10.5194/acpd-12-18459-2012.

Thomas, G. E., Poulsen, C. A., Sayer, A.M., Marsh, S. H., Dean, S. M., Carboni, E., et al. (2009). The GRAPE aerosol retrieval algorithm. *Atmospheric Measurement Techniques*, *2*, 679–701. http://dx.doi.org/10.5194/amt-2-679-2009.

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., et al. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, *118*, 847–855.

Veefkind, J. P., & de Leeuw, G. (1998). A new algorithm to determine the spectral aerosol optical depth from satellite radiometer measurements. *Journal of Aerosol Sciences*, *29*, 1237–1248.

Veefkind, J. P., de Leeuw, G., & Durkee, P. A. (1998). Retrieval of aerosol optical depth over land using two-angle view satellite radiometry during TARFOX. *Geophysical Research Letters*, *25*(16), 3135–3138.

Volten, H., Munoz, O., Rol, E., de Haan, J. F., Vassen, W., & Hovenier, J. W. (2001). Scattering matrices of mineral aerosol particles at 441.6 nm and 632.8 μm. *Journal of Geophysical Research*, *106*, 17,375–17,401.

von Hoyningen-Huene, W., Freitag, M., & Burrows, J. P. (2003). Retrieval of aerosol optical thickness over land surfaces from top-of-atmosphere radiances. *Journal of Geophysical Research*, *108*, 4260. http://dx.doi.org/10.1029/2001JD002018.

von Hoyningen-Huene, W., Yoon, J., Vountas, M., Istomina, L., Rohen, G., Dinter, T., et al. (2011). Retrieval of spectral aerosol optical thickness over land using ocean colour sensors MERIS and SeaWiFS. *Atmospheric Measurement Techniques Discussions*, *3*, 2107–2164. http://dx.doi.org/10.5194/amtd-3-2107-2010.

WMO/GAW (2003). *Aerosol measurement procedures guidelines and recommendations.* GAW report No. 153. Geneva, Switzerland: World Meteorological Organization Global Atmosphere Watch.

Yu, H., Kaufman, Y. J., Chin, M., Feingold, G., Remer, L. A., Anderson, T. L., et al. (2006). A review of measurement-based assessments of the aerosol direct radiative effect and forcing. *Atmospheric Chemistry and Physics*, *6*, 613–666. http://dx.doi.org/10.5194/acp-6-613-2006.

Zelazowski, P., Sayer, A.M., Thomas, G. E., & Grainger, R. G. (2011). Reconciling satellite-derived atmospheric properties with fine-resolution land imagery: Insights for atmospheric correction. *Journal of Geophysical Research*, *116*, D18308. http://dx.doi.org/10.1029/2010JD015488.

Zieger, P., Weingartner, E., Henzing, J., Moerman, M., de Leeuw, G., Mikkilä, J., et al. (2011). Comparison of ambient aerosol extinction coefficients obtained from in-situ, MAX-DOAS and LIDAR measurements at Cabauw. *Atmospheric Chemistry and Physics*, *11*, 2603–2624. http://dx.doi.org/10.5194/acp-11-2603-2011.