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

Article

Accepted Version

de Leeuw, G., Holzer-Popp, T., Bevan, S., Davies, W. H., Descloitres, J., Grainger, R. G., Griesfeller, J., Heckel, A., Kinne, S., Klüser, L., Kolmonen, P., Litvinov, P., Martynenko, D., North, P., Ovigneur, B., Pascal, N., Poulsen, C., Ramon, D., Schulz, M., Siddans, R., Sogacheva, L., Tanré, D., Thomas, G. E., Virtanen, T. H., von Hoyningen Huene, W., Vountas, M. and Pinnock, S. (2013) Evaluation of seven European aerosol optical depth retrieval algorithms for climate analysis. Remote Sensing of Environment, 162. pp. 295-315. ISSN 0034-4257 doi: 10.1016/j.rse.2013.04.023 Available at https://centaur.reading.ac.uk/36777/

Elsevier Editorial System(tm) for Remote Sensing of Environment
Manuscript Draft

Manuscript Number:

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

Article Type: Climate Change Initiative Special Issue

Corresponding Author: * Climate Change Initiative Guest Editors,

Corresponding Author's Institution:

First Author: Climate Change Initiative Guest Editors

Order of Authors: Climate Change Initiative Guest Editors; Thomas Holzer-Popp, Dr.; Suzanne Bevan; William Davies; Jacques Descloitres, Dr.; Roy G Grainger, Dr.; Jan Griesfeller, Dr.; Andreas Heckel, Dr.; Stefan Kinne, Dr.; Lars Klueser; Pekka Kolmonen, Dr.; Pavel Litvinov, Dr.; Dmytro Martynenko, Dr.; Peter R North, Prof. Dr.; Bertrand Ovigneur; Nicolas Pascal; Caroline Poulsen, Dr.; Didier Ramon, Dr.; Michael Schulz, Dr.; Richard Siddans, Dr.; Larisa Sogacheva, Dr.; Didier Tanre, Prof. Dr.; Gareth E Thomas, Dr.; Timo H., Dr.; Wolfgang von Hoyningen Huene, Dr.; Marco Vountas, Dr.; Simon Pinnock, Dr.

Highlights

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

- Seven European aerosol retrieval algorithms have been evaluated against independent data
- Scores for AATSR are similar to those for MODIS and MISR
- Further improvements for AATSR aerosol retrieval algorithms are recommended
- AATSR algorithms can be used for the production of AOD over land and ocean
- PARASOL and one MERIS algorithm can be used for the production of AOD over ocean

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

Gerrit de Leeuw[1,2,3,*], Thomas Holzer-Popp[4], Suzanne Bevan[5], William Davies[5], Jacques Descloitres[6], Roy G. Grainger[7], Jan Griesfeller[8], Andreas Heckel[5], Stefan Kinne[9], Lars Klüser[4,10], Pekka Kolmonen[1], Pavel Litvinov[13], Dmytro Martynenko[4], Peter North[5], Bertrand Ovigneur [6], Nicolas Pascal[6,12], Caroline Poulsen[11], Didier Ramon[l2], Michael Schulz[8], Richard Siddans[11], Larisa Sogacheva[1], Didier Tanré[13], Gareth E. Thomas[7], Timo H. Virtanen[1], Wolfgang von Hoyningen Huene[14], Marco Vountas[14], Simon Pinnock[15]

1. Finnish Meteorological Institute (FMI), Erik Palmenin Aukio 1, P.O. Box 501, FI-00101, Helsinki, Finland, email: gerrit.leeuw@fmi.fi
2. Department of Physics, University of Helsinki, Erik Palmenin Aukio 1, P.O. Box 64, FI- 00014, Finland Helsinki, Finland
3. TNO Environment and Geosciences, Dept. of Air Quality and Climate, P.O.Box 80015, 3508 TA Utrecht, The Netherlands
4. DLR German Aerospace Center, German Remote Sensing Data Center (DFD), D-82234 Oberpfaffenhofen, Germany
5. Global Environmental Modelling and Earth Observation (GEMEO), Department of Geography, College of Science, Swansea University, Singleton Park, Swansea SA2 8PP, U.K.
6. ICARE/CNRS, ICARE Data and Services Center, University Lille 1, Polytech Lille E206, 59655 Villeneuve d´Ascq Cedex, France
7. Atmospheric, Oceanic & Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, England
8. Research and Development Department, Norwegian Meteorological Institute, Postboks 43 Blindern, 0313 Oslo, Norway

9. Max-Planck-Institut für Meteorologie (MPI), Hamburg, Germany

10. Augsburg University, Experimental Physics II, Universitätsstr. 1, D- 86135Augsburg, Germany

11. STFC Rutherford Appleton Laboratory, Chilton OX11 0QX, U.K.

12. HYGEOS, Euratechnologies, 165 Avenue de Bretagne, 59000 LILLE, France

13. Laboratoire d'Optique Atmosphérique, CNRS, Universite Lille-1, Bat P5 Cite Scientifique, 59655 Villeneuve d'Ascq, France

14. Universität Bremen, Institute of Environmental Physics, Bremen, Germany

15. ESA-ESRIN, Frascati, Italy

*Corresponding author:

Prof. Dr. G. de Leeuw, Finnish Meteorological Institute, Climate Change Unit, P.O. Box 503, FI-00101 Helsinki, Finland

Mobile: +358 50 919 5458

FAX: +358 29 539 3146

email: Gerrit.Leeuw@fmi.fi

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. Seven algorithms, developed for the retrieval of aerosol properties using data from AATSR (3), 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 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 10x10km^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.

Keywords: Aerosol retrieval algorithms, Aerosol optical depth, AATRS, MERIS, PARASOL

1. Introduction

Satellite-based radiometers and spectrometers have been used for the observation of aerosol properties from space since more than three decades (e.g., de Leeuw and Kokhanovsky, 2009; Lee et al., 2009). The data have increasingly been used for purposes such as air quality assessment (Hoff and Christopher, 2009; van Donkelaar et al., 2010), emission estimates (Huneeus et al., 2012), forest fires applications (Kaufman et al., 1998; Labonne et al., 2007, Sofiev et al., 2009), atmospheric correction of oceanic (Müller et al., 2013) and terrestrial (Zelazowski et al., 2011) observations, etc.. In this paper we focus on the use of satellite instruments to provide aerosol observations for climate and climate change studies. In particular seven 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., 2012) 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.

After a brief introduction to aerosols, their effects on climate, the use of satellite-based instruments to obtain information on aerosols, algorithms developed for this purpose and a short overview of the Aerosol-cci project, the participating aerosol retrieval algorithms will be summarized followed by a summary of recent improvements which are described in detail in Holzer-Popp et al. (2013). The main focus of this paper is on methods used for the validation and evaluation of the aerosol retrieval algorithms in a round-robin (RR) exercise, the protocol used in this RR exercise to select the most suitable algorithm, or combination of algorithms, and the results from this exercise.

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

95 characterized by a certain radius (specified at a certain relative humidity, RH: dry (RH < 30-40%
96 (WMO-GAW, 2003), at RH=80%, or their *in situ* value at ambient RH. Satellites observe aerosol
97 properties in situ, usually integrating over the whole atmospheric column in which both RH and
98 aerosol concentrations may vary strongly. Ground-based measurements are prescribed to be made at
99 dry conditions (WMO-GAW, 2003). Aerosol particles may have sizes ranging from a few nm to
100 several tens of μm, can be composed of a wide range of chemical species (organic matter, inorganic
101 salts) which are either internally mixed (different species occur in one particle) or externally mixed
102 (each particle is composed of one single species) and mixed forms of these. Each size range may
103 have its own physical and chemical properties and based on these different 'modes' are considered,
104 such as cluster (a few nm), nucleation (ca. 5 nm), Aitken (some tenths of nm), accumulation (a few
105 hundreds of nm) and coarse (larger than 500 nm) particle modes, where the numbers in parenthesis
106 indicate dry mode radius (see eq. 1). The particle size distribution describes the variation of the
107 particle concentrations with size. The concentrations may be as large as 10^4 to 10^5 cm^{-3} for
108 accumulation mode particles in polluted conditions or as small as 10^{-5} cm^{-3} for the largest particles
109 (radius some tens of μm). Total concentrations, i.e. integrated over the whole size distribution, may
110 vary from a few 10s cm^{-3} in very clean conditions to up to 10^5 cm^{-3} in polluted conditions. Particles
111 can be directly produced by, e.g. mechanical (wind-blown dust, sea spray aerosol), biological
112 (pollen) or combustion (traffic, industry, fires) processes, or they can be produced from precursor
113 gases by gas-to-particle conversion processes and nucleation. Atmospheric aerosol particles have a
114 life time varying from hours to days, depending on their size, during which they undergo physical
115 and chemical changes which in turn changes their chemical composition and their optical and
116 physical properties. Of importance for climate and climate change are particles with dry radii
117 between ca. 30 nm to several μm because these particles are most effective for scattering of
118 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 and precipitation.

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

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

where each log-normal mode is defined by three parameters: aerosol number concentration $\boldsymbol{N_i}$, 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 (Yu et al., 2006; Thomas et al., 2012). 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

143 the oceans are sparsely covered. Satellite observations offer in principle global coverage, depending
144 on swath width, in about one day to a week.
145 The effect of aerosol particles on solar radiation are determined by the particle size distribution and
146 their size-segregated chemical composition, which together determine the angular scattering
147 (expressed as the phase function), absorption and single scattering albedo (ssa, the ratio of
148 scattering and the sum of total scattering and absorption), and the vertical variation of these
149 parameters. Scattering and absorption together determine the extinction of solar light by aerosol
150 particles and the extinction coefficient is the sum of the scattering and absorption coefficients.
151 Changes in global, regional and local effects of aerosol particles can thus be determined by changes
152 in these properties or a combination of them. The basic aerosol parameter retrieved from satellite-
153 based observations is the aerosol optical depth (AOD, or τ), i.e. the column-integrated extinction
154 coefficient specified for a certain wavelength, λ. AOD time series could thus be used to determine
155 trends indicating changes on regional to global scales. However, this requires that AOD can be
156 determined with sufficient accuracy to provide statistically significant trends. Such requirements
157 have been formulated by GCOS (ref) and were further formulated as part of the aerosol-cci project
158 described see in Section 2. In addition to AOD, other parameters are sometimes made available
159 from satellite observations with a varying degree of reliability and accuracy. These parameters
160 include the Ångström exponent (AE) describing the wavelength dependence of the AOD, the fine
161 mode fraction (FMF) describing the contribution of particles with dry radii smaller than 0.5 μm to
162 the total AOD, coarse mode fraction (CMF) describing the contribution of larger particles to the
163 total AOD, aerosol type (i.e. parameters describing the aerosol size distribution and optical
164 properties), absorbing aerosol index (AAI), ssa, absorbing aerosol optical depth (AAOD), aerosol
165 layer height. The determination of these other parameters usually requires an AOD exceeding a
166 certain value to obtain a reasonable value (e.g., Holzer-Popp et al. 2002a; 2002b; Kahn et al., 2010)

Instruments used for aerosol retrieval include spectrometers and radiometers with one or more wavebands in the UV/VIS and 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 such as SEVIRI (Spinning Enhanced Visibile and Infrared Imager) flying on geostationary satellites are currently used for aerosol retrieval. However, the results are often less accurate in comparison

192 with dedicated aerosol retrieval instruments. This may be somewhat surprising in cases where the
193 instrument characteristics are not limiting factors. For instance, the ATSR-2 / AATSR instruments
194 should potentially provide good results because of the dual view capability which allows for more
195 effective decoupling of the surface and atmospheric contributions to the top of atmosphere (TOA)
196 radiance than is possible with a single view, and because of the availability of wavebands from the
197 visible (VIS) to the thermal infrared (TIR) facilitating effective cloud screening.
198 All instruments, also those dedicated for the retrieval of aerosol and cloud properties, do provide
199 insufficient information to accurately determine all relevant aerosol properties, i.e. particle size
200 distribution, size-dependent particle shape and chemical composition, mixing state, from which the
201 optical properties could be determined. This is in part due to the lack of vertical resolution of
202 spectrometers and radiometers. These instruments observe the effect of aerosol particles integrated
203 over the whole atmospheric column while usually not only their concentrations may change with
204 height but also their chemical composition. In addition, as indicated above, particle sizes change
205 with varying relative humidity. Furthermore, the atmosphere may be stratified and in disconnected
206 layers with different origin and different history the aerosol properties may be different. This
207 situation is further complicated by the occurrence of absorbing particles, the effect of which on the
208 AOD depends on the altitude at which they occur.
209 As a result, the retrieval problem is underdetermined, i.e. there are more unknowns than
210 independent pieces of information to solve the radiative transfer equations and assumptions need to
211 be made. These include assumptions on the aerosol properties, using simplified descriptions of size
212 distributions and optical parameters and aerosol layer height. Furthermore the treatment of the
213 surface is very important, in particular over reflecting surfaces where the surface contribution to the
214 upwelling TOA radiance may be as strong as, or even much stronger than, the atmospheric
215 contribution. Over ocean the retrieval is often relatively simple because the ocean surface is dark at
216 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 often relatively dark at shorter wavelengths in the UV/VIS and at wavelengths in the UV all surfaces are dark. The latter has been used in the MODIS deep blue algorithm (Hsu et al., 2004). However, when UV wavelengths are not available or over brighter surfaces, other assumptions need to be made.

Aerosol retrieval algorithms utilizing 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, comparison of the AOD obtained from different algorithms may vary widely and some algorithms may perform better than others. 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 et al., 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 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 but 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 much less good 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 cloud so that it does not provide a good test of how well an algorithm behaves in the case where cloud has not 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 Section 3, where also brief descriptions are provided of the most important characteristics of each algorithm. Other instruments (MODIS, MISR) and algorithms 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 different 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 5. 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 (March, June, September and December) representing the different seasons. This paper describes the RR tests and results. Based on the RR exercise, the best possible algorithm, or combinations 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/.

2. The Aerosol-CCI project

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 engineer 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 modelling 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 brings the experience of data centres and experience with data format and data access. The Aerosol-cci project started in July 2010 and has duration of 3 years with a potential extension to 6 years.

3. Aerosol retrieval algorithms

The aerosol retrieval algorithms included in the Aerosol-cci project, 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 AATSR and 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 et al. (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 form 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 neighbouring 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 Section 4. 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. Default resolution is 10x10 km^2, but also 1x1 km^2 is

316 used in certain studies. The latest version of ADV/ASV including many improvements made at
317 FMI/UHEL and uncertainty characterization is described in Kolmonen et al. (2013).

318 ## AATSR ORAC

319 The Oxford-RAL Retrieval of Aerosol and Cloud (ORAC V1) algorithm is an optimal estimation
320 (OE) retrieval scheme designed to provide estimates of aerosol optical depth and effective radius,
321 cloud top pressure, height and temperature, cloud particle effective radius, cloud optical depth and
322 cloud type (generally liquid water or ice) from multispectral imagery (Thomas et al. 2009, Poulsen
323 et al., 2011; Sayer et al., 2011). The method fits all the shortwave forward and nadir radiances
324 simultaneous using a forward model based on the DISORT radiadive transfer code (Stamnes et al.
325 1998). It is worth noting that the simultaneous retrieval of all state parameters provided by the OE
326 method ensures that a physically consistent and numerically optimal estimate of the state is
327 produced. The quality of fit to the radiances allows the quality of the retrieval to be judged a
328 posteriori. In addition the error in the retrieved aerosol parameters is estimated by propagating both
329 the measurement and forward model uncertainties into state space. Note that the dataset described
330 here was produced by the ORAC V1 algorithm an *a priori* surface BRDF is set using MODIS
331 MCD43B BRDF products (Jin et al., 2003) over land and an ocean surface reflectance model over
332 the ocean (Sayer et al., 2010). More recent processing with an updated surface model is currently
333 under evaluation but initial indications show a substantial improvement when compared to V1.

334 ## SU ATSR algorithm

335 The SU-ATSR algorithm has been developed at Swansea University for estimation of atmospheric
336 aerosol and surface reflectance for the ATSR-2 and AATSR sensors. Over land, the algorithm
337 employs a parameterised model of the surface angular anisotropy, and uses the dual-view capability
338 of the instrument to allow estimation without a priori assumptions on surface spectral reflectance.
339 Over ocean, the algorithm uses a simple model to exploit the low ocean leaving radiance at red and
340 infra-red channels at both nadir and along-track view angles. The surface models are used to invert

341 the 6SV model (Kotchenova, et al., 2006; 2007) to perform retrieval at 10km resolution. The
342 algorithm has been implemented on the ESA Grid Processing on Demand (GPOD) system to allow
343 global processing and free download of AOD and surface reflectance. The method is documented in
344 North et al. (1999), North (2002), Grey et al. (2006a; b) and Bevan et al. (2009, 2012).

345 **SYNAER**

346 The synergistic aerosol retrieval method SYNAER delivers aerosol optical depth (AOD) and an
347 estimation of the type of aerosols in the lower troposphere over both land and ocean by exploiting a
348 combination of a radiometer and a spectrometer. The type of aerosol is estimated as percentage
349 contribution of 4 representative aerosol components (sea salt, mineral dust, weakly absorbing
350 accumulation mode and strongly absorbing accumulation mode aerosol). The high spatial resolution
351 including thermal spectral bands of the radiometer permits accurate cloud detection. The SYNAER
352 aerosol retrieval algorithm comprises of two major parts. In step 1 a dark field method exploits
353 single wavelength radiometer reflectances (at 670 nm over land, at 870 nm over ocean) to determine
354 36 values of the aerosol optical depth and surface reflectance over automatically selected and
355 characterized dark pixels for a set of 36 different pre-defined boundary layer aerosol mixtures. In
356 step 2 the parameters retrieved in the first step are used to simulate spectra for the same set of 36
357 different aerosol mixtures with the same radiative transfer code after spatial integration to the larger
358 pixels of the spectrometer. A least square fit of these calculated spectra at 10 wavelengths to the
359 measured spectrum delivers the correct AOD value (the one AOD retrieved in step 1 for the aerosol
360 type selected in step 2) and - if a uniqueness test is passed - the most plausible spectrum and its
361 underlying aerosol mixture. (Holzer-Popp et al. 2002a; 2008). Using a combination of 2 instruments
362 with different scan patterns SYNAER can only provide global cloudfree coverage every 12 days
363 and with large pixels of 60x30 km^2. However the combination of the 2 instruments has the potential
364 to provide aerosol type information (qualitatively shown in Holzer-Popp et al., 2008). Although
365 these method-inherent limitations mean a significant drawback in comparison to AATSR AOD

366 products, SYNAER has been included into the Aerosol_cci project in order to qualify and improve
367 its quantitative AOD results and thus eventually strengthen the aerosol type information.

368 **MERIS ESA Standard**

369 The MERIS standard aerosol retrieval over land algorithm was designed to work over Dense Dark
370 Vegetation (DDV) targets (Santer et al., 1999, Ramon and Santer, 2001). A set of DDV
371 Bidirectional Reflectance Function (BRF) models was assembled for 11 different biomes on Earth.
372 DDV detection is based on a threshold on the Atmospherically Resistant Vegetation Index (ARVI)
373 computed from Rayleigh corrected reflectances at 443, 665 and 865 nm. As DDV spatial cover is
374 low, the aerosol inversion was extended to brighter surfaces called Land Aerosol Remote Sensing
375 (LARS) targets (Santer et al., 2007). LARS spectral albedo can be predicted as it is linearly related
376 to ARVI. Slopes and offsets of these linear regressions are stored in Look Up Tables for 1°x1°
377 boxes and on a monthly basis. The aerosol retrieval consists in the inversion of the AOD at 443 and
378 665 nm that allow to reproduce the measured TOA reflectances at 443 and 665 nm using pre-
379 calculated aerosol scattering functions for aerosol models described by a Junge Power-Law (JPL)
380 size distribution and a constant refractive index of 1.45-0.0i. The outputs of the algorithm are the
381 AOD at 443 nm and the aerosol Ångström exponent derived between 443 and 665 nm.
382 Cloud contamination is the biggest issue of the product that is delivered at the same spatial
383 resolution as the level 1B data (i.e. 1.2 km). The product, with a good spatial coverage now, has
384 been validated only for the AOD at 443 nm. The Ångström exponent is not validated since the
385 retrieved AOT at 665 nm is noisy. It is mandatory to move toward spatial resolution of 10x10 km^2
386 for the aerosol product in order to reduce cloud contamination and enhance the Signal to Noise
387 Ratio (SNR) for the Ångström exponent retrieval. Finally there is a need for improving the LARS
388 BRDF model.

389 **MERIS ALAMO**

390 The MERIS ALAMO (Aerosol Load and Altitude from MERIS over Ocean) algorithm has been

391 primarily developed for aerosol altitude retrievals using MERIS data. Necessary inputs for altitude
392 retrievals, such as aerosol optical properties, are derived in a first step with an initial assumption on
393 the layer altitude. The cloud masking and AOD retrieval schemes are a close adaptation of the
394 MODIS algorithm (Tanré et al, 1997; Remer et al., 2005), using only the following MERIS bands:
395 510, 560, 665, 753.75 and 865 nm. Due to spectral characteristics of MERIS, ALAMO is limited to
396 a maximum wavelength of 865 nm and only two pieces of information on aerosol properties can
397 therefore be retrieved instead of three parameters with MODIS. MERIS aerosols products are
398 retrieved with a spatial resolution of 10x10 pixels (12x12 km^2). This resolution allows (i) an
399 adequate signal-to-noise ratio (SNR) for a better characterisation of the aerosols type and (ii)
400 rejection of pixels considered as non-valid through statistics criteria, in order to ensure the quality
401 of the aerosol product. The aerosol products of ALAMO include the optical thickness and the
402 mixing ratio of fine and coarse modes. Aerosol models used for ALAMO are the same as the ones
403 used for the most current version of MODIS products.
404 In a second step the altitude of the aerosol layer is estimated using the MERIS O_2 A absorption
405 channel and following the algorithm described in Dubuisson et al. (2009). A pixel reclassification is
406 done after the altitude retrieval to remove high thin clouds based on a threshold on altitude and
407 spatial variance of altitude.

408 **MERIS BAER**

409 The Bremen Aerosol Retrieval, BAER, has been developed to derive spectral AOD from
410 multispectral satellite imagery such as from MERIS over ocean and land. It separates the spectral
411 aerosol reflectance from the surface and Rayleigh path reflectances for the short-wave ($\leq$ 0.67 μm)
412 TOA reflectance over land. Over ocean the whole spectral range of MERIS is utilized for the AOD
413 retrieval.
414 The surface reflectance is estimated by a linear mixing of vegetation and non-vegetation spectra
415 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 et al., 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 et al., 2003; 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 look up tables (LUT) 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 et al, 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 at 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 and Munk, 1954) and a wind speed of 5m/s. The foam contribution is calculated according to Koepke's model

441 (Koepke , 1984) and a constant value of 0.22 for the foam reflectance.

442 The PARASOL aerosol algorithm over land (Deuzé at al., 2001) is based on a best fit between
443 polarized measurements at 650 and 865nm and LUT's simulated for aerosols within the
444 accumulation mode only. The surface contribution is estimated from a relationship using empirical
445 coefficients adjusted for the different classes of land surfaces according to the main IGBP biotypes
446 and the NDVI (Nadal and Bréon, 1999).

447 4. Algorithm improvement

448 Aerosol retrieval is an underdetermined problem since the number of degrees-of-freedom, i.e.
449 parameters describing the aerosol properties which determine the observed TOA radiances, is
450 smaller than the number of observations. Hence assumptions need to be made. The most important
451 assumptions made in aerosol retrieval concern:

452 • Cloud screening
453 • Surface treatment
454 • Aerosol optical properties and size distribution

455 Aerosol retrieval can only be made for cloud-free sky because the high reflectance of clouds at
456 wavelengths in the UV-NIR interferes with the aerosol reflectance and hence prohibits accurate
457 retrieval of aerosol properties. Therefore, an accurate cloud mask has to be applied to screen all
458 pixels for the occurrence of clouds and exclude them from retrieval. Currently all algorithms
459 participating in Aerosol-cci use their own cloud detection procedures as described in section 3 and
460 the literature referenced there. The use of a common cloud flag for similar products is under study
461 (Holzer-Popp et al., 2013). To further eliminate cloud-contaminated data, a post-processing step has
462 been developed to effectively detect cloud edges as described in Section 3 for the AATSR ADV and
463 ASV algorithms. This post-processing step results in a smoothly varying AOD across extended
464 areas without sudden transitions. This post-processing step has been implemented in other
465 algorithms (ORAC, SU) as well.

466 The treatment of the surface and accounting for contributions of surface reflectance to the radiance
467 measured at TOA depends on the instrument properties, and how they are used. An overview of
468 surface treatment and application to different algorithms such as the AATSR algorithms used in
469 Aerosol-cci has been presented in Kokhanovsky and de Leeuw (2009), for MERIS BAER in de
470 Leeuw et al. (2011), and for the other MERIS algorithms in the respective ATBDs. Therefore
471 surface treatment will not be discussed here.
472 Apart from improved cloud screening, most progress has been made in harmonizing aerosol models
473 and their use in the various retrieval algorithms. For the Aerosol-cci project, a simple set of four
474 aerosol components has been developed consisting of two fine mode components, one of which has
475 a complex refractive index representative for weakly absorbing aerosol particles and the other one
476 represents strongly absorbing aerosol particles. The other two components describe coarse mode
477 aerosol components, one with the characteristics of desert dust and the other one describing sea salt
478 aerosol. Each component is thus described by a lognormal size distribution, defined by mode radius,
479 effective radius, geometric standard deviation and variance, and by the complex refractive index
480 (Table 2).
481 The two fine mode-types are extremes in terms of absorption and reality (in terms of absorption) is
482 always a combination of these two types. The choice of the fine mode radius is based on an analysis
483 of AERONET sun-photometer data which shows that the most frequent fine mode size (in terms of
484 the effective radius) is near 0.14 μm. The coarse mode is dominated by two quite different aerosol
485 types: spherical largely non-absorbing sea-salt and non-spherical absorbing dust. Based on an
486 AERONET probability distribution for the coarse mode, the effective radius was set to 1.94 μm for
487 these two coarse mode aerosol types. See Holzer-Popp et al. (2013) for more detail.
488 The optical properties of aerosol particles are usually calculated by application of a Mie code (Mie,
489 1908), which applies to spherical particles. However, for dust Mie codes cannot be applied because
490 of the non-spherical shape of dust particles. In Aerosol-cci a T-matrix method was used assuming

491 randomly oriented spheroids with aspect ratios between 1.44 and 3.0 (Dubovik et al. 2002, Sinyuk
492 et al. 2003). Although spheroids may be unable to represent the entire shape complexity for dust,
493 this spheroid method is certainly preferable over methods for spheres. The choice of the refractive
494 index for dust is based on Volten et al. (2001). Observational data (Dubovik et al. 2002, Sinyuk et
495 al., 2003) demonstrate that the dust absorbing strength is wavelength dependent, and decreases from
496 the UV (imaginary refractive index, RFi, near 0.005) to the near-IR (RFi near 0.001). To avoid
497 time-consuming computations during the retrieval, radiative transfer is computed in atmospheres
498 with different aerosol components, for discrete AOD values and a range of discrete configurations
499 (e.g., solar zenith angle, viewing angle), and the results are stored in a look-up table (LUT). During
500 the retrieval the optical properties for the relevant configuration are obtained by simple
501 extrapolation of the LUT values.
502 For successful retrieval of the aerosol type by using a mixture of the four basic aerosol components
503 presented in Table 2, additional information may be required on relationships between fine and
504 coarse mode, between less and more absorbing fine mode and between dust and sea-salt
505 components in the coarse mode. This information is supplied in terms of monthly $1^{o}x1^{o}$
506 climatological data derived from two sources, modelling and observations.
507 MODELING: Output of 14 different global models with complex aerosol component that
508 participated in AeroCom exercises are combined into 'AeroCom' median maps (Kinne et. al.,
509 2006). Based on these median maps, ratios between different aerosol components are defined. Dust
510 and sea salt generally define the coarse mode and sulfate, organic matter and black carbon the fine
511 mode.
512 CLIMATOLOGY: To improve this AeroCom model median, AERONET (Holben et al., 1998)
513 quality data are added in a merging process for AOD, Ångström exponent (describing the AOD
514 spectral dependence) and single scattering albedo (describing the absorption potential). With
515 observational ties data of this 'climatology' are recommended over data from 'modelling'.

516 This climatology is used as a priori for the occurrence of aerosol types /mixtures, per region and per
517 month. In general the coarse mode component selected would be sea salt, except in the presence of
518 desert dust which occurs in certain areas. The choice of the fine mode component would also be
519 based on the climatology and the two fine mode components, with equal microphysical properties,
520 could be mixed to obtain the desired absorption properties (as provided by the ssa in the
521 climatology). Using the occurrence of aerosol types, the retrieval algorithm computes the radiances
522 at the top of the atmosphere which are compared with the satellite measurements. Based on this
523 comparison the aerosol mixtures are adjusted and the procedure is iterated until convergence is
524 reached and the most likely aerosol model providing the measured radiance is selected. With this
525 model the AOD is computed. It is emphasized that the climatological AOD is not used in the
526 retrieval process, and the aerosol mixtures are only used as *a priori*, except in sensitivity studies.
527 The actual AOD and aerosol mixtures are retrieved based on the measured radiances at 3 (over
528 land) or 4 (over water) wavelengths.
529 Algorithm improvement was measured by application of the validation and evaluation exercises
530 described in Section 5. These exercises were made for only one month, September 2008, a
531 necessary restriction because of the time it takes to run the retrieval with different aerosol mixtures.
532 Success was identified by comparison with the baseline algorithm and successive improvement
533 after implementation of different aerosol models, the use of the AEROCOM median with different
534 degrees of comprehensiveness (i.e. varying from completely free retrieval without any use of the
535 climatology, to a full prescription of the aerosol mixing, and combinations thereof) and different
536 cloud masks. In addition to these experiments, algorithms were also improved as regards coding and
537 debugging and the retrieval products were improved by application of post-processing. Results from
538 this study for 1 month are presented in Holzer-Popp et al. (2013).

5. 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) as described in the introduction. All satellite results, both those participating in the Aerosol-cci RR and the reference satellite data sets, are evaluated versus AERONET.

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 10x10km^2 and L3 are daily or monthly aggregates (also referred to as mean or averaged data) provided on a spatial scale of 1^ox1^o. 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

564 and analysis tool has been developed. All key parameters of each sensor product can be selected
565 independently for visualization. For each product, a link to the product documentation is
566 provided. For a given parameter, a unique colour scale is used for direct visual inter-
567 comparison. The geographic selection, date selection, and product selection, can be modified
568 independently, while the other two selection criteria remain unchanged. All data sets are displayed
569 in Plate-Carree projection to make inter-comparison and geographic selection straightforward.
570 Aerosol-cci daily and monthly L3 products don't require any reprojection. Aerosol-cci L2 products
571 are originally produced in sinusoidal grid, so they are re-gridded on-the-fly upon selection in the
572 graphical interface. Additional interactive capabilities are available, such as display of data values,
573 or X/Y plot comparison. The multi-sensor visualization and analysis tool is available from
574 http://www.icare.univ-lille1.fr/browse/?project=cci.
575 An extract tool has been developed to interactively extract Aerosol-cci product values in the vicinity
576 of validation sites. Several validation networks are supported, including AERONET. Single or
577 multiple parameters from Aerosol-cci aerosol products can be selected for extraction. A time period
578 and a search radius can be specified. For each selected validation site, and each overpass of the
579 satellite, all data values found within the specified range are displayed, if any, along with the
580 corresponding acquisition time and pixel location. Those extract values can be directly compared to
581 validation data off-line. The Aerosol-cci extract tool is available from http://www.icare.univ-
582 lille1.fr/extract/cci.

583 ## 5.1 L2 statistics

584 AOD and Ångström exponent of L2 data sets were compared with AERONET data using scatter
585 plots and least squares fits to the data. The comparisons were made for collocated satellite and
586 AERONET observations, i.e. satellite pixels were selected within a spatial threshold of +/-35 km
587 and a time frame of +/-30 minutes from AERONET measurements. Where available (ORAC, ADV
588 and SYNAER), quality flags or confidence indicators in the products were taken into account to

589 select best pixels. Furthermore a distinction was made between retrievals over land and water.
590 Round Robin MERIS datasets do not have a water/land flag, therefore the pixels over land and
591 ocean for MERIS Standard and MERIS BAER were selected using the ORAC water/land flag.

592 ## 5.2 L3 scoring

593 For L3 scoring, an evaluation routine has been developed to determine for a test data set a
594 performance error, for cases when trusted reference data are available. Here the test data are daily
595 L3 satellite data, the reference data are AERONET observations within half an hour of the
596 particular satellite overpass. To simplify comparisons, all sun photometer data were gridded on the
597 spatial 1^{o} x1^{o} resolution of the satellite data. Although in theory satellites should locally offer more
598 than 100 samples for the four months, the available number of valid data points is smaller due to the
599 presence of clouds. The number of samples is further reduced due to a limited swath (e.g. AATSR
600 and MISR), stringent quality control measures (e.g. SU) or due to limited temporal coverage (e.g.
601 SYNAER). In addition, also AERONET data were not available each day.
602 The selected performance error for L3 evaluation is based on a combination score, which separately
603 investigates temporal variation, spatial variation and bias. Errors E are defined to range from 0 for
604 'perfect' to 1 for 'poor'. Conversely, associated scores S (S=1-E) range from 0 for 'poor' to 1 for
605 'perfect'. This definition for the scores allows via sub-score multiplication the determination of an
606 overall score S_T and of an overall error E_T.
607
608 $$\mathbf{E_T} = 1 - |\mathbf{S_T}| \quad \text{with} \quad \mathbf{S_T} = \text{sign}(\mathbf{E_B}) * (1-|\mathbf{E_B}|) * (1-\mathbf{E_V}) * (1-\mathbf{E_S}) , \qquad (2)$$
609
610 where $\mathbf{E_B}$ is the error for bias, $\mathbf{E_V}$ is the error for spatial variability and $\mathbf{E_S}$ is the error from temporal
611 or seasonal variability. Note, that the sign of the bias defines the sign of the total score $\mathbf{S_T}$. Each of
612 the three sub-scores is based on statistics. Hereby, valid sub-scores require a minimum number of
613 samples. Given sufficient data-pairs for test-data D and reference-data R, the bias score $\mathbf{E_B}$

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.

$$\mathbf{E_B} = w^* [(D_{SUM}-R_{SUM}) / (D_{SUM}+R_{SUM})], \quad w=[_{IQ\text{-}R}D+_{IQ\text{-}R}R]/[_{IQ\text{-}A}D+_{IQ\_A}R] \tag{3}$$

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 error, to avoid an error 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 error $\mathbf{E_V}$ and the temporal variability error $\mathbf{E_S}$. The spatial variability error 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 error.

$$\mathbf{E_V} = w^* (1-R_C)/2 \tag{4}$$

$$\mathbf{E_S} = w^* (1-R_C/)/2. \tag{5}$$

With this definition 100% correlation yields no error, whereas 100% anti-correlation yields the maximum error of 1. Note, that randomness for temporal and spatial variability yields still scores of 0.5 (not zero).

Errors 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 error and bias error are determined at each location, by applying available time-series data pairs at the

639 smallest temporal resolution. In step 2, spatial error and bias error are determined by exploring data
640 pairs in their spatial context for each time step. The final bias error is averaged from both
641 processing steps.
642 Since properties usually vary with longitude / latitude and surface conditions, the evaluation is
643 regionally stratified. Later these regional scores can be easily combined via average weighting into
644 a single global score. Thus, this method offers an assessment via a single global (or regional) score,
645 while still maintaining regional diagnostics on bias the ability to match temporal and spatial
646 variability.

647 ### 5.3 Level3 validation using AEROCOM methods

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

655 ## 6. Round Robin exercise

656 The Round Robin exercise was set up for an independent and objective evaluation of the global
657 retrieval results (AOD, AE) provided by each of the algorithms indicated in Section 3. The versions
658 of the algorithms used to provide these products were selected by each of the retrieval groups based
659 on the exercises described in Holzer-Popp et al. (2013) and summarized in Section 4. The results
660 were evaluated using the tools described in Section 5. Based on these results, the independent
661 validation team (Section 2) provided an advice as regards the statistical quality. Other
662 considerations were data coverage and spatial patterns. In addition, the same tools (Section 5) were
663 applied to data from MODIS Aqua, MODIS Terra and MISR, for intercomparison and as a measure

of how well the Aerosol-cci algorithms are performing 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 2);
- 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 5
    - 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 3 and discussed below.

6.1 Level2 validation

For the L2 evaluation of AOD and AE provided by the participating algorithms 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

689 sites and satellite pixels used. Examples of scatterplots between the satellite-retrieved AOD and AE
690 vs. AERONET data are shown in Figures 1 and 2, together with the fit parameters. These figures
691 illustrate that there are differences amongst the several AATSR algorithms, both over ocean and
692 over land, and between the AATSR and PARASOL results over ocean. In all cases over ocean
693 satellite AOD is reasonably well correlated with AERONET, although outlyers are observed for
694 ORAC, which may be due to insufficient cloud screening. Over land SU AOD is well correlated
695 with AERONET and the slope is close to 1, but for ADV and ORAC the correlation is less good
696 than over ocean.

697 Figure 1 about here

698 Figure 2 about here

699 Correlations of AE are much smaller than for AOD, especially over land where in most cases there
700 is no correlation. Over ocean the correlations are much better and the PARASOL AE seems to
701 follow AERONET values reasonably well. It is not clear why the SU results are not at all correlated
702 with AERONET over ocean and AE's are mostly very close to zero.

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

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

710 The application of the criteria leads to the following rankings for the algorithms using AATSR or
711 MERIS data (see also Table 3):

712 AATSR over ocean: ADV, ORAC, SU, SYNAER

713 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 4. These statistics show that from all participating algorithms, over ocean PARASOL shows 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 shows good correlations, whereas MERIS has clearly weaker correlations and larger standard deviations, with only slightly larger pixel numbers.

6.2 L3 scoring

The evaluation of L3 data as described in section 5.2 was separately conducted for 25 (TransCom; Gurney, et al. 2002) sub-regions shown in Figure 3. 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^{o}x1^{o} grid boxes in each region with AERONET coverage.

Figure 3 about here

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 5 shows the resulting satellite AOD retrievals scores.

Table 5 about here

Table 5 indicates that the data volume 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 an even smaller swath of about 360km compared to about 500km for

738 AATSR). This is also illustrated by the number of samples that contribute to the scores for North
739 America, where also sub-scores for bias, temporal variability and spatial variability are listed.
740 Among the different Aerosol-cci AOD retrievals the ATSR products show the highest skill but total
741 and sub-scores vary. However, the comparison of the scores is limited. Global scores are based on
742 different numbers of regions. And also more appropriate comparisons for North America, where
743 almost all products supply a score, the underlying numbers of data-pairs differ.
744 For North America, ADV is ranked before SU and ORAC. The ADV score (.54) matches the MISR
745 score and both the ADV, MISR and SU (.48) scores are better than the MODIS scores (.42/.41)
746 which are particularly poor over America. Looking at the sub-scores, the relatively low ORAC
747 score (.39) has a bias score that is as good as in MISR or MODIS and clearly better than for SU.
748 The sub-scores also indicate that ADV and SU display spatial distributions for North America
749 which are superior among the examined data sets, even better than MISR or MODIS.
750 Calculated regional errors, as well as contributing sub-errors due to spatial variability among
751 MODIS, MISR and ADV are compared in Figure 4. The same error comparisons among the three
752 ATSR products are presented in Figure 5.
753 Figure 4 about here
754 Figure 5 about here
755 Figure 4 indicates that ADV errors for North America and Europe (where a sufficient amount of
756 AATSR data are available) are as low as for MISR (v22) and better than for MODIS (Collection
757 5.1). However, as mentioned above, the data volume of MISR and ADV is much smaller than that
758 offered by MODIS, mainly due to their narrower swaths. It is further noted that the ADV data
759 volume is similar to that of the MISR data, despite the larger AATSR swath. In that sense it should
760 be noted that the value of satellite products is not only determined by its accuracy alone but also by
761 (frequent, global) coverage.

762 The comparison of the three AATSR algorithms shows that the data coverage is poor for the SU
763 product. Clearly efforts are needed for better coverage to make these data sets more attractive to
764 users. For regions with available scores it could be concluded that the ADV product scores best and
765 that the ORAC product scores poorest, despite having a relatively low bias error. Still, this is just
766 based on an analysis for two regions dominated by urban-industrial aerosol and there are much
767 more facets to aerosol (e.g. dominance by dust or biomass burning).
768 Clearly these initial comparisons leave many open questions. Most disturbing is that there are so
769 many regions where no scores could be calculated for this limited data set. This can be addressed
770 once data are provided for one entire year or more. Also reference data over oceans are needed and
771 will be added in future assessments (e.g. using data from the marine aerosol network (Smirnov et
772 al., 2012) or using trusted and matured satellite AOD products).

773 **6.3 Level3 validation using AEROCOM tools**

774 The AeroCom tools allow for the selection of regions (World, Europe, China, India, E. Asia, N
775 Africa, N. America, S. America, World w/o mountains) and annual (only a four month average for
776 2008 in this RR), seasonal (represented by 4 different months), and monthly L3 (4 months in 2008)
777 averages. A common $1^{o}x1^{o}$ mask was established where valid data were available from all retrieval
778 algorithms (AATSR and MERIS). In all regions this info is further refined using an ocean, coastal
779 and land mask, based on whether a grid point was identified as purely ocean or land across
780 retrievals (using the ORAC L2 land/sea mask). Remaining grid points are defined as coastal.
781 Altogether 8 x 4 regional selections are possible. For comparison, similar statistics are available for
782 the annual averages of MODIS-Terra, MODIS-Aqua and MISR AOD data, with selection by
783 region. For each selection a list is produced showing the statistics, cf. Figure 6 as an example.
784 Examples of the results are presented in Figures 7 – 12.
785 Fiigure 6 about here

786 Figure 7 shows the global annual mean AOD maps for the algorithms participating in the Aerosol-
787 cci RR, as well as reference AOD maps from MODIS v5.1 (Terra and Aqua), MISR and the
788 AEROCOM median. As compared with the baseline algorithms (Holzer-Popp et al., 2013), the
789 current results are much closer to each other and also closer to the references. Yet, also quite large
790 differences are observed, both as regards the global coverage, the number of valid pixels (provided
791 with the statistics given with Figure 9), the spatial distributions and the features in each of the maps.
792 Clearly, ADV provides the smallest global coverage, which is also reflected in the number of valid
793 pixels which is smaller than for ORAC and MERIS Standard, but larger than for SU. The small
794 number of ADV pixels is due to the facts that (a) ADV limits the retrieval to solar zenith angles of
795 65^{o} and (b) no retrieval is made over bright surfaces. The even smaller number of pixels provided
796 by SU, in spite of the larger global coverage, is due to a stricter quality control. Further, there are
797 clear differences in the global mean AOD (provided for each algorithm in the legend at the top at
798 the right), which vary from 0.154 for ADV to 0.215 for SYNAER, as compared to MODIS mean
799 AOD values of 0.189 (Terra) and 0.179 (Aqua) and MISR (0.176).
800 Over land there are clear differences in the AOD distributions, such as at high northern latitudes
801 where the AOD provided by ORAC and MERIS Standard are clearly higher, SYNAER is a bit
802 higher, and ADV provides distributions similar to those from the reference satellites. SU, on the
803 other hand, provides AODs which are substantially lower. It is noted that the AEROCOM median
804 shows somewhat lower AODs at northern latitudes, with a clear gradient over Siberia, than the
805 reference satellites. Over western Europe, most AOD maps show enhanced values, higher than
806 further north, except ORAC and MERIS Standard, while the reference satellites show no
807 enhancement over western Europe with respect to northern latitudes, in contrast to the AEROCOM
808 median. Over N. America the patterns are quite different between different satellites (both Aerosol-
809 cci and references). The AOD is lower in the west for ADV, ORAC and SU, but there are clear
810 differences between these algorithms as regards the patters, and the values for ORAC are higher all

over the continent. The lower AOD in the western USA is in agreement with 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 but 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 Marine 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^{o} and 60^{o}, which is reproduced to some extent by ORAC, somewhat less clear by ALAMO and weakly by the reference satellites.

835 The overall picture emerging from these maps is that the ADV AOD distribution is closest to that of
836 the reference satellites, both over land and ocean, but ADV does not provide any data at the higher
837 latitudes resulting in a the global coverage which is much less (ca. 30%) than for some other
838 Aerosol-cci algorithms. The global mean AOD produced by most algorithms is different but locally
839 the differences are much larger, and these local differences are to some extent cancelled in the
840 global mean. Therefore it is useful to also look at regional differences to learn the strengths and
841 weaknesses of each algorithm and thus improve the algorithms. Features over land, such as forest
842 fire, desert dust and anthropogenic pollution plumes usually are smoothly extended over ocean but
843 in many cases land-sea transitions are visible. This is clearly a point for future research.
844 Figure 7 about here
845 Monthly AOD maps, for one month in each season selected for this RR exercise, are presented for
846 ADV and PARASOL in Figure 8. The features are similar as those discussed in Figure 7, but there
847 are clear differences between seasons in relation to the production and removal of different aerosol
848 types. This is most clearly illustrated with the biomass and desert dust plumes generated over Africa
849 and transported over the Atlantic Ocean. There are also clear differences in the AOD distributions
850 over the continents such as over Asia (China, India, deserts) and adjacent downwind oceans.
851 Differences are also visible over N. America (features discussed in connection with Figure 7) and
852 over S. America which is likely connected with biomass burning in Amazonia. In addition there are
853 differences between coverage caused by the seasonal variation of the solar zenith angle.
854 Fiigure 8 about here
855 Examples of the AEROCOM statistical analysis of the Aerosol-cci results for the 4 months in 2008
856 are presented in Figures 9, 10 and 11. Figures 9 and 11 include MODIS Terra evaluation results as a
857 reference. MODIS Terra was selected here because the overpass time is close to that of ENVISAT
858 with AATSR and MERIS. Figure 9 shows scatterplots of the retrieved AOD vs. AERONET values.
859 The statistics are provided in the legend in the upper left corner of each plot. 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. Figure 9 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 Figure 10 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 Figure 10 have been used to provide a ranking between the four AATSR algorithms, Table 6. The numbers in Table 6 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.

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.

Fiigure 9 about here

Fiigure 10 about here

Table 6 about here

Figure 11 shows the statistics in a different way, as histograms of the frequency of occurrence of the AOD values retrieved form the satellite observations, compared with collocated AERONET observations. Ideally, the two curves should exactly coincide. However, even MODIS, with 1468 collocations (note that these data are available for the whole year 2008, i.e. no selection was made to cover only the 4 months used for the RR algorithms, or to select on collocations with Aerosol-cci satellite data) 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. For the Aerosol-cci algorithms, covering only 4 months and thus having much less collocations, the

885 histograms show larger variations between bins. Yet, ALAMO and PARASOL, with over-ocean
886 retrieval only, follow the AERONET pattern quite well, with a tendency for PARASOL to
887 underestimate the lowest AODs. It is noted here that only AERONET data (i.e. land based) were
888 used in this analysis, i.e. sun photometers situated at or near the coasts, which may result in some
889 bias. From the other algorithms, the ADV-retrieved AOD histogram follows that of AERONET
890 quite well; this can also be said from SYNAER although the histograms are very flat. In the other
891 algorithms deviations are visible with either overestimation of the lower AOD values (SU) or
892 underestimation (ORAC) whereas MERIS Standard appears to have the largest deviations for both
893 small and large AOD.
894 Fiigure 11 about here
895 Another way to use the statistics is to evaluate where algorithms perform well and where
896 improvements are needed. An example is presented in Figure 12 where the difference between the
897 satellite and AERONET AOD observations, given by (Satellite -AERONET)/AERONET*100, is
898 colour-coded on the map for individual AERONET stations across the world. Blue indicates that the
899 satellite is underestimating; red indicates that the satellite is overestimating. Light colours indicate
900 that the differences are very small and as the colour is darker the under of over-estimation is larger.
901 For these plots, the globe has been gridded into boxes of 10^{o}x10^{o}. For each grid box, the stations
902 located there are taken into account, and their data is then averaged depending on the time period to
903 be plotted. When smaller regions are considered instead of the whole globe, the grid is reduced to a
904 1^{o}x1^{o} grid. The plot does not show how many stations are within the grid box. The locations of the
905 grid boxes (especially at the plot for the whole globe) are not entirely correct, because while the
906 outline of the continents and countries are correct according to the map projection, the grid boxes
907 are not. This would require a reverse map projection which would destroy the analysis grid because
908 of associated interpolations. The error becomes greater as the boxes are further away from the
909 equator and the zero meridian.

910 The maps show that none of the algorithms, including the MODIS reference, is perfect everywhere.
911 Improvements are needed, but areas for which the improvements are needed, and in which direction
912 (under- or over-estimation), are different for each algorithm. Taking the AATSR algorithms as an
913 example, ADV and SU appear to perform reasonably well over Europe, even though they tend to
914 underestimate, whereas ORAC and SYNAER have large overestimation. The same pattern emerges
915 over the USA, except that SYNAER appears to work quite well over the eastern USA. Almost all
916 algorithms show a large overestimation at high latitudes, except ALAMO in the northern
917 hemisphere. Because AERONET stations are located over land, or in coastal areas, this evaluation
918 cannot be made over ocean.
919 Figure 12 about here
920 Figure 13 shows the zonal mean AOD for the Aerosol-cci algorithms and MODIS Terra, with
921 AERONET for comparison. Together with Figure 11, this figure illustrates the performance of each
922 algorithm. As in Figure 11, ideally the satellite-retrieved AOD would follow the AERONET
923 observations, as for MODIS Terra in Figure 13. Also the over-ocean only AODs provided by
924 ALAMO and PARASOL show a quite good behaviour. However, for the other algorithms, which
925 include both land and ocean in the plots in Figure 13, the trends are reproduced well with high AOD
926 north of the equator and lower AODs toward the poles, but quantitatively there are differences.
927 MERIS Standard is in the (sub-) tropics very close to AERONET, but at mid-latitudes the AODs
928 are much higher than those from AERONET. Similar observations can be made for SYNAER, but
929 the SYNAER AOD shows an increasing trend from south to north as opposed to all other
930 observations. From the other AATSR algorithms, ADV deviates quantitatively most from
931 AERONET, whereas ORAC follows AERONET quite well but peaks right at the equator and has
932 much higher values both at -60^{o} and in the far north. SU seems to give the best performance in this
933 comparison except for the very high values in the far north.
934 The overall ranking resulting from the evaluation with the AEROCOM tools is given in Table 3.

935 Figure 13 about here

936 # 7. Discussion

937 The combined effort of European aerosol retrieval teams, supported by MODIS and MISR retrieval
938 specialists participating in workshops and discussion meetings, has resulted in an enormous
939 improvement of the retrieval algorithms and the products resulting from them. These efforts have
940 been described in Holzer-Popp et al. (2013) and were briefly summarized in Section 4 of this paper
941 which is focused on further improvement and algorithm inter-comparison with the goal to use the
942 algorithms for climate studies. This application requires a very high accuracy as formulated by
943 climate users, and the inclusion of uncertainties per pixel. To evaluate algorithm performance, as
944 judged by the evaluation of their products, in this case mainly the AOD and to a lesser extend the
945 wavelength dependence of the AOD expressed by the AE, methods have been developed as
946 described in section 5. The results from the application of these algorithms are presented in Section
947 6. The evaluation of these results shows, in a quantitative way, based on results from statistical
948 methods and additional evaluation using subjective but informed methods based on existing
949 knowledge of how aerosol varies on regional and global scales, the good performance of the
950 PARASOL (v0.23a) and the MERIS ALAMO (v1.0) algorithms over ocean, and the improvement
951 of the AATSR ADV, SU and ORAC algorithms for use over both land and ocean. The other
952 algorithms, SYNAER and MERIS Standard, need further improvement before they can be used to
953 provide parameters useful for climate studies. MERIS BAER needs further improvement with
954 respect to cloud screening and the consideration of absorbing aerosols. This situation led to the
955 conclusion that, in view of their good performance, the PARASOL and the MERIS ALAMO
956 algorithms can be used for the retrieval of AOD over ocean and thus provide 10 years (MERIS) and
957 7 years and more (PARASOL) global data series.
958 For AATSR, all three algorithms using only AATSR data (ADV, ORAC, SU) show good
959 performance, although there are regional and seasonal differences. However, there is not one

960 algorithm which performs best everywhere, as shown by the rankings provided in Section 6 from
961 each of the three different methods. Overall, ADV appears to provide the best scores, and compares
962 most favourably with the reference satellite data sets, but it does not provide any retrieval over
963 highly reflecting surfaces. SU does provide retrievals over highly reflecting surfaces but the number
964 of data points is very small, mainly due to the application of stricter quality control than ORAC and
965 ADV. However, this stricter quality control does not lead to the highest scores everywhere. ORAC
966 is potentially the most consistent algorithm in a mathematical sense, however both statistically and
967 as regards the reproduction of features it performs less well than ADV and SU. Yet, the low AOD
968 over ocean seems to be in line with results published by Smirnov et al. (2012). Nevertheless, ORAC
969 does not rank highest over ocean, which may be due to the lack of independent validation data over
970 open ocean which could confirm the low AOD observed by ORAC. Based on the current RR
971 results, ADV ranks first, followed by SU with ORAC as third. Yet, the differences are so small, that
972 the ranking may change when further improvements are implemented. Furthermore, the ranking
973 may be influenced by uncertainties introduced by L3 sampling methods as discussed in Sayer et al.
974 (2010).
975 Much of the difference between algorithms and their scoring may be due to cloud masking. The
976 different cloud masks used by each of the algorithms slightly complicates the like for like
977 comparison especially as common filtering may not completely account for possible differences in
978 cloud masks or thresholds used at the 10x10 km^2 retrieval level. As pointed out in the introduction,
979 the comparison with AERONET data, which are well screened for cloud occurrence, does not
980 provide a good test for how well clouds have been detected in satellite data and the results may be
981 influenced by the occurrence of residual clouds.
982 Several recommendations resulted from the RR exercise. One of them was that, although the ADV
983 algorithm overall ranking was best, the coverage was a problem and needed to be improved. Such
984 improvement could be found from increasing the maximum solar zenith angle used in the retrieval

985 from 60 to 75 degrees which would give a similar coverage as other algorithms and from the
986 implementation of a module to model the reflectance of bright surfaces such as applied by SU. The
987 latter has been implemented in ORAC V2, together with the ADV post-processing step and initial
988 results are better than those described here. SU in turn has increased the number of pixels retrieved
989 by using a less strict quality control together with the ADV post-processing. The implementation of
990 these changes requires thorough testing and evaluation of the results to avoid loss of accuracy and
991 production of erroneous results. Hence such improvements will be reported in subsequent papers.
992 Thus a similar round robin exercise should be repeated with improved algorithms and the current
993 conclusions should be regarded as a snapshot evaluation of continuous algorithm development.

994 # 8. Conclusions

995 A validation protocol and necessary tools to implement the protocol have been developed and were
996 applied to 7 algorithms for aerosol retrieval using AATSR (4, one synergistic with SCIAMACHY),
997 MERIS (2) and PARASOL data. For reference, these tools were also applied to MODIS and MISR
998 data. The application of these tools, to L2 and to L3 data using different statistical methods and
999 scoring based on a combination of methods, revealed the strengths and weaknesses of each
1000 algorithm as well as a scoring of both the Aerosol-cci and the reference algorithms. A crucial issue
1001 is the dependence of validation scores on data filtering – this led to the development of a common
1002 point filter to assure the comparison of equivalent datasets.
1003 The results show that PARASOL has the highest accuracy over ocean and covers features well. The
1004 AATSR algorithm ranking depends critical on filtering. Overall (features, validation) ADV and SU
1005 seem better than ORAC which provides some unrealistic high features. SU and ADV scores are
1006 similar over land with SU providing data over bright surfaces and ADV having a better coverage of
1007 features. Over ocean ADV seems best (except coasts). A combination ADV generally + SU over
1008 bright surfaces + coasts could provide best products. MERIS ALAMO performs well over ocean;
1009 MERIS standard has large overestimations of AOD over land. SYNAER overestimates the AOD,

1010 has lower coverage and accuracy but a rather good coverage of features (except in central Asia and
1011 high latitudes).
1012 The scoring method shows that the AATSR algorithm results are close to or somewhat better than
1013 those from MODIS (and close or similar to MISR), but the number of points retrieved is much
1014 smaller than MODIS due to swath width and the availability of only one instrument as opposed to
1015 two MODIS instruments. Obviously, this gap cannot be closed. However, the dual view provided
1016 by AATSR makes this instrument potentially better suited for aerosol retrieval over land. Also, it
1017 provides one of the longer time series with the combination of ATSR-2 and AATSR (1995-April
1018 2012), with an extension by SLSTR (Sea and Land Surface Temperature Radiometer) planned to be
1019 launched in April 2014 as part of Sentinel-3. Sentinel-3 also has OLCI (Ocean Land Colour
1020 Instrument), which will extend the 10 years of MERIS observations.
1021 Taking into account the results from the RR exercise, the improved algorithms will be used to
1022 provide a 1-year data set (2008) of Aerosol ECV products which, after validation using similar tools
1023 as described in this paper, will be offered to the climate modelling community for their validation
1024 and feedback as regards the use for climate studies. Taking these into account, the full 17 years of
1025 ATSR-2 / AATSR is planned to be processed.
1026 A round robin exercise for aerosol ECVs cannot be conducted using a fully automatic scoring since
1027 trade-offs between coverage and accuracy or between added value and accuracy need to be made.
1028 This requires scientific expertise and a team dialogue to come up with conclusions which meet the
1029 standards of peer review by the scientific community. A strong user involvement in the whole
1030 validation and selection process is crucial to understand and take into account the user needs.
1031 The cooperation of the EO community with the global modelling community has proven to be very
1032 important, in particular as regards the production of a data set in such a way that they are indeed
1033 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.

Acknowledgements

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 sevral 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"*

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. *J. Geophys. Res*., 114, D09204, doi: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. *J. Mar. Res.*, 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, G. (Eds.), *Satellite Aerosol Remote Sensing Over Land* (pp. 135-159). Berlin: Springer-Praxis. ISBN 978-3-540-69396-3.

de Leeuw, G., & Kokhanovsky, A. (2009). Introduction. In A.A. Kokhanovsky & G. de Leeuw, G. (Eds.), *Satellite Aerosol Remote Sensing Over Land* (pp. 1-18). Berlin: Springer-Praxis. ISBN 978-3-540-69396-3.

de Leeuw, G., Kinne, S., Leon, J.F., Pelon, J., Rosenfeld, D., Schaap, M., Veefkind, P.J., Veihelmann, B., Winker, D.M., & von Hoyningen-Huene, W. (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-Verlag. ISBN: 978-3-642-14790-6 doi: 10.1007/978-3-642-14791-3.

Deuzé J.L., Bréon, F.M., Devaux, C., Goloub, P., Herman, M., Lafrance, B., Maignan, F., Marchand, A., Perry, G., & Tanré, D. (2001). Remote Sensing of aerosols over land surfaces from POLDER/ADEOS-1 polarized measurements. *J. Geophys. Res.*, 106, 4913-4926.

Dinter, T., von Hoyningen-Huene, W., Burrows, J.P., Kokhanovsky, A.A., Bierwirth, E., Wendisch, M., Müller, D., Kahn, R., & Diouri, M. (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., Tanré, D., & Slutsker, I.(2002). Variability of absorption and optical properties of key aerosol types observed in worldwide locations. *J. Atmos. Sci.*, 59, 590–608.

Dubuisson, P., Frouin, R., Dessailly, D., Duforêt, L., Léon, J.-F., Voss, K., & Antoine, D. (2009). Estimation of aerosol altitude from reflectance ratio measurements in the O2 A-band. *Rem. Sens. of Env.*, 113, 1899-1911.

Eck, T.F., Holben, B.N., Reid, J.S., Dubovik, O., Smirnov, A., O'Neill, N.T., Slutsker, I., & Kinne, S. (1999). Wavelength dependence of the optical depth of biomass burning, urban, and

1083 desert dust aerosols. *J. Geophys. Res.*, 104(D24), 31,333–31,349,
1084 doi:10.1029/1999JD900923.
1085 Flowerdew, R.J. & Haigh, J.D. (1995). An approximation to improve accuracy in the derivation of
1086 surface reflectances from multi-look satellite radiometers. *Geophys. Res. Lett.*, 22,
1087 1693–1696.
1088 GCOS (2011). Systematic observation requirements for satellite-based products for climate, 2011
1089 update. *GCOS Report* 154. WMO.
1090 Grey., W.M.F., North., P.R.J., Los, S.O., & Mitchell, R.M. (2006). Aerosol optical depth and land
1091 surface reflectance from multi-angle AATSR measurements: Global validation and
1092 inter-sensor comparisons. *IEEE Transactions on Geoscience and Remote Sensing,* 44(8),
1093 2184 – 2197.
1094 Grey., W.M.F, North., P.R.J., and Los, S. (2006b). Computationally efficient method for retrieving
1095 aerosol optical depth from ATSR-2 and AATSR data, *Appl. Optics*, 45(12), 2786-2795.
1096 Gurney, K. R., Law, R.M., Denning, A.S., Rayner, P.J., Baker, D., Bousquet, P., Bruhwiler, L.,
1097 Chen, Y.-H., Ciais, P., Fan, S., Fung, I.Y., Gloor, M., Heimann, M., Higuchi, K.,
1098 John,J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B.C.,
1099 Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T. & Yuen, C.-W. (2002), Towards
1100 robust regional estimates of CO2 sources and sinks using atmospheric transport models,
1101 Nature, 415, 626–630, doi:10.1038/415626a.
1102 Herman M., Deuzé, J.-L., Marchand, A., Roger, B., & Lallart, P. (2005) . Aerosol remote sensing
1103 from POLDER/ADEOS over the ocean: Improved retrieval using a nonspherical particle
1104 model, *J. Geophys. Res.*, 110, D10S2, doi:10.1029/2004JD004798.
1105 Hoff, R.M., & Christopher, S.A. (2009). Remote sensing of particulate pollution from space: have
1106 we reached the promised land? *J. Air Waste Manag. Assoc.*, 59, 645–675.

1107 Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A.,
1108 Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., & Smirnov, A. (1998).
1109 AERONET – A federated instrument network and data archive for aerosol
1110 characterization. *Rem. Sens. Environ.*, 66, 1–16.
1111 Hollmann, R., Merchant, C., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E.,
1112 Defourny, P., de Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S.,
1113 Sathyendranath, S., van Roozendael, M., & Wagner, W. (2012). The ESA climate
1114 change initiative: satellite data records for essential climate variables. *Submitted for*
1115 *publication in BAMS.*
1116 Holzer-Popp, T., Schroedter, M., & Gesell, G. (2002a). Retrieving aerosol optical depth and type in
1117 the boundary layer over land and ocean from simultaneous GOME spectrometer and
1118 ATSR-2 radiometer measurements, 1, Method description. *J. Geophys. Res.*, 107, 4578,
1119 doi:10.1029/2001JD002013.
1120 Holzer-Popp, T., Schroedter, M., & Gesell, G. (2002a). Retrieving aerosol optical depth and type in
1121 the boundary layer over land and ocean from simultaneous GOME spectrometer and
1122 ATSR-2 radiometer measurements, 2, Case study application and validation, *J.*
1123 *Geophys. Res.*, 107, 4770, doi:10.1029/2002JD002777.
1124 Holzer-Popp T., Schroedter-Homscheidt, M., Breitkreuz, H., Klüser, L., & Martynenko, D. (2008).
1125 Improvements of synergetic aerosol retrieval for ENVISAT, *Atmos. Chem. Phys.*, 8,
1126 7651-7672.
1127 Holzer-Popp. T., de Leeuw, G., Martynenko, D., Klüser, L., Bevan, S., Davies, W., Ducos, F.,
1128 Deuzé, J. L., Graigner, R.G., Heckel, A., von Hoyningen Huene, W., Kolmonen, P.,
1129 Litvinov, P., North, P., Poulsen, C.A., Ramon, D., Siddans, R., Sogacheva, L., Tanre,
1130 D., Thomas, G.E., Breon, F.-M., Descloitres, J., Griesfeller, J., Kinne, S. & Schulz, M.

(2013). Extensive aerosol retrieval algorithm evaluation within ESA aerosol_cci project. *In preparation for submission to AMT*.

Hsu, N.C., Tsay, S.C., King, M.D., & Herman, J.R. (2004). Aerosol properties over bright-reflecting source regions. *IEEE Trans. Geosci. and Rem. 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, *Atmos. Chem. Phys.*, 12, 4585-4606, doi:10.5194/acp-12-4585-2012.

Jin, Y., Schaaf, C.B., Woodcock, C.E., Gao, F., Li, X., Strahler, A.H., Lucht, W., & Liang, S. (2003). Consistency of MODIS surface BRDF/Albedo retrievals: 1. Algorithm performance. *J. Geophys. Res.*, 108, D54158, doi:10.1029/2002JD002803.

Kaufman, Y.J., Justice, C.O., Flynn, L.P., Kendall, J.D., Prins, E.M., Giglio, L., Ward, D.E., Menzel, W.P., & Setzer, A.W. (1998). Potential global fire monitoring from EOS-MODIS, *J. Geophys. Res.*, 103, 32215–21238.

Kahn, R.A., Nelson, D.L., Garay, M.J., Levy, R.C., Bull, M.A., Diner, D.J., Martonchik, J.V., Paradise, S.R., Hansen, E.G.. & Remer, L.A. (2009). MISR Aerosol Product Attributes and Statistical Comparisons With MODIS. *IEEE Trans. On Geosc. and Remote Sensing*, 47, 4095–4114.

Kahn, R.A., Gaitley, B.J., Garay, M.J., Diner, D.J., Eck, T., Smirnov, A., & Holben, B.N. (2010). Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network. *J. Geophys. Res*. 115, D23209, doi: 10.1029/2010JD014601.

Kinne, S., Schulz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S.E., Berntsen, T., Berglen, T.F., Boucher, O., Chin, M., Collins, W., Dentener, F., Diehl, T., Easter, R., Feichter, J., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Herzog, M., Horowitz, L., Isaksen, I., Iversen, T., Kirkevåg, A., Kloster, S., Koch, D., Kristjansson,

1156 J.E., Krol, M., Lauer, A., Lamarque, J.F., Lesins, G., Liu, X., Lohmann, U., Montanaro,
1157 V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, Ø., Stier, P., Takemura, T., &
1158 Tie, X. (2006). An AeroCom initial assessment optical properties in aerosol component
1159 modules of global models. *Atmos. Chem. Phys.*, 6, 1815-1834.
1160 Koepke, P. (1984). Effective reflectance of oceanic whitecaps. *Appl. Opt.*, 23, 1816-1823.
1161 Kokhanovsky, A.A., & de Leeuw, G. (2009). *Satellite Aerosol Remote Sensing Over Land.* Berlin:
1162 Springer-Praxis. ISBN 978-3-540-69396-3, 388 pp.
1163 Kolmonen, P., Sundström, A.-M., Sogacheva, L., Rodriguez, E., Virtanen, T.H., & de Leeuw, G.
1164 (2013). The uncertainty characterization of AOD for the AATSR ADV/ASV retrieval
1165 algorithm – Towards the assimilation of the satellite retrieved aerosol properties.
1166 *Submitted for publication in AMT.*
1167 Kotchenova, S.Y., Vermote, E.F., Matarrese, R., & Klemm, F.J., Jr. (2006). Validation of a vector
1168 version of the 6S radiative transfer code for atmospheric correction of satellite data. Part
1169 I: Path radiance. *Appl. Opt.* 45(26), 6762-6774.
1170 Kotchenova, S.Y. & Vermote, E.F. (2007). Validation of a vector version of the 6S radiative
1171 transfer code for atmospheric correction of satellite data. Part II. Homogeneous
1172 Lambertian and anisotropic surfaces. *Appl. Opt.*, 46 (20), 4455-4464.
1173 Labonne, M., Bréon, F.-M., & Chevallier, F. (2007). Injection height of biomass burning aerosols
1174 as seen from a spaceborne lidar. *Geophys. Res. Lett.*, 34, 1–5,
1175 doi:10.1029/2007GL029311.
1176 Lee, K.H., Li, Z., Kim, Y.J., & Kokhanovsky, A. (2009). Atmospheric aerosol monitoring from
1177 satellite observations: a history of three decades. In Y.J. Kim, U. Platt, M.B. Gu & H.
1178 Iwahashi (Eds.), *Atmospheric and Biological Environmental Monitoring* (pp. 13–38).
1179 Springer Science+Business Media B.V.

1180 Lenoble, J., Herman, M., Deuzé, J.L., Lafrance, B., Santer, R., & Tanré, D. (2007). A Successive
1181 Order of Scattering Code for Solving the Vector Equation of Transfer in the Earth's
1182 Atmosphere with Aerosols. *JQSRT*, 107, 479-507.
1183 Levy, R.C., Remer, L.A., Mattoo, S., Vermote, E., and Kaufman, Y.J. (2007). Second-generation
1184 operational algorithm: Retrieval of aerosol properties over land from inversion of
1185 Moderate Resolution Imaging Spectroradiometer spectral reflectance. *J. Geophys. Res.*,
1186 112(D13), D13211, 10.1029/2006JD007811.
1187 Levy, R.C., Remer, L.A., Kleidman, R.G., Mattoo, S., Ichoku, C., Kahn, R., & Eck, T.F. (2010).
1188 Global evaluation of the Collection 5 MODIS dark-target aerosol products over land,
1189 *Atmos. Chem. Phys.*, 10, 10399-10420, doi:10.5194/acp-10-10399-2010.
1190 Maignan, F., Bréon, F.-M., & Lacaze, R. (2004). Bidirectional reflectance of Earth targets:
1191 Evaluation of analytical models using a large set of spaceborne measurements with
1192 emphasis on the Hot Spot. *Rem. Sens. Env.*, 90, 210–220.
1193 Mie, G. (1908). Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.*,
1194 25, 377–445.
1195 Müller, D., Krasemann, H., Brewin, R.J.W., Brockmann, C., Deschamps, P.-Y., Doerffer, R.,
1196 Fomferra, N., Franz, B.A., Grant, M.G., Groom, S.B., Mélin, F., Platt, T., Regner, P.,
1197 Sathyendranath, S., Steinmetz, F., & Swinton, J. (2013). The Ocean Colour Climate
1198 Change Initiative: Spatial and Seasonal Homogeneity of Atmospheric Correction
1199 Algorithms. *Submitted for publiscation in RSE (this issue)*.
1200 Nadal, F., & Bréon, F.M. (1999). Parameterization of surface polarized reflectance derived from
1201 POLDER spaceborne measurements. *IEEE Trans. Geosci. Remote Sens.*, 37 , 1709-
1202 1718.

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 Trans. Geosci. Remote Sens.*, 37(1), 526-537.

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

Poulsen, C.A., Watts, P.D., Thomas, G.E., Sayer, A.M., Siddans, R., Grainger, R.G., Lawrence, B.N., Campmany, E., Dean, S.M., & Arnold, C. (2012). Cloud retrievals from satellite data using optimal estimation: evaluation and application to ATSR. *Atmos. Meas. Tech.* doi:10.5194/amt-5-1889-2012.

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

Remer, L.A., Kaufman, Y.J., Tanré, D., Mattoo, S., Chu, D.A., Martins, J.V., Li, R.R., Ichoku, C., Levy, R.C., Kleidman, R.G., Eck, T.F., Vermote, E., & Holben, B.N. (2005). The MODIS Aerosol Algorithm, Products, and Validation. *J. Atmos. Sci.*, 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. *Int. J. Rem. Sens.*, 20, 1819-1840.

Santer, R., Ramon, D., Vidot, J., & Dilligeard, E. (2007). A Surface Reflectance Model for Aerosol Remote Sensing over Land. *Int. J. Rem. Sens.,* 28, 737-760.

Sayer, A.M., Thomas, G.E., & Grainger, R.G. (2008). A sea surface reflectance model suitable for use in (A)ATSR aerosol retrieval algorithms. Atmos. Meas. Tech., 3, 813–838, doi:10.5194/amt-3-813-2010.

1227 Sayer, A.M., Thomas, G.E., Palmer, P.I., & Grainger, R.G. (2010). Some implications of sampling
1228 choices on comparisons between satellite and model aerosol optical depth fields. *Atmos.*
1229 *Chem. Phys.*, 10, 10705-10716, doi:10.5194/acp-10-10705-2010.
1230 Sayer, A.M., Poulsen, C.A., Arnold, C., Campmany, E., Dean, S., Ewen, G.B.L., Grainger, R.G.,
1231 Lawrence, B.N., Siddans, R., Thomas, G.E., & Watts, P.D. (2011). Global retrieval of
1232 ATSR cloud parameters and evaluation (GRAPE): dataset assessment. *Atmos. Chem.*
1233 *Phys.*, 11, 3913-3936, doi:10.5194/acp-11-3913-2011.
1234 Seinfeld, J.H., & Pandis, S.N. (1998). *Atmospheric chemistry and physics: from air pollution to*
1235 *climate change*. New York: Wiley&Sons, ISBN 0-471-17815-2.
1236 Sinyuk, A., Torres, O., & Dubovik, O. (2003). Combined use of satellite and surface observations
1237 to infer the imaginary part of refractive index of Saharan dust. *Geophys. Res. Lett.* , 30 ,
1238 1081, doi:10.1029/2002GL016189.
1239 Smirnov, A., Sayer, A.M., Holben, B.N., Hsu, N.C., Sakerin, S.M., Macke, A., Nelson, N.B.,
1240 Courcoux, Y., Smyth, T.J., Croot, P., Quinn, P.K., Sciare, J., Gulev, S.K., Piketh, S.,
1241 Losno, R., Kinne, S., & Radionov, V.F. (2012). Effect of wind speed on aerosol optical
1242 depth over remote oceans, based on data from the Maritime Aerosol Network. *Atmos.*
1243 *Meas. Tech.*, 5, 377-388, doi:10.5194/amt-5-377-2012.
1244 Sofiev, M., Vankevich, R., Lotjonen, M., Prank, M., Petukhov, V., Ermakova, T., Koskinen, J., &
1245 Kukkonen, J. (2009). An operational system for the assimilation of the satellite
1246 information on wild-land fires for the needs of air quality modelling and forecasting,
1247 *Atmos. Chem. Phys.*, 9, 6833–6847, doi:10.5194/acp-9-6833-2009.
1248 Stamnes, K., Tsay, S.-C., Wiscombe W., & Jayweera, K. (1988). Numerically stable algorithm for
1249 discrete-ordinate-method radiative transfer in multiple scattering and emitting layered
1250 media. *Appl. Optics*, 24, 2502–2509.

1251 Tanré, D., Kaufman, Y.J., Herman, M., & Mattoo, S. (1997). Remote sensing of aerosol properties
1252 over oceans using the MODIS/EOS spectral radiances. *J. Geophys. Res.*, 102, 16 971–16
1253 988.
1254 Thomas, G.E., Poulsen, C.A., Sayer, A.M., Marsh, S.H., Dean, S.M., Carboni, E., Siddans, R.,
1255 Grainger, R.G., & Lawrence, B.N. (2009). The GRAPE aerosol retrieval algorithm,
1256 *Atmos. Meas. Tech.*, 2, 679-701, doi:10.5194/amt-2-679-2009.
1257 Thomas, G.E., Chalmers, N., Harris, B., Grainger, R.G., & Highwood, E.J. (2012). Regional and
1258 monthly and clear-sky aerosol direct radiative effect (and forcing) derived from the
1259 GlobAEROSOL-AATSR satellite aerosol product. *Atmos. Chem. Phys. Discuss.*, 12,
1260 18459-18497, doi:10.5194/acpd-12-18459-2012.
1261 van Donkelaar, A., Martin, R.V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P.
1262 (2010). Global Estimates of Ambient Fine Particulate Matter Concentrations from
1263 Satellite-Based Aerosol Optical Depth: Development and Application. *Environmental*
1264 *Health Perspectives*, 118, 847-855.
1265 Veefkind, J.P., de Leeuw, G., & Durkee, P.A. (1998). Retrieval of aerosol optical depth over land
1266 using two-angle view satellite radiometry during TARFOX. *Geophys. Res. Letters*,
1267 25(16), 3135-3138.
1268 Veefkind, J.P. & de Leeuw, G. (1998). A new algorithm to determine the spectral aerosol optical
1269 depth from satellite radiometer measurements. *Journal of Aerosol Sciences,* 29, 1237-
1270 1248.
1271 Volten, H., Munoz, O., Rol, E., de Haan, J.F., Vassen, W., & Hovenier, J.W. (2001). Scattering
1272 matrices of mineral aerosol particles at 441.6 nm and 632.8 μm. *J. Geophys. Res.*, 106,
1273 17,375-17,401.

1274 von Hoyningen-Huene, W., Freitag, M., & Burrows, J.P. (2003). Retrieval of Aerosol Optical
1275 Thickness over Land Surfaces from Top-of-Atmosphere Radiances, *J. Geophys. Res.*,
1276 108, 4260, doi:10.1029/2001JD002018.
1277 von Hoyningen-Huene, W., Yoon, J., Vountas, M., Istomina, L., Rohen, G., Dinter, T.,
1278 Kokhanovsky, A.A., & Burrows, J.P. (2011). Retrieval of spectral Aerosol Optical
1279 Thickness over Land using Ocean Colour Sensors MERIS and SeaWiFS, Atmos. Meas.
1280 Tech. Discuss., 3, 2107-2164, doi:10.5194/amtd-3-2107-2010.
1281 WMO/GAW (2003). *Aerosol Measurement Procedures Guidelines and Recommendations*.
1282 GAWReport No. 153, World Meteorological Organization Global Atmosphere Watch,
1283 Geneva, Switzerland.
1284 Yu, H., Kaufman, Y.J., Chin, M., Feingold, G., Remer, L.A., Anderson, T.L., Balkanski, Y.,
1285 Bellouin, N., Boucher, O., Christopher, S., DeCola, P., Kahn, R., Koch, D., Loeb, N.,
1286 Reddy, M.S., Schulz, M., Takemura, T., & Zhou, M. (2006). A review of measurement-
1287 based assessments of the aerosol direct radiative effect and forcing. *Atmos. Chem. Phys.*,
1288 6, 613-666, doi:10.5194/acp-6-613-2006.
1289 Zelazowski P, Sayer, A.M., Thomas, G.E., & Grainger, R.G. (2011). Reconciling satellite-derived
1290 atmospheric properties with fine-resolution land imagery: Insights for atmospheric
1291 correction, *J. Geophys. Res.*, 116, D18308, doi:10.1029/2010JD015488.
1292 Zieger, P., Weingartner, E., Henzing, J., Moerman, M., de Leeuw, G., Mikkilä, J., Ehn, M.,
1293 Petäjä, T., Clémer, K., van Roozendael, M., Yilmaz, S., Frieß, U., Irie, H., Wagner, T.,
1294 Shaiganfar, R., Beirle, S., Apituley, A., Wilson, K., and Baltensperger, U. (2011).
1295 Comparison of ambient aerosol extinction coefficients obtained from in-situ, MAX-
1296 DOAS and LIDAR measurements at Cabauw, Atmos. Chem. Phys.. 11, 2603-2624,
1297 doi:10.5194/acp-11-2603-2011.
1298

Tables

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 section 3.

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	HYGEOS	+	+	3	1	-	-	-	-	-	-	-	+
	BAER	Univ Bremen	+	+	1	0	+	-	-	-	+	-	-	+
	ALAMO	HYGEOS	-	+	2	1	+	-	-	-	-	+	-	-
POLDER	PARASOL	LOA	-	+	3	2	+	-	+	-	+	-	-	-

AOD(n), n= nr of wavelengths

Type: number of independent aerosol components which potentially can be retrieved

Table 2. 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. (55μm)	refract index imag p. (.55μm)	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	2-4km
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

1312 Table 3. Rankings of the Aerosol-cci algorithms: summary of the results from the three independent

1313 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

1314

1315

1316 Table 4. L2 validation statistics.

			AATSR				MERIS		PARASOL
			ADV	ORAC	SU	SYNAER	ESA Standard	ALAMO	
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		
		nr 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*	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
		nr of pixels	221	285	99	61	262	103	384

1317 AE(555-865)

1318 *For ADV, AE(555-1610) yields similar statistics (cc=0.66, st dev= 0.37, bias=0.05), but the
1319 average AE is lower.

1320

Table 5. Comparison of scores (1-error) for different AOD satellite retrievals for year 2008 data of 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 global scores are difficult to compare as the number of contributing regions differs, thus there is a focus for North America, where most data provide a score. Note that the total number of areas for which a score would be possible is 25 (Figure 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, ALAMO and ESA Standard because the number of samples was too small.

	10+ samples	global scores					North American scores				
		ocean & land	ocean	Land	nr. of areas		total	Bias	Temporal	Spatial	nr of pairs
reference	MISR v22	.62	.66	.59	3		.54	.87	.84	.74	25
	MODIS aqua	.55	.60	.50	8		.42	.86	.79	.62	93
	MODIS terra	.61	.63	.58	10		.41	.86	.79	.61	101
	SEAWIFS	.56	.58	.55	6		.47	.83	.80	.71	50
	OMI	.46	.48	.40	9						
AATSR	ADV v13	-.57	-.60	-.55	2		.54	.84	.79	.82	28
	SU v30	-.46	-.48	-.45	1		.48	.77	.75	.83	15
	ORAC v11	.39	.40	-.39	2		.39	.86	.64	.70	37
	Parasol v23	-.14	-.13	-.19	3		.25	.80	.56	.55	21

1335 Table 6. Ranking by level3 AEROCOM analysis for land and coastal areas. For ocean the number

1336 of collocations is too small to provide meaningful scores.

	best	**2ndbest**	**third**	**fourth**
Based on correlation				
coast				
SU	3	0	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	0	3	1	0
ORAC	0	1	2	1
SYNAER	1	0	0	3
land				
ADV	3	1	0	0
SU	2	1	1	0
ORAC	1	1	2	0
SYNAER	0	0	0	4

1337

1338

Figures

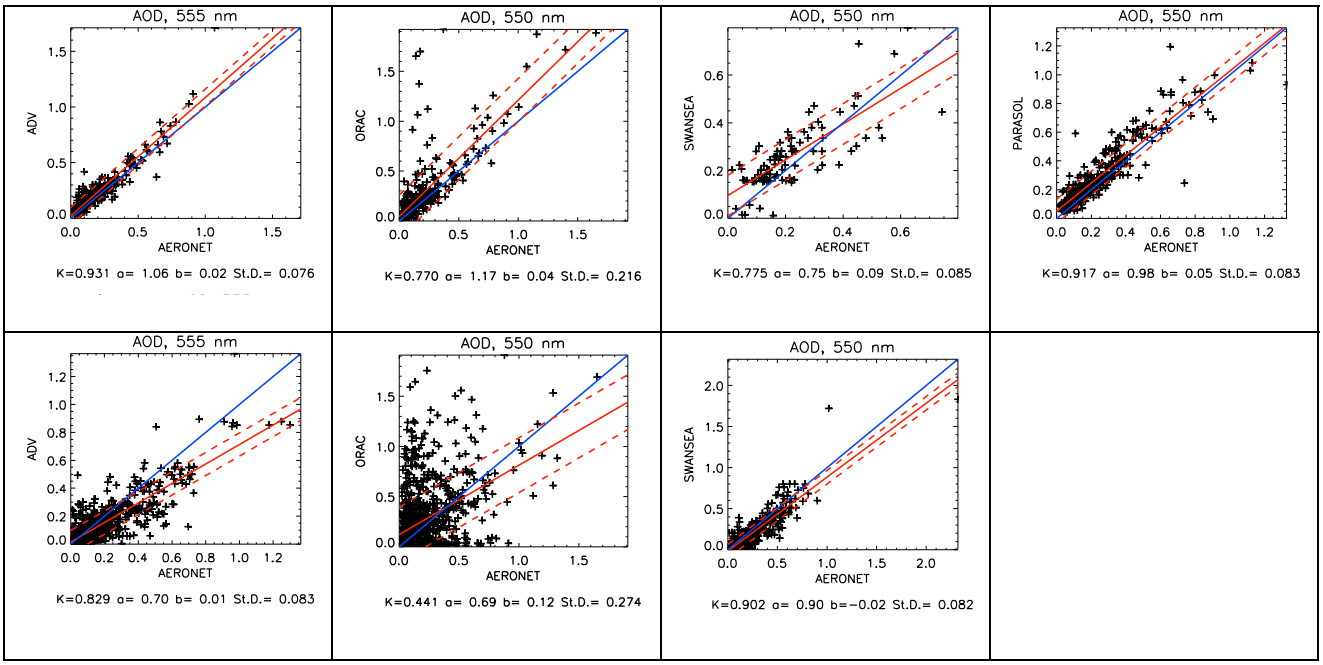

Figure 1. Examples of scatterplots between the satellite-retrieved AOD and AE vs. AERONET data. AOD over ocean (top row) and land (bottom row) were separated using the ORAC land/sea flag for all retrievals. The algorithm is indicated along the vertical axis. Statistics from a least squares fit of the type y=ax+b are indicated in the legends at the bottom (see also Table 4): K is the correlation coefficient, a is the slope, b is the bias and St.D. is the standard deviation.

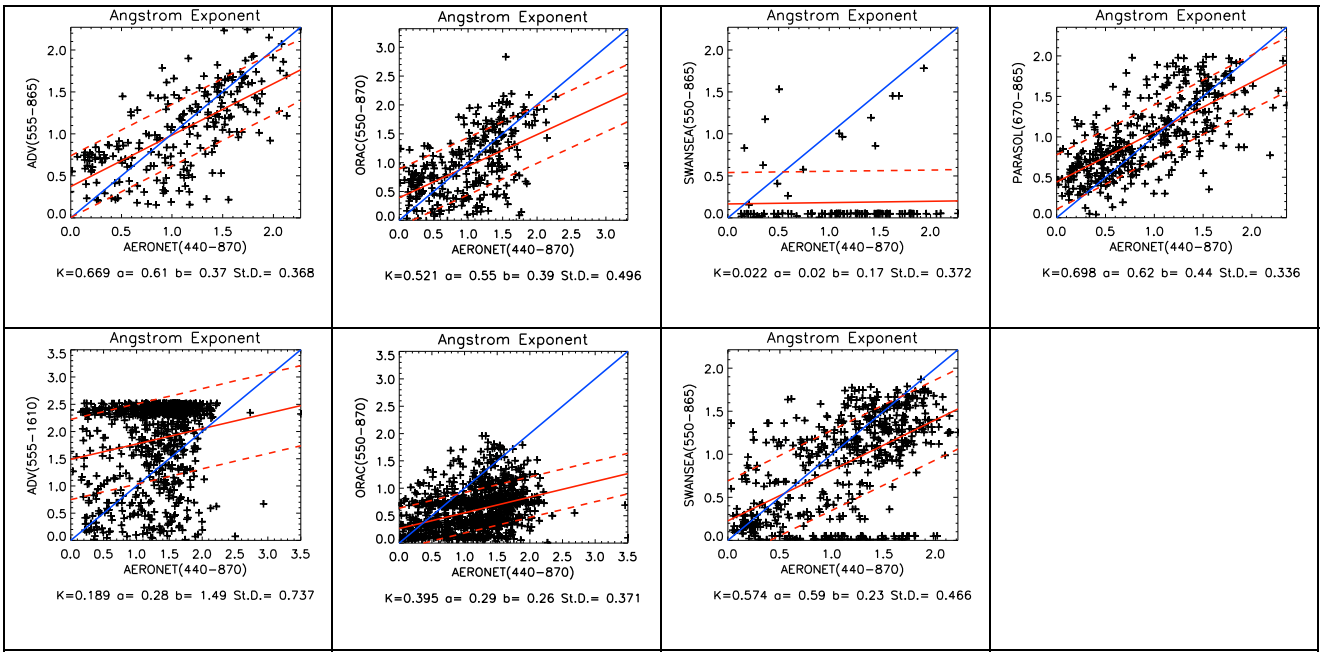

Figure 2. Examples of scatterplots for Ångström exponents vs. AERONET data, for further explanation see Figure 1.

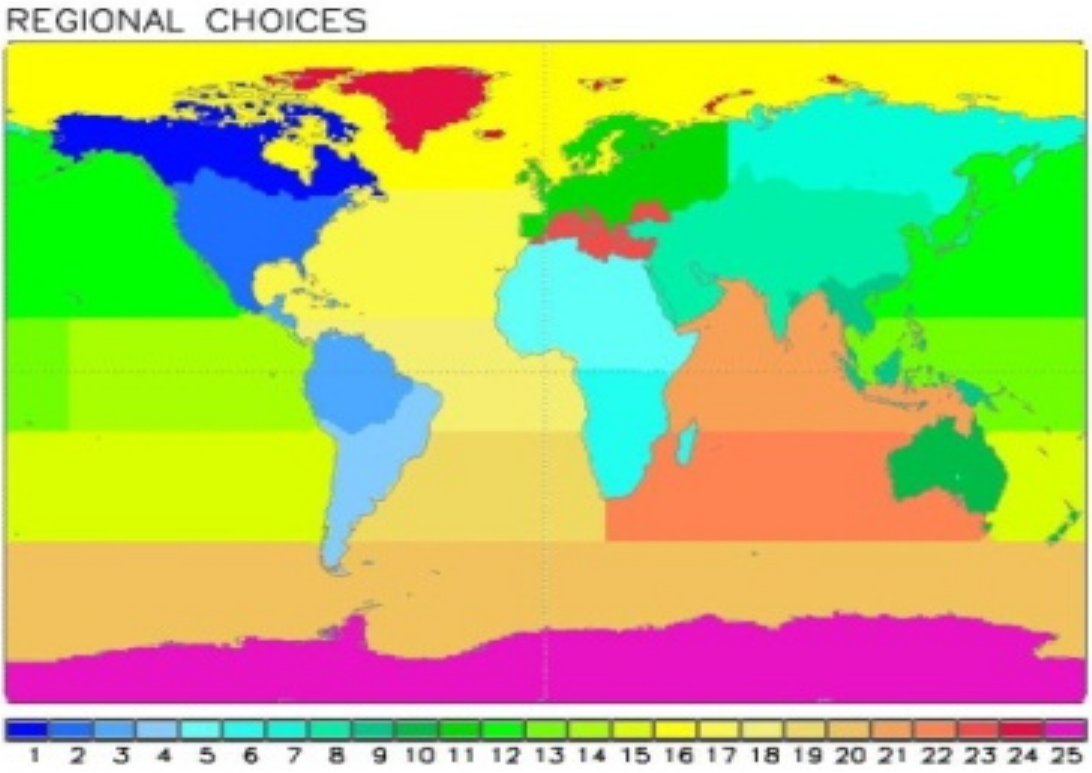

Figure 3. Regional stratification of the globe following TransCom (Gurney, et al. 2002).

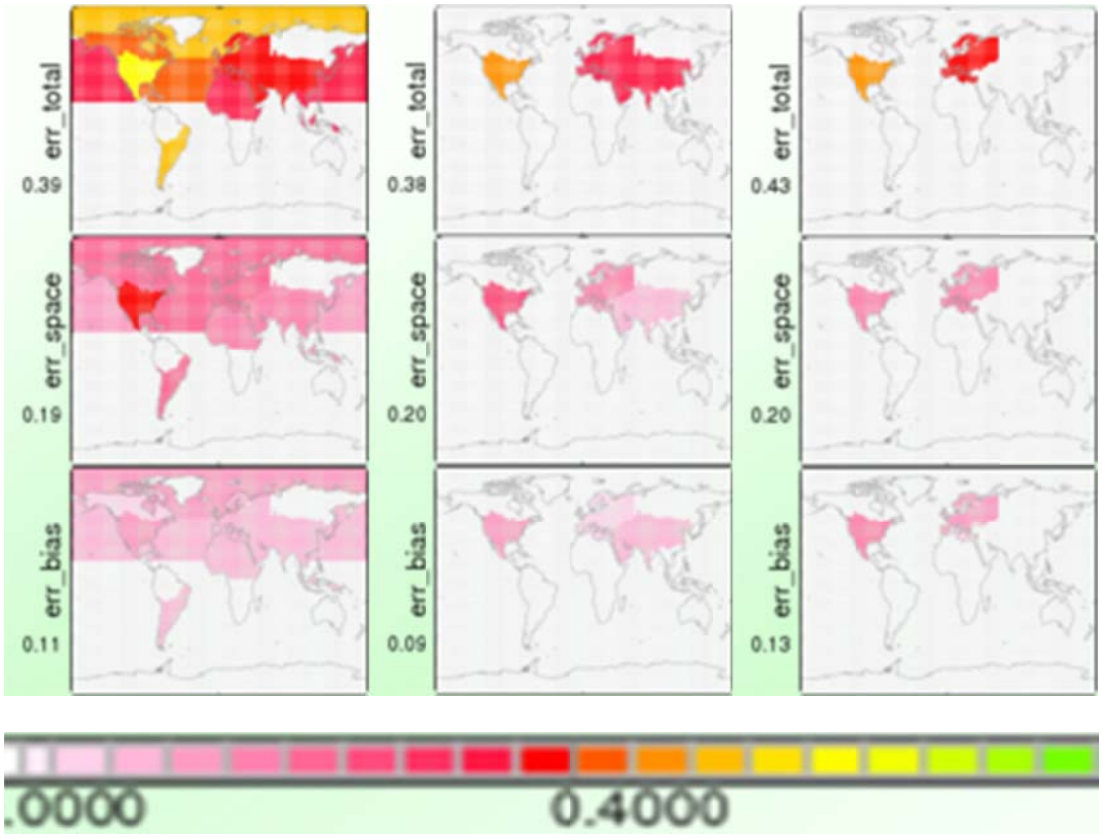

Figure 4. Regional error comparisons between MODIS (left), MISR(center) and ADV (right). Errors are only displayed for regions with sufficient data-pairs to co-incident AERONET sun-photometer samples. Aside from total regional average errors (top row), also contributing regional sub-errors for spatial variability and bias are presented.

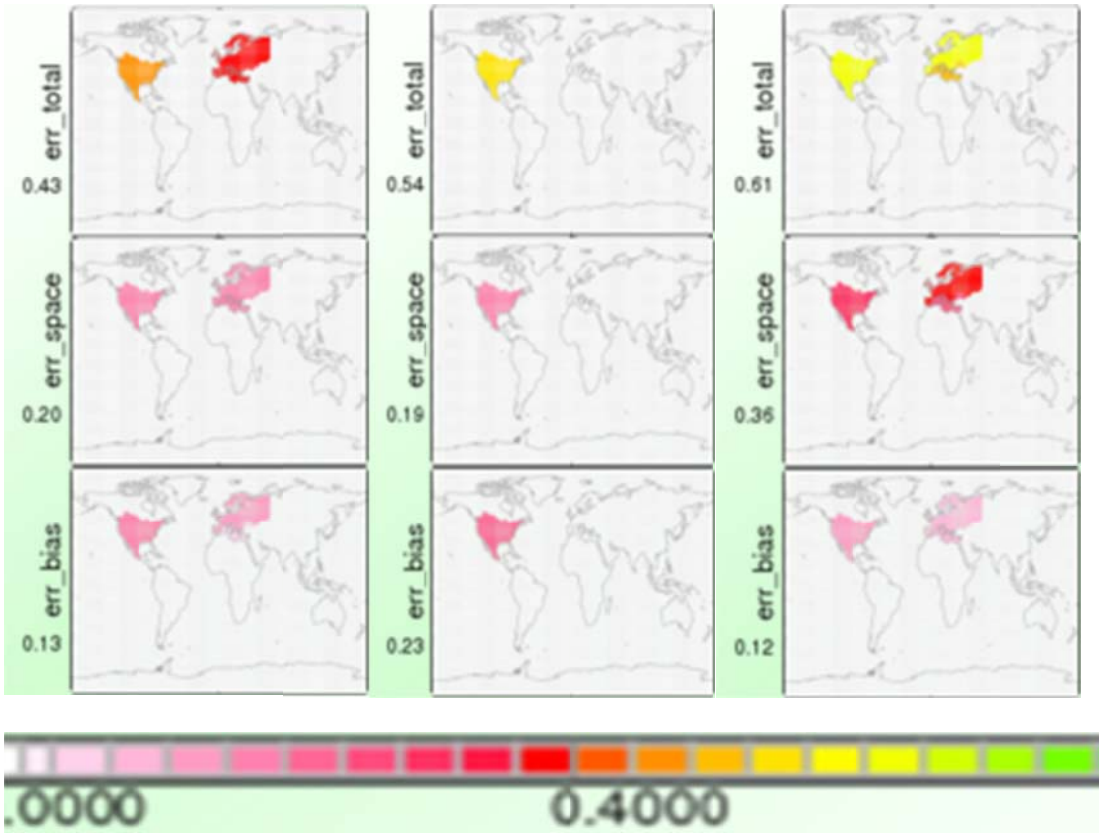

Figure 5. Regional error comparisons between AATSR algorithms ADV (left), SU (center) and ORAC (right). Errors are only displayed for regions with sufficient data-pairs to co-incident AERONET sun-photometer samples. Aside from total regional average errors (top row), also contributing regional sub-errors for spatial variability and bias are presented.

1369

AATSR_ADV.v1.3-Set3D 2008 Obs: AERONETSun 2008	
only Stations WORLDm	
# of valid observations:	296
OBS mean	0.217
MODEL mean	0.178
Spearman Rank Correlation	0.730
Pearson Correlation Coefficient	0.820
Spatial yearly mean Corr Coeff	0.784
Seasonal Anomaly Corr Coeff	0.679
RMS error	0.102
Slope fit forced through zero	1.164
Regression coefficient, Slope	1.060
Regression Constant, Offset:	0.028
STDDEV(Model)/STDDEV(Data):	0.773
Score (mean relative bias)	32%
Taylor Score	0.897

1370

1371 Figure 6. Satistics from the L3 AEROCOM validation for the World Annual Average (4 months in 2008)

1372 AATSR ADV.v1.3-Set3D AOD data.

1373

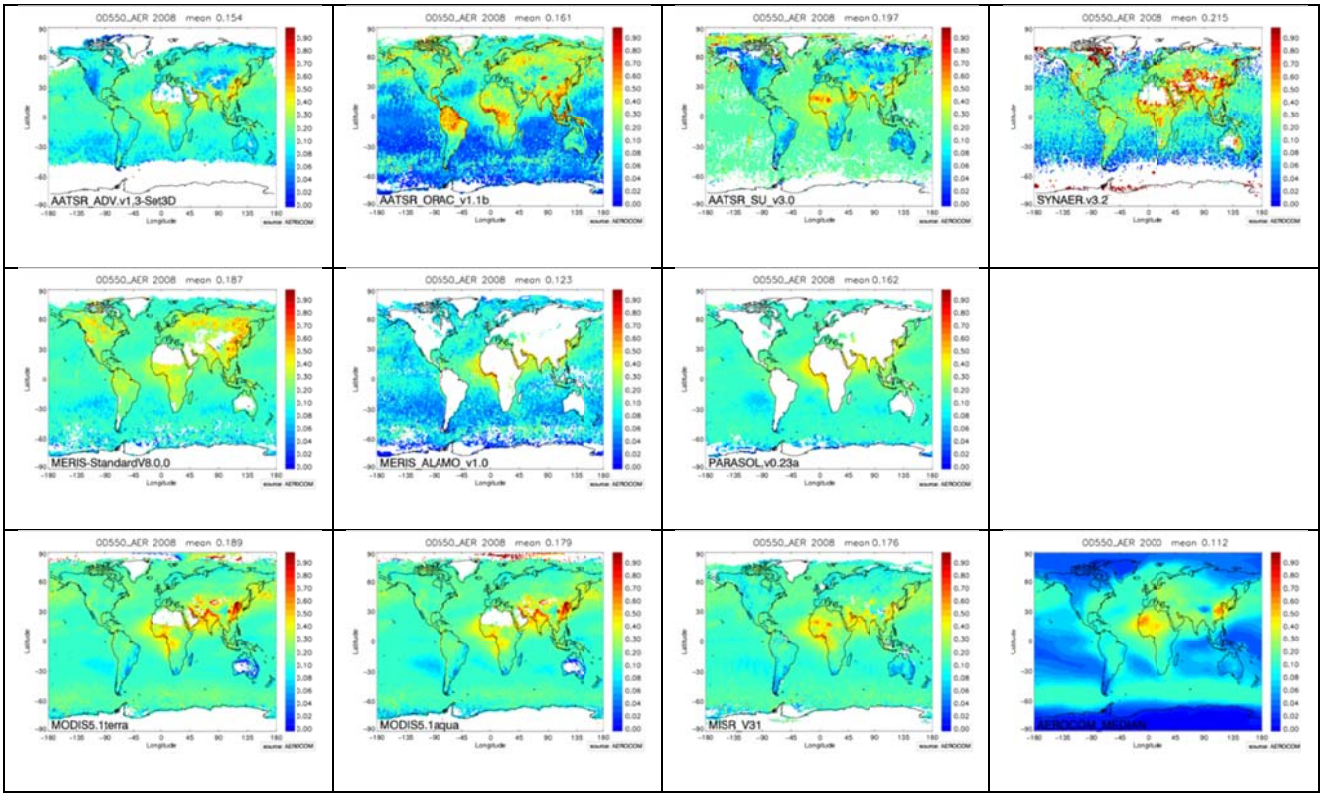

Figure 7. Global annual mean AOD maps for all algorithms participating in Aerosol-cci as well as reference maps. Top: AATSR ADV, ORAC, SU, SYNAER; Middle: MERIS Standard, ALAMO, PARASOL; Bottom, MODIS Terra, MODIS Aqua; MISR, AEROCOM median. Note that ALAMO and PARASOL are only used over ocean.

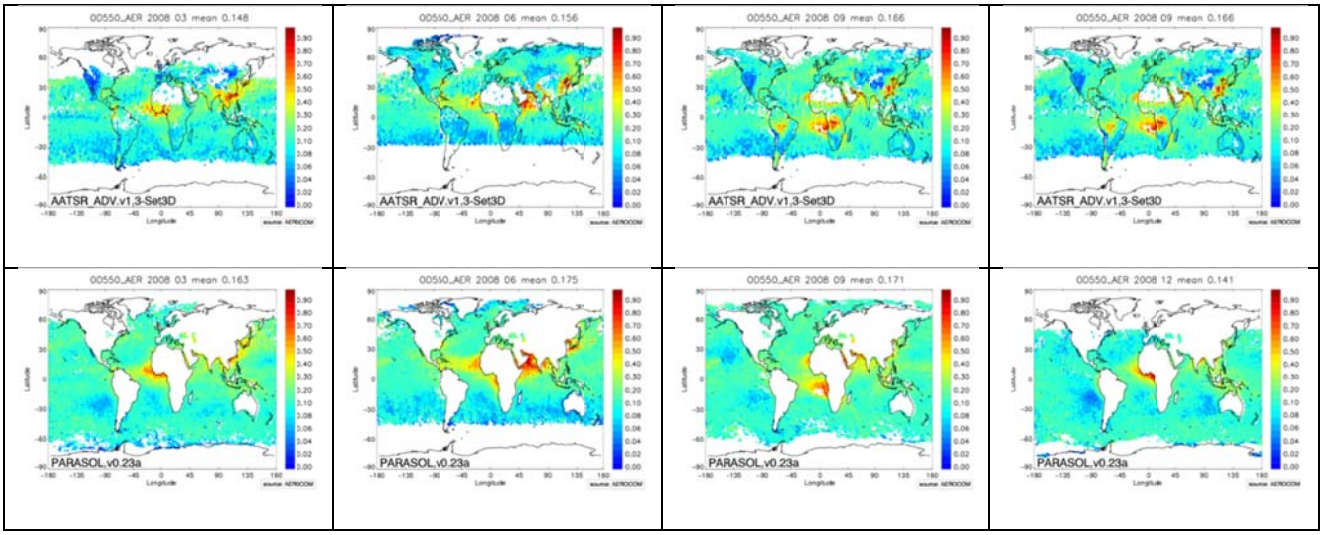

Figure 8. Global AOD maps of the AOD retrieved using ADV for AATSR data (top row) and for PARASOL data (bottom row) for the months March, June, September and December (left to right), selected in this RR exercise to cover one month in each season.

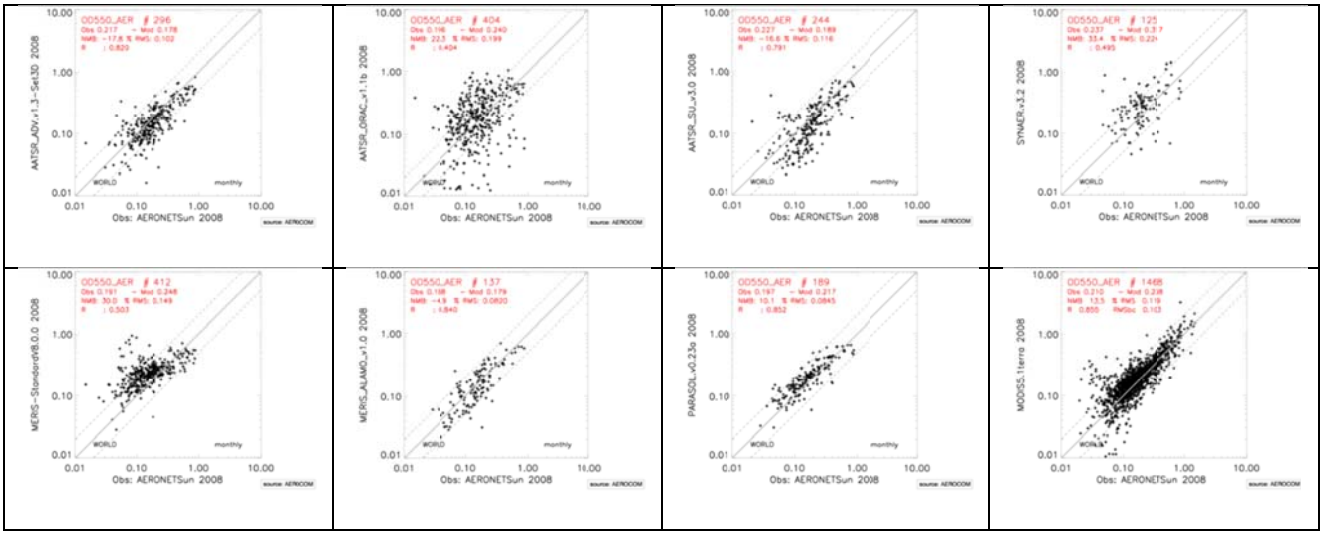

Figure 9. Scatterplots of the AOD provided by each Aerosol-cci algorithm and by MODIS Terra, identified on the vertical axis of each figure plotted on a log-log scale versus AERONET AOD. The lines indicate the 1:1 and the factor 2 limits. Results from a least squares fit are provided in the legend at the upper left corner of each figure.

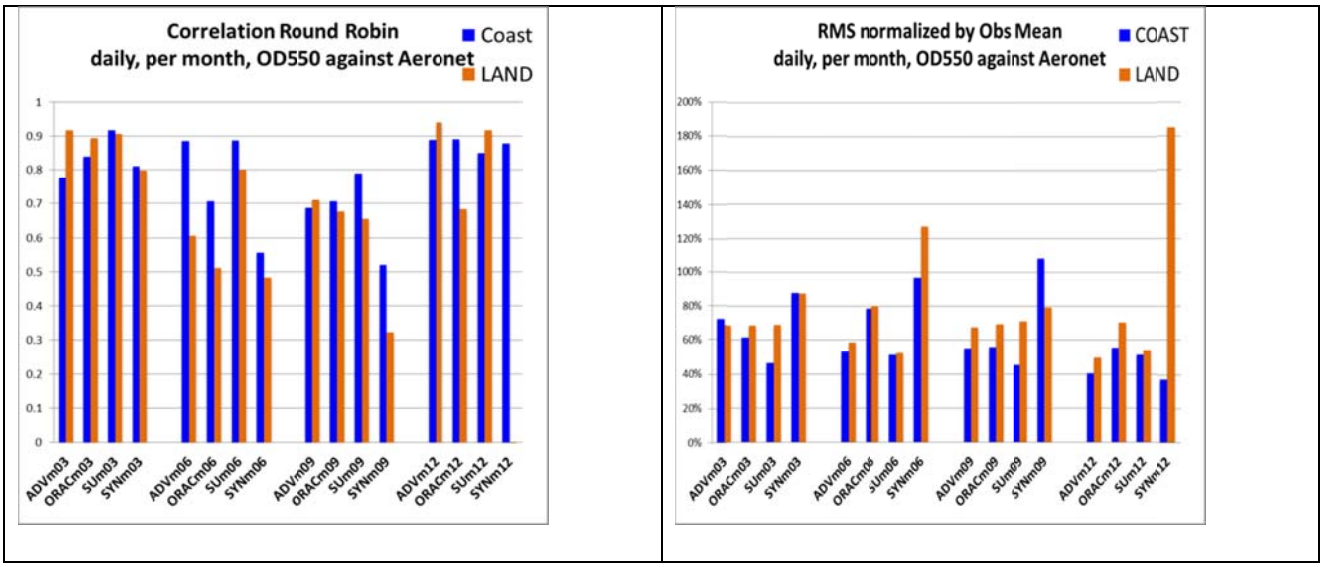

Figure 10. Bar charts showing the seasonal variation (as given by 1 month in each season) of the correlation coefficient and RMS over land and in coastal regions for the 4 algorithms using AATSR data. Over ocean there are too few collocations to provide meaningful statistics for each month.

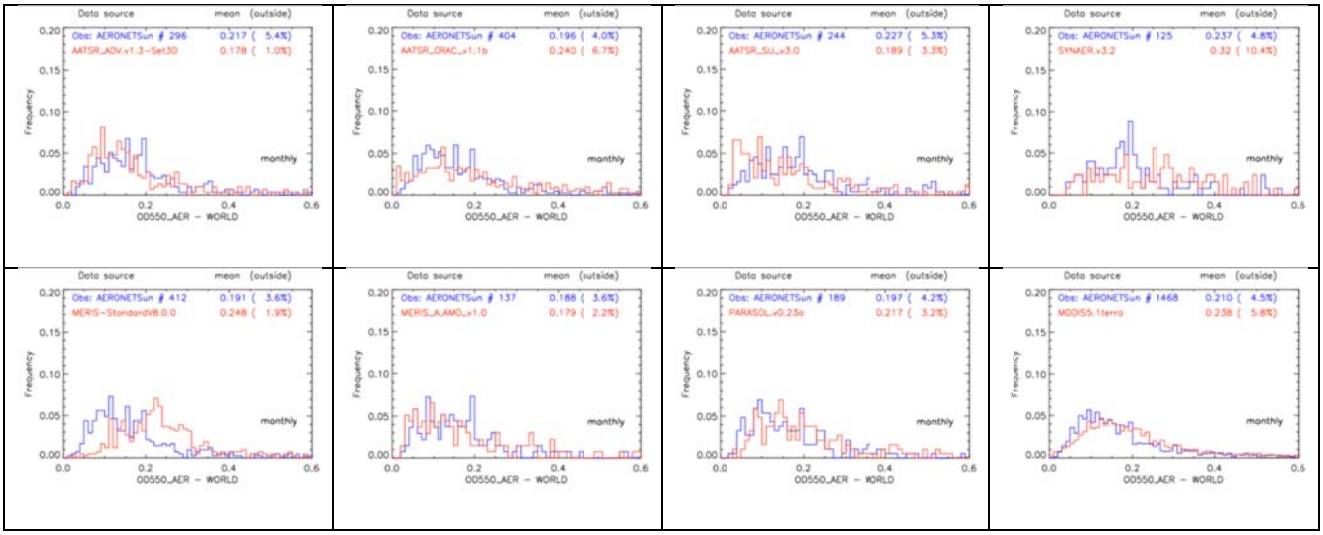

Figure 11. Histograms showing the frequency of occurrence of the AOD values provided by each Aerosol-cci RR algorithm and MODIS Terra (in red), together with that of the AERONET reference values (in blue), together with the global mean value and the percentage of points with an AOD of greater than 0.6 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.

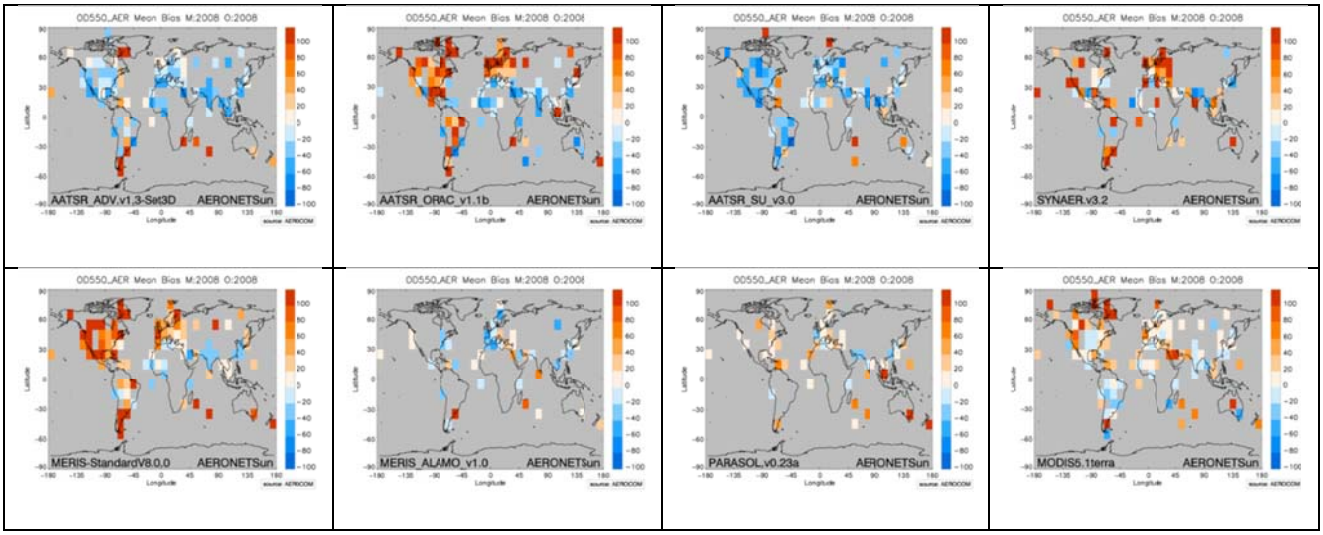

Figure 12. Evaluation of the performance of the Aerosol-cci algorithms and, for comparison, MODIS Terra. The plots show the difference between the satellite-derived AOD and the AERONET AOD, as explained in the text, the colour scales are given to the right of each map.

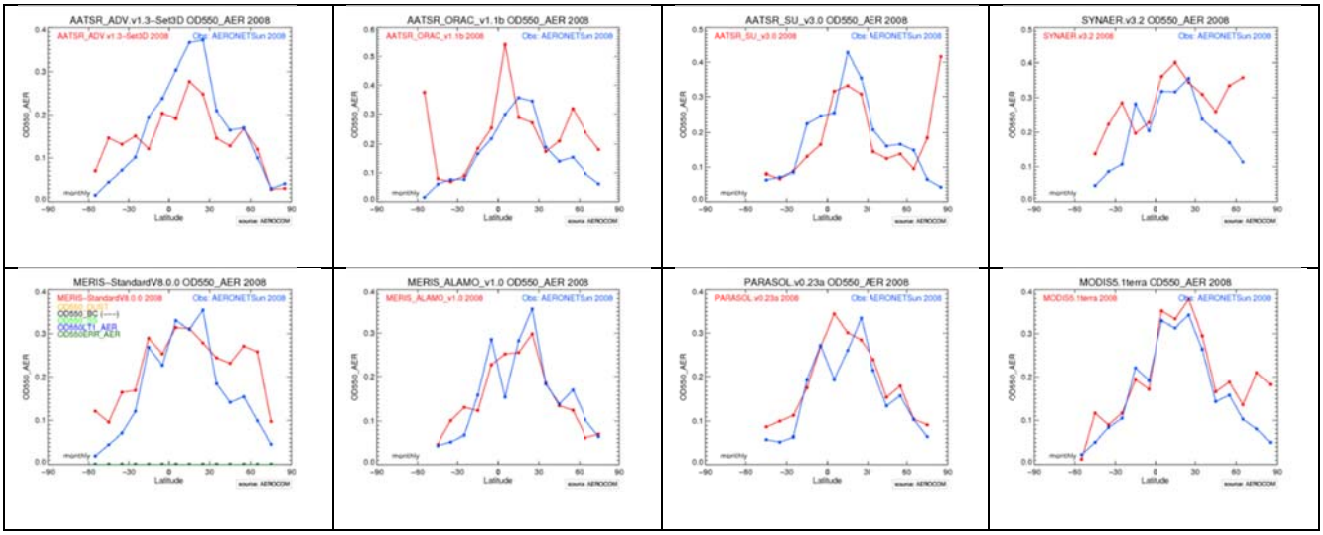

Figure 13. Zonal mean AOD for the 4 months in 2008 for the Aerosol-cci algorithms and MODIS Terra (all in blue) and AERONET (in red).